ATTACHMENT TO:

Microbial Commercialization Activity Notice:

Modified *Saccharomyces cerevisiae* Material submitted per 40 C.F.R. § 725.155

to the
U.S. Environmental Protection Agency
Office of Pollution Prevention and Toxic Chemical Control Division
New Chemicals Notice Management Branch

TS Number: YM95NC

Submitted by: Tara Eskandari, M.S.
Deputy Chief, Biosystems and Biomaterials Division
Material Measurement Laboratory
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899-8543

Submitted to: TSCA Document Processing Center (7407)
Room L-100
Office of Pollution Prevention and Toxics
U.S. Environmental Protection Agency
401 M St., SW.
Washington, D.C. 20460

This attachment contains all the documents referenced in the main submission.

Table of Contents

1. Mortimer RK, Johnston JR. Genealogy of principal strains of the yeast genetic stock center. Genetics. 1986, 113:35-45 1
2. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast. 1998, 14(2):115-132 10
3. Dharmacon GE Healthcare #YSC1062 product insert, 2014 28
4. ATCC #200901 product insert, 2013 29
5. See https://www.ncbi.nlm.nih.gov/sra/SRP150091 for Ion Torrent sequencing data (not included in this attachment)
6. Freel KC, Friedrich A, Schacherer J. Mitochondrial genome evolution in yeasts: an all-encompassing view. 2015, FEMS Yeast Research, 15:fov023 31
7. Bult CJ et al. Complete Genome Sequence of the Methaogenic Archaeon, *Methanococcus jannaschii*. Science. 1996, 273(5278):1058-1073 40
8. FASTA file of the inserted construct as determined using whole genome sequencing 52
9. Results from BLAST of ERCC00095 (GenBank Accession KC702204), performed on 06/08/2018 53
10. Nikata T, Sakai Y, Shibat K, Kato U, Kuroda A, Ohtake H. Molecular analysis of the phosphate-specific transport (pst) operon of Pseudomonas aeruginosa. Molecular and General Genetics. 1996, 250(6):692-698 62
11. Oshima Y. The phosphatase system in *Saccharomyces cerevisiae*. Genes and Genetic Systems. 1997, 72(6):323-334 69
12. See https://www.ncbi.nlm.nih.gov/sra/SRP150091 for Illumina MiSeq sequencing data (not included in this attachment)
13. Olson ND, et al. PEPR: pipelines for evaluating prokaryotic references. Anal Bioanal Chem. 2016, 408:2975-2983 81
14. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics. 2016, 17:333-351 90
15. Agarwal AK, et al. Genome-wide expression of profiling of the response to polyene, pyrimidine, azole, and echinocandin antifungal agents in Saccharomyces cerevisiae. J Biol Chem. 2003, 278(37)34998-35015 109
16. Wang QM, Liu WQ, Liti Gianni, Wang SA, Bai FY. Surprisingly diverged populations of Saccharomyces cerevisiae in natural environments remote from human activity. Molecular Ecology. 2012, 21:5404-5417 128

17. NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules, Department of Health and Human Services, April 2016 156

18. EPA. "*Saccharomyces cerevisiae* Final Risk Assessment," 1997 285

19. Cordero-Bueso G, Arroyo T, Serrano A, Valero, E. Remanence and survival of commercial yeast in different ecological niches of the vineyard. FEMS Microbiol Ecol. 2011, 77(2):429-437. 298

20. Valero E, Schuller D, Cambon B, Casal M, Dequin, S. Dissemination and survival of commercial wine yeast in the vineyard: A large-scale, three-years study. FEMS Yeast Research. 2005, 5:959-969 307

21. Da Silva SM, Vang LK, Olson ND, Lund SP, Downey AS, Kelman Z, Salit ML, Lin NJ, Morrow JB. Evaluation of microbial qPCR workflows using engineered *Saccharomyces cerevisiae*. Biomol Detect Quant. 2016, 7:27-33 318

22. ASTM E2458-17 Standard Practices for Bulk Sample Collection and Swab Sample Collection of Visible Powders Suspected of Being Biological Agents and Toxins from Nonporous Surfaces 325

23. CDC/NIOSH. Surface sampling procedures for *Bacillus anthracis* spores from smooth, non-porous surfaces, revised April 26, 2012..... 341

24. U.S. Department of Health and Human Services. Biosafety in Microbiological and Biomedical Laboratories (BMBL) 5th Edition. 2009. HHS Publication No. (CDC) 21-1112. https://www.cdc.gov/biosafety/publications/bmbl5/bmbl.pdf (not included in this attachment)

25. Liu H, Styles CA, Fink GR. *Saccharomyces cerevisiae* S288C has a mutation in *FLO8*, a gene required for filamentous growth. Genetics. 1996, 144:967-978..... 351

26. EPA. "*Saccharomyces cerevisiae* Final Risk Assessment," 1997 363

27. Microbiologics Inc. "Safety Data Sheet on Microbiologics Lyophilized or Dried Biological Material Preparations," SDS.244.ENG Rev E, 9/21/2016 376

Copyright © 1986 by the Genetics Society of America

GENEALOGY OF PRINCIPAL STRAINS OF THE YEAST GENETIC STOCK CENTER

ROBERT K. MORTIMER* AND JOHN R. JOHNSTON†

*Department of Biophysics and Medical Physics, University of California, Berkeley, California 94720, and †Department of Bioscience and Biotechnology, University of Strathclyde, Glasgow, G1 1XW, Scotland

Manuscript received August 23, 1985
Revised copy accepted December 26, 1985

ABSTRACT

We have constructed a genealogy of strain S288C, from which many of the mutant and segregant strains currently used in studies on the genetics and molecular biology of *Saccharomyces cerevisiae* have been derived. We have determined that its six progenitor strains were EM93, EM126, NRRL YB-210 and the three baking strains Yeast Foam, FLD and LK. We have estimated that approximately 88% of the gene pool of S288C is contributed by strain EM93. The principal ancestral genotypes were those of segregant strains EM93-1C and EM93-3B, initially distributed by C. C. LINDEGREN to several laboratories. We have analyzed an isolate of a lyophilized culture of strain EM93 and determined its genotype as *MATa/MATα SUC2/SUC2 GAL2/gal2 MAL/MAL mel/mel CUP1/cup1 FLO1/flo1*. Strain EM93 is therefore the probable origin of genes *SUC2*, *gal2*, *CUP1* and *flo1* of S288C. We give details of the current availability of several of the progenitor strains and propose that this genealogy should be of assistance in elucidating the origins of several types of genetic and molecular heterogeneities in Saccharomyces.

THE recent explosion of interest in the genetics and molecular biology of the yeast *Saccharomyces cerevisiae* has resulted in a rapid expansion of our knowledge of this organism. Several lines of evidence have recently shown various heterogeneities among strains in, for example, the distribution of restriction sites (OLSON, LOUGHNEY and HALL 1979), in position and numbers of transposable elements (EIBEL *et al.* 1980) and in chromosome polymorphisms (CARLE and OLSON 1985). To assess the relative contributions to these heterogeneities from recent events, as opposed to differences introduced in the development of the initial breeding stocks, we have developed a pedigree of some of the most common laboratory strains.

During the 1930s and 1940s O. WINGE of the Carlsberg Laboratories in Denmark pioneered research on the genetics of yeast. Although his scientific career extended back to 1907, he started work on yeast in 1933, using strains that had been stored in 10% sucrose many years earlier (as early as 1886) by the microbiologist EMIL HANSEN (WINGE and HJORT 1935). Along with O. LAUSTEN and C. ROBERTS, WINGE elegantly elucidated the life cycle of several

yeasts and began the genetic analyses of sucrose and maltose fermentation (WINGE and ROBERTS 1952). WINGE and ROBERTS (1949) characterized the first gene for homothallism (symbolized *D* for diploidization and now designated *HO*). CARL LINDEGREN began research on the genetics of Saccharomyces in the 1940s following his study of the genetics and life cycle of *Neurospora crassa*, upon which BEADLE and TATUM later based their pioneering studies in microbial genetics. LINDEGREN used yeast strains from a limited number of sources in the United States and, in contrast to the breeding strains used by WINGE, developed strains that were heterothallic, thus greatly facilitating genetic research in yeast. We have attempted to review these earlier studies in our effort to develop a pedigree of many of the current breeding stocks of *S. cerevisiae*, concentrating primarily on the studies of LINDEGREN and later workers. Part of the studies described here has appeared in earlier abstracts (JOHNSTON and MORTIMER 1984; MORTIMER and JOHNSTON 1984).

ANALYSIS AND RESULTS

CARL LINDEGREN (1949) published a book entitled *The Yeast Cell: Its Genetics and Cytology*, in which he presented much of his early work on yeast. Included in this book is a pedigree of the breeding stocks that he and GERTRUDE LINDEGREN used in many of their early studies. Part of this pedigree, showing in particular the ancestry of strains 1426 and 1428 (LINDEGREN numbers) and including some information given elsewhere in the book, is redrawn as Figure 1. The equivalence of strains that were designated by LINDEGREN as 4 and 93-1C (recommended designation EM93-1C) is assumed, but seems likely from the information recorded. Strain FLD is listed by LINDEGREN as a commercial baking yeast. Strains EM93 and EM126 were isolated by EMIL MRAK of the University of California, Davis, in 1938 and 1939, respectively, from rotting figs near Merced, California (H. PHAFF and M. MIRANDA, personal communication). Whether these strains were originally part of the figs' flora or were spoilage organisms originating as commercial baking and/or brewing yeasts is unknown. Strain NRRL-210 was isolated from rotting bananas from Costa Rica in 1942 (C. KURTZMAN, personal communication) and was originally classified as *S. microellipsoideus*. It has since been identified as *S. cereviseae* (D. YARROW, personal communication) or *S. uvarum* (*S. bayanus*) (C. KURTZMAN, personal communication).

The LINDEGRENS also made a major contribution to yeast genetics by distributing haploid breeding strains to several other early yeast geneticists; namely, POMPER and BURKHOLDER (1949); REAUME and TATUM (1949); ROMAN, HAWTHORNE and DOUGLAS (1951); and ZIRKLE and TOBIAS (1949). Our investigations have revealed that in every case those haploids were derived from strain EM93 and, in particular, were most probably the meiotic segregant strains EM93-1C (α) and EM93-3B (**a**). POMPER and BURKHOLDER (1949) simply identified these latter strains as "α" and "**a**" (strains 63 and 62; S. POMPER, personal communication). ROMAN, HAWTHORNE and DOUGLAS (1951) designated EM93-1C as "clone 1" (H. ROMAN, personal communication) and as strain 287 (HAWTHORNE 1956) and ZIRKLE and TOBIAS (1949) used both

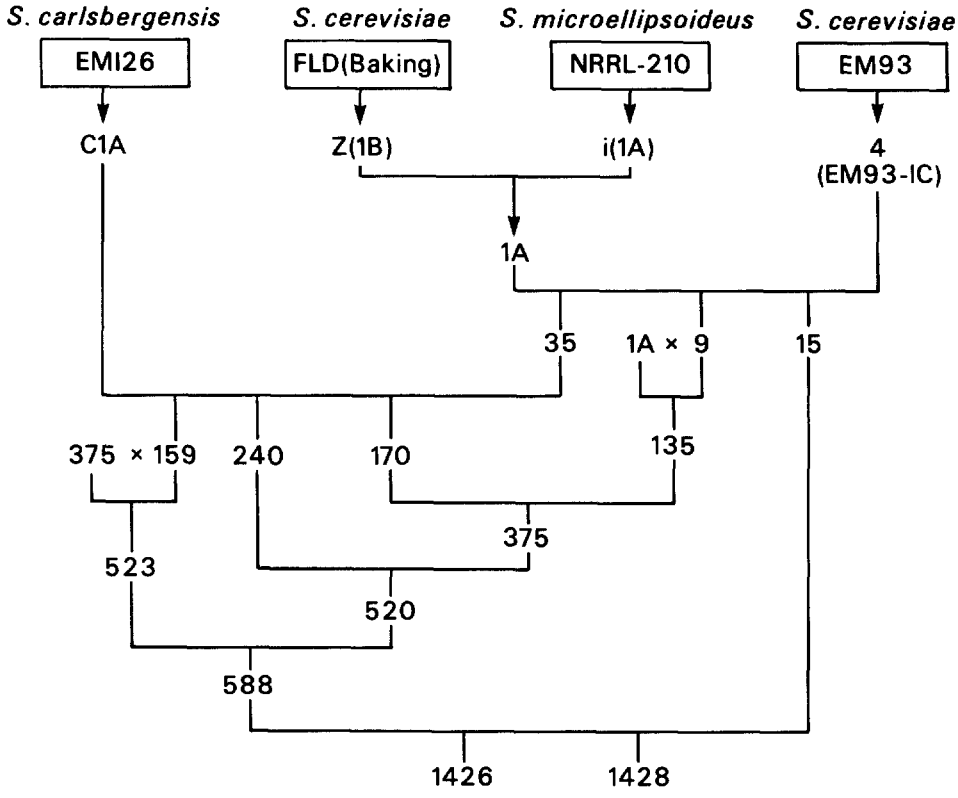


FIGURE 1.—Pedigree of the LINDEGREN strains 1426 and 1428. The diagram is essentially a reconstruction from LINDEGREN (1949). We have concluded that the LINDEGREN strains designated 4 and EM93-1C (distributed to other laboratories) are equivalent. Strain NRRL-210, originally classified as *S. microellipsoideus*, is now classified under either *S. cereviseae* (D. YARROW, personal communication) or *S. uvarum* (C. KURTZMAN, personal communication). It has been proposed that *S. carlsbergensis* be included in *S. uvarum* and that *S. uvarum* be included in *S. cerevisiae* (YARROW 1984). However, it has also been recently proposed that *S. cerevisiae*, *S. bayanus* (including *S. uvarum*) and possibly also *S. carlsbergensis* constitute different species (MARTINI and KURTZMAN 1986).

strains EM93 (SC6) and EM93-1C (SC7). (An error in denoting strain EM93-3B as 93-3P was perpetuated over a number of years.) One of us (R.K.M.) began research in yeast genetics in the TOBIAS laboratory, Berkeley, using strain EM93-1C (SC7) together with some strains obtained from S. E. REAUME (in the TATUM laboratory, Stanford). These additional strains were Y02022 and Y02587, and from information originally obtained from REAUME and TATUM, and more recently from POMPER, we have determined that these strains also derive entirely from strain EM93.

The strain S288C was isolated through genetic crosses by one of us (R.K.M.) as a parental strain to be used for isolation of biochemical mutants. Conditions established for this strain were that it be nonclumpy (nonflocculent)—*i.e.*, dispersed as single cells in liquid culture—and that it have a minimal number of nutritional requirements. In fact, S288C requires only biotin, a nitrogen

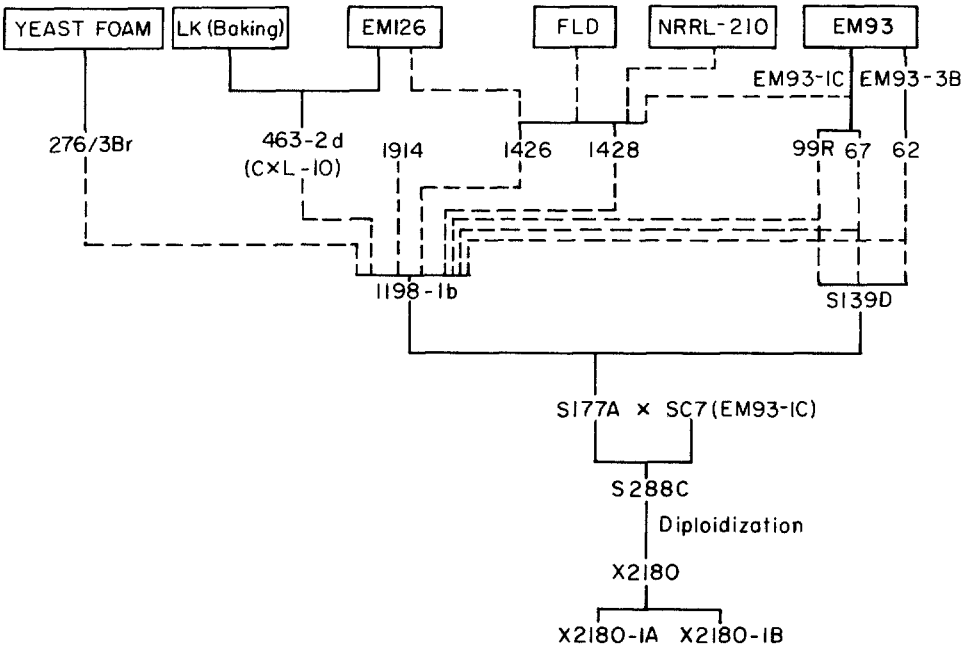


FIGURE 2.—Progenitor strains of S288C, X2180-1A and X2180-1B. In addition to the four strains, EM126, FLD, NRRL-210 and EM93 (Figure 1), from which the LINDEGREN strains 1426 and 1428 were derived, are strains Yeast Foam and LK, both baking yeasts. Solid lines show direct crosses; broken lines indicate that there were intervening crosses. The origins of laboratory strains can be allocated as follows: MRAK: EM93 and EM126; WICKERHAM: NRRL-210; LINDEGREN: EM93-1C, EM93-3B, 1426, 1428, C × L-10; POMPER: 62 and 67; REAUME: 99R; EPHRUSSI: 276/3Br; ROMAN and DOUGLAS: 463-2d; RAUT: 1914; HAWTHORNE: 1198-1b; MORTIMER: S139D, S177A, S288C.

source, glucose and an asssortment of salts and trace elements. The ancestry, with the omission of many intervening crosses, of strain S288C is shown in Figure 2. The immediate parents were strains EM93-1C (SC7) and S177A. One parent, S139D, of the latter strain was derived (by R.K.M.) entirely from the LINDEGREN strains EM93-1C and EM93-3B via strains 99R (REAUME and TATUM), 67 and 62 (POMPER). Strain 67 (POMPER and BURKHOLDER 1949) was obtained from REAUME, but as far as we can determine was strain 93-1C. The other parent of S177A, *1198-1b*, resulted from a large number of crosses made by D. C. HAWTHORNE in order to study genetic segregation for the fermentation of sucrose, maltose, α-methylglucoside and galactose (HAWTHORNE 1956). These crosses involved (D. HAWTHORNE, personal communication) the progenitor strains of POMPER, REAUME and TATUM, LINDEGREN's 1426 and 1428 (Figure 1), strain 463-2d (ROMAN and DOUGLAS) derived from the LINDEGREN diploid strain C × L-10 (LINDEGREN 1949) strain 276/3Br (EPHRUSSI, HOTTINGUER and TAVLITZKI 1949) and the strain 1914. This final strain was received from C. RAUT, and although we have been unable to trace its specific pedigree, it appears likely (RAUT 1953) that it was derived from the same ancestral strains as were LINDEGREN strains 1426 and 1428 (Figure 1). The

other seven progenitor strains of 1198-1b were descendants of the same four original strains shown in Figure 1 and, in addition, the baking yeasts LK (LINDEGREN 1949), where L apparently symbolizes Lallemand (C. KURTZMAN, personal communication) and "Yeast Foam" (EPHRUSSI, HOTTINGUER and TAVLITZKI 1949). This latter yeast is heterothallic and originated as a bakers' yeast from Northwestern Yeast Company, Chicago (WINGE and ROBERTS 1984). Figure 2 also shows the derivation of strain X2180, by self-diploidisation of S288C, and the isogenic *MAT***a** and *MATα* haploid strains, X2180-1A and X2180-1B, derived from this diploid.

Strain S288C or an *ade2-1* mutant of it, DV147, were used to isolate the original alleles of most of the nutritional mutants currently used in yeast genetics research (HAWTHORNE and MORTIMER 1960; MORTIMER and HAWTHORNE 1966). For example, all, or nearly all, of the mutants in the arginine, histidine, tryptophan, uracil, threonine-methionine, leucine, lysine and serine pathways were isolated from these strains. Analysis of the pedigrees illustrated in Figures 1 and 2 shows that strain EM93 contributes approximately 90% of the genome of S288C. Strains EM126, NRRL-210, FLD, LK and Yeast Foam contribute the balance.

Mutants of S288C (*MATα*) were brought together in various combinations to provide many of the strains for the Yeast Genetic Stock Center, which has been in operation formally since 1970 and informally since 1960. Backcrosses were made using *MAT***a** strains Y02587, 1198-1b and 4209A. The first of these strains is derived entirely from strain EM93. We have estimated that the contribution of strain EM93 to strain 1198-1b is approximately 30%. Strain 4209A was obtained from H. ROMAN and was derived from the same progenitor strains as shown in Figure 2, but with an increased contribution of Yeast Foam through the mutant 276/3Br (*ade2*).

In the course of our investigations we were fortunate to find samples of the diploid strain EM93 (SC6) that had been lyophilized in the TOBIAS laboratory, Berkeley, in 1951. Several isolates of these samples were cultured and classified. They all sporulated well, but varied in their level of ascospore viability. In particular, a few apparently segregated for a recessive lethal mutation, perhaps resulting from the lyophilization process. Isolate 10 yielded the highest recovery of viable tetrads, and we selected this strain for tetrad analysis. LINDEGREN (1949) reported heterozygosity of strain EM93 for mating type, galactose fermentation and the characteristic "dispersed" *vs.* "flaky" growth. Tetrad analysis of our isolate of strain EM93 confirms these segregations. We have shown by linkage and complementation tests that the genes segregating are *MAT*, *gal2* and *FLO1*; in addition, strain EM93 is heterozygous for *CUP1*/*cup1* (R. CONTOPOULOU, unpublished results; DE ZOYSA 1985).

DISCUSSION

In elucidating the genealogy of strain S288C, and its derivative strains X2180-1A and X2180-1B, we have determined that strain EM93, through its segregant strains EM93-1C and EM93-3B, is the principal progenitor strain. We have calculated the contribution of these segregants of strain EM93 to the

gene pool of strain S288C as approximately 88%. We have also concluded that, in turn, strain S288C (or strains closely related to it) is a major source of the gene pool of most laboratory strains. Our tetrad analysis of an isolate of strain EM93, lyophilized in Berkeley in 1951, showing it to be heterozygous for *gal2* and *CUP1* and homozygous for *SUC2* is evidence that strain EM93 is the origin of these genes which are present in strain S288C.

The genotype of strain 1198-1b was *mal mel gal2 suc*, but it is likely that the source of *gal2* in this strain was also strain EM93. The gene *gal1* originated in the LINDEGREN strain 1426 (HAWTHORNE 1956). We have determined that the source of the *mal* gene in strains 1198-1b and S288C is progenitor strain NRRL-210 (C. KURTZMAN, personal communication).

Fortunately, several progenitor strains and strains derived from them are still available in different stock collections. Strain EM93 is in culture at the University of California, Davis (M. MIRANDA and H. PHAFF, personal communication) and at the Food Research Institute, Norwich, England [National Collection of Yeast Cultures (NCYC), strain NCYC 292]. The latter strain was deposited by S. JACKSON, who received the strain from C. LINDEGREN in 1951 (NCYC Catalog, June 1981). Strain EM126 is also in culture at the University of California, Davis. The isolate of strain EM93, on which we have performed tetrad analysis, is available from the Yeast Genetic Stock Center, Berkeley, and the University of Strathclyde, Glasgow. Although the strains EM93-1C and EM93-3B are apparently no longer available, mutant strains 99R (REAUME and TATUM 1949) and 200, which we believe are derivatives of the former strains, respectively, are also available from the Yeast Genetic Stock Center (R. K. MORTIMER and C. R. CONTOPOULOU, YGSC Catalog, January 1984). Of the remaining four progenitor strains of S288C, two are available. Strain NRRL YB-210, also designated as NRRL Y-1350, is in culture at the Northern Regional Research Center, Peoria, Illinois, and (as strain CBS 6333) at the Centraalbureau voor Schimmelcultures, Delft, The Netherlands (C. KURTZMAN, D. YARROW, personal communications). The strain "Yeast Foam" (or "American Yeast Foam") is listed as culture NCYC 232 (NCYC Catalog, June, 1981) and as culture CBS 1428 (D. YARROW, personal communication).

In addition to strain S288C, several other strains have played a key role in the development of yeast genetics. These include FL100 (LACROUTE 1968, Σ 1278b (GRENSON *et al.* 1966) and A364A (HARTWELL 1967). The relationships of these strains to strain EM93 and S288C remain to be fully determined, although our preliminary analyses indicate that FL100 and A364A are closely related to S288C. Strain Σ 1278b is a segregant of a cross between a segregant of Yeast Foam and strain 1422-11D (HAWTHORNE), which we believe was derived from the same progenitor strains as shown in Figure 2. Another bakers' yeast, referred to as Boulangerie II, was the source of the nuclear petite mutant, *pet1* (CHEN, EPHRUSSI and HOTTINGUER 1950).

A major frustration in our analysis was the practice of several investigators to change the designations of strains sent to them. For example, strain EM93-1C has been assigned several synonyms by different investigators, including LINDEGREN (Table 1). Table 1 presents a record of synonyms and culture

TABLE 1

Recommended and other strain designations of several progenitor strains of YGSC

Recommended strain designation	Saccharomyces species	Original isolate by (from)	Synonymous strain designation	Reference	Culture collection number(s)
EM93	*S. cerevisiae*	E. MRAK	MRAK 93	LINDEGREN (1949)	NCYC 292
			93; SC6	ZIRKLE and TOBIAS (1949)	
EM93-1C	*S. cerevisiae*	C. LINDEGREN	93-1C; Wild type	REAUME and TATUM (1949)	
			4 (assumed)	LINDEGREN (1949)	
			α	POMPER and BURKHOLDER (1949)	
			63	S. POMPER (personal communication)	
			Clone 1	ROMAN, HAWTHORNE and DOUGLAS (1951)	
			287	HAWTHORNE (1956)	
			93-1C; SC7	ZIRKLE and TOBIAS (1949)	
EM93-3B	*S. cerevisiae*	C. LINDEGREN	**a**	POMPER and BURKHOLDER (1949)	
			62	POMPER (1952)	
			93-3B; SC8	ZIRKLE and TOBIAS (1949)	
EM126	*S. carlsbergensis*[a]	E. MRAK	MRAK 126	LINDEGREN (1949)	
NRRL Y-1350[b]	*S. microellipsoideus*[a]	L. WICKERHAM	NRRL YB-210	C. KURTZMAN (personal communication)	CBS 6333 (formerly CBS 428)
			NRRL-210		
			NRRL-B210	LINDEGREN (1949)	
Yeast foam	*S. cerevisiae*	Northwestern Yeast Company, Chicago	American Yeast Foam	WINGE and ROBERTS (1948)	NCYC 232 CBS 1428 (formerly CBS 237)

[a]For comments on current taxonomy, see the caption to Figure 1.
[b]Recommended current collection number (C. KURTZMAN, personal communication).

YM95NC Attachment -- Page 7

collection numbers, along with recommended designations for several of the progenitor strains of S288C. We would recommend most strongly that investigators, if they choose to use new strain designations, also present the previous or original strain designations as synonyms in publications or communications.

We hope that the genealogy described in this paper will be of value in the analyses of various strain heterogeneities. For example, structural heterology of a region of chromosome *III* has been shown between the species *S. cerevisiae* and *S. carlsbergensis* by the techniques of classical and molecular genetics (NILSSON-TILLGREN *et al.* 1981; HOLMBERG, 1982). DNA sequence polymorphisms of the *RDN1* gene (PETES 1979), on chromosome *XII*, and the *HIS4* gene (chromosome *III*) among various yeast strains, including those of the species *S. cereviseae*, *S. uvarum* and *S. carlsbergensis* have also been described (PEDERSEN 1983). In addition, polymorphisms in chromosome size (CARLE and OLSON 1985), restriction sites (OLSON, LOUGHNEY and HALL 1979) and *TY1* number and chromosome location (EIBEL *et al.* 1980) have been found.

This work was supported by Public Health Service grant GM-30990 from the National Institutes of Health and a grant from the Office of Health and Environmental Research of the United States Department of Energy under contract DE-AC03-76SF00098. We thank DONALD HAWTHORNE, CLETUS KURTZMAN, MARY MIRANDA, HERMAN PHAFF, SEYMOUR POMPER, HERSCHEL ROMAN and DAVID YARROW for providing information essential to our study.

LITERATURE CITED

CARLE, G. F. and M. V. OLSON, 1985 An electrophoretic karyotype for yeast. Proc. Natl. Acad. Sci. USA **82:** 3756–3760.

CHEN, S., B. EPHRUSSI and H. HOTTINGUER, 1950 Nature genetique des mutants a deficience respiratoire de la souche B-11 de la levure de boulangerie. Heredity **4:** 337–351.

DE ZOYSA, P., 1985 Genetic analysis and mapping of genes for flocculence in *Saccharomyces cerevisiae*. M.Sc. Thesis, University of Strathclyde, Glasgow, Scotland.

EIBEL, H., J. GAFNER, A. STOTZ and P. PHILIPPSEN, 1980 Characterization of the yeast mobile element TY1. Cold Spring Harbor Symp. Quant. Biol. **45:** 609–617.

EPHRUSSI, B., H. HOTTINGUER and J. TAVLITZKI, 1949 Action de l'acriflavine sur les levures. Ann. Inst. Pasteur (Paris) **76:** 419–450.

GRENSON, M., M. MOUSSET, J. M. WIAME and J. BECHET, 1966 Multiplicity of the amino acid permeases in *Saccharomyces cereviseae* 1. Evidence for a specific arginine-transporting system. Biochim. Biophys. Acta **127:** 325–338.

HARTWELL, L. H., 1967 Macromolecular synthesis in temperature-sensitive mutants of yeast. J. Bacteriol. **93:** 1662–1670.

HAWTHORNE, D. C., 1956 The genetics of galactose fermentation in Saccharomyces hybrids. C. R. Trav. Lab. Carlsberg, Ser. Physiol. **26:** 149–160.

HAWTHORNE, D. C. and R. K. MORTIMER, 1960 Chromosome mapping in Saccharomyces:centromere-linked genes. Genetics **45:** 1085–1110.

HOLMBERG, S., 1982 Genetic differences between *Saccharomyces carlsbergensis* and *Saccharomyces cerevisiae* 2. Restriction endonuclease analysis of genes in chromosome *III*. Carlsberg Res. Commun. **47:** 233–244.

JOHNSTON, J. R. and R. K. MORTIMER, 1984 Pedigree of the Yeast Genetics Stock Center "wild type" strains S288c and X2180 of *S. cerevisiae*. Heredity **52:** 459.

LACROUTE, F., 1968 Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. J. Bacteriol. **95:** 824–832.

LINDEGREN, C. C., 1949 *The Yeast Cell: Its Genetics and Cytology*, Chap. 18, p. 2. Education Publishers Inc., St. Louis.

MARTINI, A. V. and C. P. KURTZMAN, 1986 Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces sensu stricto*. Int. J. Syst. Bacteriol. In press.

MORTIMER, R. K. and D. C. HAWTHORNE, 1966 Genetic mapping in Saccharomyces. Genetics **53:** 165–173.

MORTIMER, R. K. and J. R. JOHNSTON, 1984 Genealogy of YGSC strains S288c and X2180 (Abstr.). p. 301. 12th International Conference on Yeast Genetics and Molecular Biology, Edinburgh, Scotland.

NILSSON-TILLGREN, T., C. GJERMANSEN, M. C. KIELLAND-BRANDT, J. G. L. PETERSEN and S. HOLMBERG, 1981 Genetic differences between *Saccharomyces carlsbergensis* and *Saccharomyces cerevisiae*. Analysis by single chromosome transfer. Carlsberg Res. Commun. **46:** 65–76.

OLSON, M. V., K. LOUGHNEY and B. D. HALL, 1979 Identification of the yeast DNA sequences that correspond to specific tyrosine-inserting nonsense suppressor loci. J. Mol. Biol. **132:** 387–410.

PEDERSEN, M. B., 1983 DNA sequence polymorphisms in the genus *Saccharomyces* 1. Comparison of the *HIS4* and ribosomal RNA genes in lager strains, ale strains and various species. Carlsberg Res. Commun. **48:** 485–504.

PETES, T. D., 1979 Meiotic mapping of yeast ribosomal DNA on chromosome *XII*. J. Bacteriol. **138:** 185–192.

POMPER, S., 1952 Purine-requiring and pyrimidine-requiring mutants of *Saccharomyces cerevisiae*. J. Bacteriol. **63:** 707–713.

POMPER, S. and P. R. BURKHOLDER, 1949 Studies on the biochemical genetics of yeast. Proc. Natl. Acad. Sci. USA **35:** 456–464.

RAUT, C., 1953 A cytochrome deficient mutant of *Saccharomyces cerevisiae*. Exp. Cell Res. **4:** 295–305.

REAUME, S. E. and E. L. TATUM, 1949 Spontaneous and nitrogen mustard-induced nutritional deficiencies in *Saccharomyces cerevisiae*. Arch. Biochem. **22:** 331–338.

ROMAN, H., D. HAWTHORNE and H. DOUGLAS, 1951 Polyploidy in yeast and its bearing on the occurrence of irregular genetic ratios. Proc. Natl. Acad. Sci. USA **37:** 79–84.

WINGE, O. and A. HJORT, 1935 On some Saccharomycetes and other fungi still alive in the pure cultures of Emil Chr. Hansen and Alb. Klocker. C. R. Trav. Lab. Carlsberg, Ser. Physiol. **21:** 51–58.

WINGE, O. and C. ROBERTS, 1948 Inheritance of enzymatic characters in yeast, and the phenomenon of long-term adaptation. C. R. Trav. Lab. Carlsberg, Ser. Physiol. **24:** 263–315.

WINGE, O. and C. ROBERTS, 1949 A gene for diploidization in yeasts. C. R. Trav. Lab. Carlsberg, Ser. Physiol. **24:** 341–346.

WINGE, O. and C. ROBERTS, 1952 The relation between the polymeric genes for maltose, raffinose, and sucrose fermentation in yeasts. C. R. Trav. Lab. Carlsberg, Ser. Physiol. **25:** 141–172.

YARROW, D., 1984 *Saccharomyces* Meyen ex Reess. pp. 379–395. In: *The Yeasts: A Taxonomic Study*, Edited by N. J. W. KREGER-VAN RIJ. Elsevier Nederland BV, Amsterdam.

ZIRKLE, R. E. and C. A. TOBIAS, 1949 Effects of ploidy and linear energy transfer on radiobiological survival curves. Arch. Biochem. Biophys. **47:** 282–306.

Communicating editor: I. HERSKOWITZ

YEAST VOL. **14:** 115–132 (1998)

Designer Deletion Strains derived from *Saccharomyces cerevisiae* S288C: a Useful set of Strains and Plasmids for PCR-mediated Gene Disruption and Other Applications

CARRIE BAKER BRACHMANN, ADRIAN DAVIES, GREGORY J. COST, EMERITA CAPUTO, JOACHIM LI†, PHILIP HIETER AND JEF D. BOEKE*

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, U.S.A.

Received 9 June 1997; accepted 18 June 1997

A set of yeast strains based on *Saccharomyces cerevisiae* S288C in which commonly used selectable marker genes are deleted by design based on the yeast genome sequence has been constructed and analysed. These strains minimize or eliminate the homology to the corresponding marker genes in commonly used vectors without significantly affecting adjacent gene expression. Because the homology between commonly used auxotrophic marker gene segments and genomic sequences has been largely or completely abolished, these strains will also reduce plasmid integration events which can interfere with a wide variety of molecular genetic applications. We also report the construction of new members of the pRS400 series of vectors, containing the *kanMX*, *ADE2* and *MET15* genes. © 1998 John Wiley & Sons, Ltd.

Yeast 14: 115–132, 1998.

KEY WORDS — *Saccharomyces cerevisiae*; yeast; gene disruption; S288C; bacteria-yeast shuttle vectors; auxotrophic markers

INTRODUCTION

Deletion mutations are valuable because they never revert. This is particularly true with regard to DNA transformation because there is no background of revertants to sort through; any organism that acquires the wild-type phenotype is by definition a transformant. A problem with most deletion alleles is that they rarely precisely affect a single gene; there may be 'collateral' effects on one or more adjacent genes that can confound genetic analysis. The availability of whole-genome sequence information together with modern recombinant technologies like the polymerase chain reaction (PCR) allows the precise design of deletions that are much less likely to have such problems. We have designed a strategy exploiting these technologies to create a set of precise deletion alleles of commonly used auxotrophic markers in the yeast *Saccharomyces cerevisiae*. As additional complete genomes are sequenced, this strategy can be used to create high quality deletion alleles in other organisms.

Traditionally the yeast *S. cerevisiae* has served as one of the best genetically tractable eukaryotic systems. Researchers have for years been able to disrupt genes in yeast in order to study their mutant phenotypes; however, the techniques used for gene disruption have had limitations. First, the gene to be disrupted had to be cloned. Second, because available restriction sites were used to delete a portion of the gene and insert a selectable marker, 5′ or 3′ fragments of the gene were often

*Correspondence to: J. D. Boeke, Hunterian 617, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, U.S.A. Tel: (+1) 410 955 2481; fax: (+1) 410 614 2987; e-mail: jef.boeke@qmail.bs.jhu.edu.
†Present address: Department of Biochemistry, University of California, San Francisco, CA, U.S.A.
Contract grant sponsor: NIH.

CCC 0749–503X/98/020115–18 $17.50
© 1998 John Wiley & Sons, Ltd.

left in the genome, raising potential doubt about the null phenotype. A method has recently been developed that allows complete deletion of the gene of interest and its replacement with an auxotrophic marker gene (Baudin *et al.*, 1993; Lorenz *et al.*, 1995). This procedure, called PCR-mediated gene disruption, requires only knowledge of the genomic sequence of the gene of interest, allowing for precise gene deletion prior to its cloning and immediate study of the null phenotype. Along with the availability of the entire *S. cerevisiae* genomic sequence, the method of PCR-mediated gene disruption has changed the face of yeast genetics, enabling researchers to quickly assess a gene's possible function through evaluation of the null phenotype, and to efficiently tackle the more complex problems inherent in the analysis of multi-gene families.

Like its predecessor method, one-step gene disruption (Rothstein, 1983), PCR-mediated gene disruption relies on the fact that homologous recombination in yeast can be efficiently mediated by linear fragments of DNA. The method requires two ~60 nucleotide (nt) PCR primers; the 20 nts of sequence at the 3′ ends of each primer is specific for the amplification of an auxotrophic marker gene, and the 40 nts of sequence at the 5′ ends is identical to the left and right genomic sequence flanking the gene of interest. Using these targeting primers, the auxotrophic marker gene is amplified and the resulting PCR product contains the marker gene and 40 bp of genomic DNA sequence to the left and right of the gene deletion of interest. This product is then transformed into yeast, and stable transformants are selected on appropriate medium. These transformants include those cells in which two crossover events occurred, one between each end of the 40 bp target sequences of the PCR product and the corresponding genomic DNA sequence, thereby replacing the gene of interest with the auxotrophic marker gene.

A limitation of this method has been that the genomic alleles of many of the most commonly used auxotrophic markers are point mutations, small internal deletions, or Ty1 insertions, and therefore retain large regions of homology to the internal selectable marker segment of the PCR product. This results in a high background of positive transformants in which the marker gene has replaced the auxotrophic mutation by marker locus gene conversion, leaving the gene of interest unaffected. This problem is circumvented by use of the set of 'designer' deletion strains we have made, in which some commonly used auxotrophic marker genes have been completely deleted. These strains eliminate the background of undesired marker gene convertants. Because the homology between commonly used auxotrophic marker gene segments and genomic sequences has been largely or completely abolished, these strains will also reduce plasmid integration events which can interfere with a wide variety of molecular genetic applications. We have also constructed a designer deletion allele of the *MET15* gene, a selectable marker gene which is useful both because it is counterselectable and because it can be used as a colony color marker (Cost and Boeke, 1996).

Additionally, we have increased the repertoire of the pRS plasmids by constructing new versions containing the *kanMX* marker (Wach *et al.*, 1994), *ADE2* and *MET15* markers. Finally, we have designed universal PCR primers for PCR-mediated gene disruption. This strategy allows the amplification of each of the auxotrophic marker genes found in the pRS set of plasmids (Sikorski and Hieter, 1989). Thus, the same primers can be used to amplify any of the auxotrophic marker genes, allowing for much greater flexibility in the design of gene knockout experiments.

MATERIALS AND METHODS

Strains and media

The yeast strains are all directly descended from FY2 (Winston *et al.*, 1995), which is itself a direct descendant of S288C, and are described in Table 1. The genealogy of the strains in this paper is summarized in Figure 1. Media used were as described (Boeke *et al.*, 1984; Rose *et al.*, 1990), or as described in Table 2 for scoring the *met15* marker. Efficient sporulation of diploids required growth as a patch on freshly poured GNA pre-sporulation plates (5% glucose, 3% Difco nutrient broth, 1% Difco yeast extract, 2% agar) for 1 day at 30°C prior to transfer to liquid sporulation medium (1% potassium acetate, 0·005% zinc acetate). Typically, a generous matchhead of cells was resuspended in 2 ml of liquid sporulation medium. Sporulation cultures were incubated on a roller wheel for 1 day at 25°C followed by 3–5 days at 30°C and could be readily dissected. In general, sporulation efficiencies were of the order of 20–30% (pre-growth of diploids on YPD instead of GNA resulted in somewhat lower sporulation efficiencies).

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 11

Table 1. Yeast strains.

Strain	ATCC number[a]	Genotype						
FY2		*MAT*α						*ura3-52*
FY3		*MAT***a**						*ura3-52*
FY4		*MAT***a**						
FY5		*MAT*α						
YCB436		*MAT***a**	*ade2Δ::hisG*	*his3Δ200*	*leu2Δ1*	*lys2Δ202*	*trp1Δ63*	*ura3-52*
BY378		*MAT*α		*his3Δ200*				*ura3-52*
BY379		*MAT***a**		*his3Δ200*				
BY388		*MAT***a**	*ade2Δ::hisG*	*his3Δ200*			*trp1Δ63*	*ura3-52*
BY389		*MAT*α	*ade2Δ::hisG*	*his3Δ200*			*trp1Δ63*	*ura3-52*
BY397		*MAT***a**	*ade2Δ::hisG*	*his3Δ200*			*trp1Δ63*	
BY398		*MAT*α	*ade2Δ::hisG*	*his3Δ200*			*trp1Δ63*	
BY399		*MAT***a**	*ade2Δ::hisG*	*his3Δ200*			*trp1Δ63*	*ura3Δ0*
BY401		*MAT*α	*ade2Δ::hisG*	*his3Δ200*			*trp1Δ63*	*ura3Δ0*
BY404		*MAT***a**	*ade2Δ::hisG*	*his3Δ200*	*leu2Δ0*		*trp1Δ63*	*ura3Δ0*
BY406		*MAT*α	*ade2Δ::hisG*	*his3Δ200*		*lys2Δ0*	*trp1Δ63*	*ura3Δ0*
BY411		*MAT*α	*ade2Δ::hisG*	*his3Δ200*		*lys2Δ0*	*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY413		*MAT*α		*his3Δ200*		*lys2Δ0*	*met15Δ0*		
BY418		*MAT*α	*ade2Δ::hisG*	*his3Δ200*	*leu2Δ0*	*lys2Δ0*	*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4700	200866	*MAT***a**							*ura3Δ0*
BY4702	200867	*MAT*α			*leu2Δ0*	*lys2Δ0*	*met15Δ0*		
BY4704	200868	*MAT***a**	*ade2Δ::hisG*	*his3Δ200*	*leu2Δ0*	*lys2Δ0*	*met15Δ0*	*trp1Δ63*	
BY4705	200869	*MAT*α	*ade2Δ::hisG*	*his3Δ200*	*leu2Δ0*	*lys2Δ0*	*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4706	200870	*MAT***a**					*met15Δ0*		
BY4707	200871	*MAT*α					*met15Δ0*		
BY4709	200872	*MAT*α							*ura3Δ0*
BY4710	200873	*MAT***a**						*trp1Δ63*	
BY4711	200874	*MAT*α						*trp1Δ63*	
BY4712	200875	*MAT***a**			*leu2Δ0*				
BY4713	200876	*MAT*α			*leu2Δ0*				
BY4714	200877	*MAT***a**		*his3Δ200*					
BY4715	200878	*MAT***a**				*lys2Δ0*			
BY4716	200879	*MAT*α				*lys2Δ0*			
BY4717	200880	*MAT***a**	*ade2Δ::hisG*						
BY4718	200881	*MAT***a**					*met15Δ0*		*ura3Δ0*
BY4719	200882	*MAT***a**						*trp1Δ63*	*ura3Δ0*
BY4720	200883	*MAT*α				*lys2Δ0*		*trp1Δ63*	*ura3Δ0*
BY4722	200884	*MAT*α			*leu2Δ0*				*ura3Δ0*
BY4723	200885	*MAT***a**		*his3Δ200*					*ura3Δ0*
BY4724	200886	*MAT***a**				*lys2Δ0*			*ura3Δ0*
BY4725	200887	*MAT***a**	*ade2Δ::hisG*						*ura3Δ0*
BY4726	200888	*MAT*α	*ade2Δ::hisG*						*ura3Δ0*
BY4727	200889	*MAT*α		*his3Δ200*	*leu2Δ0*	*lys2Δ0*	*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4728	200890	*MAT***a**		*his3Δ200*				*trp1Δ63*	*ura3Δ0*
BY4729	200891	*MAT*α		*his3Δ200*				*trp1Δ63*	*ura3Δ0*
BY4730	200892	*MAT***a**			*leu2Δ0*		*met15Δ0*		*ura3Δ0*
BY4731	200893	*MAT*α			*leu2Δ0*		*met15Δ0*		*ura3Δ0*
BY4732	200894	*MAT*α		*his3Δ200*			*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4733	200895	*MAT***a**		*his3Δ200*	*leu2Δ0*		*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4734	200896	*MAT*α		*his3Δ200*	*leu2Δ0*		*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4735	200897	*MAT*α	*ade2Δ::hisG*	*his3Δ200*	*leu2Δ0*		*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4736	200898	*MAT***a**	*ade2Δ::hisG*	*his3Δ200*			*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4737	200899	*MAT*α	*ade2Δ::hisG*	*his3Δ200*			*met15Δ0*	*trp1Δ63*	*ura3Δ0*
BY4738	200900	*MAT*α						*trp1Δ63*	*ura3Δ0*
BY4739	200901	*MAT*α			*leu2Δ0*	*lys2Δ0*			*ura3Δ0*
BY4740	200902	*MAT***a**			*leu2Δ0*	*lys2Δ0*			*ura3Δ0*
BY4741[b]	201388	*MAT***a**		*his3Δ1*	*leu2Δ0*		*met15Δ0*		*ura3Δ0*
BY4742[b]	201389	*MAT*α		*his3Δ1*	*leu2Δ0*	*lys2Δ0*			*ura3Δ0*

Data in this table are maintained electronically at http://www.welch.jhu.edu/~gregory/MET15.html.
[a]ATCC, American Type Culture Collection, Rockville, MD.
[b]Strains selected as parent strains for the international systematic *Saccharomyces cerevisiae* gene disruption project.

© 1998 John Wiley & Sons, Ltd.

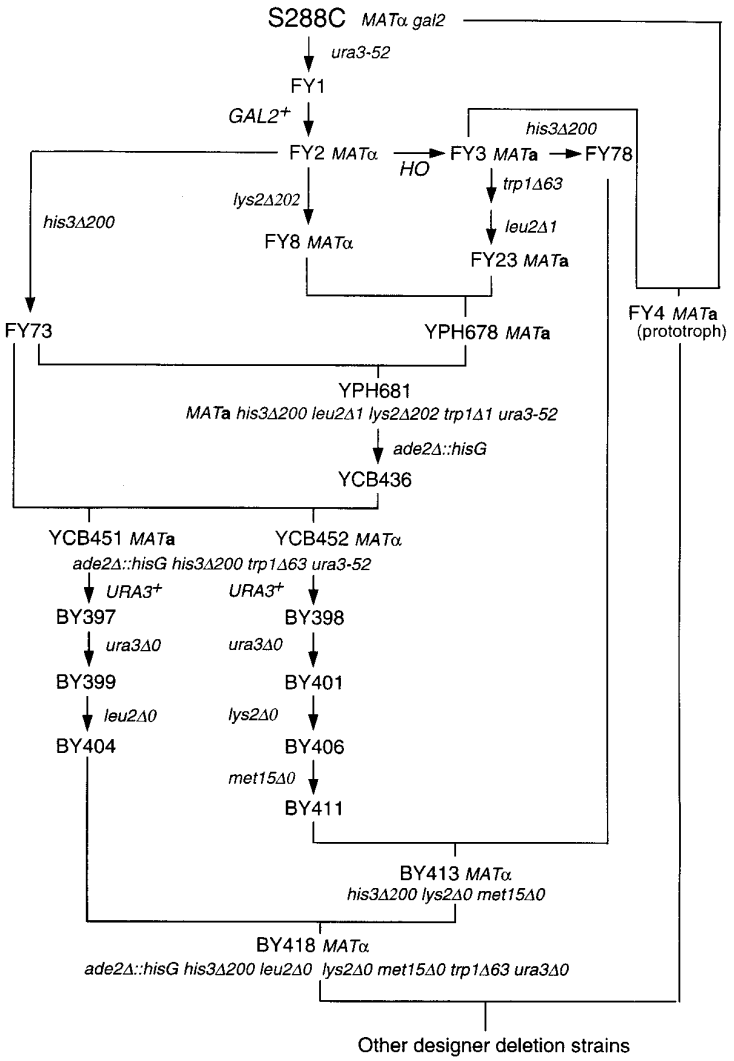


Figure 1. Genealogy of the designer deletion strains. A diagram presenting the lineage of the designer deletion strains and defining them as direct descendants of the original S288C isolate. Lines with arrows indicate that the identified allele was introduced by transformation. Lines connecting two strains define a mating between the two strains, and lines without arrows indicate that a spore from the mating was identified. In those instances where spores were identified, their genotypes are indicated. Further details of the genealogy are presented in Materials and Methods.

PCR of marker flanking sequences

Oligos LEU2-1 (cggggtaccACAGAGTACTTT ATACGTAC—small letters indicate restriction sites added) and LEU2-2 (cgcggatccGAGAACA TTCATGATTAGAGG) were used to amplify a 1010 bp region to the left of *LEU2* and oligos LEU2-3 (cgcggatccCGACACGAAATTACAAA ATG) and LEU2-4 (cggctggagctcCAAGTGTGTC TTGGAAGCCG) were used to amplify a 1220 bp region to the right of *LEU2*. Oligos LYS2-1 (cggggtaccGTACCTTTTTGAACTTCGTC) and LYS2-2 (cgcggatccAGAAGCGGTCAGGAAGA AG AAA) were used to amplify a 917 bp region to the left of *LYS2* and oligos LYS2-3 (cgcggatccTC CATGTACAATAATTAAATATGAATTAGG) and LYS2-4 (cggctggagctcCATCATGCTGCGA

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 13

Table 2. Medium used for scoring *met15*.

Lead plates[a] (per liter)
3 g peptone
5 g yeast extract
200 mg ammonium sulfate
20 g agar
q.s. 800 ml with water
Autoclave
1 g lead nitrate (add 2 ml 0·5 g/ml filter-sterilized stock after cooling above mixture to 55°C)
40 g glucose (add 200 ml 20% stock)

[a]Lead plates are useful because *met15* strains are dark brown to black, whereas wild-type strains are white. To score the Met$^-$ auxotrophic phenotype, use SD plates supplemented with the requirements of the strain. Note that SC − met medium cannot be used unless cysteine is also omitted, as either methionine or cysteine allows growth of *met15* mutants (Ono *et al*., 1991; Cost and Boeke, 1996).

AGAACTA) were used to amplify a 1178 bp region to the right of *LYS2*. Oligos MET15-1 (cggggtaccAAGTTCTCGTCGAATGCTAGG) and MET15-2 (cgcggatccTGCCAACCACCACA G TTCCCC) were used to amplify a 1034 bp region to the left of *MET15* and oligos MET15-3 (cgcggatccTCAGATATAGTCGGATTGCCC) and MET15-4 (cggctggagctcGGTGTTGACACC TTCTCCGC) were used to amplify a 2084 bp region to the right of *MET15*. In all cases, the left fragment included a *Kpn*I site at its 5′ end and a *Bam*HI site at its 3′ end, and the right fragment included a *Bam*HI site at its 5′ end and a *Sac*I site at its 3′ end.

Plasmids

pJL164 contains the *Bgl*II–*Bam*HI genomic fragment containing *URA3* cloned into the *Bam*HI and *Xho*I sites, respectively, of pRS305; the *URA3* ORF has been deleted from this construct by removing a 1·1 kb segment from *Hin*dIII to *Sma*I, leaving a *Sma*I linker at the deletion junction. pAD1 is pRS406 containing the left and right flanking fragments of *LEU2*. The left fragment was first digested with *Kpn*I and *Bam*HI and the right fragment was digested with *Bam*HI and *Sac*I. Both fragments were then purified and ligated into pRS406 linearized with *Kpn*I and *Sac*I. Correct 3 piece ligations were confirmed by restriction digestion. Plasmids pAD2 (pRS406 containing the left and right flanking fragments of *LYS2*) and pAD4 (pRS406 containing the left and right flanking fragments of *MET15*) were constructed in an identical manner. These plasmids are available from ATCC (Table 3).

New pRS plasmids

In these plasmids, the composition, orientation and position of insertion of the *CEN/ARS* and 2 μ cassettes are identical to those described by Sikorski and Hieter (1989) and Christianson *et al*. (1992), respectively. Throughout pRS plasmid

Table 3. Description of Δ*0* deleter plasmids.

Plasmid	ATCC number[a]	Resulting allele	Enzyme for linearization[b]	Selection(s)[c]
pAD1	87468	*leu2Δ0*	*Sal*I	Ura$^+$, Foa^R
pAD2	87469	*lys2Δ0*	*Cla*I	Ura$^+$, Foa^R
pAD4	87470	*met15Δ0*	*Cla*I	Ura$^+$, Foa^R
pJL164	87471	*ura3Δ0*	*Spe*I, *Xho*I	One-step Foa^R
Other cloned deletions				
YRp14-*his3Δ200*	77146	*his3Δ200*		Ura$^+$, Foa^R
YRp14-*leu2Δ1*	77147	*leu2Δ1*		Ura$^+$, Foa^R
YRp14-*trp1Δ63*	77148	*trp1Δ63*		Ura$^+$, Foa^R
pΔADE2	99604	*ade2::hisG::URA3::hisG*		One-step Foa^R

[a]The set of four designer deletion plasmids in this table is available from ATCC as a kit (87472) of four strains.
[b]Cutting with the indicated enzyme will target the plasmid to the desired locus. In the case of pJL164, two enzymes are used to release the fragment containing the deleted genomic region of *URA3*. Deletion of *URA3* requires two homologous recombination events.
[c]Two-step gene replacements (indicated as Ura$^+$, Foa^R) are required for most of these plasmids. For pJL164, which is a single-step replacement, cells are outgrown on YPD at 30°C prior to replica-plating to Foa medium. See Materials and Methods for details.

© 1998 John Wiley & Sons, Ltd.

construction, all 5′ overhangs created by restriction digests were blunted by filling in with the Klenow large fragment of *Escherichia coli* DNA polymerase, and all 3′ overhangs were made blunt using the 3′ to 5′ exonuclease activity of T4 DNA polymerase. All sites indicated on the restriction maps were verified by digestion.

Construction of the MET15 *plasmids, pRS401, pRS411 and pRS421* The pRS40X series precursor plasmid pJK142 (Keeney and Boeke, 1994) was digested with *Nde*I, blunted and the minimal complementing 1621 bp *Ehe*I–*Sca*I fragment of the *MET15* genomic locus was inserted to create pRS401 (*Ehe*I recognizes the same site as *Nar*I but leaves a blunt end). pRS401 was digested with *Aat*II, blunted, and a 514 bp *Hin*cII–*Pst*I (blunt) fragment of pRSS84 containing the *CEN6/ARSH4* cassette (Sikorski and Hieter, 1989) inserted to create pRS411. pRS401 was digested with *Aat*II, blunted, and the 1345 bp *Hpa*I–*Nde*I (blunt) fragment of YEp24 containing the 2 μ *ori* (A form; Christianson *et al*., 1992) was inserted to create pRS421.

Construction of the ADE2 *plasmids, pRS402, pRS412 and pRS422* pJK142 was digested with *Nde*I, blunted, and a 2252 bp *Bgl*II (blunted) fragment of pASZ11 (kindly provided by Patrick Linder) containing a minimal complementing region of the *ADE2* genomic locus (Stotz and Linder, 1990) was inserted to create pRS402. The *CEN6/ARSH4* plasmid pRS412 and the 2 μ plasmid pRS422 were created by ligation of the 1246 bp *Apa*LI fragment of pRS402 and the 3781 bp *Apa*LI fragment of pRS402 with either the 1011 bp *CEN6/ARSH4 Apa*LI fragment of pRS411 or the 1837 bp 2 μ *Apa*LI fragment of pRS421. It should be noted that these *ADE2* clones bear a silent A to G mutation at nucleotide 1243 of the GenBank sequence designed to destroy a *Bgl*II restriction site (Stotz and Linder, 1990).

Construction of the kanMX4 *pRS400 plasmid* pJK142 was digested with *Nde*I, blunted, and a 1483 bp *Sma*I to *Eco*RV fragment of pFA6A (Wach *et al*., 1994) containing *kanMX4* was inserted, and kanamycin-resistant transformants were isolated. The initial clone was found to contain two copies of the vector sequence ligated in tandem with a single copy of *kanMX4*. This clone was digested with *Aat*II and re-ligated to create pRS400 which contains a single copy of the vector and *kanMX4*.

Construction of markerless pRS vectors The *CEN6/ARSH4* plasmid pGC25 and the 2 μ plasmid pGC26 were created by ligation of the 1246 bp *Apa*LI fragment of pRS402 and the 1527 bp *Apa*LI fragment of pJK142 with either the 1011 bp *CEN6/ARSH4 Apa*LI fragment of pRS411 or the 1837 bp 2 μ *Apa*LI fragment of pRS421. These plasmids contain either the *CEN6/ARSH4* or the 2 μ sequence, but lack a selectable marker for use in *S. cerevisiae*. pGC25 (*CEN*) and pGC26 (2 μ) should be useful as the basis for a further extension of the pRS line to encompass an even greater diversity of markers.

Introduction of gene disruption mutations into yeast and strain history

We introduced the *ade2Δ::hisG* mutation into YPH681 (a direct descendant of FY2) to create YCB436 using the previously described *ADE2* disrupter pΔADE2 (Aparicio *et al*., 1991), which results in deletion of the *ADE2* ORF (with the exception of six C-terminal amino acids) and replacement with a copy of the bacterial *hisG* gene. To create the correct starting strains for the designer deletions, YCB436 was crossed to BY378 and the resulting diploid was sporulated. Two spores from this cross, YCB451 and YCB452 were then gene converted to $URA3^+$ by transformation with the 13 kb *Eco*RI fragment containing $URA3^+$ (derived from pSK179 provided by S. Kunes) and selection for Ura^+, thus creating strains BY397 and BY398, respectively. To introduce the *ura3Δ0* deletion, the *Spe*I–*Xho*I fragment from pJL164 was transformed into these two strains, transformants were grown for 1 day at 30°C on YPD plates to allow loss of pre-existing Ura3 protein and development of the FOA^R phenotype (Ronne and Rothstein, 1988) and then replica-plated to 5-FOA plates. 5-FOA^R colonies were identified and patched onto YPD. These were screened for *URA3* deletions by a reversion analysis. The YPD plates were replica-plated to SC − ura plates and irradiated with UV-light at 100 J/m^2 using a Stratalinker (Stratagene; setting of 100 μJ × 100/cm^2). Following irradiation, plates were wrapped in foil to prevent photoreactivation and incubated for several days at 30°C. Using this test, *ura3Δ0* deletions were identified by their inability to revert to

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 15

Ura$^+$, whereas point mutations reverted. Finally, correct *ura3Δ0* deletions in strains BY399 and BY401 were confirmed by Southern blot analysis (Ausubel *et al.*, 1987).

Figure 2 describes the method used to introduce the remaining designer deletion mutations. First, deleter plasmids pAD1, pAD2 and pAD4 were linearized at unique sites within each right flanking fragment (pAD1 with *Sal*I, pAD2 and pAD4 with *Cla*I) and transformed into the *ura3Δ0* strains. Genomic integrants were selected on SC − ura plates. Ura$^+$ colonies were replica-plated to YPD plates to allow recombinational loss of the integrated *URA3* marker and pRS406 backbone and then replica-plated to 5-FOA plates to identify those that had become Ura$^-$. 5-FOA-resistant colonies were picked and checked for loss of the targeted auxotrophic marker by replica-plating to SC − leu, SC − lys or lead-containing plates (*met15* strains are brown on these plates). Finally, deletions were confirmed by Southern analysis. The pAD1 transformation was performed in strain BY399 to create BY404, the pAD2 transformation was performed in strain BY401 to create BY406 and the pAD4 transformation was performed in strain BY406 to create BY411.

To create a set of strains consisting of a varying array of deletion alleles, BY411 (*MATα ura3Δ met15Δ0 trp1Δ63 his3Δ200 ade2Δ lys2Δ0*) was crossed to BY379 (*MAT***a** *his3Δ200*) and a spore was selected with the following genotype: *MATα met15Δ0 his3Δ200 lys2Δ0*. This strain, BY413, was mated to BY404 (*MAT***a** *ura3Δ leu2Δ0 trp1Δ63 his3Δ200 ade2Δ*). Tetrad dissection of four resulting diploids yielded an unexpected result: two spores of each tetrad grew slowly on rich medium, irrespective of genotype. Suspecting that these might have resulted from a single nuclear petite mutation, we tested all tetrads for growth on glycerol- and ethanol-containing rich medium (petite strains cannot utilize these carbon sources and so fail to grow). We found that these slow-growing spores did not exhibit a petite phenotype since all spores grew on these plates. We back-crossed one of the faster-growing spores from this cross (BY418; *MATα ura3Δ met15Δ0 trp1Δ63 his3Δ200 ade2Δ lys2Δ*0 leu2*Δ0*) to the parent FY4 (*MAT***a**) prototrophic strain. All spores obtained from tetrad dissection of this cross grew equally well, indicating that the mutation causing slow growth observed in the previous cross had been eliminated. The strain history is presented schematically in Figure 1).

Northern analysis of flanking genes

RNA was prepared from 10 ml of a log-phase culture (OD_{600}=1–2) grown in YPD at 30°C. The cells were washed once in water and resuspended in 0·3 ml RNA buffer (0·1 M-NaCl, 0·1 M-Tris base, 0·03 M-EDTA, 1% w/v *N*-lauryl sarcosine, pH 8·9–9·0) and frozen at − 70°C. Upon thawing, 0·4 g glass beads, 0·15 ml phenol and 0·15 ml chloroform were added, and the cells were broken by vortexing for 10 min. Following a short spin to separate the phases, 1/10th volume 3 M-NaOAc and 2·5 volumes ice-cold EtOH were added to the aqueous layer to precipitate the nucleic acids (mostly RNA). The precipitate was dissolved in 100 µl of loading buffer (50% formamide, 25% water, 15% formaldehyde, 10% 10 × MOPS—0·2 M-morpholinopropane-sulfonic acid, 0·05 M-NaOAc, 0·01 M-Na_2EDTA; pH adjusted to 7·0 with NaOH). 10 µg RNA was first heated to 55°C for 15 min and then electrophoresed on a 1·2% agarose/formaldehyde gel as described (Ausubel *et al.*, 1987. The separated RNA was transferred to Genescreen Plus (ICN) and hybridized to the appropriate probe (Ausubel *et al.*, 1987) and Genescreen Plus protocol.

PCR protocol for PCR-mediated gene disruptions

Using primers consisting of 40 nts of gene-specific sequence at the 5′ end followed by: 5′-CTGTGCGGTATTTCACACCG-3-′ (left primer) and 5′-AGATTGTACTGAGAGTGCAC-3′ (right primer), any auxotrophic marker can be amplified from a pRS40X integrating plasmid using the following PCR protocol: 94°C 2 min, [94°C 1 min, 55°C 1 min, 72°C 2 min] × 10 cycles, [94°C 1 min, 65°C 1 min, 72°C 2 min] × 20 cycles, 72°C 10 min. This protocol usually results in amplification of only the specific product. If yields are high enough, it is possible to directly transform yeast, but generally, 10-fold concentration by EtOH precipitation before transformation is desirable.

RESULTS AND DISCUSSION

A compilation of commonly found alleles of auxotrophic marker genes

To develop a strategy for construction of designer deletion strains, we first had to compile relevant information for existing auxotrophic selectable marker alleles. Many commonly used strains contain mutations in the auxotrophic

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 16

Table 4. A compilation of commonly used auxotrophic marker mutations.

Allele	ΔORF?	Reverts?	Notes	Molecular description[a]	References
ade2-101	No	Yes	Ochre mutation, red colonies		
*ade2-Bgl*II	No	No	Red colonies	Frameshift (*Bgl*II site filled in at pos. 592)	Engebrecht and Roeder (1990)
his3Δ200	Yes	No	Cold sensitive; high frequency of petite formation, especially during transformation	Δ 1 kb (− 205–835)	Struhl (1985); Fasullo and Davis (1988); Siram *et al.* YGM RNA processing mtg (1993)
his3Δ1	Partial	No		Δ 187 bp *Hin*dIII–*Hin*dIII internal (305–492)	Scherer and Davis (1979)
his3-11,15	No	No	Double mutant		Lau and G. R. Fink, unpublished
leu2Δ1	Partial	No		Δ 0·6 kb, *Eco*RI–*Cla*I internal (163–649)	Sikorski and Hieter (1989)
leu2-3,112	No	No	Double mutant	*leu2-3* is a +1 frameshift mutation	Hinnen *et al.* (1978); Gaber and Culbertson (1982)
lys2-801	No	Yes	Amber mutation		
lys2Δ202	Partial	No		Δ 1·0 kb, *Xho*I–*Hpa*I internal (1813–2864)	Winston *et al.* (1995)
trp1Δ1	Yes	No	Cold sensitive[b], weak galactose inducer (deletes *GAL3* UAS), removes *ARS1*, also called *trp1-901*	Δ 1·45 kb, *Eco*RI–*Eco*RI (− 102 to 1352)	Sikorski and Hieter (1989)
trp1Δ63	Partial	No	Cold sensitive[b]	Δ 0·6 kb, *Eco*RI–*Hin*dIII (− 102 to 513)	Sikorski and Hieter (1989)
trp1-289	No	Yes	Cold sensitive[b]		
ura3-52	No	No		Ty1 insertion (transcribing left to right) at pos. 121	Rose and Winston (1984)
ura3-1	No	Yes			
Alleles described in this study					
ade2Δ::hisG	No	No		Aparicio *et al.* (1991)	
leu2Δ0	Yes	No	Designer deletion	This study	
lys2Δ0	Yes	No	Designer deletion	This study	
met15Δ0	Yes	No	Designer deletion	This study	
ura3Δ0	Yes	No	Designer deletion	This study	

An updatable version of this table is maintained by SGD (http://genome-www.stanford.edu/Saccharomyces/).
[a]The sequence coordinates are relative to the first ATG of the selectable marker ORF, in which the A residue is defined as +1.
[b]All Trp$^-$ strains are cold sensitive (Singh and Manney, 1974).

marker genes—namely, point mutations or small internal deletions which are not conducive to successful PCR-mediated gene disruption. These auxotrophic loci retain extensive homology with the prototrophic marker gene segment of the PCR product, leading to high levels of mistargeting. We have compiled a table of many of these commonly used alleles (Table 4) which should be useful in determining a strain's adaptability for PCR-mediated gene disruption.

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 17

PCR-mediated gene disruption is variably successful in its degree of targeting the locus of interest, even when complete deletion alleles corresponding to the marker gene being introduced are used. The reason for this variability in efficiency is not entirely clear but appears to be related primarily to the targeting sequences that flank the target gene (Q. Feng and J.D.B., unpublished data), and to a lesser extent to the nature of the disrupting auxotrophic marker. It is important to note that the commonly used segment of *LEU2* (the *Xho*I–*Sal*I fragment found in the pRS vectors) contains 240 bp of δ (Ty2 LTR) sequence at its 5′ end. This may lead to a higher frequency of mistargeting than with the other auxotrophic markers.

Complete deletion alleles of URA3*,* LEU2*,* LYS2 *and* MET15

Commonly used strains of the S288C lineage contain only four auxotrophic alleles that are well-defined deletions, namely *his3Δ200*, *leu2Δ1*, *trp1Δ1*, and *trp1Δ63*. Both the *his3Δ200* and *trp1Δ1* mutations completely remove the protein coding regions, whereas *trp1Δ63* leaves intact the 3′ end of the gene and *leu2Δ1* is an internal deletion (Table 4). All of these mutations except *leu2Δ1* work well for the method of PCR-mediated gene disruption, resulting in proper gene disruptions 30–100% of the time. Thus, new designer deletion alleles of *TRP1* and *HIS3* have not been constructed. However, PCR-mediated gene disruptions using other commonly used auxotrophic alleles have been difficult to obtain because of the high background of gene conversion. For this reason we set out to construct an isogenic set of strains in which *URA3*, *LYS2*, *LEU2* and *MET15* were completely deleted.

The strategy for the designer deletion alleles (designated *Δ0*) is depicted in Figure 2. Approximately 1 kb 'left' and 'right' flanking regions of the auxotrophic marker to be deleted were cloned into a *URA3* integrating vector. Linearization of the resulting deleter plasmid within the right flanking sequence and transformation of a *ura3Δ0* strain results in integration of the plasmid at the auxotrophic marker locus. Subsequent recombination between the duplicated left or right flanks results in 5-FOA resistance (due to excision of the *URA3* plasmid) and leads to acquisition of the appropriate auxotrophy (in principle, 50% of the time).

Because this strategy involves integration of a *URA3* containing plasmid, it was important that our starting strain was not only Ura$^-$, but also lacked sequences overlapping with the plasmid *URA3* gene. To completely delete *URA3*, the plasmid pJL164, containing a genomic 4·0 kb *Bgl*II–*Bam*HI fragment in which the region from *Hin*dIII to *Sma*I encompassing the *URA3* gene has been deleted (Figure 3), was digested with *Spe*I and *Xho*I to liberate a DNA fragment spanning the deletion (Table 3). This was transformed into the *URA3*$^+$ strains BY397 and BY398 and Ura$^-$ colonies were selected on 5-FOA. Reversion tests and Southern blotting confirmed the acquisition of the *ura3Δ0* allele in strains BY399 and BY401.

All four designer deletion plasmids completely remove the open reading frame encoding the appropriate auxotrophic marker. They also remove several hundred bp of 5′ flanking region, and as much 3′ flanking region as is practical. Ideally, there should be no homology between the marker segment in the plasmid to be used and the deletion in the chromosome. In two cases (*ura3* and *met15*) it was possible to remove all homology between plasmid marker segment and chromosomal deletion. The details of the design of each deletion were dictated by what was known about the nature of flanking genes; often, as at the 3′ end of *LYS2*, there was not much room to maneuver due to the proximity of the adjacent gene (Figure 3). Both *MET15* and *LEU2* are flanked by tRNA genes and δ elements and in the case of *LEU2*, an entire Ty2 element. Because the segment of *LEU2* commonly used in cloning vectors contains part of this Ty2 element as well as a tRNA^{Leu} gene, it was necessary to delete the Ty2 element, the tRNA gene, and *LEU2* in order to eliminate overlap with the vector segment, at least at one end of the marker gene. Thus the *leu2Δ0* allele is considerably larger than the rest. No unusual phenotypes have resulted from this deletion.

To create the *leu2Δ0* and *lys2Δ0* strains, pAD1 was linearized with *Sal*I and pAD2 with *Cla*I (Table 3 and Figure 2) and transformed into BY399 and BY401 respectively. Ura$^+$ colonies were patched onto SC − ura, replica-plated to YPD (to allow loss of the *URA3*$^+$ and flanking vector sequences) and replica-plated to 5-FOA plates. Papillae from the 5-FOA plates were streaked onto YPD and replica-plated to either SC − leu or SC − lys to identify auxotrophs (Figure 2). Of 24 5-FOA^R papillae from the pAD1

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 18

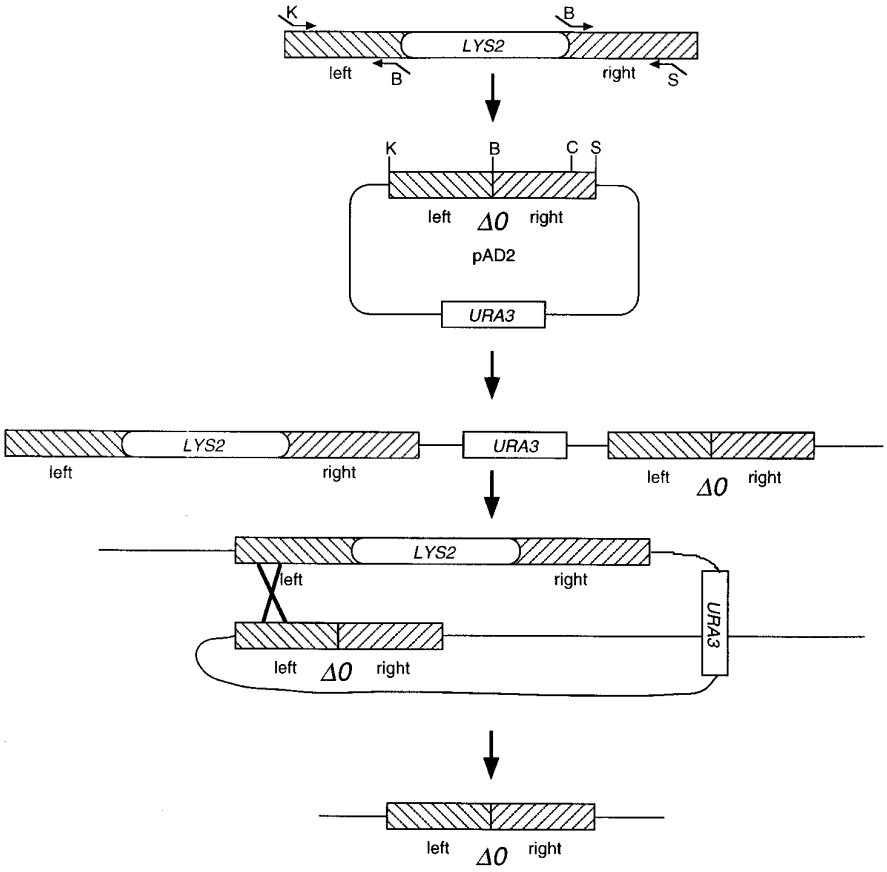


Figure 2. Method of construction of designer deletion strains. An example is given in which *LYS2*$^+$ is deleted to create *lys2Δ0*. The *LYS2* ORF is indicated by a white oval and the hatched regions indicate left and right flanking sequences. Lines with arrows indicate oligonucleotide primers for PCR-amplification of the flanking regions, and the restriction sites engineered into these oligos are indicated by K, *Kpn*I; B, *Bam*HI or S, *Sac*I. Using these oligos, the flanking regions are PCR-amplified, cut with the indicated enzymes and cloned into pRS406 to create a deleter plasmid; in this case, pAD2. pAD2 is then linearized at the unique *Cla*I (C) site in the right flanking region and transformed into yeast. Integration of the linearized vector within the genomic *LYS2* right region results in the configuration of sequences shown in the middle panel of the diagram. Recombination between the two left regions results in deletion of the *LYS2*$^+$ and vector sequences. Not shown, but equally likely, is recombination between the two right regions which results in restoration of the wild-type *LYS2*$^+$ allele.

transformation, three were Leu$^-$, and of 24 5-FOA^R papillae from the pAD2 transformation, six were Lys$^-$. The deletions were confirmed by genomic Southern analysis (Figure 4).

The *met15Δ0* mutation was introduced into BY406 (the *lys2Δ0* derivative) in a similar fashion. Plasmid pAD4 was linearized with *Cla*I and transformed. Following 5-FOA selection, candidate strains were streaked onto YPD plates and replica-plated to lead-containing plates. When grown on these plates *met15* strains are brown in color, identifying those strains in which recombinational loss of the *URA3*$^+$ and vector sequences had resulted in a Met$^-$ phenotype. The *met15Δ0* allele was confirmed by Southern analysis (Figure 4).

Using the *ADE2* disrupter plasmid pΔADE2 (Aparicio *et al.*, 1991, the *ade2Δ::hisG::URA3::hisG* (originally referred to as *ade2Δ* by Aparicio *et al.*) mutation was introduced into these strains; these were converted to *ade2Δ::hisG* alleles by plating on 5-FOA as above. Since six C-terminal

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 19

amino acids remain of the *ADE2* ORF, the *ade2Δ::hisG* mutation is not technically a designer deletion, and it has been shown to be useful in PCR-mediated gene disruption (D. Gottschling, pers. comm.).

Because these strains are intended to be used for, among other things, the systematic genetic analysis of genes of unknown function, it was important to establish that expression of the ORFs neighboring the deleted auxotrophic markers was unaffected in these deletion strains. This seemed especially important because many yeast workers have found that two of the commonly used deletion alleles, *his3Δ200* and *trp1Δ1*, have 'collateral' phenotypes that are sometimes undesirable due to effects on adjacent genes. *his3Δ200* strains display an increased rate of petite formation, presumably due to effects on expression of the neighboring *PET56* gene. *trp1Δ1* removes ARS1 (with no obvious phenotype) and partially disables the neighboring *GAL3* gene by removing a portion of its UAS. Finally, both *trp1Δ63* and *trp1Δ1* strains share a cold-sensitivity phenotype which all Trp^- strains exhibit (Singh and Manney, 1974).

It is impossible to totally rule out effects on flanking genes but the deletions were designed to minimize the possibility of such effects. In the construction of the *leu2* and *met15* deletions, tRNA genes that are part of multigene families were removed in addition to the auxotrophic marker. The phenotypic consequence of this is likely to be minimal, based on experiences with other multi-gene tRNA families (Byström and Fink, 1989; Åström and Byström, 1993). All of the designer deletions remove marker genes that are flanked by convergently transcribed neighboring genes (Figure 3); effects on the flanking genes are thus expected to be minimal. Nevertheless, the possibility of altered transcript sizes and transcriptional level of neighboring genes was investigated by Northern analysis. RNA was prepared from BY418 (which has all four designer deletion mutations) and FY2 (the wild-type strain), and left and right flanking probes (Figure 3) were used to detect the transcription of neighboring ORFs; in no case was the expression of a neighboring ORF significantly affected (Figure 5; compare the neighboring gene specific transcript with the Ty1 loading control transcript). However, our analyses did detect one minor but interesting expression pattern change. The *LEU2* neighboring ORF, *YCL17C* (reading toward *LEU2*), appears to direct a major transcript and numerous larger minor transcripts. In the *leu2Δ0* strain, the overall levels of the transcripts are slightly reduced and in particular, the largest transcripts appear to be missing. Such an observation is consistent with the possibility that the larger transcripts normally terminate within the deleted region.

A set of designer deletion strains

To obtain strains with numerous different combinations of auxotrophic alleles, 206 tetrads were dissected from a cross between FY3 and BY418 (Figure 1). This cross gave ~30% sporulation and 96·5% spore viability using the conditions described in Materials and Methods. A subset of the strains from this cross identified in Table 1, which should prove generally useful to the yeast community, has been deposited with the ATCC.

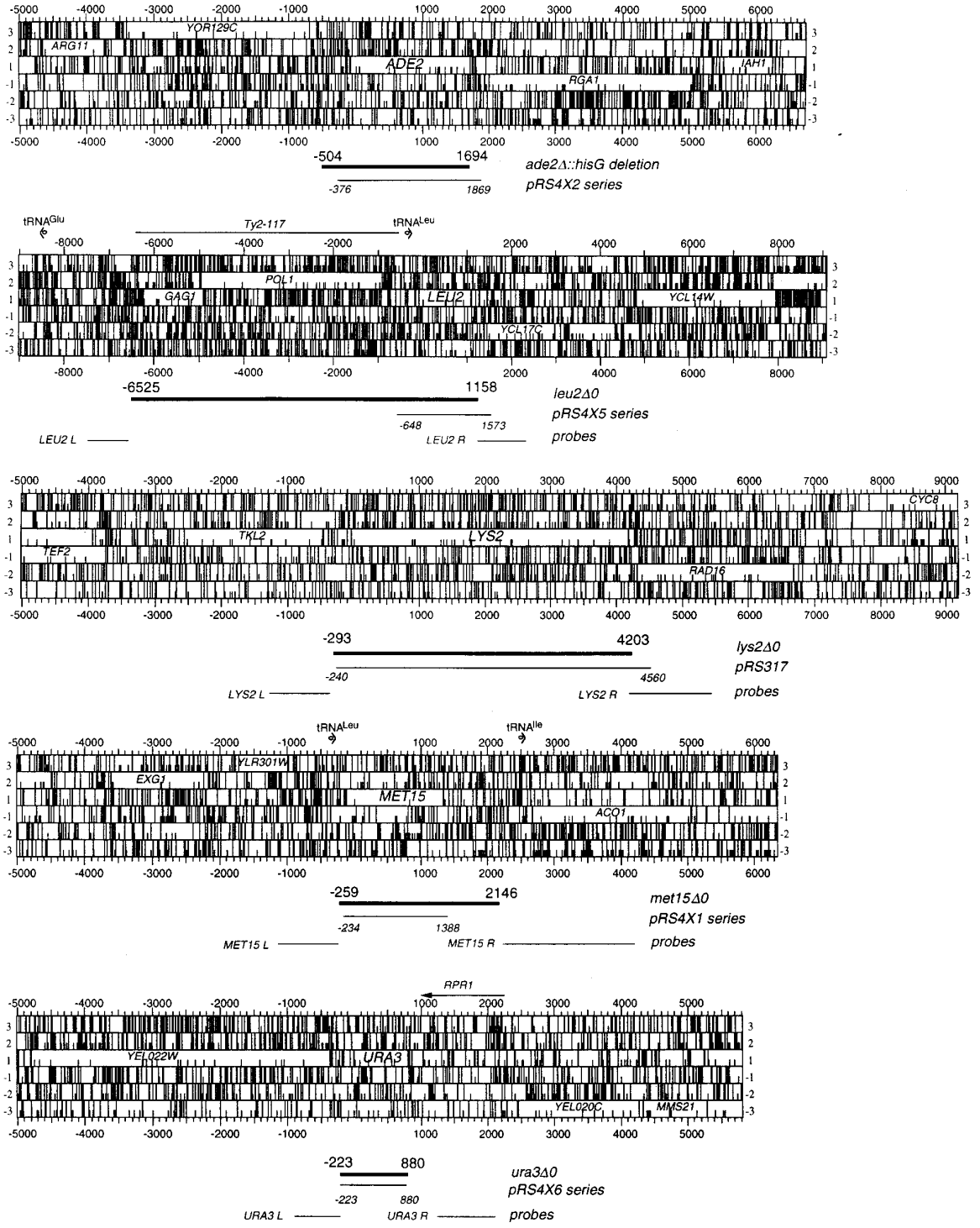
New pRS plasmids with ADE2, MET15 *and* kanMX4 *markers*

The pRS vector series is a set of *S. cerevisiae* shuttle plasmids that are in wide use. Much of their appeal derives from their small size, modular design, minimal length marker segments, useful polylinker and consistency of structure. Also, complete DNA sequence files exist for these plasmids, greatly facilitating construction and analysis of recombinant plasmids. The original set of pRS vectors (the pRS300 series) were based on pBluescript KS+ (Sikorski and Hieter, 1989) and included both integrating and centromere-containing versions; the pRS400 series is based on pBSII SK+ (Stratagene; Christianson *et al*., 1992) and also includes 2 μ versions. This latter series is more generally useful for several reasons: (1) the *lacZ* α-complementation is improved over the pRS300 series; (2) the polylinker is flanked by *Bss*HII sites, allowing a simple assessment of insert size to be carried out and facilitating swapping of inserts from one vector to another; and (3) the pRS400 series is more complete, including integrating, *CEN*, and high copy versions. The 'code' for the pRS vectors is simple: the second digit (0, 1 or 2) specifies the plasmid type (integrating, *CEN* or 2 μ, respectively) and the third digit specifies the selectable marker (see Table 5).

We constructed new integrating, *CEN* and 2 μ plasmids for use with the *ade2Δ* and *met15Δ0* strains (Figure 6). In all cases, the plasmid backbone is identical to the pRS400 series of plasmids, and the selectable markers were inserted into the same restriction site as had been utilized for

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 20



© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 21

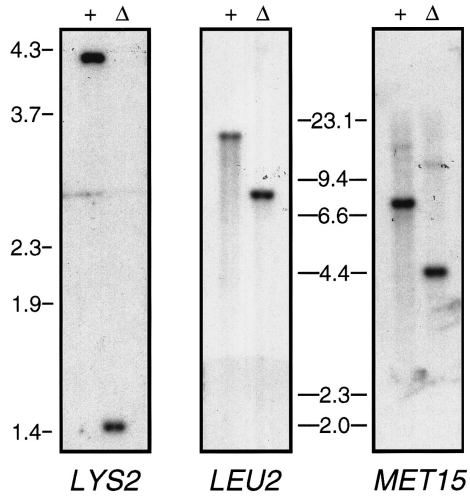


Figure 4. Genomic DNA blot analysis of deletion alleles. DNA was isolated from the wild-type strain FY5 (+) and the multiple deletion strain BY418 (Δ). Approximately 1 μg of genomic DNA was digested with *Bst*XI (*LYS2*), *Pst*I (*LEU2*) or *Kpn*I (*MET15*), and fractionated on a 1% agarose gel. Following transfer to a nylon filter, each blot was probed with a gene-specific flanking probe (as depicted in Figure 3): *LYS2*, left flank; *LEU2*, right flank; and *MET15*, right flank. Expected product sizes are: *LYS2*, wild-type genomic *Bst*XI fragment is 4 kb and the deletion *Bst*XI fragment is 1·4 kb; *LEU2*, wild-type genomic *Pst*I fragment is 12 kb and the deletion *Pst*I fragment is 7·6 kb; *MET15*, wild-type genomic *Kpn*I fragment is 6·8 b and the deletion *Kpn*I fragment is 4·4 kb. The observed *LEU2* fragment sizes do not agree with the expected; this is probably because chromosome III sequence was not obtained from this strain background (Oliver *et al*., 1992). However, the difference in size between the deleted and wild-type *LEU2* fragments agrees with the sequence. DNA marker sizes are in kb as indicated.

the pRS plasmids (Sikorski and Hieter, 1989; Christianson *et al*., 1992). In the case of the *ADE2* plasmids, pRS402, pRS412 and pRS422, the fragment inserted represents a minimal complementing fragment as previously determined by Stotz and Linder (1990). For the *MET15* plasmids, pRS401, pRS411 and pRS421, a minimal length complementing fragment as defined by a complementation analysis (Figure 7). The *Nar*I–*Sca*I fragment was the smallest strongly-complementing fragment and was cloned into the pRS plasmid backbones. The *ADE2* and the *MET15* plasmids were demonstrated to fully complement the *ade2*Δ and *met15Δ0* mutations as judged by colony growth on appropriate selective media. In the course of these experiments we noted that unlike all of the other auxotrophic designer deletion markers, the *met15Δ0* strain eventually (and at a low frequency) spawns slow-growing pseudo-Met$^+$ papillae when replica-plated as a thick patch on YPD medium to SD plates. Such growth is maximal at 22°C, drops dramatically at 25°C and decreases linearly from 30–37°C, where little growth is observed. These pseudo-Met$^+$ papillae are not true revertants because they form black colonies on lead plates, like the parental *met15* mutant strain. Most importantly, these colonies do not grow when restreaked on methionine-free medium, so they are not true Met$^+$ cells. Our working hypothesis is that these colonies represent mutants that are more efficient at scavenging methionine or cysteine from adjacent cells. This phenomenon will be described in more detail elsewhere, but does not create much of a practical problem. Using standard LiOAc transformation procedures, these pseudo-Met$^+$ colonies did not cause any detectable background (i.e. no visible colonies on a control plate to which no pRS4X1 plasmid DNA had been added). In fact no colonies appeared on the no DNA plate even after 10 days of growth. Finally, we constructed the integrating plasmid pRS400, which contains the *kanMX4* gene (Wach *et al*., 1994). This is useful for use as a template for PCR-mediated gene disruption. Available information on the

Figure 3. Genomic structure of deletion alleles and overlap with vector segments. The genomic region encompassing each of the five deletions is presented as a six-phase ORF analysis. To the left of each are indicated the three possible reading frames of the plus and minus strands. Lines from the top to the bottom of each frame indicate stop codons in that frame, and lines from the middle to the bottom of each frame indicate methionine residues in that frame (output is from the DNA Strider 1.2 ORF Map function). Each ORF is identified by its gene name or ORF designation, and above each diagram, tRNA and Ty elements have been identified. All nucleotide numbers are defined by +1 being assigned to the A of the first ATG in the auxotrophic marker ORF to prevent possible confusion resulting from renumbering of sequences in databases. The bold line under each diagram indicates the extent of each deletion. Below this is a thin line indicating the extent of the fragment of genomic sequence encoding the auxotrophic marker cloned into the corresponding pRS vector. Finally, the left and right flanking regions which were used as probes for the Southern and Northern analyses in this paper are indicated. In the cases of *LEU2*, *LYS2* and *MET15*, these are identical to the left and right flanking regions that were PCR-amplified for the deleter constructs (Figure 2). Note that the scale is the same for all regions except for the *LEU2* region, which had to be compressed to accommodate the very large deletion which removed an adjacent Ty2 element as well as *LEU2* itself.

© 1998 John Wiley & Sons, Ltd.

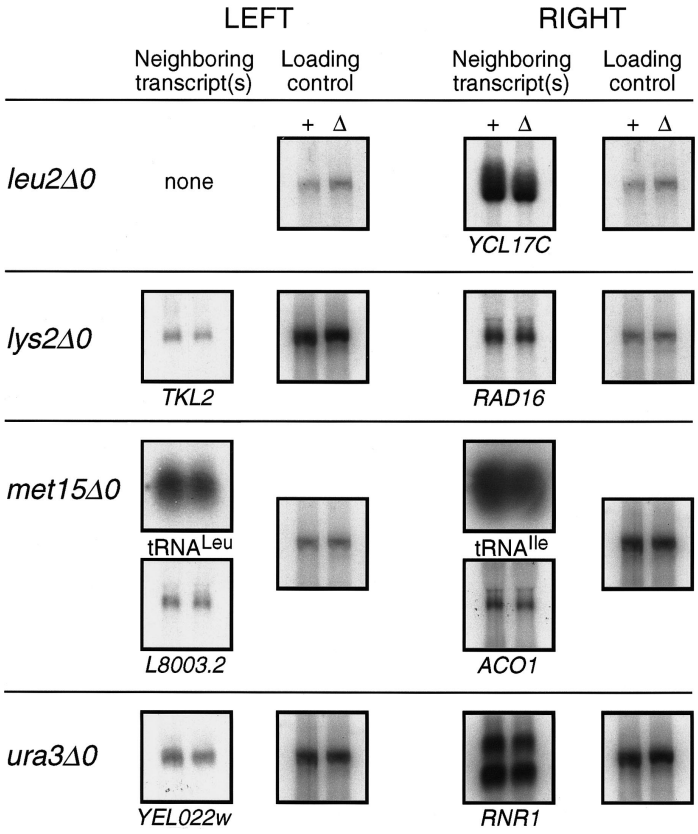


Figure 5. RNA blot analysis for effects on expression of neighboring genes. RNA was isolated from the wild-type strain FY5 (+) and the multiple deletion strain BY418 (Δ). Multiple samples containing 9·8 µg total RNA from each strain were fractionated on a 1·2% formaldehyde agarose gel. Following transfer to a nylon filter, the blot was cut into strips and each was probed with a gene-specific left flank probe (probes are indicated in Figure 3). Following this, the blots were reprobed with gene-specific right probes. For all panels, the wild-type sample is presented on the left as indicated for *LEU2*. In all cases, transcript sizes were consistent with the sizes of encoded ORFs or known transcripts; the inferred identities of the transcripts are indicated below each panel. To allow comparison between samples, the Ty transcript is included as a loading control. Fortuitously, two left probes (*LEU2* and *MET15*) contained Ty LTR sequences. The *LEU2* left probe was used to reprobe those blots in which the Ty1 transcript had not yet been visualized. It should be noted that based on the sequence no transcript was expected (i.e. there were no known ORFs or genes) to the left of *leu2Δ0* and no hybridizing bands were seen with this probe.

sequences and availability of the complete series of pRS400 series vectors is summarized in Table 5.

Universal primers for PCR-mediated gene disruption using pRS plasmid templates

A pair of primer segments was designed which can be universally used to amplify any selectable marker from the pRS set of plasmids (Figure 8). Addition of 40 nts of sequence specific to the gene of interest at the 5′ ends of the primers results in a single set of 60 nt primers which can be used to delete the gene of interest with either *URA3*, *TRP1*, *HIS3*, *LEU2*, *LYS2*, *ADE2*, *MET15* or *kanMX4*. We have used this strategy to disrupt a number of genes in our laboratories.

© 1998 John Wiley & Sons, Ltd.

Table 5. The pRS400 series of vectors.

Plasmid and type[a]	Selectable marker	Accession numbers[b]	Available from[c]	References
pRS400 (I)	*kanMX4*	U93713	Pending	Wach *et al.* (1994), this work
pRS401 (I)	*MET15*	U93714	ATCC 87473	This work
pRS411 (C)	*MET15*	U93715	ATCC 87474	This work
pRS421 (2)	*MET15*	U93716	ATCC 87475	This work
pRS402 (I)	*ADE2*	U93717	ATCC 87477	This work
pRS412 (C)	*ADE2*	U93718	ATCC 87478	This work
pRS422 (2)	*ADE2*	U93719	ATCC 87479	This work
pRS403 (I)	*HIS3*	U03443	ST 217403	
pRS413 (C)	*HIS3*	U03447	ST 217413	
pRS423 (2)	*HIS3*	U03454	ATCC 77104	Christianson *et al.* (1992)
pRS404 (I)	*TRP1*	U03444	ST 217404	
pRS414 (C)	*TRP1*	U03448	ST 217414	
pRS424 (2)	*TRP1*	U03453	ATCC 77105	Christianson *et al.* (1992)
pRS405 (I)	*LEU2*	U03445	ST 217405	
pRS415 (C)	*LEU2*	U03449	ST 217415	
pRS425 (2)	*LEU2*	U03452	ATCC 77106	Christianson *et al.* (1992)
pRS406 (I)	*URA3*	U03446	ST 217406	
pRS416 (C)	*URA3*	U03450	ST 217416	
pRS426 (2)	*URA3*	U03451	ATCC 77107	Christianson *et al.* (1992)
(pRS317[d]) (C)	*LYS2*		ATCC 77157	Sikorski and Boeke (1990)

[a]Plasmid type is in parentheses: I, integrating plasmid; C, centromeric (*CEN/ARS*) plasmid; 2, 2 μ origin-containing plasmid.
[b]DNA sequence accession number. DNA sequences are available from VecBase (URL: http://biology.queensu.ca/~miseners/vector.html) or GenBank.
[c]ATCC, American Type Culture Collection, Rockville, MD; ST, Stratagene, La Jolla, CA. Sets of pRS vectors are available in kit form from ATCC (pRS401, 411, 421, 87476; pRS402, 412, 422, 87480; pRS423, 424, 425, 426, 77108).
[d]No pRS400 series plasmids exist for *LYS2* at this time.

Other applications for the designer deletion strains and new plasmids

Although we have focused on the application of PCR-mediated gene disruption in this paper, we envision that the designer deletion alleles will be the alleles of choice for a very wide range of applications. They are ideal as transformation markers because they cannot revert (with the special exception of *met15* noted above). The ability to knock out genes of interest by the PCR approach is enhanced by expanding the set of available markers in pRS vectors. This is especially useful in the study of multigene families (Brachmann *et al.*, 1995) many of which have been revealed by the completion of the *S. cerevisiae* genome sequence. By synthesizing a single pair of primers for each target gene of interest, the investigator has the ability to knock out a member of the gene family with any of eight currently available markers in the pRS vectors (Table 5). If a heavily marked strain such as BY418 is used as the recipient, up to eight different genes can be knocked out before running out of selectable markers. In addition, the alleles that have no overlap (or minimal overlap only at one end of the marker) will be ideal for studies of DNA recombination in which lack of homology between e.g. a plasmid and the chromosome is critical. It should be pointed out that an alternative and elegant strategy to solve the same problem was developed by Wach *et al.* (1994), who constructed the *kanMX4* marker, which has no homology with the yeast genome and confers dominant resistance to G418, and hence can be used with strains of any background. Although *kanMX4* does work as a selectable marker, G418 is an expensive reagent to use routinely, and expression of the drug resistance requires an outgrowth period prior to plating on G418 and so is somewhat less convenient. Both strategies further enrich the already extensive molecular genetic toolkit available for *S. cerevisiae*.

© 1998 John Wiley & Sons, Ltd.

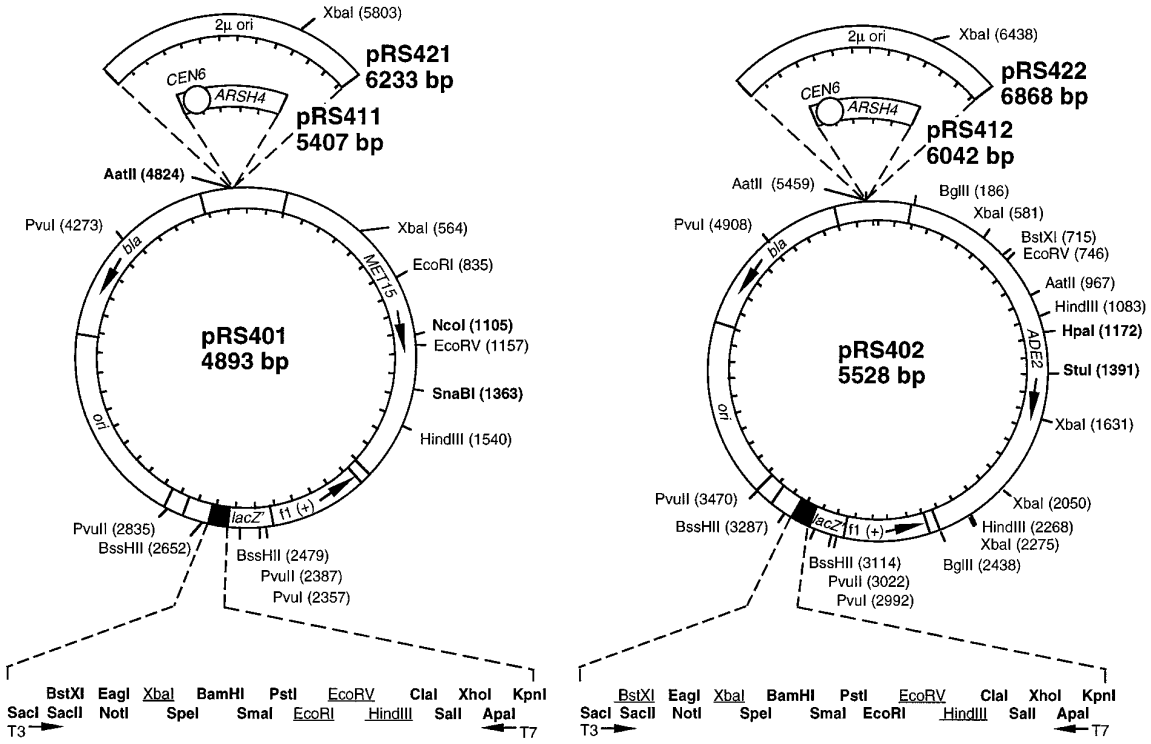


Figure 6. New pRS400 series *S. cerevisiae–E. coli* plasmid vectors. Restriction maps of the pRS vectors. Unique restriction sites are shown in bold letters. For emphasis, sites in the polylinker that are no longer unique have been underlined. Maps of the pRS40X set may be converted to either the pRS41X set by the insertion of the *CEN/ARS* cassette or to the pRS42X set by insertion of the 2 μ cassette (in either case, the *Aat*II site is destroyed). The numbering system is the same for all three sets. The direction of T3 or T7 polymerase transcription is as labeled. Numbers of parentheses refer to base pairs. Maps are drawn to scale.

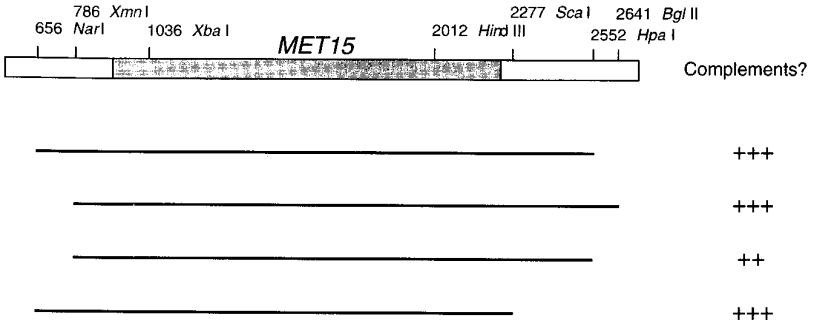


Figure 7. Definition of the minimum complementing region of *MET15*. A schematic diagram of the *MET15* genomic region indicating relevant restriction sites. Lines below the restriction site diagram indicate the extent of the fragment tested for complementation of a *met15* strain, and the column at the right indicates a qualitative analysis of the ability of each fragment to complement the *met15* allele for growth. In all cases, the fragments were cloned into the polylinker of pRS413 and yeast (strain 31-10C; Cost and Boeke, 1996) transformants were selected on SD+Ura medium, and complementation analyses were performed by comparing colony sizes on selective medium between a wild-type *MET15*$^+$ strain and the transformed strain.

© 1998 John Wiley & Sons, Ltd.

YM95NC Attachment -- Page 25

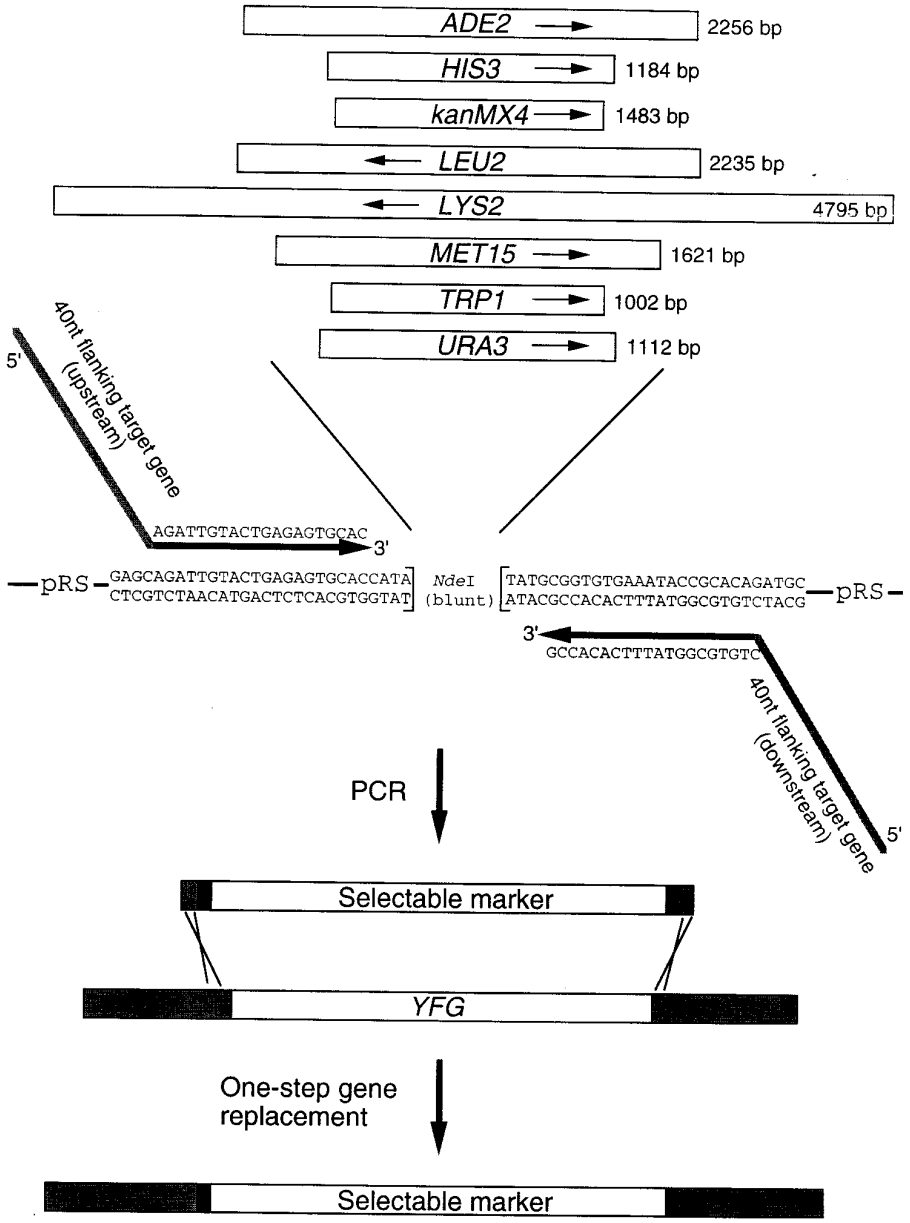


Figure 8. A universal primer set for producing PCR-mediated gene disruption cassettes. The double-stranded sequence shown represents about 50 bp surrounding a unique *Nde*I site in the parental plasmid backbone of all pRS400 series vectors [the parental backbone is a pBLUESCRIPT/pBLUESCRIBE hybrid made by ligating a *Pvu*I fragment of pBLUESCRIPT II–SK (containing the polylinker region) to a *Pvu*I fragment of pBLUESCRIBE (containing unique *Nde*I and *Aat*II sites)]. Minimal DNA segments that encode each of the selectable markers (arrows indicate transcriptional orientation) shown above were blunt-end ligated into the blunted *Nde*I site for each pRS400 series vector. A single pair of oligos can therefore be used to amplify by PCR each of the selectable marker genes. Sequences of oligos used for amplification of the selectable marker genes are as indicated. Both oligos are drawn 5′ to 3′. As indicated, 40 nts of sequence from either the upstream or downstream flanking region of the gene to be deleted, in this case YFG, is added to the 5′ end of each oligo. PCR-amplification of the auxotrophic marker allele is performed using any pRS integrating vector as a template (this is possible because the oligos are specific to the sequence just upstream and downstream of the *Nde*I vector site into which all markers were introduced). The resulting double-stranded PCR product is then transformed into yeast, replacing the genomic YFG allele by a double cross-over event. The resulting disruptions contain any of eight selectable markers transcribed in the orientations indicated.

© 1998 John Wiley & Sons, Ltd.

ACKNOWLEDGEMENTS

We thank our colleagues in the yeast community for providing information on alleles and for plasmids. We thank Hugh Rienhoff, Fred Winston, Marc Vidal and Mike Fasullo for helpful discussions and Kexin Yu for technical assistance. We thank the NIH for supporting research in our laboratories.

REFERENCES

Aparicio, O. M., Billington, B. L. and Gottschling, D.E. (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell* **66,** 1279–1287.

Åström, S. and Byström, A. (1993). The yeast initiator $tRNA^{Met}$ can act as an elongator $tRNA^{Met}$ *in vivo*. *J. Mol. Biol.* **233,** 43–58.

Ausubel, F., Brent, R., Kingston, R. E., *et al*. (1987). *Current Protocols in Molecular Biology*. John Wiley and Sons, New York.

Baudin, A., Ozier-Kalogeropoulous, O., Deanouel, A., LaCroute, F. and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucl. Acids Res.* **21,** 3329–3330.

Boeke, J. D., Lacroute, F. and Fink, G. R. (1984). A positive selection for mutants lacking orotidine 5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molec. Gen. Genet.* **197,** 345–346.

Brachmann, C. B., Sherman, J. M., Devine, S. E., *et al*. (1995). The *SIR2* gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes Dev.* **9,** 2888–2902.

Byström, A. S. and Fink, G. R. (1989). Molecular and functional analysis of the methionine initiator tRNA genes (*IMT*) in yeast. *Mol. Gen. Genet.* **216,** 276–286.

Christianson, T. R., Sikorski, R., Dante, M., Schero, J. and Hieter, P. (1992). Multifunctional yeast high copy-number shuttle vectors. *Gene* **110,** 119–122.

Cost, G. C. and Boeke, J. D. (1996). A useful colony colour phenotype associated with the yeast selectable/counterselectable marker *MET15*. *Yeast* **12,** 939–941.

Engebrecht, J. and Roeder, G. S. (1990). *MER1*, a yeast gene required for chromosome pairing and genetic recombination, is induced in meiosis. *Mol. Cell. Biol.* **10,** 2379–2389.

Fasullo, M. T. and Davis, R. W. (1988). Direction of chromosome rearrangements in *Saccharomyces cerevisiae* by use of *his3* recombinational substrates. *Mol. Cell. Biol.* **8,** 4370–4380.

Gaber, R. and Culbertson, M. R. (1982). The yeast frameshift suppressor gene SUF16-1 encodes an altered glycine tRNA containing the four-base anticodon 3′CCCG-5′. *Gene* **19,** 163–172.

Hinnen, A., Hicks, J. B. and Fink, G. R. (1979). Transformation of yeast. *Proc. Natl. Acad. Sci. USA* **75,** 1929–1933.

Keeney, J. B. and Boeke, J. D. (1994). Efficient targeted integration at *leu1-32* and *ura4-294* in *Schizosaccharomyces pombe*. *Genetics* **136,** 849–856.

Lorenz, M. C., Muir, R. S., Lim, E., *et al*. (1995). Gene disruption with PCR products in *Saccharomyces cerevisiae*. *Gene* **158,** 113–117.

Oliver, S. G., van der Aart, Q. J. M., Agostoni-Carbone, M. L., *et al*. (1992). The complete DNA sequence of yeast chromosome III. *Nature* **357,** 38–46.

Ono, B., Ishii, N., Fujino, S. and Aoyama, I. (1991). Role of hydrosulfide ions in methylmercury resistance in *Saccharomyces cerevisiae*. *Appl. Env. Microbiol.* **57,** 3183–3186.

Ronne, H. and Rothstein, R. (1988). Mitotic sectored colonies: evidence of heteroduplex DNA formation during direct repeat recombination. *Proc. Natl. Acad. Sci. USA* **85,** 2696–2700.

Rose, M. and Winston, F. (1984). Identification of a Ty insertion within the coding sequence of the *S. cerevisiae URA3* gene. *Molec. Gen. Genet.* **193,** 557–560.

Rose, M. D., Winston, F. and Hieter, P. (1990). *Methods in Yeast Genetics. A Laboratory Course Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Rothstein, R. J. (1983). One-step gene disruption in yeast. *Meth. Enzymol.* **101,** 202–211.

Scherer, S. and Davis, R. W. (1979). Replacement of chromosome segments with altered DNA sequences constructed *in vitro*. *Proc. Natl. Acad. Sci. USA* **76,** 4951–4955.

Sikorski, R. S. and Boeke, J. D. (1990). *In vitro* mutagenesis and plasmid shuffling: from cloned gene to mutant yeast. *Meth. Enzymol.* **194,** 302–318.

Sikorski, R. S. and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122,** 19–27.

Singh, A. and Manney, T. R. (1974). Genetic analysis of mutations affecting growth of *Saccharomyces cerevisiae* at low temperatures. *Genetics* **77,** 651–659.

Stotz, A. and Linder, P. (1990). The *ADE2* gene from *Saccharomyces cerevisiae*: sequence and new vectors. *Gene* **95,** 91–98.

Struhl, K. (1985). Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. USA* **82,** 8419–8423.

Wach, A., Brachat, A., Poehlmann, R. and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10,** 1793–1808.

Winston, F., Dollard, C. and Ricupero-Hovasse, S. L. (1995). Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* **11,** 53–55.

© 1998 John Wiley & Sons, Ltd.

Dharmacon™
RNAi, Gene Expression & Gene Editing

Product Insert

YKO Parental Strains

Cat. #YSC1048, YSC1049, YSC1050, YSC1061, YSC1062

A nearly complete set of yeast open reading frame (ORF) knock-outs has been produced by the *Saccharomyces* Genome Deletion Project (SGDP) using five Yeast Knockout (YKO) Parental Strains derived from *Saccharomyces cerevisiae* S288C.[1] A PCR-based strategy was used to replace each ORF with a KanMX cassette containing unique tags, "barcodes", for each deletion. Four different mutant collections have been generated: haploids of mating types MATa and MATalpha, homozygous diploids for non-essential genes, and heterozygous diploids, which contain the essential and non essential ORFs.

Storage:

- 4 °C for up to one week
- -80 °C indefinitely

Product Description:

Culture of *S. cerevisiae* in YPD broth with an inert growth indicator + 15% glycerol.

Genotypes of Specific Yeast Reference Strains[1]:

BY4730	MATa leu2Δ0 met15Δ0 ura3Δ0
BY4739	MATα leu2Δ0 lys2Δ0 ura3Δ0
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
BY4742	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0
BY4743	4741/4742

Homozygous diploids are in the BY4743 background unless 4730/4739 is indicated.

Useful Websites and References:

Saccharomyces Genome Deletion Project, click here.

1. Winzeler, E.A., et al., Functional characterization of the *Saccharomyces cerevisiae* Genome by gene deletion and parallel analysis, *Science, 285*:901-906, 1999.

GE, imagination at work and GE monogram are trademarks of General Electric Company. Dharmacon is a trademark GE Healthcare companies. All other trademarks are the property of General Electric Company or one of its subsidiaries. ©2014 General Electric Company—All rights reserved. Version published June 2014. GE Healthcare UK Limited, Amersham Place, Little Chalfont, Buckinghamshire, HP7 9NA, UK

GE Healthcare

Orders can be placed at:
gelifesciences.com/dharmacon

Customer Support: cs.dharmacon@ge.com
Technical Support: ts.dharmacon@ge.com or
1.800.235.9880; 303.604.9499 if you have any questions.

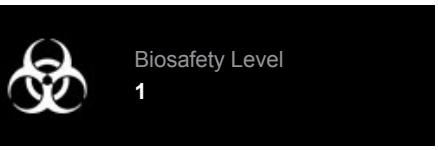
V1-0614

ATCC

Product Sheet

Saccharomyces cerevisiae (ATCC® 200901™)

Please read this FIRST



Intended Use

This product is intended for research use only. It is not intended for any animal or human therapeutic or diagnostic use.

Citation of Strain

If use of this culture results in a scientific publication, it should be cited in that manuscript in the following manner: *Saccharomyces cerevisiae* (ATCC® 200901™)

American Type Culture Collection
PO Box 1549
Manassas, VA 20108 USA
www.atcc.org

800.638.6597 or 703.365.2700
Fax: 703.365.2750
Email: Tech@atcc.org

Or contact your local distributor

Description

Strain Designation: BY4739 [ATCC 4040003]
Deposited Name: *Saccharomyces cerevisiae* Hansen, teleomorph
Genotype: MATalpha leu2delta0 lys2delta0 ura3delta0

Propagation

Techniques for cultivation and preservation of yeasts at ATCC are structured to conserve the genotype of a strain as it was deposited, including mutations specified by the depositor. To minimize selection pressure for revertants, strains are grown on media rich in amino acids and other nutrients. Strains are preserved by storage in liquid nitrogen vapor, the method most likely to stabilize genotype over long periods of time.

However, you have received this strain containing millions of cells. Some cells in this population may have genotypes different from that specified in the reference guide. Within a given culture, the exact proportion of cells possessing the designated genotype can vary**. Customers are advised to subject strains to suitable selection in order to recover a population containing the designated genotype in high frequency**. Methods for selection are often detailed in the literature cited by the reference guide. Although ATCC provides literature citations for specific strains whenever possible, it is the responsibility of the customer to obtain such literature.

ATCC® Medium 1245: YEPD

Growth Conditions
Temperature: 25.0°C

Notes

Additional information on this culture is available on the ATCC web site at www.atcc.org.

References

References and other information relating to this product are available online at www.atcc.org.

Biosafety Level: 1

Appropriate safety procedures should always be used with this material. Laboratory safety is discussed in the current publication of the *Biosafety in Microbiological and Biomedical Laboratories* from the U.S. Department of Health and Human Services Centers for Disease Control and Prevention and National Institutes for Health.

ATCC Warranty

The viability of ATCC® products is warranted for 30 days from the date of shipment, and is valid only if the product is stored and cultured according to the information included on this product information sheet. ATCC lists the media formulation that has been found to be effective for this strain. While other, unspecified media may also produce satisfactory results, a change in media or the absence of an additive from the ATCC recommended media may affect recovery, growth and/or function of this strain. If an alternative medium formulation is used, the ATCC warranty for viability is no longer valid.

Disclaimers

This product is intended for laboratory research purposes only. It is not intended for use in humans.

While ATCC uses reasonable efforts to include accurate and up-to-date information on this product sheet, ATCC makes no warranties or representations as to its accuracy. Citations from scientific literature and patents are provided for informational purposes only. ATCC does not warrant that such information has been confirmed to be accurate.

This product is sent with the condition that you are responsible for its safe storage, handling, and use. ATCC is not liable for any damages or injuries arising from receipt and/or use of this product. While reasonable effort is made to insure authenticity and reliability of strains on deposit, ATCC is not liable for damages arising from the misidentification or misrepresentation of cultures.

Please see the enclosed Material Transfer Agreement (MTA) for further details regarding the use of this product. The MTA is also available on our Web site at www.atcc.org

ATCC

Product Sheet

Saccharomyces cerevisiae (ATCC® 200901™)

Please read this FIRST

Biosafety Level
1

Intended Use

This product is intended for research use only. It is not intended for any animal or human therapeutic or diagnostic use.

Citation of Strain

If use of this culture results in a scientific publication, it should be cited in that manuscript in the following manner: *Saccharomyces cerevisiae* (ATCC® 200901™)

American Type Culture Collection
PO Box 1549
Manassas, VA 20108 USA
www.atcc.org

800.638.6597 or 703.365.2700
Fax: 703.365.2750
Email: Tech@atcc.org

Or contact your local distributor

Additional information on this culture is available on the ATCC web site at www.atcc.org.
© ATCC 2013. All rights reserved. ATCC is a registered trademark of the American Type Culture Collection. [05/06]

FEMS Yeast Research, 15, 2015, fov023

doi: 10.1093/femsyr/fov023
Advance Access Publication Date: 12 May 2015
Minireview

MINIREVIEW

Mitochondrial genome evolution in yeasts: an all-encompassing view

Kelle C. Freel, Anne Friedrich and Joseph Schacherer*

Department of Genetics, Genomics and Microbiology, University of Strasbourg/CNRS, UMR7156 Strasbourg 67083, France

*Corresponding author: Department of Genetics, Genomics and Microbiology, University of Strasbourg/CNRS, UMR7156, 28 rue Goethe Strasbourg 67083, France. Tel: 00-33-3-68-85-18-21; Fax: 00-33-3-68-85-20-28; E-mail: schacherer@unistra.fr

One sentence summary: In this review, the authors summarize the current state of yeast mitogenomics.
Editor: Jens Nielsen

ABSTRACT

Mitochondria are important organelles that harbor their own genomes encoding a key set of proteins that ensure respiration and provide the eukaryotic cell with energy. Recent advances in high-throughput sequencing technologies present a unique opportunity to explore mitochondrial (mt) genome evolution. The Saccharomycotina yeasts have proven to be the leading organisms for mt comparative and population genomics. In fact, the explosion of complete yeast mt genome sequences has allowed for a broader view of the mt diversity across this incredibly diverse subphylum, both within and between closely related species. Here, we summarize the present state of yeast mitogenomics, including the currently available data and what it reveals concerning the diversity of content, organization, structure and evolution of mt genomes.

Keywords: yeast; mitochondrial genomes; genomics; population genomics

INTRODUCTION

Mitochondria play an essential role in multiple cellular processes including the production of energy (Hatefi 1985; Amunts *et al.* 2014), apoptosis (Green and Reed 1998; Chandra, Samali and Orrenius 2000) and the generation of reactive oxygen species (ROS) (Starkov 2008). Long considered remnants of α-proteobacteria symbionts (Altmann 1890; Wallin 1927), they harbor their own DNA, encoding genes necessary for the production of the oxidative phosphorylation (OXPHOS) complexes that span the mt inner membrane (Lang, Gray and Burger 1999). The small size of the mtDNA reflects the fact that the genes it harbors are not sufficient for the production of the OXPHOS pathway, indeed many of the key components are instead found in the nuclear genome. This is because that over time most of the mt genes have been transferred to the nucleus or lost (Amunts *et al.* 2014). This ongoing process has shaped the evolution of nuclear genomes and appears to have a species-specific distribution in yeasts (Sacerdot *et al.* 2008). Indeed, it has been hypothesized that the observed divergence, in the now nuclear-encoded genes, has actually contributed to speciation in *Saccharomyces* spp. (Chou and Leu 2010).

The existence of mtDNA was first inferred in the 1940s in *Saccharomyces cerevisiae* as a heritable cytoplasmic element called the rho factor (Ephrussi, Hottinguer and Tavlitzki 1949), and was later formally recognized in chicken cells (Nass and Nass 1963) and yeast (Schatz, Haslbrunner and Tuppy 1964). Relatively recently, sequencing of both mt and nuclear genomes has become rapid and affordable. The main organism that has emerged as the focus of mt studies is *S. cerevisiae* (Solieri 2010), with its mtDNA completely sequenced in 1998 (Foury *et al.* 1998). This universally recognized species is easy to manipulate in the laboratory making it an ideal model for a variety of genetic studies. There is an estimated 50 to 200 copies of mtDNA in the haploid cell of *S. cerevisiae* (Solieri 2010), which is approximately 15% of the total DNA content (Williamson 2002). However, this varies among yeast isolates and also depends on the environment the cell is in (Moraes 2001; Hori *et al.* 2009). Additionally, contingent on if the cell is in stationary phase or grown anaerobically,

Received: 19 December 2014; **Accepted:** 4 May 2015
© FEMS 2015. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

the nucleoids (aggregates of multiple mt genomes) contain between 1.5 and 20 copies of the genome (Miyakawa *et al*. 1984; Nosek *et al*. 2006). Interestingly, *S. cerevisiae* can grow without functional mitochondria, in which case it produces respiratory-deficient smaller colonies, termed '*petite*' mutants, allowing for the study of the repercussions of a variety of severe mitochondrial defects.

Indeed *S. cerevisiae* has been used to model mt-tRNA mutations in humans (Tuppen *et al*. 2010) since mt transformation is possible (Bonnefoy and Fox 2007). Mitochondria are, in fact, implicated in a variety of human diseases (Holt, Harding and Morgan-Hughes 1988; Wallace 2005; Tuppen *et al*. 2010), and investigation of mutations in mitochondrial DNA (mtDNA), also principally using *S. cerevisiae*, has advanced the understanding of a variety of mt disorders (Mishra and Chan 2014). Furthermore, yeast mtDNA is biparentally inherited after mating, providing a means to study the role of recombination and inheritance in mt genome evolution. While *S. cerevisiae* is one of the most thoroughly studied eukaryotic organisms and has laid the groundwork for a suite of different studies, it is clearly not representative of all yeast lineages.

Yeast genetics has historically focused on *S. cerevisiae*; however, there have been attempts to move towards a broader view of yeast mt genomes. Currently, there are over 80 species within the Saccharomycotina that have mt genomes available (Fig. 1). Indeed, this vast amount of sequence data has demonstrated that there is incredible diversity in mt genome structure and organization. While interspecific analysis of mt genomes has allowed for a better understanding of the evolution of genomic architecture, intraspecific case studies provide a means to assess the extent to which selection pressures act on isolates within distinct lineages. In this brief review, we focus on the variability and evolution of mt genomes in yeasts, in particular the Saccharomycotina subphylum, and describe the existing knowledge concerning mtDNA, emphasizing the importance of expanding mt genome sequencing to include additional species.

YEAST AT THE FOREFRONT OF MITOCHONDRIAL COMPARATIVE AND POPULATION GENOMICS

Although some work was done on mt genomes in the 1980s, it wasn't until 1995 that the first Saccharomycotina mtDNA was sequenced from the species *Wickerhamomyces canadensis* (Sekito *et al*. 1995). A few years later the *S. cerevisiae* mt genome was completed (Foury *et al*. 1998); however, since that point, the availability of genomes from various species in the Saccharomycotina has grown, with 81 now publicly available (Fig. 1, Table S1, Supporting Information). These genomes represent a diverse range of yeast lineages (Fig. 1) and are an important dataset for the investigation mtDNA evolution.

While the majority of research on mt genome diversity has focused on *S. cerevisiae*, it is important to also expand studies to other diverse yeast genera in order to fully clarify mechanisms of evolution. There have been attempts to assess the roles that gene duplication and functional divergence have had on the evolution of the essential genes of the OXPHOS pathway found in the mt genome (Lavín *et al*. 2008; Marcet-Houben, Marceddu and Gabaldon 2009). These studies revealed that even in this essential pathway both gene loss and gene duplication, as well as subsequent functional divergence, are rampant (Marcet-Houben, Marceddu and Gabaldon 2009). In particular, mt ribosomal proteins appear to be quite distinct in the Saccharomycotina and Pezizomycotina subphylums (Arvas *et al*. 2007). In the Pezizomycotina, analysis of mt genomes has demonstrated a suite of interesting characteristics, including the presence of introns carrying ORFs encoding ribosomal proteins, plasmids that play a role in senescence, and finally diverse intergenic regions that affect replication, transcription and recombination (Pantou, Kouvelis and Typas 2008). The mt genomes of other fungi in the Ascomycetes, including the medically relevant genera *Aspergillus* and *Penicillium*, also represent interesting lineages for study that have implications in human health, since they are responsible for the production of antibiotics, enzymes and other pharmaceuticals (Joardar *et al*. 2012). Furthermore, mt genomes can be used to estimate both the time of divergence as well as the time since the last common ancestor of two sister clades (Jeffares *et al*. 2015). The few rare attempts to assess interspecies variation have revealed a core set of conserved genes and a wide diversity among accessory genes (Joardar *et al*. 2012), likely shuffled by horizontal gene transfer of mt plasmids and intron homing (Joardar *et al*. 2012).

Studying mt genomes provides a means to uncover diverse processes in genome evolution. Yeasts, specifically, present ideal model systems for mitogenomics research since they are easily manipulated in the laboratory and a broad diversity of different lineages is available for study. Ultimately, it is essential to recognize the importance of yeast mitogenomics for its implications in basic, applied and medical research (Lee *et al*. 2008; Chou and Leu 2010; Lipinski, Kaniak-Golik and Golik 2010; Albertin *et al*. 2013; Mishra and Chan 2014).

MITOCHONDRIAL GENOME GENE CONTENT

Yeast mtDNA encodes a set of 30 to 40 genes including those that code for proteins, the small (*rns*) and large (*rnl*) rRNA subunits, and approximately 24 tRNA genes. The core set of coding genes which are found in all yeast mt genomes are comprised of the subunits of ATP synthase (*atp*6, *atp*8 and *atp*9), units of the cytochrome c oxidase (*cox*1, *cox*2 and *cox*3) and apocytochrome b (*cob*) (Zamaroczy and Bernardi 1986; Wolf and Giudice 1988; Foury *et al*. 1998; Solieri 2010), all of which are involved in oxidative phosphorylation. Additionally, the mt genome can also include the genes coding for the seven NADH: ubiquinone oxidoreductase subunits (*nad*1-6 and *nad*4-*L*) which are part of complex I, the gene coding for the ribosomal protein *var*1 and finally the RNA subunit of the mt RNaseP (*rpm*1). Interestingly, the presence and absence of both the NADH and *var*1 genes are consistent with the main phylogenetic clusters (Fig. 1). The *var*1 gene was lost from all genomes in the CTG group (Fig. 1), which are yeasts which have a CUG codon that encodes serine, representing a deviation in the genetic code (Massey *et al*. 2003). The *var*1 gene is also missing from the extremely compact genome of *Hanseniaspora uvarum*, which has a notably unique organization (Pramateftaki *et al*. 2006). Moreover, none of the NADH subunits are found in the species of the *Saccharomycetaceae* (Dujon 2010), indicating that the deletion of these genes occurred at some point during the early evolution of the lineage (Dujon 2010; Prochazka *et al*. 2010). There is a patchy distribution of these subunits in other species, in particular they are absent from *Ogataea philodendron*, *Wickerhamomyces pijperi*, *H. uvarum* and *Starmerella bacillaris* (Fig. 1). Interestingly, all of the NADH genes are found in *Dekkera bruxellensis,* which demonstrated that the loss of complex I is not needed to develop petite-positive, Crabtree-positive and anaerobic characteristics (Prochazka *et al*. 2010). Incidentally, we did not indicate the presence or absence of *rpm*1 in the

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

Figure 1. Phylogenetic relationships among the 81 yeast species whose mt genomes have been entirely sequenced and annotated. The whole-genome duplication (WGD) is indicated with a star. The tree was constructed with phyML (Guindon *et al.* 2010), using the Mtrev model of amino acid substitution (Adachi and Hasegawa 1996). The phylogeny is based on the concatenated alignments of the seven core protein-coding genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2* and *cox3*). The bootstrap values (100 replicates run) are listed for all nodes >50. The currently accepted clades are overlaid on the phylogeny. To the right of the tree, the presence or absence of *var1* and the NADH genes is indicated with an orange and blue bar, respectively. The GC content is then listed for each genome (both using a color map as well as listing the actual data). Finally, the bar graph scheme of the genome content demonstrates the contribution to genome size by CDS and non-coding RNAs, introns and intergenic regions.

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

81 genomes surveyed as this gene is consistently poorly annotated, and analysis might not be reliable.

Since mt gene content is well conserved, concatenated sets of mt protein-coding genes have been used as phylogenetic markers (Bullerwell *et al.* 2003; Rokas *et al.* 2003; Spatafora *et al.* 2006; Hibbett *et al.* 2007; Pantou, Kouvelis and Typas 2008). The multilocus phylogenies based on these mt genes, while largely representative of the accepted relationships, do not correlate identically with species trees (Belloch *et al.* 2000; Liu *et al.* 2008; Joardar *et al.* 2012). The phylogenetic tree based on the concatenated protein sequences coded by the set of core genes found in the 81 currently publically available mt genomes (Fig. 1) is no exception: an overall conservation in the relationships with nuclear DNA is observed, while some inconsistencies can be noticed. The most apparent difference is that in addition to the lack of resolution in the root of the phylogeny, surprisingly one of the basal branches harbors the *Dekkera* cluster, normally found more closely located to the Saccharomycetaceae as was demonstrated previously (Dujon 2010). Another difference in the phylogeny is the location of the CTG group in relation to the Dipodascaceae, which have positions that are essentially inverted in the mt tree (Fig. 1) compared to previous studies based on nuclear DNA (Dujon 2010). These changes in the phylogeny concur with what has been previously found, indeed there is discordance between the evolutionary history of the mtDNA and nuclear DNA. Ultimately, the differentiation and evolution of individual species is the product of the divergence in both the mt and nuclear genomes (Lee *et al.* 2008; Chou and Leu 2010).

GENOME SIZE VARIABILITY IS DRIVEN BY DIFFERENCES IN INTRONS AND INTERGENIC REGIONS

Despite a high level of gene content conservation, there is a striking divergence among mt genome sizes. Indeed, the lengths range from the relatively small 18.5 kb genome of *H. uvarum* (Pramateftaki *et al.* 2006) to more than 85 kb in *S. cerevisiae* (Foury *et al.* 1998), and over 105 kb in *Nakaseomyces bacillisporus* (Bouchier *et al.* 2009). Interestingly, a similar pattern is seen in other eukaryotes (Gray, Burger and Lang 1999; Nosek and Tomaska 2003), including plants (Ward, Anderson and Bendich 1981; Alverson *et al.* 2010) where sizes can range from 222 to 773 kb within a family. The general overview of mt genome characteristics among species in the Saccharomycotina subphylum offered by Figs 1 and 2A demonstrates that there is no link between mt genome size and phylogeny. Some patterns can be observed with the *Saccharomyces* genus harboring a larger than average mt genome (an average of 76 kb vs 41 kb, respectively) and the protoploid species generally with smaller genomes (an average of 32 kb). Indeed, among the *Nakaseomyces* species, mt genome size ranges from 20 kb in *Candida glabrata* to 107 kb in *Kluyveromyces bacillisporus* (Bouchier *et al.* 2009). While not as dramatic, changes in length are also seen among genomes in the *Lachancea* clade, composed of a range of closely related lineages, with the smallest genome harbored by *Lachancea thermotolerans* (23.5 kb) and the largest over twice this length found in *Lachancea kluyveri* (51.5 kb) (Talla *et al.* 2005; Friedrich *et al.* 2012). This pattern is also seen among *Candida* species, with genomes ranging from 28 kb in *C. phangngensis* to 48.5 kb in *C. galli* (Gaillardin *et al.* 2012). Furthermore, intraspecific differences within the *Lachancea* clade are observed as well, with genomes among *L. thermotolerans* isolates found to be from 21.8 to 24.9 kb (Freel *et al.* 2014) and those of *L. kluyveri* from 50.1 to 53.7 kb (Jung *et al.* 2012).

The basal driver behind the differences in mt genome size is not the length of the CDS, but instead variability in the non-coding sequences (Fig. 2), as has been highlighted previously (Bouchier *et al.* 2009; Friedrich *et al.* 2012; Jung *et al.* 2012). There are two types of non-coding sequences, which contribute differentially to the changes in length: the intergenic regions and the introns (Fig. 2B). In fact, as can be inferred from Fig. 2, the intergenic regions are the most influential factors accounting for approximately 90.5% of the variance in length, in contrast to the 7 and 2% accounted for the introns and CDS, respectively. As expected, the intergenic regions are thus highly correlated to the total mt genome length, ($r^2 = 0.841$, Fig. 2D), and vary from 1 985 to 94 009 bp (with a median of 9 068 bp, Fig. 2B). The proportion of intergenic regions in relation to the whole mt genome ranges from 7.8% as found in *Millerozyma sorbitophila* (Jung *et al.* 2009) to an amazing 88% in *N. bacillisporus* (Fig. 1, Table S1, Supporting Information). Repetitive elements are found in the intergenic regions of nearly all species, which include both stretches of AT sequences and GC-rich clusters thought to enhance recombination in mtDNA (Bernardi 2005). An exception has been identified in *C. glabrata* for which the GC clusters are missing (Koszul *et al.* 2003). The proportion of AT to GC regions varies among different yeasts. For example, long AT spacers compose approximately 50% of *S. cerevisiae* mtDNA and the GC clusters just 5%. However, in some *Nakaseomyces* species, the number of GC clusters ranges from 0 to 218 (Bouchier *et al.* 2009), leading to the 5-fold change in size observed between the largest (107.1 kb in *K. bacillisporus*) and smallest genomes (20 kb in *C. glabrata*) (Bouchier *et al.* 2009). Research focusing on the clinically relevant *Candida albicans* demonstrated that phylogenetic classification based on highly variable mt intergenic regions revealed geographically distributed polymorphisms and in fact, provided a complementary method to current multilocus sequence typing (Bartelli *et al.* 2013). Similarly, the large intergenic regions found in *L. kluyveri* account for the dramatic difference in size between this lineage and its sister species *L. thermotolerans*, in which the intergenic regions represent up to 41 and 22.3% of the mt genome, respectively. Even between isolates from a single species, such as *L. kluyveri* there are changes in the intergenic content, which constitutes from 28 to 41% of the mt genome (Jung *et al.* 2012).

The second type of non-coding sequences considered are the introns. While their number and length are highly variable among species, they impart a minimal contribution to the variance in mt genome length (approximately 7%) and not surprisingly, there is no direct correlation between their length, and genome size ($r^2 = 0.044$, Fig. 2C). Among the Saccharomycotina, mt genomes harbor from 0 to 17 introns, and have a cumulative length ranging from 0 to 21 030 bp, (median of 5 455 bp, Fig. 2B). The introns in mtDNA are organized into two groups (I and II) (Lang, Laforest and Burger 2007), with group I introns classified into four families depending on their secondary structure and protein motifs (LAGLIDADG, GIY-YIG, HNH and His-Cys box), and group II introns encoding reverse transcriptases (Haugen, Simon and Bhattacharya 2005; Lang, Laforest and Burger 2007). The introns can be linked with an upstream exon or free standing, and generally are located in or around specific genes (Haugen, Simon and Bhattacharya 2005). The group I introns act as mobile elements that can be gained, lost or inverted, explaining variation in mtDNA size among isolates even within species (Dujon 1989; Foury *et al.* 1998). Among the Pezizomycotina, the intron content appears to be the main cause of differences in mt genome length (Joardar *et al.* 2012). As there is a paucity of interspecific studies, it is difficult to determine if more variability is seen within or

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

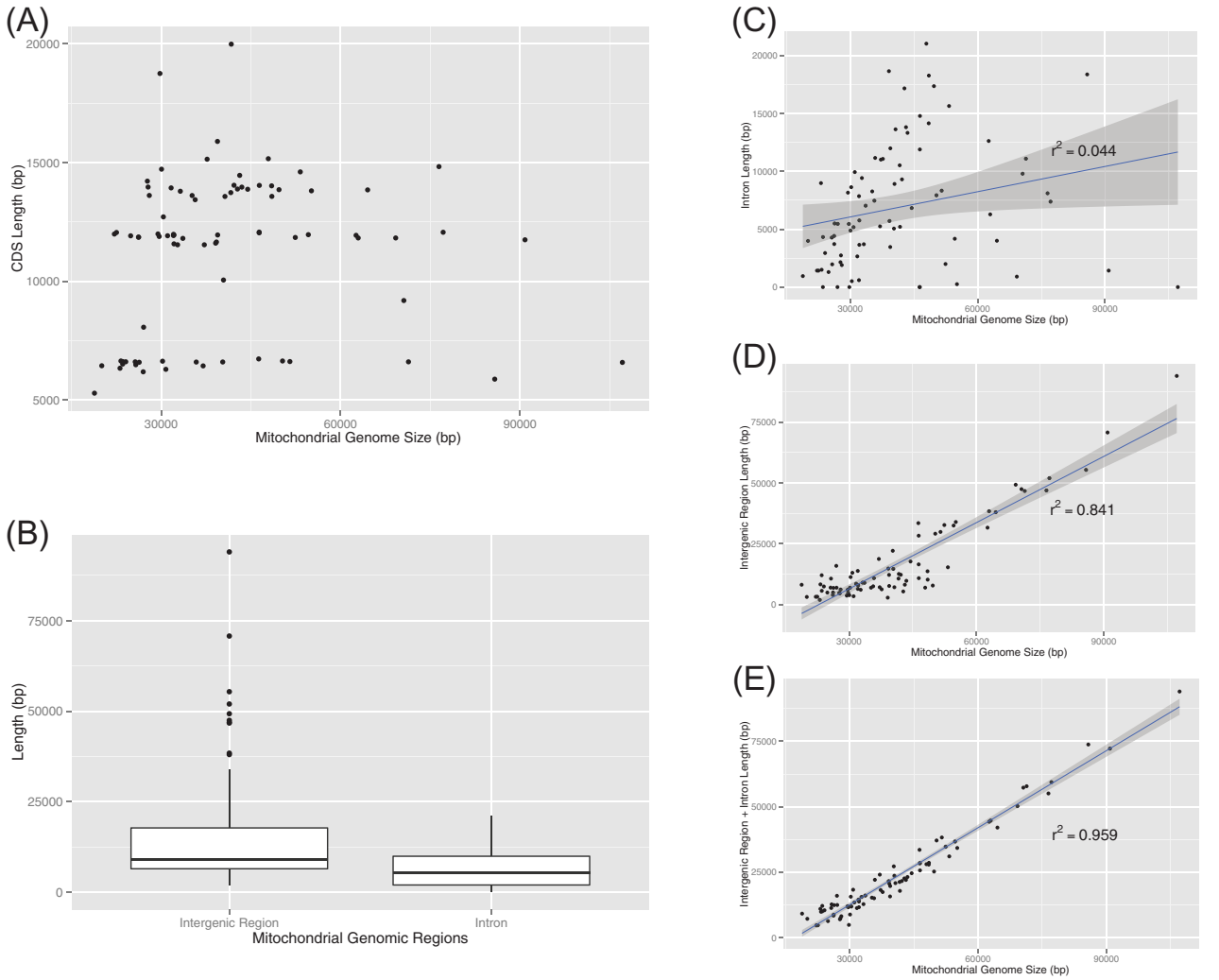


Figure 2. Graphs demonstrating the contribution that CDS length, introns and intergenic regions play in determining mt genome size, using the data available data for the 81 yeast mt genomes. (**A**) The CDS length compared to the overall mt genome size (bp), clearly demonstrating no correlation between the two. (**B**) Box plot depiction of the length of the variance and median values for both the intergenic regions and introns in mt genomes. Correlation demonstrated between the length of the mt genome and intron length (**C**), intergenic region length (**D**) and finally the combined intergenic region and intron length (**E**).

between species. For example, in the *Candida* clade, *C. alai* harbors one intron while *C. frijolesensis* has ten (Valach *et al.* 2011); clearly this contributes to the size difference in mtDNA among species. However, it must be noted that there is an inherent bias in these findings since the number of introns also varies among isolates of a single species (Jung *et al.* 2012; Freel *et al.* 2014) and only individual representatives were chosen for comparison. Finally, as expected, it is worth highlighting that when the calculation is performed using the combined length of the intergenic regions as well as the introns (Fig. 2E), the correlation is even stronger ($r^2 = 0.959$). Clearly, there is variation in the evolution of the non-coding regions within and among different species. While there is consistently variation in mt genome sizes, there are different degrees to which the synteny of gene arrangement is observed at both the inter- and intraspecies levels.

DISPARATE PATTERNS REVEALED BETWEEN INTRA- AND INTERSPECIES GENE SYNTENY

The range in genome size along with the mechanisms that shuffle introns and intergenic regions within the mt genome might suggest that mt genomes would lack gene synteny. However, different levels in conservation of genome organization are apparent throughout the Saccharomycotina. Indeed, among and within the species of the *Lachancea* (Friedrich *et al.* 2012; Jung *et al.* 2012; Freel *et al.* 2014), *Candida* (Kosa *et al.* 2006) and *Yarrowia* clades (Gaillardin *et al.* 2012) there is an extremely high degree of gene conservation and synteny. However, this is not the case between other closely related species, with variability observed among yeasts from the *Nakaseomyces* (Bouchier *et al.* 2009).

It appears that much of the incongruence in gene order is likely due to the inversion of short intergenic repeats and mobile introns (Prochazka *et al.* 2010). Interestingly, certain gene sets are consistently found together, including the most conserved region identified across multiple species including a cluster of *cox1-atp8-atp6*, in the mtDNA of species of the *Saccharomyces* complex (Pramateftaki *et al.* 2006) and *nad4L-nad5* which is conserved in all ascomycetes that have the genes for complex I (Groth, Petersen and Piskur 2000; Kosa *et al.* 2006; Solieri 2010). An analysis of *Pichia* species and other closely related lineages revealed the presence of large syntenic blocks (Jung *et al.* 2009). Indeed, while more distantly related species harbor

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

some minor differences (Jung *et al.* 2009) the intraspecific studies completed to date that have used a substantial population of isolates have both found complete synteny in gene order among the mtDNA (Jung *et al.* 2012; Freel *et al.* 2014). These differences in gene organization can, in some instances, be attributed to the circular topology of the genome, which might facilitate rearrangement.

VARIABILITY IN YEAST mtDNA TOPOLOGY

Not long after the discovery of mtDNA in yeast (Schatz, Haslbrunner and Tuppy 1964), a variety of structures in nature were identified including monomeric circles (Boore 1999), concatenated circles (Lukes *et al.* 2002; Roy *et al.* 2007), linear concatemers (Bendich 1993; Williamson 2002) and fragmented multipartite genomes (Palmer and Shields 1984; Voigt, Erpenbeck and Worheide 2008). Initial studies exploring yeast mtDNA assumed that the molecule existed solely as a circular element, which led to a false dogmatic theory that all yeast mtDNA was circular, with any linear molecules identified described as remnants of broken DNA (Williamson 2002). However, in 1991 PFGE demonstrated that although some circles were present, the majority of mtDNA in *S. cerevisiae* consists of polydisperse linear tandem arrays ranging from 75 to 150 kb in size (Ling and Shibata 2002). Another confounding aspect in understanding the evolution in mt topology is that there is asymmetry in segregation of mtDNA during cell division of *S. cerevisiae,* with daughter cells obtaining linear monomers of the genome. Ultimately these monomers are then amplified using a rolling-circle mechanism; however, mother cells generally harbor concatemeric and branched mtDNA (Ling and Shibata 2002).

The structure of genomes among species of yeast is not well conserved, and the evolutionary origin of this diversity is yet to be well understood. Various studies using *Candida* species as a model system have investigated this variability in genome architecture. In *C. glabrata* (as well as *S. cerevisiae*), linear concatamers dominate the type of topology among mtDNA, while in *C. albicans* it is present as a network of branches (Gerhold *et al.* 2010), and polydisperse molecules, which are possibly an artifact of recombination-driven replication (Bendich 2010; Gerhold *et al.* 2010). An additional study looking at three closely related *Candida* species revealed that even when gene synteny and organization is highly conserved (even with identical locations of tRNAs), the molecular forms differ. Indeed *C. parapsilosis* as well as *C. metapsilosis* have linear genomes while *C. orthopsilosis* has circular mtDNA. A careful inspection of the mt telomeres from these three species demonstrated that even with different topology, the telomeres have essentially the same organization (Kosa *et al.* 2006). This study concluded that the linear mt genome actually evolved from a circular-mapping form driven by the emergence of mt telomeres (Kosa *et al.* 2006) and an eventual defect in telomere maintenance. These findings were also supported in later work, which expanded the analysis to eight *Candida* species (Valach *et al.* 2011). Valach and colleagues identified an intriguing variability in genome architecture with forms ranging from circular to linear mapping and even multipartite forms, and suggested that the linear form evolved from an ancestral circular genome. Interestingly, it appears that genome topology actually changes depending on the life stage of the yeast cell, with more circles present in the bud as it forms, and linear concatemers detected at a higher concentration as it matures. This change in mt DNA topology is linked to both replication and recombination, which play essential roles at different points in the life cycle of the yeast cell.

REPLICATION AND RECOMBINATION IN YEAST mtDNA

Shortly after finding that mtDNA is often present in a linear form, it was suggested that mtDNA replicates through a rolling-circle mechanism (Maleszka, Skelly and Clark-Walker 1991). Within the intergenic sequences of species in the *Saccharomyces* genus, *ori* regions are located, which potentially play a key role in the initiation of replication. In *petite*-positive mtDNA, there are seven or eight regions, approximately 300 bp in length, harboring GC-rich regions and AT clusters as well as a promoter for *rpo*41 (a mt RNA polymerase) (Bernardi 2005). It has been hypothesized that mt diversity exists in a species-specific pattern linked to replication and segregation (Solieri 2010). Thus, it is essential to pursue additional analyses across and between lineages to clarify if strategies for replication are indeed unique to individual species. In *C. albicans*, possible replication origins have not been identified, but it appears that recombination drives mtDNA replication initiation (Gerhold *et al.* 2010). This finding also drew on the fact that mtDNA in *C. albicans* is found as a branched network, and the study concluded that topology is an essential factor in recombination-driven replication.

One clear difference between mammalian mtDNA and that of yeast is that it is biparentally inherited after a mating event, providing an opportunity for recombination to take place. Recombination in yeast mtDNA was first identified in 1974 (Dujon, Slonimski and Weill 1974) and occurs after mating when the mitochondria from the parental strains fuse with recombinant mtDNA then found in the diploids (Fritsch *et al.* 2014). This heteroplasmic state is fleeting, and only present in the cell for approximately 20 generations. The zygote that buds from the mid-point after mating is the most likely to possess a recombinant genotype. Previous research suggested that homologous recombination is induced after the repair enzyme Ntg1 creates a double-strand break in the DNA. Initial studies on interspecific consistency of mtDNA replication assessed recombination in the haploid mt genome of a set of *C. albicans* isolates and found that clonal proliferation of the mt genome was linked to clonal proliferation of the nuclear genome (Anderson *et al.* 2001); however, replication of the mtDNA occurs independently from the nuclear genome. Recently, mtDNA recombination was analyzed to assess the impact of intraspecies variability between a cross of *S. cerevisiae* isolates. The results indicated that the global recombination rate was approximately three events per kb (Fritsch *et al.* 2014), and the data generated allowed for the construction of the first genome-wide map of recombination in yeast. Interestingly, they uncovered 23 hotspots of recombination throughout the mt genome, 15 of which were located in intergenic regions, with only 3 found exclusively in the introns of *cox*1, *cob* and *var*1 (Fritsch *et al.* 2014). They concluded that recombination could, in fact, have a higher impact on genome reorganization than the mutation rate. The importance of the factors that affect changes in mtDNA over time might depend on the lineage in question; clearly more interspecies studies are needed to clarify this topic.

PATTERNS OF EVOLUTION IN mtDNA AND PHENOTYPIC IMPLICATIONS

Studies on yeasts, in particular concerning the Saccharomycotina, have advanced the knowledge on mt genome evolution. Lately, some studies have expanded the focus of mt research in order to include interspecific studies on single species as well as a broader diversity of yeast lineages, both of which are key

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

steps in elucidating patterns of genome evolution. The yeast *L. kluyveri* is *petite* negative, and thus is a good model for studying intraspecific mt genome evolution. The first extensive investigation of the genetic diversity and patterns of mtDNA evolution in a single species was undertaken using 18 isolates from this lineage (Jung *et al.* 2012). Among the strains analyzed, all harbored highly polymorphic mtDNA and the dN/dS ratios of the protein coding genes indicated that purifying selection was working on the genome. In fact, the only inconsistency observed was in the case of the *var*1 gene, which encodes a mt ribosomal protein of the small subunit (Hudspeth *et al.* 1982, 1983). Interestingly, *var*1 had a higher dN/dS ratio than the other genes, suggesting the functional constraints on this gene, which is less conserved and not found in all yeast mt genomes, might be unique (Jung *et al.* 2012). *L. thermotolerans*, a closely related species to *L. kluyveri* was the focus of an additional study including 50 isolates from diverse locations (Freel *et al.* 2014). These genomes were both smaller than in *L. kluyveri* and more highly conserved. Indeed, although these two lineages are closely related, stronger purifying selection and/or lower mutation rates appear to be driving the evolution of the mtDNA in *L. thermotolerans*. An additional study focused on the intergenic regions of mtDNA from a set of 21 *C. albicans* isolates also found structure linked to geographical location and suggested that the non-coding regions are under neutral evolution (Bartelli *et al.* 2013). Indeed, this highlighted that amplification of the highly variable intergenic regions of mtDNA could provide an ideal typing method for pathogens and afforded a glimpse into the microvariability of the mtDNA of a species in the CTG group. Among species of this clade, five were compared in order to characterize genome evolution. This analysis revealed a phylogeny with two distinct clades composed of either *Pichia* or *Candida* lineages, and also provided evidence for the delineation of *Pichia farinosa* and *P. sorbitophila* as different species (Jung *et al.* 2010). A study on multiple species in the *Lachancea* clade revealed consistently strong signatures of purifying selection (Friedrich *et al.* 2012) across all lineages. Purifying selection has apparently purged most indels and nonsynonymous differences from mt protein-coding genes among the genomes of the *Lachancea* clade. Additional studies are required to determine whether the results obtained are consistent with patterns of gene evolution in other groups.

Clarifying the relationship between genotype and phenotype in individuals from a single species is an essential step in elucidating the mechanisms of genome evolution. The mtDNA contributes to phenotypic diversity in yeast, which is in part generated by the interplay between the mt and nuclear genomes (Dowling, Friberg and Lindell 2008; Solieri *et al.* 2008) as many of the components for the OXPHOS pathway are located in the nuclear genome, as mentioned previously. Research has shown that variation in mt and nuclear genomes affects fitness in a diverse set of organisms ranging from insects (*Drosophila* and *Callosobruchus*) (Arnqvist *et al.* 2010) to copepods (*Tigriopus*) (Ellison and Burton 2008), and epistatic interactions between the mt and nuclear genomes have been implicated in the production of various phenotypes and possibly driving hybrid breakdown and speciation (Burton and Barreto 2012). While it is clear that mitochondria play an essential role in cell survival, the extent to which mt-nuclear DNA interactions affect phenotypic variation in yeast is not well understood. Within species of the *Saccharomyces* genus, incompatibilities between the mt and the nuclear DNA can cause respiratory deficiencies (Chou and Leu 2010), which could ultimately lead to a divergence between lineages. Recently, using *S. cerevisiae* as a model, it was demonstrated that interfering with mt-nuclear DNA interactions produced increased amounts of ROS and a variety of fitness defects (Paliwal, Fiumera and Fiumera 2014). Indeed, it appears that mt and nuclear epistasis has a large impact on determining yeast survival in different environments. Also, since these interactions apparently were not dependent on genetic distance, coevolution of the genomes can likely take place rapidly. A more thorough understanding of how mt and nuclear genomes interact to produce phenotypic changes will provide a better view of genome evolution and speciation and could potentially have implications in understanding mt diseases.

CONCLUSIONS AND PERSPECTIVES

It is essential to expand genomic research to a broad range of yeast species since genome organization and architecture are variable and the evolutionary reasons behind these differences are not well understood. Although *S. cerevisiae* has proven to be a helpful model for many studies, including those elucidating different aspects of mt genome function and structure as well as clarifying epistasis between the mt and nuclear DNA, additional species need to be included in future research. There is a clear lack of intraspecific mt genome data, and thus the extent to which mt genome organization, topology, diversity and selective pressure, vary between isolates of a single species is not clear. As the appreciation for the essential and complex role that mtDNA has in cell survival develops, understanding what changes in mt genomes affect selection pressures, phenotype and ultimately the evolution of the genome will be elucidated.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSYR online.

ACKNOWLEDGMENTS

The authors thank the Université de Strasbourg (IdEx 2012 Attractivité) and the Agence Nationale de la Recherche (2010-BLAN-1606-05 and 2011-JSV6-004-01) for their financial support.

Conflict of interest. None declared.

REFERENCES

Adachi J, Hasegawa M. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol* 1996;**42**:459–68.

Albertin W, da Silva T, Rigoulet M, *et al.* The mitochondrial genome impacts respiration but not fermentation in interspecific *Saccharomyces* hybrids. *PLoS One* 2013;**8**:e75121.

Altmann R. *Die Elementarorganismen und ihre Beziehungen zu den Zellen*. Liepzig, Germany: Veit & Comp., 1890.

Alverson AJ, Wei X, Rice DW, *et al.* Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol* 2010;**27**:1436–48.

Amunts A, Brown A, Bai XC, *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* 2014;**343**:1485–9.

Anderson JB, Wickens C, Khan M, *et al.* Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans*. *J Bacteriol* 2001;**183**:865–72.

Arnqvist G, Dowling DK, Eady P, *et al.* Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. *Evolution* 2010;**64**:3354–6.

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

Arvas M, Kivioja T, Mitchell A, *et al*. Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina. *BMC Genomics* 2007;**8**:325.

Bartelli TF, Ferreira RC, Colombo AL, *et al*. Intraspecific comparative genomics of *Candida albicans* mitochondria reveals non-coding regions under neutral evolution. *Infect Genet Evol* 2013;**14**:302–12.

Belloch C, Querol A, Garcia MD, *et al*. Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cyctochrome-c oxidase II gene. *Int J Syst Evol Micr* 2000;**50**:405–16.

Bendich AJ. Reaching for the ring: the study of mitochondrial genome structure. *Curr Genet* 1993;**24**:279–90.

Bendich AJ. The end of the circle for yeast mitochondrial DNA. *Mol Cell* 2010;**39**:831–2.

Bernardi G. Lessons from a small, dispensable genome: the mitochondrial genome of yeast. *Gene* 2005;**354**:189–200.

Bonnefoy N, Fox TD. Directed alteration of *Saccharomyces cerevisiae* mitochondrial DNA by biolistic transformation and homologous recombination. *Methods Mol Biol* 2007;**372**:153–66.

Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res* 1999;**27**:1767–80.

Bouchier C, Ma L, Creno S, *et al*. Complete mitochondrial genome sequences of three *Nakaseomyces* species reveal invasion by palindromic GC clusters and considerable size expansion. *FEMS Yeast Res* 2009;**9**:1283–92.

Bullerwell CE, Leigh CE, Forget L, *et al*. A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res* 2003;**31**:759–68.

Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Mol Ecol* 2012;**21**:4942–57.

Chandra J, Samali A, Orrenius S. Triggering and modulation of apoptosis by oxidative stress. *Free Radical Bio Med* 2000;**29**:323–33.

Chou JY, Leu JY. Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays* 2010;**32**:401–11.

Dowling DK, Friberg U, Lindell J. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol* 2008;**23**:546–54.

Dujon B. Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene* 1989;**82**:91–114.

Dujon B. Yeast evolutionary genomics. *Nat Rev Genet* 2010;**11**:512–24.

Dujon B, Slonimski PP, Weill L. Mitochondrial genetics XI: a model for recombination and segregation of mitochondrial genomes in *Saccharomyces cerevisiae*. *Genetics* 1974;**78**:415.

Ellison CK, Burton RS. Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *P Natt Acad Sci USA* 2008;**105**:15831–6.

Ephrussi B, Hottinguer H, Tavlitzki J. Action de l'acriflavine sur les levures. II. Etude genetique du mutant 'petite colonies'. *Ann Inst Pasteur* 1949;**79**:419–50.

Foury F, Roganti T, Lecrenier N, *et al*. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett* 1998;**32**:325–31.

Freel KC, Friedrich A, Hou J, *et al*. Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species *Lachancea thermotolerans*. *Genome Biol Evol* 2014;**6**:2586–94.

Friedrich A, Jung PP, Hou J, *et al*. Comparative mitochondrial genomics within and among yeast species of the *Lachancea* genus. *PLoS One* 2012;**7**:e47834.

Fritsch ES, Chabbert CD, Klaus B, *et al*. A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics* 2014;**198**:755–71.

Gaillardin C, Neuveglise C, Kerscher S, *et al*. Mitochondrial genomes of yeasts of the *Yarrowia* clade. *FEMS Yeast Res* 2012;**12**:317–31.

Gerhold JM, Aun A, Sedman T, *et al*. Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. *Mol Cell* 2010;**39**:851–61.

Gray MW, Burger G, Lang BF. Mitochondrial evolution. *Science* 1999;**283**:1476–81.

Green DR, Reed JC. Mitochondria and apoptosis. *Science* 1998;**281**:1309–12.

Groth C, Petersen RF, Piskur J. Diversity in organization and the origin of gene orders in the mitochondrial DNA molecules of the genus *Saccharomyces*. *Mol Biol Evol* 2000;**17**:1833–41.

Guindon S, Dufayard JF, Lefort V, *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307–21.

Hatefi Y. The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem* 1985;**54**:1015–69.

Haugen P, Simon DM, Bhattacharya D. The natural history of group I introns. *Trends Genet* 2005;**21**:111–9.

Hibbett DS, Binder M, Bischoff JF, *et al*. A higher-level phylogenetic classification of the Fungi. *Mycol Res* 2007;**111**:509–47.

Holt IJ, Harding AE, Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 1988;**331**:717–9.

Hori A, Yoshida M, Shibata T, *et al*. Reactive oxygen species regulate DNA copy number in isolated yeast mitochondria by triggering recombination-mediated replication. *Nucleic Acids Res* 2009;**37**:749–61.

Hudspeth MES, Ainley MW, Schumard DS, *et al*. Location and structure of the *var*1 gene on yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. *Cell* 1982;**30**:617–26.

Hudspeth MES, Vincent RD, Perlman PS, *et al*. Expandable *var*1 gene of yeast mitochondrial DNA: In-frame insertions can explain the strain-specific protein size polymorphisms. *P Natl Acad Sci USA* 1983;**81**:3148–52.

Jeffares DC, Raillis C, Rieux A, *et al*. The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nat Genet* 2015;**47**:235–41.

Joardar V, Abrams NF, Hosteler J, *et al*. Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genome size variability. *BMC Genomics* 2012;**13**:698.

Jung PP, Friedrich A, Reisser C, *et al*. Mitochondrial genome evolution in a single protoploid yeast species. *G3 (Bethesda)* 2012;**2**:1103–11.

Jung PP, Friedrich A, Souciet JL, *et al*. Complete mitochondrial genome sequence of the yeast *Pichia farinosa* and comparative analysis of closely related species. *Curr Genet* 2010;**56**:507–15.

Jung PP, Schacherer J, Souciet JL, *et al*. The complete mitochondrial genome of the yeast *Pichia sorbitophila*. *FEMS Yeast Res* 2009;**9**:903–10.

Kosa P, Valach M, Tomaska L, *et al*. Complete DNA sequences of the mitochondrial genomes of the pathogenic yeasts *Candida orthopsilosis* and *Candida metapsilosis*: insight into the evolution of linear DNA genomes from mitochondrial telomere mutants. *Nucleic Acids Res* 2006;**34**:2472–81.

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

Koszul R, Malpertuy A, Frangeul L, *et al.* The complete mitochondrial genome sequence of the pathogenic yeast *Candida (Torulopsis) glabrata*. *FEBS Lett* 2003;**534**:39–48.

Lang BF, Gray MW, Burger G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 1999;**33**:351–97.

Lang BF, Laforest MJ, Burger G. Mitochondrial introns: a critical view. *Trends Genet* 2007;**23**:119–25.

Lavín JL, Oguiza JA, Ramírez L, *et al.* Comparative genomics of the oxidative phosphorylation system in fungi. *Fungal Genet Biol* 2008;**45**:1248–56.

Lee H, Chou JY, Cheong L, *et al.* Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 2008;**135**:1065–73.

Ling F, Shibata T. Recombination-dependent mtDNA partitioning. In vivo role of Mhr1p to promote pairing of homologous DNA. *EMBO J* 2002;**21**:4730–40.

Lipinski KA, Kaniak-Golik A, Golik P. Maintenance and expression of the *S. cerevisiae* mitochondrial genome—from genetics to evolution and systems biology. *Biochim Biophys Acta* 2010;**1797**:1086–98.

Liu Y, Leigh JW, Brinkmann H, *et al.* Phylogenomic analyses support the monophyly of taphrinomycotina, including schizosaccharomyces fission yeasts. *Mol Biol Evol* 2008;**26**:27–34.

Lukes J, Guilbride DL, Votypka J, *et al.* Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot Cell* 2002;**1**:495–502.

Maleszka R, Skelly PJ, Clark-Walker GD. Rolling circle replication of DNA in yeast mitochondria. *EMBO J* 1991;**10**:3923–9.

Marcet-Houben M, Marceddu G, Gabaldon T. Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evol Biol* 2009;**9**:295.

Massey SE, Moura G, Beltrão P, *et al.* Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. *Genome Res* 2003;**13**:544–57.

Mishra P, Chan DC. Mitochondrial dynamics and inheritance during cell division, development and disease. *Nat Rev Mol Cell Bio* 2014;**15**:634–46.

Miyakawa I, Aoi H, Sando N, *et al.* Fluorescence microscopic studies of mitochondrial nucleoids during meiosis and sporulation in the yeast *Saccharomyces cerevisiae*. *J Cell Sci* 1984;**66**:21–38.

Moraes CT. What regulates mitochondrial DNA copy number in animal cells. *Trends Genet* 2001;**17**:199–205.

Nass MMK, Nass S. Intramitochondrial fibers with DNA characteristics. I. Fixation and electron staining reactions. *J Cell Biol* 1963;**19**:593–612.

Nosek J, Tomaska L. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Curr Genet* 2003;**44**:73–84.

Nosek J, Tomaska L, Bolotin-Fukuhara M, *et al.* Mitochondrial chromosome structure: an insight from analysis of complete yeast genomes. *FEMS Yeast Res* 2006;**6**:356–70.

Paliwal S, Fiumera AC, Fiumera HL. Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics* 2014;**198**:1251–65.

Palmer JD, Shields CR. Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature* 1984;**307**:437–40.

Pantou MP, Kouvelis VN, Typas MA. The complete mitochondrial genome of *Fusarium oxysporum*: insights into fungal mitochondrial evolution. *Gene* 2008;**419**:7–15.

Pramateftaki PV, Kouvelis VN, Lanaridis P, *et al.* The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. *FEMS Yeast Res* 2006;**6**:77–90.

Prochazka E, Polakova S, Piskur J, *et al.* Mitochondrial genome from the facultative anaerobe and petite-positive yeast *Dekkera bruxellensis* contains the NADH dehydrogenase subunit genes. *FEMS Yeast Res* 2010;**10**:545–57.

Rokas A, Williams BL, King N, *et al.* Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 2003;**425**:798–804.

Roy J, Faktorova D, Lukes J, *et al.* Unusual mitochondrial genome structures throughout the Euglenozoa. *Protist* 2007;**158**:385–96.

Sacerdot C, Casaregola S, Lafontaine I, *et al.* Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. *FEMS Yeast Res* 2008;**8**:846–57.

Schatz G, Haslbrunner E, Tuppy H. Deoxyribonucleic acid associated with yeast mitochondria. *Biochem Bioph Res Co* 1964;**15**:127–32.

Sekito T, Okamoto K, Kitano H, *et al.* The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characteristics of yeast mitochondria. *Curr Genet* 1995;**28**:39–53.

Solieri L. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends Microbiol* 2010;**18**:521–30.

Solieri L, Antunez O, Perez-Ortin JE, *et al.* Mitochondrial inheritance and fermentative : oxidative balance in hybrids between *Saccharomyces cerevisiae* and *Saccharomyces uvarum*. *Yeast* 2008;**25**:485–500.

Spatafora JW, Sung G, Johnson DB, *et al.* A five-gene phylogeny of Pezizomycotina. *Mycologia* 2006;**98**:1018–28.

Starkov AA. The role of mitochondria in reactive oxygen species metabolism and signaling. *Ann NY Acad Sci* 2008;**1147**:37–52.

Talla E, Anthouard V, Bouchier C, *et al.* The complete mitochondrial genome of the yeast *Kluyveromyces thermotolerans*. *FEBS Lett* 2005;**579**:30–40.

Tuppen HA, Blakely EL, Turnbull DM, *et al.* Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta* 2010;**1797**:113–28.

Valach M, Farkas Z, Fricova D, *et al.* Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Res* 2011;**39**:4202–19.

Voigt O, Erpenbeck D, Worheide G. A fragmented metazoan organellar genome: the two mitochondrial chromosomes of *Hydra magnipapillata*. *BMC Genomics* 2008;**19**:904–12.

Wallace DC. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* 2005;**39**:359–407.

Wallin I. *Symbionticism and the Origin of Species*. Baltimore: Williams & Wilkins Co, 1927.

Ward B, Anderson R, Bendich AJ. The size of the mitochondrial genome is large and variable in a family of plants. *Cell* 1981;**25**:793–803.

Williamson D. The curious history of yeast mitochondrial DNA. *Nat Rev Genet* 2002;**3**:1–7.

Wolf K, Giudice LD. The variable mitochondrial genome of ascomycetes: organization, mutations, alterations, and expression. *Adv Genet* 1988;**25**:186–308.

Zamaroczy M, Bernardi G. The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae* - a review. *Gene* 1986;**47**:155–77.

Downloaded from https://academic.oup.com/femsyr/article-abstract/15/4/fov023/644438
by The NIST Virtual Library (NVL) user
on 08 June 2018

Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*

Carol J. Bult, Owen White, Gary J. Olsen, Lixin Zhou, Robert D. Fleischmann, Granger G. Sutton, Judith A. Blake, Lisa M. FitzGerald, Rebecca A. Clayton, Jeannine D. Gocayne, Anthony R. Kerlavage, Brian A. Dougherty, Jean-Francois Tomb, Mark D. Adams, Claudia I. Reich, Ross Overbeek, Ewen F. Kirkness, Keith G. Weinstock, Joseph M. Merrick, Anna Glodek, John L. Scott, Neil S. M. Geoghagen, Janice F. Weidman, Joyce L. Fuhrmann, Dave Nguyen, Teresa R. Utterback, Jenny M. Kelley, Jeremy D. Peterson, Paul W. Sadow, Michael C. Hanna, Matthew D. Cotton, Kevin M. Roberts, Margaret A. Hurst, Brian P. Kaine, Mark Borodovsky, Hans-Peter Klenk, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese, J. Craig Venter*

The complete 1.66–megabase pair genome sequence of an autotrophic archaeon, *Methanococcus jannaschii*, and its 58– and 16–kilobase pair extrachromosomal elements have been determined by whole-genome random sequencing. A total of 1738 predicted protein-coding genes were identified; however, only a minority of these (38 percent) could be assigned a putative cellular role with high confidence. Although the majority of genes related to energy production, cell division, and metabolism in *M. jannaschii* are most similar to those found in Bacteria, most of the genes involved in transcription, translation, and replication in *M. jannaschii* are more similar to those found in Eukaryotes.

The discovery of the Archaea in 1977 (*1*) created a quandary for biologists because it was then widely believed that the deepest, most significant evolutionary distinctions were those between Prokaryotes and Eukaryotes. Yet the Archaea, although cytologically prokaryotic, are not specifically related to the Bacteria; at the molecular level, the Archaea are in many respects more like Eukaryotes and may be specifically related to them (*2*). The nature of the Archaea and their relationships to Eukaryotes and Bacteria have posed an intriguing and incompletely resolved puzzle, one that until now has been addressed on the basis of evidence from individual genes (*2*). We now report the first complete genome sequence for a representative of the Archaea, *Methanococcus jannaschii*. The M. *jannaschii* genome sequence provides the first opportunity to compare complete genetic complements and biochemical pathways among the three domains of life from which all extant life forms evolved. *Methanococcus jannaschii* also represents the first complete genome of an autotrophic organism. Its genome sequence, therefore, should provide valuable information on the genetic basis for encoding the metabolic capacity to synthesize de novo all of the building blocks essential for cellular life from inorganic constituents.

The era of true comparative genomics has been ushered in by complete genome sequencing and analysis. We recently described the first two complete bacterial genome sequences, those of *Haemophilus influenzae* and *Mycoplasma genitalium* (*3*). In addition, the complete genome of a Eukaryote, *Saccharomyces cerevisiae*, was recently reported to have been completed (*4*). Large-scale DNA sequencing also has produced an extensive collection of sequence data from *Homo sapiens* (*5*) and *Caenorhabditis elegans* (*5*). The lack of archaeal sequence data has hampered construction of a comprehensive comparative evolutionary framework for assessing the molecular basis of the origin and diversification of cellular life.

Methanococcus jannaschii was originally isolated by J. A. Leigh from a sediment sample collected from the sea floor surface at the base of a 2600-m-deep "white smoker" chimney located at 21°N on the East Pacific Rise (*6*). *Methanococcus jannaschii* grows at pressures of up to more than 200 atm and over a temperature range of 48° to 94°C, with an optimum temperature near 85°C (*6*). It is a strict anaerobe, and, as the name implies, it produces methane.

A whole-genome random sequencing method (*3*) was used to obtain the complete genome sequence for M. *jannaschii*. A small-insert plasmid library [average insert size, 2.5 kilobase pairs (kbp)] and a large-insert λ library (average insert size, 16 kbp) were used as substrates for sequencing. The λ library was used to form a genome scaffold and to verify the orientation and integrity of the contigs formed from the assembly of sequences from the plasmid library. All clones were sequenced from both ends to aid in ordering of contigs during the sequence assembly process. The average length of sequencing reads was 481 bp. A total of 36,718 sequences were assembled by means of the TIGR Assembler (*3*, *7*). Sequence and physical gaps were closed by a combination of strategies (*3*). The colinearity of the in vivo genome to the genome sequence was confirmed by comparison of restriction fragments from six rare-cutter restriction enzymes (Aat II, Bam HI, Bgl II, Kpn I, Sma I, and Sst II) to those predicted from the sequence data. Additional confidence in the colinearity was provided by the genome scaffold produced by sequence pairs from 339 large-insert λ clones, which covered 88% of the main chromosome. Open reading frames (ORFs) and predicted protein-coding regions were identified as described (*3*) with modification (*8*).

The M. *jannaschii* genome consists of three physically distinct elements: (i) a large circular chromosome of 1,664,976 base pairs (bp) (Fig. 1), which contains 1682 predicted protein-coding regions and has a G+C content of 31.4%; (ii) a large circular extrachromosomal element (ECE) (*9*) of 58,407 bp, which contains 44 predicted protein-coding regions and has a G+C content of 28.2% (Fig. 2); and (iii) a small circular ECE (*9*) of 16,550 bp, which

G. J. Olsen, C. I. Reich, B. P. Kaine, and C. R. Woese are in the Microbiology Department, University of Illinois, Champaign-Urbana, IL 61801, USA. R. Overbeek is with the Division of Mathematics and Computer Science, Argonne National Laboratory, Argonne, IL 60439, USA. J. M. Merrick is in the Department of Microbiology, State University of New York, Buffalo, NY 14214, USA. M. Borodovsky is with the School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA. H. O. Smith is in the Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. All other authors are with The Institute for Genomic Research (TIGR), Rockville, MD 20850, USA.

*To whom correspondence should be addressed at The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA.

YM95NC Attachment -- Page 40

SCIENCE

The Genome of *Methanococcus jannaschii*

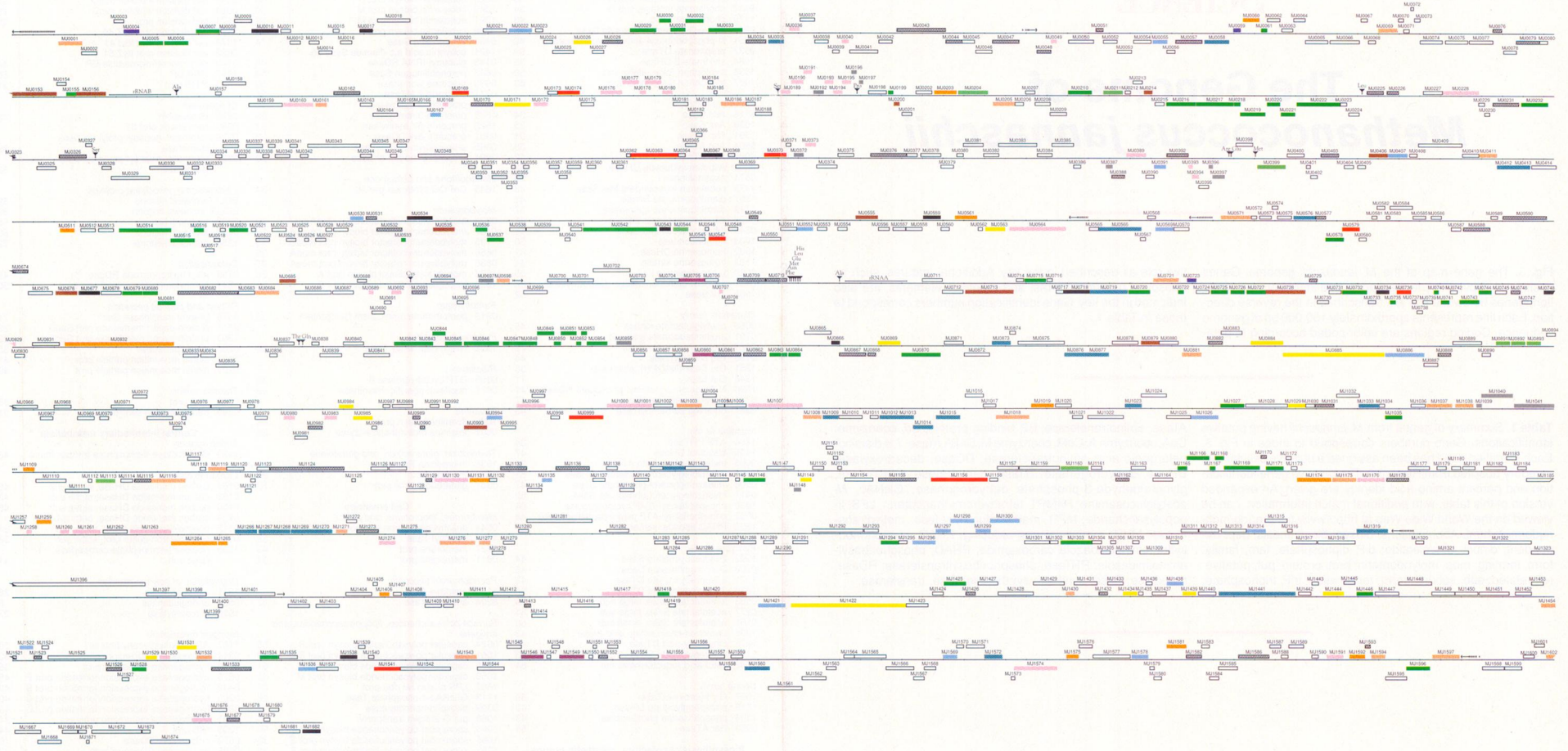
Fig. 3. The gene map of the *M. jannaschii* genome. Genes positioned above the line are transcribed in the forward direction; genes below the line are transcribed in the reverse direction. Each line represents approximately 150,000 bp of genome sequence. Coding regions are color-coded according to putative role as described in the key. Additional features, such as ribosomal operons, tRNAs, and repetitive elements, are also indicated in the key. Gene identification numbers correspond to those in Table 1.

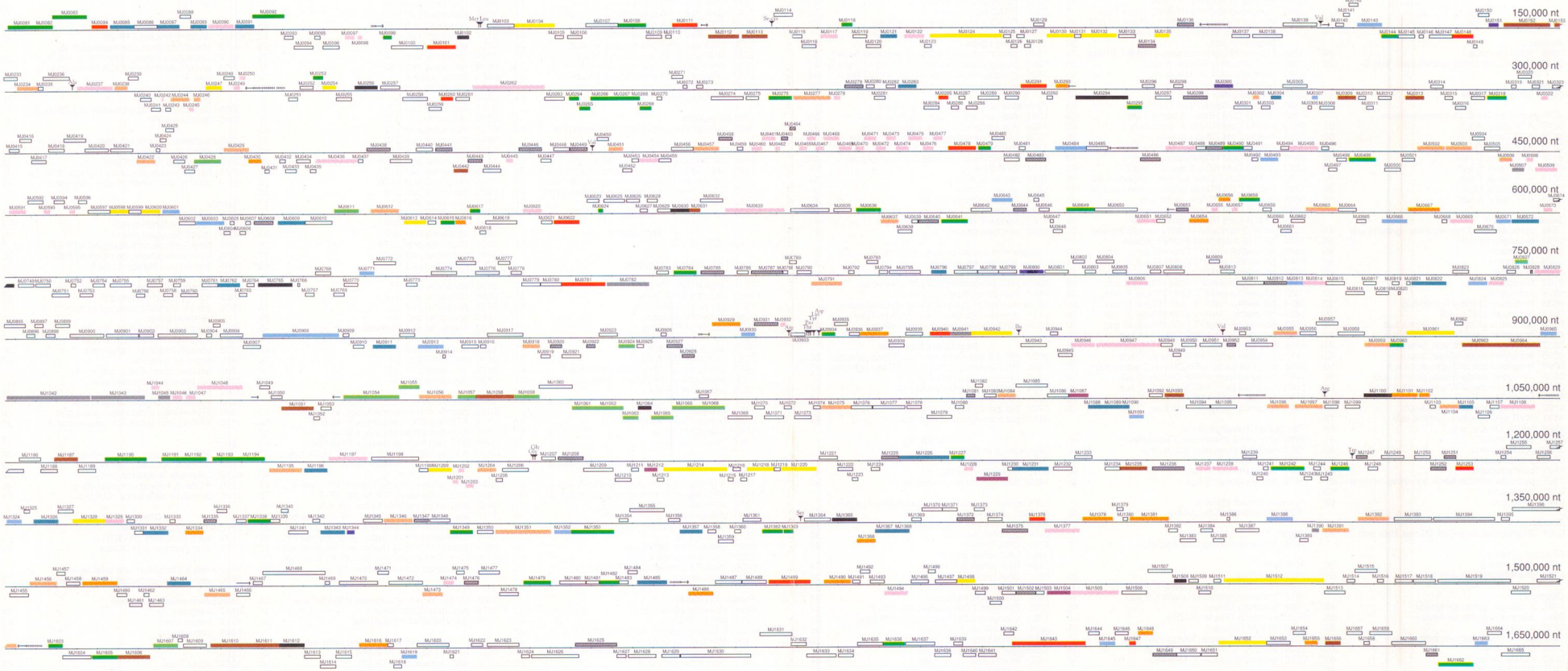
Table 1. Summary of genes from *M. jannaschii* having putative identifications. Gene numbers correspond to those in Fig. 3. Each identified gene has been assigned a putative cellular role category adapted from Riley (*34*). The putative gene identification and percent amino acid identity are given. An expanded version of this table, with additional match information, can be viewed on the World Wide Web at http://www.tigr.org/tdb/mdb/mjdb/mjdb.html. Abbreviations: biosyn, biosynthesis; dep, dependent; dinucl, dinucleotide; diP, diphosphate; fam, family; form, forming; mop, molybdopterin; prot, protein; put, putative; rel, related; sub, subunit; term, termination; triP, triphosphate; ATase, aminotransferase; BP, binding protein; Co, coenzyme; CoA, coenzyme A; CoM, coenzyme M; DAPA ATase, 7,8-diaminopelargonic acid aminotransferase; DCase, decarboxylase; DHase, dehydrogenase; DTase, dehydratase; G3PDHase, glyceraldehyde-3-phosphate dehydrogenase; GlcNAc, *N*-acetylglucosamine; H4MPT, tetrahydromethanopterin; LPS, lipopolysaccharide; MFR, methanofuran; MTase, methyltransferase; MVR, methylviologen-reducing; PRAC, phosphoribosylformiminoimidazole carboxamide; PRAD, phosphoribosylaminoimidazole; PRTase, phosphoribosyltransferase; RDase, reductase; Sase, synthase-synthetase; Tase, transferase.

MJ#	Gene description	%Id
	Amino acid biosynthesis	
	Aromatic amino acid family	
1454	3-dehydroquinate Dtase	34
0502	5-enolpyruvylshikimate 3-phosphate Sase	37
1075	anthranilate Sase, sub I	49
0234	anthranilate Sase, sub II'	45
0238	anthranilate Sase, sub II''	52
0246	chorismate mutase, sub A	38
0612	chorismate mutase, sub B	34
1175	chorismate Sase	49
0918	indole-3-glycerol phosphate Sase	43
0451	N-phosphoribosyl anthranilate isomerase	42
0637	prephenate DTase	39
1084	shikimate 5-DHase	35
1038	tryptophan Sase, alpha sub	50
1037	tryptophan Sase, beta sub	63
	Aspartate family	
1116	Asn Sase	34
1056	Asn Sase	35
1391	Asp ATase	31
0684	Asp ATase	38
0001	Asp ATase	42
0205	Asp-semialdehyde DHase	52
0571	aspartokinase I	39
1473	cobalamin-independent Met Sase	48
1097	diaminopimelate Dcase	42
1119	diaminopimelate epimerase	37
0422	dihydrodipicolinate RDase	45
0244	dihydrodipicolinate Sase	48
1003	homoaconitase	36
1602	homoserine DHase	41
1104	homoserine kinase	31
0020	L-asparaginase I	35
0457	succinyl-diaminopimelate desuccinylase	28
1465	Thr Sase	52
	Glutamate family	
0069	acetylglutamate kinase	45
0791	argininosuccinate lyase 41	
0429	argininosuccinate Sase	72
0186	Glu N-acetylTase	44
1351	Glu Sase (NADPH), alpha sub	38
1346	Gln Sase	71
1096	N-Ac-gamma-glutamyl-phosphate RDase	41
0721	N-acetylornithine ATase	47
0881	ornithine carbamoylTase	43
	Histidine family	
1204	ATP PRTase	34
1456	histidinol DHase	47
0955	histidinol-phosphate ATase	38
0698	imidazoleglycerol-phosphate DHase	51
0506	imidazoleglycerol-phosphate Sase	47
0411	imidazoleglycerol-phosphate Sase	62
1430	phosphoribosyl-AMP cyclohydrolase	64
0302	phosphoribosyl-ATP pyrophosphohydrolase	48
1532	PRAC ribotide isomerase	57
	Pyruvate family	
1392	2-isopropylmalate Sase	44
0503	2-isopropylmalate Sase	45
1271	3-isopropylmalate DTase	42
1277	3-isopropylmalate DTase	50
0663	acetolactate Sase, large sub	35
0277	acetolactate Sase, large sub	51
0161	acetolactate Sase, small sub	50
1008	branched-chain amino acid Atase	43
1276	dihydroxy-acid DTase	45
1195	isopropylmalate Sase	43
1543	ketol-acid reductoisomerase	54
	Serine family	
1597	Gly hydroxy MTase	69
1018	phosphoglycerate DHase	43
1594	phosphoserine phosphatase	43
0959	Ser ATase	55
	Biosynthesis of cofactors, prosthetic groups, and carriers	
0603	Glu-1-semialdehyde ATase	52
0569	porphobilinogen deaminase	42
0493	quinolinate PRTase	40
0407	quinolinate Sase	41
1388	S-adenosylhomocysteine hydrolase	61
	Biotin	
1297	6-carboxyhexanoate-CoA ligase	43
1298	8-amino-7-oxononanoate Sase	45
1300	DAPA ATase	40
1619	bifunctional prot	62
1296	biotin Sase	39
1299	dethiobiotin Sase	37
	Heme and porphyrin	
1438	cobalamin (5'-phosphate) Sase	28
0552	cobalamin biosyn prot J	27
1314	cobalamin biosyn prot D	35
0022	cobalamin biosyn prot D	35
1569	cobalamin biosyn prot M	30
1091	cobalamin biosyn prot M	52
0908	cobalamin biosyn prot N	38
0484	cobyric acid Sase	74
1421	cobyrinic acid a,c-diamide Sase 36	
0143	glutamyl-tRNA RDase	48
0643	porphobilinogen Sase	63
0930	precorrin isomerase	39
0771	precorrin-2 MTase	31
0813	precorrin-3 methylase	45
1578	precorrin-3 methylase	55
1522	precorrin-6Y methylase	31
0391	precorrin-8W DCase	28
0965	uroporphyrin-III C-MTase	55
0994	uroporphyrinogen III Sase	29
	Menaquinone and ubiquinone	
1645	CoPQQ synthesis prot III	34
	Molybdopterin	
0824	molybdenum cofactor biosyn moaA prot	32
0167	molybdenum cofactor biosyn prot moaB	38
1135	molybdenum cofactor biosyn prot moaC	48
0666	molybdenum cofactor biosyn prot moeA	35
0886	molybdenum cofactor biosyn prot moeA	35
1663	mop-guanine dinucl biosyn prot A	30
1324	mop-guanine dinucl biosyn prot B	33
	Pantothenate	
0913	pantothenate metabolism flavoprotein	33
	Pyridine nucleotides	
1352	NH(3)-dep NAD+ Sase	38
	Riboflavin	
0055	GTP cyclohydrolase II	39
0671	riboflavin-specific deaminase	42
	Thiamine	
1026	thiamine biosyn prot	46
0601	thiamine biosynthetic enzyme	36
	Thioredoxin, glutaredoxin, and glutathione	
1536	thioredoxin RDase	39
0530	thioredoxin-2	33
0307	glutaredoxin-like prot	53
	Cell envelope	
	Membranes, lipoproteins, and porins	
0544	dolichyl-phosphate mannose Sase 35	
1057	glycosyl Tase	32
0827	membrane prot	43
0611	membrane prot	34
	Pseudomurein sacculus	
1160	amidase 41	
0204	amidoPRTase	52
	Surface polysaccharides, lipopolysaccharides, and antigens	
0924	capsular polysaccharide biosyn prot B	55
1061	capsular polysaccharide biosyn prot D	52
1055	capsular polysaccharide biosyn prot I	51
1059	capsular polysaccharide biosyn prot M	32
1607	LPS biosyn rel rfbu-prot	34
1113	GLcNAc-1-phosphate Tase	28
0399	phosphomannomutase	37
1068	put O-antigen transporter	24
1066	spore coat polysaccharide biosyn prot C	55
1065	spore coat polysaccharide biosyn prot E	38
1063	spore coat polysaccharide biosyn prot F	39
1062	spore coat polysaccharide biosyn prot G	33
0211	UDP-glucose 4-epimerase	38
1054	UDP-glucose DHase	43
0428	UDP-N-Ac-D-mannosaminuronic acid DHase	47
	Surface structures	
0891	flagellin B1	56
0892	flagellin B2	61
0893	flagellin B3	[illegible]
	Cellular processes	
	Cell division	
1489	cell division control prot 54	35
0363	cell division control prot 21	30
1156	cell division control prot 48	52
0579	cell division inhibitor minD-rel prot	32
0169	cell division inhibitor minD	35
0547	cell division inhibitor minD	37
0084	cell division inhibitor minD	29
0174	cell division prot	29
0370	cell division prot ftsZ 49	
1376	cell division prot ftsJ 41	
0622	cell division prot ftsZ 51	
0148	centromere/microtubule-BP	44
1647	DNA BP	58
1643	P115 prot	31
	Chaperones	
0999	chaperonin	73
0285	heat shock prot 31	
0278	rotamase, peptidyl-prolyl cis-trans isomerase	39
0825	rotamase, peptidyl-prolyl cis-trans isomerase	32
	Chromosome-associated proteins	
ECL17	archaeal histone	59
ECL29	archaeal histone	59
0932	archaeal histone	68
0168	archaeal histone	68
1258	archaeal histone	72
	Detoxification	
0736	alkyl hydroperoxide RDase	48
1541	N-ethylammeline chlorohydrolase	30
	Protein and peptide secretion	
0478	preprot translocase SecY	71
0111	protein-export membrane prot SecD	29
1253	protein-export membrane prot SecF	32
0260	signal peptidase	35
0101	signal recognition particle prot	41
0291	signal recognition particle prot	49
	Transformation	
0781	klbA prot	35
0940	transformation sensitive prot	31
	Central intermediary metabolism	
	Amino sugars	
1420	Gln–fructose-6-phosphate transaminase	42
	Carbon fixation	
0153	carbon monoxide DHase, alpha sub	48
0152	carbon monoxide DHase, alpha sub	43
0156	carbon monoxide DHase, alpha sub	48
0728	carbon monoxide DHase, beta sub 36	
0112	corrinoid/iron-sulfur prot, large sub	34
0113	corrinoid/iron-sulfur prot, small sub	38
1235	ribulose bisphosphate carboxylase, large sub	41
	Degradation of polysaccharides	
1611	alpha-amylase	28
0555	endoglucanase	0
1610	glucoamylase	27
	Nitrogen metabolism	
1187	ADP-ribosylglycohydrolase	30
0214	hydrogenase accessory prot	32
0713	hydrogenase accessory prot	35
0676	hydrogenase expression/formation prot E	45
0442	hydrogenase expression/formation prot B	43
0200	hydrogenase expression/formation prot C	40
0993	hydrogenase expression/formation prot D	43
0631	hydrogenase maturation protease	34
1093	nifB prot	44
0879	nitrogenase RDase	78
0685	nitrogenase RDase rel prot	32
1051	nodulation factor production prot	33
1058	nodulation factor production prot	35
	Phosphorus compounds	
0963	N-methylhydantoinase	33
0964	N-methylhydantoinase	36
	Polyamine biosynthesis	
0535	acetylpolyamine aminohydolase	34
[illegible]	[illegible]	39

YM95NC Attachment -- Page 41

Fig. 3.




150,000 nt
300,000 nt
450,000 nt
600,000 nt
750,000 nt
900,000 nt
1,050,000 nt
1,200,000 nt
1,350,000 nt
1,500,000 nt
1,650,000 nt
Amino acid biosynthesis
Biosynthesis of cofactors, prosthetic groups, carriers
Cell envelope
Cellular processes
Central intermediary metabolism
Energy metabolism
Fatty acid and phospholipid metabolism
Purines, pyrimidines, nucleosides, and nucleotides
Regulatory functions
Replication
Transport and binding proteins
Translation
Transcription
Other categories
Unknown
Hypothetical
rRNA
tRNA
Long repeat
Short repeat
1 kb

MJ#	Gene description	%Id
	Polysaccharides (cytoplasmic)	
1606	glycogen Sase	32
	Other	
1656	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	41
0406	ribokinase	24
0309	ureohydrolase	41
	Energy metabolism	
0479	adenylate kinase	100
	Aerobic	
0649	NADH oxidase	28
0520	NADH-ubiquinone oxidoRDase, sub 1	29
	Anaerobic	
0092	fumarate RDase	41
	ATP–proton motive force interconversion	
0217	ATP Sase, A sub	61
0216	ATP Sase, B sub	68
0219	ATP Sase, C sub	29
0615	ATP Sase, D sub	39
0220	ATP Sase, E sub	33
0218	ATP Sase, F sub	22
0222	ATP Sase, I sub	28
0221	ATP Sase, K sub	46
	Electron transport	
1446	cytochrome-c3 hydrogenase, gamma sub	41
0741	desulfoferredoxin	44
0722	ferredoxin	43
0099	ferredoxin	40
0061	ferredoxin	43
0199	ferredoxin	75
0578	ferredoxin	50
0533	ferredoxin 2[4Fe-4S] homolog	37
0624	ferredoxin 2[4Fe=4S]	48
0276	ferredoxin oxidoRDase, alpha sub	45
0267	ferredoxin oxidoRDase, alpha sub	33
0537	ferredoxin oxidoRDase, beta sub	41
0266	ferredoxin oxidoRDase, beta sub	33
0268	ferredoxin oxidoRDase, delta sub	59
0536	ferredoxin oxidoRDase, gamma sub	32
0269	ferredoxin oxidoRDase, gamma sub	64
0732	flavoprotein	41
1192	MVR hydrogenase, alpha sub	78
1191	MVR hydrogenase, gamma sub	72
1362	NADH DHase, sub 1	26
0934	polyferredoxin	41
1303	polyferredoxin	40
0514	polyferredoxin	36
1193	polyferredoxin	62
1227	pyruvate formate-lyase activating enzyme	31
0735	rubredoxin	60
0740	rubredoxin	62
	Fermentation	
0007	2-hydroxyglutaryl-CoA DTase, beta sub	23
	Gluconeogenesis	
1479	Ala ATase 2	30
0542	phosphoenolpyruvate Sase	61
	Glycolysis	
1482	2-phosphoglycerate kinase	48
0641	3-phosphoglycerate kinase	58
0232	enolase	59
1605	glucose-6-phosphate isomerase	33
1146	G3PDHase	60
0490	lactate DHase	40
1411	NADP-dep G3PDHase	40
0108	pyruvate kinase	39
1528	triosephosphate isomerase	30
	Methanogenesis	
0253	F420-reducing hydrogenase, delta sub	50
1035	F420-dep N5,N10-methylene-H4MPT DHase	68
0030	F420-reducing hydrogenase, alpha sub	66
0727	F420-reducing hydrogenase, alpha sub	28
0029	F420-reducing hydrogenase, alpha sub	51
0725	F420-reducing hydrogenase, beta sub	43
1349	F420-reducing hydrogenase, beta sub	36
0032	F420-reducing hydrogenase, beta sub	72
0870	F420-reducing hydrogenase, beta sub	43
0726	F420-reducing hydrogenase, gamma sub	44
0031	F420-reducing hydrogenase, gamma sub	77
0295	formate DHase (fdhD)	36
0006	formate DHase, alpha sub	42
1353	formate DHase, alpha sub	56
0005	formate DHase, beta sub GB:J02581_2 0.0	49
0155	formate DHase, iron-sulfur sub	38
0264	formate hydrogenlyase, sub 2	40
0265	formate hydrogenlyase, sub 2	41
0515	formate hydrogenlyase, sub 5	32
1027	formate hydrogenlyase, sub 5	35
1363	formate hydrogenlyase, sub 7	38
0516	formate hydrogenlyase, sub 7	49
0318	formylmethanofuran:H4MPT formylTase	71
1338	H2-dep methylene-H4MPT DHase-rel prot	30
0715	H2-form N5,N10-methylene-H4MPT DHase-rel prot	30
0784	H2-form N5,N10-methylene-H4MPT DHase	75
1190	heterodisulfide RDase, A sub	60
0743	heterodisulfide RDase, B sub	61
0863	heterodisulfide RDase, B sub	64
0864	heterodisulfide RDase, C sub	53
0744	heterodisulfide RDase, C sub	56
0118	methyl CoM RDase II operon, prot D	54
0083	methyl CoM RDase II, alpha sub	88
0081	methyl CoM RDase II, beta sub	80
0082	methyl CoM RDase II, gamma sub	83
0844	methyl CoM RDase operon, prot C	83
0843	methyl CoM RDase operon, prot D	60
1662	methyl CoM RDase system, component A2	38
1242	methyl CoM RDase system, component A2	61
0846	methyl CoM RDase, alpha sub	86
0842	methyl CoM RDase, beta sub	76
0845	methyl CoM RDase, gamma sub	79
1636	N5,N10-methenyl-H4MPT cyclohydrolase	69
1534	N5,N10-methylene-H4MPT RDase	67
0849	N5-methyl-H4MPT:CoM MTase, C sub	40
0848	N5-methyl-H4MPT:CoM MTase, D sub	64
0850	N5-methyl-H4MPT:CoM MTase, B sub	37
0847	N5-methyl-H4MPT:CoM MTase, E sub	62
0852	N5-methyl-H4MPT:CoM MTase, F sub	38
0854	N5-methyl-H4MPT:CoM MTase, H sub	63
0851	N5-methyl-H4MPT:CoM MTase, A sub	56
0853	N5-methyl-H4MPT:CoM MTase, G sub	50
1169	tungsten formyl-MFR DHase, A sub	70
1194	tungsten formyl-MFR DHase, B sub	70
1171	tungsten formyl-MFR DHase, C sub	52
0658	tungsten formyl-MFR DHase, C sub rel prot	36
1168	tungsten formyl-MFR DHase, D sub	58
1165	tungsten formyl-MFR DHase, E sub	45
1166	tungsten formyl-MFR DHase, F sub	48
1167	tungsten formyl-MFR DHase, G sub	60
	Pentose phosphate pathway	
0680	pentose-5-phosphate-3-epimerase	46
1603	ribose 5-phosphate isomerase	42
0960	transaldolase	60
0681	transketolase, A sub	42
0679	transketolase, B sub	38
	Pyruvate dehydrogenase	
0636	dihydrolipoamide DHase	29
	Sugars	
1418	fuculose-1-phosphate aldolase	30
	TCA cycle	
0499	aconitase	30
1294	fumarate hydratase, class I, A sub	35
0617	fumarate hydratase, class I, B sub	44
1596	isocitrate DHase	43
0720	isocitrate DHase (NADP)	48
1425	malate DHase	61
0033	succinate DHase, flavoprotein sub	41
1246	succinyl-CoA Sase, alpha sub	59
0210	succinyl-CoA Sase, beta sub	49
	Fatty acid and phospholipid metabolism	
0705	3-hydroxy-3-methylglutaryl CoA RDase	48
1546	acyl carrier prot Sase	65
0860	bifunctional short chain isoprenyl diP Sase	49
1229	biotin carboxylase	59
1212	CDP-diacylglycerol–Ser O-phosphatidylTase	44
1504	lipopolysaccharide biosyn prot D	43
1087	melvalonate kinase	34
1549	nonspecific lipid-transfer prot	47
	Purines, pyrimidines, nucleosides, and nucleotides	
	2'-Deoxyribonucleotide metabolism	
0832	anaerobic ribonucleoside-triP RDase	28
0430	deoxycytidine triP deaminase	39
1102	put deoxycytidine triP deaminase	32
0511	deoxyuridylate hydroxymethylase	40
0937	glycinamide ribonucleotide Sase	38
	Purine ribonucleotide biosynthesis	
0929	adenylosuccinate lyase	43
0561	adenylosuccinate Sase	43
1131	GMP Sase	53
1575	GMP Sase	41
1616	inosine-5'-monophosphate DHase	62
1265	nucleoside diP kinase	55
0616	PRAD carboxylase	57
1592	PRAD succinocarboxamide Sase	48
0203	phosphoribosylformylglycinamidine cyclo-ligase	40
1648	phosphoribosylformylglycinamidine Sase I	52
1264	phosphoribosylformylglycinamidine Sase II	43
1486	phosphoribosylglycinamide formylTase 2	64
1366	ribose-phosphate pyrophosphokinase	35
	Pyrimidine ribonucleotide biosynthesis	
1581	Asp carbamoyl Tase, catalytic sub	50
1406	Asp carbamoyl Tase, regulatory sub	37
1378	carbamoyl-phosphate Sase, large sub	60
1381	carbamoyl-phosphate Sase, large sub	55
1019	carbamoyl-phosphate Sase, small sub	49
1174	CTP Sase	58
0656	cytidylate kinase	34
1490	dihydroorotase	35
0654	dihydroorotase DHase	42
0293	thymidylate kinase	32
1109	uridine 5'-monophosphate Sase	39
1259	uridylate kinase	31
	Salvage of nucleosides and nucleotides	
1459	adenine deaminase	36
1655	adenine PRTase	34
0060	methylthioadenosine phosphorylase	42
0667	thymidine phosphorylase 31	
	Sugar-nucleotide biosynthesis and conversions	
1101	glucose-1-phosphate thymidylyl Tase	34
1334	UDP-glucose pyrophosphorylase	47
	Regulatory functions	
0800	(R)-2-hydroxyglutaryl-CoA DTase activator	32
0004	(R)-2-hydroxyglutaryl-CoA DTase activator	38
0059	nitrogen regulatory prot P-II	57
1344	nitrogen regulatory prot P-II	57
0300	put transcriptional regulator	32
0723	put transcriptional regulator	51
0151	put transcriptional regulator	52
	Replication	
	Degradation of DNA	
1434	endonuclease III	27
0613	endonuclease III	42
1439	thermonuclease precursor	37
	DNA replication, restriction, modification, recombination, and repair	
1029	dimethyladenosine Tase	40
0104	put DNA helicase	36
0171	DNA ligase	36
0869	DNA repair prot 45	
1444	DNA repair prot RAD2	38
0254	DNA repair prot RAD51	34
0961	DNA replication initiator prot	33
1652	DNA topoisomerase I	34
0885	DNA-dep DNA polymerase, fam B	47
1529	methylated DNA protcysteine MTase	37
0698	modification methylase	36
4328	modification methylase	36
1206	modification methylase	41
1498	modification methylase	31
0563	modification methylase	35
0985	modification methylase	54
1149	mutator mutT prot	41
0942	put ATP-dep helicase	35
0247	proliferating-cell nuclear antigen	32
0026	proliferating-cell nucleolar antigen, 120 kD	47
1422	replication factor C	46
0884	replication factor C, large sub	38
1220	restriction modification enzyme, M1 sub	33
0132	type I restriction enzyme	37
0130	restriction modification system S sub	29
1512	reverse gyrase	42
0135	ribonuclease HII	40
ECL42	type I restriction enzyme ECOR124/3 I M prot	40
1214	type I restriction enzyme	28
0124	type I restriction enzyme	32
ECL40	type I restriction enzyme	37
1531	type I restriction enzyme CfrI, specificity sub	39
1218	type I restriction-modification enzyme, S sub	30
0984	type II restriction enzyme	48
0600	type II restriction enzyme DPNII	41
	Transcription	
	DNA-dependent RNA polymerase	
1042	DNA-dep RNA polymerase, A' sub	75
1043	DNA-dep RNA polymerase, A'' sub	65
1041	DNA-dep RNA polymerase, B' sub	74
1040	DNA-dep RNA polymerase, B'' sub	71
0192	DNA-dep RNA polymerase, D sub	41
0397	DNA-dep RNA polymerase, E' sub	42
0396	DNA-dep RNA polymerase, E'' sub	36
1039	DNA-dep RNA polymerase, H sub	50
1390	DNA-dep RNA polymerase, I sub	54
0197	DNA-dep RNA polymerase, K sub	44
0387	DNA-dep RNA polymerase, L sub	36
0196	DNA-dep RNA polymerase, N sub	54
	RNA processing	
0697	fibrillarin-like pre-rRNA processing prot	76
	Transcription factors	
0941	put transcription initiation factor IIIC	21
1045	put transcription term-antiterm factor nusA	48
0372	put transcription term-antiterm factor nusG	25
0507	TATA-binding transcription initiation factor	48
0782	transcription initiation factor IIB	64
1148	transcription-associated prot 'TFIIS'	59
	Translation	
0160	PET112 prot	34
	Amino acyl tRNA synthetases	
0564	alanyl-tRNA Sase	28
0237	arginyl-tRNA Sase	32
1555	aspartyl-tRNA Sase	58
1377	glutamyl-tRNA Sase	52
0228	glycyl-tRNA Sase	46
1000	histidyl-tRNA Sase	36
0947	isoleucyl-tRNA Sase	53
0633	leucyl-tRNA Sase	36
1263	methionyl-tRNA Sase	37
0487	phenylalanyl-tRNA Sase, alpha sub	41
1108	phenylalanyl-tRNA Sase, beta sub	32
1238	prolyl-tRNA Sase	40
1197	threonyl-tRNA Sase	30
1415	tryptophanyl-tRNA Sase	31
0389	tyrosyl-tRNA Sase	39
1007	valyl-tRNA Sase	37
1077	seryl-tRNA Sase	18
	Degradation of proteins, peptides, and glycopeptides	
1176	ATP-dep 26S protease regulatory sub 4	47
1494	ATP-dep 26S protease regulatory sub 8	54
1417	ATP-dep protease La	30
0090	collagenase	33
1130	O-sialoglycoprotein endopeptidase	51
0651	protease IV	35
0591	proteasome, alpha sub	58
1237	proteasome, beta sub	49
0806	XAA-PRO dipeptidase, M24B fam	34
0996	Zn protease	34
	Protein modification	
0814	deoxyhypusine Sase	50
1274	diphthine Sase	42
0172	L-isoaspartyl prot carboxyl MTase	46
1329	Met aminopeptidase	36
1530	N-terminal acetylTase complex, ARD1 sub	40
1591	selenium donor prot	35
	Ribosomal proteins: synthesis and modification	
0509	acidic ribosomal prot P0 (L10E)	63
0242	ribosomal prot HG12	64
1203	ribosomal prot HS6-type	47
0510	ribosomal prot L1	65
0373	ribosomal prot L11	47
0508	ribosomal prot L12	73
0194	ribosomal prot L13	46
0466	ribosomal prot L14	75
0657	ribosomal prot L14B	37
0477	ribosomal prot L15	65
0983	ribosomal prot L15B	55
0474	ribosomal prot L18	74
0473	ribosomal prot L19	69
0179	ribosomal prot L2	74
0040	ribosomal prot L21	55
0460	ribosomal prot L22	40
0178	ribosomal prot L23	70
0467	ribosomal prot L24	73
1201	ribosomal prot L24E	54
0462	ribosomal prot L29	52
0193	ribosomal prot L29E	49
0176	ribosomal prot L3	47
1044	ribosomal prot L30	65
0049	ribosomal prot L31	41
0472	ribosomal prot L32	57
0655	ribosomal prot L34	37
0098	ribosomal prot L37	52
0593	ribosomal prot L37a	45
0177	ribosomal prot L4	50
0707	ribosomal prot L40	58
0249	ribosomal prot L44	39
0689	ribosomal prot L46	52
0469	ribosomal prot L5	72
0471	ribosomal prot L6	67
0476	ribosomal prot L7	72
0595	ribosomal prot LX	39
0322	ribosomal prot S10	68
0191	ribosomal prot S11	62
1046	ribosomal prot S12	83
0036	ribosomal prot S13	50
1474	ribosomal prot S15A	62
0465	ribosomal prot S17	72
0245	ribosomal prot S17B	52
1489	ribosomal prot S18	43
0180	ribosomal prot S19	57
0692	ribosomal prot S19S	46
0394	ribosomal prot S24	43
0250	ribosomal prot S27	41
0393	ribosomal prot S27A	58
0461	ribosomal prot S3	47
1202	ribosomal prot S33	63
0980	ribosomal prot S3a	29
0190	ribosomal prot S4	52
0468	ribosomal prot S4E	71
0475	ribosomal prot S5	74
1260	ribosomal prot S6	37
0620	ribosomal prot S6 modification prot	35
1001	ribosomal prot S6 modification prot II	25
1047	ribosomal prot S7	63
0470	ribosomal prot S8	74
0673	ribosomal prot S8E	50
0195	ribosomal prot S9	51
	Translation factors	
0829	peptide chain release factor, eRF, sub 1	33
1574	ATP-dep RNA helicase, eIF-4A fam	34
1505	ATP-dep RNA helicase, eIF-4A fam	32
0669	ATP-dep RNA helicase, eIF-4A fam	44
0495	put translation factor, EF-TU/1 alpha fam	37
0262	put translation initiation factor, FUN12/IF-2 fam	40
0324	translation elongation factor, EF-1 alpha	80
1048	translation elongation factor, EF-2	75
0445	translation initiation factor, eIF-1A	49
0117	translation initiation factor, eIF-2, alpha sub	34
0097	translation initiation factor, eIF-2, beta sub	33
1261	translation initiation factor, eIF-2, gamma sub	53
0454	translation initiation factor, eIF-2B, alpha sub	38
0122	translation initiation factor, eIF-2B, delta sub	30
1228	translation initiation factor, eIF-5A	50
	tRNA modification	
0946	N2,N2-dimethylguanosine tRNA MTase	33
1675	pseudouridylate Sase I	34
0436	queuine tRNA ribosylTase	30
	Transport and binding proteins	
1572	ABC transporter ATP-BP	36
0719	ABC transporter ATP-BP	50
1023	ABC transporter ATP-BP	50
0035	ABC transporter sub	38
1508	ABC transporter, probable ATP-binding sub	44
1326	GTP-BP	52
1332	GTP-BP	40
1408	GTP-BP, GTP1/OBG-fam	31
1464	put GTP-BP	35
1033	magnesium and cobalt transport prot	43
0091	sodium-calcium exchanger prot	32
0283	nucleotide-BP	45
	Amino acids, peptides, and amines	
0609	amino acid transporter	22
1343	ammonium transport prot AMT1	36
0058	ammonium transporter	35
1269	branched-chain amino acid transport prot livH	55
1266	branched-chain amino acid transport prot livJ	28
1270	branched-chain amino acid transport prot livM	31
1196	cationic amino acid transporter MCAT-2	25
0304	ferripyochelin BP	53
0796	Gln transport ATP-BP Q	48
1267	branched-chain amino acid transport ATP-BP	35
1268	branched-chain amino acid transport ATP-BP	40
	Anions	
0412	nitrate transport ATP-BP	45
0413	nitrate transport permease prot 35	
1012	phosphate transport system ATP-BP	60
1013	phosphate transport system permease prot A	37
1014	phosphate transport system permease prot C	39
1009	phosphate transport system regulatory prot	29
1015	phosphate-BP	42
	Carbohydrates, organic alcohols, and acids	
0576	malic acid transport prot	24
0762	malic acid transport prot	25
0121	SN-glycerol-3-phosphate transport ATP-BP	33
1319	sodium-dep noradrenaline transporter	40
	Cations	
1088	cobalt transport ATP-BP O	46
1090	cobalt transport prot N	46
1089	cobalt transport prot Q	29
0089	ferric enterobactin transport ATP-BP	34
0873	ferric enterobactin transport ATP-BP	36
0566	ferrous iron transport prot B	36
0827	hemin permease	34
0087	hemin permease	38
0085	iron transport system BP	33
0876	iron(III) dicitrate transport system permease prot	32
1441	magnesium chelatase sub 36	
0911	magnesium chelatase sub 56	
1275	sodium-hydrogen antiporter	30
0672	sodium transporter	40
1231	oxaloacetate DCase, alpha sub	53
1357	put potassium channel prot	30
1367	sulfate permease	38
1368	sulfate/thiosulfate transport prot	30
1485	TRK system potassium uptake prot	29
1105	TRK system potassium uptake prot A	35
	Other	
1142	ATPase, arsenical pump-driving	35
0822	ATPase, vanadate-senstive	49
0718	chromate resistance prot A	28
1226	ATPase, hydrogen transporting	45
1560	quinolone resistance norA prot	29
	Other categories	
	Drug and analog sensitivity	
1538	toxin sensitivity prot KTI12	29
0102	phenylacrylic acid DCase	46
	Phage-related functions and prophages	
0630	sodium-dep phosphate transporter	33
	Transposon-related functions	
0367	integrase	31
0017	transposase	30
1466	transposase	30
	Other	
1064	acetylTase	47
1612	phosphonopyruvate DCase	32
0677	ethylene-inducible prot homolog	67
0534	flavoprotein	35
0748	flavoprotein	68
0256	phosphonopyruvate DCase	30
1682	heat shock prot X	31
0866	HIT prot, member of the HIT-fam	40
0294	large helicase rel prot, LHR	32
0010	phosphonopyruvate DCase	28
0734	rubrerythrin	49
0559	survival prot surE	35
1100	urease operon prot	34
0543	Wilm's tumor suppressor homolog	44
0765	[6Fe-6S] prismane-containing prot	61
1365	pheromone shutdown prot	31
ECL24	SOJ prot	36

YM95NC Attachment -- Page 44

contains 12 predicted protein-coding regions and has a G+C content of 28.8% (Fig. 2). The sequences of the M. *jannaschii* chromosome and of the large and small ECEs have been deposited in the Genome Sequence DataBase with the accession numbers L77117, L77118, and L77119, respectively. The annotated genome sequence data and clone information for M. *jannaschii* are available on the World Wide Web (http://www.tigr.org/tdb/mdb/mjdb/mjdb.html).

Of the 1743 predicted protein-coding regions reported previously for *H. influenzae*, 78% had a match in the public sequence database (*3*). Of these, 58% were matches to genes with reasonably well defined function, whereas 20% were matches to genes whose function was undefined. Similar observations were made for the M. *genitalium* genome (*3*). Of the predicted protein-coding regions from M. *genitalium*, 83% have a counterpart in the *H. influenzae* genome. In contrast, only 38% of the predicted protein-coding regions from M. *jannaschii* match a gene in the database that could be assigned a putative cellular role with high confidence; 6% of the predicted protein-coding regions had matches to hypothetical proteins (Fig. 3 and Table 1). Approximately 100 genes in M. *jannaschii* had marginal similarity to genes or segments of genes from the public sequence databases and could not be assigned a putative cellular role with high confidence. Only 11% of the predicted protein-coding regions from *H. influenzae* and 17% of the predicted protein-coding regions from M. *genitalium* matched a predicted protein-coding region from M. *jannaschii*.

Energy production in M. *jannaschii* occurs by the reduction of CO_2 with H_2 to produce methane. Genes for all of the known enzymes and enzyme complexes associated with methanogenesis (*10*) were identified in M. *jannaschii*, the sequence and order of which are typical of methanogens. *Methanococcus jannaschii* appears to use both H_2 and formate as substrates for methanogenesis, but lacks the genes to use methanol or acetate. The ability to fix nitrogen has been demonstrated in a number of methanogens (*11*), and all the genes necessary for this pathway have been identified in M. *jannaschii* (Table 1). In addition to its anabolic pathways, several scavenging molecules have been identified in M. *jannaschii* that probably play a role in importing small organic compounds, such as amino acids, from the environment (Table 1).

Three different pathways control the fixation of CO_2 into organic carbon: the noncyclic, reductive acetyl–coenzyme A–carbon monoxide dehydrogenase pathway (Ljungdahl-Wood pathway), the reductive trichloroacetic acid cycle, and the Calvin cycle. Methanogens fix carbon by the Ljungdahl-Wood pathway (*12*), which is facilitated by the carbon monoxide dehydrogenase enzyme complex (*13*). The complete Ljungdahl-Wood pathway, encoded in the M. *jannaschii* genome, depends on the methyl carbon in methanogenesis; however, methanogenesis can occur independently of carbon fixation.

Although genes encoding two enzymes required for gluconeogenesis (glucopyruvate oxidoreductase and phosphoenolpyruvate synthase) were found in the M. *jannaschii* genome, genes encoding other key intermediates of gluconeogenesis (fructose bisphosphatase and fructose 1,6-bisphosphate aldolase) were not identified. Glucose catabolism by glycolysis also requires the aldolase, as well as phosphofructokinase, an enzyme that also was not found in M. *jannaschii* and has not been detected in any of the Archaea. In addition, genes specific for the Entner-Doudoroff pathway, an alternative pathway used by some microbes for the catabolism of glucose, were not identified in the genomic sequence. The presence of a number of nearly complete metabolic pathways suggests that some key genes are not recognizable at the sequence level, although we cannot exclude the possibility that M. *jannaschii* may use alternative metabolic pathways.

In general, the M. *jannaschii* genes that encode proteins involved in the transport of small inorganic ions into the cell are homologs of bacterial genes. The genome includes many representatives of the ABC transporter family, as well as genes for exporting heavy metals (for example, the chromate-resistance protein) and other toxic compounds (for example, the *norA* drug efflux pump locus).

More than 20 predicted protein-coding regions have sequence similarity to polysaccharide biosynthetic enzymes. These genes have only bacterial homologs or are most closely related to their bacterial counterparts. The identified polysaccharide biosynthetic genes in M. *jannaschii* include those for the interconversion of sugars, activation of sugars to nucleotide sugars, and glycosyltransferases for the polymerization of nucle-

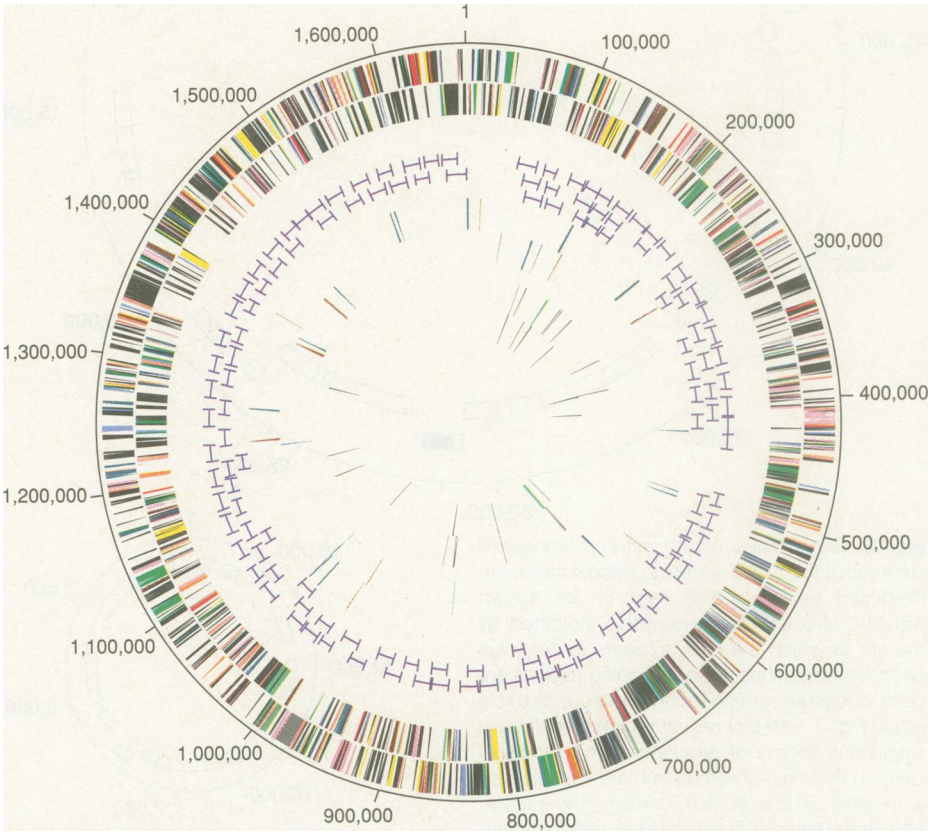


Fig. 1. A circular representation of the *M. jannaschii* chromosome illustrating the location of each predicted protein-coding region as well as selected features of the genome. Outer concentric circle: predicted protein-coding regions on the plus strand color-coded according to role as indicated in Fig. 3. Second concentric circle: predicted protein-coding regions on the minus strand color-coded according to role as indicated in Fig. 3. Third concentric circle: coverage by λ clones (three levels of blue range bars). Fourth and fifth concentric circles representing the plus and minus strands, respectively: members of the IS*AMJ1* family (red) and repetitive elements (cyan). Sixth and seventh concentric circles representing the plus and minus strands, respectively: transfer RNAs (black) and ribosomal RNAs (green).

YM95NC Attachment -- Page 45

otide sugars into oligo- and polysaccharides that are subsequently incorporated into surface structures (*14*). In an arrangement similar to that of bacterial polysaccharide biosynthesis genes, many of the genes for M. *jannaschii* polysaccharide production are clustered together (Table 1 and Fig. 3). The G+C content in this region is <95% of that in the rest of the M. *jannaschii* genome. A similar observation was made in *Salmonella typhimurium* (*15*), in which the gene cluster for lipopolysaccharide O antigen has a significantly lower G+C ratio than that in the rest of the genome. In that case, the difference in G+C content was interpreted as meaning that the region originated by lateral transfer from another organism.

Of the three main multicomponent information-processing systems (transcription, translation, and replication), translation appears to be the most universal in its overall makeup in that the basic translation machinery is similar in all three domains of life. *Methanococcus jannaschii* has two ribosomal RNA operons, designated A and B, and a separate 5S RNA gene that is associated with several transfer RNAs (tRNAs). Operon A has the organization 16S-23S-5S, whereas operon B lacks the 5S component. An alanine tRNA is situated in the spacer region between the 16S and 23S subunits in both operons. The majority of proteins associated with the ribosomal subunits (especially the small subunit) are present in both Bacteria and Eukaryotes. However, the relatively protein-rich eukaryotic ribosome contains additional ribosomal proteins not found in the bacterial ribosome. A smaller number of Bacteria-specific ribosomal proteins exist as well. The M. *jannaschii* genome contains all ribosomal proteins that are common to Eukaryotes and Bacteria. It shows no homologs of the bacterial-specific ribosomal proteins, but does possess homologs of a number of the eukaryotic-specific ones. Homologs of all archaeal-specific ribosomal proteins that have been reported to date (*16*) are found in M. *jannaschii*.

As shown for other Archaea (*2*), the *Methanococcus* translation elongation factors EF-1α (EF-Tu in Bacteria) and EF-2 (EF-G in Bacteria) are most similar to their eukaryotic counterparts. In addition, the M. *jannaschii* genome contains 11 translation-initiation factor genes. Three of these genes encode the subunits homologous to those of the eukaryotic IF-2 and are reported here in the Archaea for the first time. A fourth initiation factor gene that encodes a second IF-2 is also found in M. *jannaschii*. This additional IF-2 gene is most similar to the yeast protein FUN12 (*17*) which, in turn, appears to be a homolog of the bacterial IF-2. It is not known which of the two IF-2–like initiation factors identified in M. *jannaschii* plays a role in directing the initiator tRNA to the start site of the mRNA. The fifth identified initiation factor gene in M. *jannaschii* encodes IF-1A, which has no bacterial homolog. The sixth gene encodes the hypusine-containing initiation factor eIF-5a. Two subunits of the translation initiation factor eIF-2B were identified in M. *jannaschii*. Finally, three putative adenosine triphosphate–dependent helicases were identified that belong to the eIF-4a family of translation initiation factors.

Thirty-seven tRNA genes were identified in the M. *jannaschii* genome. Almost all amino acids encoded by two codons have a single tRNA, except for glutamic acid, which has two. Both an initiator and an internal methionyl tRNA are present. The two pyrimidine-ending isoleucine codons are covered by a single tRNA, whereas the third (AUA) seems covered by a related tRNA having a CAU anticodon. A single tRNA appears to cover the three isoleucine codons. Those amino ac-

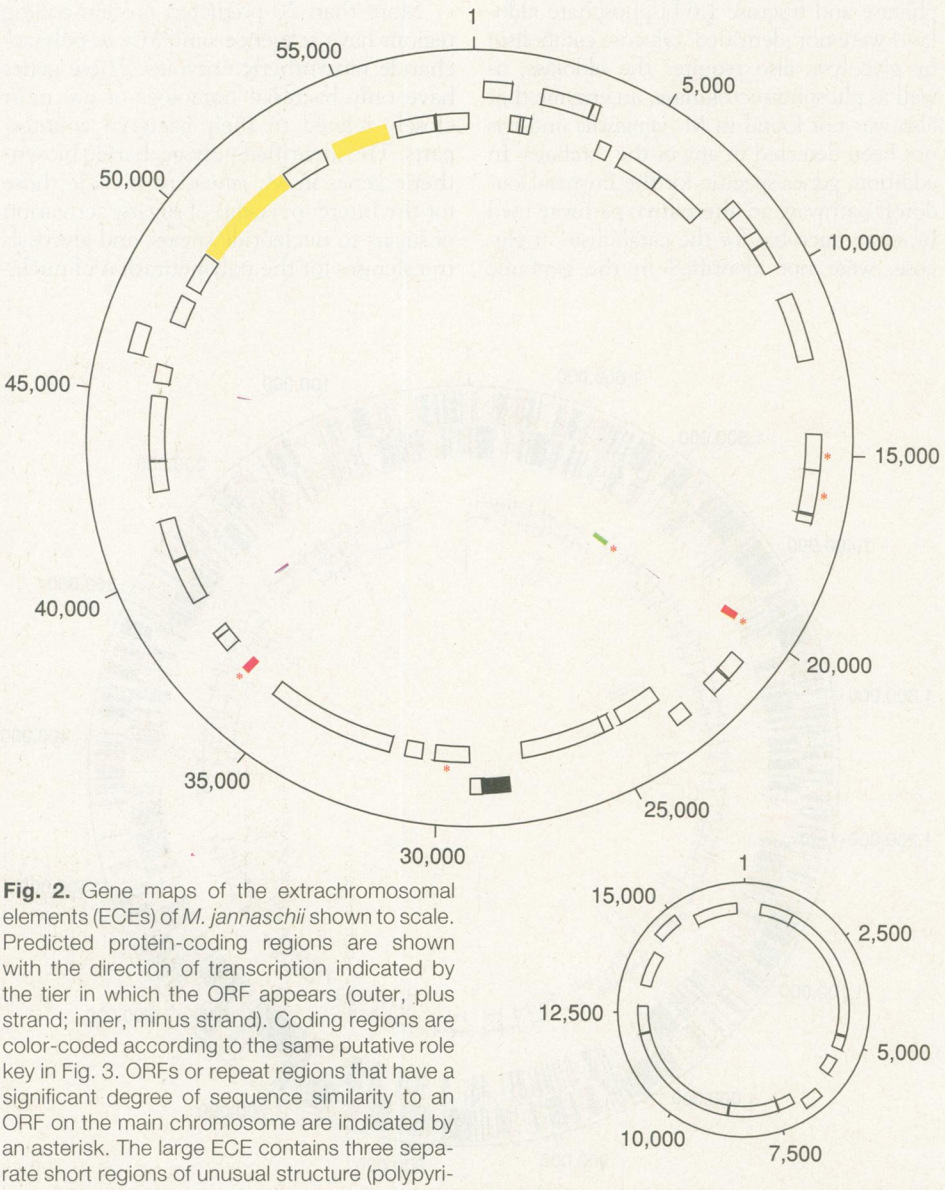


Fig. 2. Gene maps of the extrachromosomal elements (ECEs) of *M. jannaschii* shown to scale. Predicted protein-coding regions are shown with the direction of transcription indicated by the tier in which the ORF appears (outer, plus strand; inner, minus strand). Coding regions are color-coded according to the same putative role key in Fig. 3. ORFs or repeat regions that have a significant degree of sequence similarity to an ORF on the main chromosome are indicated by an asterisk. The large ECE contains three separate short regions of unusual structure (polypyrimidine or polypurine stretches of high G+C content) that show mirror symmetry. These regions are indicated in the third concentric circle by blue rectangles. In all three instances, the core of the mirror structure has the sequence CCCTCTCGGG-CTCTCCC (or its complement). Approximate mirror symmetry extends beyond this core, for stretches of (total length) 15, 19, and 21 pyrimidines (or purines) on either side of the center of symmetry. Mirror symmetry of this sort is characteristic of DNA capable of forming triple-stranded structures (*33*). The green rectangle in the innermost concentric circle of the large ECE indicates the location of the group C member of the IS*AMJ1* family of insertion elements.

YM95NC Attachment -- Page 46

ids encoded by four codons each have two tRNAs, one to cover the Y-, the other the R-ending, codons. Valine has a third tRNA, which is specific for the GUG codon; and alanine has three tRNAs (two of which are in the spacer regions separating the 16S and 23S subunits in the two ribosomal RNA operons). Leucine, serine, and arginine, all of which have six codons, each possess three corresponding tRNAs. The genes for the internal methionine and tryptophan tRNAs contain introns in their anticodon loops.

A tRNA also exists for selenocysteine (UGA codon). At least four genes in M. *jannaschii* contain internal stop codons that are potential selenocysteine codons: the α chain of formate dehydrogenase, coenzyme F420-reducing hydrogenase, B-chain tungsten formyl–methanofuran dehydrogenase, and a heterodisulfide reductase. Three genes with a putative role in selenocysteine metabolism were identified by their similarity to the *sel* genes from other organisms (Table 1).

Recognizable homologs for four of the aminoacyl-tRNA synthetases (glutamine, asparagine, lysine, and cysteine) were not identified in the M. *jannaschii* genome. The absence of a glutaminyl-tRNA synthetase is not surprising given that a number of organisms, including at least one archaeon, have none (*18*). In these instances, glutaminyl tRNA charging involves a post-charging conversion mechanism whereby the tRNA is charged by the glutamyl-tRNA synthetase with glutamic acid, which then is enzymatically converted to glutamine. A post-charging conversion is also involved in selenocysteine charging by the seryl-tRNA synthetase. A similar mechanism has been proposed for asparagine charging, but has not been demonstrated (*18*). The inability to find homologs of the lysine and cysteine aminoacyl-tRNA synthetases is surprising because bacterial and eukaryotic versions in each instance show clear homology.

Aminoacyl-tRNA synthetases of M. *jannaschii* and other Archaea resemble eukaryotic synthetases more closely than they resemble bacterial forms. The tryptophanyl synthetase is one of the more notable examples, because the M. *jannaschii* and eukaryotic versions do not appear to be specifically related to the bacterial version (*19*). Two versions of the glycyl synthetase are present in Bacteria, one that is very unlike the version found in Archaea and Eukaryotes and one that is an obvious homolog of it (*20*).

Eleven genes encoding subunits of the DNA-dependent RNA polymerase were identified in the M. *jannaschii* genome. The sequence similarity between the subunits and their homologs in *Sulfolobus acidocaldarius* supports the evolutionary unity of the archaeal polymerase complex (*21*). All of the subunits found in M. *jannaschii* show greater similarity to their eukaryotic counterparts than to the bacterial homologs. The genes encoding the five largest subunits (A′, A″, B′, B″, D) have homologs in all organisms. Six genes encode subunits shared only byArchaea and Eukaryotes (E, H, K, L, and N). The M. *jannaschii* homolog of the S. *acidocaldarius* subunit E is split into two genes designated E′ and E″. *Sulfolobus acidocaldarius* also contains two additional small subunits of RNA polymerase, designated G and F, that have no counterparts in either Bacteria or Eukaryotes. No homolog of these F subunits was identified in M. *jannaschii*.

The archaeal transcription initiation system is essentially the same as that found in Eukaryotes and is radically different from the bacterial version (*22*). The central molecules in the former systems are the TATA-binding protein (TBP) and transcription factor B (TFIIB and TFIIIB in Eukaryotes, or simply TFB). In the eukaryotic systems, TBP and TFB are parts of larger complexes, and additional factors (such as TFIIA and TFIIF) are used in the transcription process. However, the M. *jannaschii* genome does not contain obvious homologs of TFIIA and TFIIF.

Several components of the replication machinery were identified in M. *jannaschii*. The M. *jannaschii* genome appears to encode a single DNA-dependent polymerase that is a member of the B family of polymerases (*23*). The polymerase shares sequence similarity and three motifs with other family B polymerases, including eukaryotic α, γ, and ε polymerases, bacterial polymerase II, and several archaeal polymerases. However, it is not homologous to bacterial polymerase I and has no homologs in *H. influenzae* or M. *genitalium*.

Primer recognition by the polymerase takes place through a structure-specific DNA binding complex, the replication factor complex (rfc) (*23*). In humans and yeast, the rfc is composed of five proteins: a large subunit and four small subunits that have an associated adenosine triphosphatase (ATPase) activity stimulated by proliferating cell nuclear antigen (PCNA). Two genes in M. *jannaschii* are putative members of a eukaryotic-like replication factor complex. One of the genes in M. *jannaschii* is a putative homolog of the large subunit of the rfc, whereas the second is a putative homolog of one of the small subunits. Among Eukaryotes, the rfc proteins share sequence similarity in eight signature domains (*23*). Domain I is conserved only in the large subunit among Eukaryotes and is similar in sequence to DNA ligases. This domain is missing in the large-subunit homolog in M. *jannaschii*. The remaining domains in the two M. *jannaschii* genes are well conserved relative to the eukaryotic homologs. Two features of the sequence similarity in these domains are of particular interest. First, domain II (an ATPase domain) of the small-subunit homolog is split between two highly conserved amino acids (lysine and threonine) by an intervening sequence of unknown function. Second, the sequence of domain VI has regions that are useful for distinguishing between bacterial and eukaryotic rfc proteins (*23*); the rfc sequence for M. *jannaschii* shares the characteristic eukaryotic signature in this domain.

We attempted to identify an origin of replication by searching the M. *jannaschii* genome sequence with a variety of bacterial and eukaryotic replication-origin consensus sequences. Searches with oriC, ColE1, and autonomously replicating sequences from yeast (*23*) did not identify an origin of replication. With respect to the related cellular processes of replication initiation and cell division, the M. *jannaschii* genome contains two genes that are putative homologs of Cdc54, a yeast protein that belongs to a family of putative DNA replication initiation proteins (*24*). A third potential regulator of cell division in M. *jannaschii* is 55% similar at the amino acid level to *pelota*, a *Drosophila* protein involved in the regulation of the early phases of meitoic and mitotic cell division (*25*).

In contrast to the putative rfc complex and the initiation of DNA replication, the cell division proteins from M. *jannaschii* most resemble their bacterial counterparts (*26*). Two genes similar to that encoding FtsZ, a ubiquitous bacterial protein, are found in M. *jannaschii*. FtsZ is a polymer-forming, guanosine triphosphate (GTP)–hydrolyzing protein with tubulin-like elements; it is localized to the site of septation and forms a constricting ring between the dividing cells. One gene similar to FtsJ, a bacterial cell division protein of undetermined function, also is found in M. *jannaschii*. Three additional genes (MinC, MinD, and MinE) function in concert in Bacteria to determine the site of septation during cell division. In M. *jannaschii*, three MinD-like genes were identified, but none for MinC or MinE. Neither spindle-associated proteins characteristic of eukaryotic cell division nor bacterial mechanochemical enzymes necessary for partitioning the condensed chromosomes were detected in the M. *jannaschii* genome. Taken together, these observations raise the possibility that cell division in M. *jannaschii* might occur by

YM95NC Attachment -- Page 47

a mechanism specific for the Archaea.

The structural and functional conservation of the signal peptide of secreted proteins in Archaea, Bacteria, and Eukaryotes suggests that the basic mechanisms of membrane targeting and translocation may be similar among all three domains of life. The secretory machinery of M. *jannaschii* appears to be a rudimentary apparatus relative to that of bacterial and eukaryotic systems and consists of (i) a signal peptidase (SP) that cleaves the signal peptide of translocating proteins, (ii) a preprotein translocase that is the major constituent of the membrane-localized translocation channel, (iii) a ribonucleoprotein complex (signal recognition particle, SRP) that binds to the signal peptide and guides nascent proteins to the cell membrane, and (iv) a docking protein that acts as a receptor for the SRP. The 7S RNA component of the SRP from M. *jannaschii* shows a highly conserved structural domain shared by other Archaea, Bacteria, and Eukaryotes (*27*). However, the predicted secondary structure of the 7S RNA SRP component in Archaea is more like that found in Eukaryotes than in Bacteria (*27*). The SP and docking proteins from M. *jannaschii* are most similar to their eukaryotic counterparts; the translocase is most similar to the SecY translocation-associated protein in *Escherichia coli*.

A second distinct signal peptide is found in the flagellin genes of M. *jannaschii*. Alignment of flagellin genes from M. *voltae* (*28*) and M. *jannaschii* reveals a highly conserved NH_2-terminus (31 of the first 50 residues are identical in all of the mature flagellins). The peptide sequence of the M. *jannaschii* flagellin indicates that the protein is cleaved after the canonical Gly-12 position, and it is proposed to be similar to type-IV pilins of Bacteria (*28*).

Five histone genes are present in the M. *jannaschii* genome—three on the main chromosome and two on the large ECE. These genes are homologs of eukaryotic histones (H2a, H2b, H3, and H4) and of the eukaryotic transcription-related CAAT-binding factor CBF-A (*29*). The similarity between archaeal and eukaryotic histones suggests that the two groups of organisms resemble one another in the roles histones play both in genome supercoiling dynamics and in gene expression. The five M. *jannaschii* histone genes show greatest similarity among themselves even though a histone sequence is available from the closely related species, *Methanococcus voltae*. This intraspecfic similarity suggests that the gene duplications that produced the five histone genes occurred on the M. *jannaschii* lineage per se.

Self-splicing portions of a peptide sequence that generally encode a DNA endonuclease activity are called inteins, in analogy to introns (*30*). The sequences remaining after an intein is excised are called exteins, in analogy to exons. Exteins are spliced together after the excision of one or more inteins to form functional proteins. The biological significance and role of inteins are not clearly understood (*30*). Fourteen genes in the M. *jannaschii* genome contain 18 putative inteins, a significant increase in the approximately 10 intein-containing genes that have been described (*30*) (Table 2). The only previously described inteins in the Archaea are in the DNA polymerase genes of the Thermococcales (*30*). The M. *jannaschii* DNA polymerase gene has two inteins in the same locations as those in *Pyrococcus* sp. strain KOD1. In this case, the exteins exhibit 46% amino acid identity, whereas intein 2 of the two organisms has only 33% identity. This divergence suggests that intein 2 has

Table 2. Genes of *M. jannaschii* that contain inteins.

Gene no.	Putative identification	No. of inteins
MJ0043	Hypothetical protein (*Bacillus subtilis*)	1
MJ0262	Putative translation initiation factor, FUN12/bIF-2 family	1
MJ0542	Phosphoenolpyruvate synthase	1
MJ0682	Hypothetical protein (*Escherichia coli*)	1
MJ0782	Transcription initiation factor IIB	1
MJ0832	Anaerobic ribonucleoside-triphosphate reductase	2
MJ0885	DNA-dependent DNA polymerase, family B	2
MJ1042	DNA-dependent RNA polymerase, subunit A′	1
MJ1043	DNA-dependent RNA polymerase, subunit A″	1
MJ1054	UDP-glucose dehydrogenase	1
MJ1124	Hypothetical protein (*Saccharomyces cerevisiae*)	1
MJ1420	Glutamine–fructose-6-phosphate transaminase	1
MJ1442	Replication factor C, 37-kD subunit	3
MJ1512	Reverse gyrase	1

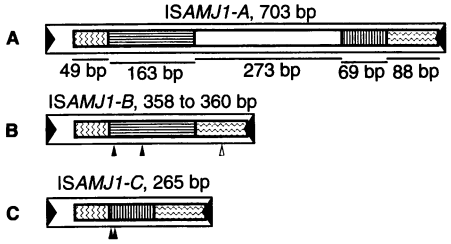


Fig. 4. Structure of a putative family of insertion sequence (IS) elements in the *M. jannaschii* genome. The family of elements has been named IS*AMJ1* and contains 11 members distributed among three groups (**A**, **B**, and **C**). The outer rectangle indicates the entire IS element; the interior rectangles indicate the predicted coding regions, oriented with the NH_2-termini to the left. DNA immediately adjacent to the NH_2-termini is 75 to 100% identical over 50 bp; DNA sequence similarity at the COOH-termini ends immediately after the stop codon. Black triangles indicate terminal inverted repeats. Fill patterns indicate which regions are missing from the elements in groups B and C. (A) Two copies of this family are 642 bp long and are 97% similar to each other at the nucleotide level. They appear to encode a protein 214 amino acids in length (ORFs MJ0017 and MJ1466) that are 27% identical to the IS*240* transposase of *B. thuringiensis* (GenBank accession number: M23741). (B) Eight copies of the family range in length from 358 to 360 bp and are missing a 342-bp internal region relative to the two members of group A. Some members of group B have putative frameshifts (indicated by solid arrows) and in-frame UGA codons (indicated by open arrows). (C) The single copy in group C is 265 bp in length and occurs on the large ECE. The 436-bp internal region missing from this element is different than that of the members of group B.

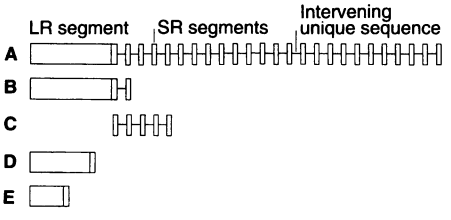


Fig. 5. Structure of a multicopy repetitive element in the *M. jannaschii* genome. Of the 18 copies identified on the main chromosome, 7 are oriented in one direction (plus strand) and 11 are oriented in the opposite strand. Each element consists of a long, 391- to 425-bp repeat segment (designated LR) followed by up to 25 short, 27- to 28-bp repeat segments (designated SR). Each SR segment is separated by 31 to 51 bp of sequence that is unique within and between each complete repeat element. (**A**) The longest repeat element has an LR segment followed by 25 SR segments and spans more than 2 kbp, and (**B**) the shortest complete element has an LR segment followed by two SR segments. (**C**) One element is present in the genome with five SR segments and no LR component. (**D** and **E**) The LR segments of two elements in the genome are truncated at the end adjacent to the SR segments; both are followed by a single SR segment.

YM95NC Attachment -- Page 48

not been recently (laterally) transferred between the Thermococcales and *M. jannaschii*. In contrast, the intein 1 sequences are 56% identical, more than that of the gene containing them, and comparable to the divergence of inteins within the Thermococcales. This high degree of sequence similarity might be the result of an intein transfer more recent than the splitting of these species. The large number of inteins found in *M. jannaschii* led us to question whether these inteins have been increasing in number by moving within the genome. If this were so, we would expect to find some pairs of inteins that are particularly similar. Comparisons of these and other available intein sequences showed that the closest relationships are those noted above linking the DNA polymerase inteins to correspondingly positioned elements in the Thermococcales. Within *M. jannaschii*, the highest identity observed was 33% for a 380-bp portion of two inteins. This finding suggests that the diversification of the inteins predates the divergence of the *M. jannaschii* and *Pyrococcus* DNA polymerases.

Three families of repeated genetic elements were identified in the *M. jannaschii* genome. Within two of the families, at least two members were identified as ORFs with a limited degree of sequence similarity to bacterial transposases. Members of the first family, designated ISAMJ1, are repeated 10 times on the main chromosome and once on the large ECE (Fig. 4). There is no sequence similarity between the IS elements in *M. jannaschii* and the ISM1 mobile element described previously for *Methanobrevibacter smithii* (*31*). Two members of this family were identified as ORFs and are 27% identical (at the amino acid sequence level) to a transposase from *Bacillus thuringiensis* (IS*240*; GenBank accession number M23741). Relative to these two members, the remaining members of the ISAMJ1 family are missing an internal region of several hundred nucleotides (Fig. 4). With one exception, all members of this family end with 16-bp terminal inverted repeats typical of insertion sequences. One member is missing the terminal repeat at its 5′ end. The second family consists of two ORFs that are identical across 928 bp. The ORFs are 23% identical at the amino acid sequence level to the COOH-terminus of a transposase from *Lactococcus lactis* (IS982; GenBank accession number L34754). Neither of the members of the second family contains terminal inverted repeats.

Eighteen copies of the third family of repeated genetic structures (Fig. 5) are distributed fairly evenly around the *M. jannaschii* genome (Fig. 3). Unlike the genetic elements described above, none of the components of this repeat unit appears to have coding potential. The repeat structure is composed of a long segment followed by 1 to 25 tandem repetitions of a short segment. The short segments are separated by sequence that is unique within and among the complete repeat structure. Three similar types of short segments were identified; however, the type of short repeat is consistent within each repeat structure, except for variation of the last short segment in six repeat structures. Similar tandem repeats of short segments have been observed in Bacteria and other Archaea (*32*) and have been

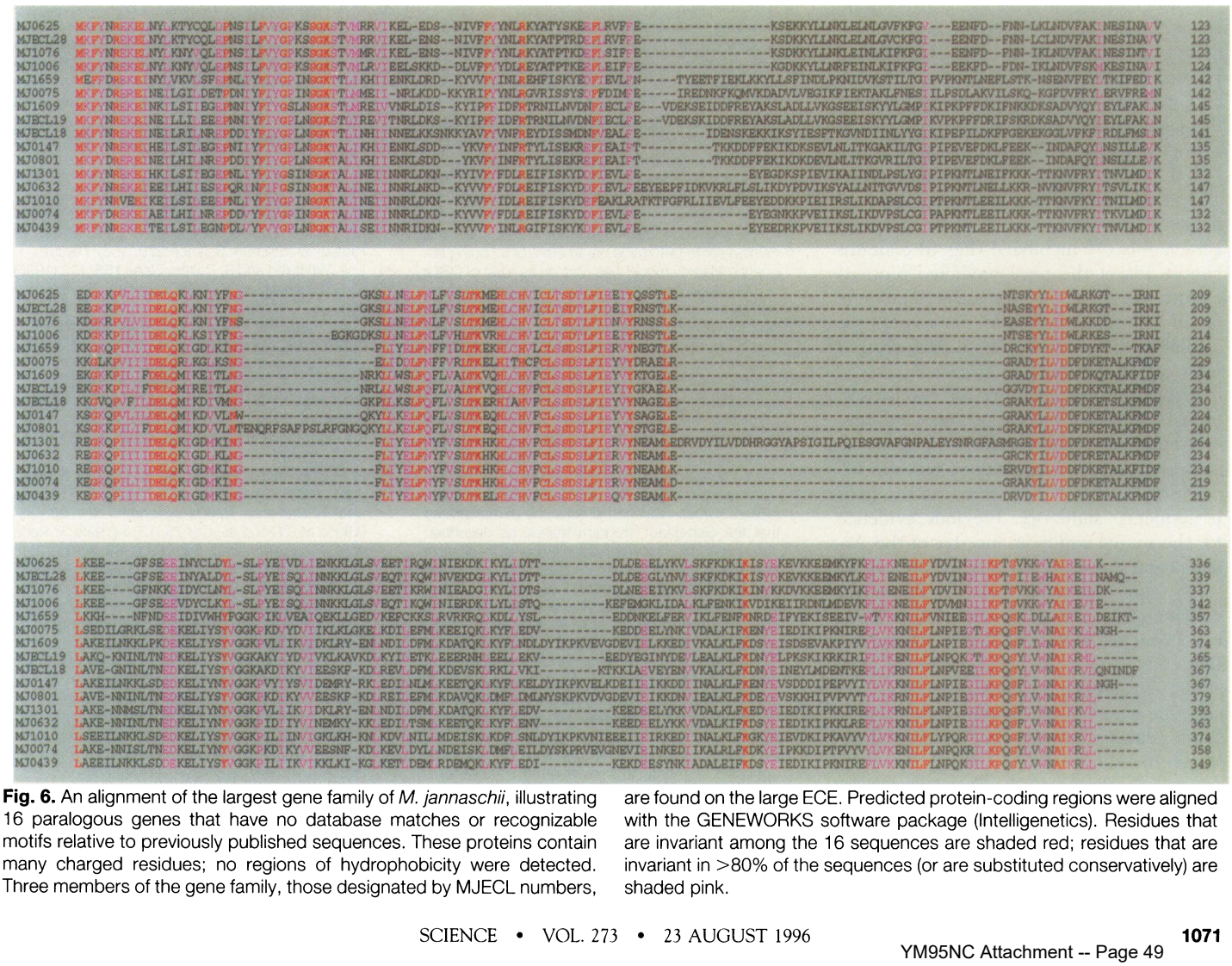


Fig. 6. An alignment of the largest gene family of *M. jannaschii*, illustrating 16 paralogous genes that have no database matches or recognizable motifs relative to previously published sequences. These proteins contain many charged residues; no regions of hydrophobicity were detected. Three members of the gene family, those designated by MJECL numbers, are found on the large ECE. Predicted protein-coding regions were aligned with the GENEWORKS software package (Intelligenetics). Residues that are invariant among the 16 sequences are shaded red; residues that are invariant in >80% of the sequences (or are substituted conservatively) are shaded pink.

YM95NC Attachment -- Page 49

hypothesized to participate in chromosome partitioning during cell division.

The 16-kbp ECE from M. *jannaschii* contains 12 ORFs, none of which had a significant full-length match to any published sequence (Fig. 2). The 58-kbp ECE contains 44 predicted protein-coding regions, 5 of which had matches to genes in the database. Two of the genes are putative archael histones, one is a sporulation-related protein (SOJ protein), and two are type I restriction modification enzymes. There are several instances in which predicted protein-coding regions or repeated genetic elements on the large ECE have similar counterparts on the main chromosome of M. *jannaschii* (Fig. 2). The degree of nucleotide sequence similarity between genes present on both the ECE and the main chromosome ranges from 70 to 90%, suggesting that there has been relatively recent exchange of at least some genetic material between the large ECE and the main chromosome.

All the predicted protein-coding regions from M. *jannaschii* were searched against each other in order to identify families of paralogous genes (genes related by gene duplication, not speciation). The initial criterion for grouping paralogs was >30% amino acid sequence identity over 50 consecutive amino acid residues. Groups of predicted protein-coding regions were then aligned and inspected individually to ensure that the sequence similarity extended over most of their lengths. This curatorial process resulted in the identification of more than 100 gene families, half of which have no database matches. The largest identified gene family (16 members) (Fig. 6) contains almost 1% of the total predicted protein-coding regions in M. *jannaschii*. The gene family alignments for M. *jannaschii* are available on the World Wide Web (http://www.tigr.org/tdb/mdb/mjdb/).

Despite the availability for comparison of two complete bacterial genomes and several hundred megabase pairs of eukaryotic sequence data, the majority of genes in M. *jannaschii* cannot be identified on the basis of sequence similarity. Previous evidence for the shared common ancestry of the Archaea and Eukaryotes was based on a small set of gene sequences (*2*). The complete genome of M. *jannaschii* allows us to move beyond a "gene by gene" approach to one that encompasses the larger picture of metabolic capacity and cellular systems. The anabolic genes of M. *jannaschii* (especially those related to energy production and nitrogen fixation) reveal an ancient metabolic world shared largely by Bacteria and Archaea. That many basic autotrophic pathways appear to have a common evolutionary origin suggests that the most recent universal common ancestor to all three domains of extant life had the capacity for autotrophy. The Archaea and Bacteria also share structural and organizational features that the most recent universal prokaryotic ancestors also likely possessed, such as circular genomes and genes organized as operons. In contrast, the cellular information-processing and secretion systems in M. *jannaschii* demonstrate the common ancestry of Eukaryotes and Archaea. Although components of these systems are present in all three domains, their apparent refinement over time—especially transcription and translation—indicate that the Archaea and Eukaryotes share a common evolutionary trajectory independent of the lineage of Bacteria.

REFERENCES AND NOTES

1. G. E. Fox *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4537 (1977); C. R. Woese and G. E. Fox, *ibid.*, p. 5088; C. R. Woese *et al.*, *ibid.* **87**, 4576 (1990).
2. N. Iwabe *et al.*, *ibid.* **86**, 9355 (1989); J. P. Gogarten *et al.*, *ibid.*, p. 6661; W. Zillig *et al.*, *Endocytobiosis Cell Res.* **6**, 1 (1989); J. R. Brown and W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2441 (1995).
3. R. D. Fleischmann *et al.*, *Science* **269**, 496 (1995); C. M. Fraser *et al.*, *ibid.* **270**, 397 (1995).
4. N. Williams, *ibid.* **272**, 481 (1996).
5. M. D. Adams *et al.*, *Nature* **377**, 3 (1995); R. Wilson *et al.*, *ibid.* **368**, 32 (1994).
6. W. Jones *et al.*, *Arch. Microbiol.* **136**, 254 (1983).
7. G. Sutton *et al.*, *Genome Sci. Tech.* **1**, 9 (1995).
8. The statistical prediction of *M. jannaschii* genes was performed with GeneMark [M. Borodovsky and J. McIninch, *Comput. Chem.* **17**, 123 (1993)]. Regular GeneMark uses nonhomogeneous Markov models derived from a training set of coding sequences and ordinary Markov models derived from a training set of noncoding sequences. Only a single 16*S* ribosomal RNA sequence of *M. jannaschii* was available in the public sequence databases before the whole genome sequence described here. Thus, the initial training set to determine parameters of a coding sequence Markov model was chosen as a set of ORFs >1000 nucleotides (nt). As an initial model for noncoding sequences, a zero-order Markov model with genome-specific nucleotide frequencies was used. The initial models were used at the first prediction step. The results of the first prediction were then used to compile a set of putative genes used at the second training step. Alternate rounds of training and predicting were continued until the set of predicted genes stabilized and the parameters of the final fourth-order model of coding sequences were derived. The regions predicted as noncoding were then used as a training set for a final model for noncoding regions. Cross-validation simulations demonstrated that the GeneMark program trained as described above was able to correctly identify coding regions of at least 96 nt in 94% of the cases and noncoding regions of the same length in 96% of the cases. These values assume that the self-training method produced correct sequence annotation for compiled control sets. Comparison with the results obtained by searches against a nonredundant protein database (*3*) demonstrated that almost all genes identified by sequence similarity were predicted by the GeneMark program as well. This observation provides additional confidence in genes predicted by GeneMark whose protein translations did not show significant similarity to known protein sequences. The predicted protein-coding regions were searched against the Blocks database [S. Henikoff and J. G. Henikoff, *Genomics* **19**, 97 (1994)] by means of BLIMPS [J. C. Wallace and S. Henikoff, *Comput. Appl. Biosci.* **8**, 249 (1992)] to verify putative identifications and to identify potential functional motifs in predicted protein-coding regions that had no database match. Genes were assigned to known metabolic pathways. When a gene appeared to be missing from a pathway, the unassigned ORFs and the complete *M. jannaschii* genome sequence were searched with specific query sequences or motifs from the Blocks database. Hydrophobicity plots were performed on all predicted protein-coding regions by means of the Kyte-Doolittle algorithm [J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982)] to identify potentially functionally relevant signatures in these sequences. The results of the Blocks and Kyte-Doolittle analyses are available on the World Wide Web (http://www.tigr. org/tdb/mdb/mjdb/mjdb.html).
9. H. Zhao *et al.*, *Arch. Microbiol.* **150**, 178 (1988).
10. A. A. DiMarco *et al.*, *Annu. Rev. Biochem.* **59**, 355 (1990).
11. N. Belay *et al.*, *Nature* **312**, 286 (1984).
12. H. G. Wood *et al.*, *Trends Biochem. Sci.* **11**, 14 (1986).
13. M. Blaat, *Antonie Leewenhoek* **66**, 187 (1994).
14. E. Hartmann and H. König, *Arch. Microbiol.* **151**, 274 (1989).
15. X. M. Jiang *et al.*, *Mol. Microbiol.* **5**, 695 (1991).
16. K. Lechner *et al.*, *J. Mol. Evol.* **29**, 20 (1989); A. K. E. Köpke and B. Wittmann-Liebold. *Can. J. Microbiol.* **35**, 11 (1989).
17. P. Keeling *et al.*, *Syst. Appl. Microbiol.*, in press.
18. M. Wilcox, *Eur. J. Biochem.* **11**, 405 (1969); N. C. Martin *et al.*, *J. Mol. Biol.* **101**, 285 (1976); N. C. Martin *et al.*, *Biochemistry* **16**, 4672 (1977); A. Schon *et al.*, *Biochimie* **70**, 391 (1988); D. Soll and U. Raj Bhandary, Eds. *tRNA: Structure, Biosynthesis, and Function* (American Society for Microbiology, Washington, DC, 1995).
19. R. de Pouplana *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 166 (1996).
20. E. A. Wagner *et al.*, *J. Bacteriol.* **177**, 5179 (1995); D. T. Logan *et al.*, *EMBO J.* **14**, 4156 (1995).
21. C. R. Woese and R. S. Wolfe, Eds. *The Bacteria* (Academic Press, New York, 1985), vol. 8; D. Langer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5768 (1995); M. Lanzendoerfer *et al.*, *Syst. Appl. Microbiol.* **16**, 656 (1994).
22. H.-P. Klenk and W. F. Doolittle, *Curr. Biol.* **4**, 920 (1994).
23. A. Bernard *et al.*, *EMBO J.* **6**, 4219 (1987); G. Cullman *et al.*, *Mol. Cell. Biol.* **15**, 4661 (1995); T. Uemori *et al.*, *J. Bacteriol.* **177**, 2164 (1995); M. Delarue *et al.*, *Protein Eng.* **3**, 461 (1990); K. A. Gavin, M. Hidaka, B. Stillman, *Science* **270**, 1667 (1995).
24. L. A. Whitbred and S. Dalton, *Gene* **155**, 113 (1995).
25. C. G. Eberhart and S. A. Wasserman, *Development* **121**, 3477 (1995).
26. L. Rothfield and C.-R Zhao, *Cell* **84**, 183 (1996); J. Lutkenhaus, *Curr. Opin. Genet. Dev.* **3**, 783 (1993).
27. B. P. Kaine and V. L. Merkel, *J. Bacteriol.* **171**, 4261 (1989); M. A. Poritz *et al.*, *Cell* **55**, 4 (1988).
28. D. M. Faguy *et al.*, *Can. J. Microbiol.* **40**, 67 (1994): M. L. Kalmokoff *et al.*, *Arch. Microbiol.* **157**, 481 (1992).
29. K. Sandman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5788 (1990).
30. P. M. Kane *et al.*, *Science* **250**, 651 (1990); R. Hirata *et al.*, *J. Biol. Chem.* **265**, 6726 (1990); A. A. Cooper and T. Stevens, *Trends Biochem. Sci.* **20**, 351 (1995); M.-Q Xu *et al.*, *Cell* **75**, 1371 (1993); F. Perler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5577 (1992); Cooper *et al.*, *EMBO J.* **12**, 2575 (1993); F. Michel *et al.*, *Biochimie* **64**, 867 (1982); S. Pietrokovski, *Protein Sci.* **3**, 2340 (1994). Most inteins in the *M. jannaschii* genome were identified by (i) similarity of the bounding exteins to other proteins, (ii) similarity of the inteins to those previously described, (iii) presence of the dodecapeptide endonuclease motifs, and (iv) canonical intein-extein junction sequences. In two instances (MJ0832 and MJ0043), the similarity to other database sequences did not unambiguously define the NH_2-terminal extein-intein junction, so it was necessary to rely on consensus sequences to select the putative site. The inteins in MJ1042 and MJ0542 have previously uncharacterized COOH-terminal splice junctions, GNC and FNC, respectively.
31. P. T. Hamilton *et al.*, *Mol. Gen. Genet.* **200**, 47 (1985).
32. F. J. M. Mojica *et al.*, *Mol. Microbiol.* **17**, 85 (1995).

YM95NC Attachment -- Page 50

33. G. Felsenfeld *et al.*, *J. Am. Chem. Soc.* **79**, 2023 (1957); A. G. Letai *et al.*, *Biochemistry* **27**, 9108 (1988).
34. M. Riley, *Microbiol. Rev.* **57**, 862 (1993).
35. Supported in part by Department of Energy Cooperative Agreements DE-FC02-95ER61962 (J.C.V.) and DEFC02-95ER61963 (C.R.W. and G.J.O), NASA grant NAGW 2554 (C.R.W.), and a core grant to TIGR from Human Genome Sciences. G.J.O. is the recipient of the National Science Foundation Presidential Young Investigator Award (DIR 89-57026). M.B. is supported by National Institutes of Health grant GM00783. We thank M. Heaney, C. Gnehm, R. Shirley, J. Slagel, and W. Hayes for software and database support; T. Dixon and V. Sapiro for computer system support; K. Hong and B. Stader for laboratory assistance; and B. Mukhopadhyay for helpful discussions. The *M. jannaschii* source accession number is DSM 2661, and the cells were a gift from P. Haney (Department of Microbiology, University of Illinois).

RESEARCH ARTICLES

Universal Quantum Simulators

Seth Lloyd

Feynman's 1982 conjecture, that quantum computers can be programmed to simulate any local quantum system, is shown to be correct.

Over the past half century, the logical devices by which computers store and process information have shrunk by a factor of 2 every 2 years. A quantum computer is the end point of this process of miniaturization—when devices become sufficiently small, their behavior is governed by quantum mechanics. Information in conventional digital computers is stored on capacitors. An uncharged capacitor registers a 0 and a charged capacitor registers a 1. Information in a quantum computer is stored on individual spins, photons, or atoms. An atom can itself be thought of as a tiny capacitor. An atom in its ground state is analogous to an uncharged capacitor and can be taken to register a 0, whereas an atom in an excited state is analogous to a charged capacitor and can be taken to register a 1.

So far, quantum computers sound very much like classical computers; the only use of quantum mechanics has been to make a correspondence between the discrete quantum states of spins, photons, or atoms and the discrete logical states of a digital computer. Quantum systems, however, exhibit behavior that has no classical analog. In particular, unlike classical systems, quantum systems can exist in superpositions of different discrete states. An ordinary capacitor can be either charged or uncharged, but not both: A classical bit is either 0 or 1. In contrast, an atom in a quantum superposition of its ground and excited state is a quantum bit that in some sense registers both 0 and 1 at the same time. As a result, quantum computers can do things that classical computers cannot.

Classical computers solve problems by using nonlinear devices such as transistors to perform elementary logical operations on the bits stored on capacitors. Quantum computers can also solve problems in a similar fashion; nonlinear interactions between quantum variables can be exploited to perform elementary quantum logical operations. However, in addition to ordinary classical logical operations such as AND, NOT, and COPY, quantum logic includes operations that put quantum bits in superpositions of 0 and 1. Because quantum computers can perform ordinary digital logic as well as exotic quantum logic, they are in principle at least as powerful as classical computers. Just what problems quantum computers can solve more efficiently than classical computers is an open question.

Since their introduction in 1980 (*1*) quantum computers have been investigated extensively (*2*–*29*). A comprehensive review can be found in (*15*). The best known problem that quantum computers can in principle solve more efficiently than classical computers is factoring (*14*). In this article I present another type of problem that in principle quantum computers could solve more efficiently than a classical computer—that of simulating other quantum systems. In 1982, Feynman conjectured that quantum computers might be able to simulate other quantum systems more efficiently than classical computers (*2*). Quantum simulation is thus the first classically difficult problem posed for quantum computers. Here I show that a quantum computer can in fact simulate quantum systems efficiently as long as they evolve according to local interactions.

Feynman noted that simulating quantum systems on classical computers is hard. Over the past 50 years, a considerable amount of effort has been devoted to such simulation. Much information about a quantum system's dynamics can be extracted from semiclassical approximations (when classical solutions are known), and ground state properties and correlation functions can be extracted with Monte Carlo methods (*30*–*32*). Such methods use amounts of computer time and memory space that grow as polynomial functions of the size of the quantum system of interest (where size is measured by the number of variables—particles or lattice sites, for example—required to characterize the system). Problems that can be solved by methods that use polynomial amounts of computational resources are commonly called tractable; problems that can only be solved by methods that use exponential amounts of resources are commonly called intractable. Feynman pointed out that the problem of simulating the full time evolution of arbitrary quantum systems on a classical computer is intractable: The states of a quantum system are wave functions that lie in a vector space whose dimension grows exponentially with the size of the system. As a result, it is an exponentially difficult problem merely to record the state of a quantum system, let alone integrate its equations of motion. For example, to record the state of 40 spin-½ particles in a classical computer's memory requires $2^{40} \approx 10^{12}$ numbers, whereas to calculate their time evolution requires the exponentiation of a $2^{40} \times 2^{40}$ matrix with $\approx 10^{24}$ entries. Feynman asked whether it might be possible to bypass this exponential explosion by having one quantum system simulate another directly, so that the states of the simulator obey the same equations of motion as the states of the simulated system. Feynman gave simple examples of one quantum system simulating another and conjectured that there existed a class of universal quantum simulators capable of simulating any quantum system that evolved according to local interactions.

The answer to Feynman's question is, yes. I will show that a variety of quantum systems, including quantum computers, can be "programmed" to simulate the behavior of arbitrary quantum systems whose dynamics are determined by local interactions. The programming is accomplished by inducing interactions between the variables of the simulator that imitate the interactions between the variables of the system to be simulated. In effect, the dynamics of the properly programmed simulator and the dynamics of the system to be simulated are one and the same to within any desired accuracy. So, to simulate the time evolution of 40 spin-½ particles over time t requires a simulator with 40 quantum bits evolving

The author is at the D'Arbeloff Laboratory for Information Systems and Technology, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: slloyd@mit.edu

>Concatenated_sequences Concatenation of 2 sequences

CAAGAATTCATGGAATTAAGGATAAAAAAGAATTGGATAAGATTGTTGAGTGGGCTTTAAAGAAAGCGGCTTTGTGGGAT
GAGGTTAAAGATGAACTGCATAAAAACGCTCTATCTCTCTCTGGAGGACAACAGCAGAGGTTATGTATAGCGAGAGCGAT
AGCGGTTAAGCCAGAGGTTTTATTGACGGATGAACCAACATCTGCCTTAGACCCTATCTCCACATTAAAGATAGAGGAGT
TAATGGTTGAGTTAGCTAAAGATTATACGATTGTTGTTGTTACCCACAACATGCAGCAGGCAAGTAGGGTTTCTGATTAC
ACTGCCTTTTTCTTAATGGGGAAATTAATTGAGTTTGGAGAGACAGAGCAGATATTCCTAAATCCACAGAAGAAGGAGAC
AGATGACTACATTAGTGGTAGGTTTGGTTAAGTATCATGCTAGCTTTTCAATTCAATTCATCATTTTTTTTTTATTCTTT
TTTTTGATTTCGGTTTCTTTGAAATTTTTTTGATTCGGTAATCTCCGAACAGAAGGAAGAACGAAGGAAGGAGCACAGAC
TTAGATTGGTATATATACGCATATGTAGTGTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAACA
AAAACCTGCAGGAAACGAAGATAAATCATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGC
TGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAACTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTAC
TGGAGTTAGTTGAAGCATTAGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGAG
GGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGG
TAATACAGTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGG
TGGGCCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCAGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCA
GAATTGTCATGCAAGGGCTCCCTATCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTT
TGTTATCGGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGG
GTTTAGATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACATTATT
ATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGGGAAGC
ATATTTGAGAAGATGCGGCCAGCAAAACTAAAAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACAAATTAGAG
CTTCAATTTAATTATATCAGTTATTACCC

```
BLASTN 2.8.0+
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and
Webb Miller (2000), "A greedy algorithm for aligning DNA
sequences", J Comput Biol 2000; 7(1-2):203-14.


Reference for database indexing: Aleksandr Morgulis, George
Coulouris, Yan Raytselis, Thomas L. Madden, Richa Agarwala,
Alejandro A. Schaffer (2008), "Database Indexing for
Production MegaBLAST Searches", Bioinformatics 24:1757-1764.


RID: HNC4ZBCT015


Database: Nucleotide collection (nt)
           48,103,425 sequences; 181,249,278,927 total letters
Query= ERCC-00095_Plasmid_PCR_Product Concatenation of 2 sequences

Length=499


                                                                  Score     E
Sequences producing significant alignments:                      (Bits)  Value

KC702204.1  Synthetic construct external RNA control ERCC-0009...  922     0.0
DQ516759.1  Methanocaldococcus jannaschii spike-in control MJ-...  900     0.0
L77117.1  Methanocaldococcus jannaschii DSM 2661, complete genome  900     0.0
CP001787.1  Methanocaldococcus vulcanius M7, complete genome       702     0.0
CP001696.1  Methanocaldococcus fervens AG86, complete genome       619     2e-173
CP001901.1  Methanocaldococcus sp. FS406-22, complete genome       582     3e-162
CP009149.1  Methanocaldococcus bathoardescens strain JH146, co...  507     2e-139

ALIGNMENTS
>KC702204.1 Synthetic construct external RNA control ERCC-00095 sequence
Length=521

 Score = 922 bits (499),  Expect = 0.0
 Identities = 499/499 (100%), Gaps = 0/499 (0%)
 Strand=Plus/Plus

Query  1    TTGATGATGATACTTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    TTGATGATGATACTTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTG  60

Query  61   GATTTAGGAATATCTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GATTTAGGAATATCTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGG  120
```

```
Query  121  CAGTGTAATCAGAAACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTAT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CAGTGTAATCAGAAACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTAT  180

Query  181  AATCTTTAGCTAACTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  AATCTTTAGCTAACTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGG  240

Query  241  CAGATGTTGGTTCATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATAC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  CAGATGTTGGTTCATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATAC  300

Query  301  ATAACCTCTGCTGTTGTCCTCCAGAGAGAGATAGAGCGTTTTTATGCAGTTCATCTTTAA  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  ATAACCTCTGCTGTTGTCCTCCAGAGAGAGATAGAGCGTTTTTATGCAGTTCATCTTTAA  360

Query  361  CCTCATCCCACAAAGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTT  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CCTCATCCCACAAAGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTT  420

Query  421  TATCCTTAATTCCATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  TATCCTTAATTCCATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAA  480

Query  481  AAGGATTTGGTTTTGAAAT  499
            |||||||||||||||||||
Sbjct  481  AAGGATTTGGTTTTGAAAT  499


>DQ516759.1 Methanocaldococcus jannaschii spike-in control MJ-500-42 genomic
sequence
Length=500

 Score = 900 bits (487),  Expect = 0.0
 Identities = 496/500 (99%), Gaps = 1/500 (0%)
 Strand=Plus/Plus

Query  1    TTGATGATGATACTTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    TTGATGATGATACTTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTG  60

Query  61   GATTTAGGAATATCTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GATTTAGGAATATCTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGG  120

Query  121  CAGTGTAATCAGAAACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTAT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  121  CAGTGTAATCAGAAACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTAT  180

Query  181  AATCTTTAGCTAACTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  AATCTTTAGCTAACTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGG  240

Query  241  CAGATGTTGGTTCATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATAC  300
            |||||||||||||||| |||||||||||||||||||| ||||||||||||||||||||
Sbjct  241  CAGATGTTGGTTCATCCATCAATAAAACCTCTGGCTTAACTGCTATCGCTCTCGCTATAC  300

Query  301  ATAACCTCTGCTGTTGTCCTCCAGAGAGAGATAGAGCGTTTTTATGCAGTTCATCTTTAA  360
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  301  ATAACCTCTGCTGTTGTCCTCCAGAGAGAGATAAAGCGTTTTTATGCAGTTCATCTTTAA  360

Query  361  CCTCATCCCACAAAGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTT  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CCTCATCCCACAAAGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTT  420

Query  421  TATCCTTAATTCCATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  TATCCTTAATTCCATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAA  480

Query  481  AAGGATTTGG-TTTTGAAAT  499
            |||||||||| |||||||||
Sbjct  481  AAGGATTTGGCTTTTGAAAT  500


>L77117.1 Methanocaldococcus jannaschii DSM 2661, complete genome
Length=1664970

 Score = 900 bits (487),  Expect = 0.0
 Identities = 496/500 (99%), Gaps = 1/500 (0%)
 Strand=Plus/Plus

Query  1       TTGATGATGATACTTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTG  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  941913  TTGATGATGATACTTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTG  941972

Query  61      GATTTAGGAATATCTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGG  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  941973  GATTTAGGAATATCTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGG  942032

Query  121     CAGTGTAATCAGAAACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTAT  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  942033  CAGTGTAATCAGAAACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTAT  942092

Query  181     AATCTTTAGCTAACTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGG  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  942093  AATCTTTAGCTAACTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGG  942152

Query  241     CAGATGTTGGTTCATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATAC  300
               |||||||||||||||| |||||||||||||||||||| ||||||||||||||||||||
Sbjct  942153  CAGATGTTGGTTCATCCATCAATAAAACCTCTGGCTTAACTGCTATCGCTCTCGCTATAC  942212

Query  301     ATAACCTCTGCTGTTGTCCTCCAGAGAGAGATAGAGCGTTTTTATGCAGTTCATCTTTAA  360
               |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  942213  ATAACCTCTGCTGTTGTCCTCCAGAGAGAGATAAAGCGTTTTTATGCAGTTCATCTTTAA  942272

Query  361     CCTCATCCCACAAAGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTT  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  942273  CCTCATCCCACAAAGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTT  942332

Query  421     TATCCTTAATTCCATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  942333  TATCCTTAATTCCATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAA  942392

Query  481     AAGGATTTGG-TTTTGAAAT  499
               |||||||||| |||||||||
Sbjct  942393  AAGGATTTGGCTTTTGAAAT  942412


>CP001787.1 Methanocaldococcus vulcanius M7, complete genome
Length=1746329

 Score = 702 bits (380),  Expect = 0.0
 Identities = 451/486 (93%), Gaps = 1/486 (0%)
 Strand=Plus/Minus

Query  14      TTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTGGATTTAGGAATAT  73
               |||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Sbjct  810978  TTAACCAAACCTACCACTGATGTAATCATCTGTCTCCTTCTTCTGTGGATTTAGGAATAT  810919

Query  74      CTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGGCAGTGTAATCAGA  133
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  810918  CTGCTCCGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGGCAGTGTAATCAGA  810859

Query  134     AACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTATAATCTTTAGCTAA  193
               ||||||||||||||| |||||||||||||||||||| |||||||||||| || |||||
Sbjct  810858  AACCCTACTTGCCTGTTGCATGTTGTGGGTAACAACAACGATCGTATAATCCTTGGCTAA  810799

Query  194     CTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGGCAGATGTTGGTTC  253
               ||||||||||||||| ||||| ||||| || || ||||| ||||| ||||| ||||||||
Sbjct  810798  CTCAACCATTAACTCTTCTATTTTTAAGGTAGATATAGGATCTAATGCAGAAGTTGGTTC  810739

Query  254     ATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATACATAACCTCTGCTG  313
               |||| |||||||||||||||||||| |||||||||||||| |||||||| ||||||||
```

```
Sbjct  810738  ATCCATCAATAAAACCTCTGGCTTAACTGCTATCGCTCTCGCGATACATAATCTCTGCTG  810679

Query  314     TTGTCCTCCAGAGAGAGATAGAGCGTTTTTATGCAGTTCATCTTTAACCTCATCCCACAA  373
               || || |||||||||||| |||||||||||||| |||||||||||||||||||||||||
Sbjct  810678  CTGCCCCCCAGAGAGAGATAAAGCGTTTTTATGCAATTCATCTTTAACCTCATCCCACAA  810619

Query  374     AGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTTTATCCTTAATTCC  433
               || ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  810618  TGCAGCTTTTTTTAATGCCCACTCAACAATCTTATCCAATTCTTTTTTATCCTTAATTCC  810559

Query  434     ATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAAAAGGATTTGG-TT  492
               || |||||||||||||||||||||||||| || |||||||||||||| || || ||
Sbjct  810558  GTGGATTCTTGGGCCAAATGCAACATTATCATATATACTCATAGCAAAAGGGTTGGGCTT  810499

Query  493     TTGAAA  498
               ||||||
Sbjct  810498  TTGAAA  810493


>CP001696.1 Methanocaldococcus fervens AG86, complete genome
Length=1485061

 Score = 619 bits (335),  Expect = 2e-173
 Identities = 436/486 (90%), Gaps = 1/486 (0%)
 Strand=Plus/Minus

Query  14      TTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTGGATTTAGGAATAT  73
               ||||||||| |||||||||| || ||||| ||||||||||||||||||||||||||||||
Sbjct  852897  TTAACCAAATCTACCACTAATATAATCATCCGTCTCCTTCTTCTGTGGATTTAGGAATAT  852838

Query  74      CTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGGCAGTGTAATCAGA  133
               |||||| |||||||||| |||||||||||||||| ||||||||||||| ||||||||
Sbjct  852837  CTGCTCCGTCTCTCCAAATTCAATTAATTTCCCCATCAAGAAAAAGGCAGTATAATCAGA  852778

Query  134     AACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTATAATCTTTAGCTAA  193
               ||||||||| ||||| |||||||||| |||||||||| || ||||| || ||||||||
Sbjct  852777  AACCCTACTCGCCTGTTGCATGTTGTGAGTAACAACAACTATTGTATAGTCCTTAGCTAA  852718

Query  194     CTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGGCAGATGTTGGTTC  253
               |||||||| |||||||||| ||||| || || |||||||||| ||||| ||||||||
Sbjct  852717  TTCAACCATCAACTCCTCTATTTTTAAGGTAGATATAGGGTCTAATGCAGAGGTTGGTTC  852658

Query  254     ATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATACATAACCTCTGCTG  313
               |||| |||||||||||| |||||||| ||||| || |||||||||| ||||||||||
Sbjct  852657  ATCCATCAATAAAACCTCCGGCTTAACTGCTATTGCCCTCGCTATACACAACCTCTGCTG  852598

Query  314     TTGTCCTCCAGAGAGAGATAGAGCGTTTTTATGCAGTTCATCTTTAACCTCATCCCACAA  373
               ||||| || || |  |||| |||||||||||| || ||||||||||||||||||||||||
```

```
Sbjct  852597  CTGTCCCCCTGATAACGATAAAGCGTTTTTATGAAGCTCATCTTTAACCTCATCCCACAA  852538

Query  374     AGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTTTATCCTTAATTCC  433
               ||||||||| ||||||||||||||| |||||||||||||| || |||||||| ||||||||
Sbjct  852537  AGCCGCTTTTTTTAAAGCCCACTCGACAATCTTATCCAACTCCTTTTTATCTTTAATTCC  852478

Query  434     ATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAAAAGGATTTGG-TT  492
                ||||||||||||||| |||||||| |||||||| || ||||| |||||||||||||| ||
Sbjct  852477  GTGAATTCTTGGGCCGAATGCAACGTTATCATAAATACTCATTGCAAAAGGATTTGGCTT  852418

Query  493     TTGAAA  498
               ||||||
Sbjct  852417  TTGAAA  852412


>CP001901.1 Methanocaldococcus sp. FS406-22, complete genome
Length=1760939

 Score = 582 bits (315),  Expect = 3e-162
 Identities = 430/487 (88%), Gaps = 1/487 (0%)
 Strand=Plus/Plus

Query  14       TTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTGGATTTAGGAATAT  73
                ||||||||| |||||||| || || ||||| |||||||||||||||| |||| ||||||
Sbjct  1577414  TTAACCAAATCTACCACTGATATAATCATCCGTCTCCTTCTTCTGTGGGTTTAAGAATAT
1577473

Query  74       CTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGGCAGTGTAATCAGA  133
                |||||||||||||||||||||||||| |||||||||||||||| || |||||||| ||
Sbjct  1577474  CTGCTCTGTCTCTCCAAACTCAATTAACTTCCCCATTAAGAAAAATGCTGTGTAATCTGA
1577533

Query  134      AACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTATAATCTTTAGCTAA  193
                ||| |||||||| |||||||||| |||||||||||||||| |||||||| ||||||||
Sbjct  1577534  AACTCTACTTGCTTGCTGCATGTTATGGGTAACAACAACAATTGTATAATCCTTAGCTAA
1577593

Query  194      CTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGGCAGATGTTGGTTC  253
                 |||||||| || |||||||||||||  || || || ||||||||||||| ||||| ||
Sbjct  1577594  TTCAACCATCAATTCCTCTATCTTTAGGGTAGAAATTGGGTCTAAGGCAGAGGTTGGCTC
1577653

Query  254      ATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATACATAACCTCTGCTG  313
                |||| | ||||||||||||||| ||||| ||||| ||||| || || ||||| ||
Sbjct  1577654  ATCCATTAATAAAACCTCTGGTTTAACGGCTATGGCTCTTGCTATGCACAATCTCTGTTG
1577713

Query  314      TTGTCCTCCAGAGAGAGATAGAGCGTTTTTATGCAGTTCATCTTTAACCTCATCCCACAA  373
```

```
               ||| || ||||| |  |||| ||||||||| |||| |||||||||||| |||||||| ||
Sbjct  1577714  TTGCCCCCCAGATAATGATAAAGCGTTTTTGTGCAATTCATCTTTAACTTCATCCCATAA
1577773

Query  374      AGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTTTATCCTTAATTCC  433
               ||||||||||||||| ||||| ||||| |||||||| ||||||||||||| || |||||
Sbjct  1577774  AGCCGCTTTCTTTAACGCCCATTCAACTATCTTATCTAATTCTTTTTTATCTTTGATTCC
1577833

Query  434      ATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAAAAGGATTTGG-TT  492
                ||||||||||| |||||||||||||||| || ||||||||||||||||||||||| ||
Sbjct  1577834  GTGAATTCTTGGACCAAATGCAACATTATCGTATATGCTCATAGCAAAAGGATTTGGCTT
1577893

Query  493      TTGAAAT  499
               |||||||
Sbjct  1577894  TTGAAAT  1577900


>CP009149.1 Methanocaldococcus bathoardescens strain JH146, complete genome
Length=1607556

 Score = 507 bits (274),  Expect = 2e-139
 Identities = 414/483 (86%), Gaps = 4/483 (1%)
 Strand=Plus/Minus

Query  14       TTAACCAAACCTACCACTAATGTAGTCATCTGTCTCCTTCTTCTGTGGATTTAGGAATAT  73
               ||||||||| ||||| ||||| |||||||||| || ||  |||||||||||| ||||||
Sbjct  947269   TTAACCAAATCTACCGCTAATATAGTCATCTGTTTCTTTTCTCTGTGGATTTAAGAATAT  947210

Query  74       CTGCTCTGTCTCTCCAAACTCAATTAATTTCCCCATTAAGAAAAAGGCAGTGTAATCAGA  133
               ||||||||||||||||||||||||| || |||||||||||||||||||||||||| ||
Sbjct  947209   CTGCTCTGTCTCTCCAAACTCAATTAACTTACCCATTAAGAAAAAGGCAGTGTAATCTGA  947150

Query  134      AACCCTACTTGCCTGCTGCATGTTGTGGGTAACAACAACAATCGTATAATCTTTAGCTAA  193
                ||||||||||| |||||||||||||||||||||||||||| |||||||| ||||||||
Sbjct  947149   GACCCTACTTGCTTGCTGCATGTTGTGGGTAACAACAACAATTGTATAATCCTTAGCTAA  947090

Query  194      CTCAACCATTAACTCCTCTATCTTTAATGTGGAGATAGGGTCTAAGGCAGATGTTGGTTC  253
                |||||||| ||||||||||||| ||||| || || ||||||||||||||||||||||||
Sbjct  947089   TTCAACCATCAACTCCTCTATCTTCAATGTTGATATTGGGTCTAAGGCAGATGTTGGTTC  947030

Query  254      ATCCGTCAATAAAACCTCTGGCTTAACCGCTATCGCTCTCGCTATACATAACCTCTGCTG  313
               |||| |||| |  || ||||| || |||||||| | ||| ||||| || ||||||||||
Sbjct  947029   ATCCATCAAAATTACTTCTGGTTTCACCGCTATAGTTCTTGCTATGCACAACCTCTGCTG  946970

Query  314      TTGTCCTCCAGAGAGAGAT-AGAGCGTTTTTATGCA-GTTCATCTTTAACCTCATCCCAC  371
                || || ||||| | || | | |||   |||||  | |||  |||||||| ||||||||
```

```
Sbjct  946969  CTGCCCACCAGATA-AGCTTAAAGCTGATTTATCTAAGTTG-TCTTTAACTTCATCCCAA  946912

Query  372     AAAGCCGCTTTCTTTAAAGCCCACTCAACAATCTTATCCAATTCTTTTTTATCCTTAATT  431
               | ||| ||||| |||| |||||| | ||||||||||||||||||||||||||| ||||||
Sbjct  946911  AGAGCAGCTTTTTTTAGAGCCCATTTAACAATCTTATCCAATTCTTTTTTATCTTTAATT  946852

Query  432     CCATGAATTCTTGGGCCAAATGCAACATTATCATAGATGCTCATAGCAAAAGGATTTGGT  491
               || |||||||||||||  | || |||||||||| || ||||| | ||| || ||||||
Sbjct  946851  CCGTGAATTCTTGGGCCGTAAGCTACATTATCATAAATACTCATTGGAAATGGGTTTGGT  946792

Query  492     TTT  494
               |||
Sbjct  946791  TTT  946789


  Database: Nucleotide collection (nt)
    Posted date:  Jun 5, 2018  11:03 PM
  Number of letters in database: 181,249,278,927
  Number of sequences in database:  48,103,425

Lambda      K        H
    1.33    0.621     1.12
Gapped
Lambda      K        H
    1.28    0.460    0.850
Matrix: blastn matrix:1 -2
Gap Penalties: Existence: 0, Extension: 0
Number of Sequences: 48103425
Number of Hits to DB: 0
Number of extensions: 0
Number of successful extensions: 0
Number of sequences better than 10: 0
Number of HSP's better than 10 without gapping: 0
Number of HSP's gapped: 0
Number of HSP's successfully gapped: 0
Length of query: 499
Length of database: 181249278927
Length adjustment: 34
Effective length of query: 465
Effective length of database: 179613762477
Effective search space: 83520399551805
Effective search space used: 83520399551805
A: 0
X1: 13 (25.0 bits)
X2: 32 (59.1 bits)
X3: 54 (99.7 bits)
S1: 13 (25.1 bits)
S2: 23 (43.6 bits)
```

Mol Gen Genet (1996) 250:692–698
© Springer-Verlag 1996

ORIGINAL PAPER

Toshiyuki Nikata · Yukihiro Sakai · Kenya Shibata
Junichi Kato · Akio Kuroda · Hisao Ohtake

Molecular analysis of the phosphate-specific transport (*pst*) operon of *Pseudomonas aeruginosa*

Received: 24 July 1995 / Accepted: 7 November 1995

Abstract The organization of the phosphate-specific transport (*pst*) operon in *Pseudomonas aeruginosa* has been determined. The gene order of the *pst* operon is *pstC*, *pstA*, *pstB*, *phoU*, and a well-conserved Pho box sequence (16/18 bases identical) exists in the promoter region. The most striking difference from the known *Escherichia coli pst* operon is the lack of the *pstS* gene encoding a periplasmic phosphate (P_i)-binding protein. Even though the three *pst* genes were absolutely required for P_i-specific transport, expression of the *pst* operon at high levels did not increase P_i uptake in *P. aeruginosa*. DNA sequences for the *pstB* and *phoU* genes have been determined previously. The newly identified *pstC* and *pstA* genes encode possible integral membrane proteins of 677 amino acids (M_r 73 844) and 513 amino acids (M_r 56 394), respectively. The amino acid sequences of PstC and PstA predict that these proteins contain a long hydrophilic domain not seen in their *E. coli* counterparts. A chromosomal deletion of the entire *pst* operon rendered *P. aeruginosa* unable to repress P_i taxis under conditions of P_i excess. The *phoU* and *pstB* genes are essential for repressing P_i taxis. However, mutants lacking either PstC or PstA alone were able to repress P_i taxis under conditions of P_i excess.

Key words Chemotaxis · *pho* regulon · Phosphate · Phosphate transport · *Pseudomonas aeruginosa*

Communicated by J. Lengeler

T. Nikata · Y. Sakai · K. Shibata · J. Kato
A. Kuroda · H. Ohtake (✉)
Department of Fermentation Technology, Hiroshima University, Higashi-Hiroshima, Hiroshima 739, Japan

Introduction

To cope with phosphate (P_i) limitation, bacteria have evolved complex regulatory systems to assimilate P_i very efficiently (Wanner 1987). In *Pseudomonas aeruginosa*, P_i limitation results in the synthesis of several proteins, including alkaline phosphatase (AP), an outer membrane protein P, as well as a hemolytic and a nonhemolytic phospholipase C (Ostroff and Vasil 1987; Hancock et al. 1990). Several lines of evidence indicate that *P. aeruginosa* possesses a P_i-specific transport (Pst) system which is derepressed by P_i limitation (Poole and Hancock 1984). P_i transport in *P. aeruginosa* occurs with biphasic kinetics, suggesting the involvement both of a high-affinity (genes *pst*) and low-affinity (genes *pit*) component. A periplasmic P_i-binding protein has been purified to apparent homogeneity from the shock fluids of P_i-starved cells. In addition, *P. aeruginosa* mutants lacking the P_i-binding protein were shown to be deficient in high-affinity P_i transport (Poole and Hancock 1984); however, the *P. aeruginosa pst* gene have not been identified.

Under conditions of P_i limitation, *P. aeruginosa* shows chemotactic responses towards P_i (Kato et al. 1992). It has been suggested that P_i taxis may have an important role in scavenging P_i under conditions of P_i limitation. To date, there have not been any reports of P_i taxis in a bacterial species other than *P. aeruginosa*, and the regulation of bacterial P_i taxis has not yet been characterized. Recently, we showed that regulation of P_i taxis is not dependent on the *phoB* and *phoR* gene products but requires the *phoU* gene (Kato et al. 1994). In the course of a cloning project to uncover genes involved in the regulation of P_i taxis, we discovered three *P. aeruginosa pst* genes located immediately upstream of the *phoU* gene. In the present report, we describe the organization of the *P. aeruginosa pst* operon and provide evidence that the *pst* gene products are required for repressing P_i taxis under conditions of P_i excess.

Table 1 Bacterial strains and plasmids

Strain/ plasmid	Relevant characteristics	Reference/source
Escherichia coli		
MV1184	*ara-14* Δ(*lac-proAB*) *rpsL thi-1* (Φ80 *lacZ*ΔM15)Δ(*srl-recA*) 306::Tn*10* F′[*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15]	Vieira and Messing (1987)
Pseudomonas aeruginosa		
PAO1	Prototroph	Holloway et al. (1977)
PHOB1	PAO1 derivative, *phoB*::Km^r	Kato et al. (1994)
Plasmids		
pUC119	Cloning vector; Ap^r, *lacPOZ′*	Vieira and Messing (1987)
pUC4K	pUC4 containing 1.3 kb Km^r cartridge; Ap^r, Km^r	Pharmacia, Uppsala, Sweden
pCP19	Broad-host range cosmid; Tc^r, IncP	Goldberg and Ohman (1987)
pPT06.1	pCP19 harboring *P. aeruginosa pst* operon; Tc^r	Kato et al. (1994)
pMMB66EH	Broad-host range vector; Cb^r, P*tac*, *lacI*^{q}, InQ	Furste et al. (1986)
pQF50	Broad-host range transcriptional fusion vector, Cb^r, *lacZ*, IncP	Farinha and Kropinski (1990)
pEP02.2	pSTV28 harboring *E. coli pst* operon fused with *tac* promoter; Cm^r, P*tac-pst* operon	Kato et al. (1993)
pMC1871	pBR322 derivative; source of truncated *lacZ* gene for translational fusion; Tc^r	Pharmacia, Uppsala, Sweden

Materials and methods

Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in Table 1. *P. aeruginosa* and *Escherichia coli* strains were grown at 37° C with shaking in 2 × YT medium (Sambrook et al. 1989) supplemented with the appropriate antibiotics. *P. aeruginosa* cells were also grown at 37° C with shaking in T_5 minimal medium (Kato et al. 1992) containing 5 mM potassium phosphate. For P_i limitation conditions, cells grown overnight in 2 × YT medium were inoculated into T_0 medium (final inoculum dilution 1.25%), which was prepared by omitting potassium phosphate from T_5 medium, and incubated at 37° C with shaking.

Plasmid construction

The construction of plasmid pPT06.1 (Fig. 1A), which contains the entire *pst* operon, was described previously (Kato et al. 1994). *Bam*HI-*Eco*RI (2.9 kb), *Eco*RI-*Bgl*II (3.3 kb), and *Bam*HI-*Bgl*II (6.2 kb) fragments were excised from pPT06.1 and subcloned into pCP19 (Goldberg and Ohman 1987) to construct pPT06.2, pPT06.5, and pPT06.10, respectively. To construct respectively pPT06.15 and pPT06.16, plasmid pPT06.1 was digested with either *Not*I or *Eco*RI, treated with the Klenow fragment of DNA polymerase and religated to itself. Plasmid pPT06.1 was digested with *Mro*I and self-ligated to construct pPT06.18. A 6.2 kb *Bam*HI-*Bgl*II fragment from plasmid pPT06.1 was inserted downstream of the *tac* promoter of pMMB66EH (Furste et al. 1986) to construct pPT06.20. To construct plasmid pPT06.32, pPT06.1 was digested with *Not*I and religated to itself.

pPT06.21 and pPT06.22 are *phoU-lacZ* transcriptional fusions (Fig. 4A). They were constructed by cloning the 5.6 kb *Bam*HI-*Xho*I and the 4.8 kb *Bal*I-*Xho*I fragment of pPT06.1 in front of the promoterless *lacZ* gene of vector pQF50 (Farinha and Kropinski 1990), respectively. Plasmid pPT06.23 was constructed by cloning the 6.2 kb *Bam*HI-*Bgl*II fragment carrying the entire *pst* operon into *Bam*HI-digested pQF50. Plasmid pQF1871 was a derivative of pQF50 which contains the *′lacZ* gene from pMC1871 (Pharmacia). To prepare *pstC-lacZ* and *pstA-lacZ* translational fusions, a 1.2 kb *Bam*HI-*Pvu*II fragment and a 2.9 kb *Bam*HI-*Eco*RI fragment was excised from pPT06.1 and inserted immediately upstream of the *′lacZ* gene in pQF1871; the resulting plasmids were designated pPT06.24 and pPT06.25, respectively (Fig. 4B).

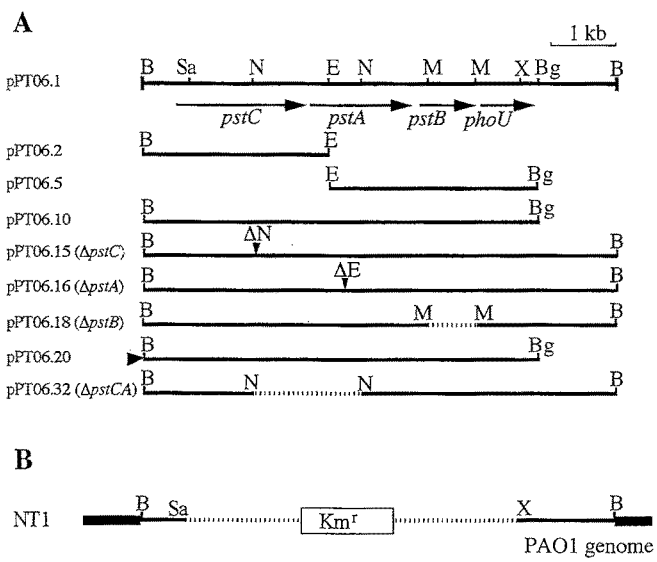


Fig. 1 Restriction maps of **A** pPT06.1, containing the 7.4 kb *Bam*HI fragment from the *Pseudomonas aeruginosa* PAO1 chromosomal DNA, and subclones of pPT06.1 and **B** the *pst* operon region of *P. aeruginosa* strain NT1 chromosomal DNA. The locations and orientations of the *pstC*, *pstA*, *pstB*, and *phoU* genes are shown by *horizontal arrows* below the map of pPT06.1. *Dashed lines* indicate DNA fragments deleted from the 7.4 kb *Bam*HI insert of pPT06.1. *Arrowhead* shows the site and direction of the *tac* promoter, and *inverted triangles* the *Not*I and *Eco*RI sites destroyed to make frameshift mutations. Restriction sites abbreviations: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; M, *Mro*I; N, *Not*I; Sa, *Sau*3AI; X, *Xho*I

Disruption of the chromosomal *pst* operon

The 7.4 kb *Bam*HI fragment was excised from pPT06.1 and subcloned into pUC119 to construct pPT06.14. Plasmid pPT06.14 partially digested with *Sau*3AI and *Xho*I was ligated with a *Bam*HI and *Sal*I-flanked kanamycin resistance (Km^r) gene cassette from pUC4K (Pharmacia). The resulting plasmid was introduced into *P. aeruginosa* PAO1 by electroporation, and Km^r transformants were selected on 2 × YT medium containing kanamycin. The disruption of the chromosomal *pst* operon was confirmed by Southern hybridization (Sambrook et al. 1989).

Immunoblotting

Cells grown overnight in either T_0 or T_5 medium were harvested by centrifugation, resuspended in lysis buffer (Lanka and Barth 1981) and boiled for 5 min. Lysates were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on an 8% acrylamide gel, transferred to a polyvinylidene difluoride membrane, and probed with mouse anti-β-galactosidase monoclonal antibody (Boehringer Mannheim). The secondary and the ternary antibodies used were sheep anti-mouse Ig $F(ab')_2$ fragment conjugated with digoxigenin (Boehringer Mannheim) and sheep anti-digoxigenin Fab fragment conjugated with alkaline phosphatase (Boehringer Mannheim), respectively. Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2′-(5′-chloro) tricyclo [3,3,1,1,3,7] decan}-4-yl) phenyl phosphate (Tropix, Bedford, Mass., USA) was used as a chemiluminescent substrate for alkaline phosphatase, and protein bands were visualized by exposure to RX film (Fuji, Tokyo, Japan).

Chemotaxis and enzyme assays

Chemotaxis assays were carried out by using the computer-assisted capillary method as described previously (Nikata et al. 1992). Digital image processing was used to count the number of bacteria accumulating around the mouth of a capillary tube containing a known concentration of P_i plus 1% agarose. The response of bacteria was dependent on the concentration of P_i in the capillary tube (Kato et al. 1992). Since the concentration-response curve showed a peak at 5–10 mM P_i, the chemotaxis assays were subsequently performed at a concentration of 10 mM P_i. The chemotaxis buffer used in the assay was 10 mM HEPES buffer (pH 7.0). Alkaline phosphatase (AP) activity was measured as previously described (Kato et al. 1992). One Unit of enzyme activity was defined as the amount of enzyme which liberated 1 nmol of *p*-nitrophenol per min. β-Galactosidase activities of cells were determined as described by Miller (1972) with the modification that the enzymatic reaction was carried out at 37° C.

Measurement of P_i transport

The procedure for measuring P_i transport was essentially as described previously (Poole and Hancock 1984). The assay was started by the addition of radioactively labeled phosphate ($^{32}P_i$) to a final concentration of 2.5 μM, and the amount of $^{32}P_i$ incorporated into the bacterial cells was measured with a liquid scintillation spectrometer (Packard).

Results and discussion

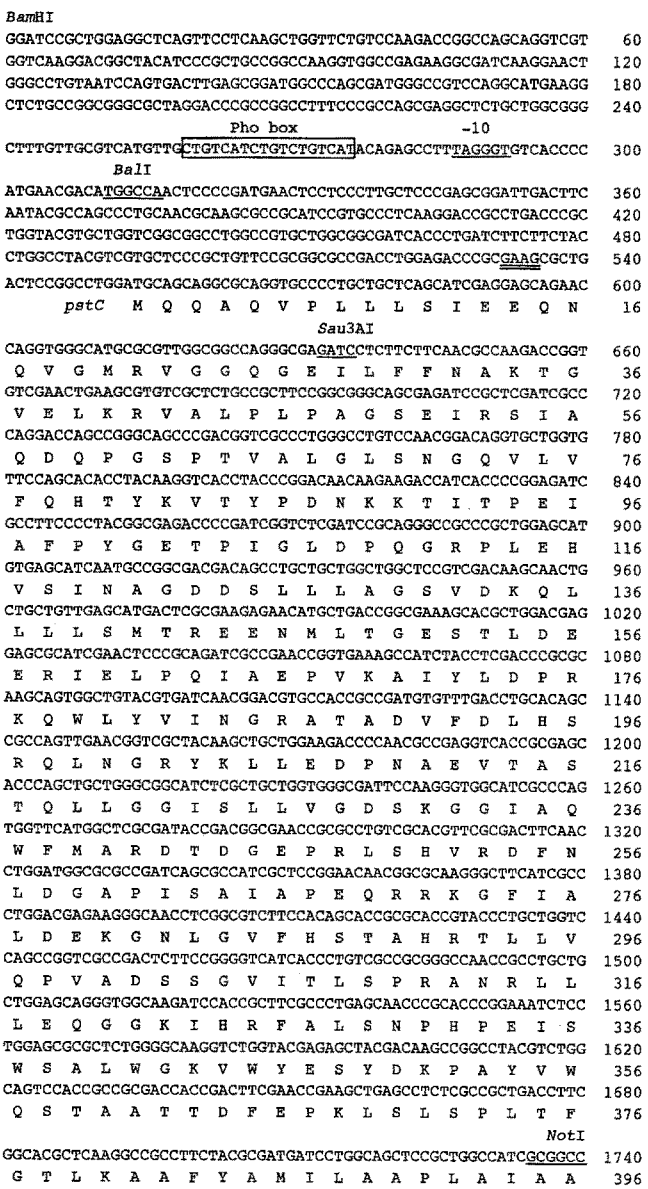
DNA sequence analysis

The nucleotide sequence of the 2.9 kb *Bam*HI-*Eco*RI region of pPT06.2 was determined (Fig. 2). Since the 3.2 kb *Eco*RI-*Bgl*II region of pPT06.1 was determined previously (Kato et al. 1994), the complete nucleotide sequence of the 6.2 kb *Bam*HI-*Bgl*II fragment of pPT06.1 has now been determined. The *pstC* gene contained a putative ATG start codon at nucleotide positions 553–555 and a stop codon at positions 2584–2586. The predicted gene product of *pstC* was composed of 677 amino acid residues (M_r 73 844). The *pstA* gene of 1539 bp started at positions 2666–2668, and a 79-bp intercistronic DNA region was located between the end of *pstC* and the proposed initiator codon for *pstA*. The *pstA* gene could code for a protein of 513 amino acid residues having a calculated M_r of 56 349.

PstC and PstA were predicted to be membrane polypeptides with potential membrane-spanning segments. These proposed membrane domains were located in the C-terminal half of the proteins (Fig. 3). PstC and PstA were also predicted to be approximately twofold larger than known *E. coli* PstC (319 amino acid residues) and PstA (296 amino acid residues), respectively. PstC and PstA appeared to have a long hydrophilic domain, which was not seen with the *E. coli* counterparts, in the N-terminal half of the proteins. Only sparse similarity with the *E. coli* counterparts was detected in the C-terminal half of the PstC (20% identical residues) and PstA (23% identical residues) proteins (Fig. 3).

Transcriptional and translational fusion studies

A putative Pho box was found at positions 259–276 (Fig. 2). There was a 16/18 bp match with the consensus Pho box sequence (Siehnel et al. 1988; Anba et al. 1990). To study the expression of the *pst* operon, we first constructed transcriptional fusions between fragments of the *pst* operon and the promoterless *lacZ* gene using vector pQF50 (Fig. 4A). *P. aeruginosa* PAO1 and PHOB1 ($PhoB^-$) were then transformed with the resulting plasmids, and the β-galactosidase levels were measured in the transformants grown under conditions of either P_i excess or P_i limitation. The enzyme levels in PAO1 (pPT06.21) increased approximately 25-fold under conditions of P_i limitation, whereas they were not elevated in PHOB1(pPT06.21) (data not shown). PAO1 carrying pPT06.22, which lacks the Pho box sequence, only showed the background enzyme activity. When the promoterless *lacZ* gene was placed at the *Bgl*II site downstream of *phoU* in pPT06.23, the β-galactosidase levels were not elevated under



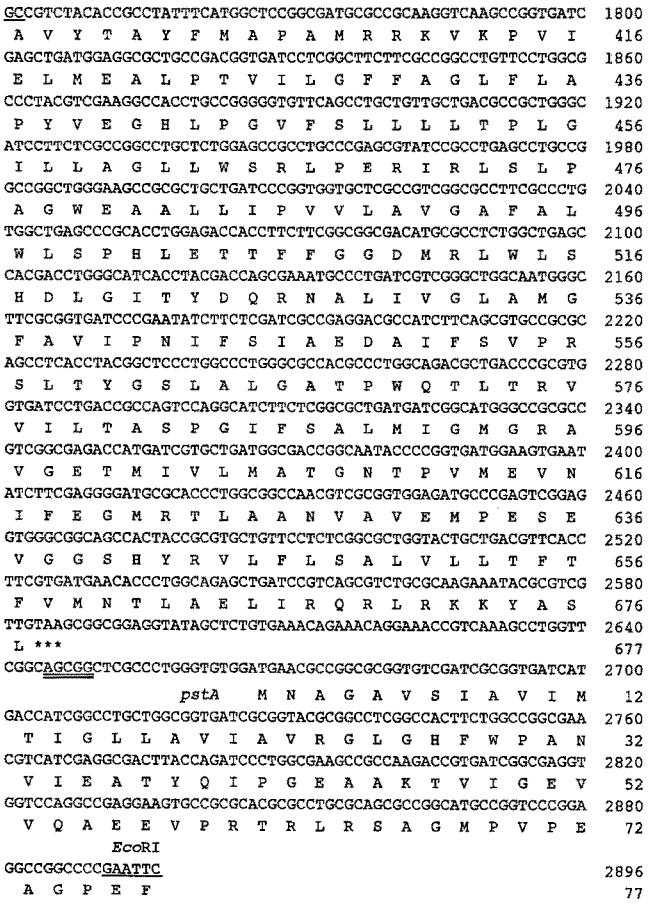


Fig. 2 DNA nucleotide and amino acid sequences of the 2.9 kb *Bam*HI-*Eco*RI region of pPT06.2. Nucleotide (upper) and amino acid (below) numbering starts at the *Bam*HI site and at each putative ATG start codon respectively. Amino acids deduced from the nucleotide sequence are specified by standard one-letter abbreviations. Putative ribosome-binding sites are *double-underlined*. Relevant restriction sites are located above the sequence. *Asterisks* indicate the *pstC* stop codon. The putative *pho* box sequence is *boxed*. The sequence of the 2.9 kb *Bam*HI-*Eco*RI fragment will appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence databases under accession number D45195

P_i limitation. These results are consistent with the conclusion that the *pstCAB* and *phoU* genes constitute an operon.

The *pstC-lacZ* and *pstA-lacZ* translational fusions were then constructed to confirm the expression of *pstC* and *pstA* by Western blot analysis (Fig. 4B). In plasmid pPT06.24, the 5′-terminal 654 bp of the *pstC* gene was fused to the *'lacZ* gene from pMC1871 (Pharmacia). Plasmid pPT06.25 contained the 5′-terminal 231 bp of *pstA* fused to *'lacZ*. The calculated masses of the *pstC-lacZ* and *pstA-lacZ* fusion products are approximately 140 and 124 kDa, respectively. *P. aeruginosa* PAO1 was transformed with either pPT06.24 or pPT06.25 and then grown under conditions of P_i excess or P_i limitation. Western blot analysis showed that both *pstC-lacZ* and *pstA-lacZ* fusion proteins were expressed under conditions of P_i limitation (Fig. 4C). Upon SDS-PAGE, *pstC-lacZ* and *pstA-lacZ* fusion proteins showed apparent molecular masses of approx. 145 and 121 kDa, respectively. These molecular masses observed with SDS-PAGE are consistent with the calculated masses of the *pstC-lacZ* and *pstA-lacZ* fusion proteins.

P_i uptake assays

The entire *pstC-phoU* region was deleted from the *P. aeruginosa* PAO1 chromosome by homologous recombination. The resulting mutant, designated NT1 (Fig. 1B), and the parental strain PAO1 were examined for their ability to take up P_i. NT1 failed to express the highly specific P_i transport system even under P_i limitation (Fig. 5A). To investigate the effects of each *pst* gene

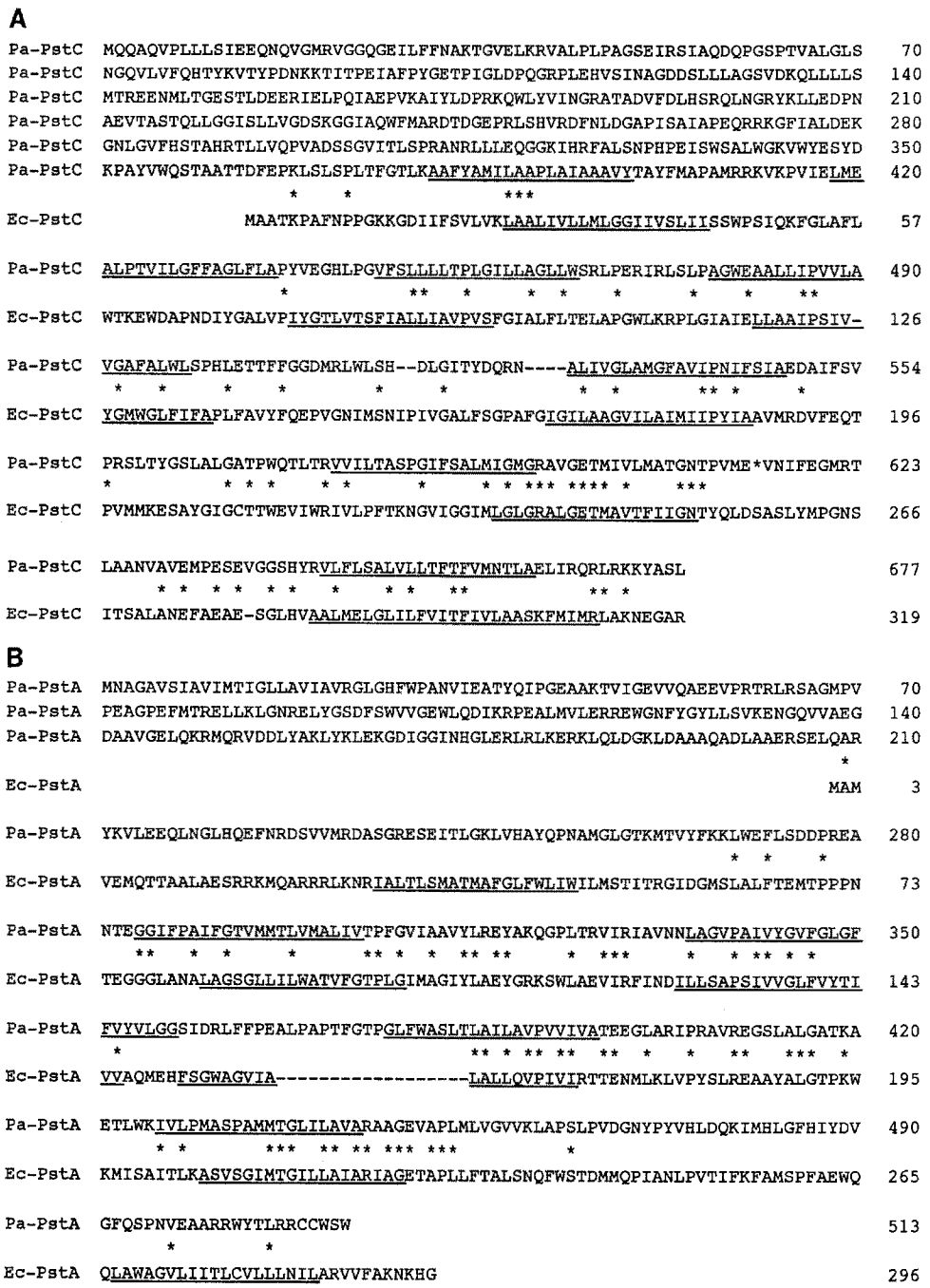


Fig. 3 Amino acid sequence similarities between **A** *P. aeruginosa* PstC (Pa–PstC) and *Escherichia coli* PstC (Ec–PstC) and **B** *P. aeruginosa* PstA (Pa–PstA) and *E. coli* PstA (Ec–PstA). The alignments were performed with FASTA and Best-Fit programs (Pearson and Lipman 1988). Identical amino acids are indicated by *asterisks*, gaps (*dashes*) are introduced into the sequence to maximize the alignments. Putative transmembrane helices are *underlined* (Cox et al. 1988)

on P_i uptake, recombinant plasmids carrying deletion and frameshift mutations in the *pst* genes were constructed (Fig. 1A). The resulting plasmids were then introduced into NT1 by electroporation. None of these plasmids could restore P_i transport in NT1 (Fig. 5A). By contrast, NT1(pPT06.10), as well as NT1(pPT06.1), exhibited P_i uptake at a rate similar to that seen with PAO1. This confirmed that the mutant phenotype of NT1 was not due to a polar effect.

The 6.2 kb *Bam*HI-*Bgl*II fragment of pPT06.1, which contains the entire *pst* operon, was placed under the control of the *tac* promoter in an expression vector pMMB66EH (Furste et al. 1986) to make pPT06.20. Introduction of this plasmid into *P. aeruginosa* PAO1 did not increase P_i uptake under conditions of P_i excess (Fig. 5B). Moreover, the addition of 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) did not improve the ability of PAO1 (pPT06.20) to take up P_i. *E. coli* MV1184 bearing pEP02.2, which carries the entire *E. coli pst* operon downstream of *tac* promoter (Kato et al. 1993), showed a linear increase in P_i uptake under conditions of P_i excess (Fig. 5B). The difference in the P_i uptake phenotype may be interpreted as due to the lack of the *pstS* gene in the *P. aeruginosa pst* operon. The *E. coli pst*

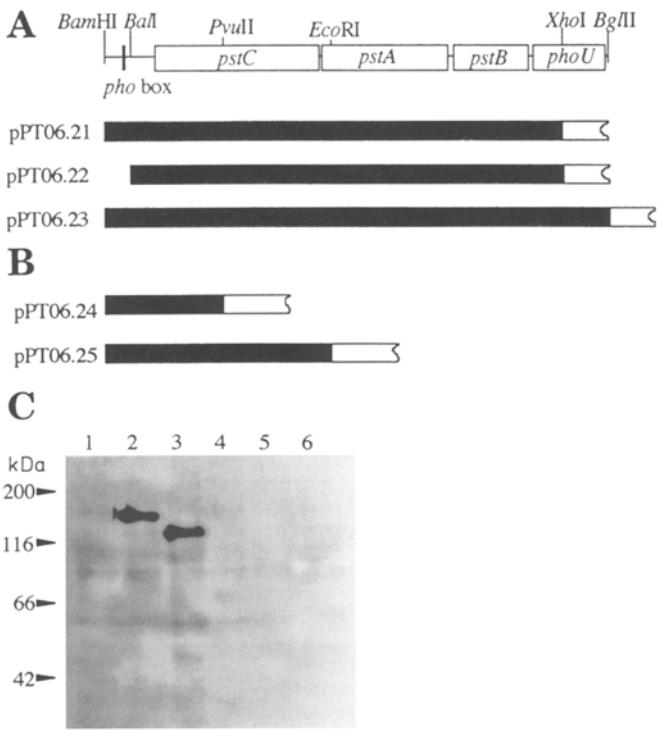


Fig. 4A–C Transcriptional and translational fusions. **A** Transcriptional *lacZ* fusions. *Filled boxes* show the *pst* operon genes, while *open boxes* indicate the fused *lacZ* gene. **B** Translational *pstC-lacZ* (pPT06.24) and *pstA-lacZ* (pPT06.25) fusions. **C** Western blot analysis. Strain PAO1 and recombinant derivatives were grown under either phosphate (P_i) limitation (lanes 1–3) or P_i excess (lanes 4–6). Lanes 1 and 4, PAO1; lanes 2 and 5, PAO1 (pPT06.24); lanes 3 and 6, PAO1 (pPT06.25)

operon consists of *pstSCAB* and *phoU*, and the *pstS* codes for a periplasmic P_i-binding protein which is required for P_i-specific transport (Amemura et al. 1985).

AP and chemotaxis assays

The chromosomal deletion mutant NT1 was constitutive for both AP synthesis and P_i taxis (Table 2). To study the effects of each *pst* gene on AP synthesis and P_i taxis, NT1 was transformed with deletion derivative plasmids of pPT06.1 (Fig. 1A). Plasmids pPT06.18 (Δ*pstB*) and pPT06.32(Δ*pstCA*) failed to repress AP synthesis and P_i taxis in NT1. However strain NT1, carrying either pPT06.15 (Δ*pstC*) or pPT06.16(Δ*pstA*) repressed both AP synthesis and P_i taxis under conditions of P_i excess. The repression of AP synthesis and P_i taxis appears to be unrelated to P_i transport, because both NT1(pPT06.15) and NT1(pPT06.16) failed to express P_i-specific transport under conditions of P_i limitation (Fig. 5A). Previous studies (Cox et al. 1988, 1989) had shown for *E. coli* that the provision of P_i to the *pst* operon is not necessary for the repression of AP synthesis. The amino acid substitution in the *E. coli* *pstC* gene, of Arg-237 to Gln and Glu-240 to Gln, caused a loss of P_i transport through the phosphate-specific transport system, while AP synthesis remained repressed.

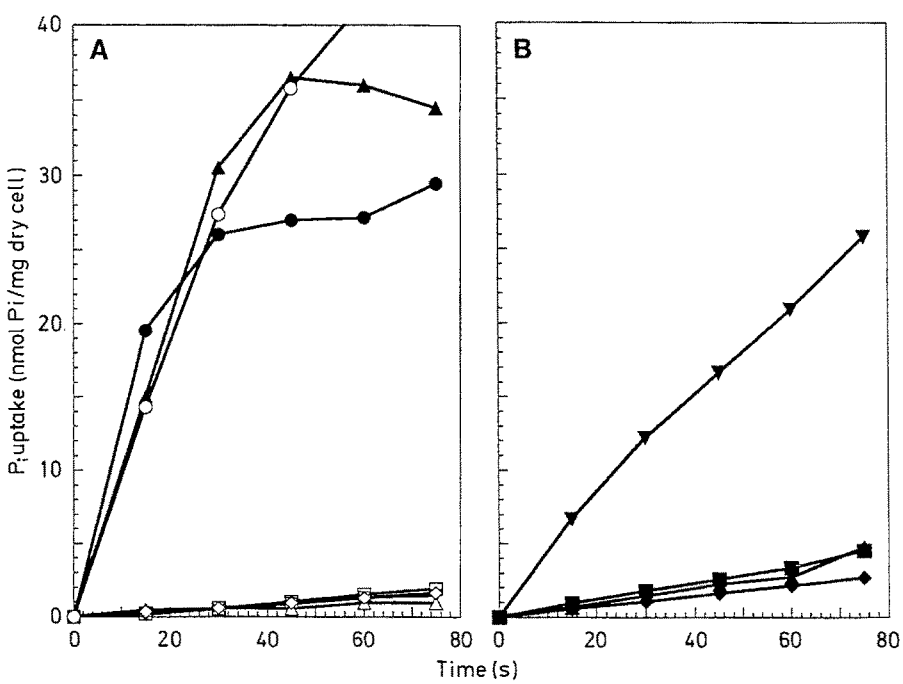


Fig. 5A, B Time course of P_i uptake by *P. aeruginosa* and *E. coli* strains. **A** P_i uptake by *P. aeruginosa* PAO1 (▲), NT1 (□), NT1(pPT06.1) (●), NT1(pPT06.10) (○), NT1(pPT06.15) (△), NT1(pPT06.16) (▽), and NT1(pPT06.18) (◇) under conditions of P_i limitation. **B** P_i uptake by *P. aeruginosa* PAO1 (▲), PAO1(pPT06.20) (◆), *E. coi* MV1184 (■) and MV1184(pEP02.2) (▼) under conditions of P_i excess. To induce expression of the *pst* genes, 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) was added to 2 × YT medium

Table 2 Alkaline phosphatase (AP) activities and phosphate (P_i) taxis in *P. aeruginosa* strains

Strain	AP activity (Units/OD_{600})[a]		P_i taxis[b]	
	High P_i[c]	P_i starvation[d]	High P_i	P_i starvation
PAO1 (Wild type)	2.2 ± 1.4	66.0 ± 2.5	−	+
NT1	32.7 ± 5.8	52.9 ± 2.2	+	+
NT1 (pPT06.15)	4.4 ± 0.6	40.5 ± 0.7	−	+
NT1 (pPT06.16)	4.2 ± 1.0	57.0 ± 1.4	−	+
NT1 (pPT06.18)	105.0 ± 7.5	100.0 ± 2.0	+	+
NT1 (pPT06.32)	44.7 ± 5.2	54.2 ± 2.5	+	+

[a] Mean ± SD of four experiments. One Unit of enzyme activity is defined as the amount of enzyme which liberated 1 nmol of *p*-nitrophenol per min. OD_{600}, Absorbance at 600 nm
[b] Positive chemotactic response to P_i; −, no response to P_i. The computer-assisted capillary assay technique was used as described (Nikata et al. 1992)
[c] The bacteria were grown in T_5 medium with 5 mM P_i to an OD_{600} of 1.0. For chemotaxis assays, cells were washed with 10 mM HEPES buffer (pH 7.0)
[d] The bacteria were grown in P_i-free T_0 medium to an OD_{600} of 0.3. For chemotaxis assays, cells were washed with 10 mM HEPES buffer (pH 7.0)

References

Amemura M, Makino K, Shinagawa H, Kobayashi A, Nakata A (1985) Nucleotide sequence of the genes involved in phosphate transport and regulation of the phosphate regulon in *Escherichia coli*. J Mol Biol 184:241–250

Anba J, Bidaud M, Vasil ML, Lazdunski A (1990) Nucleotide sequence of the *Pseudomonas aeruginosa phoB* gene, the regulatory gene for the phosphate regulon. J Bacteriol 172:4685–4689

Cox GB, Webb D, Godovac-Zimmerman J, Rosenberg H (1988) Arg-220 of the PstA protein required for phosphate transport through the phosphate-specific transport system in *Escherichia coli* but not for alkaline phosphatase repression. J Bacteriol 170:2283–2286

Cox GB, Webb D, Rosenberg H (1989) Specific amino acid residues in both the PstB and PstC proteins are required for phosphate transport by the *Escherichia coli* Pst system. J Bacteriol 171:1531–1534

Farinha MA, Kropinski AM (1990) Construction of broad-host-range plasmid vectors for easy visible selection and analysis of promoters. J Bacteriol 172:3496–3499

Furste JP, Pansegrau W, Frank R, Blocker H, Scholz P, Bagdasarian M, Lanka E (1986) Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. Gene 48:119–131

Goldberg JB, Ohman DE (1987) Construction and characterization of *Pseudomonas aeruginosa algB* mutants: role of *algB* in high-level production of alginate. J Bacteriol 169:1593–1602

Hancock RE, Siehnel R, Martin N (1990) Outer membrane proteins of *Pseudomonas*. Mol Microbiol 4:1069–1075

Holloway BW, Krishapillai V, Morgan AF (1977) Chromosomal genetics of *Pseudomonas*. Microbial Rev 43:73–102

Kato J, Ito A, Nikata T, Ohtake H (1992) Phosphate taxis in *Pseudomonas aeruginosa*. J Bacteriol 174:5149–5151

Kato J, Yamada K, Muramatsu A, Hardoyo, Ohtake H (1993) Genetic improvement of *Escherichia coli* for enhanced biological removal of phosphate from wastewater. Appl Environ Microbiol 59:3744–3749

Kato J, Sakai Y, Nikata T, Ohtake H (1994) Cloning and characterization of a *Pseudomonas aeruginosa* in the negative regulation of phosphate taxis. J Bacteriol 176:5874–5877

Lanka E, Barth PT (1981) Plasmid RP4 specifies a deoxyribonucleic acid primase involved in its conjugal transfer and maintenance. J Bacteriol 148:769–781

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Nikata T, Sumida K, Kato J, Ohtake H (1992) Rapid method for analysing bacterial behavioral responses to chemical stimuli. Appl Environ Microbiol 58:2250–2254

Ostroff RM, Vasil ML (1987) Identification of a new phospholipase C activity by analysis of an insertional mutation in the hemolytic phospholipase C structural gene of *Pseudomonas aeruginosa*. J Bacteriol 169:4597–4601

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448

Poole K, Hancock REW (1984) Phosphate transport in *Pseudomonas aeruginosa*: involvement of a periplasmic phosphate-binding protein. Eur J Biochem 144:607–612

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Siehnel RJ, Worobec EA, Hancock REW (1988) Regulation of components of the *Pseudomonas aeruginosa* phosphate-starvation-inducible regulon in *Escherichia coli*. Mol Microbiol 2:347–352

Vieira J, Messing J (1987) Production of single-stranded plasmid DNA. Methods Enzymol 153:3–11

Wanner BL (1987) Bacterial alkaline phosphatase gene regulation and the phosphate response in *Escherichia coli*. In: Torriani-Gorini A, Rothman FG, Silver S, Wright A, Yagil E (eds) Phosphate metabolism and cellular regulation in microorganisms. American Society for Microbiology Washington DC, pp 12–19

Genes Genet. Syst. (1997) **72**, p. 323–334

REVIEW

The phosphatase system in *Saccharomyces cerevisiae*

Yasuji Oshima

Department of Biotechnology, Faculty of Engineering, Kansai University, 3-3-35 Yamatecho, Suita-shi, Osaka 564, Japan

The yeast *Saccharomyces cerevisiae* has at least six species of acid and alkaline phosphatases with different cellular localizations, as well as inorganic phosphate (P_i) transporters. Most of the genes encoding these enzymes are coordinately repressed and derepressed depending on the P_i concentration in the growth medium. The P_i signals are conveyed to these genes through a regulatory circuit consisting of a set of positive and negative regulatory proteins. This phosphatase system is interested as one of the best systems for studying gene regulation in *S. cerevisiae* due to the simplicity of phenotype determination in genetic analysis. With this methodological advantage, considerable amounts of genetic and molecular evidence in phosphatase regulation have been accumulated in the past twenty-five years. This article summarizes the current progress of research into this subject.

INTRODUCTION

Saccharomyces cerevisiae has several phosphatases, with different specificities and cellular locations, and permeases, used in inorganic phosphate (P_i) uptake. The set of genes responsible for these activities is coordinately regulated repressively by P_i concentration in the growth medium. This phosphatase (*PHO*) system, consisting of several genes and regulators (Table 1), has been studied extensively due to numerous technical advantages (Oshima, 1991). Acid phosphatase activity is easily detected for each colony by specific staining, because the enzyme is located on the cell surface. Enzyme synthesis is regulated in response to a simple effector, P_i, in the medium. The same regulatory system (with minor differences) also controls the synthesis of alkaline phosphatase, located in the cell envelope, whose activity is also detectable in colonies by staining after permeabilizing the cells to the substrate. There is a 500-fold difference between the repressed and derepressed levels of acid phosphatase activity. With the above technical easiness in the mutant selection, we were able to isolate various mutants that were deficient in or constitutive for synthesis of repressible acid and alkaline phosphatases. Genetic and the subsequent molecular analysis of the *PHO* system with these mutants over the twenty-five years has provided insight into the mechanism of signal transduction for gene regulation. In this article, I attempt to summarize the current understanding of the molecular mechanism in phosphatase regulation, especially in signal transduction between negative and positive regulators and *cis*-acting regulatory sites in the *PHO* promoters. For more details on the initial genetic study of the *PHO* system, see review by Oshima (1982), and that by Johnston and Carlson (1992) on the subsequent molecular analysis.

THE GENES AND PROTEINS CONCERNING PHOSPHOROUS METABOLISM

The genes that are regulated by P_i are *PHO5*, encoding the major fraction of repressible acid phosphatase, designated p60 (rAPase; EC 3.1.3.2; Arima et al., 1983; Bajwa et al., 1984; Bostian et al., 1980; Lemire et al., 1985; Toh-e et al., 1975b), *PHO10* (or *PHO12*; de Steensma et al., 1989) and *PHO11* encoding p58 and p56 isozymes of the p60 enzyme respectively (Bostian et al., 1980; Lemire et al., 1985), *PHO8* encoding a non-specific repressible alkaline phosphatase (rALPase; EC 3.1.3.1; Kaneko et al., 1982, 1987; Toh-e et al., 1976), *PHO84* encoding a P_i-transporter (Bunya et al., 1991), and *PHO81* encoding one of the regulatory proteins in the *PHO* system (Ogawa et al., 1993; Yoshida et al., 1989b). These three rAPase isozymes, p60, p58, and p56, (optimum pH 3–4) are glycoproteins (Boer and Steyn-Parve, 1966; Onishi et al., 1979) and located in the cell wall or the periplasmic space (Arnold, 1972; Linnemans et al., 1977; Lopandic et al., 1987) and are probably responsible for acquiring phosphate from diverse extracellular sources. rALPase (optimum pH 8) encoded by *PHO8*, which is also a glycoprotein, is located in the vacuole (Clark et al., 1982) and cleaves diverse substrates to retrieve phosphate from intracellular products. It was suggested that the most important substrate of Pho8p is fructose-2, 6-bisphosphate (Plankert et al., 1991).

S. cerevisiae also has other phosphatases; one is an acid phosphatase (T-rAPase), encoded by *PHO3*, which may func-

Table 1. Genes and their encoding proteins or functions in the *PHO* system

Gene	Chromosome location	Function of gene-product	References
Structural genes			
PHO3	2R	Thiamine phosphatase (T-rAPase)	Nosaka (1990)
PHO5	2R	rAPase, p60	Toh-e et al. (1975b) Arima et al. (1983)
PHO10	8R	rAPase, p58	Bostian et al. (1980) Lemire et al. (1985) Venter and Hörz (1989)
or *PHO12*	8R	rAPase	de Steensma et al. (1989)
PHO11	1R	rAPase, p56	Bostian et al. (1980) Lemire et al. (1985) Venter and Hörz (1989) de Steensma et al. (1989)
PHO8	4R	rALPase Fructose-2, 6-bisphosphate 6-phosphatase	Kaneko et al. (1987) Plankert et al. (1991)
PHO9 (*PEP4*)	16L	Protease for Pho8p maturation	Kaneko et al. (1982)
PHO13	4L	Specific *p*-nitrophosphatase	Kaneko et al. (1989)
PHO84	13L	P_i-transporter	Bun-ya et al. (1991)
PHO86	4R	P_i-transporter associate	Yompakdee et al. (1996a)
PHO87	3R	P_i-transporter associate	Bun-ya et al. (1996)
PHO88	2R	P_i-transporter associate	Yompakdee et al. (1996b)
GTR1	13L	P_i-transporter associate	Bun-ya et al. (1992)
Regulatory genes			
PHO2 (*BAS2*)	4L	DNA-binding transcriptional activator	Sengstag and Hinnen (1987) Yoshida et al. (1989a)
PHO4	6R	DNA-binding transcriptional activator	Yoshida et al. (1989a) Ogawa and Oshima (1990)
PHO6 (*THI2*)	2R	Required for *PHO3* expression	Nishimura et al. (1992a,b)
PHO7		Required for *PHO3* expression	Toh-e et al. (1975a)
THI3	4	Required for *PHO3* expression	Nishimura et al. (1992b)
PHO80 (*TUP7*)	15L	Cyclin; inhibitor of Pho4p	Toh-e and Shimauchi (1986) Kawasaki et al. (1991) Uesono et al. (1992) Kaffman et al. (1994)
PHO81	7R	Inhibitor of Pho80p/Pho85p	Schneider et al. (1994) Ogawa et al. (1993, 1995a)
PHO85	16L	Cyclin-dependent kinase; inhibitor of Pho4p	Uesono et al. (1987) Toh-e et al. (1988) Kaffman et al. (1994)

tion to acquire thiamine (Nosaka, 1990; Schweingruber et al., 1986), and another alkaline phosphatase encoded by *PHO13* exhibiting specificity only for *p*-nitrophenylphosphate (Kaneko et al., 1989). The *PHO3* gene, which is closely linked with *PHO5* on the right arm of chromosome II (Toh-e and Oshima, 1975), is only expressed when *PHO5* is not expressed (Tait-Kamradt et al., 1986) and modestly repressed by thiamine (Nosaka, 1990; Schweingruber et al., 1986) under the positive regulation of the proteins encoded by *PHO6* (= *THI2*) (Nishimura et al., 1992a; Toh-e et al., 1975a) and *PHO7* (Toh-e et al., 1975a) or *THI3* (Nishimura et al., 1992b). The *PHO13* gene is expressed constitutively (Toh-e et al., 1976).

Beside phosphatases, *S. cerevisiae* has at least two systems for P_i uptake; one with a low K_m value (8.2 μM) and the other with a high K_m value (770 μM) for external P_i (Tamai et al., 1985). The low K_m system is repressed by P_i via the same system as that for rAPase regulation and requires the *PHO84* gene product for its function. From the nucleotide sequence of cloned *PHO84*, it was predicted that Pho84p is a 65-kDa 596-amino acid protein homologous to that of several sugar transporters of mammalian, yeast, and bacterial origin (Bun-ya et al., 1991). Pho84p has 12 putative transmembrane segments, each consisting of 21 amino acid residues and the majority of which are hydrophobic. These putative transmembrane segments are separated into two groups by a 74-amino acid segment, which is rich in charged amino acid residues and is located between the sixth and seventh segments. These architectural characteristics and the amino acid sequences of sev-

eral segments of Pho84p are highly similar to those of proteins in the human glucose transporter superfamily (Mueckler et al., 1985).

During the sequence determination of *PHO84*, another open reading frame (ORF) was detected in a cloned *PHO84* fragment at a head-to-head position of the *PHO84* ORF (Bun-ya et al., 1992). This ORF, designated *GTR1*, encodes a 310-amino acid protein containing the characteristic tripartite consensus elements for binding GTP, conserved in GTP-binding proteins of the *ras* family, in its N-terminal region. Disruption of *GTR1* resulted in a slow-growth phenotype at 30°C and no growth at 15°C, and conferred the phenotypes of constitutive synthesis of rAPase, reduced P_i uptake activity, and resistance to arsenate (a toxic P_i analog) on the cells, similar to the *pho84* mutant. Therefore, Gtr1p encoded by *GTR1* might be associated with the P_i transport mechanism.

Two other genes that may be involved in the P_i transport mechanism were detected in a mutant exhibiting arsenate resistance and constitutive synthesis of rAPase (Bun-ya et al., 1996). The mutant was shown to have double mutations designated *pho86* and *pho87*. This double mutant transcribes *PHO84* even in a high-P_i medium but has a severe defect in P_i uptake. The *pho86* single mutation conferred a substantial reduction of P_i uptake on the cells, while the *pho87* single mutant or a disruptant of it showed no apparent difference in its phenotypes with those of the wild-type cells. Sequence determination of a partial *PHO87* coding region revealed that it is identical to *YCR524* on the right arm of chromosome III (Thierry et al., 1990). The published nucleotide sequence of *YCR524* predicts that *PHO87* encodes a 923-amino acid protein with a highly charged N-terminal half and a C-terminal half consisting of twelve transmembrane segments (Bun-ya et al., 1996). The Pho86p protein, deduced from the nucleotide sequence of *PHO86*, is a 34-kDa membrane protein consisting of 311 amino acid residues with two strongly hydrophobic segments in its N-terminal half (Yompakdee et al., 1996a). Thus, Pho86p and Pho87p are likely to be involved in the P_i transport mechanism along with Pho84p and Gtr1p. Both disruption (Yompakdee et al., 1996a) and increased dosage (Yompakdee et al., 1996b) of *PHO86* activates P_i-inhibited Pho81p produced under the control of the *GAL1* promoter (described below).

Another putative membrane protein, Pho88p, was identified by screening for a multicopy suppressor of the rAPase$^-$ phenotype caused by the P_i inhibition of Pho81p (Yompakdee et al., 1996b). The nucleotide sequence of the cloned fragment revealed that the gene, encoding Pho88p, is identical to *YBR106w* (Feldmann et al., 1994). The hydrophobicity profile of the amino acid sequence predicted from the published nucleotide sequence of *YBR106w* and the cellular location suggested that Pho88p is a membrane protein. Both disruption and high dosage of *PHO88*, similarly to *PHO86*, resulted in the activation of P_i-inhibited Pho81p and reduced P_i uptake. The double disruption of *pho86* and *pho88* resulted in a somewhat higher level of rAPase activity in a high-P_i medium and conferred a higher arsenate resistance on the cells than on cells with single disruptants of these genes. These characteristics of Pho88p strongly suggest that it is also involved in P_i uptake and, together with Pho86p, modulates Pho81p activity as a primary receptor for P_i signals from the environment.

REGULATORY CIRCUITS

For the regulation of the activities of the above enzymes and the P_i uptake system, a complex interplay between numerous genes and proteins is required. In a current model for phosphatase regulation (Fig. 1), a positive regulator (or positive factor) Pho4p, coded for by *PHO4*, is indispensable for the transcriptional activation of the above *PHO* structural genes by binding at a specific cis-acting regulatory site(s), termed upstream activation site (UAS),

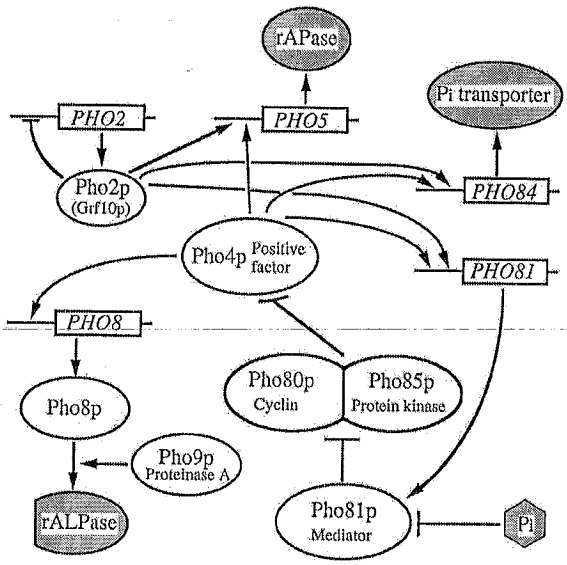


Fig. 1. The genetic regulatory circuit for phosphatase genes. The protein factors, enzymes, and enzyme precursors in the regulatory circuit are shown in open ovals, and the genes regulated by these proteins are shown in boxes. The factors in the shaded hexagon and ovals, P_i, rAPase (repressible acid phosphatase), rALPase (repressible alkaline phosphatase), and P_i-transporter are the input signal to and outputs of the regulatory circuit, respectively. The regulatory factor, for which the corresponding gene is not shown, is known to be produced constitutively in all cases, except *PHO9* (= *PEP4*), the regulation mechanism of which is unknown. The arrowheads in the regulatory circuit indicate a positive or stimulating function of the factors, and when replaced by the bars, they indicate repressive or inhibitory functions. Pho9p (= Pep4p, proteinase A) is required for processing of the Pho8p polypeptide. The *PHO3, PHO6* (= *THI2*), *PHO7, THI3, PHO10, PHO11, PHO86, PHO87, PHO88*, and *GTR1* genes and relevant proteins were not shown for simplicity.

in the promoter. An additional DNA binding protein, Pho2p (also known as Bas2p or Grf10p; Yoshida et al., 1989b), is involved in the transcription of all the *PHO* structural genes except *PHO8*, but it does not have a direct function in the transduction of P_i signals. Pho2p activates the transcription of numerous other genes, including *HIS4* (Arndt et al., 1987; Tice-Baldwin et al., 1989), *TRP4* (Braus et al., 1989), *CYC1* (Sengstag and Hinnen, 1988), genes for adenine biosynthesis (Arndt et al., 1987; Daignan-Fornier and Fink, 1992), and *HO* involved in the mating-type switching mechanism (Brazas and Stillman, 1993a,b) in combination with a cognate positive regulator.

In a high-P_i medium, a complex of negative regulators, Pho80p (cyclin) and Pho85p (cyclin-dependent protein kinase), phosphorylates five serine residues of Pho4p (Kaffman et al., 1994). The complex then turns off the transcription of *PHO* genes by reducing the nuclear location of Pho4p (O'Neill et al., 1996). When the P_i concentration in the medium is sufficiently low, Pho86p and Pho88p may stimulate Pho81p activity (Yompakdee et al., 1996b). Then the activated Pho81p inhibits the function of the Pho80p/Pho85p complex (Ogawa et al., 1995a; Schneider et al., 1994), thus allowing Pho4p, in collaboration with Pho2p, to activate the transcription of the *PHO* genes (Fig. 2; see review, Lenburg and O'Shea, 1996).

The *PHO4* (Yoshida et al., 1989a), *PHO80* (Madden et al., 1990), and *PHO85* (Madden et al., 1990; Uesono et al., 1992) genes are transcribed constitutively at low levels. *PHO2* transcription is also at a low level but is self-regulated (Yoshida et al., 1989a). The transcription of *PHO86* is under relaxed regulation of P_i, while that of *PHO87* is constitutive (Bun-ya et al., 1996), and that of *PHO88* is not affected by the P_i concentration in the medium (Yompakdee et al., 1996b). *GTR1* is transcribed at an extremely low level independent of the P_i concentration in the medium, while it shares the promoter region with *PHO84* but in the reverse direction (Bun-ya et al., 1992). In contrast, the transcription of *PHO81*, encoding Pho81p, is repressed by P_i via the same *PHO* regulatory system, indicating that the *PHO* regulatory circuit forms a positive feedback loop (Ogawa et al., 1993; Yoshida et al., 1989b).

Positive regulators Pho2p and Pho4p are DNA binding proteins. The nucleotide sequence of *PHO4* suggests that Pho4p consists of 312 amino acid residues (Yoshida et al., 1989a). It has the characteristics of a basic region beside an amphipathic helix-loop-helix (bHLH) structure known as a DNA binding domain and homodimer formation (Murre and Baltimore, 1992) in the 85-amino acid C-terminal region (Ogawa and Oshima, 1990), and binds at a specific UAS sequences in *PHO* promoters (Ogawa et al., 1994). The N-terminal region (amino acids 1 to 109), which is rich in acidic amino acids, was suggested to act as the activation domain for gene transcription and a region between amino acids 203 to 227 might be involved in oligomer formation.

The nucleotide sequence of *PHO2* predicts that Pho2p is composed of 519 (Yoshida et al., 1989a) or 559 (Sengstag and Hinnen, 1987) amino acid residues, and contains a homeodomain sequence between amino acids 76 to 135 or between amino acids 77 to 136 (Bürglin, 1988). A portion of the homeodomain sequence of Pho2p has a similar amino acid sequence to that of a short stretch found in Pho4p. Pho2p also has a stretch of glutamine residues (14 of 18 amino acids) on the N-terminal side of the homeodomain sequence. It was shown that Pho2p forms a ternary complex with Pho4p on a *PHO* promoter (Barbarić et al., 1996) by binding at A/T-rich segment(s) beside the UAS site (Magbanua et al., 1997a). It is thought that the Pho2p/Pho4p/UAS ternary complex may interact with the TBP, TFIIB, and TFIIEβ, and initiate transcription of the *PHO* genes (Magbanua et al., 1997b).

Proteins involved in the transduction of P_i signals to Pho4p. Genetic evidence indicated that the negative regulators, Pho80p and Pho85p, inhibit Pho4p function but not the expression of *PHO4* or *PHO2* (Oshima, 1982; Yoshida et al., 1989a). Pho80p, deduced from the nucleotide sequence of *PHO80*, is a 34-kDa protein (Uesono et al., 1992) composed of 293 amino acid residues (Toh-e and Shimauchi, 1986). A coordinated increase of *PHO80* dosage was found to cancel the dosage effect of *PHO4* on the expression of *PHO5* (Yoshida et al., 1989b) and a mutation occurring in *PHO80* suppressed a specific *PHO4*^{c} mutation (Okada and Toh-e, 1992). Based on these findings, it was thought that Pho80p directly interacts with Pho4p for the transduction of P_i signals. On the other hand, Pho85p, encoded by *PHO85* which contains an intron at the 6th codon, is a 36-kDa 305-amino acid protein known as a homologue of Cdc28p, a cyclin-dependent protein kinase of *S. cerevisiae* (Toh-e et al., 1988; Uesono et al., 1987, 1992). It was thought that Pho85p kinase phosphorylates Pho80p, resulting in Pho80p activation (Uesono et al., 1992). Recent reexamination of the amino acid sequence of Pho80p, however, has revealed that it has significant similarities to those of two yeast cyclins, Hcs26p (Ogas et al., 1991) and OrfDp (Fröhlich et al., 1991) in a conserved region of the cyclins (Kaffman et al., 1994). Pho80p and Pho85p form a complex and this complex, but neither protein alone, efficiently phosphorylates Pho4p in vitro and in vivo (Kaffman et al., 1994).

It was believed that the role of the cyclins and cyclin-dependent kinases is in cell cycle control. Pho80p and Pho85p, however, are thought to be unrelated to the cell cycle. This suggests a broader biochemical repertoire for cyclins and cyclin-dependent kinases in yeast (O'Neill and O'Shea, 1995). Pho85p kinase activity is apparently not restricted to the *PHO* system. A cyclin determines the specificity of the partner cyclin-dependent kinase, and Pho85p seems to associate with two other cyclins, Hsc26p and

OrfDp, besides Pho80p (Lenburg and O'Shea, 1996). Since both of these cyclins have been implicated in cell cycle regulation, Pho85p might therefore link cell-cycle events to the cell's nutritional need for phosphate.

According to the genetic analysis, the *pho81* mutation is hypostatic under the *pho80* and/or *pho85* mutation (Oshima, 1982). This fact indicates that Pho81p functions in the upstream of the Pho80p/Pho85p complex (Fig. 1). The nucleotide sequence of *PHO81* suggests that Pho81p is a 134-kDa 1179-amino acid protein which has six repeats of a 33-amino acid sequence homologous to the ankyrin repeat (Lux et al., 1990) between amino acids 424 to 656 and an asparagine-rich region (26 of 33 amino acid residues) between amino acids 222 to 254 (Ogawa et al., 1993). The transcription of *PHO81* is regulated by P_i via the *PHO* regulatory system including both Pho4p and Pho2p (Fig. 1). Experiments with a chimeric gene consisting of the *PHO81* ORF connected with the *GAL1* promoter revealed that stimulation of the *PHO* regulatory system requires both increased transcription of *PHO81* and P_i-starvation conditions (Ogawa et al., 1995a). These findings suggest that Pho81p is inactive under high-P_i conditions.

In a low-P_i medium, Pho81p binds with and inhibits the kinase activity of the Pho80p/Pho85p complex (Ogawa et al., 1995a; Schneider et al., 1994), and allows underphosphorylated Pho4p to be imported into the nucleus and to turn on the transcription of *PHO* genes (Fig. 2). In the high-P_i condition, Pho81p is, however, inactive as an inhibitor. Thus, the Pho80p/Pho85p complex is active for hyperphosphorylation of Pho4p and turns off the transcription of *PHO* genes, as described above. Hirst et al. (1994), however, claimed that Pho81p interacts directly with Pho4p, along with Pho80p, under conditions of *PHO* repression and that Pho80p and Pho81p dissociate from Pho4p under derepressive conditions.

The minimum amino acid sequence required for the above Pho81p function in vivo was narrowed down to a 141-amino acid segment (amino acids 584 to 724), which contains the fifth and sixth repeats of the ankyrin-like motif (Ogawa et al., 1995a). Schneider et al. (1994) also demonstrated that a similar short fragment of Pho81p (amino acids 400 to 720) inhibits the kinase activity of the Pho80p/Pho85p complex in vivo. The third to sixth repeats of the ankyrin-like motif of Pho81p have significant similarities to that of $p16^{INK4}$ which inhibits the activity of the human cyclin D/CDK4 kinase complex (Serrano et al., 1993). Deletion analyses revealed that the N- and C-terminal regions of Pho81p function as negative and positive regulatory domains, respectively, for the minimal 141-amino acid region (Ogawa et al., 1995a). The negative regulatory activity of the N-terminal domain was antagonized by a C-terminal segment of Pho81p supplied in trans. All four known classes of *PHO81*^{c} mutations that conferred significant rAPase activity on cells grown in a high-P_i medium occur in the N-terminal half of Pho81p.

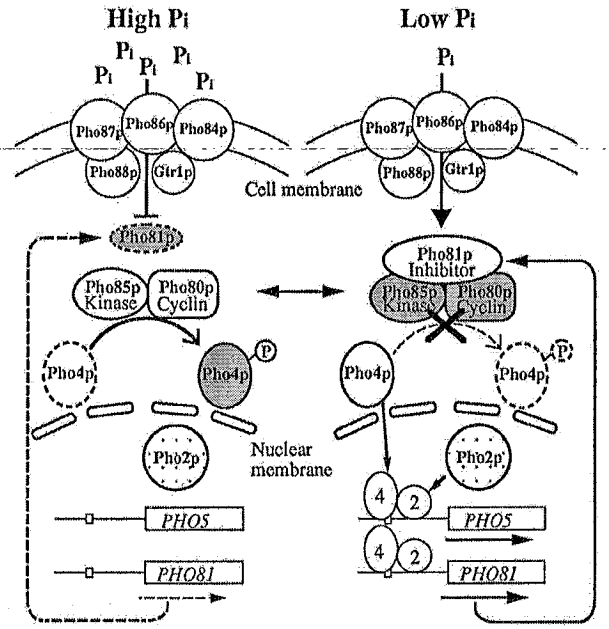


Fig. 2. The molecular mechanism for the transduction of extracellular P_i signals to the *PHO* structural genes in the nucleus. The shaded regulatory proteins are the inactivated forms of the relevant one. The arrows shown underneath the *PHO5* and *PHO81* genes indicate transcription of these genes. The broken arrow from the *PHO81* to Pho81p indicates basal expression of *PHO81*. For the detailed explanation, see text. The other structural genes in the *PHO* system were eliminated from the figure for simplicity.

Specific UAS sequence for Pho4p binding. Genetic evidence suggests that Pho2p and Pho4p bind to or interact with a specific UAS in the promoter region of the *PHO* structural genes (Oshima, 1982). Such sites of *PHO5* have been studied in several laboratories (Bergman et al., 1986; Nakao et al., 1986; Rudolph and Hinnen, 1987). The proposed nucleotide sequences for the UAS function by these authors were, however, in conflict with each other. To clarify the Pho4p-binding site, we investigated the *PHO8* promoter, because only Pho4p, and not Pho2p, is required for *PHO8* transcription. We found that a 6-bp sequence, CACGTG, at nucleotide position –535 relative to the ATG start codon (Fig. 3) is the Pho4p binding site in the *PHO8* promoter, as determined by deletion, modification, and gel retardation analyses (Hayashi and Oshima, 1991). The protection of this CACGTG sequence by purified Pho4p from DNase I digestion in vitro was also shown (Barbarić et al., 1992; Ogawa et al., 1994).

Two putative UASs in the *PHO5* promoter, designated UAS_p1 in the region from nucleotide positions –373 to –347 and UAS_p2 located from nucleotide positions –262 to –239 (Fig. 3), have been detected by in vitro binding of Pho4p (Vogel et al., 1989). These regions contain one copy each of, respectively, the CACGTT and CACGTG motifs (Fig. 4). We confirmed that purified Pho4p protects these sites

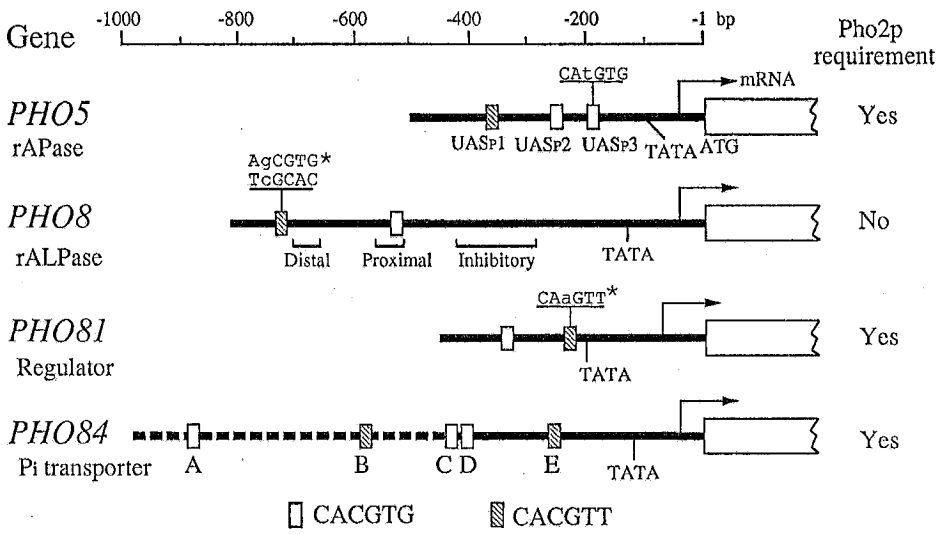


Fig. 3. Distribution of the 6-bp motif, CACGTK (K = G or T), in the promoter regions of the *PHO* genes. The open box at the right end of individual genes represents an ORF. Short open and hatched boxes indicate CACGTG and CACGTT motifs, respectively. Three 6-bp sequences, analogous to the 6-bp motif, shown in the *PHO5, PHO8*, and *PHO81* promoters are indicated with their sequences (lowercase letters indicate deviation from the 6-bp motifs). The region in the *PHO84* promoter shown with a broken line bearing sites A and B is dispensable for regulatory function. The function and Pho4p binding ability of the sequence analogous to the 6-bp motif in the *PHO8* promoter (marked with an asterisk) were described by Barbarić et al. (1992), but that of the *PHO81* promoter has not yet been examined.

from DNase I digestion in vitro (Ogawa et al., 1994). In the same experiment, we found that another 21-bp region, bearing a CATGTG sequence, in the *PHO5* promoter from nucleotide positions –202 to –182 (Figs. 3 and 4) was also protected weakly by Pho4p. Deletion of the above 21-bp region resulted in lower activity of *PHO5* transcription (Rudolph and Hinnen, 1987). Thus, this 6-bp sequence, designated as UAS_p3, might also be involved in the *PHO5* regulation.

We also examined UASs in the *PHO81* promoter (Ogawa et al., 1993). Results of a deletion analysis showed that a 95-bp region, extending from nucleotide positions –385 to –291, could sense the P_i signals. A 19-bp region in the 95-bp fragment was protected by Pho4p from DNase I digestion in vitro (Fig. 4). We found one copy of the CACGTG motif in the 19-bp region and no other CACGTG or CACGTT motifs in the 95-bp region. It is known that *PHO81* transcription is severely reduced in a *pho2* mutant (Yoshida et al., 1989b). The above 95-bp fragment, however, could not respond to a *pho2* disruption allele, but a 488-bp fragment, bearing a region of *PHO81* DNA from nucleotide positions –385 to +103, could respond to the *pho2* disruption on a YCp-type plasmid (Ogawa et al., 1993). This suggests that a specific site for interaction with Pho2p might be located in a region downstream of nucleotide position –290. A CAAGTT sequence was found at nucleotide position –233 in the *PHO81* promoter (Fig. 3), but the function of this sequence remains to be examined.

The promoter region of *PHO84* bears five copies of the 6-bp motifs at nucleotide positions –880 (site A), –587 (site B), –436 (site C), –414 (site D), and –262 (site E) relative to the ATG start codon (Fig. 3). Binding of Pho4p to these five copies of the 6-bp motif was confirmed by DNase I footprinting (Ogawa et al., 1994). We found the highest affinity for Pho4p at site D, whereas sites C and E had somewhat weaker affinities than site D. The A and B sites also had some affinity but much weaker than site E. Analyses of the promoter with deletion and base substitutions in the above 6-bp motifs revealed that the 6-bp motifs at site C or D and E are essential and sufficient for the full regulation of *PHO84* transcription, whereas sites A and B are dispensable (Ogawa et al., 1995b).

Eight effective UAS sites, each one located in the middle of the 19 to 22 bp footprinting regions, protected by Pho4p from DNase I digestion in vitro, in the promoters of *PHO5*, *PHO8*, *PHO81*, and *PHO84* (Ogawa et al., 1994), were compared along with a sequence found at the distal site in the *PHO8* promoter (Barbarić et al., 1992). We found that six of the UASs have the CACGTG sequence (type 1; Fig. 4) and three have the CACGTT sequence (type 2).

The 6-bp CACGTG core sequence has been found to have similarity to a highly conserved octanucleotide, RTCACRTG (where R is a purine residue) called CDEI, located in centromeres and in the promoters of various genes for methionine synthesis in *S. cerevisiae* (Baker and Masison, 1990; Cai and Davis, 1990; Mellor et al., 1990). The CDEI-binding protein (i.e., Cpf1p, CP1 or Cbf1p) has a bHLH structure in its DNA-binding domain. Although crosstalk was not detected between the phosphatase and methionine genes in wild-type cells, overexpression of *PHO4* was found

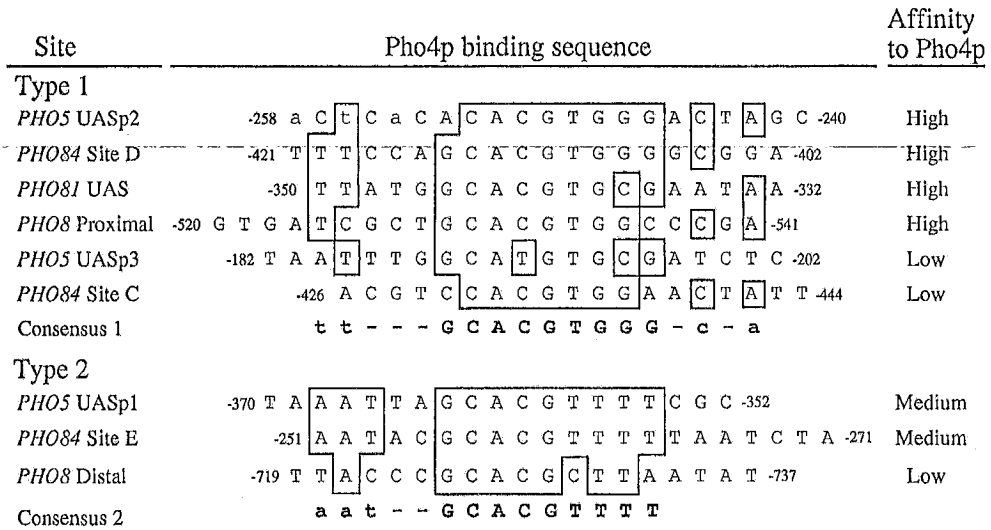


Fig. 4. Regions in the *PHO* promoters protected from DNase I digestion in vitro by Pho4p. The protected sequences observed in previous studies (Ogawa et al., 1993, 1994) are listed for strands having sequences conforming most closely to GCACGTGGG or GCACGTTTT along with the data of Barbarić et al. (1992) for the similar sequences at the distal site in the *PHO8* promoter. These sequences were classified into two groups depending on whether they contained the CACGTG (type 1) or CACGTT (type 2) sequence in the core region. The relative affinity of each sequence to Pho4p was estimated from the amount of Pho4p protein required for protection from DNase I digestion (Ogawa et al., 1994). The numerals at both ends of the sequences indicate nucleotide position of the 5'- and 3'-ends of the sequences in the promoter relative to the translation initiation codon of the respective genes.

to suppress the methionine auxotrophy of a *cep1* (= *cpf1*) mutant (O'Connell and Baker, 1992), under conditions of methionine starvation (O'Connell et al., 1995). In contrast, *CEP1* overexpression failed to suppress the rAPase$^-$ phenotype of a *pho4* mutant (O'Connell and Baker, 1992). In this context, we speculated that the 6-bp sequence, CACGTG, is not sufficient for the specific recognition of a cognate positive regulator, and that the nucleotide sequences flanking the 6-bp sequence might have important functions in the discrimination of Pho4p and Cpf1p. This view was supported by the findings that flanking by a T residue at the 5'-end of the CACGTG motif is preferable for Cpf1p interaction but not for Pho4p interaction (Fisher and Goding, 1992). We found that the 6-bp motif regulated by Pho4p was flanked by a G residue at the 5' end and GG (for type 1 core sequence) or TT (for type 2 core sequence) residues at the 3' end (Fig. 4). Therefore, we proposed that the GCACGTGGG and GCACGTTTT motifs are consensus sequences in the *PHO* promoter (Ogawa et al., 1995b). Recent analysis of a complex of the bHLH domain of Pho4p with UAS DNA by X-ray crystallography supported the significance of the G and GG residues flanking, respectively, at 5'- and 3'-ends of the CACGTG motif (Shimizu et al., 1997). The observation that *CEP1* overexpression could not suppress a *pho4* mutation (O'Connell and Baker, 1992) suggests that Cpf1p has no affinity to the GCACGTTTT sequence of UAS_p1 in the nucleosome-free region (Almer and Hörz, 1986) of the *PHO5* promoter.

Pho2p/Pho4p binding is aided by A/T-rich segment(s). The transcription of all these *PHO* structural genes, except *PHO8*, is dependent on the other positive regulator, Pho2p. Thus, Pho2p might also be involved in the *PHO* regulation. It was demonstrated by use of the two-hybrid system that Pho2p interacts with Pho4p depending on the P_i concentration in the cultivation medium (Hirst et al., 1994) and the accessibility of the activation domain of Pho4p to the transcriptional machinery is enhanced by the interaction with Pho2p (Shao et al., 1996). Barbarić et al. (1996) reported that Pho4p and Pho2p form a ternary complex with UAS DNA and that a possible Pho2p binding site partially overlaps the UAS_p1 site with a GCACGTTTT motif. We could also detect direct interaction between Pho4p and Pho2p using immunoprecipitation and a protein binding assay (Magbanua et al., 1997b). This interaction results in an increased binding affinity of Pho4p to the UAS. Both the Pho4p and Pho2p proteins bind or interact with the basal transcription factors TBP, TFIIB, and TFIIEβ. These observations suggest that the *PHO* gene transcription is initiated by direct contact of the Pho4p/Pho2p complex with the transcription machinery.

In the evaluation of the above 9-bp motifs on their transcriptional activity, we observed that a short A/T-rich segment(s) flanking the 9-bp motif has a role in the binding of Pho2p (Magbanua et al., 1997a), as suggested by Parent et al. (1994). Various 30-bp fragments of different *PHO* promoters, each bearing a 9-bp motif in the middle, and connected with 6-bp *Xho*I cohesive ends, were inserted

into a *CYC1-lacZ* fusion gene on a YCp- or YEp-type plasmid. Then the β-galactosidase activities of the yeast cells with the wild-type or *pho2* disruptant allele and harboring one of the above plasmids were examined. Although the *CYC1* promoter was reported to respond to a *pho2* mutation (Sengstag and Hinnen, 1988), we confirmed that the *CYC1* promoter fragment used was unresponsive to Pho2p and P_i signals. No expression of the *lacZ* gene was detected with the above 36-bp fragment bearing UAS_p2 of *PHO5* which possessed the ACACGTGGG sequence (Magbanua et al., 1997a). Similar 36-bp fragments each bearing UAS_p1 (GCACGTTTT) of *PHO5*, site D (GCACGTGGG) or E (GCACGTTTT) of *PHO84* showed UAS activity in response to P_i concentration in the medium in the wild-type cells, but not in the *pho2* disruptant cells. We found that the Pho2p-responsive UASs are flanked by one or two copies of an A/T-rich segment, whereas UAS_p2 is not (Fig. 5). Gel retardation and competition experiments with Pho2p, tagged with T7 epitope on N-terminal and six histidine residues at C-terminal, showed that the chimeric protein binds to the 36-bp fragments bearing A/T-rich segment(s) but not appreciably to that without A/T-rich segment(s). These facts strongly suggest that the A/T-rich segments flanking the *PHO* UASs are Pho2p binding sites and play an important role in *PHO* regulation.

Pho2p also binds to the *HIS4, ADE2, ADE5,7* and *TRP4* promoters. The Bas1p and Bas2p (= Pho2p) proteins are both required for activation of basal transcription of the *HIS4* gene. The results of DNase I footprinting experiments (Tice-Baldwin et al., 1989) demonstrated that Pho2p binding site, which is rich in A/T pairs, partially overlaps over the Bas1p binding site (Fig. 5). It was known that the core TGACTC sequence alone, to which Bas1p binds, is not sufficient for the proper regulation of the *ADE2* and *ADE5,7* promoters unless a Pho2p binding site is also present (Daignan-Fornier and Fink, 1992). The Bas1p binding sites, shown by the above authors in the *ADE2* and *ADE5,7* promoters, are flanked by one or two copies of A/T-rich segment(s) (Fig. 5). Although the Pho2p binding of the *ADE1* and *ADE8* promoters has not been demonstrated

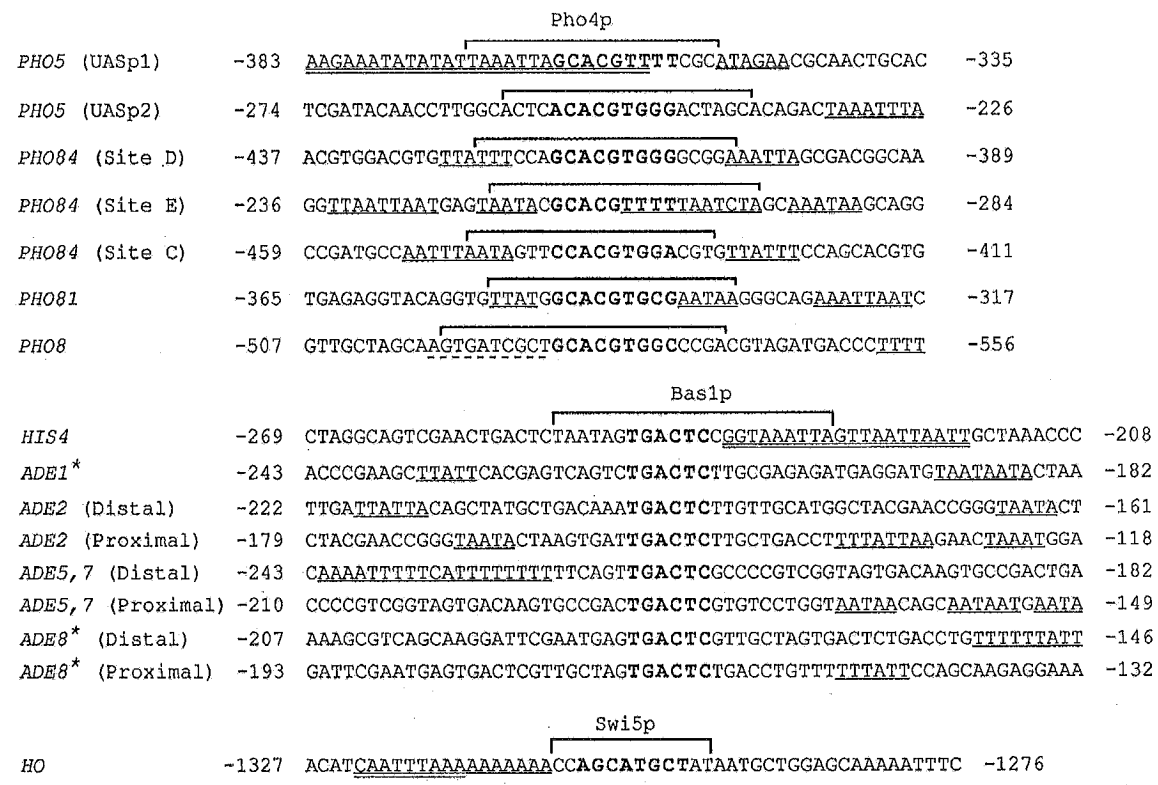


Fig. 5. Alignment of UAS sequences to which relevant regulatory proteins bind together with Pho2p. The consensus sequences in the promoter for Bas1p (Daignan-Fornier and Fink, 1992), Pho4p (Barbarić et al., 1996; Ogawa et al., 1994, 1995b), and Swi5p (Brazas and Stillman, 1993b) were as described. The core UAS sequences for binding of the indicated proteins are written in boldface letters and the regions protected in the footprinting experiments are indicated by brackets. The asterisks indicate that the Bas1p binding sites in the *ADE1* and *ADE8* promoters were identified by sequence comparison (Daignan-Fornier and Fink, 1992), Pho2p binding sites identified in footprinting experiments are indicated by double underlines, while A/T-rich segments are singly underlined. The broken underline indicates the palindromic sequence flanking the 9-bp UAS motif of *PHO8*. The numerals at both ends of the sequences indicate nucleotide position of the 5'- and 3'-ends of the sequences in the promoter of the respective genes.

in vitro, the alignment of their promoter sequences indicates the presence of A/T-rich segment(s) flanking the TGACTC motif (Fig. 5). Pho2p also binds to the promoter of the tryptophan biosynthetic *TRP4* gene at the A/T rich UAS_1 site, to which the general control regulator, Gcn4p, binds (Braus et al., 1989). In this case, Pho2p and Gcn4p, however, bind to the UAS_1 site in a mutually exclusive manner. Pho2p recognizes similar A/T-rich segments at a region adjacent to the Swi5p binding site in the *HO* promoter (Brazas and Stillman, 1993a,b). This was also supported by the experiments with increased spacing between putative Pho2p and Swi5p binding sites with A/T pair of nucleotides in the *HO* promoter (Brazas et al., 1995).

The 36-bp *PHO81* UAS fragment which was found to have 4- or 5-bp A/T segments flanking the 9-bp UAS motif had weak affinity to Pho2p and could not exhibit appreciable UAS activity under the normal dosage of Pho4p (Magbanua et al., 1997a). We speculated that the 4- and 5-bp A/T segments at the *PHO81* UAS are covered completely by Pho4p as it protects a 19-bp region, bearing these A/T segments, from the DNase I digestion (see Fig. 5). Thus, it may result in a reduced level of UAS activity due to the ineffective binding of Pho4p and/or Pho2p to the *PHO81* UAS. The GCAAGTTTT sequence at nucleotide position –234 in the *PHO81* promoter (Ogawa et al., 1993) is flanked with a long A/T-rich segment at the 5' side, but not on the 3' side. In contrast, the similar 36-bp *PHO8* UAS fragment has no apparent A/T-rich segment and no appreciable binding affinity to Pho2p was observed. However, the level of expression of the *CYC1-lacZ* fusion gene inserted with the 36-bp *PHO8* UAS fragment under high-P_i conditions was relatively high and the differential ratio of the *lacZ* expression in low- and high-P_i media was about 2.5 (Magbanua et al., 1997a) as in the case of the genomic *PHO8* expression (Kaneko et al., 1987; Toh-e et al., 1976). This may be due to the fact that the 9-bp UAS of *PHO8* has a 10-bp palindromic sequence, AGTGATCGCT, adjacent to it (Fig. 5). This sequence may respond to another unknown regulatory protein.

CONCLUDING REMARKS

We can now attain substantial understanding into the broad outline of function of the regulatory genes and proteins in the *PHO* system. However, many questions remain unanswered. One of the major unsolved mechanisms is of how extracellular P_i signals are detected. In the recent genetic study of mutants exhibiting a defect in P_i uptake, we could identify three putative membrane proteins, Pho86p, Pho87p, and Pho88p, which might be involved in the P_i uptake mechanism in association with Pho84p and Gtr1p (Bun-ya et al., 1996; Yompakdee et al., 1996a,b). Among them, Pho86p and Pho88p are particularly interesting as they were suggested to modulate Pho81p function in accordance with extracellular P_i concentration.

Another question remaining to be addressed is "what is the consequence of the Pho4p/Pho2p/UAS ternary complex formation in recruiting the transcriptional machinery to the *PHO* genes?". It was argued that nucleosome remodeling is evidently a direct consequence of the recruitment of the transcriptional machinery (Ptashne and Gann, 1997). Hörz and his colleagues have examined this idea using the *PHO5* gene (see review, Svaren and Hörz, 1997). These authors proposed that the *PHO5* promoter is protected by positioned nucleosomes in repressed conditions, except for the UAS_p1 site which is positioned at a nucleosome-free site. Upon derepression of *PHO5*, disruption of nucleosomes occurs due to the binding of Pho4p at the UAS_p1 site, possibly in association with Pho2p. Then the binding of an additional Pho4p molecule at another site, UAS_p2, induces removal of four nucleosomes from the promoter, thereby facilitating the binding of the transcriptional machinery to the promoter. Thus, these authors claimed that the chromatin disruption at the promoter is the prerequisite for transcription of the *PHO5* gene. The powerful combination of genetic and biochemical studies available in *S. cerevisiae* will lead to a more detailed and comprehensive understanding of the above questions in the near future.

Our studies cited in this article were carried out in the Department of Biotechnology (formerly Department of Fermentation Technology), Osaka University in collaboration with A. Toh-e, S. Harashima, Y. Kaneko, K. Yoshida, N. Ogawa, N. Hayashi, M. Bunya, and many students, and were supported by grants for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan.

REFERENCES

Almer, A., and Hörz, W. (1986) Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. EMBO J. **5**, 2681–2687.

Arima, K., Oshima, T., Kubota, I., Nakamura, N., Mizunaga, T., and Toh-e, A. (1983) The nucleotide sequence of the yeast *PHO5* gene: a putative precursor of repressible acid phosphatase contains a signal peptide. Nucleic Acids Res. **11**, 1657–1672.

Arndt, K., Styles, C., and Fink, G. R. (1987) Multiple global regulators control *HIS4* transcription in yeast. Science **237**, 874–880.

Arnold, W. N. (1972) Location of acid phosphatase and β-fructofuranosidase within yeast cell envelopes. J. Bacteriol. **112**, 1346–1352.

Bajwa, W., Meyhack, B., Rudolph, H., Schweingruber, A. -M., and Hinnen, A. (1984) Structural analysis of the two tandemly repeated acid phosphatase genes in yeast. Nucleic acids Res. **12**, 7721–7739.

Baker, R. E., and Masison, D. C. (1990) Isolation of the gene encoding the *Saccharomyces cerevisiae* centromere-binding protein CP1. Mol. Cell. Biol. **10**, 2458–2467.

Barbarić, S., Fascher, K. -D., and Hörz, W. (1992) Activation of the weakly regulated *PHO8* promoter in *S. cerevisiae*: chromatin transition and binding sites for the positive regulatory protein PHO4. Nucleic Acids Res. **20**, 1031–1038.

Barbarić S., Münsterkötter, M., Svaren, J., and Hörz, W. (1996) The

homeodomain protein Pho2 and the basic-helix-loop-helix protein Pho4 bind DNA cooperatively at the yeast *PHO5* promoter. Nucleic Acids Res. **24**, 4479–4486.

Bergman, L. W., McClinton, D. C., Madden, S. L., and Preis, L. H. (1986) Molecular analysis of the DNA sequences involved in the transcriptional regulation of the phosphate-repressible acid phosphatase gene (*PHO5*) of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **83**, 6070–6074.

Boer, P., and Steyn-Parve, E. P. (1966) Isolation and purification of an acid phosphatase from baker's yeast (*Saccharomyces cerevisiae*). Biochim. Biophys. Acta **128**, 400–402.

Bostian, K. A., Lemire, J. M., Cannon, L. E., and Halvorson, H. O. (1980) *In vitro* synthesis of repressible yeast acid phosphatase: identification of multiple mRNAs and products. Proc. Natl. Acad. Sci. USA **77**, 4504–4508.

Braus, G., Mösch, H. -U., Vogel, K., Hinnen, A., and Hütter, R. (1989) Interpathway regulation of the *TRP4* gene of yeast. EMBO J. **8**, 939–945.

Brazas, R. M., Bhoite, L. T., Murphy, M. D., Yu, Y., Chen, Y., Neklason, D. W., and Stillman, D. J. (1995) Determining the requirements for cooperative DNA binding by Swi5p and Pho2p (Grf10p/Bas2p) at the *HO* promoter. J. Biol. Chem. **270**, 29151–29161.

Brazas, R. M., and Stillman, D. J. (1993a) Identification and purification of a protein that binds DNA cooperatively with the yeast SWI5 protein. Mol. Cell. Biol. **13**, 5524–5537.

Brazas, R. M., and Stillman, D. J. (1993b) The Swi5 zinc-finger protein and Grf10 homeodomain proteins bind DNA cooperatively at the yeast *HO* promoter. Proc. Natl. Acad. Sci. USA **90**, 11237–11241.

Bun-ya, M., Harashima, S., and Oshima, Y. (1992) Putative GTP-binding protein, Gtr1, associated with the function of the Pho84 inorganic phosphate transporter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12**, 2958–2966.

Bun-ya, M., Nishimura, M., Harashima, S., and Oshima, Y. (1991) The *PHO84* gene of *Saccharomyces cerevisiae* encodes an inorganic phosphate transporter. Mol. Cell. Biol. **11**, 3229–3238.

Bun-ya, M., Shikata, K., Nakade, S., Yompakdee, C., Harashima, S., and Oshima, Y. (1996) Two new genes, *PHO86* and *PHO87*, involved in inorganic phosphate uptake in *Saccharomyces cerevisiae*. Curr. Genet. **29**, 344–351.

Bürglin, T. R. (1988) The yeast regulatory gene *PHO2* encodes a homeo box. Cell **53**, 339–340.

Cai, M., and Davis, R. W. (1990) Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. Cell **61**, 437–446.

Clark, D. W., Tkacz, J. S., and Lampen, J. O. (1982) Asparagine-linked carbohydrate does not determine the cellular location of yeast vacuolar nonspecific alkaline phosphatase. J. Bacteriol. **152**, 865–873.

Daignan-Fornier, B., and Fink, G. R. (1992) Coregulation of purine and histidine biosynthesis by the transcriptional activators BAS1 and BAS2. Proc. Natl. Acad. Sci. USA **89**, 6746–6750.

de Steensma, H. Y., de Jonge, P., Kaptein, A., and Kaback, D. B. (1989) Molecular cloning of chromosome I DNA from *Saccharomyces cerevisiae*: localization of a repeated sequence containing an acid phosphatase gene near a telomere of chromosome I and chromosome VIII. Curr. Genet. **16**, 131–137.

Feldmann, H., Aigle, M., Aljinovic, G., André, B., Baclet, M.C., Barthe, C., Baur, A., Bécam, A. -M., Biteau, N., Boles, E., Brandt, T., Brendel, M., Brückner, M., Bussereau, F., Christiansen, C., Contreras, R., Crouzet, M., Czieplucl, C., Démolis, N., Delaveau, Th., Doignon, F., Domdey, H., Düsterhus, S., Dubois, E., Dujon, B., El Bakkoury, M., Entian, K. -D., Feuermann, M., Fiers, W., Fobo, G. M., Fritz, C., Gassenhuber, H., Glansdorff, N., Goffeau, A., Grivell, L. A., de Haan, M., Hein, C., Herbert, C. J., Hollenberg, C. P., Holmstrφm, K., Jacq, C., Jacquet, M., Jauniaux, J. C., Jonniaux, J. -L., Kallesφe, T., Kiesau, P., Kirchrath, L., Kötter, P., Korol, S., Liebl, S., Logghe, M., Lohan, A. J. E., Louis, E. J., Li, Z. Y., Maat, M. J., Mallet, L., Mannhaupt, G., Messenguy, F., Miosga, T., Molemans, F., Müller, S., Nasr, F., Obermaier, B., Perea, J., Piérard, A., Piravandi, E., Pohl, F. M., Pohl, T. M., Potier, S., Proft, M., Purnelle, B., Ramezani Rad, M., Rieger, M., Rose, M., Schaaff-Gerstenschläger, I., Scherens, B., Schwarzlose, C., Skala, J., Slonimski, P. P., Smits, P. H. M., Souciet, J. L., Steensma, H. Y., Stucka, R., Urrestarazu, A., van der Aart, Q. J. M., van Dyck, L., Vassarotti, A., Vetter, I., Vierendeels, F., Vissers, S., Wagner, G., de Wergifosse, P., Wolfe, K. H., Zagulski, M., Zimmermann, F. K., Mewes, H. W., and Kleine, K. (1994) Complete DNA sequence of yeast chromosome II. EMBO J. **13**, 5795–5809.

Fisher, F., and Goding, C. R. (1992) Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. EMBO J. **11**, 4103–4109.

Fröhlich, K.-U., Fries, H.-W., Rüdiger, M., Erdmann, R., Botstein, D., and Mecke, D. (1991) Yeast cell cycle protein CDC48p shows full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. J. Cell Biol. **114**, 443–453.

Hayashi, N., and Oshima, Y. (1991) Specific *cis*-acting sequence for *PHO8* expression interacts with PHO4 protein, a positive regulatory factor, in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11**, 785–794.

Hirst, K., Fisher, F., McAndrew, P. C., and Goding, C. R. (1994) The transcription factor, the Cdk, its cyclin and their regulator: directing the transcriptional response to a nutritional signal. EMBO J. **13**, 5410–5420.

Johnston, M., and Carlson, M. (1992) Regulation of carbon and phosphate utilization: In: The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Gene Expression (eds.: E. W. Jones, J. R. Pringle, and J. R. Broach), pp. 193–281. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Kaffman, A., Herskowitz, I., Tjian, R., and O'Shea, E. K. (1994) Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80-PHO85. Science **263**, 1153–1156.

Kaneko, Y., Hayashi, N., Toh-e, A., Banno, I., and Oshima, Y. (1987) Structural characteristics of the *PHO8* gene encoding repressible alkaline phosphatase in *Saccharomyces cerevisiae*. Gene **58**, 137–148.

Kaneko, Y., Toh-e, A., Banno, I., and Oshima, Y. (1989) Molecular characterization of a specific *p*-nitrophenylphosphatase gene, *PHO13*, and its mapping by chromosome fragmentation in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **220**, 133–139.

Kaneko, Y., Toh-e, A., and Oshima, Y. (1982) Identification of the genetic locus for the structural gene and a new regulatory gene for the synthesis of repressible alkaline phosphatase in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **2**, 127–137.

Kawasaki, H., Matsuzaki, H., Nakajima, R., and Oshima, Y. (1991) The *PHO80/TUP7* locus in *Saccharomyces cerevisiae* is on the left arm of chromosome XV: mapping by chromosome engineering. Yeast **7**, 859–865.

Lemire, J. M., Willcocks, T., Halvorson, H. O., and Bostian, K. A. (1985) Regulation of repressible acid phosphatase gene transcription in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **5**, 2131–2141.

Lenburg, M. E., and O'Shea, E. K. (1996) Signaling phosphate starvation. Trends Biochem. Sci. **21**, 383–387.

Linnemans, W. A. M., Boer, P., and Elbers, P. F. (1977) Localization of acid phosphatase in *Saccharomyces cerevisiae*: a clue to cell wall formation. J. Bacteriol. **131**, 638–644.

Lopandic, K., Deana, A. D., Barbarić, S., and Pinna, L. A. (1987)

Biochemical and genetic evidence that yeast extracellular protein phosphatase activity is due to acid phosphatase. Biochem. Int. **14**, 627–633.

Lux, S. E., John, K. M., and Bennett V. (1990) Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. Nature **344**, 36–42.

Madden, S. L., Johnson, D. L., and Bergman, L. W. (1990) Molecular and expression analysis of the negative regulators involved in the transcriptional regulation of acid phosphatase production in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **10**, 5950–5957.

Magbanua, J. P. V., Fujisawa, K., Ogawa, N., and Oshima, Y. (1997a) The homeodomain protein Pho2p binds at an A/T-rich segment flanking the binding site of the basic-helix-loop-helix protein Pho4p in the yeast *PHO* promoters. Yeast **13**, 1299–1308.

Magbanua, J. P. V., Ogawa, N., Harashima, S., and Oshima, Y. (1997b) The transcriptional activators of the *PHO* regulon, Pho4p and Pho2p, interact directly with each other and with components of the basal transcription machinery in *Saccharomyces cerevisiae*. J. Biochem. **121**, 1182–1189.

Mellor, J., Jiang, W., Funk, M., Rathjen, J., Barnes, C. A., Hinz, T., Hegemann, J. H., and Philippsen, P. (1990) CPF1, a yeast protein which functions in centromeres and promoters. EMBO J. **9**, 4017–4026.

Mueckler, M., Caruso, C., Baldwin, S. A., Panico, M., Blench, I., Morris, H. R., Allard, W. J., Lienhard, G. E., and Lodish, H. F. (1985) Sequence and structure of a human glucose transporter. Science **229**, 941–945.

Murre, C., and Baltimore, D. (1992) The helix-loop-helix motif: structure and function: In: Transcriptional Regulation (eds.: S. L. McKnight and Y. R. Yamamoto), pp. 861–879. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nakao, J., Miyanohara, A., Toh-e, A., and Matsubara, K. (1986) *Saccharomyces cerevisiae PHO5* promoter region: location and function of the upstream activation site. Mol. Cell. Biol. **6**, 2613–2623.

Nishimura, H., Kawasaki, Y., Kaneko, Y., Nosaka, K., and Iwashima, A. (1992a) Cloning and characteristics of a positive regulatory gene, *THI2* (*PHO6*), of thiamine biosynthesis in *Saccharomyces cerevisiae*. FEBS Lett. **297**, 155–158.

Nishimura, H., Kawasaki, Y., Kaneko, Y., Nosaka, K., and Iwashima, A. (1992b) A positive regulatory gene, *THI3*, is required for thiamine metabolism in *Saccharomyces cerevisiae*. J. Bacteriol. **174**, 4701–4706.

Nosaka, K. (1990) High affinity of acid phosphatase encoded by *PHO3* gene in *Saccharomyces cerevisiae* for thiamin phosphates. Biochim. Biophys. Acta **1037**, 147–154.

O'Connell, K. F., and Baker, R. E. (1992) Possible cross-regulation of phosphate and sulfate metabolism in *Saccharomyces cerevisiae*. Genetics **132**, 63–75.

O'Connell, K. F., Surdin-Kerjan, Y., and Baker, R. E. (1995) Role of the *Saccharomyces cerevisiae* general regulatory factor CP1 in methionine biosynthetic gene transcription. Mol. Cell. Biol. **15**, 1879–1888.

Ogas, J., Andrews, B. J., and Herskowitz, I. (1991) Transcriptional activation of *CLN1, CLN2*, and a putative new G1 cyclin (*HCS26*) by SWI4, a positive regulator of G1-specific transcription. Cell **66**, 1015–1026.

Ogawa, N., Hayashi, N., Saito, H., Noguchi, K., Yamashita, Y., and Oshima, Y. (1994) Regulatory circuit for phosphatase genes in *Saccharomyces cerevisiae*: specific *cis*-acting sites in *PHO* promoters for binding the positive regulator Pho4p: In: Phosphate in Microorganisms: Cellular and Molecular Biology (eds.: A. Torriani-Gorini, E. Yagil, and S. Silver), pp. 56–62. American Society for Microbiology, Washington, D. C.

Ogawa, N., Noguchi, K., Sawai, H., Yamashita, Y., Yompakdee, C., and Oshima, Y. (1995a) Functional domains of Pho81p, an inhibitor of Pho85p protein kinase, in the transduction pathway of P_i signals in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **15**, 997–1004.

Ogawa, N., Noguchi, K., Yamashita, Y., Yasuhara, T., Hayashi, N., Yoshida, K., and Oshima, Y. (1993) Promoter analysis of the *PHO81* gene encoding a 134 kDa protein bearing ankyrin repeats in the phosphatase regulon of *Saccharomyces cerevisiae*. Mol. Gen. Genet. **238**, 444–454.

Ogawa, N., and Oshima, Y. (1990) Functional domains of a positive regulatory protein, PHO4, for transcriptional control of the phosphatase regulon in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **10**, 2224–2236.

Ogawa, N., Saitoh, H., Miura, K., Magbanua, J.P.V., Bun-ya, M., Harashima, S., and Oshima, Y. (1995b) Structure and distribution of specific *cis*-elements for transcriptional regulation of *PHO84* in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **249**, 406–416.

Okada, H., and Toh-e, A. (1992) A novel mutation occurring in the *PHO80* gene suppresses the *PHO4*^{c} mutations of *Saccharomyces cerevisiae*. Curr. Genet. **21**, 95–99.

O'Neill, E. M., Kaffman, A., Jolly, E.R., and O'Shea, E. K. (1996) Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. Science **271**, 209–212.

O'Neill, E. M., and O'Shea, E. K. (1995) Cyclins in initiation. Nature **374**, 121–122.

Onishi, H. R., Tkacz, J. S., and Lampen, J. O. (1979) Glycoprotein nature of yeast alkaline phosphatase: formation of active enzyme in the presence of tunicamycin. J. Biol. Chem. **254**, 11943–11952.

Oshima, Y. (1982) Regulatory circuits for gene expression: the metabolism of galactose and of phosphate: In: The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression (eds.: J. N. Strathern, E. W. Jones, and J. R. Broach), pp. 159–180. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Oshima, Y. (1991) Impact of the Douglas-Hawthorne model as a paradigm for elucidating cellular regulatory mechanisms in fungi. Genetics **128**, 195–201.

Parent, S. A., Justice, M. C., Yuan, L. -W., Hopper, J. E., and Bostian, K. A. (1994) Protein-DNA and protein-protein interactions regulating the phosphatase multigene family of *Saccharomyces cerevisiae*: In: Phosphate in Microorganisms: Cellular and Molecular Biology (eds.: A. Torriani-Gorini, E. Yagil, and S. Silver), pp. 63–69. American Society for Microbiology, Washington, D. C.

Plankert, U., Purwin, C., and Holzer, H. (1991) Yeast fructose-2,6-bisphosphate 6-phosphatase is encoded by *PHO8*, the gene for nonspecific repressible alkaline phosphatase. Eur. J. Biochem. **196**, 191–196.

Ptashne, M., and Gann, A. (1997) Transcriptional activation by recruitment. Nature **386**, 569–577.

Rudolph, H., and Hinnen, A. (1987) The yeast *PHO5* promoter: phosphate-control elements and sequences mediating mRNA start-site selection. Proc. Natl. Acad. Sci. USA **84**, 1340–1344.

Schneider, K. R., Smith, R. L., and O'Shea, E. K. (1994) Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. Science **266**, 122–126.

Schweingruber, M. E., Fluri, R., Maudrell, K., Schweingruber, A. -M., and Dumermuth, E. (1986) Identification and characterization of thiamin repressible acid phosphatase in yeast. J. Biol. Chem. **261**, 15877–15882.

Sengstag, C., and Hinnen, A. (1987) The sequence of the *Saccharomyces cerevisiae* gene *PHO2* codes for a regulatory protein with unusual aminoacid composition. Nucleic Acids Res. **15**, 233–

246.

Sengstag, C., and Hinnen, A. (1988) A 28-bp segment of the *Saccharomyces cerevisiae PHO5* upstream activator sequence confers phosphate control to the *CYC1-lacZ* gene fusion. Gene **67**, 223–228.

Serrano, M., Hannon, G. J., and Beach, D. (1993) A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature **366**, 704–707.

Shao, D., Creasy, C. L., and Bergman, L. W. (1996) Interaction of *Saccharomyces cerevisiae* Pho2 with Pho4 increases the accessibility of the activation domain of Pho4. Mol. Gen. Genet. **251**, 358–364.

Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997) Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. EMBO J. **16**, 4689–4697.

Svaren, J., and Hörz, W. (1997) Transcription factors vs nucleosomes: regulation of the *PHO5* promoter in yeast. Trends Biochem. Sci. **22**, 93–97.

Tait-Kamradt, A. G., Turner, K. J., Kramer, R. A., Elliott, Q. D., Bostian, S. J., Thill, G. P., Rogers, D. T., and Bostian, K. A. (1986) Reciprocal regulation of the tandemly duplicated *PHO5*/*PHO3* gene cluster within the acid phosphatase multigene family of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **6**, 1855–1865.

Tamai, Y., Toh-e, A., and Oshima, Y. (1985) Regulation of inorganic phosphate transport systems in *Saccharomyces cerevisiae*. J. Bacteriol. **164**, 964–968.

Thierry, A., Fairhead, C., and Dujon, B. (1990) The complete sequence of the 8.2 kb segment left of *MAT* on chromosome III reveals five ORFs, including a gene for a yeast ribokinase. Yeast **6**, 521–534.

Tice-Baldwin, K., Fink, G. R., and Arndt, K. T. (1989) BAS1 has a Myb motif and activates *HIS4* transcription only in combination with BAS2. Science **246**, 931–935.

Toh-e, A., Kakimoto, S., and Oshima, Y. (1975a) Two new genes controlling the constitutive acid phosphatase synthesis in Saccharomyces cerevisiae. Mol. Gen. Genet. **141**, 81–83.

Toh-e, A., Kakimoto, S., and Oshima, Y. (1975b) Genes coding for the structure of the acid phosphatase in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **143**, 65–70.

Toh-e, A., Nakamura, H., and Oshima, Y. (1976) A gene controlling the synthesis of non specific alkaline phosphatase in *Saccharomyces cerevisiae*. Biochim. Biophys. Acta **428**, 182–192.

Toh-e, A., and Oshima, Y. (1975) Regulation of acid phosphatase synthesis in *Saccharomyces cerevisiae*: In: Proc. 1st Intersectional Congress of IAMS, Vol. 1 (ed.: T. Hasegawa), pp. 396–399. Science Council of Japan, Tokyo.

Toh-e, A., and Shimauchi, T. (1986) Cloning and sequencing of the *PHO80* gene and *CEN15* of *Saccharomyces cerevisiae*. Yeast **2**, 129–139.

Toh-e, A., Tanaka, K., Uesono, Y., and Wickner, R. B. (1988) *PHO85*, a negative regulator of the PHO system, is a homolog of the protein kinase gene, *CDC28*, of *Saccharomyces cerevisiae*. Mol. Gen. Genet. **214**, 162–164.

Uesono, Y., Tanaka, K., and Toh-e, A. (1987) Negative regulators of the PHO system in *Saccharomyces cerevisiae*: isolation and structural characterization of *PHO85*. Nucleic Acids Res. **15**, 10299–10309.

Uesono, Y., Tokai, M., Tanaka, K., and Toh-e, A. (1992) Negative regulators of the PHO system of *Saccharomyces cerevisiae*: characterization of *PHO80* and *PHO85*. Mol. Gen. Genet. **231**, 426–432.

Venter, U., and Hörz, W. (1989) The acid phosphatase genes PHO10 and PHO11 in *S. cerevisiae* are located at the telomeres of chromosomes VIII and I. Nucleic Acids Res. **17**, 1353–1369.

Vogel, K., Hörz, W., and Hinnen, A. (1989) The two positively acting regulatory proteins PHO2 and PHO4 physically interact with *PHO5* upstream activation regions. Mol. Cell. Biol. **9**, 2050–2057.

Yompakdee, C., Bun-ya, M., Shikata, K., Ogawa, N., Harashima, S., and Oshima, Y. (1996a) A putative new membrane protein, Pho86p, in the inorganic phosphate uptake system of *Saccharomyces cerevisiae*. Gene **171**, 41–47.

Yompakdee, C., Ogawa, N., Harashima, S., and Oshima, Y. (1996b) A putative membrane protein, Pho88p, involved in inorganic phosphate transport in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **251**, 580–590.

Yoshida, K., Kuromitsu, Z., Ogawa, N., and Oshima, Y. (1989a) Mode of expression of the positive regulatory genes *PHO2* and *PHO4* of the phosphatase regulon in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **217**, 31–39.

Yoshida, K., Ogawa, N., and Oshima, Y. (1989b) Function of the *PHO* regulatory genes for repressible acid phosphatase synthesis in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **217**, 40–46.

Anal Bioanal Chem (2016) 408:2975–2983
DOI 10.1007/s00216-015-9299-5

RESEARCH PAPER

PEPR: pipelines for evaluating prokaryotic references

Nathan D. Olson[1] · Justin M. Zook[1] · Daniel V. Samarov[2] · Scott A. Jackson[1] · Marc L. Salit[1,3]

Received: 30 September 2015 / Revised: 21 December 2015 / Accepted: 23 December 2015 / Published online: 2 March 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The rapid adoption of microbial whole genome sequencing in public health, clinical testing, and forensic laboratories requires the use of validated measurement processes. Well-characterized, homogeneous, and stable microbial genomic reference materials can be used to evaluate measurement processes, improving confidence in microbial whole genome sequencing results. We have developed a reproducible and transparent bioinformatics tool, PEPR, Pipelines for Evaluating Prokaryotic References, for characterizing the reference genome of prokaryotic genomic materials. PEPR evaluates the quality, purity, and homogeneity of the reference material genome, and purity of the genomic material. The quality of the genome is evaluated using high coverage paired-end sequence data; coverage, paired-end read size and direction, as well as soft-clipping rates, are used to identify mis-assemblies. The homogeneity and purity of the material relative to the reference genome are characterized by comparing base calls from replicate datasets generated using multiple sequencing technologies. Genomic purity of the material is assessed by checking for DNA contaminants. We demonstrate the tool and its output using sequencing data while developing a *Staphylococcus aureus* candidate genomic reference material. PEPR is open source and available at https://github.com/usnistgov/pepr.

Keywords Microbiology · Whole genome sequencing · Bioinformatics

Published in the topical collection featuring *Young Investigators in Analytical and Bioanalytical Science* with guest editors S. Daunert, A. Baeumner, S. Deo, J. Ruiz Encinar, and L. Zhang.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-015-9299-5) contains supplementary material, which is available to authorized users.

✉ Nathan D. Olson
nolson@nist.gov

[1] Biosystems and Biomaterials Division, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

[2] Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

[3] Department of Bioengineering, Stanford University, Stanford, CA, USA

Introduction

Over the past decade, the availability of affordable and rapid Next-Generation Sequencing (NGS) technology has revolutionized the field of microbiology. Arguably the most discriminatory typing method available, whole genome sequencing (WGS), has been adopted by the research community, as well as public health laboratories, clinical testing laboratories, and the forensic community. High stakes decisions are often made based on the outcome of a WGS assay. To increase confidence in WGS assay, results a critical assessment of the errors inherent to the measurement processes is required. A number of sources of error associated with the WGS measurement process have been identified, but the degree to which they can be predicted, controlled, or compensated varies significantly [1].

Well-characterized, homogeneous, and stable genomic materials can be used to evaluate methods and aid in establishing confidence in results from a measurement process. For example, we recently characterized a whole human genome reference material (National Institute of Standards

YM95NC Attachment -- Page 81

and Technology, NIST, Reference Material 8398) to assess the performance of variant calling in human genomic samples [2], but no reference materials for microbial sequencing currently exist. NIST is developing four microbial genomic DNA candidate reference materials to meet this need. When considering the extensive genomic diversity of prokaryotic organisms as well as the rapidly evolving and diverse DNA sequencing applications, we envision the need for a wide variety of application-specific genomic materials for use in method validation and benchmarking. Currently, many laboratories and sequencing centers are using in-house materials as part of a regular method validation and quality control system. However, the degree to which these materials are characterized varies significantly, and, therefore, a common pipeline for characterizing prokaryotic genomic materials is needed.

PEPR, Pipelines for Evaluating Prokaryotic References, a set of reproducible and transparent bioinformatic pipelines, was developed to characterize genomic materials for use in WGS method validation. Using the pipeline increases confidence in method validation through the ability to develop better-characterized control materials. PEPR characterizes prokaryotic genomic material for purity and homogeneity of the genome sequence, as well as the presence of genomic material other than the material genus. The general approach to material characterization that guided the development of PEPR is the use of orthogonal sequencing methods along with technical replicates to obtain consensus values for the characterized properties. These consensus values are our best current estimates of the true values. We do not assert probabilistic estimates of confidence or confidence classification values with the sequence data, as we lack good models of biases or systematic errors of current sequencing technologies. Here we will first describe PEPR then show how PEPR was used to characterize NIST *Staphylococcus aureus* genomic DNA candidate reference material.

Methods

Pipelines for evaluating prokaryotic references: PEPR

PEPR consists of three bioinformatic pipelines written in Python (Fig. 1). The three bioinformatic pipelines are genome evaluation, genome characterization, and genomic purity. A YAML file (http://yaml.org) is used to define pipeline inputs. The pipeline coordinates the execution of a number of command line tools, logging the standard output and standard error for each executed command in time-stamped files for reference and debugging. Pipeline code is available at (https://github.com/usnistgov/pepr). To reduce the barrier for reuse, two Docker (https://www.docker.com/) containers are available with pre-installed pipeline dependencies. Docker is a lightweight virtual environment that facilitates the sharing and distribution of computing environments and can be run on any desktop, cloud, or high-performance computing environment, regardless of the operating system. The `pepr` container (https://registry.hub.docker.com/u/natedolson/pepr) includes dependencies for the genome evaluation and characterization pipelines, excluding the Genome Analysis Toolkit (due to licensing restrictions). The `docker-pathoscope` container has dependencies for the genomic purity pipeline installed (https://registry.hub.docker.com/u/natedolson/docker-pathoscope/).

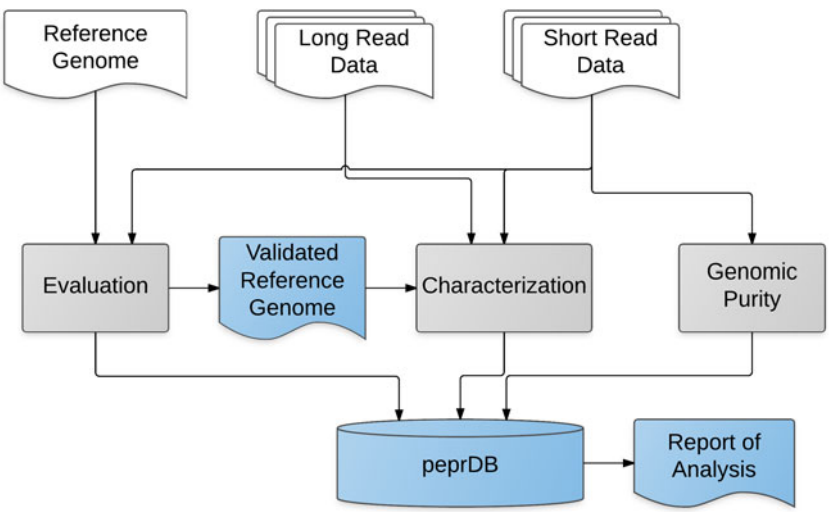


Fig. 1 PEPR workflow. *White objects* are pipeline inputs, *grey objects* are the three pipeline components, and *light blue* objects are the pipeline products

YM95NC Attachment -- Page 82

A software package, `peprr`, was developed for the statistical computing language R [3] to compile the output from the genome evaluation, characterization, and genomic purity pipelines. The compiled data was formatted into a series of data tables within an SQLite, `peprDB`, database to facilitate downstream analysis [4]. The package includes functions to generate a number of summary tables and figures, including those in this publication.

Genome evaluation pipeline

The Genome Evaluation Pipeline is the first step in the PEPR workflow and is used to reduce errors in the user-provided genome assembly prior to characterization. The `evaluation` pipeline consists of three steps. Illumina sequencing data are retrieved from the Genbank Sequence Read Archive (SRA) using the sratoolkit `fastq-dump` command (http://ncbi.github.io/sra-tools/). Users can also run the pipeline using fastq files by including file paths in the pipeline parameters file. Next, sequencing reads are mapped to the reference genome using BWA mem algorithm [5]. Finally, Pilon is used to evaluate and polish the reference assembly [6]. The corrected reference genome is then used as input for the Genome Characterization Pipeline.

Genome characterization pipeline

The Genome Characterization Pipeline uses replicate sequence dataset from multiple sequencing platforms to characterize the corrected reference genome produced by the Genome Evaluation Pipeline at the individual base level. Illumina data are aligned to the reference genome using the same methods as the `evaluation` pipeline. Ion Torrent PGM data are mapped to the reference using the TMAP algorithm [7]. If Pacific Biosciences (PacBio) sequencing data are used to generate the input reference assembly, the data are mapped to the reference genome using the BWA mem algorithm [5]. Sequence alignment files are processed prior to downstream analysis by marking duplicates with Picard's MarkDuplicates command (http://broadinstitute.github.io/picard) and realigning reads mapping to regions with insertions or deletion using the GenomeAnalysisToolKit [8, 9]. After refining the alignment files, base level analysis is performed using the short-read sequencing data. For each platform a VCF (variant call format) file with a number of summary statistics is generated using SAMtools mpileup [10]. A base purity metric is calculated from the resulting VCF files. The base purity metric is the number of high-quality bases (quality score $\geq$ 20) in reads aligned to a genome position that are in agreement with the reference base divided by the total number of reads high-quality bases supporting the reference and alternate base called by SAMtools. The metric is calculated from the SAMtools DP4 INFO tag in the vcf output generated with the mpileup command. Homogeneity analysis, a measure of genomic content similarity between vials of the reference material, is performed by first generating a pileup file using SAMtools mpileup for each dataset then performing pairwise tumor-normal variant calling using VarScan [11]. In this work, VarScan looks specifically for differences between vials in the proportion of reads containing variants. A standard Benjamini-Hochberg procedure was used to assess the power of the homogeneity analysis (Electronic Supplemental Material, https://github.com/DanSBS/NGSPower). Additionally, a number of summary statistics are calculated for the sequencing datasets using Picard's Collect Multiple Metrics (http://broadinstitute.github.io/picard).

Genome purity pipeline

The Genomic Purity Pipeline assesses the purity of the genomic material, defined as the presence of DNA from sources other than the expected genus. Material genomic purity was assessed using the metagenomic taxonomic read classification algorithm PathoScope 2.0 [12]. This method uses an expectation-maximization algorithm where the sequence data are first mapped to a database comprised of all sequence data in the Genbank nt database. Then, through an iterative process, PathoScope re-assigns ambiguously mapped reads to a taxonomic group based on the proportion of reads mapped unambiguously to individual taxonomic groups in the database. Using short-read sequencing data as input, PathoScope 2.0 first filters and trims low-quality reads (PathoQC), followed by mapping reads to a reference database (PathoMap - a wrapper for bowtie2 [13]), and then the expectation-maximization algorithm (PathoID) is used for the taxonomic classification. The annotated Genbank nt database provided by the PathoScope developers was used as the reference database (ftp://pathoscope.bumc.bu.edu/data/nt_ti.fa.gz).

Candidate reference material *S. aureus* sequencing data

Sequencing data and the reference assembly for the NIST candidate reference material *S. aureus* was used to demonstrate how PEPR is used to characterize a genomic material. A *de novo* genome assembly from Pacific Biosciences (PacBio) long-read sequencing data was used as input for PEPR. Prior to being used as input, the assembly was validated using optical mapping data. Eight replicate vials of the candidate reference material were sequenced on the Illumina MiSeq and Ion Torrent PGM sequencing platforms (Electronic Supplemental Material).

Results and discussion

Pipelines for Evaluating Prokaryotic References (PEPR) uses biological and technical replicate sequencing data from orthogonal sequencing platforms to characterize the reference genome of a prokaryotic material. The prokaryotic material is a batch of genomic DNA extracted from a prokaryotic culture. There are two primary reasons for using replicate sequencing datasets. One is to test for homogeneity within the batch of DNA. The second is to minimize the impact of library specific biases. The resulting characterized genome is suitable for evaluating and benchmarking whole genome sequencing methods. PEPR consists of three pipelines: genome evaluation, genome characterization, and genomic purity assessment (Fig. 1). The following section includes the characterization results for the NIST *S. aureus* candidate reference material along with a discussion of PEPR's assumptions and limitations.

Preparation of reference assembly

A user-provided high-quality closed reference assembly free of large assembly errors should be used as input when running PEPR. Optical mapping, as well as large insert mate-pair and synthetic long-read library preparation methods [14], are a few orthogonal methods that can be used to identify large mis-assemblies. The long DNA fragments used in optical mapping (average size < 200 Mb) allows for the identification of large mis-assemblies (< 3 kb)

Table 1 Summary of sequencing datasets

Acc.	Plat	Vial	Lib.	Reads	Length (bp)	Insert (bp)	Cov.
SRR1979039	miseq	0	1	3305082	230	257	247
SRR1979040	miseq	0	2	3732088	216	233	263
SRR1979041	miseq	1	1	3973320	218	242	279
SRR1979042	miseq	1	2	3941040	223	247	285
SRR1979043	miseq	2	1	3442554	234	268	261
SRR1979070	miseq	2	2	3226726	232	268	240
SRR1979044	miseq	3	1	3025028	233	264	229
SRR1979045	miseq	3	2	4796382	200	210	303
SRR1979046	miseq	4	1	3338456	239	278	260
SRR1979047	miseq	4	2	2995090	237	277	231
SRR1979048	miseq	5	1	3495384	225	255	255
SRR1979049	miseq	5	2	3116128	241	281	244
SRR1979050	miseq	6	1	3129282	237	271	240
SRR1979060	miseq	6	2	2976312	242	280	233
SRR1979064	miseq	7	1	2630544	241	283	204
SRR1979065	miseq	7	2	3416580	225	248	247
SRR2002412	pgm	0	1	556903	231		42
SRR2002413	pgm	1	1	530117	224		38
SRR2002414	pgm	2	1	437527	231		33
SRR2002415	pgm	3	1	552692	232		42
SRR2002416	pgm	4	1	498479	232		37
SRR2002418	pgm	5	1	390070	235		30
SRR2002419	pgm	6	1	426196	232		32
SRR2002420	pgm	7	1	439119	238		34
SRR2056302	pacbio	9	1	163475	10510		108
SRR2056306	pacbio	9	2	163471	10436		103
SRR2056310	pacbio	9	3	163474	9863		91

Acc. - Sequence read archive (SRA) database accessions. Plat. - sequencing platform, miseq: Illumina MiSeq, pgm: Ion Torrent PGM, pacbio: Pacific Biosciences RSII. Lib. - library replicate number for miseq and pgm, smartcell replicate for pacbio. Reads - number of sequencing reads in the dataset. Length - median read length in base pairs. Insert - median insert size in base pairs for paired-end reads. Cov. - median sequence coverage across the genome

YM95NC Attachment -- Page 84

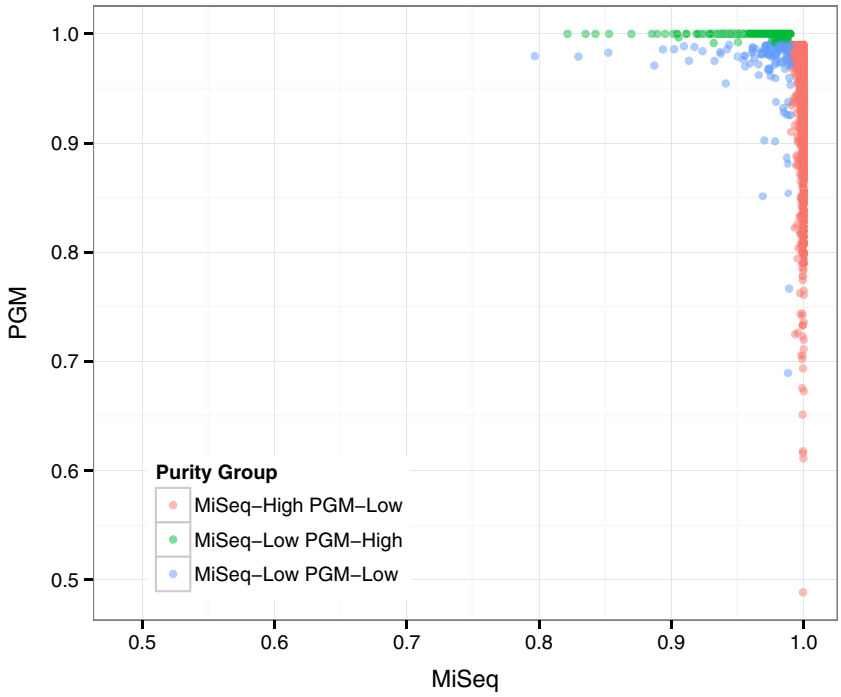


Fig. 2 Comparison of base purity values for PGM and MiSeq. Positions are colored based of high and low purity values for the two sequencing platforms, MiSeq - Illumina MiSeq and PGM - Ion Torrent PGM. A purity value of 0.99 was used to differentiate between high and low purity positions. Positions with high purity for both platforms were excluded from the figure

that are not easily identified using short-read sequencing data [15]. For the *S. aureus* RM, the reference assembly was constructed from Pacific Biosciences (PacBio) RSII long-read sequencing data (Electronic Supplemental Material). To validate the *S. aureus* reference assembly prior to evaluation with PEPR, OpGen optical mapping technology was used. The *S. aureus* optical mapping results were in agreement with the PacBio assembly, indicating no large mis-assemblies, and that the assembly was suitable for use in PEPR.

Genome evaluation

The *Genome Evaluation* pipeline is the first step in PEPR. Pilon was chosen for the evaluation step as it assesses the accuracy of the genome and corrects errors in the assembly [6]. While currently not part of the PEPR, PAGIT, and REAPR are two alternative methods for evaluating and correcting genome assemblies similar to Pilon [16, 17]. Other methods are available for evaluating reference genomes, e.g. amosValidate [18] and ALE [19], however, these methods only assess assembly accuracy without correcting misassemblies. The resulting reference assembly represents the consensus genome of the population of cells used to generate the material. The genome evaluation pipeline does not attempt to identify or characterize low-frequency structural variants within the material or vial-to-vial variability of the reference genome. The evaluation pipeline failed to identify any assembly or base call errors in the *S. aureus* candidate genome assembly. The candidate genome, once evaluated, and, if necessary, refined during the *Genome Evaluation* pipeline, is used as input for the *Genome Characterization* pipeline.

Genome characterization

The *Genome Characterization* pipeline calculates whole genome base level statistics using replicate sequencing data from orthogonal measurement methods. As part of the characterization pipeline, summary statistics are generated for the user provided sequencing datasets. While PEPR only uses short-read sequencing data to evaluate the material, dataset summary statistics for additional sequence data, such as long-read data used to generate the reference assembly, can be calculated. The results from our analysis of the *S. aureus* candidate reference material using PEPR provides an example of the type of information, summary figures, and tables that can be generated with PEPR.

Table 2 Number of genome positions with high and low purity, purity metric values higher and lower than 0.99 respectively, for the Illumina MiSeq and Ion Torrent PGM sequencing platforms

	PGM-High	PGM-Low
MiSeq-High	2864925	44534
MiSeq-Low	394	115

YM95NC Attachment -- Page 85

Sequencing data summary statistics

Summary statistics were calculated including number of reads, mapped read length, insert size for paired-end datasets, and coverage for *S. aureus* datasets (Table 1). The MiSeq sequencing run had an average of 1.7 million paired-end reads per library with a median read length of 232 bp, whereas the PGM sequencing run produced 0.2 million reads per library on average with a median read length of 232 bp. Based on the sequencing methods used (Electronic Supplemental Material), longer reads were expected for PGM. The shorter read length is potentially due to the low GC content, which is known to challenge current sequencing technologies [20]. The higher throughput and paired-end reads resulted in higher coverage for MiSeq compared to PGM (251 X vs. 36 X). The three PacBio datasets are library replicates run on different SMRT cells. The replicate libraries had a median subread length 10,436 bp and 302 X total coverage. Between the three platforms a total coverage of 4,611 X. The dataset summary statistics provide general information about the sequencing datasets and identify potential biases in the sequence methods.

Base level purity

A base purity metric was used to evaluate how well the sequencing data supports the reference base call. Through comparison of the base purity for two orthogonal sequencing methods, we identified genome positions with low purity values due to platform specific systematic sequencing errors. The reference base is identified using a third orthogonal sequencing method (Pacific Biosciences RSII), which only chooses the dominant base and does not identify small impurities. Thus, a low purity (below 50 %) for one of the two short-read sequencing platforms and a high purity value for the other means that two technologies (one short-read and one long-read) agree that the dominant base is the reference base. It is important to acknowledge that even if the two short-read sequencing platforms indicate an impurity, they are potentially susceptible to the same unknown bias.

We compared purity metric values between two orthogonal sequencing methods, MiSeq and PGM, for all positions in the genome (Fig. 2). The purity metric was used to categorize genomic positions as high (> 0.99) or low purity (< 0.99). Out of 2,909,968 positions in the genome 2,864,925 positions had purity values greater than 0.99 for both short-read sequencing platforms (Table 2). Further, 2,909,853 and 2,909,965 positions had purity values greater than 0.99 and 0.97, respectively, for one of the two platforms. Only 115 positions had purity values less than 0.99 for both platforms, and no positions had purity value less than 0.95 for both platforms. The positions with low purity for MiSeq were non-uniformly distributed whereas

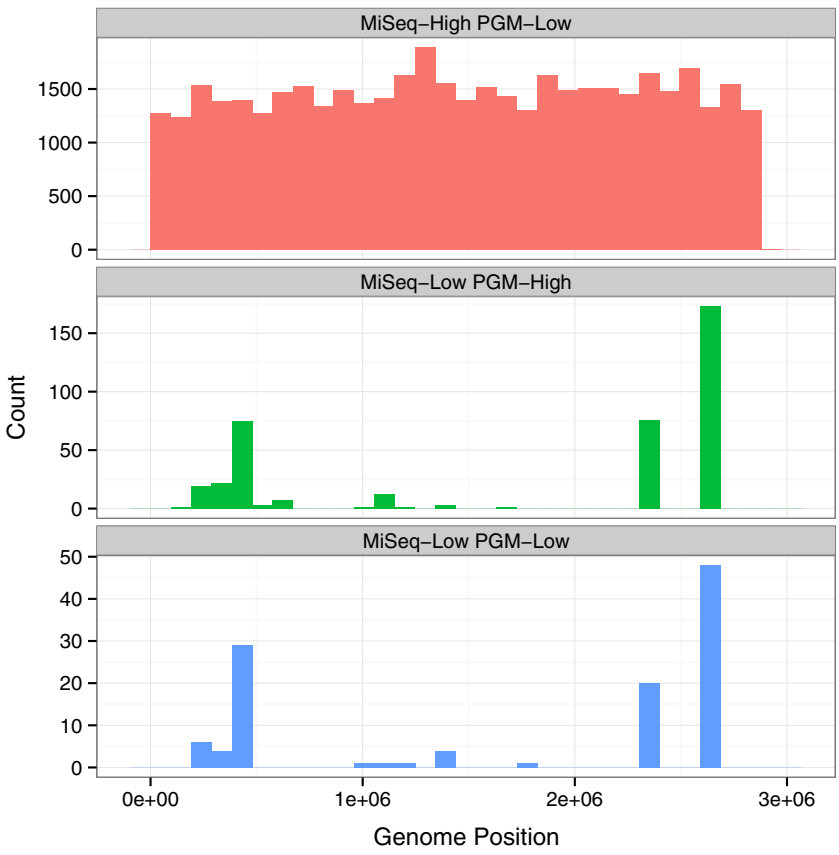


Fig. 3 Distribution of genome positions by purity group. Bases with high and low purity and purity values greater than and less than 0.99 respectively for the two platforms, MiSeq - Illumina MiSeq and PGM - Ion Torrent PGM. Positions with high purity for both platforms were excluded from the figure

positions with low purity for PGM were uniformly distributed (Fig. 3). The difference in low purity position distributions is due to differences in the systematic sequencing error profiles for the two platforms. PGM has a higher error rate for homopolymers, whereas MiSeq has a more context specific sequencing error profile [20].

The sequencing technologies used to characterize the material are still maturing, and an incomplete understanding of platform-specific biases limits our ability to provide a confidence value for the base calls. A number of base level metrics, such as strand bias, are calculated as part of the PEPR Genome Characterization Pipeline and are included in the pipeline results database. These metrics can be used to differentiate positions with low purity due to measurement error and those due to biological variability. Use of additional metrics and algorithms developed for the identification of low-frequency variants, such as LoFreq [21], could help identify positions with low levels of biological variability, but are not currently implemented in PEPR.

Base level homogeneity

Material homogeneity was assessed through pairwise statistical analysis of the replicate MiSeq datasets using the VarScan somatic variant caller [11]. The pairwise variant analysis failed to identify any statistically significant base level differences among the replicates (Table 3). Only Illumina data was used to assess the homogeneity of the material as the higher coverage increased the statistical power of the test, and replicate libraries provide information regarding the method error rate. The Ion Torrent dataset did not include replicate libraries for the eight vials sequenced, and therefore, library specific sequencing errors were confounded with vial-to-vial variability. No statistically significant variants were identified between all pairwise comparisons indicating that the material is homogeneous. If potential inhomogeneities were found, then the PGM sequencing data could be examined for additional support for the inhomogeneities. Even without replicate libraries for the different vials, it is unlikely that any library specific bias will correlate with vial-to-vial variability observed in the Illumina data by chance.

Genomic purity pipeline

The *Genomic Purity* pipeline is used to identify DNA within the material that belongs to a genus other than the material genus. Short-read sequencing data was used to identify the proportion of DNA in the material from an organism other than the material genus, in this case, *Staphylococcus*, using PathoScope 2.0 [12]. The genus level cutoff was selected based on results from a previous study characterizing the specificity of the PathoScope 2.0 classification algorithm (Olson et al. *in-prep*). Genomic contaminants can be from the culture itself or reagents and materials used to prepare the material or during sequencing [22–24]. Contaminants identified by the Genomic Purity Pipeline may not be present in the material. For example, reagents used during library preparation may include contaminants [24–27]. Additionally, bioinformatic errors may lead to false positives, either due to errors in the database or errors by the classification algorithm.

Based on analysis of the MiSeq and PGM sequencing data, the reference material has minimal if any genomic contaminants, with a maximum of 0.0039 % reads in any dataset classified as not belonging to the genus *Staphylococcus*. The most abundant contaminant was *Escherichia coli* (Fig. 4). *E. coli* is a well-documented contaminant of molecular biology reagents, and not likely a true contaminant [24]. Lower abundant contaminants may be bioinformatic errors and not true contaminants. While, contaminants identified by the *Genomic Purity* pipeline are most likely from reagents or due to bioinformatic errors, a conservative estimate of the material purity, assuming all contaminants are real, is reported by the pipeline. Users will want to

Table 3 Pairwise variant analysis results

Position	Proportion of Pairs	Median Frequency	Minimum P-value	N Significant
244332	0.01	21.31	0.51	0.00
2615986	0.03	20.48	0.45	0.00
2616058	0.08	25.29	0.15	0.00
2619808	0.01	20.78	0.61	0.00
2619886	0.01	21.54	0.50	0.00

Position is the position in the genome where differences in variant frequency for at least one of the 16 pairwise comparisons were reported. Proportion of pairs is the fraction of the pairwise comparisons between the 16 Illumina MiSeq datasets where VarScan reported a difference in variant frequency. Median frequency is the median variant frequency for datasets with reported difference at that genome position. Minimum p-value is the lowest p-value reported by VarScan for all pairwise dataset comparisons with reported differences in variant frequency. N Significant is the number of datasets with reported statistically significant differences at that genome position

YM95NC Attachment -- Page 87

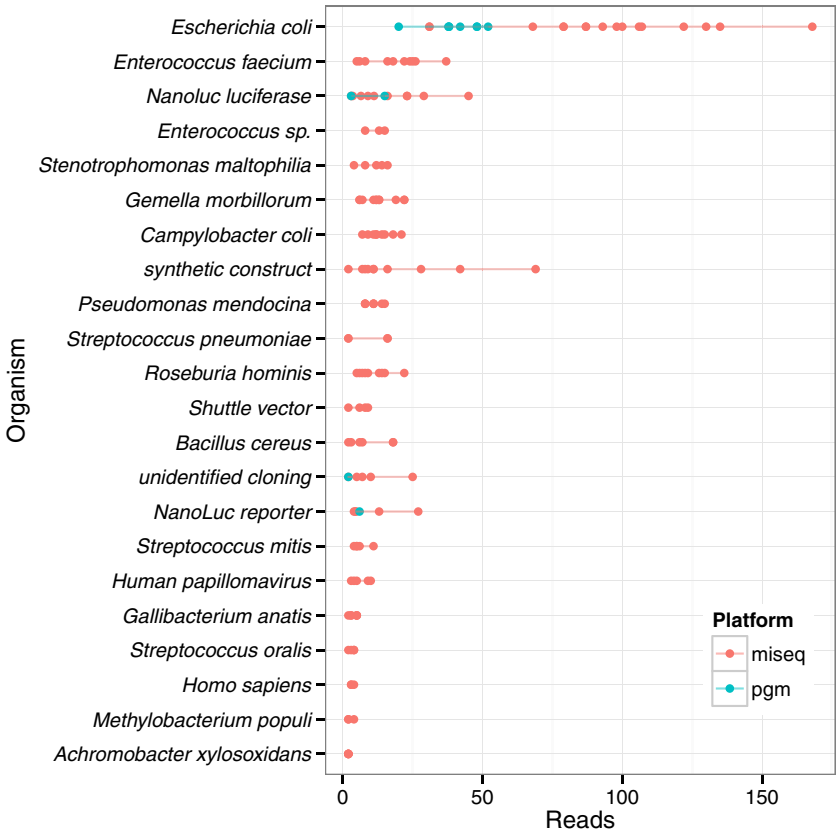


Fig. 4 Breakdown of contaminants by organism

consider the limited specificity of the taxonomic classification method. For example, if the intended use of the genomic DNA is for use as part of an inclusivity exclusivity panel, additional genomic purity assessment in addition to the PEPR Genomic Purity Pipeline is required to validate the material.

Conclusions

PEPR provides a framework for characterizing microbial genomic reference materials, for instance, a homogenized batch of DNA from a single prokaryotic strain. The objective in developing PEPR was to provide a reproducible and transparent workflow for characterizing of prokaryotic genomic materials. The pipeline can be used to characterize reference materials as well as in-house quality control materials for which replicate sequencing datasets from multiple platforms are available. If another lab wishes to characterize a new reference or quality control material, they could follow this process:

1. Ideally, the user should generate a large batch of material and aliquot it to reduce inhomogeneity.
2. Identify a high-quality genome assembly. If a good reference assembly does not exist for the sample, then long-read sequencing like PacBio may be required to generate an assembly, and ideally mapping technologies would be used to validate the assembly.
3. Short-read whole genome sequencing, preferably from two orthogonal sequencing technologies, should be generated from multiple vials of the material, ideally with technical replicate libraries from at least six randomly selected vials.
4. PEPR can then be run to assess base level purity and homogeneity, genomic contaminants, and misassemblies.

PEPR outputs include a corrected reference genome assembly, genome positions with high and low purity based on biological and technical variation, base level homogeneity of the material, as well as the percentage and identity of genus level genomic contaminants. The resulting characterization values are intentionally conservative and without uncertainty or confidence estimates, as sources of bias and error associated with the measurement process are currently not fully understood. As the scientific community's understanding of the measurement process matures new algorithms can be incorporated into the pipeline to increase the quality of material characterization process. The genomic materials characterized using PEPR will help increase confidence in WGS measurement methods and improve our understanding of the sequencing and data analysis process.

Acknowledgments The authors would like to thank Jenny McDaniel, Lindsay Vang, and David Catoe for performing the MiSeq and PGM sequencing, and Tim Muruvanda for performing the PacBio sequencing. Additionally, we would like to thank Drs. Nancy Lin and Steven Lund for conversations and suggestions while developing PEPR and drafting the manuscript. Department of Homeland Security (DHS) Science and Technology Directorate funded this project under Interagency Agreement HSHQPM-14-X-00078 with NIST. This work was additionally supported by two interagency agreements between NIST and the FDA. Opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of the DHS, NIST, or affiliated venues. Certain commercial equipment, instruments, or materials are identified in this paper only to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose. Official contribution of NIST; not subject to copyrights in USA.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Compliance with ethical standards

Conflict of interests The authors declare no competing interests.

References

1. Olson N, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et. al. Front Genet. 2015;6. doi:10.3389/fgene.2015.00235
2. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, et. al. Nat Biotechnol. 2014;32:246–5
3. R Core Team. R: A language and environment for statistical computing. 2015 Vienna, Austria
4. Wickham H. J Stat Softw. 2014;59:10
5. Li H. 2013 arXiv preprint arXiv:1303.3997
6. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et. al. PLoS ONE. 2014;9:e112963–11. doi:10.1371/journal.pone.0112963
7. Homer N, Tmap MB the torrent mapping alignment program. 2015 https://github.com/iontorrent/TMAP
8. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et. al. Genome Res. 2010;20:1297–9. doi:10.1101/gr.107524.110
9. DePristo M, Banks E, Poplin R, Garimella1 K, Maguire J, Hartl C, et. al. Nat Genet. 2011;43:491–5. doi:10.1038/ng.806.A
10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et. al. Bioinformatics (Oxford England). 2009;25:2078–16. doi:10.1093/bioinformatics/btp352
11. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et. al. Bioinformatics (Oxford, England). 2009;25:2283–17. doi:10.1093/bioinformatics/btp373
12. Hong C, Manimaran S, Shen Y, Perez-Rogers JF, Byrd AL, Castro-Nallar E, et. al. Microbiome. 2014;2:33–1. doi:10.1186/2049-2618-2-33
13. Langmead B, Salzberg SL. Nat Methods. 2012;9:357–4. doi:10.1038/nmeth.1923
14. McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, Pushkarev D, et. al. 2014
15. Mendelowitz L, Pop M. 2014;3:33–1. 10.1186/2047-217X-3-33
16. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. Nat. Protocols. 2012;7:1260–7
17. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. Genome Biol. 2013;14:R47–5. doi:10.1186/gb-2013-14-5-r47
18. Phillippy AM, Schatz MC, Pop M. Genome Biol. 2008;9:R55–3. doi:10.1186/gb-2008-9-3-r55
19. Clark SC, Egan R, Frazier PI, Wang Z. Bioinformatics (Oxford, England). 2013;29:435–4. doi:10.1093/bioinformatics/bts723
20. Quail M, Smith ME, Coupland P, Otto TD, Harris SR, Connor TR, et. al. BMC Genomics. 2012;13:341–1. doi:10.1186/1471-2164-13-341
21. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et. al. Nucleic Acids Res. 2012;40:11189–22. doi:10.1093/nar/gks918
22. Shrestha PM, Nevin KP, Shrestha M, Lovley DR. MBio. 2013;4:e00591–2
23. Tang JS Int J Syst Evol Microbiol. 2003;53:995–4. doi:10.1099/ijs.0.02372-0
24. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et. al. BMC Biol. 2014;12:87–1. doi:10.1186/s12915-014-0087-z
25. Tanner MA, Goebel BM, Dojka MA, Pace NR. Appl Environ Microbiol. 1998;64:3110–8
26. Newsome T, Li Bj, Zou N, Lo Sc. 2004;42:2264–5. doi:10.1128/JCM.42.5.2264
27. Motley S, Picuri J, Crowder C, Minich J, Hofstadler S, Eshoo M. BMC Genomics. 2014;15:443–1. doi:10.1186/1471-2164-15-443

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Coming of age: ten years of next-generation sequencing technologies

Sara Goodwin[1], John D. McPherson[2] and W. Richard McCombie[1]

Abstract | Since the completion of the human genome project in 2003, extraordinary progress has been made in genome sequencing technologies, which has led to a decreased cost per megabase and an increase in the number and diversity of sequenced genomes. An astonishing complexity of genome architecture has been revealed, bringing these sequencing technologies to even greater advancements. Some approaches maximize the number of bases sequenced in the least amount of time, generating a wealth of data that can be used to understand increasingly complex phenotypes. Alternatively, other approaches now aim to sequence longer contiguous pieces of DNA, which are essential for resolving structurally complex regions. These and other strategies are providing researchers and clinicians a variety of tools to probe genomes in greater depth, leading to an enhanced understanding of how genome sequence variants underlie phenotype and disease.

Read
The sequence of bases from a single molecule of DNA.

Sanger sequencing
An approach in which dye-labelled normal deoxynucleotides (dNTPs) and dideoxy-modified dNTPs are mixed. A standard PCR reaction is carried out and, as elongation occurs, some strands incorporate a dideoxy-dNTP, thus terminating elongation. The strands are then separated on a gel and the terminal base label of each strand is identified by laser excitation and spectral emission analysis.

[1]*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.*
[2]*Department of Biochemistry and Molecular Medicine; and the Comprehensive Cancer Center, University of California, Davis, California 95817, USA.*
Correspondence to W.R.M.
mccombie@cshl.edu

doi:10.1038/nrg.2016.49
Published online 17 May 2016

Starting with the discovery of the structure of DNA[1], great strides have been made in understanding the complexity and diversity of genomes in health and disease. A multitude of innovations in reagents and instrumentation supported the initiation of the Human Genome Project[2]. Its completion revealed the need for greater and more advanced technologies and data sets to answer the complex biological questions that arose; however, limited throughput and the high costs of sequencing remained major barriers. The release of the first truly high-throughput sequencing platform in the mid-2000s heralded a 50,000-fold drop in the cost of human genome sequencing since the Human Genome Project[3] and led to the moniker: next-generation sequencing (NGS). Over the past decade, NGS technologies have continued to evolve — increasing capacity by a factor of 100–1,000 (REF. 4) — and have incorporated revolutionary innovations to tackle the complexities of genomes. These advances are providing read lengths as long as some entire genomes, they have brought the cost of sequencing a human genome down to around US$1,000 (as reported by Veritas Genomics)[5], and they have enabled the use of sequencing as a clinical tool[3,6].

Although exciting, these advancements are not without limitations. As new technologies emerge, existing problems are exacerbated or new problems arise. NGS platforms provide vast quantities of data, but the associated error rates (~0.1–15%) are higher and the read lengths generally shorter (35–700 bp for short-read approaches)[7] than those of traditional Sanger sequencing platforms, requiring careful examination of the results, particularly for variant discovery and clinical applications. Although long-read sequencing overcomes the length limitation of other NGS approaches, it remains considerably more expensive and has lower throughput than other platforms, limiting the widespread adoption of this technology in favour of less-expensive approaches. Finally, NGS is also competing with alternative technologies that can carry out similar tasks, often at lower cost (BOX 1); it is not clear how these disparate approaches to genomics, medicine and research will interact in the years to come.

This Review evaluates various approaches used in NGS and how recent advancements in the field are changing the way genetic research is carried out. Details of each approach along with its benefits and drawbacks are discussed. Finally, various emerging applications within this field and its exciting future are explored.

Short-read NGS

Overview of clonal template generation approaches. Short-read sequencing approaches fall under two broad categories: sequencing by ligation (SBL) and sequencing by synthesis (SBS). In SBL approaches, a probe sequence that is bound to a fluorophore hybridizes to a DNA fragment and is ligated to an adjacent oligonucleotide for imaging. The emission spectrum of the fluorophore indicates the identity of the base or bases complementary to specific positions within the probe. In SBS approaches, a polymerase is used and a signal, such as a fluorophore or a change in ionic concentration, identifies the incorporation of a nucleotide into

© 2016 Macmillan Publishers Limited. All rights reserved.

Box 1 | **Alternative genomic strategies**

There are other technologies that either compete with or complement next-generation sequencing (NGS). This section outlines these technologies and their relationship with NGS.

DNA microarrays

DNA microarrays have been used for genetic research since the early 1980s[122] (see the figure, part **a**). In DNA microarrays, single-stranded DNA (ssDNA) probes are immobilized on a substrate in a discrete location with spots as small as 50 μm[123]. Target DNA is labelled with a fluorophore and hybridized to the array. The intensity of the signal is used to determine the number of bound molecules.

Microarrays are used in many applications. Single-nucleotide polymorphism (SNP) arrays identify common polymorphisms associated with disease and phenotypes, including cardiovascular disease[124], cancer[125–127], pathogens[128,129], ethnicity[130,131] and genome-wide association study (GWAS) analysis[132,133]. Additionally, lower-resolution arrays are used to identify structural variation, copy number variants (CNVs) and DNA–protein interactions[134–137]. Expression arrays measure expression levels by measuring the amount of gene-specific cDNA[138].

Microarrays remain widely used in genomic research. They are used to identify SNPs at costs far below NGS routines. This is also true for expression studies, in which arrays inexpensively measure expression levels of thousands of genes. Variations in hybridization and normalization are problematic, leading some people to recommend RNA sequencing (RNA-seq) over gene expression microarrays[139].

NanoString

Similar to microarrays, the nCounter Analysis System from NanoString relies on target–probe hybridization (see the figure, part **b**). Probes target a gene of interest; one probe is bound to a fluorophore 'barcode' and the other anchors the target for imaging. The number and type of each barcode is counted. NanoString is unique in that the probes are labelled molecules that are bound together in a discrete order, which can be changed to create hundreds of different labels.

nCounter applications are similar to those of microarrays and quantitative PCR (qPCR; see below), including gene expression analysis[145,146], CNV and SNP detection[147,148], and fusion gene detection[149]. This approach provides exceptionally high resolution, less than one copy per cell[145], far below microarrays and approaching TaqMan in sensitivity. Unlike most NGS applications, neither template enrichment nor reverse transcription are required. Around 800 targets can be read at a time, far below either microarrays or NGS.

qPCR

Real-time qPCR utilizes the PCR reaction to detect targets of interest (see the figure, part **c**). Gene-specific primers are used and the target is detected either by the incorporation of a double-stranded DNA (dsDNA)-specific dye or by the release of a TaqMan FRET (fluorescence resonance energy transfer) probe through polymerase 5′–3′ exonuclease activity.

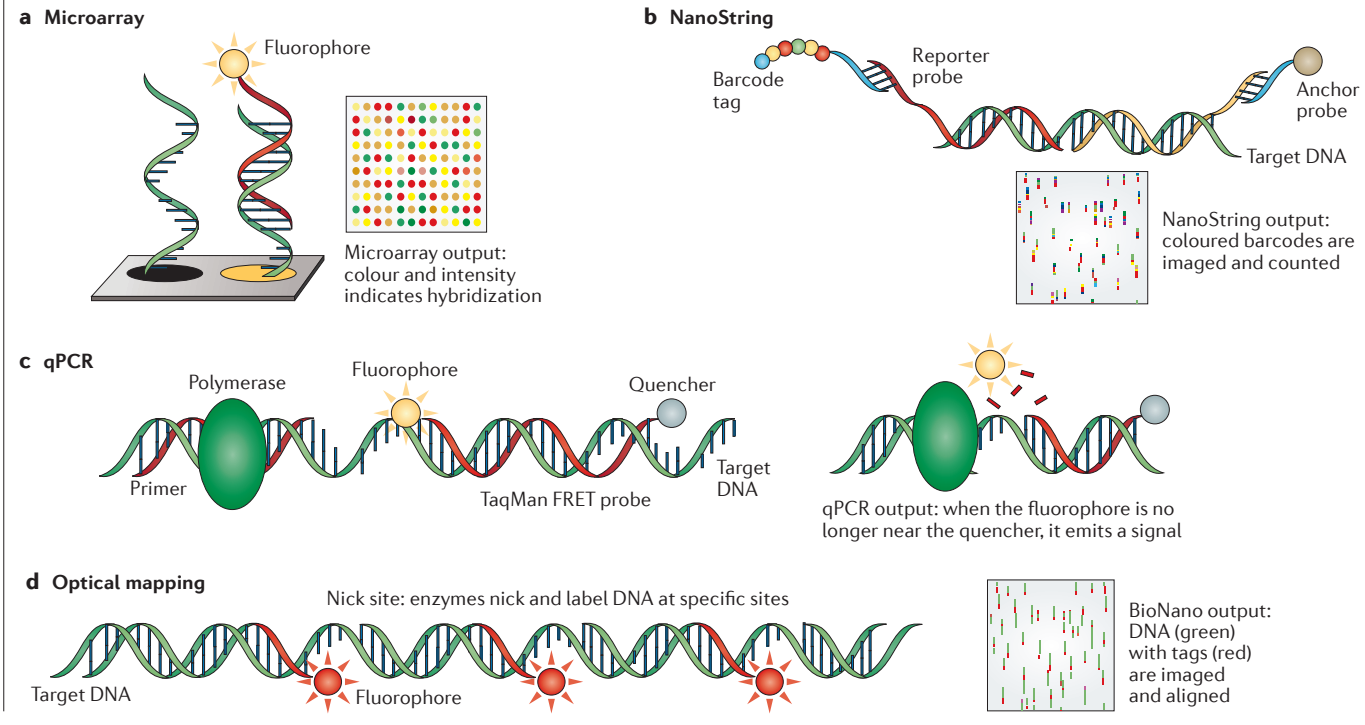
Developed in the early 1990s[140], qPCR is widely used in both clinical and research settings for genotyping[141], gene expression analysis[142], CNV assays[143] and pathogen detection[144]. qPCR is extremely rapid and robust, which is beneficial for point-of-care applications. Its high sensitivity and specificity make it the gold standard for clinical gene detection with several US Food and Drug Administration (FDA)-approved tests. The number of simultaneous targets that can be detected is in the hundreds rather than the thousands for microarrays and NGS. This method also requires primers and/or probes designed for specific targets.

Optical mapping

Optical mapping combines long-read technology with low-resolution sequencing (see the figure, part **d**). Originally a method for ordering restriction enzyme sites[150] through digestion and size separation, this technology now uses fluorescent markers to tag particular sequences within DNA fragments that are up to ~1 Mb long. The results are imaged and aligned to each other, and/or a reference, to map the locations of the probes relative to each other.

A central application of this technology is the generation of genome maps that are used in *de novo* assembly and gap filling[94,151]. This technology can be used to detect structural variations that are up to tens of kilobases in length[94,152]. Haplotype blocks that are several hundred kilobases in size can also be resolved[153].

Optical mapping can either be an alternative to NGS or a complementary approach. As an alternative, it provides a low-cost option for understanding structural and copy number variation, but it does not provide base-level resolution. As a complementary technology, optical mapping improves *de novo* genome assemblies by providing a long-range scaffold on which to align short-read data.



© 2016 Macmillan Publishers Limited. All rights reserved.

an elongating strand. In most SBL and SBS approaches, DNA is clonally amplified on a solid surface. Having many thousands of identical copies of a DNA fragment in a defined area ensures that the signal can be distinguished from background noise. Massive parallelization is also facilitated by the creation of many millions of individual SBL or SBS reaction centres, each with its own clonal DNA template. A sequencing platform can collect information from many millions of reaction centres simultaneously, thus sequencing many millions of DNA molecules in parallel.

There are several different strategies used to generate clonal template populations: bead-based, solid-state and DNA nanoball generation (FIG. 1). The first step of DNA template generation is fragmentation of the sample DNA, followed by ligation to a common adaptor set for clonal amplification and sequencing. For bead-based preparations, one adaptor is complementary to an oligonucleotide fragment that is immobilized on a bead (FIG. 1a). Using emulsion PCR (emPCR)[8], the DNA template is amplified such that as many as one million clonal DNA fragments are immobilized on a single bead[9]. These beads can be distributed onto a glass surface[10] or arrayed on a PicoTiterPlate (Roche Diagnostics)[11]. Solid-state amplification[12] eschews the use of emPCR in favour of amplification directly on a slide[13] (FIG. 1b,c). In this approach, forward and reverse primers are covalently bound to the slide surface, either randomly or on a patterned slide. These primers provide complementary ends to which single-stranded DNA (ssDNA) templates can bind. Precise control over template concentration enables the amplification of templates into localized, non-overlapping clonal clusters, thus maintaining spatial integrity. Recently, several NGS platforms have utilized patterned flow cells. By defining precisely where primers are bound to the slide, more DNA templates can be spatially resolved, enabling higher densities of reaction centre clusters and increasing sequencing throughput.

The Complete Genomics technology used by the Beijing Genomics Institute (BGI) is currently the only approach that achieves template enrichment in solution. In this case, DNA undergoes an iterative ligation, circularization and cleavage process to create a circular template, with four distinct adaptor regions. Through the process of rolling circle amplification (RCA), up to 20 billion discrete DNA nanoballs are generated (FIG. 1d). The nanoball mixture is then distributed onto a patterned slide surface containing features that allow a single nanoball to associate with each location[14].

Sequencing by ligation (SOLiD and Complete Genomics). Fundamentally, SBL approaches involve the hybridization and ligation[15] of labelled probe and anchor sequences to a DNA strand. The probes encode one or two known bases (one-base-encoded probes or two-base-encoded probes) and a series of degenerate or universal bases, driving complementary binding between the probe and template, whereas the anchor fragment encodes a known sequence that is complementary to an adapter sequence and provides a site to initiate ligation. After ligation, the template is imaged and the known base or bases in the probe are identified[16]. A new cycle begins after complete removal of the anchor–probe complex or through cleavage to remove the fluorophore and to regenerate the ligation site.

The SOLiD platform utilizes two-base-encoded probes, in which each fluorometric signal represents a dinucleotide[17]. Consequently, the raw output is not directly associated with the incorporation of a known nucleotide. Because the 16 possible dinucleotide combinations cannot be individually associated with spectrally resolvable fluorophores, four fluorescent signals are used, each representing a subset of four dinucleotide combinations. Thus, each ligation signal represents one of several possible dinucleotides, leading to the term colour-space (rather than base-space), which must be deconvoluted during data analysis. The SOLiD sequencing procedure is composed of a series of probe–anchor binding, ligation, imaging and cleavage cycles to elongate the complementary strand (FIG. 2a). Over the course of the cycles, single-nucleotide offsets are introduced to ensure every base in the template strand is sequenced.

Complete Genomics performs DNA sequencing using combinatorial probe–anchor ligation (cPAL)[14] or combinatorial probe–anchor synthesis (cPAS; see the BGISEQ-500 website). In cPAL (FIG. 2b), an anchor sequence (complementary to one of the four adaptor sequences) and a probe hybridize to a DNA nanoball at several locations. In each cycle, the hybridizing probe is a member of a pool of one-base-encoded probes, in which each probe contains a known base in a constant position and a corresponding fluorophore. After imaging, the entire probe–anchor complex is removed and a new probe–anchor combination is hybridized. Each subsequent cycle utilizes a probe set with the known base in the n + 1 position. Further cycles in the process also use adaptors of variable lengths and chemistries, allowing sequencing to occur upstream and downstream of the adaptor sequence. The cPAS approach is a modification of cPAL intended to increase read lengths of Complete Genomics' chemistry; however, at present, details about the approach are limited.

Sequencing-by-synthesis categories. SBS is a term used to describe numerous DNA-polymerase-dependent methods in the literature, but it does not delineate the different mechanisms involved in SBS approaches. For this article, SBS approaches will be classified either as cyclic reversible termination (CRT) or as single-nucleotide addition (SNA)[18].

Sequencing by synthesis: CRT (Illumina, Qiagen). CRT approaches are defined by their use of terminator molecules that are similar to those used in Sanger sequencing, in which the ribose 3′-OH group is blocked, thus preventing elongation[19,20]. To begin the process, a DNA template is primed by a sequence that is complementary to an adapter region, which will initiate polymerase binding to this double-stranded DNA (dsDNA) region. During each cycle, a mixture of all four individually labelled and 3′-blocked deoxynucleotides (dNTPs) are

Template
A DNA fragment to be sequenced. The DNA is typically ligated to one or more adapter sequences where DNA sequencing will be initiated.

Fragmentation
The process of breaking large DNA fragments into smaller fragments. This can be achieved mechanically (by passing the DNA through a narrow passage), by sonication or enzymatically.

Clusters
Groups of DNA templates in close spatial proximity, generated either though bead-based amplification or by solid-phase amplification. Bead-based approaches rely on emulsions to maintain template isolation during amplification. Solid-phase approaches rely on the template-to-bound-adapter ratio to probabilistically bind template molecules at a sufficient distance from each other.

Flow cells
Disposable parts of a next-generation sequencing routine. Template DNA is immobilized within the flow cell where fluid reagents can be streamed into the cell and flushed away.

Rolling circle amplification
(RCA). A method of DNA amplification using a circular template. Briefly, DNA polymerase binds to a primed section of a circular DNA template. As the polymerase traverses the template, a new strand is synthesized. When the polymerase completes a full circle and encounters the double-stranded DNA (dsDNA) template, it displaces the template without degradation, thus creating a long ssDNA fragment composed of many copies of the template sequence.

© 2016 Macmillan Publishers Limited. All rights reserved.

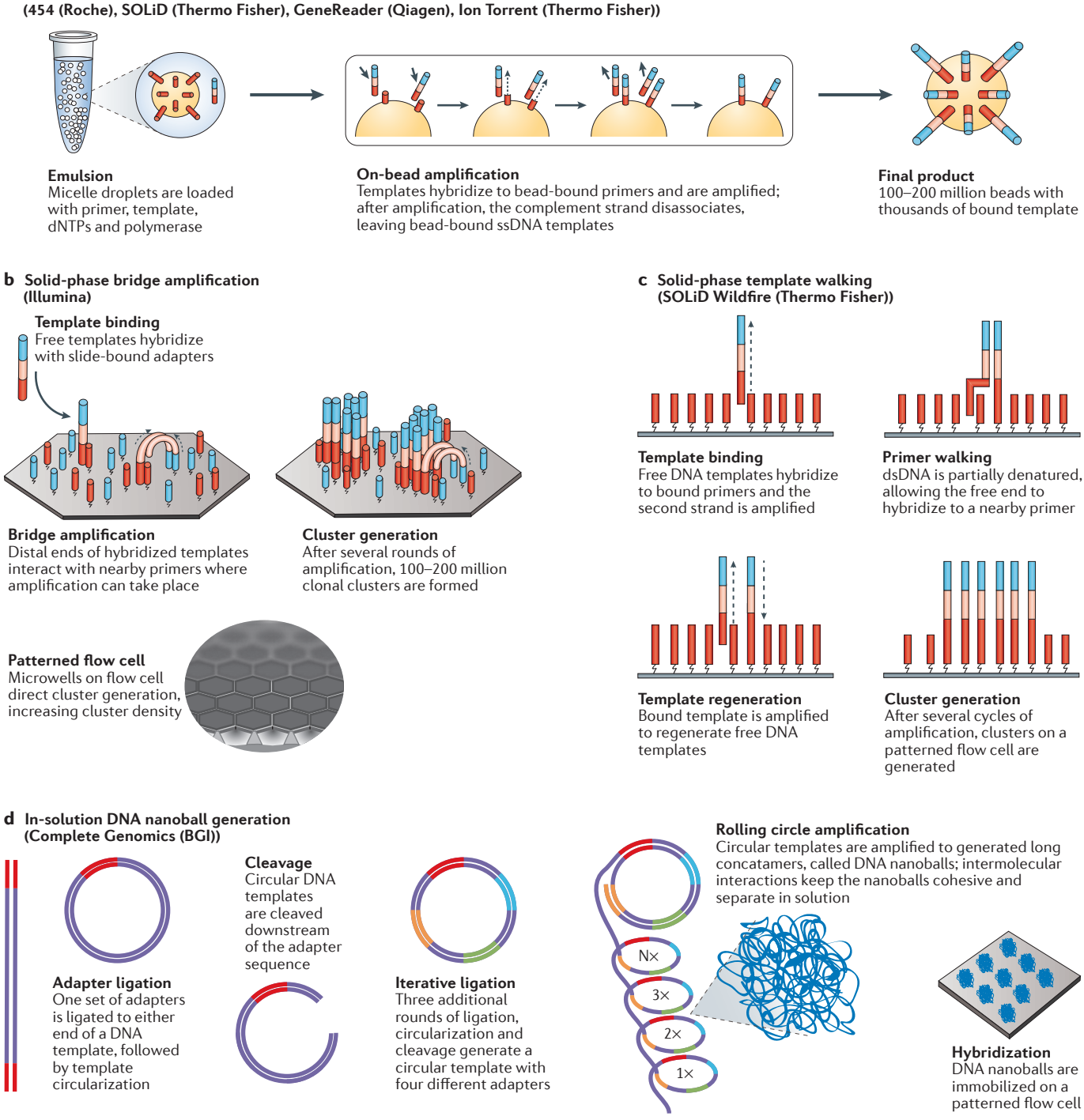


Figure 1 | **Template amplification strategies. a** | In emulsion PCR, fragmented DNA templates are ligated to adapter sequences and are captured in an aqueous droplet (micelle) along with a bead covered with complementary adapters, deoxynucleotides (dNTPs), primers and DNA polymerase. PCR is carried out within the micelle, covering each bead with thousands of copies of the same DNA sequence. **b** | In solid-phase bridge amplification, fragmented DNA is ligated to adapter sequences and bound to a primer immobilized on a solid support, such as a patterned flow cell. The free end can interact with other nearby primers, forming a bridge structure. PCR is used to create a second strand from the immobilized primers, and unbound DNA is removed. **c** | In solid-phase template walking[154], fragmented DNA is ligated to adapters and bound to a complementary primer attached to a solid support. PCR is used to generate a second strand. The now double-stranded template is partially denatured, allowing the free end of the original template to drift and bind to another nearby primer sequence. Reverse primers are used to initiate strand displacement to generate additional free templates, each of which can bind to a new primer. **d** | In DNA nanoball generation, DNA is fragmented and ligated to the first of four adapter sequences. The template is amplified, circularized and cleaved with a type II endonuclease. A second set of adapters is added, followed by amplification, circularization and cleavage. This process is repeated for the remaining two adapters. The final product is a circular template with four adapters, each separated by a template sequence. Library molecules undergo a rolling circle amplification step, generating a large mass of concatamers called DNA nanoballs, which are then deposited on a flow cell. Parts **a** and **b** are adapted from REF. 18, Nature Publishing Group.

© 2016 Macmillan Publishers Limited. All rights reserved.

added. After the incorporation of a single dNTP to each elongating complementary strand, unbound dNTPs are removed and the surface is imaged to identify which dNTP was incorporated at each cluster. The fluorophore and blocking group can then be removed and a new cycle can begin.

The Illumina CRT system (FIG. 3a) accounts for the largest market share for sequencing instruments compared to other platforms[21]. Illumina's suite of instruments for short-read sequencing range from small, low-throughput benchtop units to large ultra-high-throughput instruments dedicated to population-level whole-genome sequencing (WGS). dNTP identification is achieved through total internal reflection fluorescence (TIRF) microscopy using either two or four laser channels. In most Illumina platforms, each dNTP is bound to a single fluorophore that is specific to that base type and requires four different imaging channels, whereas the NextSeq and Mini-Seq systems use a two-fluorophore system.

In 2012, Qiagen acquired the Intelligent BioSystems CRT platform, which was commercialized and relaunched in 2015 as the GeneReader[22] (FIG. 3b). Unlike other systems, this platform is intended to be an all-in-one NGS platform, from sample preparation to analysis. To accomplish this, the GeneReader system is bundled with the QIAcube sample preparation system and the Qiagen Clinical Insight platform for variant analysis. The GeneReader uses virtually the same approach as that used by Illumina; however, it does not aim to ensure that each template incorporates a fluorophore-labelled dNTP[23]. Rather, GeneReader aims to ensure that just enough labelled dNTPs are incorporated to achieve identification.

Sequencing by synthesis: SNA (454, Ion Torrent). Unlike CRT, SNA approaches rely on a single signal to mark the incorporation of a dNTP into an elongating strand. As a consequence, each of the four nucleotides must be added iteratively to a sequencing reaction to ensure only one dNTP is responsible for the signal. Furthermore, this does not require the dNTPs to be blocked, as the absence of the next nucleotide in the sequencing reaction prevents elongation. The exception to this is homopolymer regions where identical dNTPs are added, with sequence identification relying on a proportional increase in the signal as multiple dNTPs are incorporated.

The first NGS instrument developed was the 454 pyrosequencing[24] device. This SNA system distributes template-bound beads into a PicoTiterPlate along with beads containing an enzyme cocktail. As a dNTP is incorporated into a strand, an enzymatic cascade occurs, resulting in a bioluminescence signal. Each burst of light, detected by a charge-coupled device (CCD) camera, can be attributed to the incorporation of one or more identical dNTPs at a particular bead (FIG. 4a).

The Ion Torrent was the first NGS platform without optical sensing[25]. Rather than using an enzymatic cascade to generate a signal, the Ion Torrent platform detects the H^+ ions that are released as each dNTP is incorporated. The resulting change in pH is detected by an integrated complementary metal-oxide-semiconductor (CMOS) and an ion-sensitive field-effect transistor (ISFET) (FIG. 4b). The pH change detected by the sensor is imperfectly proportional to the number of nucleotides detected, allowing for limited accuracy in measuring homopolymer lengths.

Comparison of short-read platforms. Individual short-read sequencing platforms vary with respect to throughput, cost, error profile and read structure (TABLE 1). Despite the existence of several NGS technology providers, NGS research is increasingly being conducted within the Illumina suite of instruments[21]. Although this implies high confidence in their data, it also raises concerns about systemic biases derived from using a single sequencing approach[26–28]. As a consequence, new approaches are being developed and researchers increasingly have the choice to integrate multiple sequencing methods with complementary strengths.

The SBL technique used by both the SOLiD and Complete Genomics systems affords these technologies a very high accuracy (~99.99%)[7,14], as each base is probed multiple times. Although accurate, both platforms also show evidence of a trade-off between sensitivity and specificity, such that true variants are missed while few false variants are called[29–31]. There is also evidence that the platforms share some under-representation of AT-rich regions[26,32], and the SOLiD platform displays some substitution errors and some GC-rich under-representation[32]. Perhaps the feature most limiting to the widespread adoption of these technologies is the very short read lengths. Although both platforms can generate single-end and paired-end sequencing reads, the maximum read length is just 75 bp for SOLiD and 28–100 bp for Complete Genomics[33], limiting their use for genome assembly and structural variant detection applications. Unfortunately, owing to these limitations, along with runtimes on the order of several days, the SOLiD system has been relegated to a small niche within the industry. Furthermore, although the cPAL-based Revolocity system was intended to compete with the Illumina HiSeq in terms of cost and throughout, its launch was suspended in 2016 and it is now only available as a service platform for human WGS[33,34], whereas the cPAS-based BGISEQ-500 platform is limited to mainland China.

Illumina dominates the short-read sequencing industry owing, in part, to its maturity as a technology, a high level of cross-platform compatibility and its wide range of platforms. The suite of instruments available ranges from the low-throughput MiniSeq to the ultra-high-throughput HiSeq X, which is capable of sequencing ~1,800 human genomes to 30× coverage per year. Further diversification is derived from the many options available for runtime, read structure and read length (up to 300 bp). As the Illumina platform relies on a CRT approach, it is much less susceptible to the homopolymer errors observed in SNA platforms. Although it has an overall accuracy rate of >99.5%[35], the platform does display some under-representation in AT-rich[32,36] and GC-rich regions[32,37], as well as a tendency towards substitution errors[38]. In 2008, Bentley

One-base-encoded probes
Oligonucleotides that contain a single interrogation base in a known position. The base corresponds to a fluorescent label on each probe. The remaining bases are either degenerate (any of the four bases) or universal (unnatural bases with nonspecific hybridization), allowing the probe to interact with many different possible template sequences.

Two-base-encoded probes
Oligonucleotides that contain two adjacent interrogation bases in a known position. The bases correspond to a fluorescent label on each probe. The remaining bases are either degenerate (any of the four bases) or universal (unnatural bases with nonspecific hybridization) allowing the probe to interact with many different possible template sequences.

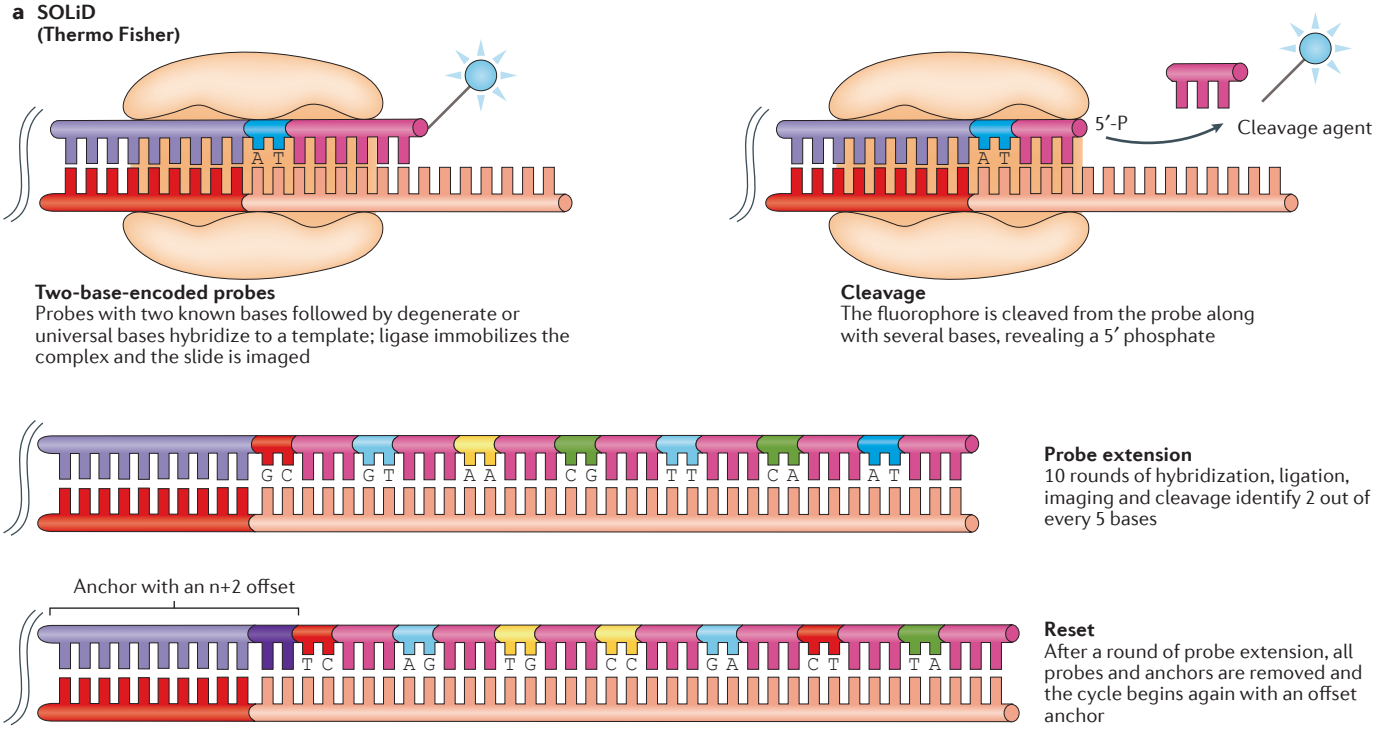
Colour-space
A system exclusively used by SOLiD. When a two-base-encoded probe is used, the bound label corresponds to two bases rather than one. Thus, the signal derived from a SOLiD run is in a series of colours that represent overlapping dinucleotides, rather than each colour being directly correlated to a single base. A reference-based alignment is the most efficient way to translate colour-space into base-space. For example, in the sequence ATGT the first probe will match AT, the second will match TG and the third GT. If the AT is known, then the subsequent colour order is uniquely solved as TG and GT, leading to a readout of ATGT. Final sequence deconvolution of colour-space is achieved with the knowledge of the second base identity in one round and the colour of the subsequent round in which the ligation is offset by one nucleotide, allowing for the identification of the next base.

Base-space
A system used by most next-generation sequencing platforms. When a one-base-encoded probe or a sequencing-by-synthesis approach is used, each signal is correctly correlated to a base.

© 2016 Macmillan Publishers Limited. All rights reserved.

Whole-genome sequencing (WGS). Sequencing of the entire genome without using methods for sequence selection.

et al.[35] reported a very high concordance rate between human single-nucleotide polymorphisms (SNPs) identified with Illumina and SNPs identified from genotyping microarrays[35]. However, this high sensitivity came with a false-positive rate of around 2.5%, leading this and other groups to consider using Sanger sequencing to resequence the called SNPs in order to distinguish between true SNPs and false positives[35,39,40]. With all of the possible options available, the Illumina suite allows for a wide range of applications: genome sequencing through WGS or exome sequencing; epigenomics applications, such as ChIP–seq (chromatin



Two-base-encoded probes
Probes with two known bases followed by degenerate or universal bases hybridize to a template; ligase immobilizes the complex and the slide is imaged

Cleavage
The fluorophore is cleaved from the probe along with several bases, revealing a 5′ phosphate

Probe extension
10 rounds of hybridization, ligation, imaging and cleavage identify 2 out of every 5 bases

Reset
After a round of probe extension, all probes and anchors are removed and the cycle begins again with an offset anchor

Single-base-encoded probes
A probe with a single known base and degenerate bases hybridizes to a template and is imaged

Reset
After each imaging step, both the probe and anchor are removed

Paired-end sequencing
Sequencing is performed for both the left and right sides of the adapter

Offset anchors
Subsequent rounds of hybridization and ligation use offset anchors to sequence more-distant bases

© 2016 Macmillan Publishers Limited. All rights reserved.

Two-fluorophore system
A system in which bases are discriminated by labelling Cs and Ts with a red or green fluorophore, respectively. Each A base is labelled with either a red or green fluorophore, but the two populations are mixed. During base discrimination, clusters that are either red or green are called either C or T, whereas clusters with a red and green mixed signal are called A. The G base is unlabelled, thus any cluster without a fluorophore signal is called G.

Homopolymer
A sequence run of identical bases.

Charge-coupled device
(CCD). A device composed of an integrated circuit that forms light-sensitive elements: pixels. When a photon interacts with the device, the light generates a charge that can be interpreted by an electronic device.

Integrated complementary metal-oxide-semiconductor
(CMOS). An integrated circuit design that is printed on a microchip that contains different types of semiconductor transistors to create a circuit that both uses very little power and is resistant to high levels of electronic noise.

immunoprecipitation followed by sequencing)[41], ATAC–seq (assay for transposase-accessible chromatin using sequencing)[42] or DNA methylation sequencing (methyl-seq)[43]; and transcriptomics applications through RNA sequencing (RNA-seq)[44], to name a few. The two-colour labelling system used by the NextSeq and MiniSeq platforms increases speed and reduces costs by reducing scanning to two colour channels and reducing fluorophore usage. However, the two-channel system results in a slightly higher error profile and underperformance for low-diversity samples owing to more ambiguous base discrimination[45]. HiSeq X is currently the highest-throughput instrument available; however, as a consequence of its optimization, it is limited to just a few applications, such as WGS and whole-genome bisulfite sequencing. HiSeq X is further limited as an all-purpose instrument owing to a required initial purchase of five or ten instruments (additional single instruments can be purchased after the initial commitment), placing this system out of reach of most facilities.

The Qiagen GeneReader is intended to be a clinical device with an explicit focus on cancer gene panels[46]; although this severely limits its possible applications, it is well optimized within its niche. With a reported several-day runtime, and the use of validated gene panels, it fulfils a similar role to the Illumina MiSeq[46]. Although no user data are available at this point, it is likely that the Qiagen GeneReader will share many of the same advantages and disadvantages as the Illumina MiSeq platform at a potentially lower cost per gigabase sequenced.

Both the 454 and the Ion Torrent systems offer superior read lengths compared to other short-read sequencers with reads up to an average of 700 bp and 400 bp, respectively, providing some advantages for applications that focus on repetitive or complex DNA. However, as both of these platforms rely on SNA, they share many of the same drawbacks. Insertion and deletion (indel) errors dominate, although the overall error rate is on par with other NGS platforms in non-homopolymer regions. Homopolymer regions are problematic for these platforms, which lack single-base accuracy in measuring homopolymers larger than 6–8 bp[47,48]. Unfortunately, whereas the Ion Torrent platform has kept pace with the rapidly evolving NGS field, the 454 platform has been unable to complete with other platforms in terms of yield or cost. This limitation has led Roche to discontinue the platform in 2016 (REF. 49).

The Ion Torrent platform offers several different types of chips and instruments to tune sequencer performance to the needs of the researcher. The throughput of these chips ranges from ~50 Mb to 15 Gb, with runtimes between 2 and 7 hours, making it faster than most other current platforms. This makes the device well suited for gene-panel sequencing and for point-of-care clinical applications[50], including transcriptome profiling[51] and splice site identification (although not to the level of long-read sequencers)[51]. Ion Torrent is attempting to capitalize on the growing interest in clinical sequencing with the release of its dedicated diagnostic instruments: the Ion Personal Genome Machine (PGM) Dx and the Ion S5 series. When paired with the Ion Chef library preparation and chip loading device, the S5 series in particular aims to be one of the simplest platforms to operate, eliminating the need for the argon required by other Ion Torrent instruments and establishing plug-and-play protocols. An important disadvantage, however, is that although the Ion PGM Dx sequencer can support paired-end sequencing[52], the higher-throughput Ion Proton and S5 devices lack the ability to perform paired-end sequencing, thus limiting their utility for elucidating long-range genomic or transcriptomic structure[53].

Figure 2 | **Sequencing by ligation methods. a** | SOLiD sequencing. Following cluster generation or bead deposition onto a slide, fragments are sequenced by ligation, in which a fluorophore-labelled two-base-encoded probe, which is composed of known nucleotides in the first and second positions (dark blue), followed by degenerate or universal bases (pink), is added to the DNA library. The two-base probe is ligated onto an anchor (light purple) that is complementary to an adapter (red), and the slide is imaged to identify the first two bases in each fragment. Unextended strands are capped by unlabelled probes or phosphatase to maintain cycle synchronization. Finally, the terminal degenerate bases and the fluorophore are cleaved off the probe, leaving a 5 bp extended fragment. The process is repeated ten times until two out of every five bases are identified. At this point, the entire strand is reset by removing all of the ligated probes and the process of probe binding, ligation, imaging and cleavage is repeated four times, each with an n + 1, n + 2, n + 3 or n + 4 offset anchor. **b** | Complete Genomics. DNA is sequenced using the combinatorial probe–anchor ligation (cPAL) approach. After DNA nanoball deposition, an anchor complementary to one of four adapter sequences and a fluorophore-labelled probe are bound to each nanoball. The probe is degenerate at all but the first position. The anchor and probe are then ligated into position and imaged to identify the first base on either the 3′ or the 5′ side of the anchor. Next, the probe–anchor complex is removed and the process begins again with the same anchor but a different probe with the known base at the n + 1 position. This is repeated until five bases from the 3′ end of the anchor and five bases from the 5′ end of the anchor are identified. Another round of hybridization occurs, this time using anchors with a five-base offset identifying an additional five bases on either side of the anchor. Finally, this whole process is repeated for each of the remaining three adapter sequences in the nanoball, generating 100 bp paired-end reads.

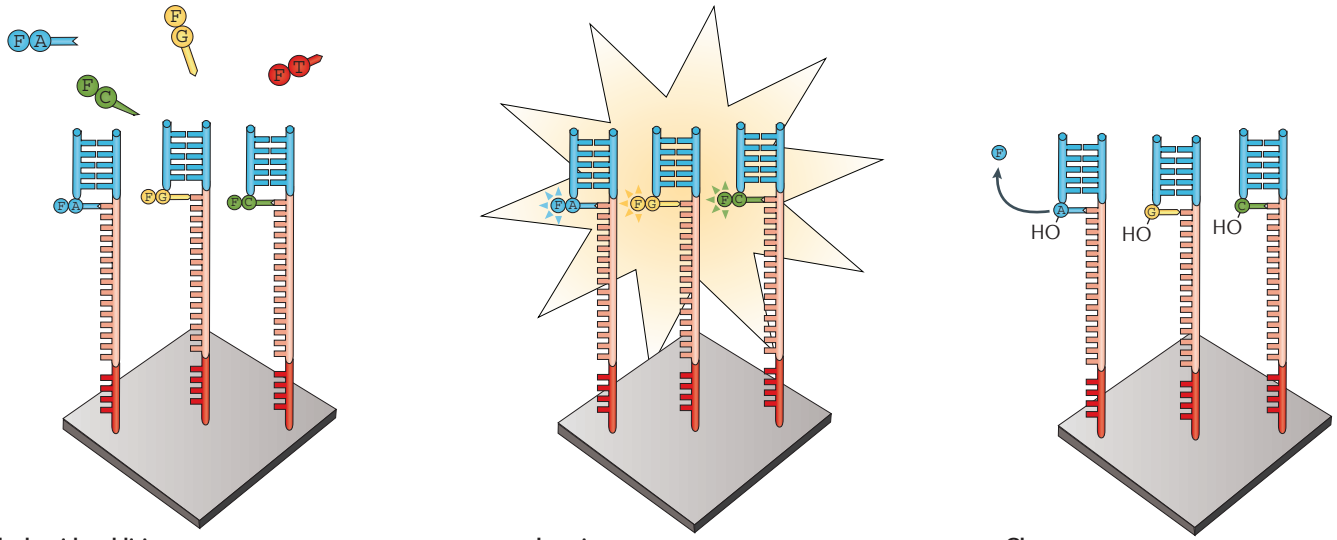
Long-read sequencing

Overview. It has become apparent that genomes are highly complex with many long repetitive elements, copy number alterations and structural variations that are relevant to evolution, adaptation and disease[54–56]. However, many of these complex elements are so long that short-read paired-end technologies are insufficient to resolve them. Long-read sequencing delivers reads in excess of several kilobases, allowing for the resolution of these large structural features. Such long reads can span complex or repetitive regions with a single continuous read, thus eliminating ambiguity in the positions or size of genomic elements. Long reads can also be useful for transcriptomic research, as they are capable of spanning entire mRNA transcripts, allowing researchers to identify the precise connectivity of exons and discern gene isoforms.

Currently, there are two main types of long-read technologies: single-molecule real-time sequencing approaches and synthetic approaches that rely on existing short-read technologies to construct long reads *in silico*. The single-molecule approaches differ from short-read approaches in that they do not rely on a clonal population of amplified DNA fragments to generate detectable

© 2016 Macmillan Publishers Limited. All rights reserved.

a Illumina

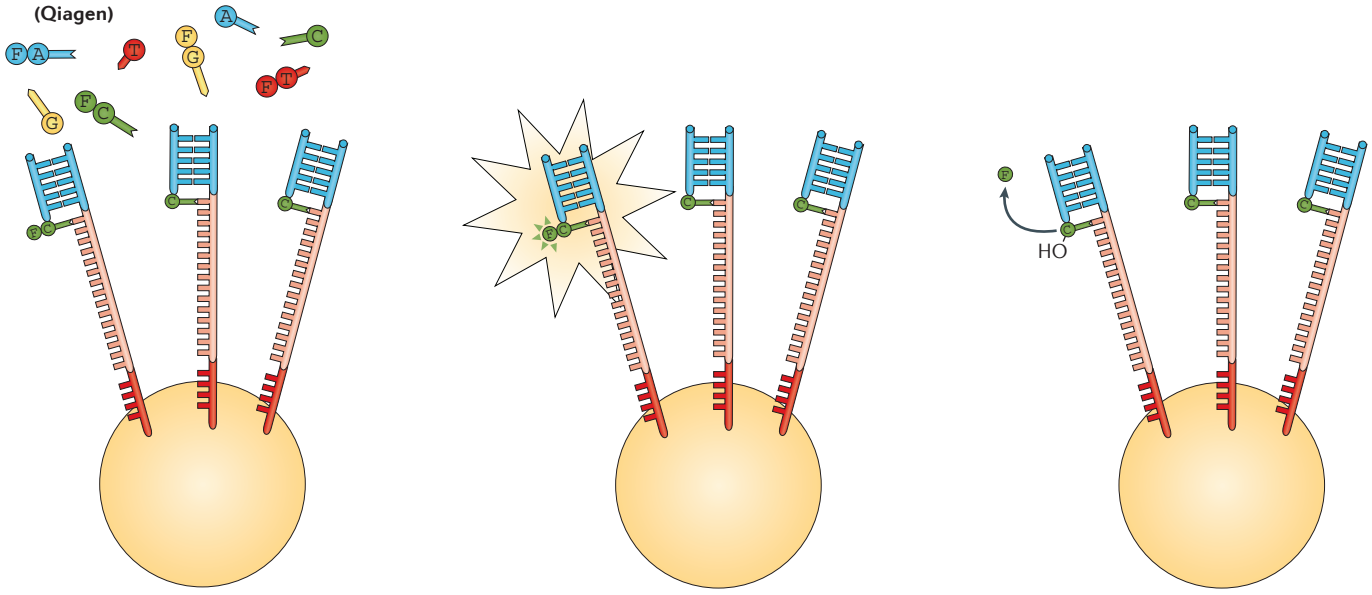


Nucleotide addition
Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.

Imaging
Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.

Cleavage
Fluorophores are cleaved and washed from flow cells and the 3′-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

b GeneReader (Qiagen)



Nucleotide addition
A mixture of fluorophore-labelled, terminally blocked nucleotides and unlabelled, blocked nucleotides hybridize to complementary bases. Each bead on a slide can incorporate a different base.

Imaging
Slides are imaged with four laser channels. Each bead emits a colour corresponding to the base incorporated during this cycle, but only labelled bases emit a signal.

Cleavage
Fluorophores are cleaved and washed from flow cells and the 3′-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

Figure 3 | **Sequencing by synthesis: cyclic reversible termination approaches. a** | Illumina. After solid-phase template enrichment, a mixture of primers, DNA polymerase and modified nucleotides are added to the flow cell. Each nucleotide is blocked by a 3′-*O*-azidomethyl group and is labelled with a base-specific, cleavable fluorophore (F). During each cycle, fragments in each cluster will incorporate just one nucleotide as the blocked 3′ group prevents additional incorporations. After base incorporation, unincorporated bases are washed away and the slide is imaged by total internal reflection fluorescence (TIRF) microscopy using either two or four laser channels; the colour (or the lack or mixing of colours in the two-channel system used by NextSeq) identifies which base was incorporated in each cluster. The dye is then cleaved and the 3′-OH is regenerated with the reducing agent tris(2-carboxyethyl)phosphine (TCEP). The cycle of nucleotide addition, elongation and cleavage can then begin again. **b** | Qiagen. After bead-based template enrichment, a mixture of primers, DNA polymerase and modified nucleotides are added to the flow cell. Each nucleotide is blocked by a 3′-*O*-allyl group and some of the bases are labelled with a base-specific, cleavable fluorophore. After base incorporation, unincorporated bases are washed away and the slide is imaged by TIRF using four laser channels. The dye is then cleaved and the 3′-OH is regenerated with the reducing agent mixture of palladium and $P(PhSO_3Na)_3$ (TPPTS).

© 2016 Macmillan Publishers Limited. All rights reserved.

Ion-sensitive field-effect transistor
(ISFET). A type of transistor that is sensitive to changes in ion concentration.

Single-end and paired-end sequencing
In single-end sequencing, a DNA template is sequenced only in one direction. In paired-end sequencing, a DNA template is sequenced from both sides; the forward and reverse reads may or may not overlap. A deviation in the expected genome alignment between two ends of a paired-end read can indicate astructural variation.

Structural variant
A variation larger than single-nucleotide polymorphisms (SNPs). This can include the insertion or deletion of blocks of DNA, inversions or translocations of DNA segments, and copy-number differences.

ChIP–seq
(Chromatin immunoprecipitation followed by sequencing). A method used to analyse protein interactions with DNA by combining ChIP with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins.

ATAC–seq
(Assay for transposase-accessible chromatin with high-throughput sequencing). A method that uses the activity of a hyperactive transposase to cleave exposed DNA and add sequencing adapters. Regions that cannot be sequenced are inferred to be chromatin interacting.

RNA sequencing
(RNA-seq). A method of sequencing cDNA derived from RNA. This approach can be used to sequence both coding and non-coding RNA.

Real-time sequencing
A sequencing strategy used in the Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) platforms. In these approaches there is no pause after the detection of a base or series of bases, thus the sequence is derived in real-time.

signal, nor do they require chemical cycling for each dNTP added. Alternatively, the synthetic approaches do not generate actual long-reads; rather, they are an approach to library preparation that leverages barcodes to allow computational assembly of a larger fragment.

Single-molecule long-read sequencing (PacBio and ONT). Currently, the most widely used long-read platform is the single-molecule real-time (SMRT) sequencing approach used by Pacific Biosciences (PacBio)[57] (FIG. 5a). The instrument uses a specialized flow cell

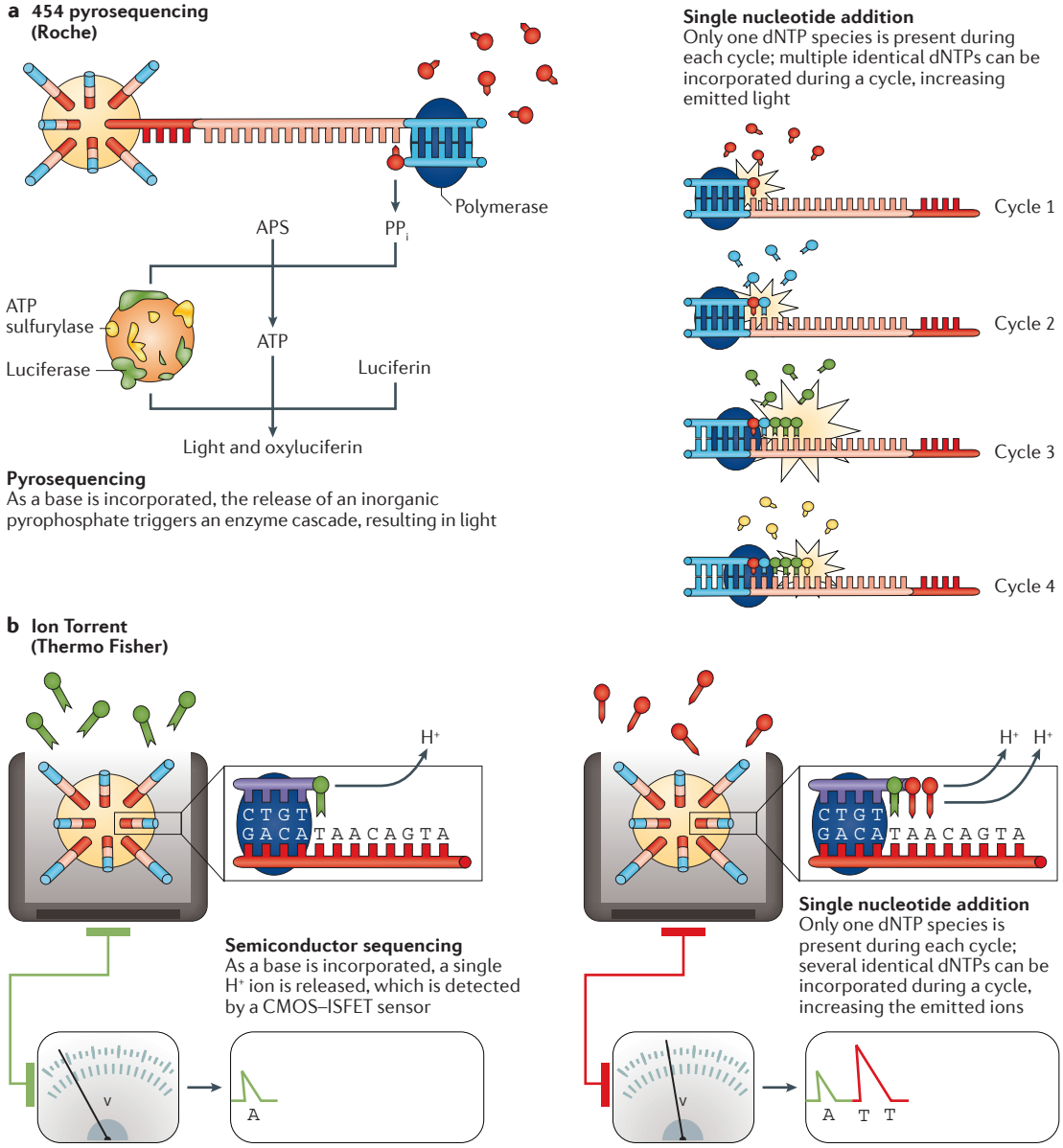


Figure 4 | **Sequencing by synthesis: single-nucleotide addition approaches. a** | 454 pyrosequencing. After bead-based template enrichment, the beads are arrayed onto a microtitre plate along with primers and different beads that contain an enzyme cocktail. During the first cycle, a single nucleotide species is added to the plate and each complementary base is incorporated into a newly synthesized strand by a DNA polymerase. The by-product of this reaction is a pyrophosphate molecule (PP_i). The PP_i molecule, along with ATP sulfurylase, transforms adenosine 5′ phosphosulfate (APS) into ATP. ATP, in turn, is a cofactor for the conversion of luciferin to oxyluciferin by luciferase, for which the by-product is light. Finally, apyrase is used to degrade any unincorporated bases and the next base is added to the wells. Each burst of light, detected by a charge-coupled device (CCD) camera, can be attributed to the incorporation of one or more bases at a particular bead. **b** | Ion Torrent. After bead-based template enrichment, beads are carefully arrayed into a microtitre plate where one bead occupies a single reaction well. Nucleotide species are added to the wells one at a time and a standard elongation reaction is performed. As each base is incorporated, a single H^+ ion is generated as a by-product. The H^+ release results in a 0.02 unit change in pH, detected by an integrated complementary metal-oxide semiconductor (CMOS) and an ion-sensitive field-effect transistor (ISFET) device. After the introduction of a single nucleotide species, the unincorporated bases are washed away and the next is added. Part **a** is adapted from REF. 18, Nature Publishing Group.

© 2016 Macmillan Publishers Limited. All rights reserved.

Table 1 | **Summary of NGS platforms**

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US$)	Cost per Gb (US$, approx.)
Sequencing by ligation							
SOLiD 5500 Wildfire	50 (SE)	80 Gb	~700 M*	6 d*	≤0.1%, AT bias‡	NA§	$130‡
	75 (SE)	120 Gb					
	50 (SE)*	160 Gb*					
SOLiD 5500xl	50 (SE)	160 Gb	~1.4 B*	10 d*	≤0.1%, AT bias‡	$251,000‡	$70‡
	75 (SE)	240 Gb					
	50 (SE)*	320 Gb*					
BGISEQ-500 FCS[155]	50–100 (SE/PE)*	8–40 Gb*	NA‖	24 h*	≤0.1%, AT bias‡	$250 (REF. 155)	NA‖
BGISEQ-500 FCL[155]	50–100 (SE/PE)*	40–200 Gb*	NA‖	24 h*	≤0.1%, AT bias‡	$250,000 (REF. 155)	NA‖
Sequencing by synthesis: CRT							
Illumina MiniSeq Mid output	150 (SE)*	2.1–2.4 Gb*	14–16 M*	17 h*	<1%, substitution‡	$50,000 (REF. 118)	$200–300 (REF. 118)
Illumina MiniSeq High output	75 (SE)	1.6–1.8 Gb	22–25 M (SE)*	7 h	<1%, substitution‡	$50,000 (REF. 118)	$200–300 (REF. 118)
	75 (PE)	3.3–3.7 Gb	44–50 M (PE)*	13 h			
	150 (PE)*	6.6–7.5 Gb*		24 h*			
Illumina MiSeq v2	36 (SE)	540–610 Mb	12–15 M (SE)	4 h	0.1%, substitution‡	$99,000‡	~$1,000
	25 (PE)	750–850 Mb	24–30 M (PE)*	5.5 h			$996
	150 (PE)	4.5–5.1 Gb		24 h			$212
	250 (PE)*	7.5–8.5 Gb*		39 h*			$142‡
Illumina MiSeq v3	75 (PE)	3.3–3.8 Gb	44–50 M (PE)*	21–56 h*	0.1%, substitution‡	$99,000‡	$250
	300 (PE)*	13.2–15 Gb*					$110‡
Illumina NextSeq 500/550 Mid output	75 (PE)	16–20 Gb	Up to 260 M (PE)*	15 h	<1%, substitution‡	$250‡	$42
	150 (PE)*	32–40 Gb*		26 h*			$40‡
Illumina NextSeq 500/550 High output	75 (SE)	25–30 Gb	400 M (SE)*	11 h	<1%, substitution‡	$250‡	$43
	75 (PE)	50–60 Gb	800 M (PE)*	18 h			$41
	150 (PE)*	100–120 Gb*		29 h*			$33‡
Illumina HiSeq2500 v2 Rapid run	36 (SE)	9–11Gb	300 M (SE)*	7 h	0.1%, substitution‡	$690‡	$230
	50 (PE)	25–30 Gb	600 M (PE)*	16 h			$90
	100 (PE)	50–60 Gb		27 h			$52
	150 (PE)	75–90 Gb		40 h			$45
	250 (PE)*	125–150 Gb*		60 h*			$40‡
Illumina HiSeq2500 v3	36 (SE)	47–52 Gb	1.5 B (SE)	2 d	0.1%, substitution‡	$690‡	$180
	50 (PE)	135–150 Gb	3 B (PE)*	5.5 d			$78
	100 (PE)*	270–300 Gb		11 d*			$45‡
Illumina HiSeq2500 v4	36 (SE)	64–72 Gb	2 B (SE)	29 h	0.1%, substitution‡	$690‡	$150
	50 (PE)	180–200 Gb	4 B (PE)*	2.5 d			$58
	100 (PE)	360–400 Gb		5 d			$45
	125 (PE)*	450–500 Gb*		6 d*			$30‡
Illumina HiSeq3000/4000	50 (SE)	105–125 Gb	2.5 B (SE)*	1–3.5 d*	0.1%, substitution‡	$740/$900 (REF. 156)	$50
	75 (PE)	325–375 Gb					$31
	150 (PE)*	650–750 Gb*					$22 (REF. 157)

© 2016 Macmillan Publishers Limited. All rights reserved.

Table 1 (cont.) | **Summary of NGS platforms**

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US$)	Cost per Gb (US$, approx.)
Sequencing by synthesis: SNA (cont.)							
Illumina HiSeq X	150 (PE)*	800–900 Gb per flow cell*	2.6–3 B (PE)*	<3 d*	0.1%, substitution‡	$1,000‡,¶	$7.0‡
Qiagen GeneReader	NA‖	12 genes; 1,250 mutations[22]	NA‖	Several days[22]	Similar to other SBS systems[22]	NA‖	$400–$600 per panel[22]
Sequencing by synthesis: SNA							
454 GS Junior	Up to 600; 400 average (SE, PE)*	35 Mb*	~0.1 M*	10 h*	1%, indel‡	NA§	$40,000‡
454 GS Junior+	Up to 1,000; 700 average (SE, PE)*	70 Mb*	~0.1 M*	18 h*	1%, indel‡	$108,000‡	$19,500‡
454 GS FLX Titanium XLR70	Up to 600; 450 mode (SE, PE)*	450 Mb*	~1 M*	10 h*	1%, indel‡	NA§	$15,500‡
454 GS FLX Titanium XL+	Up to 1,000; 700 mode (SE, PE)*	700 Mb*	~1 M*	23 h*	1%, indel‡	$450,000‡	$9,500‡
Ion PGM 314	200 (SE)	30–50	400,000–550,000*	23 h	1%, indel‡	$49‡	$25–3,500‡
	400 (SE)	60–100 Mb*		3.7 h*			
Ion PGM 316	200 (SE)	300–500 Mb	2–3 M*	3 h	1%, indel‡	$49‡	$700–1,000‡
	400 (SE)*	600 Mb–1 Gb*		4.9 h*			
Ion PGM 318	200 (SE)	600 Mb–1 Gb	4–5.5 M*	4 h	1%, indel‡	$49‡	$450–800‡
	400 (SE)*	1–2 Gb*		7.3 h*			
Ion Proton	Up to 200 (SE)	Up to 10 Gb*	60–80 M*	2–4 h*	1%, indel‡	$224‡	$80‡
Ion S5 520	200 (SE)	600 Mb–1 Gb	3–5 M*	2.5 h	1%, indel‡	$65 (REF. 158)	$2,400*
	400 (SE)*	1.2–2 Gb*		4 h*			$1,200*
Ion S5 530	200 (SE)	3–4 Gb	15–20 M*	2.5 h	1%, indel‡	$65 (REF. 158)	$950*
	400 (SE)*	6–8 Gb*		4 h*			$475*
Ion S5 540	200 (SE)*	10–15 Gb*	60–80 M*	2.5 h*	1%, indel‡	$65 (REF. 158)	$300*
Single-molecule real-time long reads							
Pacific BioSciences RS II	~20 Kb	500 Mb–1 Gb*	~55,000*	4 h*	13% single pass, ≤1% circular consensus read, indel‡	$695‡	$1,000‡
Pacific Biosciences Sequel	8–12 Kb[69]	3.5–7 Gb*	~350,000*	0.5–6 h*	NA‖	$350 (REF. 69)	NA‖
Oxford Nanopore MK 1 MinION	Up to 200 Kb[159]	Up to 1.5 Gb[159]	>100,000 (REF. 159)	Up to 48 h[160]	~12%, indel[159]	$1,000*	$750*
Oxford Nanopore PromethION	NA‖	Up to 4 Tb*	NA‖	NA‖	NA‖	$75*	NA‖
Synthetic long reads							
Illumina Synthetic Long-Read	~100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	No additional instrument required	~$1,000*
10X Genomics	Up to 100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	$75 (REFS 72,161)	See HiSeq 2500 +$500 per sample[161]

Approx., approximate; AT, adenine and thymine; B, billion; bp, base pairs; d, days; Gb, gigabase pairs; h, hours; indel, insertions and deletions; Kb, kilobase pairs; M, million; Mb, megabase pairs; NA, not available; PE, paired-end sequencing; SBS, sequencing by synthesis; SE, single-end sequencing; Tb, terabase pairs. *Manufacturer's data. ‡Rounded from Field Guide to next-generation DNA sequencers[160] and 2014 update. §Not available as this instrument will be discontinued or only available as an upgraded version. ‖As this product has been developed only recently, this information is not available. ¶Not available as a single instrument.

© 2016 Macmillan Publishers Limited. All rights reserved.

A Real-time long-read sequencing

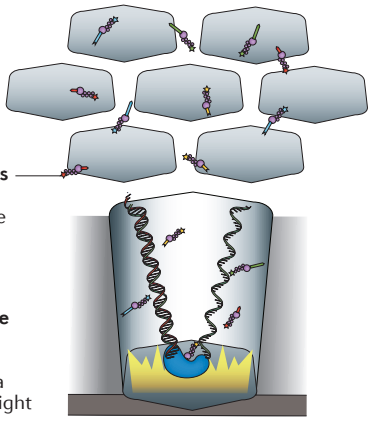
Aa Pacific Biosciences

SMRTbell template
Two hairpin adapters allow continuous circular sequencing

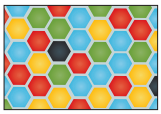
ZMW wells
Sites where sequencing takes place

Labelled nucleotides
All four dNTPs are labelled and available for incorporation

Modified polymerase
As a nucleotide is incorporated by the polymerase, a camera records the emitted light

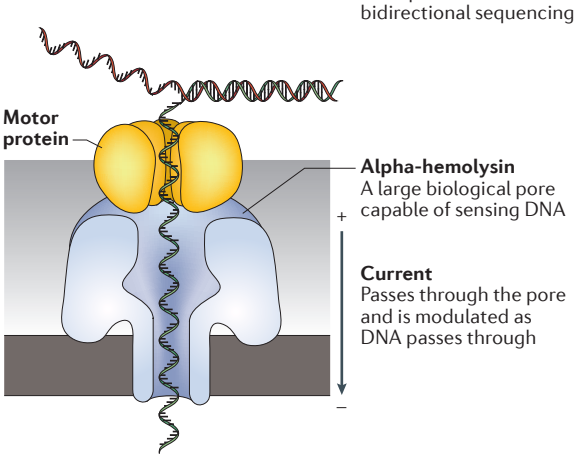


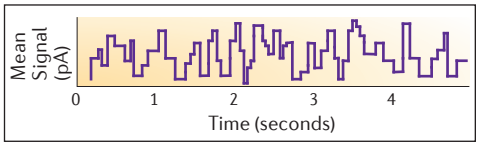
PacBio output
A camera records the changing colours from all ZMWs; each colour change corresponds to one base



Ab Oxford Nanopore Technologies

Leader–Hairpin template
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing





ONT output (squiggles)
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

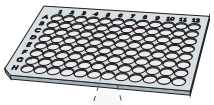
B Synthetic long-read sequencing

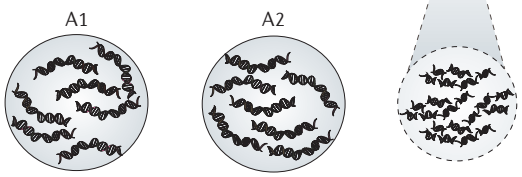
Ba Illumina

DNA fragment
DNA is fragmented and selected to ~10 kb

~3,000 molecules per well

Enzymatic cleavage
DNA is barcoded and fragmented to ~350 bp





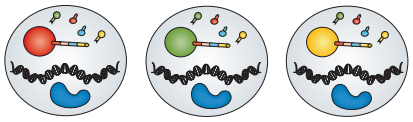
Barcodes
DNA from the same well shares the same barcode

Pooling
DNA from each well is pooled and undergoes a standard library preparation

Sequencing
DNA is sequenced on a standard short-read sequencer

Bb 10X Genomics

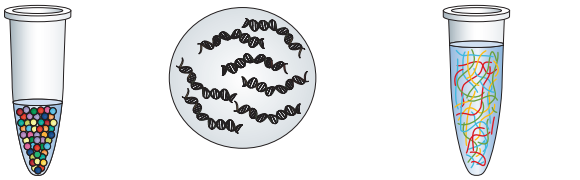
Emulsion PCR
Arbitrarily long DNA is mixed with beads loaded with barcoded primers, enzyme and dNTPs

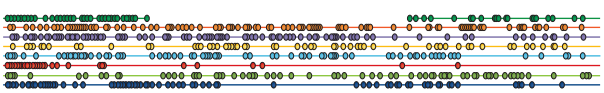


GEMs
Each micelle has 1 barcode out of 750,000

Amplification
Long fragments are amplified such that the product is a barcoded fragment ~350 bp

Pooling
The emulsion is broken and DNA is pooled, then it undergoes a standard library preparation





Linked reads
- All reads from the same GEM derive from the long fragment, thus they are linked
- Reads are dispersed across the long fragment and no GEM achieves full coverage of a fragment
- Stacking of linked reads from the same loci achieves continuous coverage

© 2016 Macmillan Publishers Limited. All rights reserved.

Barcodes
A series of known bases added to a template molecule either through ligation or amplification. After sequencing, these barcodes can be used to identify which sample a particular read is derived from.

with many thousands of individual picolitre wells with transparent bottoms — zero-mode waveguides (ZMW)[58]. Whereas short-read SBS technologies bind the DNA and allow the polymerase to travel along the DNA template, PacBio fixes the polymerase to the bottom of the well and allows the DNA strand to progress through the ZMW. By having a constant location of incorporation owing to the stationary enzyme, the system can focus on a single molecule. dNTP incorporation on each single-molecule template per well is continuously visualized with a laser and camera system that records the colour and duration of emitted light as the labelled nucleotide momentarily pauses during incorporation at the bottom of the ZMW. The polymerase cleaves the dNTP-bound fluorophore during incorporation, allowing it to diffuse away from the sensor area before the next labelled dNTP is incorporated. The SMRT platform also uses a unique circular template that allows each template to be sequenced multiple times as the polymerase repeatedly traverses the circular molecule. Although it is difficult for DNA templates longer than ~3 kb to be sequenced multiple times, shorter DNA templates can be sequenced many times as a function of template length[57,59]. These multiple passes are used to generate a consensus read of insert, known as a circular consensus sequence (CCS).

In 2014, the first consumer prototype of a nanopore sequencer — the MinION from Oxford Nanopore Technologies (ONT) — became available. Unlike other platforms, nanopore sequencers do not monitor incorporations or hybridizations of nucleotides guided by a template DNA strand. Whereas other platforms use a secondary signal, light, colour or pH, nanopore sequencers directly detect the DNA composition of a native ssDNA molecule. To carry out sequencing, DNA is passed through a protein pore as current is passed through the pore[60] (FIG. 5b). As the DNA translocates through the action of a secondary motor protein, a voltage blockade occurs that modulates the current passing through the pore. The temporal tracing of these charges is called squiggle space, and shifts in voltage are characteristic of the particular DNA sequence in the pore, which can then be interpreted as a k-mer. Rather than having 1–4 possible signals, the instrument has more than 1,000 — one for each possible k-mer, especially when modified bases present on native DNA are taken into account. The current MK1 MinION flow cell structure is composed of an application-specific integrated circuit (ASIC) chip with 512 individual channels that are capable of sequencing at ~70 bp per second, with an expected increase to 500 bp per second in 2016. The upcoming PromethION instrument is intended to be an ultra-high-throughput platform reported to include 48 individual flow cells, each with 3,000 pores running at 500 bp per second. This works out to ~2–4 Tb for a 2-day run on a fully loaded device, placing this device in potential competition with Illumina's HiSeq X. Similar to the circular template used by PacBio, the ONT MinION uses a leader–hairpin library structure. This allows the forward DNA strand to pass through the pore, followed by a hairpin that links the two strands, and finally the reverse strand. This generates 1D and 2D reads in which both '1D' strands can be aligned to create a consensus sequence '2D' read.

Synthetic long-reads. Unlike true sequencing platforms, synthetic long-read technology relies on a system of barcoding to associate fragments that are sequenced on existing short-read sequencers[61]. These approaches partition large DNA fragments into either microtitre wells or an emulsion such that very few molecules exist in each partition. Within each partition the template fragments are sheared and barcoded. This approach allows for sequencing on existing short-read instrumentation, after which data are split by barcode and reassembled with the knowledge that fragments sharing barcodes

Figure 5 | **Real-time and synthetic long-read sequencing approaches.** **A** | Real-time long-read sequencing platforms. **Aa** | Single-molecule real-time (SMRT) sequencing from Pacific Biosciences (PacBio). Template fragments are processed and ligated to hairpin adapters at each end, resulting in a circular DNA molecule with constant single-stranded DNA (ssDNA) regions at each end with the double-stranded DNA (dsDNA) template in the middle. The resulting 'SMRTbell' template undergoes a size-selection protocol in which fragments that are too large or too small are removed to ensure efficient sequencing. Primers and an efficient φ29 DNA polymerase are attached to the ssDNA regions of the SMRTbell. The prepared library is then added to the zero-mode waveguide (ZMW) SMRT cell, where sequencing can take place. To visualize sequencing, a mixture of labelled nucleotides is added; as the polymerase-bound DNA library sits in one of the wells in the SMRT cell, the polymerase incorporates a fluorophore-labelled nucleotide into an elongating DNA strand. During incorporation, the nucleotide momentarily pauses through the activity of the polymerase at the bottom of the ZMW, which is being monitored by a camera. **Ab** | Oxford Nanopore Technologies (ONT). DNA is initially fragmented to 8–10 kb. Two different adapters, a leader and a hairpin, are ligated to either end of the fragmented dsDNA. Currently, there is no method to direct the adapters to a particular end of the DNA molecule, so there are three possible library conformations: leader–leader, leader–hairpin and hairpin–hairpin. The leader adapter is a double-stranded adapter containing a sequence required to direct the DNA into the pore and a tether sequence to help direct the DNA to the membrane surface. Without this leader adapter, there is minimal interaction of the DNA with the pore, which prevents any hairpin–hairpin fragments from being sequenced. The ideal library conformation is the leader–hairpin. In this conformation the leader sequence directs the DNA fragment to the pore with current passing through. As the DNA translocates through the pore, a characteristic shift in voltage through the pore is observed. Various parameters, including the magnitude and duration of the shift, are recorded and can be interpreted as a particular k-mer sequence. As the next base passes into the pore, a new k-mer modulates the voltage and is identified. At the hairpin, the DNA continues to be translocated through the pore adapter and onto the complement strand. This allows the forward and reverse strands to be used to create a consensus sequence called a '2D' read. **B** | Synthetic long-read sequencing platforms. **Ba** | Illumina. Genomic DNA templates are fragmented to 8–10 kb pieces. They are then partitioned into a microtitre plate such that there are around 3,000 templates in a single well. Within the plate, each fragment is sheared to around 350 bp and barcoded with a single barcode per well. The DNA can then be pooled and sent through standard short-read pipelines. **Bb** | 10X Genomics' emulsion-based sequencing. With as little as 1 ng of starting material, the GemCode can partition arbitrarily large DNA fragments, up to ~100 kb, into micelles (also called 'GEMs') along with gel beads containing adapter and barcode sequences. The GEMs typically contain ~0.3× copies of the genome and 1 unique barcode out of 750,000. Within each GEM, the gel bead dissolves and smaller fragments of DNA are amplified from the original large fragments, each with a barcode identifying the source GEM. After sequencing, the reads are aligned and linked together to form a series of anchored fragments across a span of ~50 kb. Unlike the Illumina system, this approach does not attempt to get full end-to-end coverage of a single DNA fragment. Instead, the reads from a single GEM are dispersed across the original DNA fragment and the cumulative coverage is derived from multiple GEMs with dispersed — but linked — reads. Part **Aa** is adapted from REF. 18, Nature Publishing Group. Part **Ba** is adapted from REF. 62.

© 2016 Macmillan Publishers Limited. All rights reserved.

Zero-mode waveguides
(ZMW). Nanostructure devices used in the Pacific Biosciences (PacBio) platform. Each ZMW well (also called a waveguide) is several nanometres in diameter and is anchored to a glass substrate. The size of each well does not allow for light propagation, thus the fluorophores bound to bases can only be visualized through the glass substrate in the bottom-most portion of the well, a volume in the zeptolitre range.

Read of insert
The highest-quality single sequence for an insert, regardless of the number of passes.

Consensus sequence
In next-generation sequencing (NGS) routines that allow multiple overlapping reads from a single molecule of DNA, all related reads are aligned to each other and the most likely base at each position is determined. This process helps to overcome high, single-pass error rates. A high-quality consensus sequence derived from the circular template from Pacific Biosciences (PacBio) is called a circular consensus sequence (CCS).

Squiggle space
A system exclusively used by Oxford Nanopore Technologies (ONT). As DNA translocates through the pore, a shift in voltage occurs that is directly correlated to a k-mer within the pore. Thus, the signal derived from a nanopore run is a continuous series of voltage shifts (squiggles) that represent a series of overlapping k-mers.

K-mer
A substring within a sequence of bases of some (k) length. Currently, k-mer sizes of Oxford Nanopore Technologies (ONT) range from 3 to 6 bases.

1D and 2D reads
Oxford Nanopore Technologies (ONT) sequencing allows for both the full forward and full reverse strand of a double-stranded DNA (dsDNA) molecule to be sequenced and associated. A 1D read is the sequence of DNA bases derived from either the forward or reverse DNA strand. A 2D read is a consensus sequence derived from both the forward and the reverse reads.

are derived from the same original large fragment[62]. Similar to an earlier technology, BAC-by-BAC sequencing, synthetic barcoded reads provide an association among small fragments derived from a larger one. By segregating the fragments, repetitive or complicated regions can be isolated, allowing each to be assembled locally. This prevents unresolvable branch points in the assemblies, which lead to breaks (gaps) and shorter assembled contiguous sequences.

There are currently two systems available for generating synthetic long-reads: the Illumina synthetic long-read sequencing platform (FIG. 5c) and the 10X Genomics emulsion-based system (FIG. 5d). The Illumina system (formerly Moleculo) partitions DNA into a microtitre plate and does not require specialized instrumentation. However, the 10X Genomics instruments (GemCode and Chromium) use emulsion to partition DNA and require the use of a microfluidic instrument to perform pre-sequencing reactions. With as little as 1 ng of starting material, the 10X Genomics instruments can partition arbitrarily large DNA fragments, up to ~100 kb, into micelles called 'GEMs', which typically contain ≤0.3× copies of the genome and one unique barcode. Within each GEM, a gel bead dissolves and smaller fragments of DNA are amplified from the original large fragments, each with a barcode identifying the source GEM. After sequencing, the reads are aligned and linked together to form a series of anchored fragments across the span of the original fragment. Unlike the Illumina system, this approach does not attempt gapless, end-to-end coverage of a single DNA fragment. Instead it relies on linked reads, in which dispersed, small fragments that are derived from a single long molecule share a communal barcode. Although these fragments leave segments of the original large molecule without any coverage, the gaps are overcome by ensuring that there are many long fragments from the same genomic region in the initial preparation, thus generating a read cloud wherein linked reads from each long fragment can be stacked, combining their individual coverage into an overall map (FIG. 5d).

Comparison of single-molecule and synthetic long-read sequencing. There is growing interest in the field of long-read sequencing, and each system has its own advantages and drawbacks (TABLE 1). Currently, the most widely used instrument in long-read sequencing is the PacBio RS II instrument. This device is capable of generating single polymerase reads in excess of 50 kb with average read lengths of 10–15 kb for a long-insert library. Such properties are ideal for *de novo* genome assembly applications[63], for revealing complex long-range genomic structures[64] and for full-length transcript sequencing. There are, however, several notable limitations. The single-pass error rate for long reads is as high as 15% with indel errors dominating[65], raising concerns about the utility of the instrument[66]. Fortunately, these errors are randomly distributed within each read and hence sufficiently high coverage can overcome the high error rate[67]. The use of a circular template by PacBio also provides a level of error correction. The more frequently a single molecule is sequenced, the higher the resulting accuracy — up to ~99.999% for insert sequences derived from at least 10 subreads[59,68].This high accuracy rivals that of Sanger sequencing, leading researchers to speculate that this technology can be used in a manner analogous to Sanger-based SNP validation[65]. The runtimes and throughput of this instrument can be tuned by controlling the length of time for which the sensor monitors the ZMW; longer templates require longer times. For example, a 1 kb library that is run for 1 hour will generate around 7,500 bases of sequence per molecule, with an average of 8 passes, whereas a 4-hour run will generate around 30,000 bases per molecule and ~30 passes. Conversely, a 10 kb library requires a 4-hour run to generate ~30,000 bases with ~3 passes. The limited throughput and high costs of PacBio RS II (around $1,000 per Gb), in addition to the need for high coverage, place this instrument out of reach of many small laboratories. However, in an attempt to ameliorate these concerns, PacBio has launched the Sequel System, which reportedly has a throughput 7× that of the RS II, thus halving the cost of sequencing a human genome at 30× coverage[69].

The ONT MinION is a small (~3 cm × 10 cm for the MK1) USB-based device that runs off a personal computer, giving it the smallest footprint of any current sequencing platform. This affords the MinION superior portability, highlighting its utility for rapid clinical responses and hard-to-reach field locations. Although substantial adjunct equipment is still required for library preparation (for instance, a thermocycler), improvements in library preparation and equipment optimization could conceivably reduce the space required for a fully functional sequencing system to the size of a single bag of luggage. Unlike other platforms, the MinION has few constraints on the size of the fragments to be sequenced. In theory, a DNA molecule of any size can be sequenced on the device, but in practice there are some limitations when dealing with ultra-long fragments[70]. As a consequence of the unique nature of the ONT technology, in which there are more than 1,000 distinct signals, ONT MinION has a large error rate — up to 30% for a 1D read — and is dominated by indel errors. Effective homopolymer sequencing also remains a challenge for ONT MinION. When a homopolymer exceeds the k-mer length, it can be difficult to identify when one k-mer leaves the pore and another k-mer enters. Modified bases also pose a challenge to the device, as a modified base will alter the typical voltage shift for a given k-mer. Fortunately, recent improvements in the chemistry and the base calling algorithms are improving accuracy[71].

The Illumina synthetic long-read approaches are a direct response to the costs, error rates and throughput of true long-read sequencers. Relying on the existing Illumina infrastructure affords researchers the ability to simply purchase a kit for long-read sequencing. Accordingly, the throughput and error profile are identical to those of current Illumina devices. However, as a consequence of how the DNA is partitioned, the system requires more coverage than is required

© 2016 Macmillan Publishers Limited. All rights reserved.

BAC-by-BAC sequencing
A sequencing method where a physical map is generated from overlapping bacterial artificial chromosome (BAC) clones tiled across a chromosome. Each BAC is then fragmented and sequenced. The sequenced fragments are aligned with the knowledge of the originating BAC.

Linked reads
Reads derived from the 10X Genomics synthetic long-read platform. These are discontinuous reads each sharing the same barcode, thus they are derived from the same original long molecule.

Read cloud
The means by which the 10X Genomics platform determines a synthetic long read. Discontinuous linked reads from the same genomic region are aligned to each other. No single linked read contains the entire long sequence; however, when they are stacked, full coverage is achieved.

Polymerase reads
Contiguous sequences of nucleotides incorporated by the DNA polymerase while reading a template. These reads include sequences from adapters and can represent sequences from multiple passes around a circular template.

Single-pass
The single-molecule real-time (SMRT) sequencing approach from Pacific Biosciences (PacBio) enables a single molecule of DNA to be sequenced multiple times. A single pass is one single iteration through a molecule.

Subreads
The sequences derived from a single pass as a polymerase traverses a DNA molecule multiple times. A subread is trimmed to exclude any adapter sequence.

Whole-exome and targeted sequencing
Sequencing of only exons or other selected regions. A system of capture or amplification is used to isolate or enrich for only exons or target regions. This is done by designing probes or primers for the regions of interest.

for a typical short-read project, thus increasing the costs associated with this technology relative to other Illumina applications[62].

Like the Illumina synthetic long-read platform, the 10X Genomics emulsion-based platform relies on an existing short-read infrastructure to provide the sequencing. The microfluidic instrument is a one-time additional equipment cost, and the emulsion approach used allows for as little as 1 ng of starting material, which can be beneficial for situations in which the DNA is precious, such as biopsy samples. Currently, data output from the GemCode instrument is partially limited by the number of barcodes used and the somewhat inefficient DNA partitioning. Inefficient partitioning can lead to a surplus of DNA fragments within a droplet, thus complicating sequence deconvolution, which is further exacerbated by the limited number of barcodes. Both of these conditions lead to ambiguity regarding the positional relationship between reads sharing the same barcode, making analysis more difficult. To rectify this, 10X Genomics plans to release the Chromium System in mid-2016; this will be an upgrade from the GemCode device. Although the chemistry will remain fundamentally the same, the number of possible micelle partitions will increase from 100,000 to 1 million, and the number of barcodes will increase from 750,000 to ~4 million[72].

Applications

WGS is becoming one of the most widely used applications in NGS. Through this technology, researchers can obtain the most comprehensive view of genomic information and associated biological implications[73]. For example, in 2012, Ellis *et al.*[74] published an exploration of the interactions between genes and aromatase inhibitor therapy in patients with breast cancer. They outlined a range of correlations between mutations, outcomes and clinical features, as well as mutational enrichment in genes linked to other cancers, providing additional support for the idea that breast cancer is a highly complex pathology with variable phenotypes that are based on different repertoires of mutations[75]. However, the recent diversification of NGS platforms has revealed new and more ambitious opportunities that were not possible just a few years ago. In 2010, the 1000 Genomes Project released its initial results from WGS of 179 individuals and targeted sequencing of 697 individuals[76]. As of 2015, the genomes of 2,504 people from 26 different populations have been reconstructed[77,78], providing an unparalleled insight into human variation on a population level, and projects to sequence even larger sets of people are nearing completion or are underway[79–81]. Population-level sequencing is proving to be an essential tool in understanding human disease, with exciting results. In one example, Sidore *et al.*[82] performed WGS on 2,120 Sardinians and discovered new loci for lipid levels and inflammatory markers, providing greater insight into the mechanisms driving blood cholesterol levels.

Whole-exome and targeted sequencing[83] are also proving invaluable to sequencing research. By limiting the size of the genomic material used, more individual samples can be sequenced within a sequencing run, which can increase both the breadth and the depth of a genomic study. Using exome sequencing, Iossifov *et al.*[84] sequenced more than 2,500 simplex families, each with a child who was diagnosed with autism spectrum disorder (ASD), and they discovered *de novo* missense mutations, *de novo* gene-disrupting mutations and copy number variants in ~30% of all cases. This and other work regarding the classes of genes mutated is providing a framework for possible causes of ASD[85,86]. There is also increasing evidence that even high-coverage WGS is inadequate to resolve all variants in complex and/or limited material clinical samples. In 2015, Griffith *et al.*[87] demonstrated the use of an integrated, cross-platform approach (including targeted sequencing) to identify high-confidence SNPs from tumour samples. In this approach, the authors demonstrated that coverage as high as 10,000× can be required to validate rare variants. Although such high coverage is prohibitive for WGS, targeted sequencing approaches are uniquely suited for clinical applications.

NGS is also providing insight into the regulatory mechanisms of the genome. Protein–DNA interactions can be probed by enriching for protein-interacting DNA fragments, often through immunoprecipitation as in the case of ChIP–seq[41]. Conversely, ATAC–seq uses a hyperactive transposase to generate short-read NGS-compatible DNA fragments from regions unprotected by proteins or nucleosomes[42]. Analysis of modified bases is also possible. For example, methyl-seq involves the capture and enrichment of methylated DNA[88], selective digestion of methylated or unmethylated regions[89,90], and/or modification of a methylated base such that it introduces a SNP into the DNA sequence[91]. Although important discoveries have been made using these approaches, limitations exist, which are primarily derived from the modification or capture process. In 2010, Flusberg *et al.*[92] published a proof-of-concept study of using PacBio to discriminate between methylated and un-methylated bases, as well as between methylated adenine and methylated cytosine. As the polymerase attempts to elongate DNA containing modified bases, it pauses for longer at modified sites compared with unmodified controls, increasing a metric called the interpulse duration and thus indicating the presence of a modified base. Similarly, nanopore platforms also show promise for the direct detection of modified bases, as the characteristic shift in voltage across the pore is modulated by base modifications, allowing for discrimination without the need for chemical manipulations[93].

A recent paradigm shift in NGS is the ability to sequence very long stretches of DNA. Repetitive and complex regions have historically been difficult to assemble and resolve using short-read sequencing approaches[94–96]. Recently, using long-read sequencing technology, Chaisson *et al.*[97] were able to add more than 1 Mb of novel sequence to the human GRCh37 reference genome through gap closure and extension, and they identified >26,000 indels that were ≥50 bp in length, providing one of the most comprehensive reference genomes available. Beyond simply improving reference genomes, long reads are proving to be more

© 2016 Macmillan Publishers Limited. All rights reserved.

effective in identifying clinically relevant structural variation than short-read approaches[98]. Furthermore, synthetic approaches are enhancing the ability of researches to phase genomes[99]. In 2014, Kuleshov *et al.*[100] showed how synthetic approaches can reduce the coverage required for genome phasing[99] by 10-fold while also phasing up to 99% of all SNPs. This allows researchers to track the history of a mutation across generations — an essential tool for family studies.

Transcriptomic research has also benefited from greater accessibility to NGS. Today, researchers are leveraging the power of NGS to deeply sequence down to single-transcript sensitivity. In 2014, Treutlein *et al.*[101] demonstrated the power of single-cell RNA-seq to characterize different cell populations in developing tissue and to discover new markers for cell subpopulations. Although long-read approaches currently lack the ability to effectively measure transcript abundance levels, long reads provide superior performance for detecting transcriptomic structure[51]. For example, a recent long-read profile of the human transcriptome showed that >10% of the reads represented novel splice isoforms[102].

The newest instrument in the NGS landscape, the nanopore sequencer, is still in the process of finding its niche in the field. Nevertheless, researchers are capitalizing on its rapid library preparation time, real-time generation of data and its small size. Recently, researchers at the Stanley Royd Hospital in the United Kingdom used MinION sequencing to monitor an outbreak of *Salmonella enterica*. Using phylogenetic placement, the authors were able to unambiguously identify the serovar within 50 minutes after the start of sequencing, indicating that the MinION device is a viable platform for rapid pathogen profiling[103]. Perhaps one of the most striking applications of MinION sequencing in the field was its use during the 2014 Ebola outbreak, which is outlined in Quick *et al.*[104]. Under the auspices of the European Mobile Laboratories in Guinea, the authors were able to monitor the transmission history and evolution of the Ebola virus as the outbreak unfolded.

Closing remarks

We are sitting at the cusp of a new revolution in NGS technologies. Rather than being a novelty, NGS technologies are now a routine part of biological research. The advent of ultra-high-throughput sequencing is propelling research that was considered impossible only a few years ago and is becoming more widespread within the clinical sector. This includes recent precision medicine initiatives[105] and plans from Illumina to develop a pan-cancer screening method using circulating tumour DNA[106], each with goals of sequencing tens of thousands of genomes. Thus, rapid and low-cost sequencing is providing physicians with the tools needed to translate genomic information into clinically actionable results.

This revolution brings with it a new set of challenges. As NGS aims to be ubiquitous in the clinical setting, time remains a challenge. In cases of serious neonatal disease and aggressive cancer, the weeks it can take for WGS data generation and analysis can be the difference between success and failure. In the case of aggressive infections, that time is reduced to mere days. Although substantial advances have been made in reducing response time, most of the current systems do not yet generate enough data fast enough for a truly rapid response.

Whereas clinics may be challenged by a paucity of rapid data, other aspects of NGS are suffering from an abundance of data. To date, more than 14,000 genomes have been deposited within the US National Center for Biotechnology Information (NCBI) genome repositories, with new genomes deposited regularly. In 2013, Schatz and Langmead reported that the world can generate ~15 petabytes of sequencing data per year, and the number and throughput of sequencers has only increased since then[107]. This wealth of data is proving challenging for both analysis and infrastructure, requiring innovative storage and bioinformatic solutions[108]. Translation of this vast pool of genetic data into biological contexts remains a challenge, requiring both integrative approaches to NGS research and well-validated guidelines for discovery[87,109,110]. The proliferation of NGS in the clinical arena also raises concerns about the utility and ethics associated with genomic information. With so many genetic tests available, including direct-to-consumer testing, questions exist regarding the consumer response to genetic results and the impact of false positives and false negatives on health care[111,112].

Recently, Illumina's highly successful suite of instruments have been the juggernaut of NGS, relegating technologies that could not keep pace to niche applications or outright dissolution. The casualties of the NGS arms race have included the Helicos Genetic Analysis System[113], the Revolocity system from Complete Genomics and 454 pyrosequencing from Roche. Yet, despite these setbacks, new applications and instruments are being developed at an astounding pace. As Illumina's market share grows, so too do challenges to its dominance. The BGISEQ-500 and the forthcoming GenoCare[114] from Direct Genomics (based on the Helicos technology) seek to gain a foothold in Asia, where NGS research is still developing. Platforms such as the ONT PromethION[115] and Illumina HiSeq X are poised to push the limits of cost and yield. With the growing interest in clinical sequencing, established NGS providers are offering rapid solutions, such as the Ion Torrent S5 and the Illumina MiniSeq, and newcomers such as Qiagen's GeneReader[22] are also competing to fill the void. Finally, pre-sequencing approaches such as 10X Genomics aim to fundamentally change how existing sequencing is carried out by providing long-range information on short-read sequencing platforms.

In the next few years, additional players seek to further democratize the field with novel sequencing solutions. GenapSys (in partnership with Sigma-Aldrich), with its electronic 'lunchbox'-sized sequencer[116]; Genia (Roche), with a new nanopore sequencing approach[117]; and the recently announced Firefly (Illumina), with its one-channel CMOS technology[118], all claim to deliver superior cost and time savings for clinical applications. Finally, NanoString Technologies with its enzyme-free hybridization method[119], GnuBio (Bio-Rad) with a fluorescence resonance energy transfer (FRET)-based approach[120], and Electron Optica with an electron

Genome phasing
A method to identify which chromosome a DNA sequence is derived from. By examining polymorphisms, the chromosome of origin can be inferred by matching the reads that share the same variation.

Family studies
A study design in which many members of a family across several generations are sequenced. These studies are used to understand how phenotypes manifest within a particular genotype background.

Helicos Genetic Analysis System
A sequencing technology based on single nucleotide addition. Each nucleotide contains a 'virtual terminator' that prevents the incorporation of multiple nucleotides per cycle.

Fluorescence resonance energy transfer
(FRET; also known as Förster resonance energy transfer). A system in which energy can be transferred from one light-sensitive molecule to another. When the two molecules are in close proximity (≤30 nm), energy transferred between the two molecules modulates the intensity of a fluorescence signal.

YM95NC Attachment -- Page 105

© 2016 Macmillan Publishers Limited. All rights reserved.

microscopy-based system[121] all aim to revolutionize sequencing with unique technologies. These existing and forthcoming NGS tools have the potential to allow for revolutionary science, including direct sequencing of RNA or proteins, real-time genomic pathogen monitoring or precision medicine based on personal genome sequencing. These current and future advances make today an astonishing time for the field of NGS.

1. Watson, J. D. & Crick, F. H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123–131 (1953).
2. Mardis, E. R. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem. (Palo Alto Calif.)* **6**, 287–303 (2013).
This article provides a concise description of technological advancements supporting NGS.
3. Wetterstrand, K. A. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). *National Human Genome Research Institute* [online], http://www.genome.gov/sequencingcosts (updated 15 Jan 2016).
4. Kircher, M. & Kelso, J. High-throughput DNA sequencing — concepts and limitations. *Bioessays* **32**, 524–536 (2010).
5. Veritas Genetics. Veritas genetics launches $999 whole genome and sets new standard for genetic testing — Press Release. *Veritas Genetics* [online], https://www.veritasgenetics.com/documents/VG-launches-999-whole-genome.pdf (updated 4 Mar 2016).
6. Veritas Genetics. Veritas genetics breaks $1,000 whole genome barrier — Press Release. *Veritas Genetics* [online], https://www.veritasgenetics.com/documents/VG-PGP-Announcement-Final.pdf (29 Sep 2015).
7. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 251364 (2012).
8. Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl Acad. Sci. USA* **100**, 8817–8822 (2003).
9. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
10. Kim, J. B. *et al.* Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* **316**, 1481–1484 (2007).
11. Leamon, J. H. *et al.* A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis* **24**, 3769–3777 (2003).
12. Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* **34**, e22 (2006).
13. Harris, T. D. *et al.* Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106–109 (2008).
14. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
This paper describes the use of DNA nanoballs to achieve clonal amplification and the use of cPAL to achieve human genome sequencing as implemented by Complete Genomics (BGI).
15. Tomkinson, A. E., Vijayakumar, S., Pascal, J. M. & Ellenberger, T. DNA ligases: structure, reaction mechanism, and function. *Chem. Rev.* **106**, 687–699 (2006).
16. Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science* **241**, 1077–1080 (1988).
17. Valouev, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**, 1051–1063 (2008).
This paper describes the use of cleavable two-base-encoded probes to achieve genome-wide nucleosome mapping in *Caenorhabditis elegans*. This technology is implemented by Applied Biosystems (Thermo Fisher) for the SOLiD platform.
18. Metzker, M. L. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
19. Ju, J. *et al.* Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl Acad. Sci. USA* **103**, 19635–19640 (2006).
20. Guo, J. *et al.* Four-color DNA sequencing with 3′-*O*-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl Acad. Sci. USA* **105**, 9145–9150 (2008).
21. Timmerman, L. DNA sequencing market will exceed $20 billion, says Illumina CEO Jay Flatley. *Forbes* [online], http://www.forbes.com/sites/luketimmerman/2015/04/29/qa-with-jay-flatley-ceo-of-illumina-the-genomics-company-pursuing-a-20b-market/#4dbd19943bf5 (29 Apr 2015).
22. Karow, J. Qiagen launches GeneReader NGS System at AMP; presents performance evaluation by broad. *GenomeWeb* [online], https://www.genomeweb.com/molecular-diagnostics/qiagen-launches-genereader-ngs-system-amp-presents-performance-evaluation (4 Nov 2015).
23. Smith, D. R. & McKernan, K. Methods of producing and sequencing modified polynucleotides. US Patent 8058030 (2011).
24. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
This paper describes the development of the first NGS technology through the use of pyrosequencing. The authors demonstrate this method through sequencing of the *Mycoplasma genitalium* genome.
25. Rothberg, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352 (2011).
This paper describes the first non-optical sequencing technology using a massively parallel semi-conductor device to monitor H^+ release during DNA synthesis, as implemented by the Ion Torrent platform (Thermo Fisher). The authors demonstrate this technology by sequencing both bacterial and human DNA.
26. Rieber, N. *et al.* Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. *PLoS ONE* **8**, e66621 (2013).
27. Zook, J. M. *et al.* Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* **32**, 246–251 (2014).
28. Nothnagel, M. *et al.* Technology-specific error signatures in the 1000 Genomes Project data. *Hum. Genet.* **130**, 505–516 (2011).
29. Shen, Y. Sarin, S., Liu, Y., Hobert, O. & Pe'er, I. Comparing platforms for *C. elegans* mutant identification using high-throughput whole-genome sequencing. *PLoS ONE* **3**, e4012 (2008).
30. Chan, M. *et al.* Development of a next-generation sequencing method for *BRCA* mutation screening: a comparison between a high-throughput and a benchtop platform. *J. Mol. Diagnost.* **14**, 602–612 (2012).
31. Wall, J. D. *et al.* Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Res.* **24**, 1734–1739 (2014).
32. Harismendy, O. *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10**, R32 (2009).
33. BGI. Revolocity Whole Genome Sequencing Service — Press Release. *BGI* [online], http://u70g92ptbyk941g21dd41fc4.wpengine.netdna-cdn.com/wp-content/uploads/2015/10/Global-WGSRevolocity-ENG-10-15.pdf (2015).
34. Karow, J. BGI halts revolocity launch, cuts complete genomics staff as part of strategic shift. *GenomeWeb* [online], https://www.genomeweb.com/sequencing-technology/bgi-halts-revolocity-launch-cuts-complete-genomics-staff-part-strategic-shift (23 Nov 2015).
35. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
This paper demonstrates the use of reversible dye-terminator chemistry for human genome sequencing. This platform is used by the Illumina suite of platforms.
36. Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105 (2008).
37. Nakamura, K. *et al.* Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* **39**, e90 (2011).
38. Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* **12**, R112 (2011).
39. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
40. Sarin, S., Prabhu, S., O'Meara, M. M., Pe'er, I. & Hobert, O. *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nat. Methods* **5**, 865–867 (2008).
41. Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).
This review provides an overview of ChIP–seq methods for detecting chromatin–DNA interactions and their importance to epigenetics research.
42. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
43. Brunner, A. L. *et al.* Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res.* **19**, 1044–1056 (2009).
44. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
This review provides an overview of advances and challenges in techniques that are used in transcriptomic research with a specific focus in methods that use NGS technologies.
45. Wang, X. *et al.* A trimming-and-retrieving alignment scheme for reduced representation bisulfite sequencing. *Bioinformatics* **31**, 2040–2042 (2015).
46. Qiagen. Oncology insights enabled by knowledge base-guided panel design and the seamless workflow of the GeneReader NGS system — Press Release. *Qiagen* [online], http://www.genereadernngs.com/PROM-9192-001_1100403_WP_GeneReader_NGS_0116_NA.pdf (2016).
47. Forgetta, V. *et al.* Sequencing of the Dutch elm disease fungus genome using the Roche/454 GS-FLX Titanium System in a comparison of multiple genomics core facilities. *J. Biomol. Tech.* **24**, 39–49 (2013).
48. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).
49. GenomeWeb. Roche shutting down 454 sequencing business. *GenomeWeb* [online], https://www.genomeweb.com/sequencing/roche-shutting-down-454-sequencing-business (15 Oct 2015).
50. Malapelle, U. *et al.* Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *J. Clin. Pathol.* **68**, 64–68 (2015).
51. Li, S. *et al.* Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat. Biotechnol.* **32**, 915–925 (2014).
52. Life Technologies. Ion semiconductor sequencing uniquely enables both accurate long reads and paired-end sequencing. *Life Technologies* [online], https://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_098680.pdf (2011)
53. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).
54. McCarroll, S. A. & Altshuler, D. M. Copy-number variation and association studies of human disease. *Nat. Genet.* **39**, S37–S42 (2007).
55. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447**, 932–940 (2007).
56. Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437–455 (2010).
57. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
The authors describe the development of a real-time sequencing method using their zero-mode waveguide sensors as implemented by the Pacific Biosciences platform. The authors demonstrate the technique by sequencing synthetic DNA templates.

© 2016 Macmillan Publishers Limited. All rights reserved.

58. Levene, M. J. *et al.* Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686 (2003).
59. Loomis, E. W. *et al.* Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res.* **23**, 121–128 (2013).
60. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
The authors demonstrate the use of a mutant alpha-hemolysin for ordered, continuous detection of free nucleotides in solution. This work provides the basis for the approach used by ONT.
61. Voskoboynik, A. *et al.* The genome sequence of the colonial chordate, *Botryllus schlosseri*. *eLife* **2**, e00569 (2013).
62. McCoy, R. C. *et al.* Illumina TruSeq synthetic long-reads empower *de novo* assembly and resolve complex, highly-repetitive transposable elements. *PLoS ONE* **9**, e106689 (2014).
63. Schatz, M. C., Delcher, A. L. & Salzberg, S. L. Assembly of large genomes using second-generation sequencing. *Genome Res.* **20**, 1165–1173 (2010).
64. English, A. C. *et al.* Assessing structural variation in a personal genome-towards a human reference diploid genome. *BMC Genomics* **16**, 286 (2015).
65. Carneiro, M. O. *et al.* Pacific Biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* **13**, 375 (2012).
66. Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).
67. Koren, S. *et al.* Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).
68. Larsen, P. A., Heilman, A. M. & Yoder, A. D. The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. *BMC Genomics* **15**, 720 (2014).
69. Heger, M. PacBio launches higher-throughput, lower-cost single-molecule sequencing system. *GenomeWeb* [online], https://www.genomeweb.com/business-news/pacbio-launches-higher-throughput-lower-cost-single-molecule-sequencing-system (01 Oct 2015).
70. Goodwin, S. *et al.* Oxford Nanopore sequencing, hybrid error correction, and *de novo* assembly of a eukaryotic genome. *Genome Res.* **25**, 1750–1756 (2015).
71. Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **12**, 351–356 (2015).
72. Heger, M. 10X Genomics, Pacific Biosciences provide business updates at JP Morgan Healthcare Conference. *GenomeWeb* [online], https://www.genomeweb.com/sequencing-technology/10x-genomics-pacific-biosciences-provide-business-updates-jp-morgan-healthcare (13 Jan 2016).
73. Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **11**, 415–425 (2010).
This review provides a comprehensive overview of advances in, and challenges of using, WGS for variant discovery in human disease.
74. Ellis, M. J. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353–360 (2012).
75. Prat, A. & Perou, C. M. Mammary development meets cancer genomics. *Nat. Med.* **15**, 842–844 (2009).
76. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
77. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
78. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
79. UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
80. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
81. Regalado, A. U.S. to develop DNA study of one million people. *MIT Technology Review* [online], http://www.technologyreview.com/news/534591/us-to-develop-dna-study-of-one-million-people (30 Jan 2015).
82. Sidore, C. *et al.* Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.* **47**, 1272–1281 (2015).
83. Hodges, E. *et al.* Genome-wide *in situ* exon capture for selective resequencing. *Nat. Genet.* **39**, 1522–1527 (2015).
This paper describes the *in situ* capture and selective enrichment of human exons for downstream NGS. This manuscript provides the methodological basis for whole-exome and targeted sequencing.
84. Iossifov, I. *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
85. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
86. Sanders, S. J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
87. Griffith, M. *et al.* Optimizing cancer genome sequencing and analysis. *Cell Syst.* **1**, 210–223 (2015).
88. Rauch, C. *et al.* Towards an understanding of DNA recognition by the methyl-CpG binding domain 1. *J. Biomol. Struct. Dyn.* **22**, 695–706 (2005).
89. Oda, M. *et al.* High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res.* **37**, 3829–3839 (2009).
90. Irizarry, R. A. *et al.* Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* **18**, 780–790 (2008).
91. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
92. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
93. Wescoe, Z. L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J. Am. Chem. Soc.* **136**, 16582–16587 (2014).
94. Lam, E. T. *et al.* Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **30**, 771–776 (2012).
95. Eichler, E. E., Clark, R. A. & She, X. An assessment of the sequence gaps: unfinished business in a finished human genome. *Nat. Rev. Genet.* **5**, 345–354 (2004).
96. Chaisson, M. J., Wilson, R. K. & Eichler, E. E. Genetic variation and the *de novo* assembly of human genomes. *Nat. Rev. Genet.* **16**, 627–640 (2015).
97. Chaisson, M. J. *et al.* Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608–611 (2015).
This article provides strong support for the utility of long-read sequencing for generating high-quality reference genomes. The authors demonstrate this by closing and/or extending gaps and resolving structural variants in the GRCh37 human reference genome.
98. Ritz, A. *et al.* Characterization of structural variants with single molecule and hybrid sequencing approaches. *Bioinformatics* **30**, 3458–3466 (2014).
99. Snyder, M. W., Adey, A., Kitzman, J. O. & Shendure, J. Haplotype-resolved genome sequencing: experimental methods and applications. *Nat. Rev. Genet.* **16**, 344–358 (2015).
100. Kuleshov, V. *et al.* Whole-genome haplotyping using long reads and statistical methods. *Nat. Biotechnol.* **32**, 261–266 (2014).
101. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
102. Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* **31**, 1009–1014 (2013).
103. Quick, J. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16**, 114 (2015).
104. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
105. GenomeWeb. White House announces efforts to accelerate precision medicine initiative. *GenomeWeb* [online], https://www.genomeweb.com/molecular-diagnostics/white-house-announces-efforts-accelerate-precision-medicine-initiative (25 Feb 2016).
106. Illumina. Illumina forms new company to enable early cancer detection via blood-based screening — Press Release. *Illumina* [online], http://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=2127903 (10 Jan 2016).
107. Schatz, M. C. & Langmead, B. The DNA data deluge: fast, efficient genome sequencing machines are spewing out more data than geneticists can analyze. *IEEE Spectr.* **50**, 26–33 (2013).
108. Pop, M. & Salzberg, S. L. Bioinformatics challenges of new sequencing technology. *Trends Genet.* **24**, 142–149 (2008).
109. Sunyaev, S. R. Inferring causality and functional significance of human coding DNA variants. *Hum. Mol. Genet.* **21**, R10–R17 (2012).
110. Gargis, A. S. *et al.* Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.* **30**, 1033–1036 (2012).
111. Chrystoja, C. C. & Diamandis, E. P. Whole genome sequencing as a diagnostic test: challenges and opportunities. *Clin. Chem.* **60**, 724–733 (2014).
112. McGuire, A. L. *et al.* Point-counterpoint. Ethics and genomic incidental findings. *Science* **340**, 1047–1048 (2013).
113. Bowers, J. *et al.* Virtual terminator nucleotides for next-generation DNA sequencing. *Nat. Methods* **6**, 593–595 (2009).
114. Heger, M. China's Direct Genomics unveils new targeted NGS system based on Helicos Tech for clinical use. *GenomeWeb* [online], https://www.genomeweb.com/business-news/chinas-direct-genomics-unveils-new-targeted-ngs-system-based-helicos-tech-clinical-use (27 Oct 2015).
115. Karow, J. Oxford Nanopore presents details on new high-throughput sequencer, improvements to MinIon. *GenomeWeb* [online], https://www.genomeweb.com/sequencing/oxford-nanopore-presents-details-new-high-throughput-sequencer-improvements-mini (16 Sep 2014).
116. Karow, J. Sigma-Aldrich enters co-marketing agreement with GenapSys for Genius sequencer. *GenomeWeb* [online], https://www.genomeweb.com/sequencing-technology/sigma-aldrich-enters-co-marketing-agreement-genapsys-genius-sequencer (1 Jul 2015).
117. Roche. Roche acquires Genia Technologies to strengthen next generation sequencing pipeline — Press Release. *Roche* [online], http://www.roche.com/media/store/releases/med-cor-2014-06-02.htm (2 Jun 2014).
118. Heger, M. Illumina unveils mini targeted sequencer, semiconductor sequencing project at JP Morgan Conference. *GenomeWeb* [online], https://www.genomeweb.com/sequencing-technology/illumina-unveils-mini-targeted-sequencer-semiconductor-sequencing-project-jp (1 Jan 2016).
119. NanoString. NanoString Technologies presents proof-of-concept data for novel massively parallel single molecule sequencing chemistry at AGBT meeting — Press Release. *NanoString* [online], http://investors.nanostring.com/releasedetail.cfm?ReleaseID=954517 (11 Feb 2016).
120. Raz, T. & Pascaline, M. Nucleic acid target detection using a detector, a probe and an inhibitor. US Patent 20130344485 (2013).
121. Mankos, M. *et al.* A novel low energy electron microscope for DNA sequencing and surface analysis. *Ultramicroscopy* **145**, 36–49 (2014).
122. Augenlicht, L. H. & Kobrin, D. Cloning and screening of sequences expressed in a mouse colon tumor. *Cancer Res.* **42**, 1088–1093 (1982).
123. Dandy, D. S., Wu, P. & Grainger, D. W. Array feature size influences nucleic acid surface capture in DNA microarrays. *Proc. Natl Acad. Sci. USA* **104**, 8223–8228 (2007).
124. Keating, B. J. *et al.* Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS ONE* **3**, e3583 (2008).
125. DeRisi, J. *et al.* Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**, 457–460 (1996).
126. Alizadeh, A. A. & Staudt, L. M. Genomic-scale gene expression profiling of normal and malignant immune cells. *Curr. Opin. Immunol.* **12**, 219–225 (2000).
127. Rhodes, D. R. *et al.* Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc. Natl Acad. Sci. USA* **101**, 9309–9314 (2004).

YM95NC Attachment -- Page 107

© 2016 Macmillan Publishers Limited. All rights reserved.

128. Vora, G. J., Meador, C. E., Stenger, D. A. & Andreadis, J. D. Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl. Environ. Microbiol.* **70**, 3047–3054 (2004).
129. Wilson, W. J. *et al.* Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol. Cell Probes* **16**, 119–127 (2002).
130. Imai, K., Kricka, L. J. & Fortina, P. Concordance study of 3 direct-to-consumer genetic-testing services. *Clin. Chem.* **57**, 518–521 (2011).
131. Dolgin, E. Personalized investigation. *Nat. Med.* **16**, 953–955 (2010).
132. Jia, P., Wang, L., Meltzer, H. Y. & Zhao, Z. Common variants conferring risk of schizophrenia: a pathway analysis of GWAS data. *Schizophr. Res.* **122**, 38–42 (2010).
133. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
134. Carter, N. P. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.* **39**, S16–S21 (2007).
135. Vrijenhoek, T. *et al.* Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am. J. Hum. Genet.* **83**, 504–510 (2008).
136. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
137. Buck, M. J. & Lieb, J. D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360 (2004).
138. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.* **23**, 716–726 (2013).
139. Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE* **9**, e78644 (2014).
140. Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5′→3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl Acad. Sci. USA* **88**, 7276–7280 (1991).
141. Morin, P. A. & McCarthy, M. Highly accurate SNP genotyping from historical and low-quality samples. *Mol. Ecol. Notes* **7**, 937–946 (2007).
142. VanGuilder, H. D., Vrana, K. E. & Freeman, W. M. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44**, 619–626 (2008).
143. Weaver, S. *et al.* Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods* **50**, 271–276 (2010).
144. Sedlak, R. H., Cook, L., Cheng, A., Magaret, A. & Jerome, K. R. Clinical utility of droplet digital PCR for human cytomegalovirus. *J. Clin. Microbiol.* **52**, 2844–2848 (2014).
145. Kulkarni, M. M. in *Current Protocols in Molecular Biology* Ch. 25 (eds Ausubel, F. M. *et al.*) (Wiley, 2011).
146. Nielsen, T. *et al.* Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer* **14**, 177 (2014).
147. Ku, B. M. *et al.* High-throughput profiling identifies clinically actionable mutations in salivary duct carcinoma. *J. Transl. Med.* **12**, 299 (2014).
148. Sailani, M. R. *et al.* The complex SNP and CNV genetic architecture of the increased risk of congenital heart defects in Down syndrome. *Genome Res.* **23**, 1410–1421 (2013).
149. Lira, M. E. *et al.* Multiplexed gene expression and fusion transcript analysis to detect ALK fusions in lung cancer. *J. Mol. Diagn.* **15**, 51–61 (2013).
150. Schwartz, D. C. *et al.* Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science* **262**, 110–114 (1993).
151. Hastie, A. R. *et al.* Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* genome. *PLoS ONE* **8**, e55864 (2013).
152. Cao, H. *et al.* Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *Gigascience* **3**, 34 (2014).
153. Pendleton, M. *et al.* Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods* **12**, 780–786 (2015).
154. Life Technologies. 5500W series genetic analyzers. *Life Technologies* [online], https://tools.thermofisher.com/content/sfs/brochures/5500-w-series-spec-sheet.pdf (2012).
155. Yuzuki, D. BGISEQ-500 debuts at the International Congress of Genomics 10. *Next Generation Technologist* [online], http://www.yuzuki.org/bgiseq-500-debut-at-the-international-congress-of-genomics-10 (24 Oct 2015).
156. Winnick, E. Illumina launches four new systems; provides financial, Dx update at JP Morgan. *GenomeWeb* [online], https://www.genomeweb.com/business-news/illumina-launches-four-new-systems-provides-financial-dx-update-jp-morgan (12 Jan 2015).
157. [No authors listed.] Illumina HiSeq 3000 Service Fees. *Oregon State University* [online], http://cgrb.oregonstate.edu/core/illumina-hiseq-3000/illumina-hiseq-3000-service-fees (updated 1 Jan 2016)
158. Heger, M. Thermo Fisher launches new systems to focus on plug and play targeted sequencing. *GenomeWeb* [online], https://www.genomeweb.com/sequencing-technology/thermo-fisher-launches-new-systems-focus-plug-and-play-targeted-sequencing (1 Sep 2015).
159. Ip, C. L. *et al.* MinION analysis and reference consortium: Phase 1 data release and analysis. *F1000Research* **4**, 1075 (2015).
160. Glenn, T. C. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* **11**, 759–769 (2011).
161. Karow, J. At AGBT, 10X Genomics launches GemCode platform; shipments slated for Q2 as firm battles IP lawsuits. *GenomeWeb* [online], https://www.genomeweb.com/sample-prep/agbt-10x-genomics-launches-gemcode-platform-shipments-slated-q2-firm-battles-ip-lawsuits (2 Mar 2015).

Competing interests statement
The authors declare competing interests: see Web version for details.

FURTHER INFORMATION

10X Genomics: http://www.10xgenomics.com
454 Sequencing: http://www.454.com
Advances in Genome Biology and Technology (AGBT): http://www.agbt.org
BGISEQ-500: http://seq500.com/en/portal/Sequencer.shtml
Illumina: http://www.illumina.com
Ion Torrent: https://www.thermofisher.com/us/en/home/brands/ion-torrent.html
Oxford Nanopore Technologies: https://www.nanoporetech.com
Pacific Biosciences: http://www.pacb.com
Personal Genome Project: http://www.personalgenomes.org
SOLiD Next-Generation Sequencing: http://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing.html
The 1000 Genomes Project: http://www.1000genomes.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2016 Macmillan Publishers Limited. All rights reserved.

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 278, No. 37, Issue of September 12, pp. 34998–35015, 2003
Printed in U.S.A.

Genome-wide Expression Profiling of the Response to Polyene, Pyrimidine, Azole, and Echinocandin Antifungal Agents in *Saccharomyces cerevisiae**[S]

Received for publication, June 15, 2003
Published, JBC Papers in Press, June 24, 2003, DOI 10.1074/jbc.M306291200

Ameeta K. Agarwal‡§, P. David Rogers¶, Scott R. Baerson‖, Melissa R. Jacob‡, Katherine S. Barker¶, John D. Cleary, Larry A. Walker‡ ‡‡, Dale G. Nagle§§¶¶, and Alice M. Clark‡§§**

*From the ‡National Center for Natural Products Research, §§Department of Pharmacognosy, the ‡‡Department of Pharmacology, and ¶¶Research Institute of Pharmaceutical Sciences, University of Mississippi School of Pharmacy, University, Mississippi 38677, the ¶Departments of Pharmacy, Pharmaceutical Sciences and Pediatrics, Colleges of Pharmacy and Medicine, University of Tennessee Health Science Center, Memphis, Tennessee 38163, the ‖United States Department of Agriculture, Agricultural Research Service, Natural Products Utilization Research Unit, University, Mississippi 38677, and the **Departments of Clinical Pharmacy Practice and Medicine, Schools of Pharmacy and Medicine, University of Mississippi Medical Center, Jackson, Mississippi 39216*

Antifungal compounds exert their activity through a variety of mechanisms, some of which are poorly understood. Novel approaches to characterize the mechanism of action of antifungal agents will be of great use in the antifungal drug development process. The aim of the present study was to investigate the changes in the gene expression profile of *Saccharomyces cerevisiae* following exposure to representatives of the four currently available classes of antifungal agents used in the management of systemic fungal infections. Microarray analysis indicated differential expression of 0.8, 4.1, 3.0, and 2.6% of the genes represented on the Affymetrix S98 yeast gene array in response to ketoconazole, amphotericin B, caspofungin, and 5-fluorocytosine (5-FC), respectively. Quantitative real time reverse transcriptase-PCR was used to confirm the microarray analyses. Genes responsive to ketoconazole, caspofungin, and 5-FC were indicative of the drug-specific effects. Ketoconazole exposure primarily affected genes involved in ergosterol biosynthesis and sterol uptake; caspofungin exposure affected genes involved in cell wall integrity; and 5-FC affected genes involved in DNA and protein synthesis, DNA damage repair, and cell cycle control. In contrast, amphotericin B elicited changes in gene expression reflecting cell stress, membrane reconstruction, transport, phosphate uptake, and cell wall integrity. Genes with the greatest specificity for a particular drug were grouped together as drug-specific genes, whereas genes with a lack of drug specificity were also identified. Taken together, these data shed new light on the mechanisms of action of these classes of antifungal agents and demonstrate the potential utility of gene expression profiling in antifungal drug development.

Invasive fungal infections cause significant morbidity and mortality in susceptible patient populations (1–4). Currently available antifungal agents are few and have limitations with regard to efficacy and toxicity (5). The need for alternative strategies in the management of fungal infections has increased with the emergence of new fungal pathogens and the development of antifungal resistance (6, 7). Novel approaches to identify unique antifungal drug targets are needed. Furthermore, a greater understanding of the mechanisms of action of available antifungal agents would aid in this effort.

Currently only four classes of antifungal agents are used to treat invasive fungal disease. Amphotericin B, the only polyene effective systemically, has been in use for over 40 years. It exerts its activity by binding to ergosterol in the fungal cell membrane, compromising membrane integrity, and ultimately leading to cell death (8, 9). The antimetabolite, 5-fluorocytosine (5-FC),[1] acts by interfering with nucleotide metabolism and inhibiting RNA, DNA, and protein synthesis (10). The more recently introduced azole antifungal agents exert their activity by inhibiting the biosynthesis of ergosterol, the major membrane sterol of fungi. The azoles inhibit the cytochrome P450 enzyme, lanosterol demethylase, by binding to the heme in the active site of the enzyme (6, 11). The newest class of antifungal agents used in the management of systemic fungal infections is the echinocandins. Caspofungin is currently the only commercially available agent in this class. It exerts its activity by inhibiting the synthesis of β-1,3-glucan in the fungal cell wall (12).

One of the ways in which cells adjust to changes in their environment is by altering gene expression patterns. Thus, measurement of changes in gene expression upon exposure to a drug can help determine how drugs and drug candidates work in cells and organisms. DNA microarray technology is a powerful tool to measure gene expression on a genomic scale, allowing simultaneous measurement of changes in the expression of thousands of genes. In the past few years, this technology has been used to discover gene function, understand biochemical pathways and regulatory mechanisms, classify dis-

* This work was supported in part by a grant from the United States Public Health Service, NIAID Grant R01 AI27094 from the National Institutes of Health, the United States Department of Agriculture, Agricultural Research Service Specific Cooperative Agreement 58-6408-2-0009, and a grant from Merck. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains Tables A–D.

§ To whom correspondence should be addressed: National Center for Natural Products Research, University, MS 38677. Tel.: 662-915-1218; Fax: 662-915-7062; E-mail: aagarwal@olemiss.edu.

[1] The abbreviations used are: 5-FC, 5-fluorocytosine; RT, reverse transcriptase; CFU, colony-forming units; DI, difference index; MOPS, 3-(*N*-morpholino)propanesulfonic acid; 5-FU, 5-fluorouracil; MMS, methylmethane sulfonate; MIC, minimum inhibitory concentration; MAP, mitogen-activated protein.

This paper is available on line at http://www.jbc.org
YM95NC Attachment -- Page 109

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

ease specimens, and discover drug targets (reviewed in Ref. 13). The yeast *Saccharomyces cerevisiae* is an excellent model organism for the study of antifungal action. Its genome has been fully sequenced and well characterized, and with the development of whole-genome microarrays, it is now possible to monitor globally gene expression changes in response to various experimental conditions. The availability of deletion mutant stocks for *S. cerevisiae* also greatly facilitates the validation of novel hypotheses generated from microarray experiments.[2] In addition, among the pathogenic fungi, *S. cerevisiae* is most closely related to *Candida albicans*, the major opportunistic fungal pathogen (14). In recent years, *S. cerevisiae*, previously considered to be non-pathogenic, has been reported as an emerging fungal pathogen in immunocompromised individuals (15). Thus, *S. cerevisiae* serves as a valuable model system in the study of fungal infections and identification of new drug targets.

Genomic profiling studies have examined the effects of amphotericin B, 5-FC, and various azoles in both *S. cerevisiae* and *C. albicans* (16–19). These studies provide an excellent overview of the genes expected to change their expression in response to treatment with these drugs. However, these studies have used concentrations that may not be optimal for appropriate between-drug comparisons with regard to their effects. Furthermore, we show here that differences in experimental conditions such as growth media, concentration of drug, and period of drug exposure can contribute to gene expression changes. Because the experimental conditions utilized in our study are different from those used in previous studies, we have been able to identify responses that were not previously reported. In addition, the minimal media conditions used in the current study perhaps better mimic the nutrient-limited environment of macrophages, which phagocytose *C. albicans* cells and allow hyphal development in this pathogen (20). In the present study, genome-wide gene expression profiling was used to compare and characterize changes in *S. cerevisiae* gene expression in response to representatives of four major classes of antifungal agents in an effort to identify drug class-specific and mechanism-independent changes in gene expression.

EXPERIMENTAL PROCEDURES

Antifungal Agents—Ketoconazole and 5-fluorocytosine were obtained from Sigma. Amphotericin B was from ICN Biomedicals (Aurora, OH). A commercially available preparation of caspofungin acetate for injection (Cancidas®) was obtained from a local pharmacy. Stock solutions of varying concentrations were made in Me_2SO (Sigma).

Yeast Strains and Media—S. cerevisiae S288C (*MATα, SUC2, mal, mel, gal2, CUP1, flo1, flo8-1*), obtained from ATCC (Manassas, VA), was used in the microarray experiments. The *upc2-1* mutant strain (SCY1012, *MATα, ade2-1, can 1-1, trp1-1, ura3-1, his3-11, 15, leu2-3, 112, upc2-1 lo)* and its isogenic wild type strain (SCY325) were generously provided by Dr. Stephen Sturley (Columbia University, New York). The *upc2*Δ strain (*MATα*, *leu2*, *ura3*, *his3*, *lys2, upc2*Δ) and its isogenic wild type strain (BY4742) were obtained from Invitrogen. Synthetic dextrose (SD) medium, containing 0.67% (w/v) yeast nitrogen base without amino acids and 2% (w/v) dextrose, was used to grow the wild type S288C strain. Synthetic complete (SC) medium, consisting of SD medium supplemented with complete supplement mixture (CSM, Qbiogene, Inc., Carlsbad, CA), was used to grow the *upc2-1* mutant strain, the *upc2*Δ strain, and their respective isogenic wild type controls. In all cases, the medium was buffered with 0.165 M MOPS (Sigma), and the pH was adjusted to 7.0 with NaOH.

IC_{50} Determinations—For IC_{50} determinations in small scale cultures, a single colony of *S. cerevisiae* was resuspended in 0.9% sterile saline. The culture was diluted in SD medium after comparison to the 0.5 McFarland standard to afford a final inoculum of $\sim 3 \times 10^6$ CFU/ml. After dilution, 175 μl of the culture was inoculated into a 96-well flat bottom culture plate (Corning Inc., Corning, NY) containing 25 μl of test compounds that were 2-fold serially diluted in 0.9% sterile saline. The starting concentrations used were 10 μg/ml for amphotericin B and caspofungin, 50 μg/ml for 5-fluorocytosine, and 100 μg/ml for ketoconazole. The culture plate was read at 630 nm prior to and after incubation (30 °C for 17 h) using an EL340 microplate reader (Bio-Tek Instruments, Inc., Winooski, VT).

For determination of IC_{50} values in large scale cultures, an overnight culture (started from a single colony) in late exponential phase was used to inoculate 100 ml of SD medium to an A_{600} of 0.1. Multiple cultures were started for each drug in order to test 4–5 different drug concentrations. The cultures were incubated at 30 °C in an environmental shaker (200 rpm) and allowed to recover from stationary phase until an A_{600} of 0.2 was reached. Drug was then added at 2–4-fold serial dilutions into each culture. Two rounds of experiments were conducted, with a broad range of drug concentrations tested in the first round and a narrow range tested in duplicates in the second round. The cultures were grown for 17 h at 30 °C, and a final A_{600} was measured using an Ultrospec 2000 spectrophotometer (Amersham Biosciences).

Cell Culture and Drug Exposure for Microarray Experiments—A single colony of *S. cerevisiae* was inoculated into 25 ml of SD medium and grown overnight at 30 °C in an environmental shaker (200 rpm) until late exponential phase. The culture was used to inoculate 100 ml of SD medium to an A_{600} of 0.1. Two independent 100-ml cultures were grown for each drug. All drug exposures were performed on the same day using the same starting culture to minimize experimental variations. The cultures were incubated at 30 °C in an environmental shaker (200 rpm) and allowed to recover from stationary phase until an A_{600} of 0.2 was reached. Drug was then added to each culture at a concentration equivalent to the IC_{50} value (concentrations used were 0.12 μg/ml amphotericin B, 0.02 μg/ml caspofungin, 0.3 μg/ml 5-fluorocytosine, and 56 μg/ml ketoconazole). Control cultures were simultaneously treated with 0.25% Me_2SO. The cultures were allowed to grow until an A_{600} of 0.5 was reached (~3 h, the doubling time in SD medium was ~2 h). Cells were harvested by centrifugation at 500 × *g* at 20 °C for 5 min in a Sorvall RC5C Plus centrifuge using an SH-3000 rotor. The cells were flash-frozen in liquid nitrogen and stored at −80 °C.

RNA Preparation—Total RNA was isolated using the Qiagen RNeasy® Midi-kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's instructions with the following modifications: frozen *S. cerevisiae* cells were ground into a powder with a mortar and pestle in liquid nitrogen to facilitate cell disruption. The cell powder was resuspended in 4 ml of lysis buffer provided in the kit and homogenized for 2 min at 30-s intervals at 25,000 rpm using a PT3100 Polytron (Brinkman Instruments). The RNA concentration and purity were determined spectrophotometrically by measuring absorbance at 230, 260, 280, and 320 nm. The purity and integrity of the RNA were confirmed by agarose gel electrophoresis.

Microarray Hybridization and Analysis—Microarray hybridization was performed using the Affymetrix GeneChip® Yeast Genome S98 Array using protocols described by Affymetrix, Inc. (Santa Clara, CA). Data were analyzed using Affymetrix® Microarray Suite 5.0 software. Genes were considered to be differentially expressed if (i) expression changed by at least 2-fold in two independent experiments performed with duplicate RNA samples, (ii) the mRNAs were assigned at least one "present" detection call by the Affymetrix software in both experiments, and (iii) the change in gene expression was in the same direction ("increased" or "decreased") in both experiments. Genes were annotated using the *Saccharomyces* Genome Data Base[3] and Yeast Proteome Data Base (Incyte Corp., Palo Alto, CA). Drug-specific responses were identified by calculating the difference index (DI) described by Nau *et al.* (22) as follows: DI = $\log_2$(fold change of gene *X* in drug exposure A) − $\log_2$(average fold change of gene *X* in all four drug exposures).

Quantitative Real Time RT-PCR Assays—An aliquot of the RNA preparations from untreated and treated samples, used in the microarray experiments, was saved for quantitative real time RT-PCR follow-up studies. The RNA samples were treated with DNase I "on column" as per manufacturer's instructions using Qiagen RNeasy® (Qiagen, Inc., Valencia, CA) columns to remove residual DNA contamination. First strand cDNAs were synthesized from 2 μg of total RNA in a 100-μl reaction volume using the TaqMan Reverse Transcription

[2] Saccharomyces Genome Deletion Project Home Page (www.sequence.stanford.edu/group/yeast_deletion_project) (accessed June, 2003).

[3] K. Dolinski, R. Balakrishnan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, L. Issel-Tarver, A. Sethuraman, C. L. Theesfeld, G. Binkley, C. Lane, M. Schroeder, S. Dong, S. Weng, R. Andrada, D. Botstein, and J. M. Cherry, Saccharomyces Genome Data Base (www.yeastgenome.org) (accessed June, 2003).

YM95NC Attachment -- Page 110

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

TABLE I
Gene-specific primers used for real time RT-PCR assays

Gene	ORF	Primer pairs	Position[a]	Amplicon size
				bp
ERG11	YHR007C	F, 5′-AATGCTCACCAATTCAACATTCAC-3′	1264	50
		R, 5′-AGAGGAGGCAGAATCTTTGTTCC-3′	1314	
ERG26	YGL001C	F, 5′-TGGACTTATGCTGGAAATGTTGC-3′	646	51
		R, 5′-GTTTCTGTGCAGCTAACACATGC-3′	697	
UPC2	YDR213W	F, 5′-TTCGAGGACTGAAACTGGACTG-3′	1983	50
		R, 5′-TCTAGGCGGTGAGATGAAACG-3′	2033	
DAN1	YJR150C	F, 5′-GTCTCCTCCAAAGCTCAATCTACAG-3′	751	50
		R, 5′-AGATGCTGACGATGTGACAGAAG-3′	801	
AUS1	YOR011W	F, 5′-TTTGCTGGTTACTACTTAACTTACGTGG-3′	4081	49
		R, 5′-AAGACCTTTGGCCATATTCTTGC-3′	4130	
GIT1	YCR098C	F, 5′-ATGCTGGACCAGGTGATATGC-3′	1133	53
		R, 5′-CGGTTGCTGACGCTTCACTAC-3′	1186	
TIS11	YLR136C	F, 5′-TCAAACATGGCGACGACAAC-3′	689	50
		R, 5′-AAGTACCAGCTTTCACGTAAACAGC-3′	739	
HXT5	YHR096C	F, 5′-GAGTCGTATCCCTTAAGAGTACGTGG-3′	1423	49
		R, 5′-GCACTTGCAATCGACATTGC-3′	1472	
YPS3	YLR121C	F, 5′-GCTGTAGATCACAGCAAGTACGAG-3′	688	50
		R, 5′-AACCAACGGGATAGTGTACAGTTG-3′	738	
PHM8	YER037W	F, 5′-AGTGGGTTGTCGAGCTTTGC-3′	649	50
		R, 5′-GCTTTCGTTGTCATCAATAAACCAG-3′	699	
GSC2	YGR032W	F, 5′-TCCATAACTTGGTTCAACCAAGG-3′	5567	47
		R, 5′-ACCCTGTATCATTGTTAGATACGTTCC-3′	5614	
MTL1	YGR023W	F, 5′-TCAGAATTTTGGTGATATCGATCC-3′	1317	50
		R, 5′-GGGTTGTCATTGTTAAGTATATCTTCTGG-3′	1367	
BAG7	YOR134W	F, 5′-TGCTTCCAAGGTCACGTGC-3′	908	50
		R, 5′-GATGTATGGTAAAATTGTTGGAGTCAG-3′	958	
SLT2	YHR030C	F, 5′-TGGAGATGAGACCTGCCACTG-3′	1340	50
		R, 5′-CATTCTGAGGCGGAATATCTGC-3′	1390	
RLM1	YPL089C	F, 5′-AAATACCGGGCTGACTCCATAC-3′	1923	50
		R, 5′-GTGCCTAGTGGTGTTTGAGCAG-3′	1973	
RNR3	YIL066C	F, 5′-GAAAGCTCAGATAATGAGAAGGATGC-3′	2389	50
		R, 5′-TGATTGTTCGGATGGAACTGG-3′	2439	
RFA2	YNL312W	F, 5′-TGGAATGATGAAACAACCTTTGG-3′	534	51
		R, 5′-TGATTGCCCATTGTTGTTGG-3′	585	
HO	YDL227C	F, 5′-ATTACGCCCGAAGGTTTGC-3′	1636	50
		R, 5′-CGCCACATTTTATACACTCTGGTC-3′	1687	
DSE4	YNR067C	F, 5′-TGACCAGTCAATGGAATTGAGG-3′	2886	51
		R, 5′-ATCCTTCATTATACTAATCATCAAGTCACC-3′	2937	
PCL1	YNL289W	F, 5′-CCGATTTAAGTTCTGCTACTCTAGTCG-3′	656	50
		R, 5′-GCCTTCTTAGATCATTACAACTGTCG-3′	706	

[a] Position of 5′-most primer nucleotide in the coding sequence of the gene.

Reagents Kit (Applied Biosystems, Foster City, CA) as per the manufacturer's instructions. Quantitative real time PCRs were performed in triplicate using the GenAmp® 5700 Sequence Detection System (Applied Biosystems). Independent PCRs were performed using the same cDNA for both the gene of interest and 18 S rRNA, using the SYBR® Green PCR Master Mix (Applied Biosystems). Gene-specific primers were designed for the gene of interest and 18 S rRNA using Primer Express® software (Applied Biosystems) and the Amplify program (23). All related coding sequences within the *S. cerevisiae* genome were identified via BLASTN queries of the complete coding sequences using the BLAST server at the *Saccharomyces* Genome Data Base.[3] All sequences thus identified were then aligned using the ClustalW alignment function of MegAlign software (DNAstar, Inc., Madison, WI). Gene-specific primer pairs were then manually selected such that at least one primer per pair contained a minimum of two consecutive mismatches at the 3′ end when compared against all related *S. cerevisiae* sequences. For almost all genes analyzed, both primers within a pair fulfilled the above criteria and contained numerous additional mismatches when compared against related sequences. The sequences of the primer pairs are listed in Table I. The PCR conditions consisted of denaturation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s and annealing/extension at 60 °C for 1 min. A dissociation curve was generated at the end of each PCR cycle to verify that a single product was amplified using software provided with the GenAmp® 5700 sequence detection system. A negative control reaction in the absence of template (no template control) was also routinely performed in triplicate for each primer pair. The change in fluorescence of SYBR® Green I dye in every cycle was monitored by the GenAmp® 5700 system software, and the threshold cycle (C_T) above background for each reaction was calculated. The C_T value of 18 S rRNA was subtracted from that of the gene of interest to obtain a ΔC_T value. The C_T value of an arbitrary calibrator (*e.g.* untreated sample in the case of up-regulated genes) was subtracted from the ΔC_T value to obtain a $\Delta\Delta C_T$ value. The gene expression level relative to the calibrator was expressed as $2^{-\Delta\Delta C_T}$.

Susceptibility Testing of upc2 Mutant Strains—The *upc2*Δ strain was confirmed as recommended by the Saccharomyces Genome Deletion Project protocols.[2] Single colony PCR was performed using the following primers: UPC2-A (5′-TATTGTCTTCAGTGAATGTGCTTGT-3′), UPC-2-B (5′-TTCCCTTCATCACACTTAACTCTTC-3′), UPC2-C (5′-AGGTC-AAGGATAAAGTCTGGTTTCT-3′), UPC2-D (5′-AATTGCCTCTCTAT-CAAAGTTTCCT-3′), KanB (5′-CTGCAGCGAGGAGCCGTAAT-3′), and KanC (5′-TGATTTTGATGACGAGCGTAAT-3′). The UPC2-A and UPC-D primers are located upstream and downstream of the *UPC2* coding region, respectively, whereas the UPC2-B and UPC-C primers are located within the coding region. The KanB and KanC primers are located in the *Kan* gene used to replace *UPC2* in the deletion strain. PCR products were obtained only in the UPC2-A/KanB and UPC2-D/KanC primer combinations, whereas the UPC2-A/UPC2-B and UPC2-C/UPC2-D primer combinations resulted in no PCR products. The *upc2-1* strain (SCY1012) was confirmed by colony PCR with UPC2-C/UPC2-D primers, followed by restriction digest of the amplified PCR product with *Hph*I restriction enzyme. The *upc2-1* mutation (G to A transition at nucleotide 2663) creates a unique *Hph*I restriction site (72). After restriction digestion, the *upc2-1* strain produced two DNA fragments (395 and 80 bp), whereas the corresponding wild type strain (SCY325) produced only one fragment that was 475 bp in size.

For susceptibility testing with ketoconazole, a single colony from a YPD agar plate was grown overnight in SC medium. The culture was diluted in SC medium after comparison to the 0.5 McFarland standard to afford a final inoculum of ~1×10^4 CFU/ml. After dilution, 180 μl of the culture was inoculated into a 96-well flat-bottom microplate (Corn-

YM99NC Attachment - Page 11

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

ing Inc., Corning, NY) containing 10 μl of ergosterol (final concentration of 10 μg/ml) and 10 μl of ketoconazole at varying concentrations. The starting concentration of ketoconazole used in the assay was 100 μg/ml. Ergosterol (Sigma) was added from a stock solution of 0.2 mg/ml prepared in Tween 80/ethanol (1:1 v/v). Ketoconazole was prepared as a stock solution of 10 mg/ml in Me_2SO, which was 3-fold serially diluted in Me_2SO, followed by a 5-fold dilution of each dilution in SC medium. For anaerobic conditions, microplates were incubated in BBL Bio-Bag CO_2-generating environmental chambers (BD Biosciences). The chambers are supplied with an anaerobic indicator that turns from pink to clear when anaerobiosis has been achieved. The microplates were read at 630 nm, prior to and after incubation (30 °C for 48 h), using an EL340 microplate reader (Bio-Tek Instruments, Inc., Winooski, VT).

RESULTS AND DISCUSSION

Experimental Design—Because four different classes of drugs were utilized in this study, treatments were restricted to a single time point and a single concentration. Based on previous reports (16–19) in which one doubling time (90 min) was sufficient to detect specific gene expression changes in response to azoles, amphotericin B, and 5-FC, cells were exposed to various drugs for a period of ~1.5 times the doubling time in this study. The concentration of drug is critical in these studies because the effect on gene expression may not be detectable if the drug concentration is too low, and secondary drug effects could mask the primary responses if the test concentration is too high. In similar microarray experiments, 0.5× minimum inhibitory concentration (MIC) has been utilized (16–19). However, for some antifungal drugs, this value is sometimes very close to the MIC and causes as much as 80% inhibition (Fig. 1, *B–D*). For this reason, the IC_{50} concentration was used in this study. The IC_{50} value was determined not only in conventional microplate assays but also in large scale cultures designed to mimic the conditions of the microarray experiments as closely as possible. A minimal medium such as synthetic dextrose medium was used because nutrient-rich media conditions may compensate for the inhibitory effects of the drug and obscure detection of some gene expression responses. "Biological replicates" were included in the experimental design, *i.e.* two independent experiments were conducted simultaneously, such that two independent cell cultures and drug treatments were initiated, two RNA samples were isolated, and two hybridizations were performed for each drug.

IC_{50} Determinations—Antifungal drug susceptibility assays routinely performed according to the National Committee for Clinical Laboratory Standards (25) require inocula at low cell densities (~5×10^3 CFU/ml). Because the aim of this study was to perform microarray analysis, which requires RNA isolations from a large number of cells, it was necessary to modify the currently available protocols. For this reason, inhibition assays were performed with inocula at high cell densities (~3×10^6 cells per ml, A_{600} of ~0.1) in SD medium. Assays were first performed in microplates as described under "Experimental Procedures." Absorbance readings were measured 17 h after the experiment was started in order to ensure that cells were in late exponential phase.

The inhibition assays were repeated in large scale experiments because 100-ml cultures would be used in the microarray experiments, and it is possible that IC_{50} values determined in microplate assays would be different under large scale conditions. The results from the microplate and large scale assays are shown in Fig. 1, *A–D*. Based on these experiments, the IC_{50} values for the four drugs were determined to be the following: 56 ± 8.18 μg/ml for ketoconazole, 0.14 ± 0.04 μg/ml for amphotericin B, 0.03 ± 0.01 μg/ml for caspofungin, and 0.30 ± 0.16 μg/ml for 5-fluorocytosine. In general, the microplate and large scale assay results were in agreement with each other (Fig. 1*E*) except for 5-fluorocytosine, which showed 2–4 times greater IC_{50} values in large scale conditions. The high IC_{50} value observed for ketoconazole is consistent with the fact that ketoconazole is a fungistatic agent. This value is also in agreement with a previous study in which it was reported that increasing the starting inoculum from 1×10^3 to 1×10^7 CFU/ml caused a dramatic reduction in fluconazole activity (26).

Gene Expression Responses to Ketoconazole—A total of 51 genes was found to be responsive to ketoconazole treatment under the experimental conditions used. Of these, 44 genes showed a significant increase in expression, and 7 genes showed a significant decrease in expression. The number of responsive genes in this study is less than the number observed in previous studies (16, 17) because two independent biological replicate experiments were conducted in this study, and only genes that responded similarly in both experiments were selected as responsive genes. In studies by Lee *et al.* (27), it was shown that the number of false positives can be quite high in a microarray study conducted with samples from only one experiment. For example, in a single experiment, 1.8% of the total genes showed a 2-fold change in expression and of these about 45% were false positives, whereas in a duplicate experiment, 0.6% of the total genes showed a 2-fold change in expression and of these only 1% were false positive (27, 28).

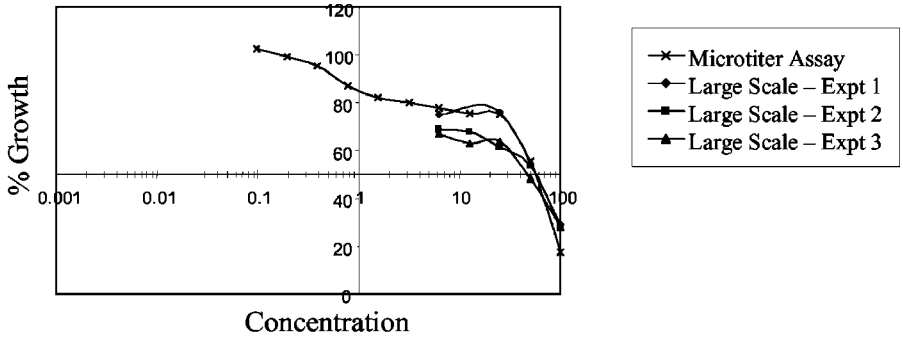
The distribution of ketoconazole-responsive genes and their biological roles are shown in Fig. 2. The category of genes with the largest number of responses is the lipid, fatty acid, and sterol metabolism group. As shown in two previous studies, genes involved in ergosterol biosynthesis were up-regulated in response to azole treatment (16, 17). The genes found to be responsive in this study include *ERG1, ERG2, ERG3, ERG4, ERG5, ERG6, ERG11, ERG24, ERG25, ERG26,* and *ERG28*. Because the enzyme encoded by *ERG11* is the direct target of azole antifungals and because its expression is increased in some azole-resistant *C. albicans* strains, it is not surprising that *ERG11* would be induced in response to ketoconazole treatment (29, 30). Except for *ERG1*, most of the responsive genes in this study are functional downstream of *ERG11* (Fig. 3), suggesting that their induction is in response to ergosterol depletion. Comparison of these results with those obtained in previous studies (16, 17) indicates a few discrepancies that could be attributed to differences in strains used, media conditions used (*e.g. ERG1* and *ERG26* not detected in the study by Bammert and Fostel (16) where rich medium was used), time of drug exposure (*ERG10* induced in the study by De Backer *et al.* (17) where cells were treated with itraconazole for 24 h), absence of genes from previous microarrays (*e.g. ERG24* not measured in the De Backer *et al.* study), and previous lack of annotations for some genes (*e.g. ERG28*).

In addition to the ergosterol biosynthesis genes, six additional genes involved in lipid metabolism were also affected in this study. These include *CYB5*, *HES1*, *YSR3*, *INO1*, *ELO1,* and *UPC2* (Fig. 2). The gene *CYB5* encodes cytochrome b_5 and is required for suppressing azole hypersensitivity in a *S. cerevisiae erg11* mutant (31). The gene *HES1* shows homology to the human oxysterol-binding protein and plays a role in ergosterol biosynthesis, as implied by reduced membrane ergosterol levels in a *hes1* mutant of *S. cerevisiae* (32). The genes *YSR3* (encoding dihydrosphingosine-1-phosphate phosphatase), *INO1* (encoding inositol-1-phosphate synthase), and *ELO1* (encoding a fatty acid elongating enzyme) have been shown to play a role in sphingolipid biosynthesis in *S. cerevisiae* (33–35). It has been shown recently (36) that a mutation in the *ERG26* gene produced defects in sphingolipid metabolism, indicating that sphingolipid metabolism is regulated by sterol levels in *S. cerevisiae* cells. Thus, the effect of ketoconazole on *CYB5*,

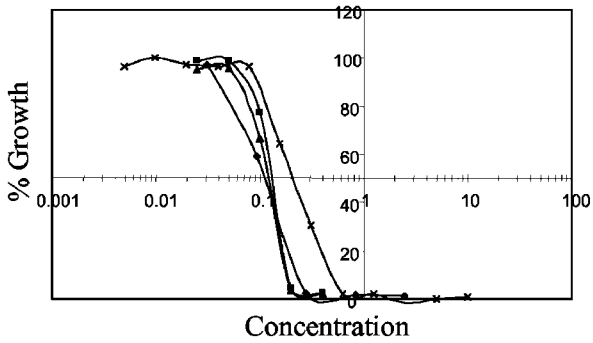
TM95NC Attachment – Page 112

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

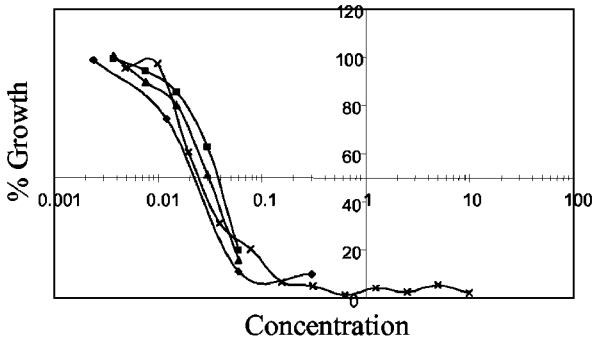
A



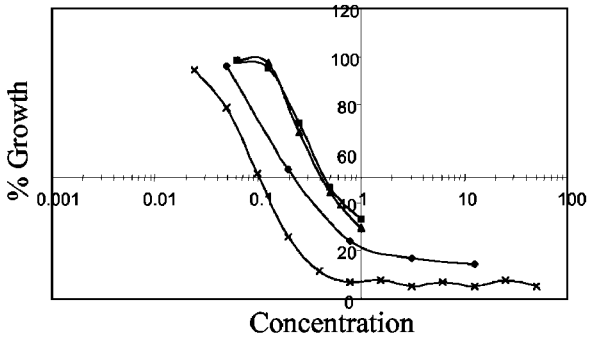
B



C



D



E

IC_{50} Values (μg/ml) in Individual Experiments

	KTZ	AmB	CPF	5-FC
Microtiter Assay	55.00	0.20	0.02	0.10
Large Scale – Expt 1	66.00	0.10	0.02	0.21
Large Scale – Expt 2	56.00	0.14	0.04	0.45
Large Scale – Expt 3	46.00	0.13	0.03	0.40

FIG. 1. Determination of IC_{50} values for ketoconazole (*A*), amphotericin B (*B*), caspofungin (*C*), and 5-fluorocytosine (*D*). Results are shown from a microtiter assay and three independent assays performed in large scale cultures. *E*, IC_{50} values obtained in the individual assays. *KTZ*, ketoconazole; *AmB*, amphotericin B; *CPF*, caspofungin.

YM95NC Attachment -- Page 113

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

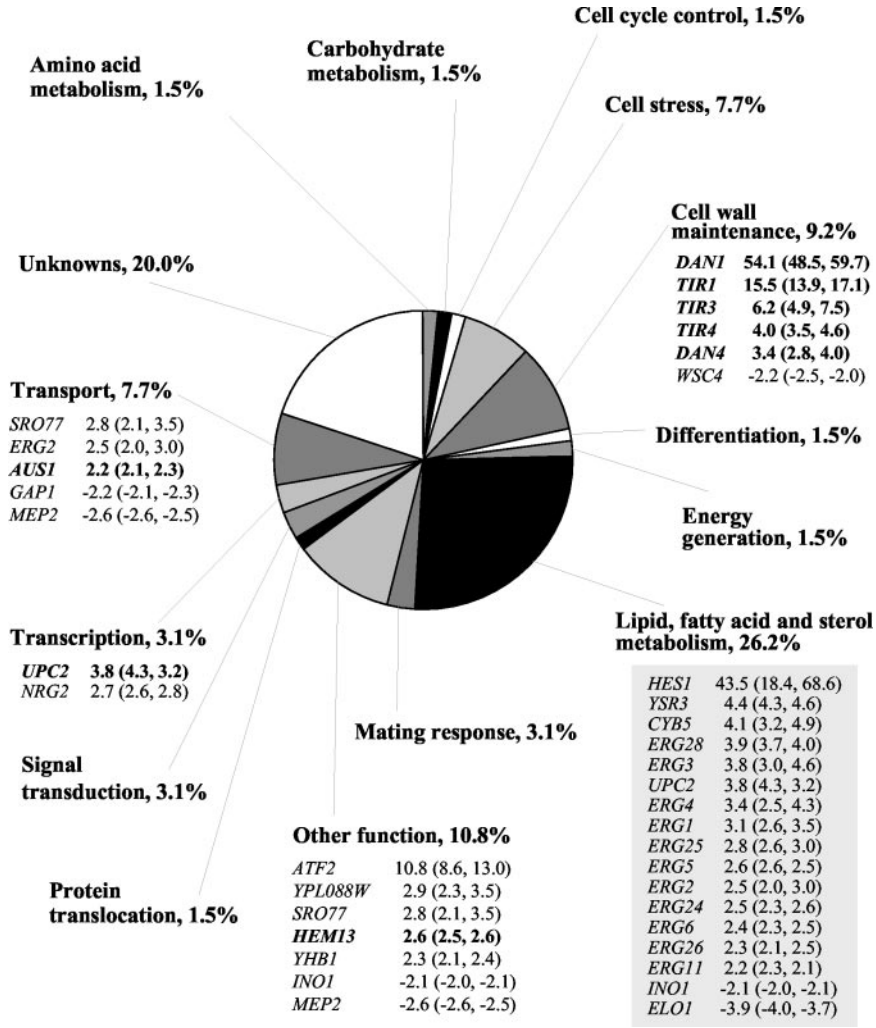


FIG. 2. **Distribution of ketoconazole-responsive genes.** Genes were annotated and assigned to various functional categories using the *Saccharomyces* Genome Data Base and Yeast Proteome Data Base. The average expression ratio from two independent experiments is shown for genes associated with the mechanism of action. In *parentheses* are the expression ratios obtained in each experiment. *Positive numbers* indicate induction, and *negative numbers* indicate repression. Genes in the *gray box* are indicative of major drug-specific responses. Genes in *boldface* are summarized in Fig. 3. An annotated list of all the genes in each functional category, including expression ratios from individual experiments, can be found in Supplemental Material Table A.

HES1, *YSR3*, *INO1,* and *ELO1* is in agreement with their roles in sterol or sphingolipid metabolism.

It was interesting to observe that the gene *UPC2* was induced in this study (Fig. 2, *transcription category*). The gene *UPC2* encodes a binucleate zinc cluster transcription factor, which is involved in the regulation of sterol biosynthesis and uptake, and sphingolipid biosynthesis (37, 38). Upc2p is a sterol regulatory element-binding protein, which regulates the transcription of ergosterol biosynthetic genes *ERG2* and *ERG3* (37). Therefore, the induction of *UPC2* in response to ketoconazole is consistent with its role in regulating ergosterol biosynthesis. In addition, *UPC2* is involved in regulating sterol uptake as indicated by the ability of the *upc2-1* mutant to accumulate exogenous sterol in *S. cerevisiae* cells (38). Comparison of normal and *upc2-1* mutant strains by microarray analysis revealed that the genes *UPC2, AUS1*, *PDR11,* and *DAN1* are up-regulated in the mutant strain (39). *AUS1* and *PDR11* encode ATP-binding cassette transporters, and *DAN1* codes for a cell wall mannoprotein. Deletions in *AUS1*, *PDR11,* and *DAN1* in the *upc2-1* strain resulted in a decrease in sterol accumulation, suggesting that these genes are required for sterol uptake when sterol biosynthesis is compromised (39).

Most of the genes that were induced in the *upc2-1* mutant strain (39) were also induced in this study in response to ketoconazole (Fig. 2 and Supplemental Material, Table A). These include *HES1*, *CYB5*, *UPC2*, *ERG25*, *ERG24*, *ERG26*, and *ERG11* (lipid metabolism); *DAN1*, *DAN4*, *TIR1*, *TIR3,* and *TIR4* (cell wall maintenance); *AUS1* and *SRO77* (transport); *ATF2* and *HEM13* (other metabolism, heme biosynthesis); *AMS1* (carbohydrate metabolism); *SCM4* (cell cycle control); and *YPL272C*, *YMR317W,* and *YGR131W* (unknowns). All of these genes, except *CYB5*, *YMR317W,* and *YGR131W* have a sterol regulatory element motif in their promoter regions, which has been identified as a Upc2p-binding site. Also, the level of induction of the highly induced genes is very similar between the two studies. For example, *DAN1* was induced ~55-fold in the *upc2-1* mutant, and it was induced ~54-fold in the current study. The induction of *HEM13*, whose product is involved in heme biosynthesis, is consistent with a previous azole microarray result (16). There are several lines of evidence which show that heme plays an important role in sterol synthesis and regulates the transcription of several genes involved in this process (40). It is interesting to note that heme deficiency has been associated with increased sterol uptake in *S. cerevisiae* cells (41).

The similarities between the gene expression responses in the *upc2-1* mutant strain and the ketoconazole-treated cells in the current study support the hypothesis that when sterol biosynthesis is inhibited, cells respond by inducing genes required for sterol uptake from the external environment. It has also been proposed that resistance to azoles could involve multiple mechanisms including alterations in sterol biosynthesis, target site, uptake, and efflux (42). The induction of *UPC2* and some *UPC2*-regulated genes (*e.g. AUS1*, *DAN1*, *DAN4*, *TIR3,* and *TIR4*) appears to be unique to our study. These genes could have been unresponsive in previous microarray studies due to

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

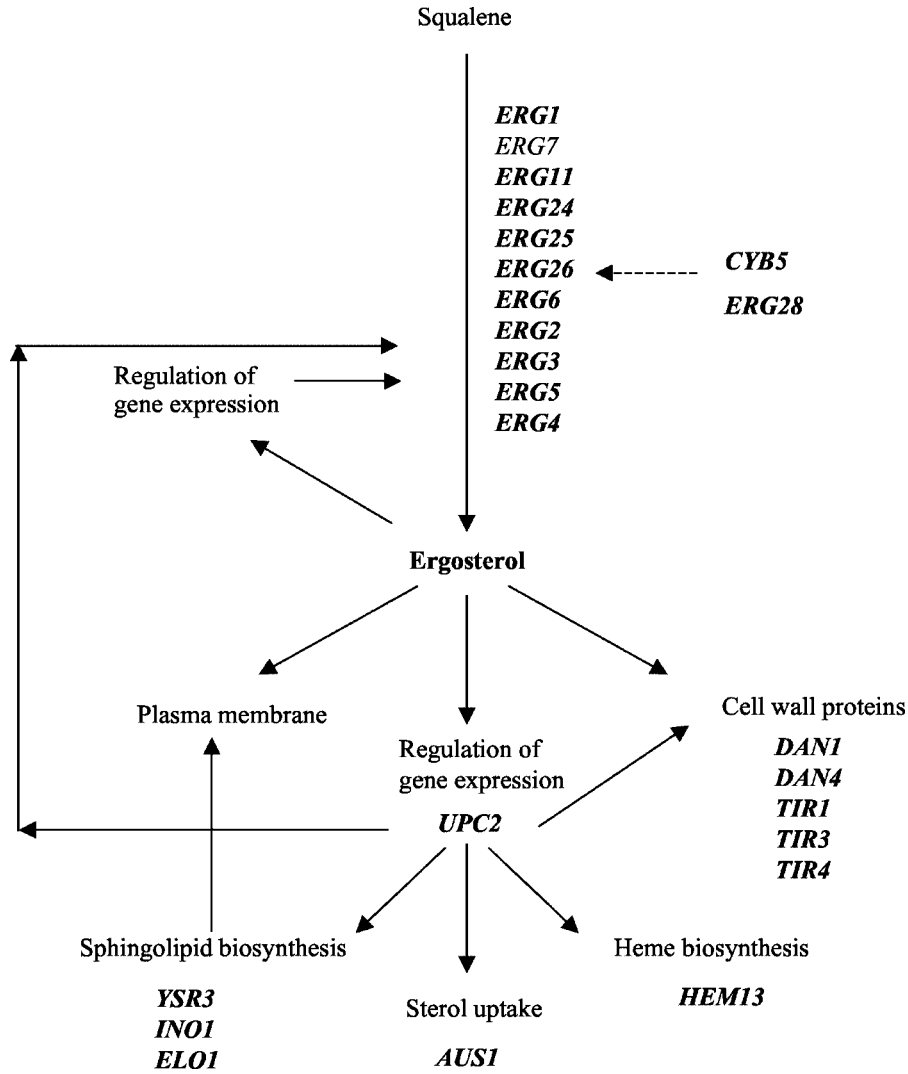


FIG. 3. **Summary of gene expression responses to ketoconazole.** Genes in *boldface* were responsive in this study. The relationship between ergosterol biosynthesis and *UPC2*-mediated gene regulation is summarized.

strain differences and media conditions used. In addition, *AUS1* was not annotated at the time these studies were published, and therefore its drug response was not documented.

Taken together, the results in the current study suggest that when *S. cerevisiae* cells grown in minimal medium are exposed for a short time to a sub-inhibitory dose of ketoconazole, they respond by inducing genes that are not only involved in ergosterol biosynthesis but also genes involved in sterol uptake from the external environment. Fig. 3 summarizes the results from this study and highlights the ketoconazole-responsive genes. It also shows the relationship between ergosterol biosynthesis and gene regulation by *UPC2*.

Gene Expression Responses to Amphotericin B—A total of 265 genes was responsive to amphotericin B treatment under the experimental conditions tested. Of these, 185 genes showed a significant increase in expression, and 80 genes showed a significant decrease in expression. Fewer responsive genes were identified in this study relative to a previous study (18), because only genes that responded similarly in two independent experiments were selected as responsive genes in this study.

The distribution of the amphotericin B-responsive genes and their biological roles are shown in Fig. 4. The category of genes with the largest number of responses was the "unknown" group, indicating that many genes that responded to amphotericin B have no previously characterized cellular role. The second largest number of responses was in the transport group, in agreement with results observed by Zhang *et al.* (18). Regulation of cellular transport is consistent with the mechanism of action of amphotericin B, which binds with membrane ergosterol to form pores that disrupt the membrane and cause leakage of ions and small molecules from the cell (43). The leakage of ions and nutrients from the cell may be at least partially counterbalanced by the induction of genes involved in membrane transport. Genes associated with membrane transport that were induced in this study include *GIT1*, *BAP2*, *BAP3*, *HXT2*, *HXT4*, *HXT5, TPO2,* and *PTR2* (Fig. 4). The most highly differentially expressed gene in this category was *GIT1* which showed an ~33-fold induction in response to amphotericin B. This is in agreement with a recent observation that *GIT1* is required for optimal growth of *S. cerevisiae* cells in the presence of 10 μM nystatin, another polyene antifungal drug (44). Several genes that encode proteins involved in membrane transport were down-regulated in response to amphotericin B, including *MEP2*, *SAM3*, *CTR1*, *SEO1*, *SUL1*, *OPT1*, *GAP1,* and *MMP1* (Fig. 4). It is possible that the induction of certain transport-related genes or the repression of others is necessary to maintain homeostasis within the cell when membranes are damaged.

The category of genes with the next largest number of responses to amphotericin B treatment was the cell stress group. As would be expected, the leakage of cell contents induces stress-related responses in the cell. The genes listed in the cell stress category in Fig. 4 have been shown to respond to various environmental stresses in *S. cerevisiae* cells (45), *e.g. HSP12*,

YM95NC Attachment - Page 15

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

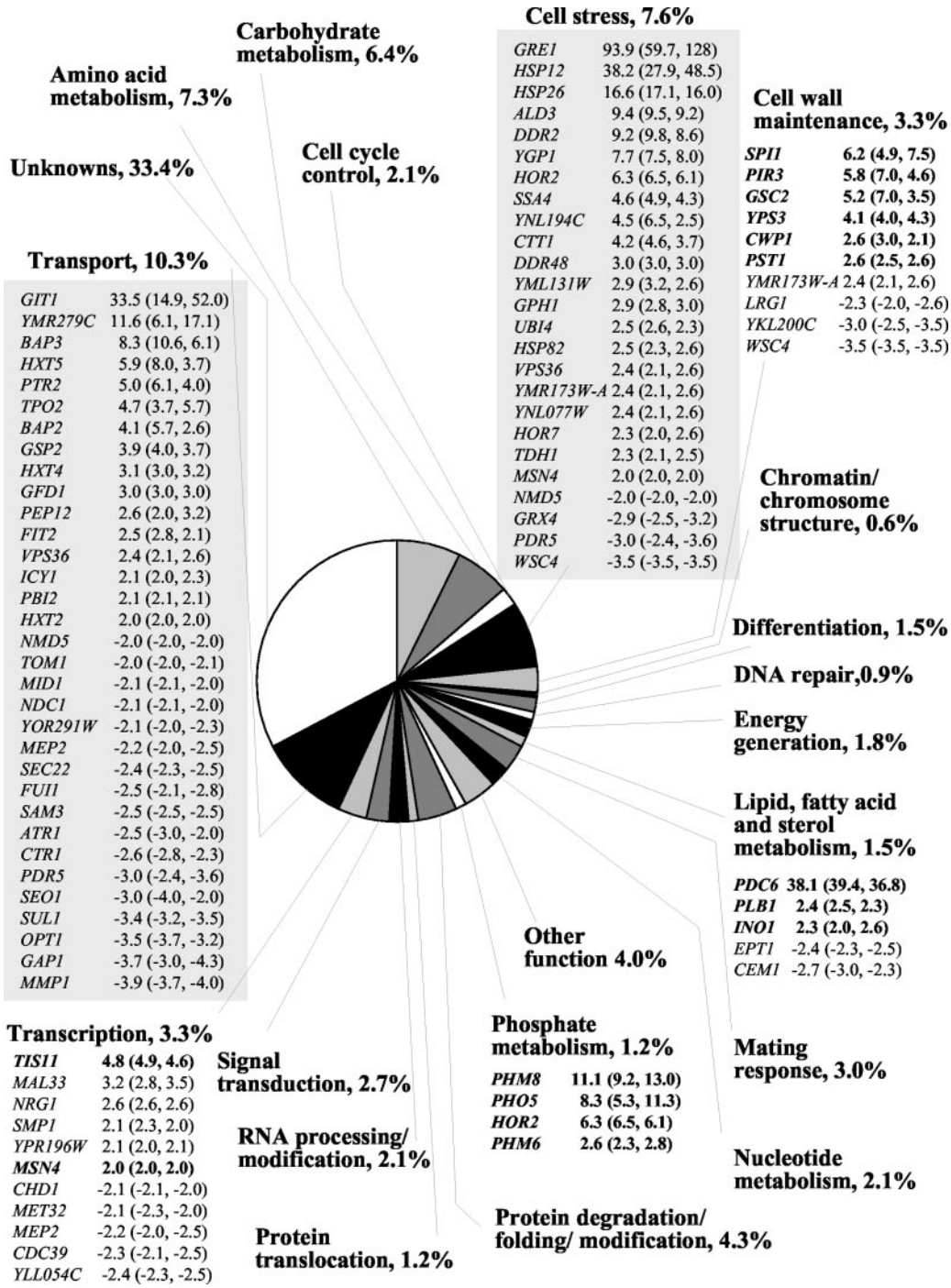


FIG. 4. **Distribution of amphotericin B-responsive genes.** Gene annotations and expression values are as in Fig. 2. Genes in the *gray boxes* are indicative of major responses to amphotericin B treatment. Genes in *boldface* are summarized in Fig. 5. An annotated list of all the genes in each functional category, including expression ratios from individual experiments, can be found in Supplemental Material Table B.

HSP26, *HSP82,* and *SSA4* (heat shock, see Ref. 45); *GRE1*, *HOR2,* and *HOR7* (osmotic stress, see Ref. 46), *CTT1* and *DDR48* (oxidative stress, see Ref. 47); *YGP1* (nutrient limitation, see Ref. 48), *ALD3*, *GPH1,* and *TDH1* (ethanol stress, see Ref. 49); and *MSN4,* which is a transcription factor regulating the expression of various stress-response related genes (50). In addition, several genes induced in the environmental stress response (ESR, see Ref. 45) were also induced in response to amphotericin B in this study (Fig. 4 and Supplemental Material, Table B), including *HSP12*, *HSP26*, *UBI4*, *SSE2*, *SSA4*, *SSA1*, *LAP4*, *YPS3,* and *PBI2* (protein folding and degradation); *PDC6*, *TKL2*, *NRG1*, *AMS1*, *HOR2*, *HXK1*, *HXT5,* and *GPH1* (carbohydrate metabolism); *SPI1*, *PIR3*, *GSC2* and *CWP1* (cell wall maintenance); and *GIT1*, *BAP2*, *BAP3,* and *HXT5* (metabolite transport). Several genes repressed in the ESR (45) were also repressed in response to amphotericin B (see Supplemental Material, Table B), including *ADE2*, *FUI1,* and *FUR1* (nucleotide metabolism); and *HAS1* (RNA processing). These results indicate that a general stress response is exerted in *S. cerevisiae* cells in response to amphotericin B treatment.

Reconstructing the cell membrane might be one of the ways in which cells would respond to membrane disruptions caused by amphotericin B. This is reflected in the induction of genes involved in lipid, fatty acid, and sterol metabolism, such as *PLB1* and *INO1* (Fig. 4). In addition, the gene *GIT1* was in-

YMSNC Attachment – Page 116

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

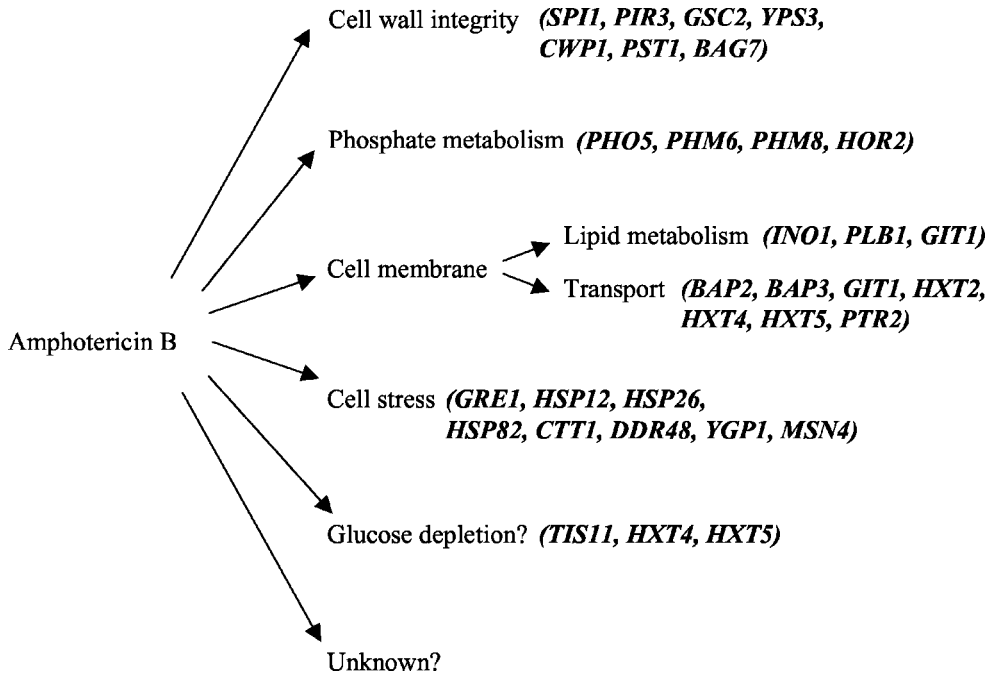


FIG. 5. **Summary of gene expression responses to amphotericin B.** The functional features of the gene expression responses to amphotericin B are summarized.

duced that is involved in the transport of glycerophosphoinositol, a precursor to several phospholipids that are essential membrane components in *S. cerevisiae* cells. The responses involved with phospholipid metabolism are in agreement with the hypothesis that amphotericin B interacts with phospholipids in the cell membrane in order to exert its effect (43).

In addition to genes involved in cell membrane preservation, six genes involved in cell wall maintenance were also induced in response to amphotericin B treatment in this study (Fig. 4), including *SPI1*, *PIR3*, *GSC2, YPS3*, *CWP1,* and *PST1*. These results are in agreement with previous observations that cell wall components can affect the interaction of amphotericin B with the cytoplasmic membrane (10). It was shown that stationary phase *C. albicans* cells were more resistant to polyenes than exponential phase cells. It has been suggested that because the turnover of cell wall constituents is slower in stationary phase cells, polyenes would have less access to the cell membrane in that phase. The cell wall maintenance genes that were induced in response to amphotericin B in this study are involved in the cell wall stress-response pathway, which reflects a compensatory reaction to cell wall damage (reviewed in Ref. 51). The gene *BAG7* (a gene in the *signal transduction category*, see Supplemental Material, Table B), which was induced ~56-fold in response to amphotericin B in this study, is also involved in regulating key components of the cell wall stress-response pathway (52).

The correlation between cell wall integrity and polyene binding to cell membrane is also consistent with a recent study conducted on a nearly complete (96% of all annotated genes) collection of gene-deletion mutants in *S. cerevisiae* (44). These mutants were analyzed in parallel, and the fitness contribution of each gene was quantitatively assessed in response to various environmental conditions, including challenge with the antifungal polyene nystatin. Interestingly, two of the deletion strains most sensitive to nystatin, *myo5*Δ and *bro1*Δ, had deletions in genes required for cell wall structure and integrity. This result supports the observation that cell wall components can affect the binding of polyene compounds to the cell membrane. In contrast, the two genes *MYO5* and *BRO1* were not affected by amphotericin B treatment in the current study. As indicated by Giaever *et al.* (44), there is little correlation between fitness profiling and transcription profiling data. Whether this is due to post-transcriptional regulation of some genes or other factors is not presently understood.

Comparison between the results from this study and the microarray study by Zhang *et al.* (18) revealed several differences in the responses observed, which could be due to differences in the experimental conditions such as media conditions (*e.g.* rich medium was used in the previous study), yeast strains (strain L1190 was used in the previous study), and drug concentration used (*e.g.* 20-fold higher concentration of amphotericin B was used in the previous study). Among the genes that responded similarly between the two studies, the gene *TIS11* is of particular interest (Fig. 4, *transcription category*). Tis11p is a glucose-repressible zinc finger protein of unknown function (53). It is possible that glucose depletion in amphotericin B-exposed cells leads to the induction of genes such as *TIS11*, *HXT4,* and *HXT5* (Fig. 4). The induction of *PHO5* in the two microarray studies indicates the need to increase phosphate uptake in response to amphotericin B, phosphate being an essential nutrient required for the synthesis of several cellular components such as lipids, proteins, nucleic acids, and sugars. In addition, the genes *PHM6, PHM8,* and *HOR2* that were induced in the current study (Fig. 4) are also regulated by the PHO regulatory system involved in the regulation of phosphate uptake from extracellular sources (54). Thus, membrane reconstruction, cell stress, improvement of cell wall integrity, phosphate uptake, and perhaps glucose depletion appear to be the major responses to amphotericin B (Fig. 5).

Gene Expression Responses to Caspofungin—A total of 192 genes responded to caspofungin treatment under the experimental conditions tested. Of these, 137 genes showed a significant increase in expression, and 55 genes showed a significant decrease in expression. The distribution of the responsive genes and their biological roles are shown in Fig. 6. The category of genes with responses most relevant to the mechanism of caspofungin action, *i.e.* inhibition of β-1,3-glucan synthase, is the cell wall maintenance group. The genes induced in this group include *PIR3*, *CWP1*, *YPS3, YLR194C, CRH1,* and *PST1* (cell wall-related proteins), *CSG2, KRE11,* and *FKS1* (cell wall

NIMSNC Attachment - Page 117

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

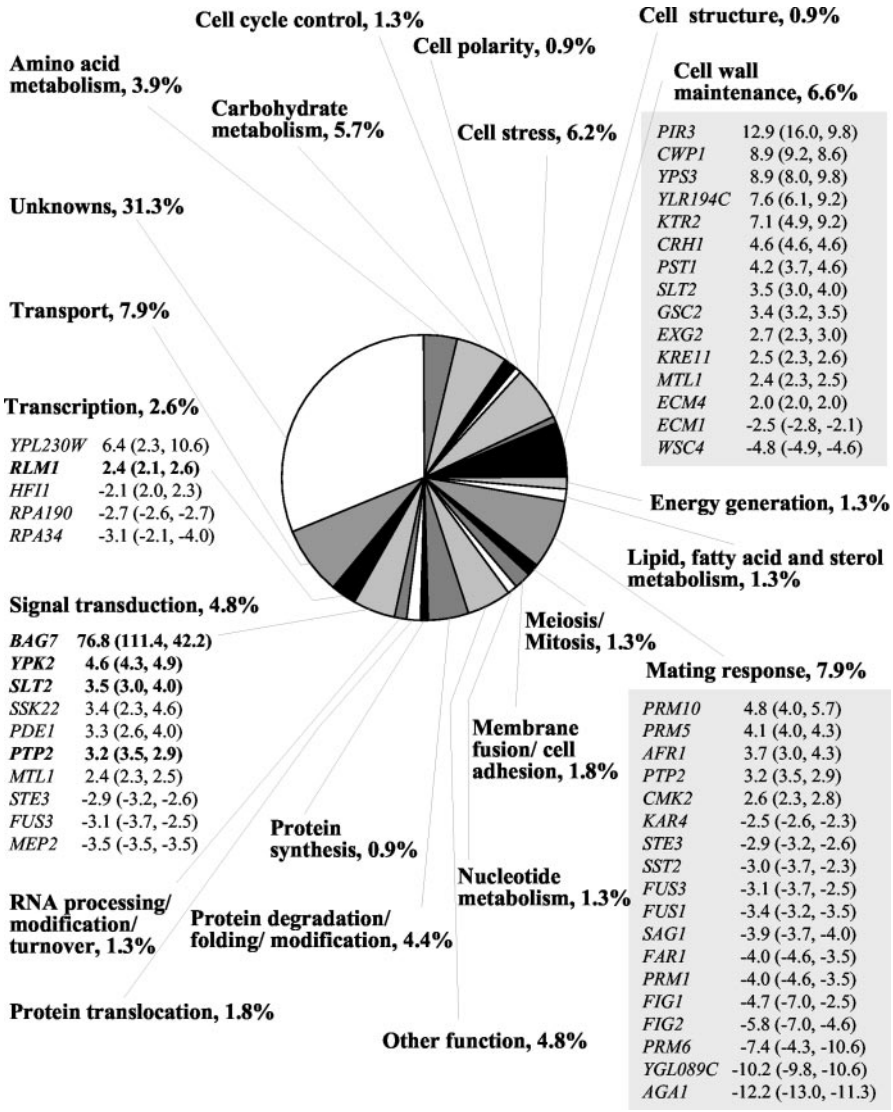


FIG. 6. **Distribution of caspofungin-responsive genes.** Gene annotations and expression values are as in Fig. 2. Genes in the *gray boxes* are indicative of major drug-specific responses. Genes in *boldface* are summarized in Fig. 7. An annotated list of all the genes in each functional category, including expression ratios from individual experiments, can be found in Supplemental Material Table C.

biosynthesis); *KTR2* and *EXG2* (cell wall modification); and *SLT2* and *MTL1* (cell wall stress response signal transduction). The genes present in this category belong to the cell wall integrity pathway, which is activated in response to cell wall disrupting agents, such as cell wall lytic enzymes, hypo-osmotic stress, heat stress, cell wall weakening mutations, and chemicals such as the anionic detergent SDS and Calcofluor (an agent that interferes with chitin crystallization) (reviewed in Refs. 51 and 55). These responses include (i) increase in the synthesis of cell wall components, (ii) utilization of alternative mechanisms for incorporation of cell wall proteins, and (iii) increase in cell wall strengthening. The increase in the synthesis of cell wall components in response to cell damage has been shown in several studies utilizing cell wall mutants (reviewed in Ref. 51). A transcription profiling study conducted in a mutant strain of *S. cerevisiae* carrying a disruption in the *FKS1* gene, which encodes the catalytic subunit of β-1,3-glucan synthase enzyme, indicated that several genes encoding glycosylphosphatidylinositol-dependent cell wall proteins and glycosylphosphatidylinositol-attached cell membrane proteins were up-regulated in *fks1*Δ cells, including *PST1*, *YPS3*, *YLR194C*, *CRH1,* and *CWP1* (56). These genes were also induced in response to caspofungin in the current study (Fig. 6). In addition, the gene *FKS2* (also known as *GSC2*) was induced in the *fks1*Δ cells. The gene *FKS2* is homologous to *FKS1* and encodes an alternative catalytic subunit of β-1,3-glucan synthases. Thus, the induction of *GSC2* in the current study is in agreement with the need to synthesize cell wall glucans when glucan biosynthesis is compromised.

Another response to cell wall stress, *i.e.* the increase in cell wall strength, is indicated not only by the induction of genes encoding cell wall proteins in this study but also by the induction of *KRE11*, a gene required for the synthesis of β-1,6-glucan (Fig. 6). The cell wall of *S. cerevisiae* is composed of three major components: glucan, mannoproteins, and chitin (57). Glucan consists of two types of polymers, β-1,3-glucan and β-1,6-glucan that account for 50 and 10% of the dry weight of the cell wall, respectively. Whereas β-1,3-glucan molecules form an internal skeletal layer, β-1,6-glucan plays an important role as a cross-linker connecting the cell wall proteins, chitin and β-1,3-glucan to each other. Thus, when β-1,3-glucan levels are low, as in caspofungin treatment, the cells may compensate by increasing β-1,6-glucan through the induction of *KRE11* and thereby also increasing the strength of the cell wall.

Although several genes involved in cell wall maintenance were induced in response to caspofungin in this study, it is worth noting that the gene *SBE2,* encoding a Golgi protein involved in the transport of cell wall components, was not induced. Overexpression of *SBE2* was shown to confer resistance, whereas deletion of *SBE2* conferred hypersensitivity to caspofungin (58). It is possible that a low level exposure to caspofungin for a short period of time, as was done in

YMSNC Attachment - Page 18

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

the current study, may not be sufficient to induce the expression of *SBE2*. It is also possible that *SBE2* is not transcriptionally responsive to caspofungin and is regulated posttranscriptionally.

In this study, the induction of the genes *SLT2* (encoding a MAP kinase) and *MTL1* (encoding a transmembrane protein) in response to caspofungin is indicative of the involvement of specific signal transduction pathway components required for cell wall stress-induced responses (Fig. 6, *cell wall maintenance and signal transduction category*). This signal transduction pathway involves a putative sensor in the plasma membrane that senses defects in the cell wall. The gene *MTL1* encodes a membrane protein and has been hypothesized to be a cell wall sensor (reviewed in Ref. 51). Thus, its induction in response to caspofungin is supportive of this proposed role. Downstream of the sensor is a protein designated as Rho1p, a GTPase, that serves as an integral regulatory subunit of the β-1,3-glucan synthase complex. Rho1p can directly activate glucan synthesis through interaction with Fks1p and Fks2p, but it can also activate protein kinase C (encoded by *PKC1*) that in turn can activate a MAP kinase (encoded by *SLT2*) (reviewed in Refs. 51 and 55). Activation of Slt2p leads to the activation of transcription factors that regulate the transcription of several cell wall stress-response genes. A genome-wide survey of *S. cerevisiae* cells in which the cell wall stress-response signaling pathway was strongly active showed that induction of this pathway resulted in the induction of genes such as *SLT2*, *FKS2*, *PST1*, *CRH1*, *CWP1,* and *PIR3* (59). It is interesting to note that one of the transcription factors activated by *SLT2* is Rlm1p. Jung and Levin (59) have also shown that the transcriptional induction, caused by overactivation of the cell wall stress-response signaling pathway, was abolished in an *rlm1*Δ mutant. This result indicates that Rlm1p plays an important role in regulating the expression of cell wall stress-response genes. Interestingly, the gene *RLM1* was induced in response to caspofungin in the current study (Fig. 6, *transcription category*). Thus, the induction of *MTL1*, *SLT2,* and *RLM1* in this study is consistent with their roles in the cell wall integrity pathway.

Another component of the cell wall stress-response pathway is the GTPase-activating protein Bag7p that activates Rho1p *in vitro* (reviewed in Ref. 51). Bag7p also suppresses the lethality of Rho1p hyperactivation in response to cell wall damage (52). It has been suggested that Bag7p interacts with Rho1p and is involved in controlling actin cytoskeleton reorganization. It is worth noting that the gene *BAG7* was induced ~76-fold in response to caspofungin in this study (Fig. 6, *signal transduction category*). The genes *YPK2* and *PTP2* were also induced in the current study (Fig. 6, *signal transduction category*). *YPK2* encodes a yeast protein kinase that acts as a regulator of the cell wall stress response MAP kinase cascade (60). Mutants in *S. cerevisiae* lacking *YPK2* display severely reduced activation of the MAP kinase pathway. The *PTP2* gene encodes a protein-tyrosine phosphatase that can also regulate this kinase pathway by inactivating the MAP kinase encoded by *SLT2* (61). Thus, *YPK2* and *PTP2* regulate the cell wall integrity pathway by controlling the phosphorylation and dephosphorylation of kinases involved in this pathway. The *PTP2* gene was also induced in *fks1*Δ mutants of *S. cerevisiae* (56).

Taken together, these results suggest that exposure to caspofungin, an inhibitor of β-1,3-glucan synthase, causes the cells to improve their cell wall integrity by triggering components of the cell wall stress-response pathway including a cell wall sensor (*MTL1*), a MAP kinase (*SLT2*), regulators of the MAP kinase (*YPK2* and *PTP2*), a GTPase activator (*BAG7*), a transcription factor (*RLM1*), and several genes regulated by Rlm1p

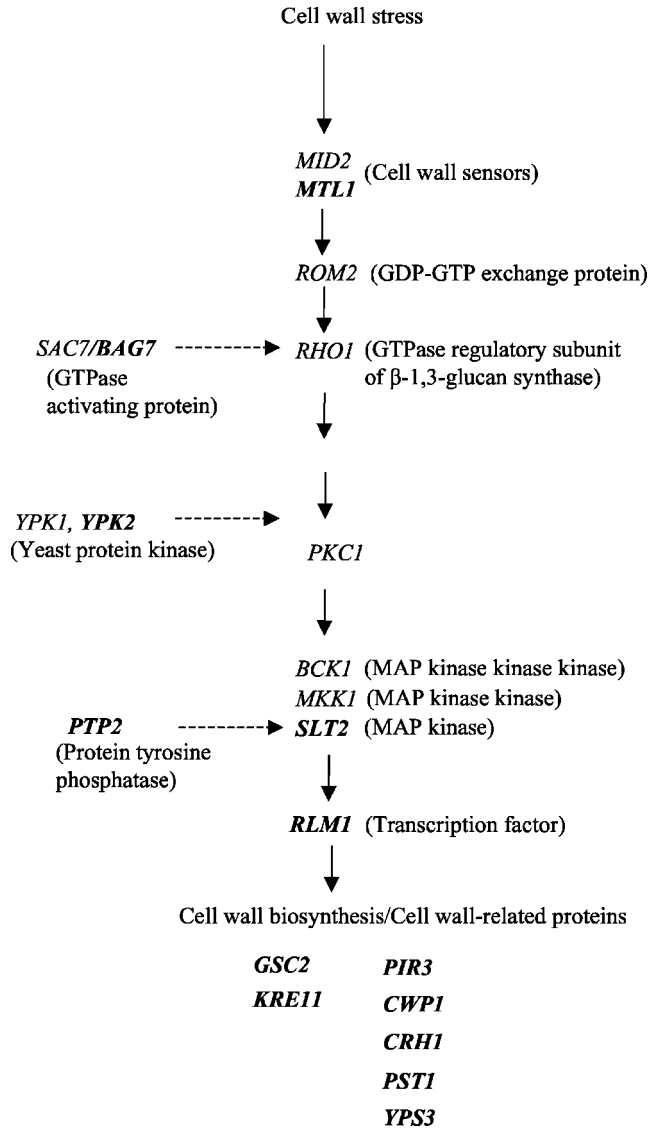


FIG. 7. **Summary of gene expression responses to caspofungin.** Genes in *boldface* were responsive in this study. Various components of the cell wall stress-response pathway are summarized.

that are either cell wall-related proteins (*PIR3*, *CWP1*, *YPS3*, *YLR194C*, *CRH1,* and *PST1*) or cell wall biosynthesis genes (*GSC2* and *KRE11*). Fig. 7 summarizes the results from this study and highlights the caspofungin-responsive genes and the various components of the cell wall integrity pathway.

Apart from the response in cell wall integrity pathway, caspofungin treatment resulted in the down-regulation of genes involved in mating response (Fig. 6). These include *AGA1*, *FIG1*, *FIG2*, *SAG1*, *FUS1*, *FUS3*, *SST2, STE3,* and *KAR4*. The mating response requires the formation of a mating projection, adhesion of cell walls between two cells, and degradation of the cell walls to allow the plasma membranes to fuse (51). This response is in opposition to the cell wall integrity response, which appears to be the major response to caspofungin treatment. Therefore, it is not surprising that mating response genes are down-regulated in response to caspofungin.

Several genes involved in small molecule transport and vesicular transport were either up- or down-regulated in response to caspofungin (Fig. 6). It is possible that damage to the cell wall would cause defects in the plasma membrane affecting transport of small molecules through the membrane. This would be compensated for by the induction of genes involved in facilitating transport.

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

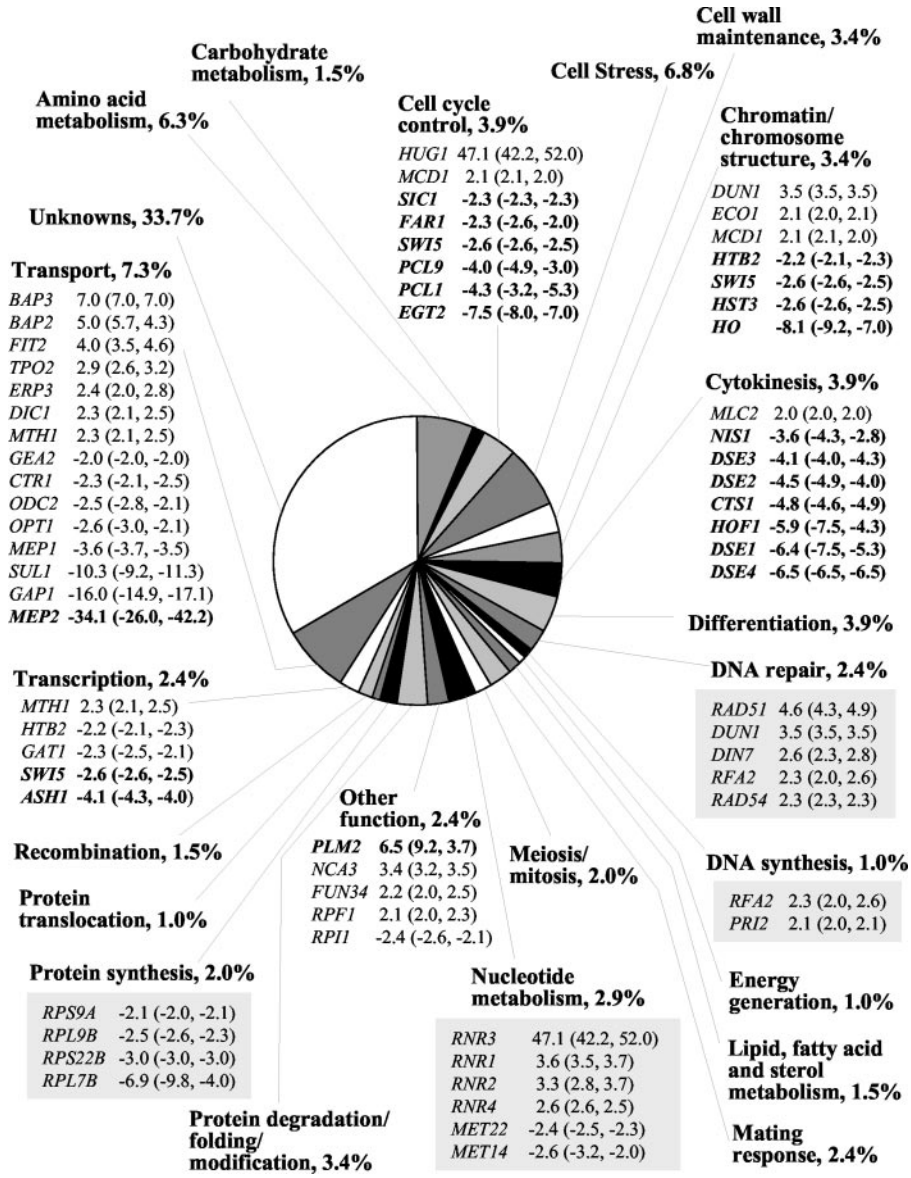


FIG. 8. **Distribution of 5-fluorocytosine-responsive genes.** Gene annotations and expression values are as in Fig. 2. Genes in the *gray boxes* are indicative of major drug-specific responses. Genes in *boldface* are summarized in Fig. 9. An annotated list of all the genes in each functional category, including expression ratios from individual experiments, can be found in Supplemental Material Table D.

As observed with amphotericin B, the exposure to caspofungin also triggered a cell stress response and several genes in the cell stress category were induced (Fig. 6 and Supplemental Material, Table C). These include *HSP12* and *UBI4* (heat stress, see Ref. 45); *GRE1* and *HOR2* (osmotic stress, see Ref. 46); *CTT1*, *DDR2,* and *NTH1* (oxidative stress, see Ref. 47); and *ALD3* and *GPH1* (ethanol stress, see Ref. 49). Also several genes induced in the ESR (45) were also induced in response to caspofungin (see Supplemental Material, Table C). These include *HSP12*, *UBI4*, *YPS6,* and *PRB1* (protein folding and degradation); *HXT5*, *AMS1*, *GPH1*, *PFK26*, *ALD4,* and *HXK1* (carbohydrate metabolism); *PIR3*, *CWP1,* and *GSC2* (cell wall maintenance); and *BAP2*, *BAP3*, *HXT5,* and *AGP2* (metabolite transport). Several genes repressed in the ESR (45) were also repressed in response to caspofungin (see Supplemental Material, Table C), including *ADE4*, *FUI1,* and *AAH1* (nucleotide metabolism) and *HAS1*, *MTR4,* and *MAK11* (RNA processing and modification). As with amphotericin B, a large number of genes that responded to caspofungin treatment belong to the "unknowns" category. Perhaps some of these genes play an important role in the cell wall stress-response pathway because several components of this pathway remain unidentified including factors involved in sensing cell wall defects, in reorganizing the cell wall, and in mediating specific responses originating from Rho1p.

Gene Expression Responses to 5-Fluorocytosine—A total of 167 genes responded to treatment with 5-FC under the experimental conditions tested. Of these, 99 genes showed a significant increase in expression, and 68 genes showed a significant decrease in expression. The distribution of the responsive genes and their biological roles are shown in Fig. 8. The responses associated with the mechanism of 5-FC action fell into the following categories: DNA synthesis, protein synthesis, nucleotide metabolism, and DNA repair. These results are in agreement with a previous microarray study in *S. cerevisiae* with 5-FC (19). The drug 5-FC enters fungal cells with the help of a permease enzyme and is converted to 5-fluorouracil (5-FU) by the enzyme cytosine deaminase. Subsequently, 5-FU is converted to 5-fluorouridine phosphate which is incorporated into RNA, resulting in the inhibition of protein synthesis. In addition, 5-FU is also converted to 5-fluorodeoxyuridine monophosphate, a strong inhibitor of thymidylate synthase, an enzyme involved in DNA synthesis and nuclear division. Thus, 5-FC acts by interfering with nucleotide metabolism, as well as RNA, DNA, and protein synthesis in the fungal cell (reviewed in Ref. 10).

YM95NC Attachment -- Page 120

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

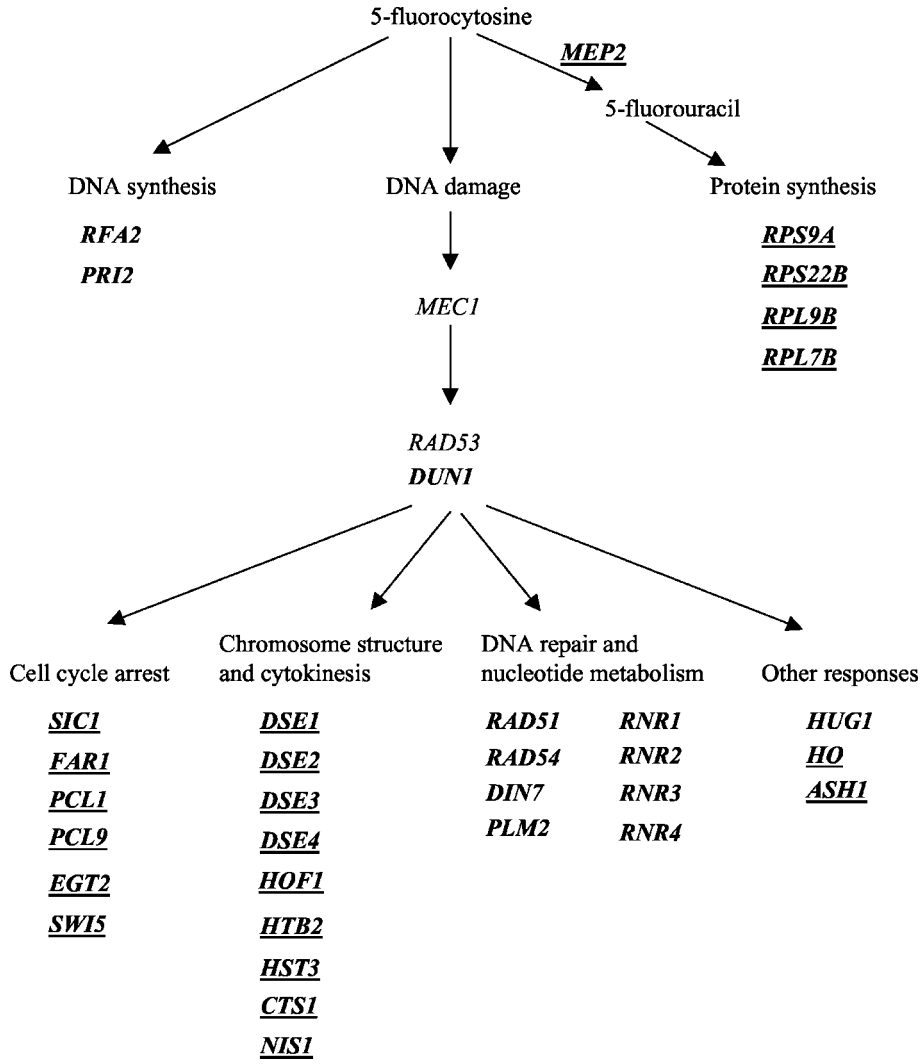


FIG. 9. **Summary of gene expression responses to 5-fluorocytosine.** Genes in *boldface* were responsive in this study. *Underlined* genes were repressed by 5-fluorocytosine treatment. Various components of the DNA damage response pathway are summarized.

The effect of 5-FC on protein synthesis was reflected in the repression of genes involved in this process, including genes encoding ribosomal proteins *RPS9A*, *RPL9B*, *RPS22B,* and *RPL7B* (Fig. 8). Similarly, the effect on DNA synthesis was reflected in the induction of genes involved in this process, including *RFA2* (encoding DNA replication factor A) and *PRI2* (encoding DNA primase) (Fig. 8). This result indicates that the cells responded to the inhibition of DNA synthesis by inducing genes required for DNA synthesis. In addition, the cells responded by inducing genes involved in DNA-damage repair. These include the DNA repair genes *RAD51*, *RAD54*, *DUN1,* and *DIN7* and the nucleotide metabolism genes *RNR1*, *RNR2*, *RNR3,* and *RNR4* (Fig. 8). These genes belong to the "DNA damage signature cluster" and are induced in response to DNA-damaging agents such as methylmethane sulfonate (MMS) and ionizing radiation (62). The gene *PLM2* has been implicated in the maintenance of the 2-μm plasmid in yeast and is also included in this cluster (62). Notably, *PLM2* was also induced in response to 5-FC in this study (represented in the *other function category*, Fig. 8). The genes *RAD51* and *RAD54* are involved in homologous recombination and the repair of double-strand DNA breaks (63). The gene *DUN1* is involved in cell cycle arrest and transcriptional regulation of genes in DNA damage response (64). The gene *DIN7* encodes a protein that is a structural homolog of the DNA repair genes *RAD2* and *RAD27* (65). The *RNR* genes encode subunits of the enzyme ribonucleotide reductase that catalyzes the rate-limiting step of deoxyribonucleotide biosynthesis (reviewed in Ref. 66). Their induction may be required to increase or alter the synthesis of nucleotide pools for DNA replication. The high level of induction of *RNR3* (~47 times) is consistent with previous observations where *RNR3* transcript levels increased up to 100-fold upon DNA damage (67). Thus, the induction of genes in the DNA repair and nucleotide metabolism category is consistent with the need to repair the DNA damage exerted by 5-FC.

In addition to DNA repair, cells also respond to DNA damage by arresting the cell cycle to provide time for repair. The DNA damage-response mechanism generates a signal that arrests cells in the G_1 phase of the cell cycle, slows down S phase (DNA synthesis), and arrests cells in the G_2 phase (reviewed in Ref. 68). Several cell cycle-regulated genes repressed by DNA-damaging agents such as MMS and ionizing radiation (60) were also repressed in response to 5-FC treatment in this study (Fig. 7). These genes include *SIC1*, *FAR1*, *PCL9,* and *EGT2* (cell cycle control); *HBT2* and *HST3* (chromosome/chromatin structure); *CTS1*, *HOF1, NIS1*, *DSE1*, *DSE2, DSE3,* and *DSE4* (cytokinesis); and *SWI5* and *ASH1* (transcription).

The genes *CTS1* and *EGT2* were also repressed in response to 5-FC in a previous microarray study (19). It was also shown that treatment of *S. cerevisiae* cells with 5-FC inhibited the separation of daughter cells from their mother cells (19), consistent with its effect on the cell cycle and cytokinesis. Although there was general agreement in the types of responses

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

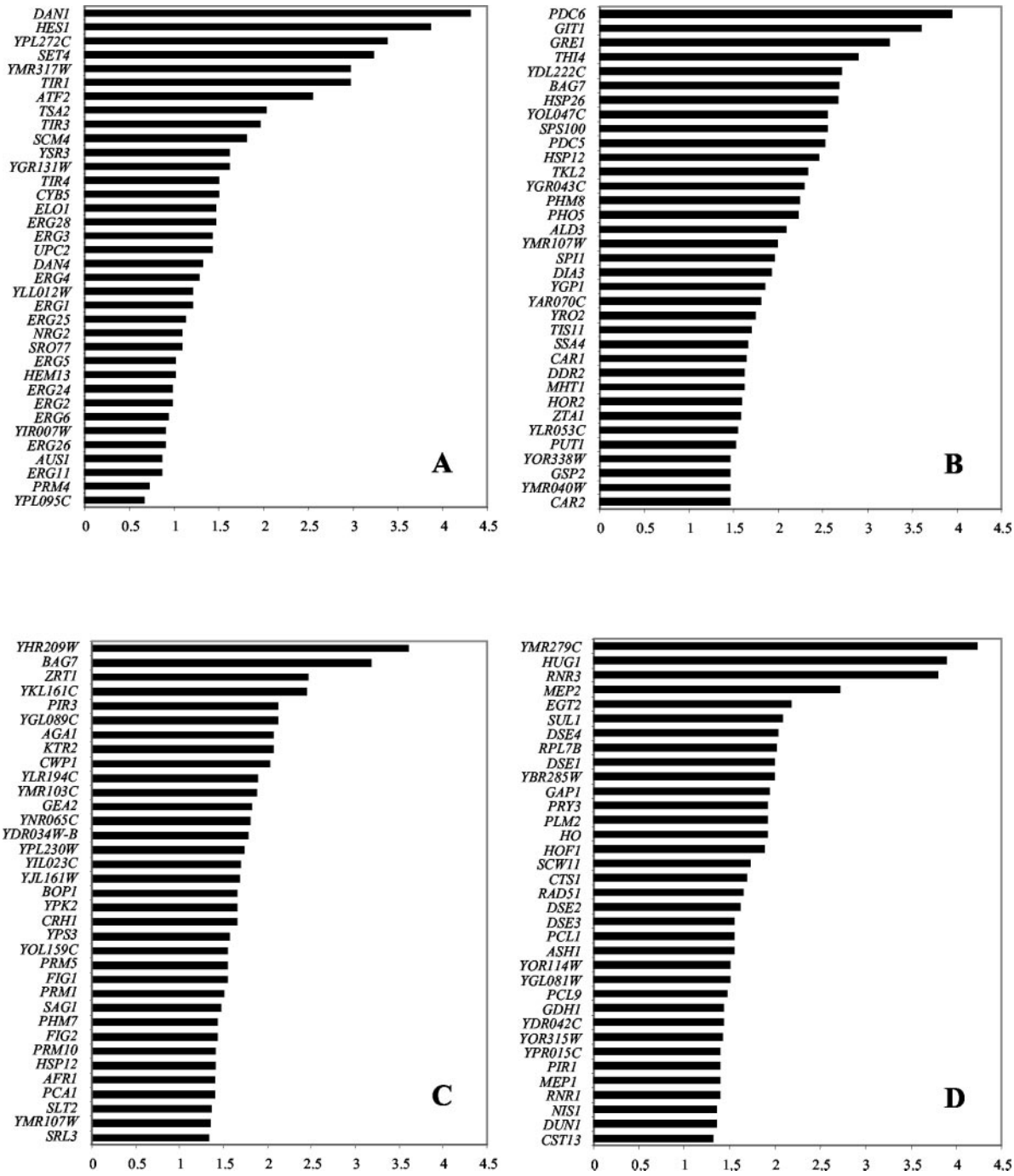


FIG. 10. **Drug-specific gene expression responses to ketoconazole (*A*), amphotericin B (*B*), caspofungin (*C*), and 5-fluorocytosine (*D*).** The *x* axis represents the DI value. The difference index was calculated as follows: DI = $\log_2$(fold change of gene *X* in drug exposure A) − $\log_2$(average fold change of gene *X* in all drug exposures). The DI represents how expression levels induced by one drug differ from the average expression level induced by all four drugs studied.

to 5-FC between the current study and the previous study by Zhang *et al.* (19), only 13 genes responded similarly in the two studies. This could be due to differences in the strains (L1190 was used in the previous study), media (rich medium was used in the previous study), and drug concentration used (~80-fold higher concentration of 5-FC was used in the previous study).

An additional gene that was down-regulated in response to 5-FC treatment in this study was the *HO* gene in the chromatin/chromosome category (Fig. 8). The *HO* gene encodes a homing endonuclease that introduces site-specific double-strand breaks in DNA at the *MAT* locus causing mating-type switching (69). Because 5-FC induces the DNA damage repair pathway, it makes sense that a gene such as *HO* that damages DNA would be down-regulated. The *HO* gene is regulated not only transcriptionally during late G_1 phase of the cell cycle but is also regulated post-transcriptionally during DNA damage response when the *HO* protein is degraded through the ubiquitin-26 S proteasome system (69). Based on the current study with 5-FC, it appears that *HO* is also regulated at the transcriptional level by the DNA damage-response pathway.

It is interesting to note that most of the cell cycle-regulated genes were down-regulated in response to 5-FC; however, the gene *HUG1* in the cell cycle control category was strongly induced (~47-fold) in this study (Fig. 8). Transcription of *HUG1* is induced in response to DNA damage and replication arrest (induced by hydroxyurea and UV and γ-radiation, see

YMSNC Attachment -- Page 122

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

TABLE II
Responses observed with all four drugs

Cellular role	ORF	Gene	Description
Cell cycle control	YJL157C	*FAR1*[a]	Factor arrest protein
Cell stress	YFL014W	*HSP12*	12-kDa heat shock protein
	YOL053C-A	*DDR2*	DNA damage-responsive
	YPL223C	*GRE1*	Induced by osmotic stress
Cell wall maintenance	YHL028W	***WSC4***[b]	Putative integral membrane protein with novel cysteine motif
	YDR055W	*PST1*	GPI[c]-attached protein
Lipid, fatty acid and sterol metabolism	YJL153C	*INO1*	L-Myo-inositol-1-phosphate synthase
Mating response	YNR044W	*AGA1*	Anchorage subunit of α-agglutinin
	YML047C	*PRM6*	Pheromone-regulated membrane protein
Protein degradation	YLR121C	*YPS3*	GPI-anchored aspartic protease
Transport	YBR068C	*BAP2*	Probable amino acid permease for leucine, valine, isoleucine
	YDR046C	*BAP3*	Valine transporter
	YKR039W	***GAP1***	General amino acid permease
	YNL142W	***MEP2***	Ammonium transport protein
	YJL212C	*OPT1*	Oligopeptide transporter
	YGR138C	*TPO2*	Polyamine transport protein
Unknown	YLR267W	*BOP2*	Bypass of PAM1
	YKR091W	***SRL3***	Suppressor of Rad53 null lethality
	YHR138C		Homologous to PBI2, proteinase inhibitor
	YGR043C		Hypothetical ORF[d]
	YPL088W		Hypothetical ORF
	YBR005W		Hypothetical ORF
	YGR146C		Hypothetical ORF
	YGL121C		Hypothetical ORF

[a] Non-highlighted genes responded in at least three of the four drug treatments.
[b] Highlighted genes responded in all four drug treatments.
[c] GPI, glycosylphosphatidylinositol.
[d] ORF, open reading frame.

Ref. 62). The *HUG1* gene has been identified as a component of the DNA-damage checkpoint response using deletion and overexpression mutants of *S. cerevisiae* (70). Thus, the induction of *HUG1* expression in response to 5-FC is consistent with its role in DNA damage response.

In addition to responses related to the mechanism of action of 5-FC, there was also a general induction of stress response (Fig. 8 and Supplemental Material, Table D). Genes induced in the cell stress category included *HSP12* and *HSP26* (heat stress, see Ref. 45); *GRE1* (osmotic stress, see Ref. 46); *DDR2* (oxidative stress, see Ref. 47); and *YGP1*, *XBP1*, *GTT1,* and *SHC1* (various stress conditions including heat, nitrogen depletion, and hydrogen peroxide, see Ref. 45). In addition, several genes induced in the ESR (45) were also induced in response to 5-FC treatment (Fig. 8 and Supplemental Material, Table D), including *HSP12*, *HSP26*, *YRO2*, *SSE2,* and *YPS3* (protein folding and degradation); *MTH1* (carbohydrate metabolism); *ECM4* and *SHC1* (cell wall maintenance); and *BAP2* and *BAP3* (metabolite transport). This result is consistent with previous observations in which the ESR was rapidly induced in response to DNA damage by MMS and ionizing radiation in *S. cerevisiae* cells (62).

There were also several genes in the transport category that were affected by 5-FC; in particular, the gene *MEP2* was repressed ~34-fold (Fig. 8). The gene *MEP2* encodes an NH_4^+ transport protein that is down-regulated when *S. cerevisiae* cells are grown in the presence of an appropriate nitrogen source (71). It is possible that the strong repression of *MEP2* in 5-FC-treated cells is due to the build up of NH_4^+ resulting from the deamination of 5-FC to 5-FU by cytosine deaminase. *MEP2* was also down-regulated in response to the DNA damage-inducing agents, MMS and ionizing radiation (62). As with caspofungin and amphotericin B, there are several genes in the unknown category that respond to 5-FC treatment. Some of these may also play a function in DNA damage response or alternatively, they may participate in an as-yet-unknown response to 5-FC. Fig. 9 summarizes the responses to 5-FC and highlights the genes whose expression was affected in this study.

Drug-specific Gene Expression Responses—In order to identify the gene expression responses specific to each drug tested, a difference index (DI) was generated (22). The difference index was calculated as follows, DI = $\log_2$(fold change of gene *X* in drug exposure A) − $\log_2$(average fold change of gene *X* in all four drug exposures). Thus, the DI represents how expression levels induced by one drug differ from the average expression level induced by all four drugs studied. Because the absolute value of fold changes were used as the log argument, genes that had negative expression values (*i.e.* were down-regulated) were assigned positive values. Thirty five genes with the highest DI values were represented in a graph (DI value *versus* gene) for each drug (Fig. 10). These graphs provide a clear indication of drug-specific responses. For example, responses specific to ketoconazole were in ergosterol pathway genes and in Upc2p-regulated genes (Fig. 10*A*). Specific responses to amphotericin B included genes in the following categories: cell stress, phosphate metabolism, transport, and cell wall integrity (Fig. 10*B*). Caspofungin-specific responses were in genes represented in the cell wall stress-response pathway and the mating response category (Fig. 10*C*). Effects specific to 5-fluorocytosine represented genes in the DNA damage repair, nucleotide metabolism, protein synthesis, cytokinesis, and cell cycle control categories (Fig. 10*D*). In addition, there are several genes in the unknown category that show specific responses to each of the four drugs. This may suggest a possible function of these genes in drug sensitivity or resistance. Thus, the DI graphs make it possible to generate "signature profiles" that represent specific gene expression responses to each of the four drugs tested. These signature profiles will be useful in determining the mechanism of action of novel compounds with antifungal activity.

Responses Observed with All Four Drugs—Table II shows a

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

YM96NC Attachment 1 Page 123

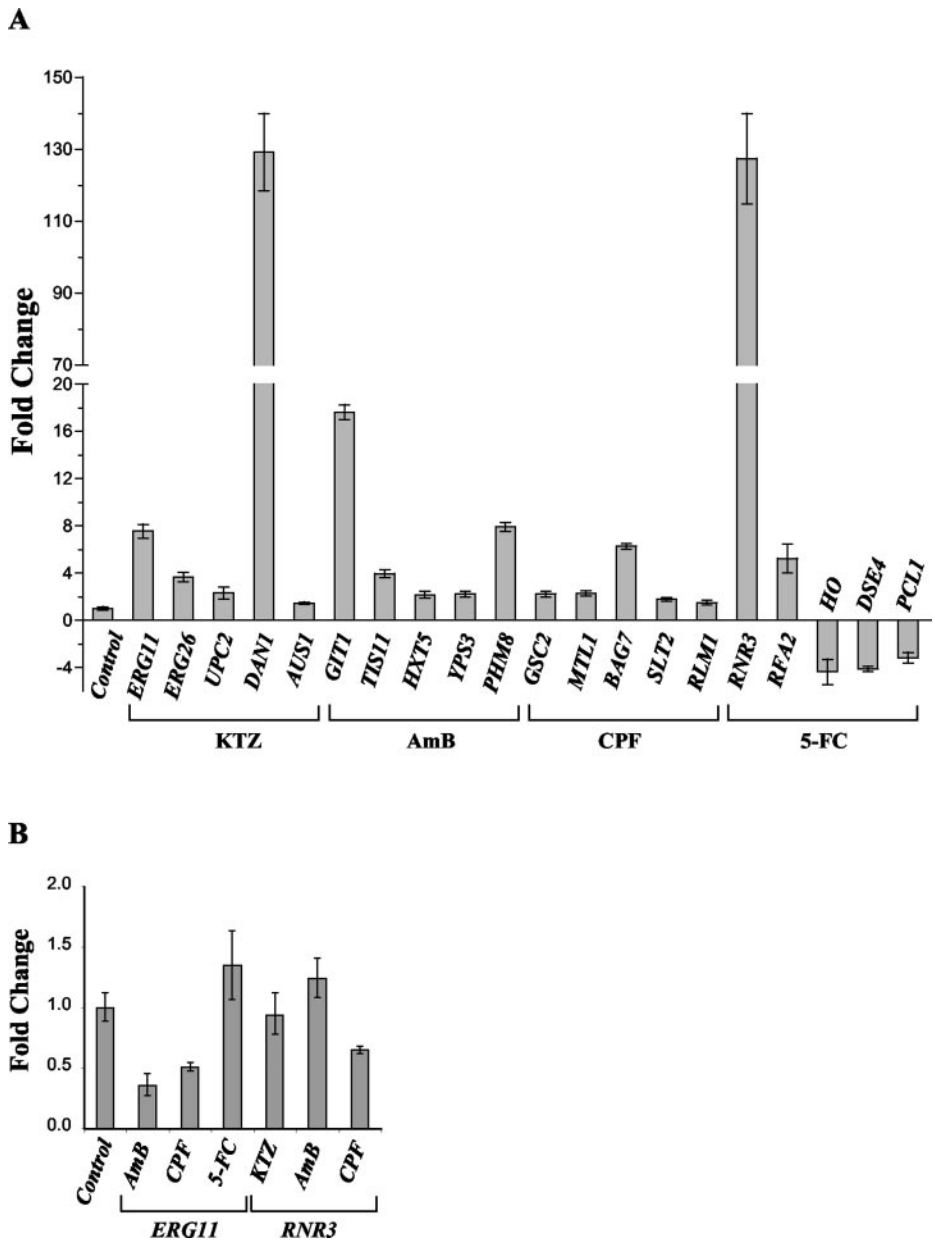


FIG. 11. **Quantitative real time RT-PCR analysis of genes identified as differentially expressed by microarray experiments.** Twenty genes identified as differentially expressed by microarray analysis were examined by quantitative real time RT-PCR with gene-specific primers. Assays were performed in triplicate with SYBR® Green I dye using the GenAmp® 5700 Sequence Detection System. Data were normalized to an internal control (18 S rRNA), and the $\Delta\Delta C_T$ method was used to obtain the relative expression level for each gene. Data are shown as mean ± S.D. *Error bars* were generated by transformation of mean $\Delta\Delta C_T$ values ± one S.D. unit, taken to 2^{-n}. *A*, gene expression changes in 20 genes (5 per drug) detected by real time RT-PCR. *B*, validation of drug-specific gene expression responses. Primers specific for *ERG11* were assayed in real time PCRs using RNA from AmB-, CPF-, and 5-FC-treated samples. Primers specific for *RNR3* were tested against RNA from ketoconazole-, amphotericin B-, and caspofungin-treated samples. *KTZ*, ketoconazole; *AmB,* amphotericin B; *CPF,* caspofungin.

list of genes that responded similarly to the four drug treatments. Non-highlighted genes represent those genes that responded to at least three of the four drugs tested, and genes in boldface italic represent those that responded to all four drug treatments. This list probably represents genes that are involved in survival mechanisms that yeast cells employ when exposed to toxic agents. It is worth noting that only 24 genes responded similarly to three different drug exposures, indicating that the majority of responses are not related to generalized survival mechanisms and therefore represent specific responses to each drug. However, some of the genes shown in Table II are represented in the DI graphs used to identify drug-specific responses (Fig. 10). These genes include *GRE1*, *HSP12,* and *DDR2* (amphotericin B); *AGA1* and *YPS3* (caspofungin); and *MEP2* and *GAP1* (5-FC). It is worth noting that these genes were strongly induced or strongly repressed in response to specific drug treatments. For example, *MEP2*, encoding an NH_4^+ transport protein, is ~34-fold repressed by 5-FC (Fig. 8), and this is possibly because of NH_4^+ accumulation in the cells due to 5-FC deamination. Similarly *GRE1*, a gene induced by osmotic stress, is ~93-fold up-regulated by amphotericin B (Fig. 4); this up-regulation indicates the loss of osmotic balance resulting from the leakage of ions due to membrane damage. These results indicate that the magnitude of the fold change for a particular gene may be specific for a given treatment, even though its expression level may show changes under a variety of conditions as a generalized stress response. Thus, in these types of drug exposure studies, it is also important to examine the data using tools such as the difference index to identify specific *versus* nonspecific responses to drug treatments.

Validation of Microarray Data by Real Time RT-PCR—To validate the differential expression of genes obtained by microarray analysis, real time RT-PCR was performed using the same RNA from the original microarray experiment. Twenty genes (5 per drug) were selected not only to confirm their importance in the mechanism of action of the respective drugs (*e.g. ERG11* and *ERG26* for ketoconazole; *GSC2* for caspofungin; and *RNR3* and *RFA2* for 5-FC) but also to verify the novel responses identified in this work (*e.g. UPC2*, *DAN1,* and *AUS1* for ketoconazole; *PHM8*, *HXT5,* and *YPS3* for amphotericin B; *MTL1*, *RLM1*, *BAG7,* and *SLT2* for caspofungin; and *HO*,

YM96NC Attachment - Page 124

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

DSE4, and *PCL1* for 5-FC). For all 20 genes, there was complete correlation between real time RT-PCR and microarray data (Fig. 11*A*). In agreement with the microarray data, 17 genes showed up-regulation and 3 genes showed down-regulation in response to drug treatment. Furthermore, genes that showed high levels of induction in the microarray experiment also showed high levels of induction in real time RT-PCR assays (*e.g. DAN1* in response to ketoconazole and *RNR3* in response to 5-FC). However, *DAN1* and *RNR3* showed a higher level of induction in real time RT-PCR assays (129- and 127-fold induction, respectively, Fig. 11*A*) compared with the microarray data (54- and 47-fold induction, respectively, Figs. 2 and 8). This discrepancy could be attributed to the fact that the Affymetrix software assigned an "absent" call to both genes in the untreated sample, *i.e.* these genes were undetectable in the untreated samples and were highly induced upon drug treatment. Fold induction values obtained from microarray analysis can be particularly misleading in cases where poor or no hybridization signals are generated in one of the samples due to low expression levels of specific genes in that sample.

For the remaining 18 genes, the levels of gene induction did not differ markedly between microarray data and real time RT-PCR data. A few minor discrepancies that were observed could be explained, in part, by the greater dynamic range of real time RT-PCR assays. In addition, cross-hybridization may occur in microarray experiments with splice variants or related genes. For example, *HXT5* shows 65–75% identity at the nucleotide level to 17 related *HXT* genes.

In order to provide further confirmation of gene-specific responses to the drugs tested, additional negative control assays were performed. Primers specific for the *ERG11* gene were used in real time RT-PCR assays with RNA from amphotericin B-, caspofungin-, and 5-FC-treated samples. For all three samples, there was no significant induction of *ERG11* compared with the untreated control (Fig. 11*B*). Similarly, primers specific for the *RNR3* gene were tested with RNA from ketoconazole-, amphotericin B-, and caspofungin-treated samples. No significant induction of *RNR3* was detected in the three samples assayed (Fig. 11*B*). These results provide further evidence that the genes assayed by real time RT-PCR respond specifically to the respective drug treatments.

Susceptibility of upc2 Mutant Strains to Ketoconazole—Because ketoconazole treatment resulted in the induction of *UPC2* (Fig. 2), a gene that encodes a transcription factor involved in regulating sterol biosynthesis and uptake, it is possible that any alteration in *UPC2* levels would have an effect on the susceptibility to ketoconazole. The *upc2-1* mutant accumulates ~6-fold more exogenous sterols compared with the wild type (39), which could reduce the level of inhibition caused by ketoconazole. Conversely, the deletion of *UPC2* results in reduced sterol uptake (39, 72) and could cause an increase in the inhibitory effect of ketoconazole. To test this hypothesis, the susceptibility of *S. cerevisiae* to ketoconazole was tested in the *upc2-1* mutant and a *upc2* deletion mutant.

Experiments were performed in the presence of erogosterol (10 μg/ml), and cells were grown under anaerobic conditions to allow exogenous sterol uptake. It is known that *S. cerevisiae* cells are not able to uptake sterols under aerobic conditions due to the phenomenon of aerobic sterol exclusion, *i.e.* under aerobic conditions *S. cerevisiae* cells synthesize ergosterol and are not dependent upon exogenous sterols for survival (73).

As can be seen in Fig. 12, the *upc2-1* mutant was less susceptible to ketoconazole under these growth conditions. The IC_{50} for ketoconazole increased from 2 μg/ml in the wild type strain to 4.8 μg/ml in the *upc2-1* mutant, and the MIC (concentration at which there was no detectable growth) increased from 3.7 to 11.1 μg/ml, respectively (Fig. 12*A*). No significant difference in ketoconazole susceptibility was detected under aerobic growth conditions even though the *upc2-1* mutant is capable of accumulating exogenous sterols during aerobic growth (data not shown). It is possible that the amount of sterol accumulated in the *upc2-1* mutant under aerobic conditions is not enough to counterbalance the effect of ketoconazole. Because *UPC2* is induced under anaerobic conditions (74, 75), it is likely that higher levels of sterol are accumulated anaerobically resulting in reduced susceptibility to ketoconazole.

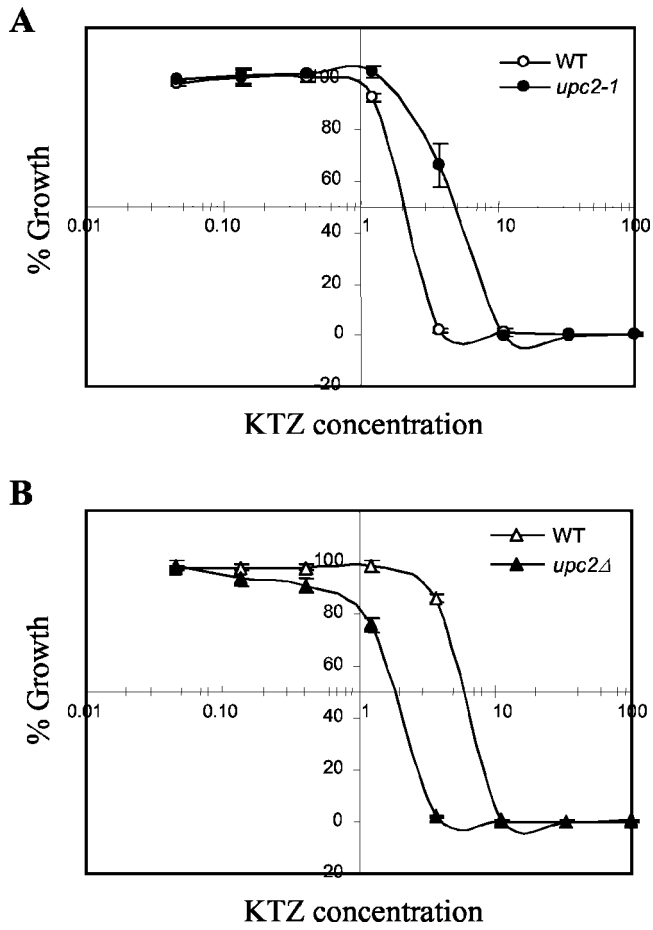


FIG. 12. **Dose-response curves of ketoconazole-treated cells.** Mutant strains and their isogenic wild type strains were grown in microplates, at 30 °C for 48 h, in SC medium containing varying concentrations of ketoconazole and 10 μg/ml ergosterol. Microplates were incubated under anaerobic conditions as described under "Experimental Procedures." Four replicates of each assay were performed. Percent growth is shown as mean ± S.D. *A*, susceptibility to ketoconazole of *upc2-1* mutant and its isogenic wild type (*WT*). *B*, susceptibility to ketoconazole of *upc2*Δ mutant and its isogenic wild type. *KTZ*, ketoconazole.

Comparison of the *upc2*Δ mutant to its isogenic wild type indicated that the *upc2*Δ mutant was more susceptible to ketoconazole under the experimental conditions tested (Fig. 12*B*). The IC_{50} decreased from 6 μg/ml in the wild type strain to 1.8 μg/ml in the *upc2*Δ strain, and the MIC decreased from 11.1 to 3.7 μg/ml, respectively. The increased susceptibility to ketoconazole of the *upc2*Δ mutant under anaerobic conditions is consistent with the hypothesis that reduced sterol uptake in the mutant would increase the inhibitory effect of ketoconazole.

Interestingly, a recent report (21) indicated that *UPC2* was induced in *S. cerevisiae* cells exhibiting reduced susceptibility to fluconazole and itraconazole. Similar results were reported in *S. cerevisiae* cells overexpressing the *SUT1* gene (encoding a second transcription factor involved in regulating sterol up-

YM5NC Attachment - Page 125

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

take), which showed increased resistance to the ergosterol biosynthesis inhibitor fenpropimorph, in medium supplemented with ergosterol (24). Taken together, the current study suggests an important role for *UPC2* in determining sensitivity to ergosterol biosynthesis inhibitors.

In conclusion, the microarray studies reported here have revealed specific changes in gene expression consistent with known mechanisms of action, changes suggesting previously unidentified effects of these agents, and nonspecific changes to antifungal agents in general. Gene expression profiling will likely be useful for screening compounds to identify candidates with specific mechanisms of antifungal activity. The approach will also be useful in identifying the mechanisms of action for compounds identified as exhibiting activity through other screens. These findings lay the groundwork for the application of functional genomics to identify gene expression profiles that will be useful in an antifungal drug development program.

Acknowledgments—We thank Dr. Divyen Patel of Genome Explorations, Inc. (Memphis, TN), for conducting the Affymetrix gene array hybridization. We are grateful to Drs. Stephen Sturley and Lisa Wilcox (Columbia University, College of Physicians and Surgeons, New York) for providing strains and for helpful discussions. We also thank Melanie Mask (United States Department of Agriculture, Natural Products Utilization Research Unit, University, MS) for technical assistance with quantitative real time RT-PCR assays.

REFERENCES

1. Groll, A. H., Shah, P. M., Mentzel, C., Schneider, M., Just-Nuebling, G., and Huebner, K. (1996) *J. Infect.* **33,** 23–32
2. Viscoli, C., Girmenia, C., Marinus, A., Collette, L., Martino, P., Vandercam, B., Doyen, C., Lebeau, B., Spence, D., Krcmery, V., De Pauw, B., and Meunier, F. (1999) *Clin. Infect. Dis.* **28,** 1071–1079
3. Wald, A., Leisenring, W., van Burik, J. A., and Bowden, R. A. (1997) *J. Infect. Dis.* **175,** 1459–1466
4. Groll, A. H., and Walsh, T. J. (2001) *Clin. Microbiol. Infect.* **7,** 8–24
5. Groll, A. H., Piscitelli, S. C., and Walsh, T. J. (1998) *Adv. Pharmacol.* **44,** 343–500
6. Sanglard, D. (2002) *Curr. Opin. Microbiol.* **5,** 379–385
7. Kontoyiannis, D. P., and Lewis, R. E. (2002) *Lancet* **359,** 1135–1144
8. Gallis, H. A., Drew, R. H., and Pickard, W. W. (1990) *Rev. Infect. Dis.* **12,** 308–329
9. Lyman, C. A., and Walsh, T. J. (1992) *Drugs* **44,** 9–35
10. Ghannoum, M. A., and Rice, L. B. (1999) *Clin. Microbiol. Rev.* **12,** 501–517
11. Kelly, S. L., Arnoldi, A., and Kelly, D. E. (1993) *Biochem. Soc. Trans.* **21,** 1034–1038
12. Groll, A. H., and Walsh, T. J. (2001) *Exp. Opin. Invest. Drugs* **10,** 1545–1558
13. Lockhart, D. J., and Winzeler, E. A. (2000) *Nature* **405,** 827–836
14. Barns, S. M., Lane, D. J., Sogin, M. L, Bibeau, C., and Weisburg, W. G. (1991) *J. Bacteriol.* **173,** 2250–2255
15. Ponton, J., Ruchel, R., Clemons, K. V., Coleman, D. C., Grillot, R., Guarro, J., Aldebert, D., Ambroise-Thomas, P., Cano, J., Carrillo-Munoz, A. J., Gene, J., Pinel, C., Stevens, D. A., and Sullivan, D. J. (2000) *Med. Mycol.* **38,** 225–236
16. Bammert, G. F., and Fostel, J. M. (2000) *Antimicrob. Agents Chemother.* **44,** 1255–1265
17. De Backer, M. D., Ilyina, T., Ma, X.-J., Vandoninck, S., Luyten, W. H. M. L., and Bossche, H. V. (2001) *Antimicrob. Agents Chemother.* **45,** 1660–1670
18. Zhang, L., Zhang, Y., Zhou, Y., An, S., Zhou, Y., and Cheng, J. (2002) *J. Antimicrob. Chemother.* **49,** 905–915
19. Zhang, L., Zhang, Y., Zhou, Y., Zhao, Y., Zhou, Y., and Cheng, J. (2002) *Int. J. Antimicrob. Agents* **20,** 444–450
20. Lorenz, M. C., and Fink, G. R. (2001) *Nature* **412,** 83–86
21. Barker, K. S., Pearson, M. M., and Rogers, P. D. (2003) *J. Antimicrob. Chemother.* **51,** 1131–1140
22. Nau, G. J., Richmond, J. F. L., Schleisinger, A., Jennings, E. G., Lander, E. S., and Young, R. A. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1503–1508
23. Engels, W. R. (1993) *Trends Biochem. Sci.* **18,** 448–450
24. Bourot, S., and Karst, F. (1995) *Gene* (*Amst.*) **165,** 97–102
25. National Committee for Clinical Laboratory Standards (1997) *Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts; Approved Standard,* NCCLS document M 27-A, NCCLS, Wayne, PA
26. Klepser, M. E., Ernst, E. J., Lewis, R. E., Ernst, M. E., and Pfaller, M. A. (1998) *Antimicrob. Agents Chemother.* **42,** 1207–1212
27. Lee, M-L. T., Kuo, F. C., Whitmore, G. A., and Sklar, J. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 9834–9839
28. Lee, J. K. (2001) *Clin. Chem.* **47,** 1350–1352
29. Yoshida, Y., and Aoyama, Y. (1987) *Biochem. Pharmacol.* **36,** 229–235
30. Morschhaüser, J. (2002) *Biochim. Biophys. Acta* **1587,** 240–248
31. Truan, G., Epinat, J. C., Rougeulle, C., Cullin, C., and Pompon, D. (1994) *Gene* (*Amst.*) **142,** 123–127
32. Jiang, B., Brown, J. L., Sheraton, J., Fortin, N., and Bussey, H. (1994) *Yeast* **10,** 341–353
33. Mao, C., and Obeid, L. M. (2000) *Methods Enzymol.* **311,** 223–232
34. Hirsch, J. P., and Henry, S. A. (1986) *Mol. Cell Biol.* **10,** 3320–3328
35. Oh, C. S., Toke, D. A., Mandala, S., and Martin, C. E. (1997) *J. Biol. Chem.* **272,** 17376–17384
36. Swain, E., Baudry, K., Stukey, J., McDonough, V., Germann, M., and Nickels, J. T. (2002) *J. Biol. Chem.* **277,** 26177–26184
37. Vik, A., and Rine, J. (2001) *Mol. Cell Biol.* **19,** 6395–6405
38. Lewis, T. L., Keesler, G. A., Fenner, G. P., and Parks, L. W. (1988) *Yeast* **4,** 93–106
39. Wilcox, L. J., Balderes, D. A., Wharton, B., Tinkelenberg, A. H., Rao, G., and Sturley, S. L. (2002) *J. Biol. Chem.* **277,** 32466–32472
40. Parks, L. W., and Casey, W. M. (1995) *Annu. Rev. Microbiol.* **49,** 95–116
41. Lewis, T. A., Taylor, F. R., and Parks, L. W. (1985) *J. Bacteriol.* **163,** 199–207
42. Joseph-Horne, T., and Hollomon, D. W. (1997) *FEMS Microbiol. Lett.* **149,** 141–149
43. Hartsel, S., and Bolard, J. (1996) *Trends Pharmacol. Sci.* **17,** 445–449
44. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A. P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K. D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kotter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C. Y., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W., and Johnston, M. (2002) *Nature* **418,** 387–391
45. Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., and Brown, P. O. (2000) *Mol. Biol. Cell* **11,** 4241–4257
46. Hirayama, T., Maeda, T., Saito, H., and Shinozaki, K. (1995) *Mol. Gen. Genet.* **249,** 127–138
47. Koerkamp, M. G., Rep, M., Bussemaker, H. J., Hardy, G. P., Mul, A., Piekarska, K., Szigyarto, C. A., De Mattos, J. M., and Tabak, H. F. (2002) *Mol. Biol. Cell* **13,** 2783–2794
48. Destruelle, M., Holzer, H., and Klionsky, D. J. (1994) *Mol. Cell. Biol.* **14,** 2740–2754
49. Alexandre, H., Ansanay-Galeote, V., Dequin, S., and Blondin, B. (2001) *FEBS Lett.* **498,** 98–103
50. Moskvina, E., Schuller, C., Maurer, C. T., Mager, W. H., and Ruis, H. (1998) *Yeast* **14,** 1041–1050
51. Smits, G. J., van den Ende, H., and Klis, F. M. (2001) *Microbiology* **147,** 781–794
52. Schmidt, A., Schmelzle, T., and Hall, M. N. (2002) *Mol. Microbiol.* **45,** 1433–1441
53. Ma, Q., and Herschman, H. R. (1995) *Oncogene* **10,** 487–494
54. Ogawa, N., DeRisi, J., and Brown, P. O. (2000) *Mol. Biol. Cell* **11,** 4309–4321
55. Popolo, L., Gualtieri, T., and Ragni, E. (2001) *Med. Mycol.* **39,** 111–121
56. Terashima, H., Yabuki, N., Arisawa, M., Hamada, K., and Kitada, K. (2000) *Mol. Gen. Genet.* **264,** 64–74
57. Lipke, P. N., and Ovalle, R. (1998) *J. Bacteriol.* **180,** 3735–3740
58. Osherov, N., May, G. S., Albert, N. D., and Kontoyiannis, D. P. (2002) *Antimicrob. Agents Chemother.* **46,** 2462–2469
59. Jung, U. S., and Levin, D. E. (1999) *Mol. Microbiol.* **34,** 1049–1057
60. Schmelzle, T., Helliwell, S. B., and Hall, M. N. (2002) *Mol. Cell Biol.* **22,** 1329–1339
61. Mattison, C. P., Spencer, S. S., Kresge, K. A., Lee, J., and Ota, I. M. (1999) *Mol. Cell. Biol.* **19,** 7651–7660
62. Gasch, A. P., Huang, M., Metzner, S., Botstein, D., Elledge, S. J., and Brown, P. O. (2001) *Mol. Biol. Cell* **12,** 2987–3003
63. Rattray, A. J., and Symington, L. S. (1995) *Genetics* **139,** 45–56
64. Zhou, Z., and Elledge, S. J. (1993) *Cell* **75,** 1119–1127
65. Mieczkowski, P. A., Fikus, M. U., and Ciesla, Z. (1997) *Mol. Gen. Genet.* **253,** 655–665
66. Stubbe, J. (1990) *J. Biol. Chem.* **265,** 5329–5332
67. Elledge, S. J., and Davis, R. W. (1990) *Genes Dev.* **4,** 740–751
68. Elledge, S. J. (1996) *Science* **274,** 1664–1672
69. Kaplun, L., Ivantsiv, Y., Kornitzer, D., and Raveh, D. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10077–10082
70. Basrai, M. A., Velculescu, V. E., Kinzler, K. W., and Hieter, P. (1999) *Mol. Cell. Biol.* **19,** 7041–7049
71. Marini, A. M., Soussi-Boudekou, S., Vissers, S., and Andre, B. (1997) *Mol. Cell. Biol.* **17,** 4282–4293
72. Crowley, J. H., Leak, F. W., Jr., Shianna, K. V., Tove, S., and Parks, L. W. (1998) *J. Bacteriol.* **180,** 4177–4183
73. Lorenz, R. T., and Parks, L. W. (1991) *Lipids* **8,** 598–603
74. ter Linde, J. J., Liang, H., Davis, R. W., Steensma, H. Y., van Dijken, J. P., and Pronk, J. T. (1999) *J. Bacteriol.* **181,** 7409–7413
75. Abramova, N. E., Cohen, B. D., Sertil, O., Kapoor, R., Davies, K. J., and Lowry, C. V. (2001) *Genetics* **157,** 1169–1177

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

Genome-wide Expression Profiling of the Response to Polyene, Pyrimidine, Azole, and Echinocandin Antifungal Agents in *Saccharomyces cerevisiae*

Ameeta K. Agarwal, P. David Rogers, Scott R. Baerson, Melissa R. Jacob, Katherine S. Barker, John D. Cleary, Larry A. Walker, Dale G. Nagle and Alice M. Clark

J. Biol. Chem. 2003, 278:34998-35015.
doi: 10.1074/jbc.M306291200 originally published online June 24, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M306291200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2003/07/23/M306291200.DC1

This article cites 71 references, 32 of which can be accessed free at
http://www.jbc.org/content/278/37/34998.full.html#ref-list-1

Downloaded from http://www.jbc.org/ at Research Library Information Grp, National Institute of Standards and Technology on June 8, 2018

MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 5404–5417 doi: 10.1111/j.1365-294X.2012.05732.x

Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity

QI-MING WANG,*[1] WAN-QIU LIU,*[1] GIANNI LITI,† SHI-AN WANG*‡ and FENG-YAN BAI*

*State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China, †Faculté de Médecine, Institute for Research on Cancer and Aging in Nice (IRCAN), INSERM U1081, CNRS UMR 7284, University of Nice, Nice, France

Abstract

The budding yeast, *Saccharomyces cerevisiae*, is a leading system in genetics, genomics and molecular biology and is becoming a powerful tool to illuminate ecological and evolutionary principles. However, little is known of the ecology and population structure of this species in nature. Here, we present a field survey of this yeast at an unprecedented scale and have performed population genetics analysis of Chinese wild isolates with different ecological and geographical origins. We also included a set of worldwide isolates that represent the maximum genetic variation of *S. cerevisiae* documented so far. We clearly show that *S. cerevisiae* is a ubiquitous species in nature, occurring in highly diversified substrates from human-associated environments as well as habitats remote from human activity. Chinese isolates of *S. cerevisiae* exhibited strong population structure with nearly double the combined genetic variation of isolates from the rest of the world. We identified eight new distinct wild lineages (CHN I –VIII) from a set of 99 characterized Chinese isolates. Isolates from primeval forests occur in ancient and significantly diverged basal lineages, while those from human-associated environments generally cluster in less differentiated domestic or mosaic groups. Basal lineages from primeval forests are usually inbred, exhibit lineage-specific karyotypes and are partially reproductively isolated. Our results suggest that greatly diverged populations of wild *S. cerevisiae* exist independently of and predate domesticated isolates. We find that China harbours a reservoir of natural genetic variation of *S. cerevisiae* and perhaps gives an indication of the origin of the species.

Keywords: biogeography, ecology, evolution, population genetics, *Saccharomyces cerevisiae*

Received 25 May 2011; revision received 6 June 2012; accepted 13 June 2012

Introduction

The budding yeast *Saccharomyces cerevisiae* is used worldwide as a microbial agent for baking, fermentation of wine, beer and other alcoholic beverages. This organism is also an extensively used model system in genetics and molecular biology and was the first sequenced eukaryote genome (Goffeau *et al.* 1996). Two-thirds of the ~6000 identified genes of the species have been characterized (Kumar & Snyder 2001; Peña-Castillo & Hughes 2007). As one of the most thoroughly studied eukaryotic species with the advantages of a short generation time, control of the sexual cycle and easy genome manipulation, *S. cerevisiae* is now becoming an important model in ecology, population and evolutionary genetics (Dickinson 2000; Zeyl 2000; Liti & Louis 2005; Landry *et al.* 2006; Ruderfer *et al.* 2006; Replansky *et al.* 2008).

Our understanding of the ecology and population structure of *S. cerevisiae* in the wild is still limited. Until recently, *S. cerevisiae* was considered a domesticated

Correspondence: Feng-Yan Bai, Fax: +86 10 6480 7406;
E-mail: baify@im.ac.cn
‡Present address: Key Laboratory of Biofuels, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, 266101, China.
[1]These authors contributed equally to this work.

© 2012 Blackwell Publishing Ltd

species and no truly natural isolates existed. The occasional isolates of *S. cerevisiae* found in nature were thought to be feral strains with origins from domestic stocks (Martini 1993; Vaughan-Martini & Martini 1995; Naumov 1996; Ciani *et al.* 2004). Because of the potential complications of domestication, its closest relative, *S. paradoxus*, is considered a better model for ecology and evolutionary biology (Greig 2007; Replansky *et al.* 2008; Tsai *et al.* 2008). *S. paradoxus* is found globally coexisting frequently with *S. cerevisiae* but rarely associates with human activity (Naumov *et al.* 1998; Sniegowski *et al.* 2002; Sampaio & Gonçalves 2008). Phylogenetic analysis of multiple genes in *S. paradoxus* populations has shown that isolates from Europe form a single recombining population within which genetic differentiation increases with physical distance (Koufopanou *et al.* 2006). No shared polymorphisms were found between European, Far East Asian and North American *S. paradoxus* populations, suggesting that there are at least three independent lineages (Johnson *et al.* 2004; Koufopanou *et al.* 2006; Liti *et al.* 2006; Kuehne *et al.* 2007). A few European isolates were also isolated in North America presumably following a transoceanic migratory event, suggesting secondary contact between European and North American populations. However, genetic admixture was not detected between the two populations, reinforcing allopatric *S. paradoxus* populations that diverged genetically and evolved reproductive barriers (Kuehne *et al.* 2007).

A different population structure scenario has been described in *S. cerevisiae*. This species exists in both wild and domesticated populations that are phylogenetically distinct (Fay & Benavides 2005). Early studies showed that genetic variation among *S. cerevisiae* isolates was limited compared with *S. paradoxus* isolates and not clearly correlated with geographical location (Sniegowski *et al.* 2002; Aa *et al.* 2006; Liti *et al.* 2006). A population genomics study based on genome resequencing of over seventy isolates of *S. cerevisiae* and *S. paradoxus* sampled from different ecological niches and from locations on different continents confirmed that *S. paradoxus* populations are well delineated along geographical boundaries, whereas the worldwide *S. cerevisiae* isolates show less differentiation and are comparable with a single *S. paradoxus* population (Liti *et al.* 2009). From a set of 38 *S. cerevisiae* isolates with worldwide origin, five well-defined, geographically isolated lineages (Malaysian, West African, Sake, North American and 'Wine/European') and many different recombinant (mosaics) strains of these lineages were characterized (Liti *et al.* 2009). Surveying of genome scale single nucleotide polymorphisms (SNPs) by microarray analysis in 63 *S. cerevisiae* isolates from similar ecological and geographical origins as those employed in Liti *et al.* (2009) revealed three distinct lineages (wine, sake and laboratory), which reflect different ecological but not geographical origins (Schacherer *et al.* 2009).

Previous studies suggested separate domestication/selection events in *S. cerevisiae* (Fay & Benavides 2005; Liti *et al.* 2009; Schacherer *et al.* 2009) and presented evidence supporting the idea that human activity provided opportunity for population mixing and recombination (Liti *et al.* 2009; Magwene *et al.* 2011). The limited population differentiation observed in *S. cerevisiae* is suggested to be the consequence of the long historic association of the species with humans (Replansky *et al.* 2008). Nevertheless, these studies have also shown that *S. cerevisiae* as a whole is not domesticated. Therefore, truly wild isolates are important to elucidate the population structure of *S. cerevisiae*. However, wild populations of the species were poorly represented in previous studies and were isolated from a limited range of ecological niches. These strains were mainly from vineyards, the bark of oak trees and associated soil (Sniegowski *et al.* 2002; Aa *et al.* 2006). The vineyard isolates may not be truly wild. The oak isolates of *S. cerevisiae* have been considered to be truly 'wild' (Fay & Benavides 2005); however, the association of these isolates with human activities cannot be excluded. The sampled oak trees were in parks or arboreta, belonging to man-made environments or environments frequently visited by humans. Therefore, the population structure of *S. cerevisiae* in the real wild remains obscure. The limited number of wild isolates in previous population genetics and genomics surveys precluded a reasonable inference of the origin of *S. cerevisiae*. Although hypotheses include that modern *S. cerevisiae* may have originated in Eastern Asia (Naumov & Nikonenko 1988) or that the Malaysian population of *S. cerevisiae* might be the most ancient representative of the species (Naumov *et al.* 2006), solid evidences supporting these hypotheses have not been presented.

In the past few years, we have successfully isolated *S. cerevisiae* from an unprecedented scale of diversified substrates. These samples were collected from human-associated environments as well as habitats rarely visited by humans, including primeval forests located in different regions covering a wide range of climate zones from northern to southern China. Population genetics analysis of these wild *S. cerevisiae* isolates offers new insights into the ecological distribution, population structure and biogeography of this species.

Materials and methods

Yeast isolation and identification

Different substrates including damaged fruit, tree bark, rotten wood and soil were collected from 2006 to 2009

© 2012 Blackwell Publishing Ltd

from markets, orchards, planted forests, original secondary forests and primeval forests. These habitats were located in 11 provinces covering temperate, subtropical and tropical climate zones from northern to southern China. Yeasts were isolated using the enrichment method as described in Sniegowski *et al.* (2002) with minor modifications. The enrichment medium contained (w/v) 0.3% yeast extract, 0.3% malt extract, 0.5% peptone, 1% glucose, 8% (v/v) ethanol, 200 μg of chloramphenicol/mL and 1 mL of 1 M HCl per litre. After 7 days' incubation at 25 °C, the enrichment medium was diluted and spread on WL nutrient agar plates (MacFaddin 1985) which were then incubated at 25 °C. Yeast colonies with different colour and morphology were isolated for further study. *S. cerevisiae* isolates were screened and identified using single-strand conformation polymorphism analysis of the internal transcribed spacer 1 (ITS1) region of ribosomal RNA genes and subsequently confirmed by sequencing the whole ITS region (Wang *et al.* 2008).

DNA sequencing and molecular phylogenetic analysis

Nine genes and four intergenic loci were selected based on previous studies (Fay & Benavides 2005; Aa *et al.* 2006; Ayoub *et al.* 2006) and sequenced in a selection of Chinese *S. cerevisiae* (Table S1, Supporting information). Corresponding sequences from the *S. cerevisiae* and *S. paradoxus* isolates released by the *Saccharomyces* Genome Resequencing Project (SGRP) (Liti *et al.* 2009) were retrieved as references from GenBank or the SGRP web site. Sequences were aligned using Clustal X program (Thompson *et al.* 1997). Phylogenetic trees were constructed from the evolutionary distance data calculated from Kimura's two-parameter model (Kimura 1980) by using the neighbour-joining method (Saitou & Nei 1987). Bootstrap analyses (Felsenstein 1985) were performed from 1000 random resamplings. Nucleotide diversity was analysed using the method of Tajima (1989). MEGA 5 was used to draw phylogenetic trees and calculate sequence diversity (Tamura *et al.* 2011).

Recombination analyses

Different methods were used to determine the extent of recombination and clonality. First, linkage disequilibrium was evaluated by measuring the standardized index of association (Smith *et al.* 1993). Two measures of the index of association (I_A and rBarD) for the 13 loci sequenced were calculated using the program MultiLocus 1.3 (Agapow & Burt 2001) and statistically tested against 1000 randomized data sets. Second, gene genealogy concordance was inspected by visual comparison of topologies of phylogenetic trees drawn from combined and individual gene sequences using the neighbour-joining method (Saitou & Nei 1987). Third, incongruence length difference/partition homogeneity test (ILD/PHT) was performed using PAUP* 4.0beta10 (Swofford 2003). The ILD/PHT analysis was achieved using maximum parsimony heuristic searches with 10 000 replicates. Finally, recombination was further detected using DnaSP 5.10 (Librado & Rozas 2009), and the minimum numbers of recombination events were obtained using the four-gamete test (Hudson & Kaplan 1985).

Population structure analyses

Population structure was inferred using the program Structure 2.3 (Pritchard *et al.* 2000; Falush *et al.* 2003) based on 827 SNPs from the 13 loci sequenced. An admixture model with correlated allele frequencies was used. The result was visualized using Distruct 1.1 (Rosenberg 2004). Patterns of allelic relatedness of multilocus genotypes were also evaluated by the principal component analysis (PCA) using the EIGENSOFT 4.2 package (http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html) which implements methods from Patterson *et al.* (2006) and Price *et al.* (2006).

PFGE karyotype analysis

Intact yeast chromosomal DNA was prepared using the method of Bai *et al.* (2000). Pulsed field gel electrophoresis (PFGE) was carried out at 6 v/cm for 16 h with a switch time of 60 s and then for 11 h with a switch time of 90 s using a contour-clamped homogeneous electric field electrophoresis apparatus (CHEF-Mapper XA; Bio-Rad) as described in Wang & Bai (2008). Chromosome DNA from *S. cerevisiae* S288c was used as molecular size marker.

Genetic cross and spore viability analysis

Highly fertile strains representing different lineages of Chinese wild *S. cerevisiae* were selected for genetic crosses (Table 1). Strains DBVPG 6765 and UWOPS03-461.4 representing Wine/European and Malaysian lineages, respectively (Liti *et al.* 2009), were also included. Haploid strains were obtained by deleting the *HO* gene using the dominant drug-resistance markers *KanMX* or *HygMX* (Goldstein & McCusker 1999). The *HO* deletion was performed by PCR-mediated gene replacement as described in the study by Cubillos *et al.* (2009). Genetic crosses were performed by mating haploid cells of opposite mating types and different resistance markers. Sporulation of hybrid diploids obtained was induced on 1% potassium acetate agar. Asci were partially digested by zymolyase and spore viabilities were tested

© 2012 Blackwell Publishing Ltd

Table 1 Crosses combinations and percentage of spore viability

Lineage/Isolate		CHN I				CHN II	CHN III	CHN IV	CHN V	CHN VIII
		HN6	FJ7	HN1	HN2	SX1	HN10	BJ20	HN15	BJ10
CHN I	HN6		97.3	20.0						
	HN2	14.6	21.2	91.8						
CHN II	SX1	31.4	46.0							
	SX6	53.0	55.0	15.2	14.1	95.6				
CHN III	HN9	29.6	43.1	12.0	9.1	10.2	84.4			
	HN10	37.8	27.5	3.0	11.8	23.9				
CHN IV	BJ6	89.1	76.8	16.7	19.0	48.2	38.9	90.0		
	BJ20	67.9	72.1	24.3	28.1	55.1	23.7			
CHN V	HN14	68.6	82.0	18.2	15.7	55.4	49.0	69.8	94.0	
	HN15	73.2	62.5	15.3	10.9	48.6	35.5	68.7		
CHN VIII	BJ10	46.1	67.3	17.0	18.9	42.4	35.2	83.3	84.0	
Wine/European	DBVPG6765	62.5	77.1	20.4	20.8	36.5	18.3	75.2	69.2	92.0
Malaysian	UWOPS03-461.4	11.7	7.0	1.0	1.9	7.0	2.0	7.6	7.7	5.7

At least 25 tetrads (100 spores) were analysed for each cross.

on YPD agar by dissecting at least 25 tetrads (100 spores).

Results

Ubiquitous distribution of wild S. cerevisiae in nature

A total of 2064 samples were collected and subjected for selective *Saccharomyces sensu stricto* enrichment (Table S2, Supporting information). *S. cerevisiae* was isolated from 226 of the samples, a success rate of 10.9%. *S. cerevisiae* isolates were distinguished from those of other *Saccharomyces sensu stricto* species by ITS sequence analysis. The ITS sequence of the Chinese *S. cerevisiae* isolates exhibited 0–4 substitutions compared with the type strain of the species (NRRL Y-12632) and 7–9 substitutions compared with *S. paradoxus* type strain (NRRL Y-17217). In addition to grapes and oak tree (*Quercus* spp.) bark, *S. cerevisiae* was successfully isolated from a variety of damaged fruit collected in orchards or markets in different provinces of China; from the bark of different deciduous trees; forest soil and rotten wood collected in primeval, original secondary and planted forests located in different regions covering temperate, subtropical and tropical climate zones from northern to southern China (Table S2, Supporting information). Unexpectedly, the success rate of *S. cerevisiae* isolation from fruit samples (6.5%) was lower than that from tree bark (16.5%), soil (10.8%) and rotten wood (9.2%) samples. *S. cerevisiae* was more frequently isolated from forest soil samples (success rate 13.7%) than from orchard soil samples (success rate 9.1%). Among the fruits giving positive *S. cerevisiae* isolation, grape samples showed the lowest success rate (Table S2, Supporting information).

Rarity of heterozygosity in wild S. cerevisiae

A selection of 102 natural Chinese *S. cerevisiae* isolates with different ecological and geographical origins were used for multiple locus sequence analysis. Heterozygosity was observed in only three of the isolates, namely BJ03 from an apple sample; CH03 from a tree bark sample collected in a park and YSHZ1 from a persimmon sample. These isolates exhibited 44 (0.34%), 10 (0.07%) and 49 (0.36%) heterozygous sites, respectively, in the 13 loci sequenced with a total length of 13 540 bp. The other wild isolates showed no signal of heterozygosity in any of the sequenced loci. Considering the scarcity of heterozygous sites in the data set and the topic of this study focusing on *S. cerevisiae* from wild environments, the three heterozygous isolates were not analysed further.

Remarkable sequence diversity in wild S. cerevisiae Chinese isolates

The sequence data from 99 *S. cerevisiae* isolates, representing all sampled regions and substrates within China (Table S3, Supporting information) and from 38 SGRP isolates representing worldwide geographical and ecological origins (Liti *et al.* 2009), were used for molecular phylogenetic and population genetic analyses. In the 13 loci (~13.5 kb) analysed, a total of 827 SNPs, including 214 (25.9%) singletons, were identified from the combined data set of the Chinese and SGRP isolates, exhibiting a nucleotide diversity (π) of 7.27×10^{-3} (Table S4, Supporting information). The Chinese isolates are approximately twice as diverse as the SGRP isolates which represent the maximum genetic variation of *S. cerevisiae* documented so far (Liti *et al.* 2009). A total

© 2012 Blackwell Publishing Ltd

of 718 SNPs, including 178 (24.8%) singleton, were identified from the Chinese isolates ($\pi = 7.71 \times 10^{-3}$), while only 278 SNPs, including 66 (22.7%) singletons, were found from the SGRP isolates ($\pi = 4.47 \times 10^{-3}$) (Table S4, Supporting information). Only 169 (60.8%) of the SNPs previously characterized in the SGRP isolates were shared in the Chinese isolates, which in turn present an additional 549 private SNPs.

Multiple novel and highly diverged S. cerevisiae lineages from China

Sequences of the 13 loci were concatenated to obtain a neighbour-joining phylogenetic tree (Fig. 1). Phylogenetic analyses using different algorithms including maximum likelihood, maximum parsimony and Bayesian resulted in similar tree topologies (data not shown). The majority (71.7%) of the 99 Chinese isolates grouped into eight new distinct lineages with strong (85–100%) bootstrap values (Fig. 1, Table S3, Supporting information). The Chinese lineages were designated as CHN I–VIII, corresponding to the order of their positions from the bottom to the top in the phylogenetic tree rooted using *S. paradoxus* (Fig. S1, Supporting information). The five lineages previously described by Liti *et al.* (2009) based on the genome wide SNP analysis were also clearly resolved, indicating the robustness of the tree topology generated in this study. Lineages CHN I–V were located as basal branches in the *S. cerevisiae* tree while lineages CHN VI–VIII were located at the upper part of the tree together with the five worldwide lineages of *S. cerevisiae* (Fig. S1, Supporting information). Lineages CHN I and II are the most significantly diverged lineages identified so far and located at the deep root of the *S. cerevisiae* tree (100% bootstrap, Figs 1 and S1, Supporting information). The average genetic distances of lineages CHN I or II from the other lineages, as calculated from Kimura's two-parameter model (Kimura 1980), were nearly double that of any other pairwise combinations (Table S5, Supporting information).

The newly identified Chinese lineages are characteristic of specific geographical areas and ecological niches (Figs 1 and S2, Table S3, Supporting information).

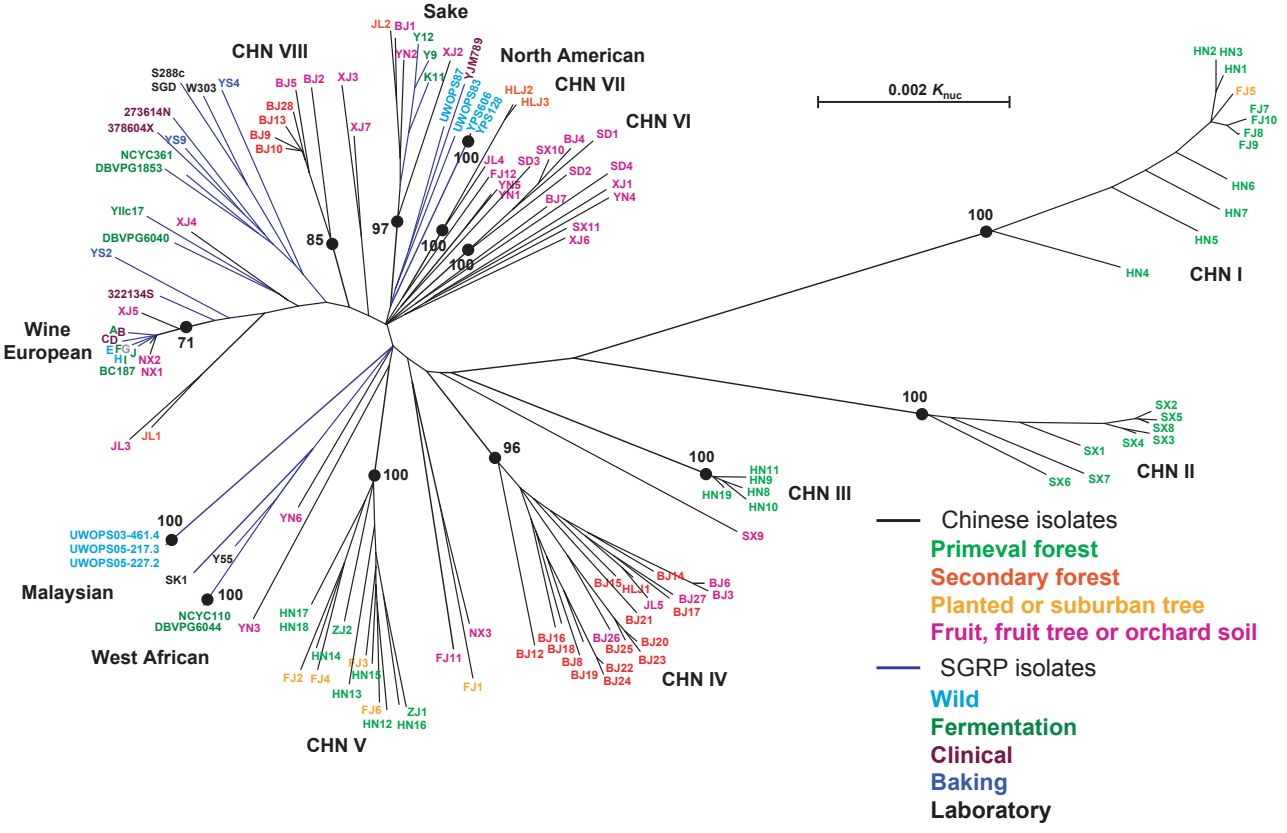


Fig. 1 Neighbour-joining tree inferred from the concatenated sequences of nine genes and four intergenic loci (total length of ~13 500 bp). Designations of Chinese and worldwide lineages are shown. Black and blue lines indicate Chinese and SGRP isolates and strain name colours indicate sources. Bootstrap percentages above 50% from 1000 bootstrap replicates are indicated. The scale bar represents 0.002 substitutions per nucleotide position calculated from Kimura's two-parameter model.

© 2012 Blackwell Publishing Ltd

Lineages CHN I, III and V were isolated from rotten wood and tree bark mostly collected in tropical and subtropical primeval forests in southern China; lineage CHN II from the tree bark exclusively collected in a temperate primeval forest in northern China; lineages CHN IV and VI–VIII mostly from tree bark and various fruit collected in secondary forests, park or orchards in northern China with temperate climate.

Close wild relatives of the domestic Wine/European and Sake lineages were also identified within the Chinese isolates. Two isolates from orchard soil and one from grape were located at basal branches close to the Wine/European lineage with 71% bootstrap support (Fig. 1). These three isolates were from Ninxia and Xinjiang autonomous regions of north-western China (Fig. S2, Supporting information). Three fruit isolates from different locations in China and one oak isolate from Jilin province, north-eastern China, clustered in the basal branches of the Sake lineage with 97% bootstrap support (Figs 1 and S2, Supporting information). Twenty-one of the 99 (21.2%) Chinese wild isolates analysed were placed in weakly supported branches (<50% bootstrap) or in singletons outside the wild and domestic lineages (Fig. 1). These strains appear to have mosaic genome and were mostly isolated from fruit or other substrates collected in environments affected by human activity (Fig. 1, Table S3, Supporting information).

Strong population structure in wild S. cerevisiae

The model-based clustering algorithm (Pritchard *et al.* 2000; Falush *et al.* 2003) analysis revealed that the majority of the identified Chinese lineages represent distinct populations (Fig. 2). Multiple analyses were performed using varying number of assumed populations (K = 2–15). Even at the lowest resolution (K = 2), lineage CHN I was clearly recognized (Fig. 2). With increasing K values, more populations were resolved. At K = 8, lineages CHN I–V which were from primeval and original secondary forests were all resolved as distinct populations. The maximum resolution was achieved at K = 13 (Fig. 2), consistently with the number of lineages that have been recognized in the whole population of *S. cerevisiae*. At the maximum resolution, lineages CHN I–VI and the five worldwide lineages recognized from the SGRP isolates (Liti *et al.* 2009) were private and monomorphic for the majority of the segregating sites. Evidence of shared genetic variants was observed among lineages CHN VI–VIII and the Chinese mosaic group, which were mainly isolated from man-made environments or secondary forest, suggesting recombination from outcrossing. Interestingly, these mosaics strains also shared polymorphisms with the worldwide lineages, whereas contributions from the diverged CHN I–V lineages appear limited (Fig. 2).

The strong population structure of *S. cerevisiae* and the demarcation of the basal Chinese lineages were also recognized by PCA (Patterson *et al.* 2006; Price *et al.* 2006). The three most significantly diverged lineages CHN I–III from primeval forests also formed three discrete clusters distant from each other and from the other lineages (Fig. S3, Supporting information). The other Chinese and international lineages were not clearly separated from each other or from the mosaic isolates (Fig. S3, Supporting information), implying genetic admixture (Patterson *et al.* 2006). These lineages except CHN V were mainly from human-associated environments (Table S3, Supporting information). The presence

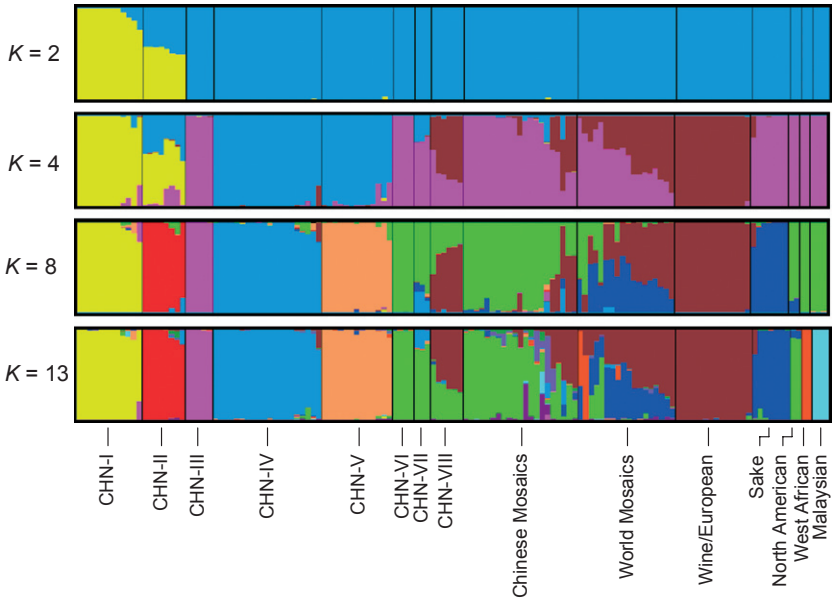


Fig. 2 Population structure of *S. cerevisiae* inferred from 827 SNPs of 138 isolates using the program Structure (version 2.3.1) using different K values (the numbers of populations assumed). An admixture model with correlated allele frequencies was used. Both the length of burnin and the MCMC repetitions after burnin were 30 000. Each strain is represented by a single vertical line broken into K coloured segments, with lengths proportional to each of the K inferred clusters. Isolates in each identified population are consistent with the lineages identified in the phylogenetic tree reported in Fig. 1.

© 2012 Blackwell Publishing Ltd

of the CHN V lineage, which was mainly isolated from primeval forests, probably suggests a gene flow between strains isolated in primeval forests and strains present in human-associated environments. The presence of CHN V isolates in planted trees (Figs 1 and S2, Table S3, Supporting information) might explain this phenomenon.

Different degree of recombination in wild populations of S. cerevisiae

The scarcity of heterozygous sites observed in the wild *S. cerevisiae* isolates suggests that the wild Chinese lineages constitute inbreeding populations (Charlesworth 2003). However, population structure analyses presented previously clearly suggest recombination exits at least in lineages characteristic of human-associated environments. We further examined recombination by linkage disequilibrium (LD) and ILD/PHT analyses. Lineages CHN III, VI and VII, including only 3–5 isolates, were not included in these analyses owing to the insufficient sample size required to detect statistically significant parameters (Xu 2005). The observed I_A and rBarD values of most of the lineages or groups analysed differed significantly from those calculated from randomized data sets (Table S4, Supporting information), rejecting the null hypothesis of recombination. However, the four-gamete test (Hudson & Kaplan 1985) detected evidence of recombination from the groups showing strong allelic associations, being consistent with the observation of Fay & Benavides (2005) that the lack of recombination can not explain the high levels of LD observed in *S. cerevisiae* populations.

Individual gene genealogies clustered the isolates belonging to the basal Chinese lineages CHN I–III and Wine/European consistently in almost all the 13 loci analysed (Fig. S4, Supporting information). Evident genealogical incongruences were observed for CHN IV–VII and the Sake lineages in five to seven of the 13 trees. Although three individual trees (*CYT1*, *MLS1* and IntMD) showed incongruence for lineage CHN VIII, the ILD/PHT analysis used to assess the statistical significance of gene genealogy congruence (Farris *et al.* 1995) did not reject the null hypothesis of clonality for this lineage ($P = 0.4014$). As expected, the ILD/PHT analysis rejected the clonality for lineages CHN IV and V ($P = 0.0001$) but did not for other lineages tested including CHN I, II and the Wine/European lineages (Table S4, Supporting information).

Karyotype differentiation between wild S. cerevisiae lineages

The karyotype of 75 isolates was analysed by pulse field gel electrophoresis. All the isolates showed chromosomal banding patterns typical for *S. cerevisiae* (Fig. 3). Generally, 13–14 chromosomal bands ranging in sizes from 220 to 2200 kb, including 2–3 small (<500 kb), 6–7 medium (500–1000 kb) and 3–4 large (>1000 kb) bands were resolved. As chromosome bands with stronger relative intensity correspond to doublets (chromosomes that co-migrate), 16 chromosomes could be recognized from each isolates by using strain S288c as a reference. Although chromosomal length polymorphism is common in *S. cerevisiae* and electrophoretic karyotyping has been shown to be more discriminative than many other approaches for strain typing (Schuller *et al.* 2004), each of the Chinese lineages, especially the basal lineages, exhibited a lineage-specific chromosomal banding pattern. Lineages CHN I–III differed from each other and from the other lineages and each has a characteristic karyotype. Notably, the molecular size of the largest chromosome of lineage CHN III (~2000 kb) was smaller compared with other lineages (usually 2200 kb). Isolates within each CHN IV–VIII lineages also generally showed homogenous banding patterns, which were distinguishable from each other. The wild isolates in the Sake lineage showed remarkably variable banding patterns for several chromosomes (Fig. 3), although the sequence variation within this lineage was not greater than that within some of other lineages (e.g. CHN IV–VI; Fig. 1, Table S4, Supporting information). This result is consistent with common industrial Sake yeast strains, which also exhibited frequent chromosomal size changes caused by structural variation (Ogihara *et al.* 2008). Finally, isolates from the Chinese mosaic group showed variable karyotypes (Fig. 3).

Partial reproductive isolation between wild S. cerevisiae lineages

Representative isolates from CHN I–V, CHN VIII, Wine/European and Malaysian lineages were used in genetic crosses to measure gametes viability (Table 1). A general correlation between cross fertility and sequence divergence was observed among the Chinese lineages tested. Crosses within the same lineages exhibited high spore viabilities (84.4–97.3%) with the exception of HN1 and HN2 from CHN I (discussed below). In contrast, crosses between different Chinese lineages showed a wide range of spore viabilities (10.2–89.1%). The crosses between the three most significantly diverged lineages CHN I–III showed 10.2–55.0% spore viabilities, indicating partial sexual reproductive isolation between them. Crosses of the Wine/European isolate DBVPG6765 with basal lineages CHN I–V exhibited 18.3–77.1% spore viabilities, while with lineage CHN VIII, which is most closely related to the Wine/European lineage (Figs 1 and S1, Supporting informa-

© 2012 Blackwell Publishing Ltd

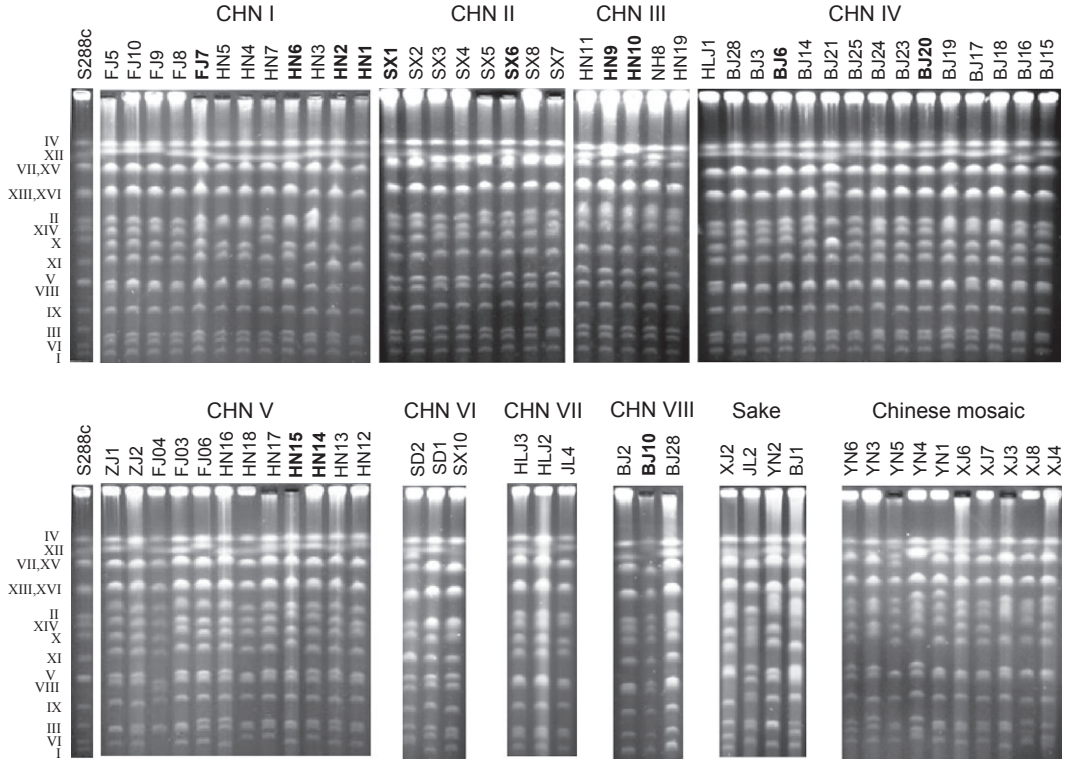


Fig. 3 Electrophoretic karyotypes of wild *S. cerevisiae* belonging to different Chinese lineages. Strain S288c is used as a reference; isolates with bold numbers were used for genetic crosses and gamete viability analysis.

tion), showed 92.0% spore viabilities (Table 1). As crosses between recognized biological species of *Saccharomyces sensu stricto* usually result in spore viabilities of less than 1% (Liti *et al.* 2006), the relatively higher and overlapping spore viabilities of crosses between the different lineages (Table 1) suggest that all the wild Chinese lineages belong to a single species *S. cerevisiae*.

The low fertility (1–11.7% spore viability) of crosses of the Chinese isolates with the Malaysian isolate UWOPS03-461.4 is consistent with previous reports, which showed 2.5–10% spore viability of crosses of the Malaysian isolate with other *S. cerevisiae* isolates (Naumov *et al.* 2006; Cubillos *et al.* 2011). Reduced spore viability was also observed in crosses of isolates with limited sequence divergence (Table 1). While the cross between isolates HN1 and HN2 of lineage CHN I exhibited 91.8% spore viability, the crosses of these two isolates with isolates FJ7 and HN6 from the same lineage showed only 14.6–21.2% spore viabilities. A similar low fertility (3.0–28.1% spore viability) was observed in the crosses of isolates HN1 and HN2 with isolates from other lineages CHN II–IV, VIII and Wine/European (Table 1). One explanation for the reduced fertility of crosses between isolates of the same *Saccharomyces* species is the presence of large-scale chromosomal rearrangements (Liti *et al.* 2006). The reduced fertility of interlineage crosses between CHN I–III corresponds to their karyotype differences as showed in Fig. 3. Interestingly, karyotype difference can also be found between the pairs of isolates in the intralineage crosses that show low spore viability. The CHEF karyotypes of isolates HN1 and HN2 differ from those of isolates FJ7 and HN6 clearly in the positions of chromosome V and XI (Figs 3 and S5, Supporting information). The contribution of chromosomal rearrangements to the reduced fertility is worth further investigation.

Discussion

Despite over 70 years of research, the ecology of the simple genetic system, *S. cerevisiae*, remains elusive (Replansky *et al.* 2008). Although recent efforts in characterizing natural variation in *S. cerevisiae* have shed new light in the population structure and human influence, the large majority of the strains available so far were isolated from man-driven environments (Fay & Benavides 2005; Liti *et al.* 2009; Schacherer *et al.* 2009). Here, we reported the result of a field survey at an unprecedented scale that allowed the gathering of a large set of wild *S. cerevisiae* isolates. Population genetics analysis revealed surprising genetic diversity and a strong population structure for the species in wild envi-

© 2012 Blackwell Publishing Ltd

ronments from multiple Chinese areas. We addressed several key issues in the ecology, population structure and biogeography of this species in nature. Furthermore, we present genetic evidence for the possible Asian origin of modern *S. cerevisiae*.

Ecological distribution of S. cerevisiae

A growing number of studies suggested that wild *S. cerevisiae* might occur in tree bark as well as vineyards (Yoneyama 1957; Banno & Mikata 1981; Naumov & Naumova 1991; Török *et al.* 1996; Naumov *et al.* 1998; Mortimer & Polsinelli 1999; Mortimer 2000; Redzepovic *et al.* 2002; Sniegowski *et al.* 2002). However, previous surveys on the distribution of wild *Saccharomyces* species in nature were usually limited to specific substrates (e.g. grapes from vineyards or the bark of oak trees) or to limited geographical ranges within the same climate zones. Extensive systematic surveys with the aim of understanding the ecological distribution of *S. cerevisiae* in nature have rarely been performed or were unsuccessful (Replansky *et al.* 2008). Our study substantially expands the ecological knowledge of *S. cerevisiae*, clearly showing that this species is ubiquitous in the wild and frequently present in environments that are not or only marginally affected by human activity (e.g. primeval forests). Furthermore, *S. cerevisiae* was isolated from many different substrates likely to offer suitable nutrients for the growth of this eukaryotic microbe.

Population structure of wild S. cerevisiae isolates

Phylogenetic analysis of *S. cerevisiae* strains isolated from novel environments and geographical areas greatly increased the genetic variation characterized in this species. We clearly show that *S. cerevisiae* populations have differentiated more in nature than previously recognized (Sniegowski *et al.* 2002; Fay & Benavides 2005; Aa *et al.* 2006; Liti *et al.* 2006, 2009; Schacherer *et al.* 2009). A strong population structure of *S. cerevisiae* in the wild was identified here. Almost all Chinese isolates from primeval and original secondary forests occur in clearly differentiated lineages. The isolates that clustered outside distinct lineages are mostly from human-associated environments. The lineages from human-associated environments generally exhibit high level of recombination or genetic admixture, as revealed by gene genealogy concordance, ILD/PHT and PCA analyses. This observation is consistent with recent studies showing that human activity creates opportunities for outcrossing (Liti *et al.* 2009; Goddard *et al.* 2010; Magwene *et al.* 2011). In contrast, wild lineages CHN I–III from primeval forests present rare interlineage recombination and thus representing distinct inbreeding populations in nature. The genetic isolation of these lineages is supported by the absence of heterozygosity, the nature of the segregating sites (private and monomorphic, Figs 2 and S3, Supporting information), consistency of individual gene genealogy (Fig. S4, Table S4, Supporting information) and lineage-specific karyotypes (Fig. 3). Our results indicate that, in remote forest environments with limited human activity, *S. cerevisiae* occurs ubiquitously and in highly diverged populations. This observation dispels the concern of the suitability of *S. cerevisiae* as a model in ecology, evolutionary biology and biogeography because of its close relationship with human activity (Greig 2007).

The lack of heterozygosity was also observed in wild Chinese lineages of *S. cerevisiae* (CHN IV–VI) showing strong evidence of recombination. This phenomenon is consistent with previous studies (Aa *et al.* 2006; Bradbury *et al.* 2006; Diezmann & Dietrich 2009; Mortimer *et al.* 1994). For examples, 25 of the 27 *S. cerevisiae* isolates, with different origins compared by Aa *et al.* (2006), were homozygous for all four loci sequenced; and none of the 24 soil isolates of *S. cerevisiae* examined by Diezmann and Dietrich (2009) showed any evidence of heterozygosity in five loci analysed. A possible explanation to this phenomenon is that homothallic selfing might occur frequently in nature through a process known as 'genome renewal' (Mortimer *et al.* 1994).

Despite the greatly expanding variation described in *S. cerevisiae*, as a whole, the variation of *S. cerevisiae* is considerably less than in *S. paradoxus* (Fig. S1, Supporting information). However, variation within lineages of wild *S. cerevisiae* is usually greater than that within lineages of *S. paradoxus* (Fig. S1, Supporting information). Our result suggests that more diverged populations of the species possibly exist in primeval forests in other geographical areas. A worldwide survey of *S. cerevisiae* populations in primeval forests unaffected by human activity will be necessary for a thorough understanding of the global ecological distribution, genetic diversity and population structure of *S. cerevisiae* in nature.

Factors shaping S. cerevisiae population structure

There is an ongoing debate as to the role of ecology vs. geography in shaping the population structure of *S. cerevisiae*. Some studies have shown that ecological factors may play a more important role than geographical factors (Versavaud *et al.* 1995; Fay & Benavides 2005; Aa *et al.* 2006; Ezov *et al.* 2006; Schacherer *et al.* 2009), while others suggested the opposite (Ezeronye & Legras 2009; Liti *et al.* 2009; Goddard *et al.* 2010). We show that among the significantly diverged populations of wild *S. cerevisiae* from primeval and original secondary forest

© 2012 Blackwell Publishing Ltd

environments, isolates from tropical and subtropical regions (southern China) are not mixed with those from temperate regions (northern China) (Table S3, Fig. S2, Supporting information). The oak isolates from different regions in northern China (Shaanxi province and Beijing) were separated into different lineages (CHN II and IV, respectively) (Table S3, Figs 1 and S2, Supporting information). The Chinese oak isolates were also clearly separated from those from North America (Fig. 1). These observations suggest the role of geographical isolation for the differentiation of *S. cerevisiae* population in nature.

On the other hand, primeval forest isolates from the same region also grouped in significantly differentiated lineages, exhibiting a phenomenon of sympatric differentiation (Table S3, Fig. S2, Supporting information). Most strikingly, isolates from the primeval rainforests in Hainan, a tropical island in southern China, exhibited a magnitude of population variation equivalent to the whole globe population (Table S4, Fig. S1, Supporting information). Hainan Island has an area comparable with the size of Belgium (33 920 km^2). The Hainan isolates were isolated from rotten wood and bark of *Fagaceae* trees collected in the rainforests of four mountains locate in the southern part of the island. The mountains are covered with similar flora and separated from each other by ~30–100 km. The isolates from the four mountains distributed in three significantly diverged basal lineages CHN I, III and V (Figs 1 and S2, Supporting information). Isolates from the same habitats and mountain may also belong to different lineages; for example, isolates from rotten wood collected in Wuzhi Mountain, clustered within lineages CHN III and V (Table S3, Supporting information). Therefore, ecological factors, which are usually considered as being responsible for sympatric differentiation, seem unable to explain the coexistence of different lineages in this island.

The mechanism of the seemingly sympatric differentiation of *S. cerevisiae* in nature is worth further study. Among the diverged lineages CHN I, III and V from Hainan Island, we observed partial reproductive isolation between CHN I and III (29.6–43.1% spore viability) and between CHN III and V (35.5–59.0% spore viability); but the reproductive isolation between CHN I and V seems weak (62.5–82.0% spore viability) (Table 1). Interestingly, karyotype analysis showed that lineage CHN III differs from lineages CHN I and V more remarkably than the latter two lineages differ from each other (Fig. S5, Supporting information). The crosses of two Hainan isolates HN1 and HN2 in lineage CHN I with other Hainan isolates from the same or different lineages showed only 3.0–21.2% spore viabilities (Table 1). Consistently, karyotype differences of HN1 and HN2 from other Hainan isolates, even those in the same lineage, were observed (Figs 3 and S5, Supporting information). These observations suggest that genetic isolation may arise between wild *S. cerevisiae* isolates from the same region with limited sequence divergence. Further studies are needed to investigate whether the partial reproductive isolation is responsible for the differentiation of these Hainan lineages and the impact of chromosomal rearrangements in the reduced fertility.

We observed *S. cerevisiae* population migration between different locations and from primeval forests to human-associated environments. In addition to Hainan Island, isolates belonging to lineages CHN I and V were found in subtropical primeval forests in Fujian and Zhejiang provinces, southern China (Figs 1 and S1, Table S3, Supporting information). Isolates of these two lineages were also found from the bark of planted chestnut trees (*Castanea mollissima*) in Wuyi Mountain, Fujian province (Figs 1 and S2, Table S3, Supporting information), indicating a possible role for human activity in the population migration of *S. cerevisiae* in the wild.

Evidence for a Far Eastern Asia origin of S. cerevisiae

Far Eastern Asia harbours higher species diversity of the *Saccharomyces sensu stricto* complex than other regions of the world. All six biological species of this complex have been isolated from this area and two of them (*S. arboricolous* and *S. mikatae*) have only been found in Asia (Naumov *et al.* 2000, 2003; Wang & Bai 2008). In addition, genetic variation between two Japanese lineages of *S. kudriavzevii* significantly exceeds the divergence between one of the Japanese lineage and the European population (Hittinger *et al.* 2010), suggesting high intraspecific diversity of *Saccharomyces* species within the Far East. These observations are consistent with the hypothesis that Far Eastern Asia is the reservoir of *Saccharomyces* natural variation (Naumov & Nikonenko 1988; Naumov *et al.* 2003), which was proposed based on Vavilov's theory of centers of diversity and origin (Vavilov 1926). Here, we provide strong evidence supporting this hypothesis.

We show that Chinese isolates contribute the majority of the genetic variation of *S.cerevisiae* documented so far. The Chinese isolates exhibit much higher sequence diversity than worldwide isolates (Fig. 1) selected from population genetic studies (Sniegowski *et al.* 2002; Fay & Benavides 2005; Aa *et al.* 2006; Liti *et al.* 2006; Naumov *et al.* 2006). In addition, the oldest lineage (CHN I) and the other basal lineages (CHN II–V) of *S. cerevisiae* were only found in China (Fig. 1). Our results support the hypothesis that *S. cerevisiae* originated from Far Eastern Asia (Naumov & Nikonenko 1988). Specifically, the *S. cerevisiae* isolates from primeval forests in Hainan

© 2012 Blackwell Publishing Ltd

Island represent the oldest lineage and exhibit a degree of genetic variation ($\pi = 8.14 \times 10^{-3}$) higher than those of the whole Chinese ($\pi = 7.71 \times 10^{-3}$) and world populations ($\pi = 7.27 \times 10^{-3}$) of the species. This implies that wild *S. cerevisiae* originated from tropical rainforests within China or Far Eastern Asia. Nevertheless, further investigation on the distribution and genetic diversity of *S. cerevisiae* in primeval forests worldwide will be necessary for a better understanding of the origin and evolution of modern *S. cerevisiae*.

The possible origin of S. cerevisiae domestic populations

Separate domestication events in *S. cerevisiae* have been proposed; however, the origin of domestic *S. cerevisiae* lineages is still uncertain (Mortimer & Polsinelli 1999; Fay & Benavides 2005; Naumov *et al.* 2006; Legras *et al.* 2007; Ezeronye & Legras 2009; Liti *et al.* 2009; Schacherer *et al.* 2009; Goddard *et al.* 2010). We clearly showed that the oldest and basal lineages of *S. cerevisiae* were from primeval forest environments, while the wild isolates from man-made environments were usually located together or near the domesticated isolates (Figs 1 and S1, Supporting information). The phylogentic tree, rooted using *S. paradoxus*, shows that lineages CHN I and CHN II diverged from each other and from the other lineages much earlier than the lineages isolated from human-associated environments diverged from each other (Fig. S1). Fay & Benavides (2005) estimated that the divergence time between the wine and sake lineages is *c.* 11 900 years ago. Applying the same method, the divergence time between lineages CHN I and CHN II will be at least double this figure. Our results indicate that diversified wild populations of *S. cerevisiae* exist in nature all along independently of and predate domesticated strains and that primeval forest environments may be the sources of *S. cerevisiae* strains associated with human activity.

Close wild relatives of the wine and sake lineages were identified in this study (Table S3, Figs 1 and S2, Supporting information). The three isolates (NX1, NX2 and XJ5) located within the Wine/European cluster were isolated from grape and orchard soil collected in north-western China (Fig. S2, Supporting information). These three Chinese wild isolates clearly belong to the Wine/European lineage as revealed by Structure and individual gene genealogy analyses (Figs 2 and S4, Supporting information). This result raises the possibility of an Asian origin of the Wine/European strains. Archaeological and historical events support this hypothesis as the earliest evidence for winemaking in the world, *c.* 9000 years ago, has been found in China (McGovern *et al.* 2004). Among the four wild isolates clustered in the Sake lineage, three were from northern and north-eastern China (Table S3, Fig. S2, Supporting information), which are geographically near Korea and Japan, suggesting the wild origin of the Sake population.

In conclusion, our work substantially expands the ecological and evolutionary knowledge of *S. cerevisiae* and probably unveils the hidden part of the iceberg of its natural variation. We present a key resource of isolates that can be exploited in quantitative trait loci (QTLs) mapping (Cubillos *et al.* 2011). Further studies of this set of isolates on population genomics, quantitative genetics and phenomics will certainly be helpful for a better understanding the life history, biogeography and speciation of *S. cerevisiae* in nature.

Acknowledgements

We thank Dr. Conrad Nieduszynski, University of Nottingham, for his suggestions. This study was supported by grants no. 30825002 and no. 91131004 from the National Natural Science Foundation of China (NSFC) and by no. KSCX2-EW-J-6 from the Knowledge Innovation Program of the Chinese Academy of Sciences. G.L. was supported by CNRS, ATIP-AVENIR and ARC.

References

Aa E, Townsend JP, Adams RI, Nielsen KM, Taylor JW (2006) Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, **6**, 702–715.

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.

Ayoub MJ, Legras JL, Saliba R, Gaillardin C (2006) Application of multi locus sequence typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon. *Journal of Applied Microbiology*, **100**, 699–711.

Bai FY, Liang HY, Jia JH (2000) Taxonomic relationships among the taxa in the *Candida guilliermondii* complex, as revealed by comparative electrophoretic karyotyping. *International Journal Systematic and Evolutionary Microbiology*, **50**, 417–422.

Banno I, Mikata K (1981) Ascomycetous yeasts isolated from forest materials in Japan. *IFO Research Communication*, **10**, 10–19.

Bradbury JE, Richards KD, Niederer HA, Lee SA, Rod Dunbar P, Gardner RC (2006) A homozygous diploid subset of commercial wine yeast strains. *Antonie van Leeuwenhoek*, **89**, 27–37.

Charlesworth D (2003) Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **358**, 1051–1070.

Ciani M, Mannazzu I, Marinangeli P, Clementi F, Martini A (2004) Contribution of winery-resident *Saccharomyces cerevisiae* strains to spontaneous grape must fermentation. *Antonie van Leeuwenhoek*, **85**, 159–164.

Cubillos FA, Louis EJ, Liti G (2009) Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains. *FEMS Yeast Research*, **9**, 1217–1225.

Cubillos FA, Billi E, Zörgö E *et al.* (2011) Assessing the complex architecture of polygenic traits in diverged yeast populations. *Molecular Ecology*, **20**, 1401–1413.

© 2012 Blackwell Publishing Ltd

Dickinson JR (2000) Yeasts: providing questions and answers for modern biology. *Science Progress*, **83**, 173–192.

Diezmann S, Dietrich FS (2009) Saccharomyces cerevisiae: population divergence and resistance to oxidative stress in clinical, domesticated and wild isolates. *PLoS One*, **4**, e5317.

Ezeronye OU, Legras JL (2009) Genetic analysis of *Saccharomyces cerevisiae* isolates isolated from palm wine in eastern Nigeria. Comparison with other African isolates. *Journal of Applied Microbiology*, **106**, 1569–1578.

Ezov TK, Boger-Nadjar E, Frenkel Z *et al.* (2006) Molecular-genetic biodiversity in a natural population of the yeast *Saccharomyces cerevisiae* from 'Evolution canyon': microsatellite polymorphism, ploidy and controversial sexual status. *Genetics*, **174**, 1455–1468.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Farris JS, Källersjö M, Kluge AG, Bult C (1995) Testing significance of incongruence. *Cladistics*, **10**, 315–319.

Fay JC, Benavides JA (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genetics*, **1**, e5.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Goddard MR, Anfang N, Tang R, Gardner RC, Jun C (2010) A distinct population of *Saccharomyces cerevisiae* in New Zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Environmental Microbiology*, **12**, 63–73.

Goffeau A, Barrell BG, Bussey H *et al.* (1996) Life with 6000 genes. *Science*, **274**, 546–567.

Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast*, **15**, 1541–1553.

Greig D (2007) Population biology: wild origins of a model yeast. *Current Biology*, **17**, R251–R253.

Hittinger CT, Gonçalves P, Sampaio JP, Dover J, Johnston M, Rokas A (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature*, **464**, 54–58.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Johnson LJ, Koufopanou V, Goddard MR, Hetherington R, Schäfer SM, Burt A (2004) Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics*, **166**, 43–52.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Koufopanou V, Hughes J, Bell G, Burt A (2006) The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 1941–1946.

Kuehne HA, Murphy HA, Francis CA, Sniegowski PD (2007) Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations. *Current Biology*, **17**, 407–411.

Kumar A, Snyder M (2001) Emerging technologies in yeast genomics. *Nature Reviews Genetics*, **2**, 302–312.

Landry CR, Townsend JP, Hartl DL, Cavalieri D (2006) Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Molecular Ecology*, **15**, 575–591.

Legras JL, Merdinoglu D, Cornuet JM, Karst F (2007) Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology*, **16**, 2091–2102.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Liti G, Louis EJ (2005) Yeast evolution and comparative genomics. *Annual Review of Microbiology*, **59**, 135–153.

Liti G, Barton DB, Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics*, **174**, 839–850.

Liti G, Carter DM, Moses AM *et al.* (2009) Population genomics of domestic and wild yeasts. *Nature*, **458**, 337–341.

MacFaddin JD (1985) Media for Isolation Cultivation-Identification-Maintenance of Medical Bacteria, Vol. 1. Willians & Wilkins, Baltimore, Maryland.

Magwene PM, Kayıkçı Ö, Granek JA, Reininga JM, Scholl Z, Murray D (2011) Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1987–1992.

Martini A (1993) Origin and domestication of the wine yeast *Saccharomyces cerevisiae*. *Journal of Wine Research*, **4**, 165–176.

McGovern PE, Zhang J, Tang J *et al.* (2004) Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17593–17598.

Mortimer RK (2000) Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Research*, **10**, 403–409.

Mortimer R, Polsinelli M (1999) On the origins of wine yeast. *Research in Microbiology*, **150**, 199–204.

Mortimer RK, Romano P, Suzzi G, Polsinelli M (1994) Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts. *Yeast*, **10**, 1543–1552.

Naumov GI (1996) Genetic identification of biological species in the *Saccharomyces* sensu stricto complex. *Journal of Industrial Microbiology*, **17**, 295–302.

Naumov GI, Naumova ES (1991) A wild yeast population of *Saccharomyces cerevisiae* found in Siberia. *Mikrobiologiia*, **60**, 537–540.

Naumov GI, Nikonenko TA (1988) The East Asia is a probable land of the cultured yeasts *Saccharomyces cerevisiae* (in Russian). *Izvestiya Sibirskogo Otdeleniya Akademii Nauk SSSR, Seriya Biologicheskikh Nauk*, **20**, 97–101.

Naumov GI, Naumova ES, Sniegowski P (1998) *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. *Canadian Journal of Microbiology*, **44**, 1045–1050.

Naumov GI, James SA, Naumova ES, Louis EJ, Roberts IN (2000) Three new species in the *Saccharomyces* sensu stricto complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *International Journal of Systematic and Evolutionary Microbiology*, **50**, 1931–1942.

Naumov GI, Gazdiev DO, Naumova ES (2003) The finding of the yeast species *Saccharomyces bayanus* in Far East Asia. *Microbiology*, **72**, 738–743.

Naumov GI, Serpova EV, Naumova ES (2006) A genetically isolated population of *Saccharomyces cerevisiae* in Malaysia. *Mikrobiologiia*, **75**, 245–249.

© 2012 Blackwell Publishing Ltd

Ogihara F, Kitagaki H, Wang Q, Shimoi H (2008) Common industrial sake yeast strains have three copies of the AQY1-ARR3 region of chromosome XVI in their genomes. *Yeast*, **25**, 419–432.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.

Peña-Castillo L, Hughes TR (2007) Why are there still over 1000 uncharacterized yeast genes? *Genetics*, **176**, 7–14.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904–909.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Redzepovic S, Orlic S, Sikora S, Majdak A, Pretorius IS (2002) Identification and characterization of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* strains isolated from Croatian vineyards. *Letters in Applied Microbiology*, **35**, 305–310.

Replansky T, Koufopanou V, Greig D, Bell G (2008) *Saccharomyces* sensu stricto as a model system for evolution and ecology. *Trends in Ecology and Evolution*, **23**, 494–501.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L (2006) Population genomic analysis of outcrossing and recombination in yeast. *Nature Genetics*, **38**, 1077–1081.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Sampaio JP, Gonçalves P (2008) Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Applied and Environmental Microbiology*, **74**, 2144–2152.

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature*, **458**, 342–345.

Schuller D, Valero E, Dequin S, Casal M (2004) Survey of molecular methods for the typing of wine yeast strains. *FEMS Microbiology Letters*, **231**, 19–26.

Smith JM, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 4384–4388.

Sniegowski PD, Dombrowski PG, Fingerman E (2002) *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research*, **1**, 299–306.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA 5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Török T, Mortimer RK, Romano P, Suzzi G, Polsinelli M (1996) Quest for wine yeasts – an old story revisited. *Journal of Industrial Microbiology*, **17**, 303–313.

Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 4957–4962.

Vaughan-Martini A, Martini A (1995) Facts, myths and legends of the prime industrial microorganism. *Journal of Industrial Microbiol.*, **14**, 514–522.

Vavilov NI (1926) Centers of Origin of Cultivated Plants. Origin and Geography of Cultivated Plants (English edition); trans Love D (1992). Cambridge University Press, Cambridge, pp. 22–135.

Versavaud A, Courcoux P, Roulland C, Dulau L, Hallet JN (1995) Genetic diversity and geographical distribution of wild *Saccharomyces cerevisiae* strains from the wine-producing area of Charentes, France. *Applied and Environmental Microbiology*, **61**, 3521–3529.

Wang SA, Bai FY (2008) *Saccharomyces arboricolus* sp. nov., a yeast species from tree bark. *International Journal of Systematic and Evolutionary Microbiology*, **58**, 510–514.

Wang QM, Li J, Wang SA, Bai FY (2008) Rapid differentiation of phenotypically similar yeast species by single-strand conformation polymorphism analysis of ribosomal DNA. *Applied and Environmental Microbiology*, **74**, 2604–2611.

Xu J (2005) Fundamentals of fungal molecular population genetic analyses. In: Evolutionary Genetics of Fungi (ed. Xu J), pp. 87–116. Horizon Bioscience, , Norfolk, UK.

Yoneyama M (1957) Studies on natural habitats of yeasts – bark inhabiting yeasts. *Journal of Science, Hiroshima University, Series B, Division*, **2**(Botany), 19–38.

Zeyl C (2000) Budding yeast as a model organism for population genetics. *Yeast*, **16**, 773–784.

F.Y.B. and Q.M.W. conceived and designed the project. Q.M.W. and S.A.W. performed sampling and yeast isolation. Q.M.W. and W.Q.L. performed sequencing of Chinese isolates. W.Q.L. performed genetic cross test. G.L. provided sequence data of SGRP isolates. F.Y.B., G.L. and Q.M.W. analysed the data and wrote the paper.

Data accessibility

DNA sequences: GenBank accession nos JQ284454–JQ285737; phylogenetic data: TreeBASE Study accession no. S12670.

Supporting information

Additional Supporting Information may be found in the online version of this article.

© 2012 Blackwell Publishing Ltd

Table S1 Genes or loci sequenced and primers for PCR and sequencing.

Table S2 Isolation of *Saccharomyces cerevisiae* strains from different substrates.

Table S3 Wild Chinese isolates of *S. cerevisiae* employed in multigene sequence analysis.

Table S4 Population genetics analyses for different groups or lineages.

Table S5 Average genetic distances between lineages of *S. cerevisiae* as calculated from Kimura's 2-parameter model.

Fig. S1 Phylogenetic tree reconstructed from neighbour-joining analysis of the combined sequences of nine genes and four intergenicloci with a total length of ~13 500 bp, depicting the relationships of lineages recognized in *S. cerevisiae* and *S. paradoxus*.

Fig. S2 Geographic origins of *S. cerevisiae* lineagesfrom nature in China.

Fig. S3 Principal component analysis (PCA) of the allelic profiles of 13 loci from different lineages or groups of *S. cerevisiae* employed.

Fig. S4 Phylogenetic trees drawn from neighbour-joining analysis of the combined sequences of 13 loci and the sequences of single genes or loci, respectively, showing concordance between gene genealogies.

Fig. S5 Comparison of CHEF electrophoretickaryotypesof wild *S. cerevisiae*isolates from lineages CHN I, III and V.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2012 Blackwell Publishing Ltd

Supporting information

Table S1. Genes or loci sequenced and primers for PCR and sequencing

Locus (Chromosome)	Gene product	Primers (5'-3')	Size of amplicon	GenBank accession nos.
CCA1 (V)	tRNA CCA-pyrophosphorylase	CCA1F: CCACGATTTCTGAAGGAG CCA1R: TTTACGACG GAG CTATCA CCA1-587R2: TTAGGGCTGAGTAACGACAA CCA1-952R: CTTGTCACGCACCCATCC	1635	JQ284454 - JQ284552
CYT1 (XV)	Cytochrome c1	CYT1F: ATGTTCTGTGAGCCAATCCC CTY1R: TTTCTGGTTTTGATACCGGC CYT1-425R: GAAAGATTGTGAAACCGATG CYT1-790R: CTGTTGCGGTAGAGTAGAA	1498	JQ284553 - JQ284651
DSN1 (IX)	Dosage suppressor of NNF1	DSN1F: GATCATATTGCGATAGATTGGG DSN1R: GAGTGGTCATCGTCCGTTTCG	596	JQ284652 - JQ284750
LAS1 (XI)	Essential nuclear protein	LAS1F: GTTGGGCCGCCAATGAAGC LAS1R: AAGACGCCAAATCGCCTAG	948	JQ285144 - JQ285242
MET4 (XIV)	Leucine-zipper transcriptional activator	F: CGAGGATAAGCCGAGCAA R: GCGCATCCACTCCATTGT	395	JQ285243 - JQ285341
MLS1 (XIV)	MaLate Synthase	MLS1F: TTCTCCAATGCGTGTCGTAG MLS1R: ATCCATGAATGGGGAAGTCA MLS1-533F: TCAGCCTTACGACGATTAT MLS1-1021F: GACTCGGGCTCCTATC	1785	JQ285342 - JQ285440
NUP116 (XIII)	Nuclear pore complex protein	NUP116F: AAGCAACTGTCACCAACACG NUP116R: CTTCCCCATCGTTCTTTGAG	501	JQ285441 - JQ285539
PDR10 (XV)	ATP-binding cassette transporter	PDR10F: ATTCCCACCGTGTCATCAAT PDR10R: GTCCCTATCGACACCCTTGA PDR10-512F: GACTGATAAAGGGATACGCTAA PDR10-1202R: TTTCGTACCCATCTGACACTC	1734	JQ285540 - JQ285638
ZDS2 (XIII)	Protein regulating Swe1p-dependent polarized growth	ZDS2F: CCCTGGTCCTCCCACAATCTA ZDS2R: ACAGAGCCCGTGCGTTTT ZDS2-473R: GTATCTGCCCTGTTGC ZDS2-813R: TGCTACCCGCACAACG ZDS2-1352R: GTCTTCGCTCCCTTGTT	1780	JQ285639 - JQ285737
IntAY* (XIV)		IntAYF: TCGCAGAATTAGGGAGAAGT IntAYR: ACTTCCCGACAGCAGATTC	600	JQ284751 - JQ284849
IntCG (XI)		IntCG F: TTTTGTACCGTCGGGACG IntCG R: GGTAAAAGCCTAACAGCAGATG	821	JQ284850 - JQ284948
IntFR (XIII)		IntFR F: CCCATTTCTGGATTTAGGTC IntFR R: GGTAACGTGTAATGCGGATG	631	JQ284949 - JQ285046
IntMD (IX)		IntMD F: TGAGTTCAGCAAGAGTGAAAC IntMD R: TGGGTTCCAGACTCATCC	620	JQ285047 - JQ285143

* IntAY, IntCG, IntFR and IntMD indicate the intergenic loci between ORFs APP1 and YPT53, CCP1 and GPT2, FAR8 and RSF1, and MSL1 and DSN1, respectively.

Table S2. Isolation of *Saccharomyces cerevisiae* strains from different substrates

Substrate	No. of samples collected	No. (%) of samples showing positive *S. cerevisiae* isolation
Fruit	**753**	**49 (6.5)**
Apple	52	12 (23.1)
Grape	452	8 (1.8)
Orange	11	1 (9.1)
Peach	21	0
Pear	42	12 (28.6)
Persimmon	122	6 (4.9)
Pineapple	11	2 (18.2)
Tomato	5	1 (20.0)
Waxberry	7	2 (28.6)
Unidentified	30	5 (16.7)
Tree bark	**747**	**123 (16.5)**
Ailanthus altissima	15	0
Alniphyllum fortunei	1	1 (100)
Carya spp.	25	3 (12.0)
Castanea mollissima	81	15 (18.5)
Castanopsis eyrei	3	2 (66.7)
Castanopsis fabri	2	1 (50.0)
Castanopsis fargesii	1	1 (100)
Castanopsis fordii	7	5 (71.4)
Castanopsis orthacantha	3	1 (33.3)
Castanopsis tibetana	3	1 (33.3)
Cyclobalanopsis glauca	2	1 (50.0)
Cyclobalanopsis nubium	7	4 (57.1)
Dendropanax dentiger	2	1 (50.0)
Diospyros glaucifolia	1	1 (100)
Diospyros kaki	27	2 (7.4)
Hovenia acerba	1	1 (100)
Itsea elongata	1	1 (100)
Juglans sp.	5	2 (40.0)
Lithocarpus hancei	1	1 (100)
Loropetalum chinense	1	1 (100)
Machilus pauhoi	2	2 (100)
Michelia maudiae	1	1 (100)
Prunus mume	15	1 (6.7)
Pyrus sp.	23	4 (17.4)
Quercus fabri	13	3 (23.1)
Quercus griffithii	18	11 (61.1)
Quercus mongokica	243	12 (4.9)
Quercus phillyraeoides	2	1 (50.0)
Quercus wutaishanica	45	20 (44.4)

Table S2 continued

Rapanea neriifolia		1	1 (100)
Rhododendron ovatum		1	1 (100)
Schima superba		1	1 (100)
Sloanea sinensis		2	1 (50.0)
Toxicodendron succedaneum		1	1 (100)
Ulmus macrocarpa		35	10 (28.6)
Vaccinium mandarinorum		1	1 (100)
Unidentified *Quercus* trees		71	4 (5.6)
Unidentified trees		83	3 (3.6)
Soil	**139**	**15 (10.8)**	
Forest soil		51	7 (13.7)
Orchard soil		88	8 (9.1)
Rotten wood	**425**	**39 (9.2)**	
Total	**2064**	**226 (10.9)**	

Table S3. Wild Chinese isolates of *S. cerevisiae* employed in multigene sequence analysis

Lineage	Strain	Source	Geographic location	Climate zone
CHN I	HN1	Rotten wood, primeval forest	Diaoluo Mountain, Hainan	Tropical
	HN2	Rotten wood, primeval forest	Diaoluo Mountain, Hainan	Tropical
	HN3	Rotten wood, primeval forest	Diaoluo Mountain, Hainan	Tropical
	HN4	Rotten wood, primeval forest	Bawangling Mountain, Hainan	Tropical
	HN5	Rotten wood, primeval forest	Bawangling Mountain, Hainan	Tropical
	HN6	Rotten wood, primeval forest	Bawangling Mountain, Hainan	Tropical
	HN7	Rotten wood, primeval forest	Bawangling Mountain, Hainan	Tropical
	FJ5	Bark of *Castanea mollissima* (chestnut), planted tree	Wuyi Mountain, Fujian	Subtropical
	FJ7	Bark of *Hovenia acerba*, primeval forest	Wuyi Mountain, Fujian	Subtropical
	FJ8	Bark of *Cyclobalanopsis nubium*, primeval forest	Wuyi Mountain, Fujian	Subtropical
	FJ9	Bark of *Castanopsis fordii*, primeval forest	Wuyi Mountain, Fujian	Subtropical
	FJ10	Bark of *Castanopsis fordii*, primeval forest	Wuyi Mountain, Fujian	Subtropical
CHN II	SX1	Bark of a Fagaceae tree, primeval forest	Qinling Mountain, Shaanxi	Temperate
	SX2	Bark of a Fagaceae tree, primeval forest	Qinling Mountain, Shaanxi	Temperate
	SX3	Bark of *Carya* sp., primeval forest	Qinling Mountain, Shaanxi	Temperate
	SX4	Bark of a Fagaceae tree, primeval forest	Qinling Mountain, Shaanxi	Temperate
	SX5	Bark of *Quercus fabri*, primeval forest	Qinling Mountain, Shaanxi	Temperate
	SX6	Bark of Quercus fabri, primeval forest	Qinling Mountain, Shaanxi	Temperate
	SX7	Bark of a Fagaceae tree, primeval forest	Qinling Mountain, Shaanxi	Temperate
	SX8	Bark of Quercus fabri, primeval forest	Qinling Mountain, Shaanxi	Temperate
CHN III	HN8	Rotten wood, primeval forest	Wuzhi Mountain, Hainan	Tropical
	HN9	Rotten wood, primeval forest	Wuzhi Mountain, Hainan	Tropical
	HN10	Rotten wood, primeval forest	Wuzhi Mountain, Hainan	Tropical
	HN11	Rotten wood, primeval forest	Jianfengling Mountain, Hainan	Tropical
	HN19	Bark of a Fagaceae tree, primeval forest	Wuzhi Mountain, Hainan	Tropical
CHN IV	BJ3	Grape, orchard	Shunyi, Beijing	Temperate
	BJ6	Persimmon, orchard	Changping, Beijing	Temperate
	BJ8	Bark of *Quercus mongokica*, secondary forest	Wuling Mountain, Beijing	Temperate
	BJ12	Bark of *Quercus mongokica*, secondary forest	Wuling Mountain, Beijing	Temperate
	BJ14	Forest soil, secondary forset	Wuling Mountain, Beijing	Temperate
	BJ15	Bark of *Quercus wutaishanica*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ16	Bark of *Quercus wutaishanica*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ17	Bark of *Quercus wutaishanica*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ18	Bark of *Quercus wutaishanica*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ19	Bark of *Quercus wutaishanica*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ20	Bark of *Quercus wutaishanica*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ21	Bark of *Carya cathayensis*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ22	Bark of *Carya cathayensis*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ23	Bark of *Ulmus macrocarpa*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ24	Bark of *Ulmus macrocarpa*, secondary forest	Dongling Mountain, Beijing	Temperate
	BJ25	Bark of Ulmus macrocarpa, secondary forest	Dongling Mountain, Beijing	Temperate

	BJ26	Grape, orchard	Dongling Mountain, Beijing	Temperate
	BJ27	Grape, orchard	Dongling Mountain, Beijing	Temperate
	HLJ1	Bark of *Quercus mongokica*, secondary forest	Jingbo lake, Heilongjiang	Temperate
	JL5	Bark of *Pyrus* sp., orchard	Yanbian, Jilin	Temperate
CHN V	FJ2	Bark of Castanea mollissima (chestnut), planted tree	Wuyi Mountain, Fujian	Subtropical
	FJ3	Bark of Castanea mollissima, planted tree	Wuyi Mountain, Fujian	Subtropical
	FJ4	Bark of Castanea mollissima, planted tree	Wuyi Mountain, Fujian	Subtropical
	FJ6	Bark of Prunus mume, planted tree	Wuyi Mountain, Fujian	Subtropical
	HN12	Rotten wood, primeval forest	Jianfengling Mountain, Hainan	Tropical
	HN13	Rotten wood, primeval forest	Jianfengling Mountain, Hainan	Tropical
	HN14	Rotten wood, primeval forest	Jianfengling Mountain, Hainan	Tropical
	HN15	Rotten wood, primeval forest	Wuzhi Mountain, Hainan	Tropical
	HN16	Forest soil, primeval forest	Wuzhi Mountain, Hainan	Tropical
	HN17	Bark of a Fagaceae tree, primeval forest	Wuzhi Mountain, Hainan	Tropical
	HN18	Bark of a Fagaceae tree, primeval forest	Wuzhi Mountain, Hainan	Tropical
	ZJ1	Rotten wood, primeval forest	Gutian Mountain, Zhejiang	Subtropical
	ZJ2	Rotten wood, primeval forest	Gutian Mountain, Zhejiang	Subtropical
CHN VI	BJ4	Intestine of a butterfly, park	Haidian, Beijing	Temperate
	SD1	Persimmon, planted tree	Anqiu, Shandong	Temperate
	SD2	Bark of *Quercus* sp., planted tree	Taian, Shandong	Temperate
	SX10	Persimmon, planted tree	Qinling Mountain, Shaanxi	Temperate
CHN VII	HLJ2	Bark of *Quercus mongokica*, secondary forest	Jingbo lake, Heilongjiang	Temperate
	HLJ3	Bark of *Quercus mongokica*, secondary forest	Jingbo lake, Heilongjiang	Temperate
	JL4	Bark of Pyrus sp., orchard	Yanbian, Jilin	Temperate
CHN VIII	BJ2	Apple, orchard	Shunyi, Beijing	Temperate
	BJ5	Apple, orchard	Wuling Mountain, Beijing	Temperate
	BJ9	Bark of *Ulmus* sp., secondary forest	Wuling Mountain, Beijing	Temperate
	BJ10	Bark of *Ulmus* sp., secondary forest	Wuling Mountain, Beijing	Temperate
	BJ13	Rotten wood, secondary forest	Wuling Mountain, Beijing	Temperate
	BJ28	Bark of *Castanea* sp., secondary forest	Wuling Mountain, Beijing	Temperate
Sake	BJ1	Pear, orchard	Shunyi, Beijing	Temperate
	JL2	Bark of *Quercus mongokica*, secondary forest	Changbai Mountain, Jilin	Temperate
	YN2	Orange, orchard	Jinghong, Yunnan	Tropical
	XJ2	Apple, orchard	Korla, Xinjiang	Temperate
Wine/ European	NX1	Melon patch soil	Zhongwei, Ningxia	Temperate
	NX2	Mulberry field soil	Helan Mountain, Ningxia	Temperate
	XJ5	Grape, orchard	Korla, Xinjiang	Temperate
Chinese mosaic	BJ7	Bark of *Pyrus* sp., ochard	Shunyi, Beijing	Temperate
	FJ1	Bark of Castanea mollissima, planted tree	Wuyi Mountain, Fujian	Subtropical
	FJ11	Waxberry, market	Wuyi Mountain, Fujian	Subtropical
	FJ12	Peach, market	Wuyi Mountain, Fujian	Subtropical
	JL1	Bark of *Quercus* sp., secondary forest	Changbai Mountain, Jilin	Temperate
	JL3	Peach, market	Changbai Mountain, Jilin	Temperate
	NX3	Melon patch soil	Zhongwei, Ningxia	Temperate

	SD3	Bark of *Quercus* sp., planted tree	Taian, Shandong	Temperate
	SD4	Grapes, orchard	Qufu, Shandong	Temperate
	SX11	Bark of walnut (*Juglans* sp.), planted tree	Qinling Mountain, Shaanxi	Temperate
	SX9	Tomato, orchard	Qinling Mountain, Shaanxi	Temperate
	XJ1	Soil, orchard	Xinjiang	Temperate
	XJ3	Apple, orchard	Korla, Xinjiang	Temperate
	XJ4	Grape, orchard	Korla, Xinjiang	Temperate
	XJ6	Pear, orchard	Korla, Xinjiang	Temperate
	XJ7	Pear, orchard	Korla, Xinjiang	Temperate
	YN1	Pineapple, orchard	Jinghong, Yunnan	Tropical
	YN3	Soil, suburban tree	Dali, Yunnan	Tropical
	YN4	Rotten wood, suburban tree	Dali, Yunnan	Tropical
	YN5	Soil, suburban tree	Dali, Yunnan	Tropical
	YN6	Bark of Castanopsis orthacantha, suburban tree	Zixi Mountain, Yunnan	Tropical

Table S4. Population genetics analyses for different groups or lineages.

Group/Lineage	No. of isolates	SNPs	$\pi \times 10^{-3}$	I_A (*P*)	rBarD (*P*)	*P* of ILD/PHT	Rm
All	138	827	7.27	2.973 (< 0.001)	0.278 (< 0.001)	nt	94
Chinese	99	718	7.71	2.722 (< 0.001)	0.256 (< 0.001)	nt	86
SGRP	38	278	4.47	3.971 (< 0.001)	0.343 (< 0.001)	nt	35
CHN-I	12	99	1.59	0.936 (0.002)	0.086 (0.002)	0.2015	8
CHN-II	8	82	1.86	1.816 (< 0.001)	0.187 (< 0.001)	0.1613	7
CHN-III	5	16	0.58	na	na	na	0
CHN-IV	20	148	3.11	0.614 (< 0.001)	0.054 (< 0.001)	0.0001	28
CHN-V	13	125	3.23	0.671 (< 0.001)	0.059 (< 0.001)	0.0001	20
CHN-VI	4	54	2.11	na	na	na	3
CHN-VII	3	34	1.67	na	na	na	0
CHN-VIII	6	56	1.59	2.270 (< 0.001)	0.326 (< 0.001)	0.4014	4
Sake	7	87	2.58	2.453 (< 0.001)	0.223 (< 0.001)	0.0001	6
Wine/European	14	36	0.56	0.428 (0.085)	0.055 (0.085)	0.8984	2
Hainan	19	276	8.14	5.127 (< 0.001)	0.435 (< 0.001)	0.0001	29

na, not applicable because of the limited sample size.

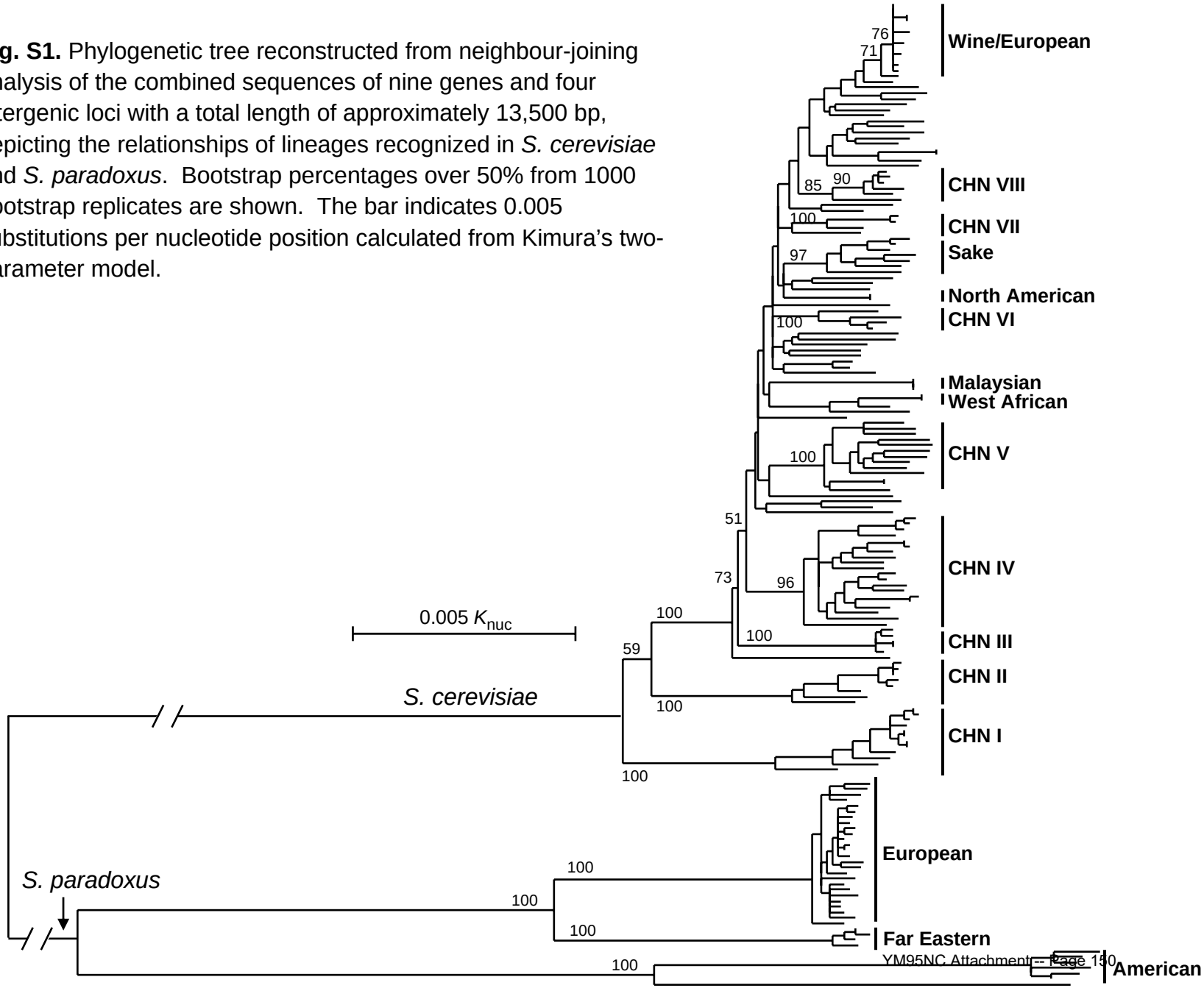
nt, not tested because of the unacceptably long time required for running the analysis.

Rm, the minimum number of recombination events.

Table S5. Average genetic distances between lineages of *S. cerevisiae* as calculated from Kimura's 2-parameter model

Lineage	1	2	3	4	5	6	7	8	9	10	11	12
1 CHN I												
2 CHN II	0.0125											
3 CHN III	0.0121	0.0108										
4 CHN IV	0.0114	0.0099	0.0074									
5 CHN V	0.0126	0.0108	0.0078	0.0063								
6 CHN VI	0.0125	0.0108	0.0071	0.0067	0.0068							
7 CHN VII	0.0120	0.0107	0.0071	0.0059	0.0061	0.0056						
8 CHN VIII	0.0122	0.0107	0.0063	0.0065	0.0070	0.0056	0.0051					
9 Malaysian	0.0135	0.0111	0.0076	0.0074	0.0068	0.0062	0.0060	0.0065				
10 Weat Africa	0.0130	0.0114	0.0074	0.0070	0.0072	0.0056	0.0064	0.0064	0.0060			
11 North American	0.0124	0.0104	0.0067	0.0066	0.0060	0.0047	0.0045	0.0056	0.0058	0.0058		
12 Sake	0.0127	0.0110	0.0072	0.0072	0.0070	0.0058	0.0052	0.0060	0.0066	0.0065	0.0047	
13 Wine/European	0.0128	0.0108	0.0066	0.0065	0.0075	0.0062	0.0059	0.0032	0.0071	0.0068	0.0058	0.0060

Fig. S1. Phylogenetic tree reconstructed from neighbour-joining analysis of the combined sequences of nine genes and four intergenic loci with a total length of approximately 13,500 bp, depicting the relationships of lineages recognized in *S. cerevisiae* and *S. paradoxus*. Bootstrap percentages over 50% from 1000 bootstrap replicates are shown. The bar indicates 0.005 substitutions per nucleotide position calculated from Kimura's two-parameter model.



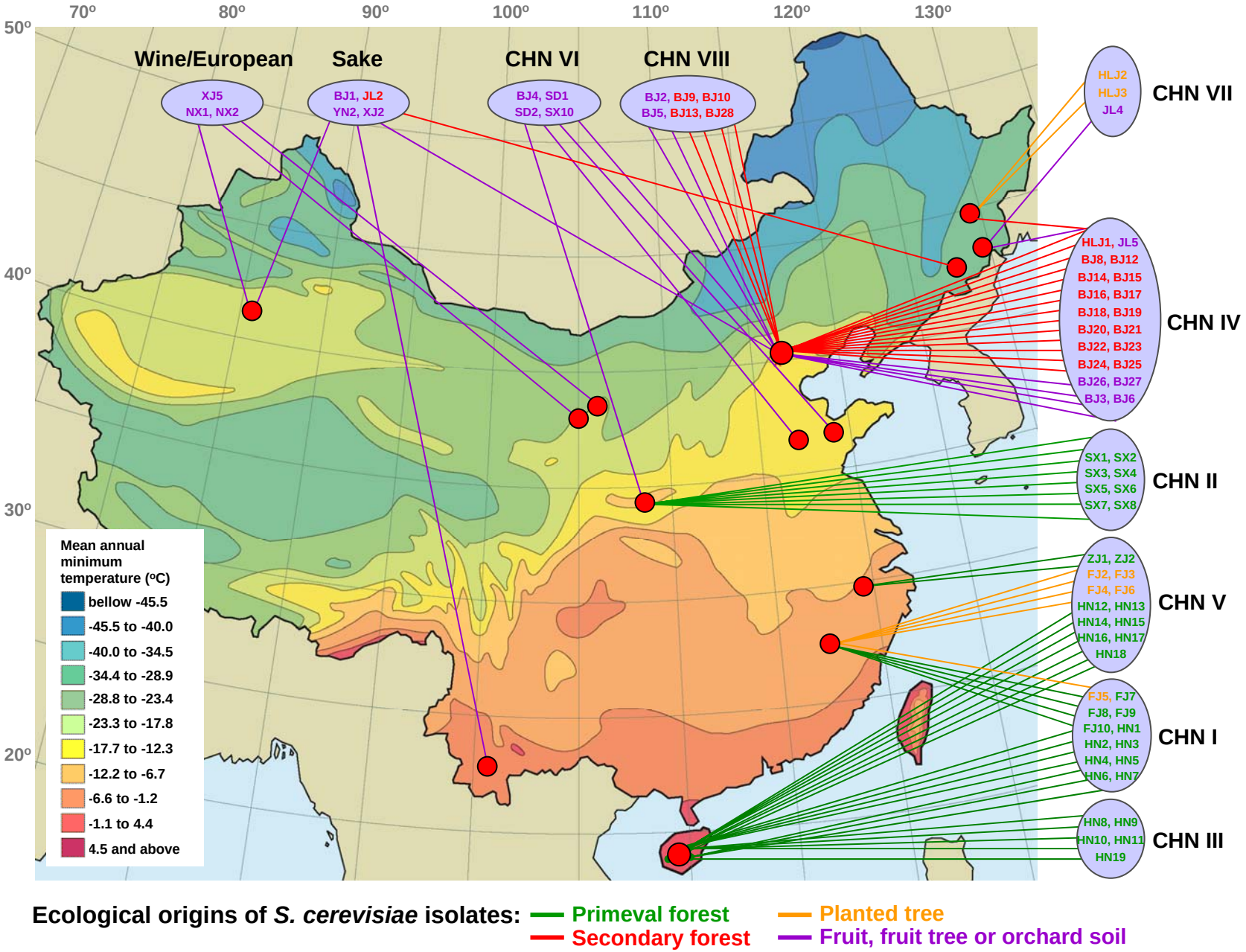


Fig. S2. Geographic origins of *S. cerevisiae* lineages from nature in China. The hardiness zones in China are labeled (the map is modified from Mark P. Widrlechner, Agricultural Research Service, U.S. Department of Agriculture).

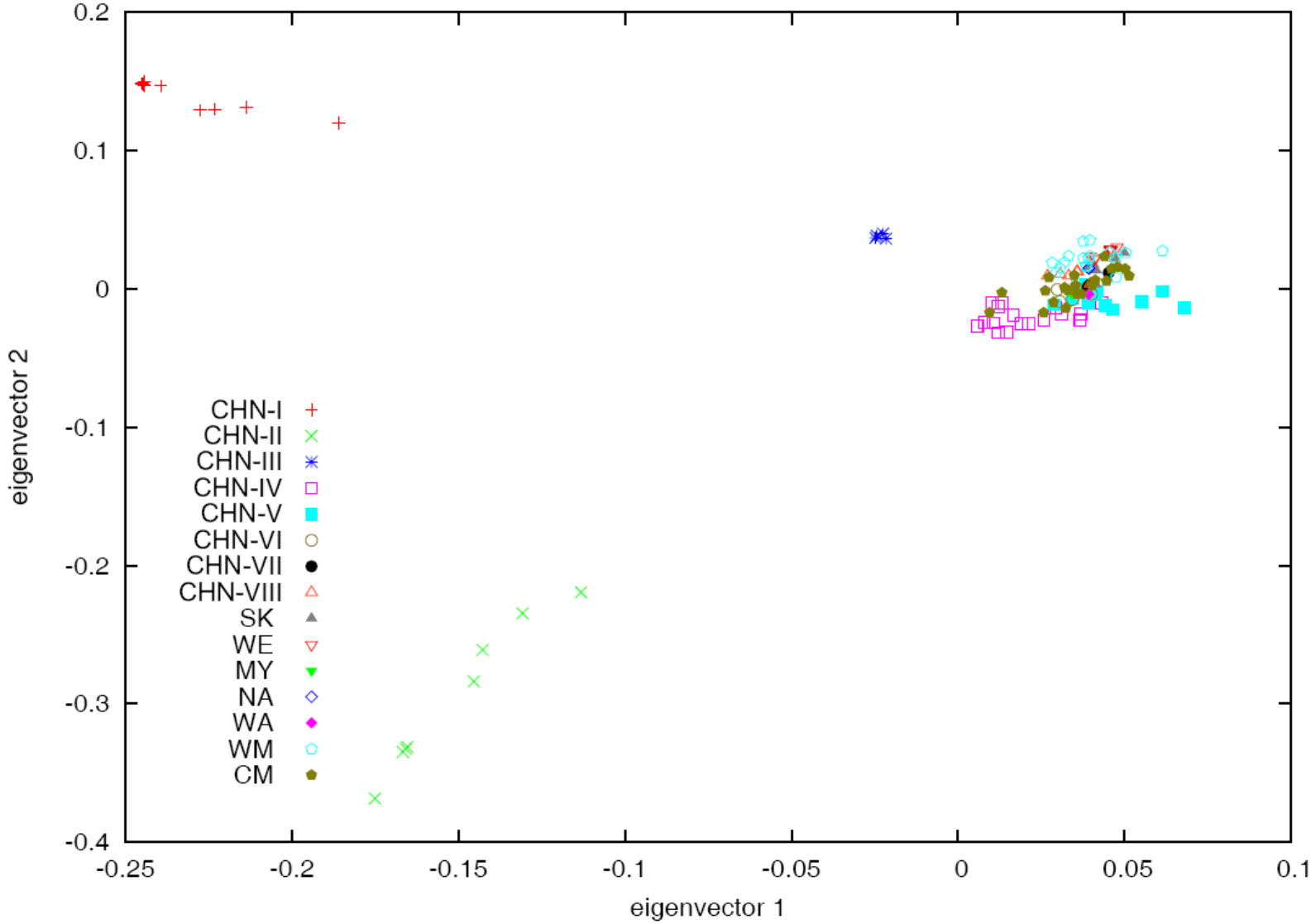
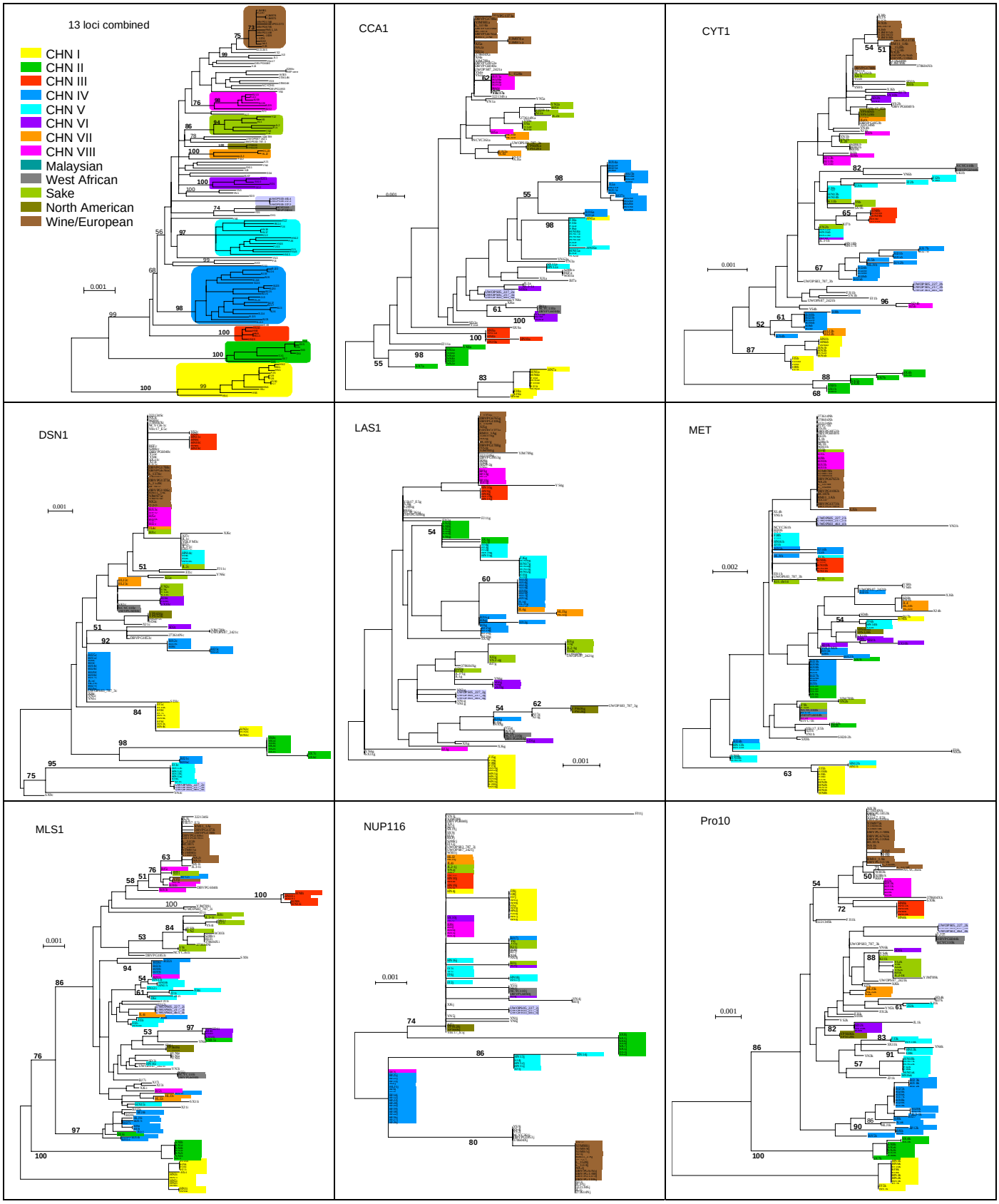


Fig. S3. Principal component analysis (PCA) of the allelic profiles of 13 loci from different lineages or groups of *S. cerevisiae* employed. SK, WE, MY, NA and WA refer to the Sake, Wine/European, Malaysian, North American and West African lineages, respecrively; WM and CM refer to the world and Chinese mosaic groups, respectively.


13 loci combined
CHN I
CHN II
CHN III
CHN IV
CHN V
CHN VI
CHN VII
CHN VIII
Malaysian
West African
Sake
North American
Wine/European
0.001
CCA1
0.001
CYT1
0.001
DSN1
0.001
LAS1
0.001
MET
0.002
MLS1
0.001
NUP116
0.001
Pro10
0.001

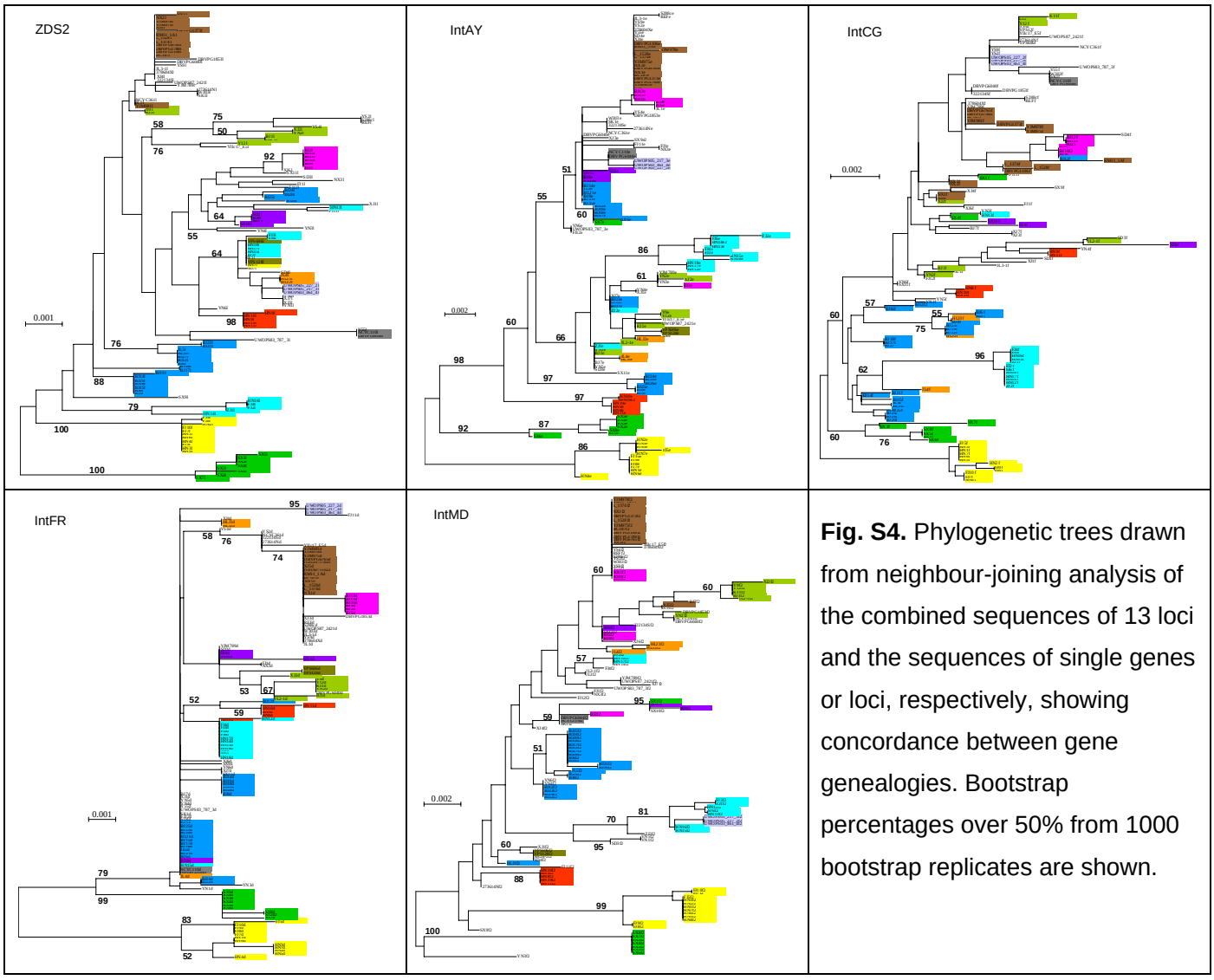


Fig. S4. Phylogenetic trees drawn from neighbour-joining analysis of the combined sequences of 13 loci and the sequences of single genes or loci, respectively, showing concordance between gene genealogies. Bootstrap percentages over 50% from 1000 bootstrap replicates are shown.

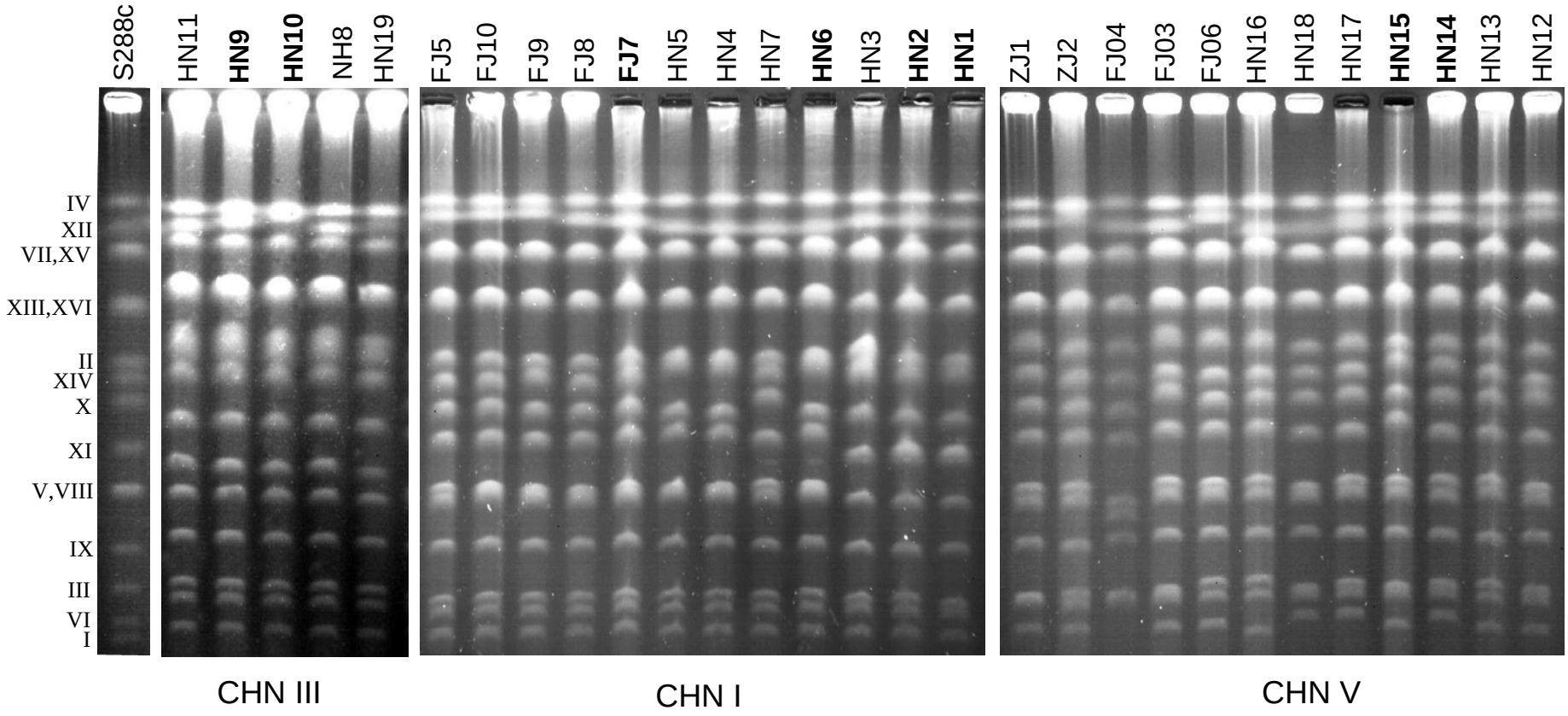


Fig. S5. Comparison of CHEF electrophoretic karyotypes of wild *S. cerevisiae* isolates from lineages CHN I, III and V. Strain S288c is used as a reference; isolates with bold numbers were selected for cross tests.

NIH GUIDELINES FOR RESEARCH INVOLVING RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULES (NIH GUIDELINES)

April 2016

DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

****

Visit the NIH OSP Web site at:
http://www.osp.od.nih.gov
For current information on Guidelines, Protocols, Principal Investigators, Meetings, and information about upcoming Gene Therapy Policy Conferences

NIH OFFICE OF SCIENCE POLICY CONTACT INFORMATION:

Office of Science Policy, National Institutes of Health, 6705 Rockledge Drive, Suite 750, MSC 7985, Bethesda, MD 20892-7985 (20817 for non-USPS mail), (301) 496-9838; (301) 496-9839 (fax).

For inquiries, information requests, and report submissions: NIHGuidelines@od.nih.gov
Human gene transfer protocol submissions: HGTprotocols@mail.nih.gov

These *NIH Guidelines* shall supersede all earlier versions until further notice.
****

FEDERAL REGISTER NOTICES

Effective June 24, 1994, Published in Federal Register, July 5, 1994 (59 FR 34472)
Amendment Effective July 28, 1994, Federal Register, August 5, 1994 (59 FR 40170)
Amendment Effective April 17, 1995, Federal Register, April 27, 1995 (60 FR 20726)
Amendment Effective December 14, 1995, Federal Register, January 19, 1996 (61 FR 1482)
Amendment Effective March 1, 1996, Federal Register, March 12, 1996 (61 FR 10004)
Amendment Effective January 23, 1997, Federal Register, January 31, 1997 (62 FR 4782)
Amendment Effective September 30, 1997, Federal Register, October 14, 1997 (62 FR 53335)
Amendment Effective October 20, 1997, Federal Register, October 29, 1997 (62 FR 56196)
Amendment Effective October 22, 1997, Federal Register, October 31, 1997 (62 FR 59032)
Amendment Effective February 4, 1998, Federal Register, February 17, 1998 (63 FR 8052)
Amendment Effective April 30, 1998, Federal Register, May 11, 1998 (63 FR 26018)
Amendment Effective April 29, 1999, Federal Register, May 11, 1999 (64 FR 25361)
Amendment Effective October 2, 2000, Federal Register, October 10, 2000 (65 FR 60328)
Amendment Effective December 28, 2000, Federal Register, January 5, 2001 (66 FR 1146)
Amendment Effective December 11, 2001, Federal Register, December 11, 2001 (66 FR 64051)
Amendment Effective December 19, 2001, Federal Register, November 19, 2001 (66 FR 57970)
Amendment Effective January 10, 2002, Federal Register, December 11, 2001 (66 FR 64052)
Amendment Effective January 24, 2002, Federal Register, November 19, 2001 (66 FR 57970)
Amendment Effective September 22, 2009, Federal Register, September 22, 2009 (74 FR 48275)
Amendment Effective January 19, 2011, Federal Register, January 19, 2011 (76 FR 3150)
Amendment Effective May 12, 2011, Federal Register, May 12, 2011 (76 FR 27653)
Amendment Effective October 11, 2011, Federal Register, October 11, 2011 (76 FR 62816)
Amendment Effective February 21, 2013, Federal Register, February 21, 2013 (78 FR 12074)
Amendment Effective March 5, 2013, Federal Register, September 5, 2012 (77 FR 54584)
Amendment Effective November 6, 2013, Federal Register, November 6, 2013 (78 FR 66751)
Amendment Effective April 27, 2016, Federal Register, March 22, 2016 (81 FR 15315)
Amendment Effective April 15, 2016, Federal Register, April 15, 2016 (81 FR 22286)

TABLE OF CONTENTS

SECTION I.	**SCOPE OF THE *NIH GUIDELINES***	**9**
Section I-A.	Purpose	9
Section I-B.	Definition of Recombinant and Synthetic Nucleic Acid Molecules	9
Section I-C.	General Applicability	10
Section I-D.	Compliance with the *NIH Guidelines*	10
Section I-E.	General Definitions	11
SECTION II.	**SAFETY CONSIDERATIONS**	**12**
Section II-A.	Risk Assessment	12
Section II-A-1.	Risk Groups	12
Section II-A-2.	Criteria for Risk Groups	12
Section II-A-3.	Comprehensive Risk Assessment	12
Section II-B.	Containment	13
SECTION III.	**EXPERIMENTS COVERED BY THE *NIH GUIDELINES***	**15**
Section III-A.	Experiments that Require Institutional Biosafety Committee Approval, RAC Review, and NIH Director Approval Before Initiation (See Section IV-C-1-b-(1), Major Actions).	15
Section III-A-1.	Major Actions under the *NIH Guidelines*	15
Section III-B.	Experiments That Require NIH OSP and Institutional Biosafety Committee Approval Before Initiation	15
Section III-B-1.	Experiments Involving the Cloning of Toxin Molecules with LD_{50} of Less than 100 Nanograms per Kilogram Body Weight	16
Section III-B-2.	Experiments that have been Approved (under Section III-A-1-a) as Major Actions under the *NIH Guidelines*	16
Section III-C.	Experiments that Require Institutional Biosafety Committee and Institutional Review Board Approvals and RAC Review Before Research Participant Enrollment	16
Section III-C-1.	Experiments Involving the Deliberate Transfer of Recombinant or Synthetic Nucleic Acid Molecules, or DNA or RNA Derived from Recombinant or Synthetic Nucleic Acid Molecules, into One or More Human Research Participants	16
Section III-D.	Experiments that Require Institutional Biosafety Committee Approval Before Initiation	17
Section III-D-1.	Experiments Using Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents as Host-Vector Systems (See Section II-A, Risk Assessment)	17
Section III-D-2.	Experiments in Which DNA From Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents is Cloned into Nonpathogenic Prokaryotic or Lower Eukaryotic Host-Vector Systems	18
Section III-D-3.	Experiments Involving the Use of Infectious DNA or RNA Viruses or Defective DNA or RNA Viruses in the Presence of Helper Virus in Tissue Culture Systems	18
Section III-D-4.	Experiments Involving Whole Animals	19
Section III-D-5.	Experiments Involving Whole Plants	19
Section III-D-6.	Experiments Involving More than 10 Liters of Culture	20
Section III-D-7.	Experiments Involving Influenza Viruses	20
Section III-E.	Experiments that Require Institutional Biosafety Committee Notice Simultaneous with Initiation	21
Section III-E-1.	Experiments Involving the Formation of Recombinant or Synthetic Nucleic Acid Molecules Containing No More than Two-Thirds of the Genome of any Eukaryotic Virus	22
Section III-E-2.	Experiments Involving Whole Plants	22
Section III-E-3.	Experiments Involving Transgenic Rodents	23
Section III-F.	Exempt Experiments	23
SECTION IV.	**ROLES AND RESPONSIBILITIES**	**24**
Section IV-A.	Policy	24
Section IV-B.	Responsibilities of the Institution	24
Section IV-B-1.	General Information	24
Section IV-B-2.	Institutional Biosafety Committee (IBC)	25
Section IV-B-2-a.	Membership and Procedures	25
Section IV-B-2-b.	Functions	27
Section IV-B-3.	Biological Safety Officer (BSO)	28
Section IV-B-4.	Plant, Plant Pathogen, or Plant Pest Containment Expert	28

Section IV-B-5. Animal Containment Expert 28
Section IV-B-6. Human Gene Therapy Expertise 28
Section IV-B-7. Principal Investigator (PI) 28
Section IV-B-7-a. General Responsibilities 29
Section IV-B-7-b. Information to Be Submitted by the Principal Investigator to NIH OSP 29
Section IV-B-7-c. Submissions by the Principal Investigator to the Institutional Biosafety Committee 30
Section IV-B-7-d. Responsibilities of the Principal Investigator Prior to Initiating Research 30
Section IV-B-7-e. Responsibilities of the Principal Investigator During the Conduct of the Research .. 30
Section IV-C. Responsibilities of the National Institutes of Health (NIH) 31
Section IV-C-1. NIH Director 31
Section IV-C-1-a. General Responsibilities 31
Section IV-C-1-b. Specific Responsibilities 31
Section IV-C-1-b-(1). Major Actions 31
Section IV-C-1-b-(2). Minor Actions 32
Section IV-C-2. Recombinant DNA Advisory Committee (RAC) 32
Section IV-C-3. Office of Science Policy (OSP) 33
Section IV-C-4. Other NIH Components 34
Section IV-D. Voluntary Compliance 34
Section IV-D-1. Basic Policy - Voluntary Compliance 34
Section IV-D-2. Institutional Biosafety Committee Approval - Voluntary Compliance 35
Section IV-D-3. Certification of Host-Vector Systems - Voluntary Compliance 35
Section IV-D-4. Requests for Exemptions and Approvals - Voluntary Compliance 35
Section IV-D-5. Protection of Proprietary Data - Voluntary Compliance 35
Section IV-D-5-a. General 35
Section IV-D-5-b. Pre-submission Review 35

SECTION V. FOOTNOTES AND REFERENCES OF SECTIONS I THROUGH IV 36

APPENDIX A. EXEMPTIONS UNDER SECTION III-F-6--SUBLISTS OF NATURAL EXCHANGERS 37

Appendix A-I. Sublist A 37
Appendix A-II. Sublist B 38
Appendix A-III. Sublist C 38
Appendix A-IV. Sublist D 38
Appendix A-V. Sublist E 38
Appendix A-VI. Sublist F 38

APPENDIX B. CLASSIFICATION OF HUMAN ETIOLOGIC AGENTS ON THE BASIS OF HAZARD 38

Appendix B-I. Risk Group 1 (RG1) Agents 39
Appendix B-II. Risk Group 2 (RG2) Agents 39
Appendix B-II-A. Risk Group 2 (RG2) - Bacterial Agents Including Chlamydia 39
Appendix B-II-B. Risk Group 2 (RG2) - Fungal Agents 40
Appendix B-II-C. Risk Group 2 (RG2) - Parasitic Agents 40
Appendix B-II-D. Risk Group 2 (RG2) - Viruses 41
Appendix B-III. Risk Group 3 (RG3) Agents 42
Appendix B-III-A. Risk Group 3 (RG3) - Bacterial Agents Including Rickettsia 42
Appendix B-III-B. Risk Group 3 (RG3) - Fungal Agents 43
Appendix B-III-C. Risk Group 3 (RG3) - Parasitic Agents 43
Appendix B-III-D. Risk Group 3 (RG3) - Viruses and Prions 43
Appendix B-IV. Risk Group 4 (RG4) Agents 44
Appendix B-IV-A. Risk Group 4 (RG4) - Bacterial Agents 44
Appendix B-IV-B. Risk Group 4 (RG4) - Fungal Agents 44
Appendix B-IV-C. Risk Group 4 (RG4) - Parasitic Agents 44
Appendix B-IV-D. Risk Group 4 (RG4) - Viral Agents 44
Appendix B-V. Animal Viral Etiologic Agents in Common Use 44
Appendix B-V-1. Murine Retroviral Vectors 45

APPENDIX C. EXEMPTIONS UNDER SECTION III-F-8 45

Appendix C-I. Recombinant or Synthetic Nucleic Acid Molecules in Tissue Culture 45
Appendix C-I-A. Exceptions 46
Appendix C-II. *Escherichia coli K-12* Host-Vector Systems 46
Appendix C-II-A. Exceptions 46
Appendix C-III. *Saccharomyces* Host-Vector Systems 46

Appendix C-III-A. Exceptions....46
Appendix C-IV. *Kluyveromyces* Host-Vector Systems....47
Appendix C-IV-A Exceptions....47
Appendix C-V. *Bacillus subtilis* or *Bacillus licheniformis* Host-Vector Systems....47
Appendix C-V-A. Exceptions....47
Appendix C-VI. Extrachromosomal Elements of Gram Positive Organisms....47
Appendix C-VI-A. Exceptions....48
Appendix C-VII. The Purchase or Transfer of Transgenic Rodents....48
Appendix C-VIII. Generation of BL1 Transgenic Rodents via Breeding....48
Appendix C-IX. Footnotes and References of Appendix C....49
APPENDIX D. MAJOR ACTIONS TAKEN UNDER THE *NIH GUIDELINES*....49
APPENDIX E. CERTIFIED HOST-VECTOR SYSTEMS (See Appendix I, Biological Containment)....68
Appendix E-I. *Bacillus subtilis*....68
Appendix E-I-A. *Bacillus subtilis* Host-Vector 1 Systems....68
Appendix E-I-B. *Bacillus subtilis* Host-Vector 2 Systems....68
Appendix E-II. Saccharomyces cerevisiae....68
Appendix E-II-A. *Saccharomyces cerevisiae* Host-Vector 2 Systems....68
Appendix E-III. *Escherichia coli*....69
Appendix E-III-A. *Escherichia coli* (EK2) Plasmid Systems....69
Appendix E-III-B. *Escherichia coli* (EK2) Bacteriophage Systems....69
Appendix E-IV. *Neurospora crassa*....69
Appendix E-IV-A. *Neurospora crassa* Host-Vector 1 Systems....69
Appendix E-V. *Streptomyces*....69
Appendix E-V-A. *Streptomyces* Host-Vector 1 Systems....69
Appendix E-VI. *Pseudomonas putida*....69
Appendix E-VI-A. *Pseudomonas putida* Host-Vector 1 Systems....69
APPENDIX F. CONTAINMENT CONDITIONS FOR CLONING OF GENES CODING FOR THE BIOSYNTHESIS OF MOLECULES TOXIC FOR VERTEBRATES....70
Appendix F-I. General Information....70
Appendix F-II. Cloning of Toxin Molecule Genes in *Escherichia coli* K-12....70
Appendix F-III. Cloning of Toxic Molecule Genes in Organisms Other Than *Escherichia coli* K-12....70
Appendix F-IV. Specific Approvals....70
APPENDIX G. PHYSICAL CONTAINMENT....71
Appendix G-I. Standard Practices and Training....71
Appendix G-II. Physical Containment Levels....71
Appendix G-II-A. Biosafety Level 1 (BL1) (See Appendix G-III-M, *Footnotes and References of Appendix G*)....72
Appendix G-II-A-1. Standard Microbiological Practices (BL1)....72
Appendix G-II-A-2. Special Practices (BL1)....72
Appendix G-II-A-3. Containment Equipment (BL1)....72
Appendix G-II-A-4. Laboratory Facilities (BL1)....72
Appendix G-II-B. Biosafety Level 2 (BL2) (See Appendix G-III-N, *Footnotes and References of Appendix G*)....72
Appendix G-II-B-1. Standard Microbiological Practices (BL2)....72
Appendix G-II-B-2. Special Practices (BL2)....73
Appendix G-II-B-3. Containment Equipment (BL2)....74
Appendix G-II-B-4. Laboratory Facilities (BL2)....74
Appendix G-II-C. Biosafety Level 3 (BL3) (See Appendix G-III-P, *Footnotes and References of Appendix G*)....74
Appendix G-II-C-1. Standard Microbiological Practices (BL3)....74
Appendix G-II-C-2. Special Practices (BL3)....75
Appendix G-II-C-2-t. Alternative Selection of Containment Equipment (BL3)....76
Appendix G-II-C-3. Containment Equipment (BL3)....76
Appendix G-II-C-4. Laboratory Facilities (BL3)....77
Appendix G-II-C-5. Biosafety Level 3 Enhanced for Research Involving Risk Group 3 Influenza Viruses. (See Appendices G-II-C-2-n, G-II-C-2-r, and G-II-C-4-i for additional guidance for facilities, waste handling, and serum collection for research involving mammaliantransmissible HPAI H5N1 virus....77
Appendix G-II-C-5-a. Containment, Practices, and Training for Research with Risk Group 3 Influenza

Viruses (BL3 Enhanced). 78
Appendix G-II-C-5-b. Containment for Animal Research. 79
Appendix G-II-C-5-c. Occupational Health 79
Appendix G-II-D. Biosafety Level 4 (BL4) 80
Appendix G-II-D-1. Standard Microbiological Practices (BL4) 80
Appendix G-II-D-2. Special Practices (BL4) 80
Appendix G-II-D-2-m. Alternative Selection of Containment Equipment (BL4) 82
Appendix G-II-D-3. Containment Equipment (BL4) 82
Appendix G-II-D-4. Laboratory Facilities (BL4) 83
Appendix G-III. Footnotes and References of Appendix G 85
APPENDIX H. SHIPMENT 86
Appendix H-III. Footnotes and References of Appendix H 87
APPENDIX I. BIOLOGICAL CONTAINMENT (See Appendix E, *Certified Host-Vector Systems*) 87
Appendix I-I. Levels of Biological Containment 87
Appendix I-I-A. Host-Vector 1 Systems 87
Appendix I-I-A-1. *Escherichia coli* K-12 Host-Vector 1 Systems (EK1) 87
Appendix I-I-A-2. Other Host-Vector 1 Systems 87
Appendix I-I-B. Host-Vector 2 Systems (EK2) 87
Appendix I-II. Certification of Host-Vector Systems 88
Appendix I-II-A. Responsibility 88
Appendix I-II-B. Data to be Submitted for Certification 88
Appendix I-II-B-1. Host-Vector 1 Systems Other than *Escherichia coli* K-12 88
Appendix I-II-B-2. Host-Vector 2 Systems 88
Appendix I-III. Footnotes and References of Appendix I 89
APPENDIX J. BIOTECHNOLOGY RESEARCH SUBCOMMITTEE 90
APPENDIX K. PHYSICAL CONTAINMENT FOR LARGE SCALE USES OF ORGANISMS CONTAINING RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULES 90
Appendix K-I. Selection of Physical Containment Levels 91
Appendix K-II. Good Large Scale Practice (GLSP) 91
Appendix K-III. Biosafety Level 1 (BL1) - Large Scale 92
Appendix K-IV. Biosafety Level 2 (BL2) - Large Scale 92
Appendix K-V. Biosafety Level 3 (BL3) - Large Scale 94
Appendix K-VI. Footnotes of Appendix K 98
Appendix K-VII. Definitions to Accompany Containment Grid and Appendix K 98
APPENDIX L. GENE THERAPY POLICY CONFERENCES (GTPCS) 99
APPENDIX M. POINTS TO CONSIDER IN THE DESIGN AND SUBMISSION OF PROTOCOLS FOR THE TRANSFER OF recombinant or synthetic NUCLEIC ACID MOLECULES INTO ONE OR MORE HUMAN RESEARCH PARTICIPANTS (POINTS TO CONSIDER) 99
Appendix M-I. Requirements for Protocol Submission, Review, and Reporting – Human Gene Transfer Experiments 100
Appendix M-I-A. Requirements for Protocol Submission 100
Appendix M-I-B. Selection of Individual Protocols for Public RAC Review and Discussion 101
Appendix M-I-C. Reporting Requirements 102
Appendix M-I-C-1. Initiation of the Clinical Investigation 102
Appendix M-I-C-2. Additional Clinical Trial Sites 102
Appendix M-I-C-3. Annual Reports 103
Appendix M-I-C-4. Safety Reporting 103
Appendix M-I-C-4-a. Safety Reporting: Content and Format 104
Appendix M-I-C-4-b. Safety Reporting: Time frames for Expedited Reports 104
Appendix M-I-C-5. Confidentiality 104
Appendix M-I-D. Safety Assessment in Human Gene Transfer Research 105
Appendix M-II. Long-Term Follow-Up 105
Appendix M-III. Footnotes of Appendix M 105
APPENDIX P. PHYSICAL AND BIOLOGICAL CONTAINMENT FOR RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULE RESEARCH INVOLVING PLANTS 105
Appendix P-I. General Plant Biosafety Levels 106
Appendix P-II. Physical Containment Levels 106

Appendix P-II-A. Biosafety Level 1 - Plants (BL1-P) 106
Appendix P-II-A-1. Standard Practices (BL1-P) 106
Appendix P-II-A-1-a. Greenhouse Access (BL1-P) 106
Appendix P-II-A-1-b. Records (BL1-P) 107
Appendix P-II-A-1-c. Decontamination and Inactivation (BL1-P) 107
Appendix P-II-A-1-d. Control of Undesired Species and Motile Macroorganisms (BL1-P) 107
Appendix P-II-A-1-e. Concurrent Experiments Conducted in the Greenhouse (BL1-P) 107
Appendix P-II-A-2. Facilities (BL1-P) 107
Appendix P-II-A-2-a. Definitions (BL1-P) 107
Appendix P-II-A-2-b. Greenhouse Design (BL1-P) 107
Appendix P-II-B. Biosafety Level 2 - Plants (BL2-P) 107
Appendix P-II-B-1. Standard Practices (BL2-P) 107
Appendix P-II-B-1-a. Greenhouse Access (BL2-P) 107
Appendix P-II-B-1-b. Records (BL2-P) 108
Appendix P-II-B-1-c. Decontamination and Inactivation (BL2-P) 108
Appendix P-II-B-1-d. Control of Undesired Species and Motile Macroorganisms (BL2-P) 108
Appendix P-II-B-1-e. Concurrent Experiments Conducted in the Greenhouse (BL2-P) 108
Appendix P-II-B-1-f. Signs (BL2-P) 108
Appendix P-II-B-1-g. Transfer of Materials (BL2-P) 108
Appendix P-II-B-1-h. Greenhouse Practices Manual (BL2-P) 109
Appendix P-II-B-2. Facilities (BL2-P) 109
Appendix P-II-B-2-a. Definitions (BL2-P) 109
Appendix P-II-B-2-b. Greenhouse Design (BL2-P) 109
Appendix P-II-B-2-c. Autoclaves (BL2-P) 109
Appendix P-II-B-2-d. Supply and Exhaust Air Ventilation Systems (BL2-P) 109
Appendix P-II-B-2-e. Other (BL2-P) 109
Appendix P-II-C. Biosafety Level 3 - Plants (BL3-P) 109
Appendix P-II-C-1. Standard Practices (BL3-P) 109
Appendix P-II-C-1-a. Greenhouse Access (BL3-P) 109
Appendix P-II-C-1-b. Records (BL3-P) 110
Appendix P-II-C-1-c. Decontamination and Inactivation (BL3-P) 110
Appendix P-II-C-1-d. Control of Undesired Species and Motile Macroorganisms (BL3-P) 110
Appendix P-II-C-1-e. Concurrent Experiments Conducted in the Greenhouse (BL3-P) 110
Appendix P-II-C-1-f. Signs (BL3-P) 110
Appendix P-II-C-1-g. Transfer of Materials (BL3-P) 110
Appendix P-II-C-1-h. Greenhouse Practices Manual (BL3-P) 111
Appendix P-II-C-1-i. Protective Clothing (BL3-P) 111
Appendix P-II-C-1-j. Other (BL3-P) 111
Appendix P-II-C-2. Facilities (BL3-P) 111
Appendix P-II-C-2-a. Definitions (BL3-P) 111
Appendix P-II-C-2-b. Greenhouse Design (BL3-P) 111
Appendix P-II-C-2-c. Autoclaves (BL3-P) 112
Appendix P-II-C-2-d. Supply and Exhaust Air Ventilation Systems (BL3-P) 112
Appendix P-II-C-2-e. Other (BL3-P) 112
Appendix P-II-D. Biosafety Level 4 - Plants (BL4-P) 112
Appendix P-II-D-1. Standard Practices (BL4-P) 112
Appendix P-II-D-1-a. Greenhouse Access (BL4-P) 112
Appendix P-II-D-1-b. Records (BL4-P) 112
Appendix P-II-D-1-c. Decontamination and Inactivation (BL4-P) 113
Appendix P-II-D-1-d. Control of Undesired Species and Motile Macroorganisms (BL4-P) 113
Appendix P-II-D-1-e. Concurrent Experiments Conducted in the Greenhouse (BL4-P) 113
Appendix P-II-D-1-f. Signs (BL4-P) 113
Appendix P-II-D-1-g. Transfer of Materials (BL4-P) 114
Appendix P-II-D-1-h. Greenhouse Practices Manual (BL4-P) 114
Appendix P-II-D-1-i. Protective Clothing (BL4-P) 114
Appendix P-II-D-2. Facilities (BL4-P) 114
Appendix P-II-D-2-a. Greenhouse Design (BL4-P) 114
Appendix P-II-D-2-b. Autoclaves (BL4-P) 115
Appendix P-II-D-2-c. Supply and Exhaust Air Ventilation Systems (BL4-P) 115
Appendix P-II-D-2-d. Other (BL4-P) 115

Appendix P-III. Biological Containment Practices 115
Appendix P-III-A. Biological Containment Practices (Plants) 115
Appendix P-III-B. Biological Containment Practices (Microorganisms) 116
Appendix P-III-C. Biological Containment Practices (Macroorganisms) 116
APPENDIX Q. PHYSICAL AND BIOLOGICAL CONTAINMENT FOR RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULE RESEARCH INVOLVING ANIMALS 116
Appendix Q-I. General Considerations 116
Appendix Q-I-A. Containment Levels 117
Appendix Q-I-B. Disposal of Animals (BL1-N through BL4-N) 117
Appendix Q-II. Physical and Biological Containment Levels 117
Appendix Q-II-A. Biosafety Level 1 - Animals (BL1-N) 117
Appendix Q-II-A-1. Standard Practices (BL1-N) 117
Appendix Q-II-A-1-a. Animal Facility Access (BL1-N) 117
Appendix Q-II-A-1-b. Other (BL1-N) 117
Appendix Q-II-A-2. Animal Facilities (BL1-N) 117
Appendix Q-II-B. Biosafety Level 2 - Animals (BL2-N) (See Appendix Q-III-A, *Footnotes and References for Appendix Q*) 117
Appendix Q-II-B-1. Standard Practices (BL2-N) 118
Appendix Q-II-B-1-a. Animal Facility Access (BL2-N) 118
Appendix Q-II-B-1-b. Decontamination and Inactivation (BL2-N) 118
Appendix Q-II-B-1-c. Signs (BL2-N) 118
Appendix Q-II-B-1-d. Protective Clothing (BL2-N) 118
Appendix Q-II-B-1-e. Records (BL2-N) 118
Appendix Q-II-B-1-f. Transfer of Materials (BL2-N) 119
Appendix Q-II-B-1-g. Other (BL2-N) 119
Appendix Q-II-B-2. Animal Facilities (BL2-N) 119
Appendix Q-II-C. Biosafety Level 3 - Animals (BL3-N) (See Appendix Q-III-B, *Footnotes and References for Appendix Q*) 120
Appendix Q-II-C-1. Standard Practices (BL3-N) 120
Appendix Q-II-C-1-a. Animal Facility Access (BL3-N) 120
Appendix Q-II-C-1-b. Decontamination and Inactivation (BL3-N) 120
Appendix Q-II-C-1-c. Signs (BL3-N) 120
Appendix Q-II-C-1-d. Protective Clothing (BL3-N) 121
Appendix Q-II-C-1-e. Records (BL3-N) 121
Appendix Q-II-C-1-f. Transfer of Materials (BL3-N) 121
Appendix Q-II-C-1-g. Other (BL3-N) 121
Appendix Q-II-C-2. Animal Facilities (BL3-N) 122
Appendix Q-II-D. Biosafety Level 4 - Animals (BL4-N) (See Appendix Q-III-C, *Footnotes and References for Appendix Q*) 123
Appendix Q-II-D-1. Standard Practices (BL4-N) 123
Appendix Q-II-D-1-a. Animal Facility Access (BL4-N) 123
Appendix Q-II-D-1-b. Decontamination and Inactivation (BL4-N) 124
Appendix Q-II-D-1-c. Signs (BL4-N) 124
Appendix Q-II-D-1-d. Protective Clothing (BL4-N) 125
Appendix Q-II-D-1-e. Records (BL4-N) 125
Appendix Q-II-D-1-f. Transfer of Materials (BL4-N) 125
Appendix Q-II-D-1-g. Other (BL4-N) 126
Appendix Q-II-D-2. Animal Facilities (BL4-N) 127
Appendix Q-III. Footnotes and References for Appendix Q 128

****

LIST OF TABLES

Appendix B - Table 1. Basis for the Classification of Biohazardous Agents by Risk Group 39

Appendix G - Table 1. Possible Alternate Combinations Of Physical And Biological Containment Safeguards 84

Appendix K - Table 1. Comparison of Good Large Scale Practice (GLSP) and Biosafety Level (BL) - Large Scale (LS) Practice..... 97

****

SECTION I. SCOPE OF THE *NIH GUIDELINES*

Section I-A. Purpose

The purpose of the *NIH Guidelines* is to specify the practices for constructing and handling: (i) recombinant nucleic acid molecules, (ii) synthetic nucleic acid molecules, including those that are chemically or otherwise modified but can base pair with naturally occurring nucleic acid molecules, and (iii) cells, organisms, and viruses containing such molecules.

Section I-A-1. Any nucleic acid molecule experiment, which according to the *NIH Guidelines* requires approval by NIH, must be submitted to NIH or to another Federal agency that has jurisdiction for review and approval. Once approvals, or other applicable clearances, have been obtained from a Federal agency other than NIH (whether the experiment is referred to that agency by NIH or sent directly there by the submitter), the experiment may proceed without the necessity for NIH review or approval. (See exception in Section I-A-1-a regarding requirement for human gene transfer protocol registration.)

Section I-A-1-a. For experiments involving the deliberate transfer of recombinant or synthetic nucleic acid molecules, or DNA or RNA derived from recombinant or synthetic nucleic acid molecules, into human research participants (human gene transfer), no research participant shall be enrolled (see definition of enrollment in Section I-E-7) until the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*); Institutional Biosafety Committee (IBC) approval (from the clinical trial site) has been obtained; Institutional Review Board (IRB) approval has been obtained; and all applicable regulatory authorization(s) have been obtained.

For a clinical trial site that is added after the completion of the NIH protocol registration process, no research participant shall be enrolled (see definition of enrollment in Section I-E-7) at the clinical trial site until IBC approval and IRB approval from that site have been obtained. Within 30 days of enrollment (see definition of enrollment in Section I-E-7) at a clinical trial site, the following documentation shall be submitted to NIH OSP: (1) Institutional Biosafety Committee approval (from the clinical trial site); (2) Institutional Review Board approval; (3) Institutional Review Board-approved informed consent document(s); and (4) NIH grant number(s) if applicable.

Section I-B. Definition of Recombinant and Synthetic Nucleic Acid Molecules

In the context of the *NIH Guidelines*, recombinant and synthetic nucleic acids are defined as:

(i) molecules that a) are constructed by joining nucleic acid molecules and b) that can replicate in a living cell, i.e., recombinant nucleic acids;
(ii) nucleic acid molecules that are chemically or by other means synthesized or amplified, including those that are chemically or otherwise modified but can base pair with naturally occurring nucleic acid molecules, i.e., synthetic nucleic acids, or
(iii) molecules that result from the replication of those described in (i) or (ii) above.

Section I-C. General Applicability

Section I-C-1. The *NIH Guidelines* are applicable to:

Section I-C-1-a. All recombinant or synthetic nucleic acid research within the United States (U.S.) or its territories that is within the category of research described in either Section I-C-1-a-(1) or Section I-C-1-a-(2).

Section I-C-1-a-(1). Research that is conducted at or sponsored by an institution that receives any support for recombinant or synthetic nucleic acid research from NIH, including research performed directly by NIH. An individual who receives support for research involving recombinant or synthetic nucleic acids must be associated with or sponsored by an institution that assumes the responsibilities assigned in the *NIH Guidelines*.

Section I-C-1-a-(2). Research that involves testing in humans of materials containing recombinant or synthetic nucleic acids developed with NIH funds, if the institution that developed those materials sponsors or participates in those projects. Participation includes research collaboration or contractual agreements, not mere provision of research materials.

Section I-C-1-b. All recombinant or synthetic nucleic acid research performed abroad that is within the category of research described in either Section I-C-1-b-(1) or Section I-C-1-b-(2).

Section I-C-1-b-(1). Research supported by NIH funds.

Section I-C-1-b-(2). Research that involves testing in humans of materials containing recombinant or synthetic nucleic acids developed with NIH funds, if the institution that developed those materials sponsors or participates in those projects. Participation includes research collaboration or contractual agreements, not mere provision of research materials.

Section I-C-1-b-(3). If the host country has established rules for the conduct of recombinant or synthetic nucleic acid molecule research, then the research must be in compliance with those rules. If the host country does not have such rules, the proposed research must be reviewed and approved by an NIH-approved Institutional Biosafety Committee or equivalent review body and accepted in writing by an appropriate national governmental authority of the host country. The safety practices that are employed abroad must be reasonably consistent with the *NIH Guidelines*.

Section I-D. Compliance with the *NIH Guidelines*

As a condition for NIH funding of recombinant or synthetic nucleic acid molecule research, institutions shall ensure that such research conducted at or sponsored by the institution, irrespective of the source of funding, shall comply with the *NIH Guidelines*.

Information concerning noncompliance with the *NIH Guidelines* may be brought forward by any person. It should be delivered to both NIH OSP and the relevant institution. The institution, generally through the Institutional Biosafety Committee, shall take appropriate action. The institution shall forward a complete report of the incident recommending any further action to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

In cases where NIH proposes to suspend, limit, or terminate financial assistance because of noncompliance with the *NIH Guidelines*, applicable DHHS and Public Health Service procedures shall govern.

The policies on compliance are as follows:

Section I-D-1. All NIH-funded projects involving recombinant or synthetic nucleic acid molecules must comply with the *NIH Guidelines*. Non-compliance may result in: (i) suspension, limitation, or termination of financial assistance for the noncompliant NIH-funded research project and of NIH funds for other recombinant or synthetic nucleic acid molecule research at the institution, or (ii) a requirement for prior NIH approval of any or all recombinant or synthetic nucleic acid molecule projects at the institution.

Section I-D-2. All non-NIH funded projects involving recombinant or synthetic nucleic acid molecule conducted at or sponsored by an institution that receives NIH funds for projects involving such techniques must comply with the *NIH Guidelines*. Noncompliance may result in: (i) suspension, limitation, or termination of NIH funds for recombinant or synthetic nucleic acid molecule research at the institution, or (ii) a requirement for prior NIH approval of any or all recombinant or synthetic nucleic acid molecule projects at the institution.

Section I-E. General Definitions

The following terms, which are used throughout the *NIH Guidelines*, are defined as follows:

Section I-E-1. An "institution" is any public or private entity (including Federal, state, and local government agencies).

Section I-E-2. An "Institutional Biosafety Committee" is a committee that: (i) meets the requirements for membership specified in Section IV-B-2, *Institutional Biosafety Committee (IBC)*, and (ii) reviews, approves, and oversees projects in accordance with the responsibilities defined in Section IV-B-2, *Institutional Biosafety Committee (IBC)*.

Section I-E-3. The "Office of Science Policy (OSP)" is the office within the NIH that is responsible for: (i) reviewing and coordinating all activities relating to the *NIH Guidelines*, and (ii) performing other duties as defined in Section IV-C-3, *Office of Science Policy (OSP)*.

Section I-E-4. The "Recombinant DNA Advisory Committee" is the public advisory committee that advises the Department of Health and Human Services (DHHS) Secretary, the DHHS Assistant Secretary for Health, and the NIH Director concerning recombinant or synthetic nucleic acid molecule research. The RAC shall be constituted as specified in Section IV-C-2, *Recombinant DNA Advisory Committee (RAC)*.

Section I-E-5. The "NIH Director" is the Director of the National Institutes of Health, or any other officer or employee of NIH to whom authority has been delegated.

Section I-E-6. "Deliberate release" is defined as a planned introduction of recombinant or synthetic nucleic acid molecule-containing microorganisms, plants, or animals into the environment.

Section I-E-7. “Enrollment” is the process of obtaining informed consent from a potential research participant, or a designated legal guardian of the participant, to undergo a test or procedure associated with the gene transfer experiment.

Section I-E-8. A “serious adverse event” is any event occurring at any dose that results in any of the following outcomes: death, a life-threatening event, in-patient hospitalization or prolongation of existing hospitalization, a persistent or significant disability/incapacity, or a congenital anomaly/birth defect. Important medical events that may not result in death, be life-threatening, or require hospitalization also may be considered a serious adverse event when, upon the basis of appropriate medical judgment, they may jeopardize the human gene transfer research subject and may require medical or surgical intervention to prevent one of the outcomes listed in this definition.

Section I-E-9. An adverse event is “associated with the use of a gene transfer product” when there is a reasonable possibility that the event may have been caused by the use of that product.

Section I-E-10. An “unexpected serious adverse event” is any serious adverse event for which the specificity or severity is not consistent with the risk information available in the current investigator’s brochure.

Section I-E-11. An “oversight body” is an institutional entity (an Institutional Biosafety Committee or an Institutional Review Board) that must review and approve a human gene transfer trial.

Section I-E-12. A “regulatory authority” in the context of human gene transfer research is a federal entity that by statute has oversight over research involving human subjects.

****

SECTION II. SAFETY CONSIDERATIONS

Section II-A. Risk Assessment

Section II-A-1. Risk Groups

Risk assessment is ultimately a subjective process. The investigator must make an initial risk assessment based on the Risk Group (RG) of an agent (see Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*). Agents are classified into four Risk Groups (RGs) according to their relative pathogenicity for healthy adult humans by the following criteria: (1) Risk Group 1 (RG1) agents are not associated with disease in healthy adult humans. (2) Risk Group 2 (RG2) agents are associated with human disease which is rarely serious and for which preventive or therapeutic interventions are *often* available. (3) Risk Group 3 (RG3) agents are associated with serious or lethal human disease for which preventive or therapeutic interventions *may be* available. (4) Risk Group 4 (RG4) agents are likely to cause serious or lethal human disease for which preventive or therapeutic interventions are *not usually* available.

Section II-A-2. Criteria for Risk Groups

Classification of agents in Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, is based on the potential effect of a biological agent on a healthy human adult and does not account for instances in which an individual may have increased susceptibility to such agents, e.g., preexisting diseases, medications, compromised immunity, pregnancy or breast feeding (which may increase exposure of infants to some agents).

Personnel may need periodic medical surveillance to ascertain fitness to perform certain activities; they may also need to be offered prophylactic vaccines and boosters (see Section IV-B-1-f, *Responsibilities of the Institution, General Information*).

Section II-A-3. Comprehensive Risk Assessment

In deciding on the appropriate containment for an experiment, the first step is to assess the risk of the agent itself. Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, classifies agents into Risk Groups based on an assessment of their ability to cause disease in humans and the available treatments for such disease. Once the Risk Group of the agent is identified, this should be followed by a thorough consideration of how the agent is to be manipulated. Factors to be considered in determining the level of containment include agent factors such as: virulence, pathogenicity, infectious dose, environmental stability, route of spread, communicability, operations, quantity, availability of vaccine or treatment, and gene product effects such as toxicity, physiological activity, and allergenicity. Any strain that is known to be more hazardous than the parent (wild-type) strain should be considered for handling at a higher containment level. Certain attenuated strains or strains that have been demonstrated to have irreversibly lost known virulence factors may qualify for a reduction of the containment level compared to the Risk Group assigned to the parent strain (see Section V-B, *Footnotes and References of Sections I-IV*).

While the starting point for the risk assessment is based on the identification of the Risk Group of the parent agent, as technology moves forward, it may be possible to develop an organism containing genetic sequences from multiple sources such that the parent agent may not be obvious. In such cases, the risk assessment should include at least two levels of analysis. The first involves a consideration of the Risk Groups of the source(s) of the sequences and the second involves an assessment of the functions that may be encoded by these sequences (e.g., virulence or transmissibility). It may be prudent to first consider the highest Risk Group classification of all agents that are the source of sequences included in the construct. Other factors to be considered include the percentage of the genome contributed by each parent agent and the predicted function or intended purpose of each contributing sequence. The initial assumption should be that all sequences will function as they did in the original host context.

The Principal Investigator and Institutional Biosafety Committee must also be cognizant that the combination of certain sequences in a new biological context may result in an organism whose risk profile could be higher than that of the contributing organisms or sequences. The synergistic function of these sequences may be one of the key attributes to consider in deciding whether a higher containment level is warranted, at least until further assessments can be carried out. A new biosafety risk may occur with an organism formed through combination

of sequences from a number of organisms or due to the synergistic effect of combining transgenes that results in a new phenotype.

A final assessment of risk based on these considerations is then used to set the appropriate containment conditions for the experiment (see Section II-B, *Containment*). The appropriate containment level may be equivalent to the Risk Group classification of the agent or it may be raised or lowered as a result of the above considerations. The Institutional Biosafety Committee must approve the risk assessment and the biosafety containment level for recombinant or synthetic nucleic acid experiments described in Sections III-A, *Experiments that Require Institutional Biosafety Committee Approval, RAC Review, and NIH Director Approval Before Initiation*; III-B, *Experiments that Require NIH OSP and Institutional Biosafety Committee Approval Before Initiation*; III-C, *Experiments that Require Institutional Biosafety Committee and Institutional Review Board Approvals and NIH OSP Registration Before Initiation*; III-D, *Experiments that Require Institutional Biosafety Committee Approval Before Initiation*.

Careful consideration should be given to the types of manipulation planned for some higher Risk Group agents. For example, the RG2 dengue viruses may be cultured under the Biosafety Level (BL) 2 containment (see Section II-B); however, when such agents are used for animal inoculation or transmission studies, a higher containment level is recommended. Similarly, RG3 agents such as Venezuelan equine encephalomyelitis and yellow fever viruses should be handled at a higher containment level for animal inoculation and transmission experiments.

Individuals working with human immunodeficiency virus (HIV), hepatitis B virus (HBV) or other bloodborne pathogens should consult the applicable Occupational Safety and Health Administration (OSHA) regulation, 29 CFR 1910.1030, and OSHA publication 3127 (1996 revised). BL2 containment is recommended for activities involving all blood-contaminated clinical specimens, body fluids, and tissues from all humans, or from HIV- or HBV-infected or inoculated laboratory animals. Activities such as the production of research-laboratory scale quantities of HIV or other bloodborne pathogens, manipulating concentrated virus preparations, or conducting procedures that may produce droplets or aerosols, are performed in a BL2 facility using the additional practices and containment equipment recommended for BL3. Activities involving industrial scale volumes or preparations of concentrated HIV are conducted in a BL3 facility, or BL3 Large Scale if appropriate, using BL3 practices and containment equipment.

Exotic plant pathogens and animal pathogens of domestic livestock and poultry are restricted and may require special laboratory design, operation and containment features not addressed in *Biosafety in Microbiological and Biomedical Laboratories* (see Section V-C, *Footnotes and References of Sections I through IV*). For information regarding the importation, possession, or use of these agents see Sections V-G and V-H, *Footnotes and References of Sections I through IV*.

Section II-B. Containment

Effective biological safety programs have been operative in a variety of laboratories for many years. Considerable information already exists about the design of physical containment facilities and selection of laboratory procedures applicable to organisms carrying additional recombinant or synthetic nucleic acid molecules (see Section V-B, *Footnotes and References of Sections I-IV*). The existing programs rely upon mechanisms that can be divided into two categories: (i) a set of standard practices that are generally used in microbiological laboratories; and (ii) special procedures, equipment, and laboratory installations that provide physical barriers that are applied in varying degrees according to the estimated biohazard. Four biosafety levels are described in Appendix G, *Physical Containment*. These biosafety levels consist of combinations of laboratory practices and techniques, safety equipment, and laboratory facilities appropriate for the operations performed and are based on the potential hazards imposed by the agents used and for the laboratory function and activity. Biosafety Level 4 provides the most stringent containment conditions, Biosafety Level 1 the least stringent.

Experiments involving recombinant or synthetic nucleic acid molecules lend themselves to a third containment mechanism, namely, the application of highly specific biological barriers. Natural barriers exist that limit either: (i) the infectivity of a vector or vehicle (plasmid or virus) for specific hosts, or (ii) its dissemination and survival in the environment. Vectors, which provide the means for recombinant or synthetic nucleic acid molecule and/or host cell replication, can be genetically designed to decrease, by many orders of magnitude, the probability of dissemination of recombinant or synthetic nucleic acid molecule outside the laboratory (see Appendix I, *Biological Containment*).

Since these three means of containment are complementary, different levels of containment can be established that apply various combinations of the physical and biological barriers along with a constant use of standard practices. Categories of containment are considered separately in order that such combinations can be conveniently expressed in the *NIH Guidelines*.

Physical containment conditions within laboratories, described in Appendix G, *Physical Containment*, may not always be appropriate for all organisms because of their physical size, the number of organisms needed for an experiment, or the particular growth requirements of the organism. Likewise, biological containment for microorganisms described in Appendix I, *Biological Containment*, may not be appropriate for all organisms, particularly higher eukaryotic organisms. However, significant information exists about the design of research facilities and experimental procedures that are applicable to organisms containing additional recombinant or synthetic nucleic acid molecules that are either integrated into the genome or into microorganisms associated with the higher organism as a symbiont, pathogen, or other relationship. This information describes facilities for physical containment of organisms used in non-traditional laboratory settings and special practices for limiting or excluding the unwanted establishment, transfer of genetic information, and dissemination of organisms beyond the intended location, based on both physical and biological containment principles. Research conducted in accordance with these conditions effectively confines the organism.

For research involving plants, four biosafety levels (BL1-P through BL4-P) are described in Appendix P, *Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Plants*. BL1-P is designed to provide a moderate level of containment for experiments for which there is convincing biological evidence that precludes the possibility of survival, transfer, or dissemination of recombinant or synthetic nucleic acid molecules into the environment, or in which there is no recognizable and predictable risk to the environment in the event of accidental release. BL2-P is designed to provide a greater level of containment for experiments involving plants and certain associated organisms in which there is a recognized possibility of survival, transmission, or dissemination of recombinant or synthetic nucleic acid molecule containing organisms, but the consequence of such an inadvertent release has a predictably minimal biological impact. BL3-P and BL4-P describe additional containment conditions for research with plants and certain pathogens and other organisms that require special containment because of their recognized potential for significant detrimental impact on managed or natural ecosystems. BL1-P relies upon accepted scientific practices for conducting research in most ordinary greenhouse or growth chamber facilities and incorporates accepted procedures for good pest control and cultural practices. BL1-P facilities and procedures provide a modified and protected environment for the propagation of plants and microorganisms associated with the plants and a degree of containment that adequately controls the potential for release of biologically viable plants, plant parts, and microorganisms associated with them. BL2-P and BL3-P rely upon accepted scientific practices for conducting research in greenhouses with organisms infecting or infesting plants in a manner that minimizes or prevents inadvertent contamination of plants within or surrounding the greenhouse. BL4-P describes facilities and practices known to provide containment of certain exotic plant pathogens.

For research involving animals, which are of a size or have growth requirements that preclude the use of conventional primary containment systems used for small laboratory animals, four biosafety levels (BL1-N through BL4-N) are described in Appendix Q, *Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Animals*. BL1-N describes containment for animals that have been modified by stable introduction of recombinant or synthetic nucleic acid molecules, or DNA derived therefrom, into the germ-line (transgenic animals) and experiments involving viable recombinant or synthetic nucleic acid molecule-modified microorganisms and is designed to eliminate the possibility of sexual transmission of the modified genome or transmission of recombinant or synthetic nucleic acid molecule-derived viruses known to be transmitted from animal parent to offspring only by sexual reproduction. Procedures, practices, and facilities follow classical methods of avoiding genetic exchange between animals. BL2-N describes containment which is used for transgenic animals associated with recombinant or synthetic nucleic acid molecule-derived organisms and is designed to eliminate the possibility of vertical or horizontal transmission. Procedures, practices, and facilities follow classical methods of avoiding genetic exchange between animals or controlling arthropod transmission. BL3-N and BL4-N describe higher levels of containment for research with certain transgenic animals involving agents which pose recognized hazard.

In constructing the *NIH Guidelines*, it was necessary to define boundary conditions for the different levels of physical and biological containment and for the classes of experiments to which they apply. These definitions do not take into account all existing and anticipated information on special procedures that will allow particular experiments to be conducted under different conditions than indicated here without affecting risk. Individual

investigators and Institutional Biosafety Committees are urged to devise simple and more effective containment procedures and to submit recommended changes in the *NIH Guidelines* to permit the use of these procedures.

****

SECTION III. EXPERIMENTS COVERED BY THE *NIH GUIDELINES*

This section describes six categories of experiments involving recombinant or synthetic nucleic acid molecules: (i) those that require Institutional Biosafety Committee (IBC) approval, RAC review, and NIH Director approval before initiation (see Section III-A), (ii) those that require NIH OSP and Institutional Biosafety Committee approval before initiation (see Section III-B), (iii) those that require Institutional Biosafety Committee and Institutional Review Board approvals and RAC review before research participant enrollment (see Section III-C), (iv) those that require Institutional Biosafety Committee approval before initiation (see Section III-D), (v) those that require Institutional Biosafety Committee notification simultaneous with initiation (see Section III-E), and (vi) those that are exempt from the *NIH Guidelines* (see Section III-F).

Note: *If an experiment falls into Sections III-A, III-B, or III-C and one of the other sections, the rules pertaining to Sections III-A, III-B, or III-C shall be followed*. If an experiment falls into Section III-F and into either Sections III-D or III-E as well, the experiment is considered exempt from the *NIH Guidelines*.

Any change in containment level, which is different from those specified in the *NIH Guidelines*, may not be initiated without the express approval of NIH OSP (see Section IV-C-1-b-(2) and its subsections, *Minor Actions*).

Section III-A. Experiments that Require Institutional Biosafety Committee Approval, RAC Review, and NIH Director Approval Before Initiation (See Section IV-C-1-b-(1), Major Actions).

Section III-A-1. Major Actions under the *NIH Guidelines*

Experiments considered as *Major Actions* under the *NIH Guidelines* cannot be initiated without submission of relevant information on the proposed experiment to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov, the publication of the proposal in the *Federal Register* for 15 days of comment, review by RAC, and specific approval by NIH. The containment conditions or stipulation requirements for such experiments will be recommended by RAC and set by NIH at the time of approval. Such experiments require Institutional Biosafety Committee approval before initiation. Specific experiments already approved are included in Appendix D, *Major Actions Taken under the NIH Guidelines*, which may be obtained from the Office of Science Policy, National Institutes of Health, preferably by submitting a request for this information to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section III-A-1-a. The deliberate transfer of a drug resistance trait to microorganisms that are not known to acquire the trait naturally (see Section V-B, *Footnotes and References of Sections I-IV*), if such acquisition could compromise the ability to control disease agents in humans, veterinary medicine, or agriculture, will be reviewed by the RAC.

Consideration should be given as to whether the drug resistance trait to be used in the experiment would render that microorganism resistant to the primary drug available to and/or indicated for certain populations, for example children or pregnant women.

At the request of an Institutional Biosafety Committee, NIH OSP will make a determination regarding whether a specific experiment involving the deliberate transfer of a drug resistance trait falls under Section III-A-1-a and therefore requires RAC review and NIH Director approval. An Institutional Biosafety Committee may also consult with NIH OSP regarding experiments that do not meet the requirements of Section III-A-1-a but nonetheless raise important public health issues. NIH OSP will consult, as needed, with one or more experts, which may include the RAC.

Section III-B. Experiments That Require NIH OSP and Institutional Biosafety Committee Approval Before Initiation

Experiments in this category cannot be initiated without submission of relevant information on the proposed

experiment to NIH OSP. The containment conditions for such experiments will be determined by NIH OSP in consultation with *ad hoc* experts. Such experiments require Institutional Biosafety Committee approval before initiation (see Section IV-B-2-b-(1), *Institutional Biosafety Committee*).

Section III-B-1. Experiments Involving the Cloning of Toxin Molecules with LD_{50} of Less than 100 Nanograms per Kilogram Body Weight

Deliberate formation of recombinant or synthetic nucleic acid molecules containing genes for the biosynthesis of toxin molecules lethal for vertebrates at an LD_{50} of less than 100 nanograms per kilogram body weight (e.g., microbial toxins such as the botulinum toxins, tetanus toxin, diphtheria toxin, and *Shigella dysenteriae* neurotoxin). Specific approval has been given for the cloning in *Escherichia coli* K-12 of DNA containing genes coding for the biosynthesis of toxic molecules which are lethal to vertebrates at 100 nanograms to 100 micrograms per kilogram body weight. Specific experiments already approved under this section may be obtained from the Office of Science Policy, National Institutes of Health, preferably by submitting a request for this information to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section III-B-2. Experiments that have been Approved (under Section III-A-1-a) as Major Actions under the *NIH Guidelines*

Upon receipt and review of an application from the investigator, NIH OSP may determine that a proposed experiment is equivalent to an experiment that has previously been approved by the NIH Director as a Major Action, including experiments approved prior to implementation of these changes. An experiment will only be considered equivalent if, as determined by NIH OSP, there are no substantive differences and pertinent information has not emerged since submission of the initial III-A-1-a experiment that would change the biosafety and public health considerations for the proposed experiments. If such a determination is made by NIH OSP, these experiments will not require review and approval under Section III-A.

Section III-C. Experiments that Require Institutional Biosafety Committee and Institutional Review Board Approvals and RAC Review (if applicable) Before Research Participant Enrollment

Section III-C-1. Experiments Involving the Deliberate Transfer of Recombinant or Synthetic Nucleic Acid Molecules, or DNA or RNA Derived from Recombinant or Synthetic Nucleic Acid Molecules, into One or More Human Research Participants

Human gene transfer is the deliberate transfer into human research participants of either:

1. Recombinant nucleic acid molecules, or DNA or RNA derived from recombinant nucleic acid molecules, or
2. Synthetic nucleic acid molecules, or DNA or RNA derived from synthetic nucleic acid molecules, that meet any one of the following criteria:
    a. Contain more than 100 nucleotides; or
    b. Possess biological properties that enable integration into the genome (e.g., *cis* elements involved in integration); or
    c. Have the potential to replicate in a cell; or
    d. Can be translated or transcribed.

No research participant shall be enrolled (see definition of enrollment in Section I-E-7) until the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*).

In its evaluation of human gene transfer protocols, the NIH will make a determination, following a request from one or more oversight bodies involved in the review at an initial site(s), whether a proposed human gene transfer experiment meets the the requirements for selecting protocols for RAC review and discussion (See Appendix M-I-B). The process of public RAC review and discussion is intended to foster the safe and ethical conduct of human gene transfer experiments. Public review and discussion of a human gene transfer experiment (and access to relevant information) also serves to inform the public about the technical aspects of the proposal, the meaning and significance of the research, and any significant safety, social, and ethical implications of the research.

Public RAC review and discussion of a human gene transfer experiment will be initiated in two exceptional

circumstances: (1) Following a request for public RAC review from one or more oversight bodies involved in the review at an initial site(s), the NIH concurs that (a) the individual protocol would significantly benefit from RAC review and (b) that the submission meets one or more of the following NIH RAC review criteria: i) the protocol uses a new vector, genetic material, or delivery methodology that represents a first-in-human experience, thus presenting an unknown risk; ii) the protocol relies on preclinical safety data that were obtained using a new preclinical model system of unknown and unconfirmed value; or iii) the proposed vector, gene construct, or method of delivery is associated with possible toxicities that are not widely known and that may render it difficult for oversight and federal regulatory bodies to evaluate the protocol rigorously. However, if one or more oversight bodies involved in the review at an initial site(s) requests public RAC review, but the NIH does not concur that (a) the individual protocol would significantly benefit from RAC review and (b) that the submission meets one or more of the RAC review criteria (listed in i, ii, or iii), then the NIH OSP will inform, within 10 working days, the requesting and other oversight bodies involved in the review at an initial site(s) that public RAC review is not warranted. (2) The NIH Director, in consultation (if needed) with appropriate regulatory authorities, determines that the submission: (a) meets one or more of the NIH RAC review criteria (listed in i, ii, or iii) and that public RAC review and discussion would provide a clear and obvious benefit to the scientific community or the public; or (b) raises significant scientific, societal, or ethical concerns.

For a clinical trial site that is added after the completion of the NIH protocol registration process, no research participant shall be enrolled (see definition of enrollment in Section I-E-7) at the clinical trial site until IBC approval and IRB approval from that site have been obtained. Within 30 days of enrollment (see definition of enrollment in Section I-E-7) at a clinical trial site, the following documentation shall be submitted to NIH OSP: (1) Institutional Biosafety Committee approval (from the clinical trial site); (2) Institutional Review Board approval; (3) Institutional Review Board-approved informed consent document(s); and (4) NIH grant number(s) if applicable.

In order to maintain public access to information regarding human gene transfer (including protocols that are not publicly reviewed by the RAC), the NIH OSP will maintain the documentation described in Appendices M-I through M-II. The information provided in response to Appendix M should not contain any confidential commercial or financial information or trade secrets, enabling all aspects of RAC review to be open to the public.

Note: For specific directives concerning the use of retroviral vectors for gene delivery, consult Appendix B-V-1, *Murine Retroviral Vectors*.

Section III-D. Experiments that Require Institutional Biosafety Committee Approval Before Initiation

Prior to the initiation of an experiment that falls into this category, the Principal Investigator must submit a registration document to the Institutional Biosafety Committee which contains the following information: (i) the source(s) of DNA; (ii) the nature of the inserted DNA sequences; (iii) the host(s) and vector(s) to be used; (iv) if an attempt will be made to obtain expression of a foreign gene, and if so, indicate the protein that will be produced; and (v) the containment conditions that will be implemented as specified in the *NIH Guidelines*. For experiments in this category, the registration document shall be dated, signed by the Principal Investigator, and filed with the Institutional Biosafety Committee. The Institutional Biosafety Committee shall review and approve all experiments in this category prior to their initiation. Requests to decrease the level of containment specified for experiments in this category will be considered by NIH (see Section IV-C-1-b-(2)-(c), *Minor Actions*).

Section III-D-1. Experiments Using Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents as Host-Vector Systems (See Section II-A, Risk Assessment)

Section III-D-1-a. Experiments involving the introduction of recombinant or synthetic nucleic acid molecules into Risk Group 2 agents will usually be conducted at Biosafety Level (BL) 2 containment. Experiments with such agents will usually be conducted with whole animals at BL2 or BL2-N (Animals) containment.

Section III-D-1-b. Experiments involving the introduction of recombinant or synthetic nucleic acid molecules into Risk Group 3 agents will usually be conducted at BL3 containment. Experiments with such agents will usually be conducted with whole animals at BL3 or BL3-N containment.

Section III-D-1-c. Experiments involving the introduction of recombinant or synthetic nucleic acid molecules into Risk Group 4 agents shall be conducted at BL4 containment. Experiments with such agents shall be conducted with whole animals at BL4 or BL4-N containment.

Section III-D-1-d. Containment conditions for experiments involving the introduction of recombinant or synthetic nucleic acid molecules into restricted agents shall be set on a case-by-case basis following NIH OSP review. A U.S. Department of Agriculture - Animal and Plant Health Inspection Service (USDA/APHIS) permit is required for work with plant or animal pathogens (see Section V-G and V-M, *Footnotes and References of Sections I-IV*). Experiments with such agents shall be conducted with whole animals at BL4 or BL4-N containment.

Section III-D-2. Experiments in Which DNA From Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents is Cloned into Nonpathogenic Prokaryotic or Lower Eukaryotic Host-Vector Systems

Section III-D-2-a. Experiments in which DNA from Risk Group 2 or Risk Group 3 agents (see Section II-A, *Risk Assessment)* is transferred into nonpathogenic prokaryotes or lower eukaryotes may be performed under BL2 containment. Experiments in which DNA from Risk Group 4 agents is transferred into nonpathogenic prokaryotes or lower eukaryotes may be performed under BL2 containment after demonstration that only a totally and irreversibly defective fraction of the agent's genome is present in a given recombinant. In the absence of such a demonstration, BL4 containment shall be used. The Institutional Biosafety Committee may approve the specific lowering of containment for particular experiments to BL1. Many experiments in this category are exempt from the *NIH Guidelines* (see Section III-F, *Exempt Experiments*). Experiments involving the formation of recombinant or synthetic nucleic acid molecules for certain genes coding for molecules toxic for vertebrates require NIH OSP approval (see Section III-B-1, *Experiments Involving the Cloning of Toxin Molecules with LD_{50} of Less than 100 Nanograms Per Kilogram Body Weight*) or shall be conducted under NIH specified conditions as described in Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*.

Section III-D-2-b. Containment conditions for experiments in which DNA from restricted agents is transferred into nonpathogenic prokaryotes or lower eukaryotes shall be determined by NIH OSP following a case-by-case review (see Section V-L, *Footnotes and References of Sections I-IV*). A U.S. Department of Agriculture permit is required for work with plant or animal pathogens (see Section V-G, *Footnotes and References of Sections I-IV*).

Section III-D-3. Experiments Involving the Use of Infectious DNA or RNA Viruses or Defective DNA or RNA Viruses in the Presence of Helper Virus in Tissue Culture Systems

Caution: Special care should be used in the evaluation of containment levels for experiments which are likely to either enhance the pathogenicity (e.g., insertion of a host oncogene) or to extend the host range (e.g., introduction of novel control elements) of viral vectors under conditions that permit a productive infection. In such cases, serious consideration should be given to increasing physical containment by at least one level.

Note: Recombinant or synthetic nucleic acid molecules or nucleic acid molecules derived therefrom, which contain less than two-thirds of the genome of any eukaryotic virus (all viruses from a single Family (see Section V-J, *Footnotes and References of Sections I-IV*) being considered identical (see Section V-K, *Footnotes and References of Sections I-IV*), are considered defective and may be used in the absence of helper under the conditions specified in Section III-E-1, *Experiments Involving the Formation of Recombinant or Synthetic Molecules Containing No More than Two-Thirds of the Genome of any Eukaryotic Virus*.

Section III-D-3-a. Experiments involving the use of infectious or defective Risk Group 2 viruses (see Appendix B-II, *Risk Group 2 Agents*) in the presence of helper virus may be conducted at BL2.

Section III-D-3-b. Experiments involving the use of infectious or defective Risk Group 3 viruses (see Appendix B-III-D, *Risk Group 3 (RG3) - Viruses and Prions*) in the presence of helper virus may be conducted at BL3.

Section III-D-3-c. Experiments involving the use of infectious or defective Risk Group 4 viruses (see Appendix B-IV-D, *Risk Group 4 (RG4) - Viral Agents*) in the presence of helper virus may be conducted at BL4.

Section III-D-3-d. Experiments involving the use of infectious or defective restricted poxviruses (see Sections V-A and V-L, *Footnotes and References of Sections I-IV*) in the presence of helper virus shall be determined on a case-by-case basis following NIH OSP review. A U.S. Department of Agriculture permit is required for work with plant or animal pathogens (see Section V-G, *Footnotes and References of Sections I-IV*).

Section III-D-3-e. Experiments involving the use of infectious or defective viruses in the presence of helper virus which are not covered in Sections III-D-3-a through III-D-3-d may be conducted at BL1.

Section III-D-4. Experiments Involving Whole Animals

This section covers experiments involving whole animals in which the animal's genome has been altered by stable introduction of recombinant or synthetic nucleic acid molecules, or nucleic acids derived therefrom, into the germ-line (transgenic animals) and experiments involving viable recombinant or synthetic nucleic acid molecule-modified microorganisms tested on whole animals. For the latter, other than viruses which are only vertically transmitted, the experiments may *not* be conducted at BL1-N containment. A minimum containment of BL2 or BL2-N is required.

Caution - Special care should be used in the evaluation of containment conditions for some experiments with transgenic animals. For example, such experiments might lead to the creation of novel mechanisms or increased transmission of a recombinant pathogen or production of undesirable traits in the host animal. In such cases, serious consideration should be given to increasing the containment conditions.

Section III-D-4-a. Recombinant or synthetic nucleic acid molecules, or DNA or RNA molecules derived therefrom, from any source except for greater than two-thirds of eukaryotic viral genome may be transferred to any non-human vertebrate or any invertebrate organism and propagated under conditions of physical containment comparable to BL1 or BL1-N and appropriate to the organism under study (see Section V-B, *Footnotes and References of Sections I-IV*). Animals that contain sequences from viral vectors, which do not lead to transmissible infection either directly or indirectly as a result of complementation or recombination in animals, may be propagated under conditions of physical containment comparable to BL1 or BL1-N and appropriate to the organism under study. Experiments involving the introduction of other sequences from eukaryotic viral genomes into animals are covered under Section III-D-4-b, *Experiments Involving Whole Animals*. For experiments involving recombinant or synthetic nucleic acid molecule-modified Risk Groups 2, 3, 4, or restricted organisms, see Sections V-A, V-G, and V-L, *Footnotes and References of Sections I-IV*. It is important that the investigator demonstrate that the fraction of the viral genome being utilized does not lead to productive infection. A U.S. Department of Agriculture permit is required for work with plant or animal pathogens (see Section V-G, *Footnotes and References of Sections I-IV*).

Section III-D-4-b. For experiments involving recombinant or synthetic nucleic acid molecules, or DNA or RNA derived therefrom, involving whole animals, including transgenic animals, and not covered by Section III-D-1, *Experiments Using Human or Animal Pathogens (Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents as Host-Vector Systems)*, or Section III-D-4-a, the appropriate containment shall be determined by the Institutional Biosafety Committee.

Section III-D-4-c. Exceptions under Section III-D-4, *Experiments Involving Whole Animals*

Section III-D-4-c-(1). Experiments involving the generation of transgenic rodents that require BL1 containment are described under Section III-E-3, *Experiments Involving Transgenic Rodents*.

Section III-D-4-c-(2). The purchase or transfer of transgenic rodents is exempt from the *NIH Guidelines* under Section III-F, *Exempt Experiments* (see Appendix C-VII, *The Purchase or Transfer of Transgenic Rodents*).

Section III-D-5. Experiments Involving Whole Plants

Experiments to genetically engineer plants by recombinant or synthetic nucleic acid molecule methods, to use such plants for other experimental purposes (e.g., response to stress), to propagate such plants, or to use plants together with microorganisms or insects containing recombinant or synthetic nucleic acid molecules, may be conducted under the containment conditions described in Sections III-D-5-a through III-D-5-e. If experiments involving whole plants are not described in Section III-D-5 and do not fall under Sections III-A, III-B, III-D or III-F, they are included in Section III-E.

NOTE - For recombinant or synthetic nucleic acid molecule experiments falling under Sections III-D-5-a through III-D-5-d, physical containment requirements may be reduced to the next lower level by appropriate biological containment practices, such as conducting experiments on a virus with an obligate insect vector in the absence of that vector or using a genetically attenuated strain.

Section III-D-5-a. BL3-P (Plants) or BL2-P + biological containment is recommended for experiments involving most exotic (see Section V-M, *Footnotes and References of Sections I-IV*) infectious agents with recognized potential for serious detrimental impact on managed or natural ecosystems when recombinant or synthetic nucleic acid molecule techniques are associated with whole plants.

Section III-D-5-b. BL3-P or BL2-P + biological containment is recommended for experiments involving plants containing cloned genomes of readily transmissible exotic (see Section V-M, *Footnotes and References of Sections I-IV*) infectious agents with recognized potential for serious detrimental effects on managed or natural ecosystems in which there exists the possibility of reconstituting the complete and functional genome of the infectious agent by genomic complementation *in planta.*

Section III-D-5-c. BL4-P containment is recommended for experiments with a small number of readily transmissible exotic (see Section V-M, *Footnotes and References of Sections I-IV*) infectious agents, such as the soybean rust fungus (*Phakospora pachyrhizi*) and maize streak or other viruses in the presence of their specific arthropod vectors, that have the potential of being serious pathogens of major U.S. crops.

Section III-D-5-d. BL3-P containment is recommended for experiments involving sequences encoding potent vertebrate toxins introduced into plants or associated organisms. Recombinant or synthetic nucleic acid molecules containing genes for the biosynthesis of toxin molecules lethal for vertebrates at an LD_{50} of <100 nanograms per kilogram body weight fall under Section III-B-1, *Experiments Involving the Cloning of Toxin Molecules with LD_{50} of Less than 100 Nanograms Per Kilogram Body Weight*, and require NIH OSP and Institutional Biosafety Committee approval before initiation.

Section III-D-5-e. BL3-P or BL2-P + biological containment is recommended for experiments with microbial pathogens of insects or small animals associated with plants if the recombinant or synthetic nucleic acid molecule-modified organism has a recognized potential for serious detrimental impact on managed or natural ecosystems.

Section III-D-6. Experiments Involving More than 10 Liters of Culture

The appropriate containment will be decided by the Institutional Biosafety Committee. Where appropriate, Appendix K, *Physical Containment for Large Scale Uses of Organisms Containing Recombinant or Synthetic Recombinant or synthetic nucleic acid Molecules*, shall be used. Appendix K describes containment conditions Good Large Scale Practice through BL3-Large Scale.

Section III-D-7. Experiments Involving Influenza Viruses

Experiments with influenza viruses generated by recombinant or synthetic methods (e.g., generation by reverse genetics of chimeric viruses with reassorted segments, introduction of specific mutations) shall be conducted at the biosafety level containment corresponding to the Risk Group of the virus that was the source of the majority of segments in the recombinant or synthetic virus (e.g., experiments with viruses containing a majority of segments from a RG3 virus shall be conducted at BL3). Experiments with influenza viruses containing genes or segments from 1918-1919 H1N1 (1918 H1N1), human H2N2 (1957-1968) and highly pathogenic avian influenza H5N1 strains within the Goose/Guangdong/96-like H5 lineage (HPAI H5N1), including, but not limited to, strains of HPAI H5N1 virus that are transmissible among mammals by respiratory droplets, as demonstrated in an appropriate animal model or clinically in humans (hereinafter referred to as mammalian-transmissible HPAI H5N1 virus), shall be conducted at BL3 enhanced containment (see Appendix G-II-C-5, Biosafety Level 3 Enhanced for Research Involving Risk Group 3 Influenza Viruses) unless indicated below.

Section III-D-7-a. Human H2N2 (1957-1968). Experiments with influenza viruses containing the H2 hemagglutinin (HA) segment shall be conducted at BL3 enhanced (see Appendix G-II-C-5, Biosafety Level 3 Enhanced for Research Involving Risk Group 3 Influenza Viruses). Experiments with the H2 HA gene in cold-adapted, live attenuated vaccine strains (e.g., A/Ann Arbor/6/60 H2N2) may be conducted at BL2 containment provided segments with mutations conferring temperature sensitivity and attenuation are not altered in the recombinant or synthetic virus. Experiments with Risk Group 2 influenza viruses containing genes from human H2N2 other than the HA gene can be worked on at BL2.

Section III-D-7-b. Highly Pathogenic Avian Influenza H5N1 strains within the Goose/Guangdong/96-like H5 lineage (HPAI H5N1). Experiments involving influenza viruses containing a majority of genes and/or segments from a HPAI H5N1 influenza virus shall be conducted at BL3 enhanced containment, (see Appendix G-II-C-5, Biosafety Level 3 Enhanced for Research Involving Risk Group 3 Influenza Viruses). Experiments involving influenza viruses containing a minority of genes and/or segments from a HPAI H5N1 influenza virus shall be conducted at BL3 enhanced unless a risk assessment performed by the IBC determines that they can be conducted safely at biosafety level 2 and after they have been excluded pursuant to 9 CFR 121.3(e). NIH OSP is available to IBCs to provide consultation with the RAC and influenza virus experts when risk assessments are being made to determine the appropriate biocontainment for experiments with influenza viruses containing a minority of gene/segments from HPAI H5N1. Such experiments may be performed at BL3 enhanced containment or containment may be lowered to biosafety level 2, the level of containment for most research with other influenza viruses. (USDA/APHIS regulations and decisions on lowering containment also apply.) In deciding to lower containment, the IBC should consider whether, in at least two animal models (e.g., ferret, mouse, Syrian golden hamster, cotton rat, non-human primates), there is evidence that the resulting influenza virus shows reduced replication and virulence compared to the parental RG3 virus at relevant doses. This should be determined by measuring biological indices appropriate for the specific animal model (e.g., severe weight loss, elevated temperature, mortality or neurological symptoms).

Section III-D-7-c. 1918 H1N1. Experiments involving influenza viruses containing any gene or segment from 1918 H1N1 shall be conducted at BL3 enhanced containment (see Appendix G-II-C-5, Biosafety Level 3 Enhanced for Research Involving Risk Group 3 Influenza Viruses).

Section III-D-7-d. Antiviral Susceptibility and Containment. The availability of antiviral drugs as preventive and therapeutic measures is an important safeguard for experiments with 1918 H1N1, HPAI H5N1, and human H2N2 (1957-1968). If an influenza virus containing genes from one of these viruses is resistant to both classes of current antiviral agents, adamantanes and neuraminidase inhibitors, higher containment may be required based on the risk assessment considering transmissibility to humans, virulence, pandemic potential, alternative antiviral agents if available, etc.

Experiments with 1918 H1N1, human H2N2 (1957-1968) or HPAI H5N1 that are designed to create resistance to neuraminidase inhibitors or other effective antiviral agents (including investigational antiviral agents being developed for influenza) would be subject to Section III-A-1 (Major Actions) and require RAC review and NIH Director approval. As per Section I-A-1 of the *NIH Guidelines,* if the agent is a Select Agent, the NIH will defer to the appropriate Federal agency (HHS or USDA Select Agent Divisions) on such experiments.

Section III-E. Experiments that Require Institutional Biosafety Committee Notice Simultaneous with Initiation

Experiments not included in Sections III-A, III-B, III-C, III-D, III-F, and their subsections are considered in Section III-E. All such experiments may be conducted at BL1 containment. For experiments in this category, a registration document (see Section III-D, *Experiments that Require Institutional Biosafety Committee Approval Before Initiation*) shall be dated and signed by the investigator and filed with the local Institutional Biosafety Committee at the time the experiment is initiated. The Institutional Biosafety Committee reviews and approves all such proposals, but Institutional Biosafety Committee review and approval prior to initiation of the experiment is not required (see Section IV-A, *Policy*). For example, experiments in which all components derived from non-pathogenic prokaryotes and non-pathogenic lower eukaryotes fall under Section III-E and may be conducted at BL1 containment.

Section III-E-1. Experiments Involving the Formation of Recombinant or Synthetic Nucleic Acid Molecules Containing No More than Two-Thirds of the Genome of any Eukaryotic Virus

Recombinant or synthetic nucleic acid molecules containing no more than two-thirds of the genome of any eukaryotic virus (all viruses from a single Family being considered identical [see Section V-J, *Footnotes and References of Sections I-IV*]) may be propagated and maintained in cells in tissue culture using BL1 containment. For such experiments, it must be demonstrated that the cells lack helper virus for the specific Families of defective viruses being used. If helper virus is present, procedures specified under Section III-D-3, *Experiments Involving the Use of Infectious Animal or Plant DNA or RNA Viruses or Defective Animal or Plant DNA or RNA Viruses in the Presence of Helper Virus in Tissue Culture Systems*, should be used. The DNA may contain fragments of the genome of viruses from more than one Family but each fragment shall be less than two-thirds of a genome.

Section III-E-2. Experiments Involving Whole Plants

This section covers experiments involving nucleic acid molecule-modified whole plants, and/or experiments involving recombinant or synthetic nucleic acid molecule-modified organisms associated with whole plants, except those that fall under Section III-A, III-B, III-D, or III-F. It should be emphasized that knowledge of the organisms and judgment based on accepted scientific practices should be used in all cases in selecting the appropriate level of containment. For example, if the genetic modification has the objective of increasing pathogenicity or converting a non-pathogenic organism into a pathogen, then a higher level of containment may be appropriate depending on the organism, its mode of dissemination, and its target organisms. By contrast, a lower level of containment may be appropriate for small animals associated with many types of recombinant or synthetic nucleic acid molecule-modified plants.

Section III-E-2-a. BL1-P is recommended for all experiments with recombinant or synthetic recombinant or synthetic nucleic acid molecule-containing plants and plant-associated microorganisms not covered in Section III-E-2-b or other sections of the *NIH Guidelines*. Examples of such experiments are those involving recombinant or synthetic nucleic acid molecule-modified plants that are not noxious weeds or that cannot interbreed with noxious weeds in the immediate geographic area, and experiments involving whole plants and recombinant or synthetic nucleic acid molecule-modified non-exotic (see Section V-M, *Footnotes and References of Sections I-IV*) microorganisms that have no recognized potential for rapid and widespread dissemination or for serious detrimental impact on managed or natural ecosystems (e.g., *Rhizobium* spp. and *Agrobacterium* spp.).

Section III-E-2-b. BL2-P or BL1-P + biological containment is recommended for the following experiments:

Section III-E-2-b-(1). Plants modified by recombinant or synthetic nucleic acid molecules that are noxious weeds or can interbreed with noxious weeds in the immediate geographic area.

Section III-E-2-b-(2). Plants in which the introduced DNA represents the complete genome of a non-exotic infectious agent (see Section V-M, *Footnotes and References of Sections I-IV*).

Section III-E-2-b-(3). Plants associated with recombinant or synthetic nucleic acid molecule-modified non-exotic microorganisms that have a recognized potential for serious detrimental impact on managed or natural ecosystems (see Section V-M, *Footnotes and References of Sections I-IV*).

Section III-E-2-b-(4). Plants associated with recombinant or synthetic nucleic acid molecule-modified exotic microorganisms that have no recognized potential for serious detrimental impact on managed or natural ecosystems (see Section V-M, *Footnotes and References of Sections I-IV*).

Section III-E-2-b-(5). Experiments with recombinant or synthetic nucleic acid molecule-modified arthropods or small animals associated with plants, or with arthropods or small animals with recombinant or synthetic nucleic acid molecule-modified microorganisms associated with them if the recombinant or synthetic nucleic acid molecule-modified microorganisms have no recognized potential for serious detrimental impact on managed or natural ecosystems (see Section V-M, *Footnotes and References of Sections I-IV*).

Section III-E-3. Experiments Involving Transgenic Rodents

This section covers experiments involving the generation of rodents in which the animal's genome has been altered by stable introduction of recombinant or synthetic nucleic acid molecules, or nucleic acids derived therefrom, into the germ-line (transgenic rodents). Only experiments that require BL1 containment are covered under this section; experiments that require BL2, BL3, or BL4 containment are covered under Section III-D-4, *Experiments Involving Whole Animals*.

Section III-E-3-a. Experiments involving the breeding of certain BL1 transgenic rodents are exempt under Section III-F, *Exempt Experiments* (See Appendix C-VIII, *Generation of BL1 Transgenic Rodents via Breeding*).

Section III-F. Exempt Experiments

The following recombinant or synthetic nucleic acid molecules are exempt from the *NIH Guidelines* and registration with the Institutional Biosafety Committee is not required; however, other federal and state standards of biosafety may still apply to such research (for example, the Centers for Disease Control and Prevention (CDC)/NIH publication *Biosafety in Microbiological and Biomedical Laboratories*).

Section III-F-1. Those synthetic nucleic acids that: (1) can neither replicate nor generate nucleic acids that can replicate in any living cell (e.g., oligonucleotides or other synthetic nucleic acids that do not contain an origin of replication or contain elements known to interact with either DNA or RNA polymerase), and (2) are not designed to integrate into DNA, and (3) do not produce a toxin that is lethal for vertebrates at an LD50 of less than 100 nanograms per kilogram body weight. If a synthetic nucleic acid is deliberately transferred into one or more human research participants and meets the criteria of Section III-C, it is not exempt under this Section.

Section III-F-2. Those that are not in organisms, cells, or viruses and that have not been modified or manipulated (e.g., encapsulated into synthetic or natural vehicles) to render them capable of penetrating cellular membranes.

Section III-F-3. Those that consist solely of the exact recombinant or synthetic nucleic acid sequence from a single source that exists contemporaneously in nature.

Section III-F-4. Those that consist entirely of nucleic acids from a prokaryotic host, including its indigenous plasmids or viruses when propagated only in that host (or a closely related strain of the same species), or when transferred to another host by well-established physiological means.

Section III-F-5. Those that consist entirely of nucleic acids from a eukaryotic host including its chloroplasts, mitochondria, or plasmids (but excluding viruses) when propagated only in that host (or a closely related strain of the same species).

Section III-F-6. Those that consist entirely of DNA segments from different species that exchange DNA by known physiological processes, though one or more of the segments may be a synthetic equivalent. A list of such exchangers will be prepared and periodically revised by the NIH Director with advice of the RAC after appropriate notice and opportunity for public comment (see Section IV-C-1-b-(1)-(c), *Major Actions*). See Appendices A-I through A-VI, *Exemptions under Section III-F-6--Sublists of Natural Exchangers*, for a list of natural exchangers that are exempt from the *NIH Guidelines*.

Section III-F-7. Those genomic DNA molecules that have acquired a transposable element, provided the transposable element does not contain any recombinant and/or synthetic DNA.

Section III-F-8. Those that do not present a significant risk to health or the environment (see Section IV-C-1-b-(1)-(c), *Major Actions*), as determined by the NIH Director, with the advice of the RAC, and following appropriate notice and opportunity for public comment. See Appendix C, *Exemptions under Section III-F-8* for other classes of experiments which are exempt from the *NIH Guidelines*.

****

SECTION IV. ROLES AND RESPONSIBILITIES

Section IV-A. Policy

The safe conduct of experiments involving recombinant or synthetic nucleic acid molecules depends on the individual conducting such activities. The *NIH Guidelines* cannot anticipate every possible situation. Motivation and good judgment are the key essentials to protection of health and the environment. The *NIH Guidelines* are intended to assist the institution, Institutional Biosafety Committee, Biological Safety Officer, and the Principal Investigator in determining safeguards that should be implemented. The *NIH Guidelines* will never be complete or final since all conceivable experiments involving recombinant or synthetic nucleic acid molecules cannot be foreseen. The utilization of new genetic manipulation techniques may enable work previously conducted using recombinant means to be accomplished faster, more efficiently, or at larger scale. These techniques have not yet yielded organisms that present safety concerns that fall outside the current risk assessment framework used for recombinant nucleic acid research. Nonetheless, an appropriate risk assessment of experiments involving these techniques must be conducted taking into account the way these approaches may alter the risk assessment. As new techniques develop, the *NIH Guidelines* should be periodically reviewed to determine whether and how such research should be explicitly addressed.

It is the responsibility of the institution and those associated with it to adhere to the intent of the *NIH Guidelines* as well as to their specifics. Therefore, each institution (and the Institutional Biosafety Committee acting on its behalf) is responsible for ensuring that all research with recombinant or synthetic nucleic acid molecules conducted at or sponsored by that institution is conducted in compliance with the *NIH Guidelines*. The following roles and responsibilities constitute an administrative framework in which safety is an essential and integral part of research involving recombinant or synthetic nucleic acid molecules. Further clarifications and interpretations of roles and responsibilities will be issued by NIH as necessary.

Section IV-B. Responsibilities of the Institution

Section IV-B-1. General Information

Each institution conducting or sponsoring recombinant or synthetic nucleic acid molecule research which is covered by the *NIH Guidelines* is responsible for ensuring that the research is conducted in full conformity with the provisions of the *NIH Guidelines*. In order to fulfill this responsibility, the institution shall:

Section IV-B-1-a. Establish and implement policies that provide for the safe conduct of recombinant or synthetic nucleic acid molecule research and that ensure compliance with the *NIH Guidelines*. As part of its general responsibilities for implementing the *NIH Guidelines*, the institution may establish additional procedures, as deemed necessary, to govern the institution and its components in the discharge of its responsibilities under the *NIH Guidelines*. Such procedures may include: (i) statements formulated by the institution for the general implementation of the *NIH Guidelines*, and (ii) any additional precautionary steps the institution deems appropriate.

Section IV-B-1-b. Establish an Institutional Biosafety Committee that meets the requirements set forth in Section IV-B-2-a and carries out the functions detailed in Section IV-B-2-b.

Section IV-B-1-c. Appoint a Biological Safety Officer (who is also a member of the Institutional Biosafety Committee) if the institution: (i) conducts recombinant or synthetic nucleic acid molecule research at Biosafety Level (BL) 3 or BL4, or (ii) engages in large-scale (greater than 10 liters) research. The Biological Safety Officer carries out the duties specified in Section IV-B-3.

Section IV-B-1-d. Appoint at least one individual with expertise in plant, plant pathogen, or plant pest containment principles (who is a member of the Institutional Biosafety Committee) if the institution conducts recombinant or synthetic nucleic acid molecule research that requires Institutional Biosafety Committee approval in accordance with Appendix P, *Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Plants*.

Section IV-B-1-e. Appoint at least one individual with expertise in animal containment principles (who is a member of the Institutional Biosafety Committee) if the institution conducts recombinant or synthetic nucleic acid molecule research that requires Institutional Biosafety Committee approval in accordance with Appendix Q,

Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Animals.

Section IV-B-1-f. Ensure that when the institution participates in or sponsors recombinant or synthetic nucleic acid molecule research involving human subjects: (i) the Institutional Biosafety Committee has adequate expertise and training (using *ad hoc* consultants as deemed necessary), (ii) all aspects of Appendix M have been appropriately addressed by the Principal Investigator; and (iii) no research participant shall be enrolled (see definition of enrollment in Section I-E-7) in a human gene transfer experiment until the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*), Institutional Biosafety Committee approval has been obtained, Institutional Review Board approval has been obtained, and all applicable regulatory authorizations have been obtained. Institutional Biosafety Committee approval must be obtained from the clinical trial site.

Section IV-B-1-g. Assist and ensure compliance with the *NIH Guidelines* by Principal Investigators conducting research at the institution as specified in Section IV-B-7.

Section IV-B-1-h. Ensure appropriate training for the Institutional Biosafety Committee Chair and members, Biological Safety Officer and other containment experts (when applicable), Principal Investigators, and laboratory staff regarding laboratory safety and implementation of the *NIH Guidelines*. The Institutional Biosafety Committee Chair is responsible for ensuring that Institutional Biosafety Committee members are appropriately trained. The Principal Investigator is responsible for ensuring that laboratory staff are appropriately trained. The institution is responsible for ensuring that the Principal Investigator has sufficient training; however, this responsibility may be delegated to the Institutional Biosafety Committee.

Section IV-B-1-i. Determine the necessity for health surveillance of personnel involved in connection with individual recombinant or synthetic nucleic acid molecule projects; and if appropriate, conduct a health surveillance program for such projects. The institution shall establish and maintain a health surveillance program for personnel engaged in large-scale research or production activities involving viable organisms containing recombinant or synthetic nucleic acid molecules which require BL3 containment at the laboratory scale. The institution shall establish and maintain a health surveillance program for personnel engaged in animal research involving viable recombinant or synthetic nucleic acid molecule-containing microorganisms that require BL3 or greater containment in the laboratory. The *Laboratory Safety Monograph* discusses various components of such a program (e.g., records of agents handled, active investigation of relevant illnesses, and the maintenance of serial serum samples for monitoring serologic changes that may result from the employees' work experience). Certain medical conditions may place a laboratory worker at increased risk in any endeavor where infectious agents are handled. Examples cited in the *Laboratory Safety Monograph* include gastrointestinal disorders and treatment with steroids, immunosuppressive drugs, or antibiotics. Workers with such disorders or treatment should be evaluated to determine whether they should be engaged in research with potentially hazardous organisms during their treatment or illness. Copies of the *Laboratory Safety Monograph* are available from the Office of Science Policy, National Institutes of Health, preferably by submitting a request for this information to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section IV-B-1-j. Report any significant problems, violations of the *NIH Guidelines*, or any significant research-related accidents and illnesses to NIH OSP within thirty days, unless the institution determines that a report has already been filed by the Principal Investigator or Institutional Biosafety Committee. Reports shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section IV-B-2. Institutional Biosafety Committee (IBC)

The institution shall establish an Institutional Biosafety Committee whose responsibilities need not be restricted to recombinant or synthetic nucleic acid molecule research. The Institutional Biosafety Committee shall meet the following requirements:

Section IV-B-2-a. Membership and Procedures

Section IV-B-2-a-(1). The Institutional Biosafety Committee must be comprised of no fewer than five members so selected that they collectively have experience and expertise in recombinant or synthetic nucleic acid molecule technology and the capability to assess the safety of recombinant or synthetic nucleic acid molecule

research and to identify any potential risk to public health or the environment. At least two members shall not be affiliated with the institution (apart from their membership on the Institutional Biosafety Committee) and who represent the interest of the surrounding community with respect to health and protection of the environment (e.g., officials of state or local public health or environmental protection agencies, members of other local governmental bodies, or persons active in medical, occupational health, or environmental concerns in the community). The Institutional Biosafety Committee shall include at least one individual with expertise in plant, plant pathogen, or plant pest containment principles when experiments utilizing Appendix P, *Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Plants*, require prior approval by the Institutional Biosafety Committee. The Institutional Biosafety Committee shall include at least one scientist with expertise in animal containment principles when experiments utilizing Appendix Q, *Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Animals*, require Institutional Biosafety Committee prior approval. When the institution conducts recombinant or synthetic nucleic acid molecule research at BL3, BL4, or Large Scale (greater than 10 liters), a Biological Safety Officer is mandatory and shall be a member of the Institutional Biosafety Committee (see Section IV-B-3, *Biological Safety Officer*). When the institution participates in or sponsors recombinant or synthetic nucleic acid molecule research involving human research participants, the institution must ensure that: (i) the Institutional Biosafety Committee has adequate expertise and training (using *ad hoc* consultants as deemed necessary); (ii) all aspects of Appendix M have been appropriately addressed by the Principal Investigator; (iii) no research participant shall be enrolled (see definition of enrollment in Section I-E-7) in a human gene transfer experiment until the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*); and (iv) final IBC approval is granted only after the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*). Institutional Biosafety Committee approval must be obtained from the clinical trial site.

Note: Individuals, corporations, and institutions not otherwise covered by the *NIH Guidelines*, are encouraged to adhere to the standards and procedures set forth in Sections I through IV (see Section IV-D, *Voluntary Compliance*. The policy and procedures for establishing an Institutional Biosafety Committee under *Voluntary Compliance*, are specified in Section IV-D-2, *Institutional Biosafety Committee Approval*).

Section IV-B-2-a-(2). In order to ensure the competence necessary to review and approve recombinant or synthetic nucleic acid molecule activities, it is recommended that the Institutional Biosafety Committee: (i) include persons with expertise in recombinant or synthetic nucleic acid molecule technology, biological safety, and physical containment; (ii) include or have available as consultants persons knowledgeable in institutional commitments and policies, applicable law, standards of professional conduct and practice, community attitudes, and the environment, and (iii) include at least one member representing the laboratory technical staff.

Section IV-B-2-a-(3). The institution shall file an annual report with NIH OSP which includes: (i) a roster of all Institutional Biosafety Committee members clearly indicating the Chair, contact person, Biological Safety Officer (if applicable), plant expert (if applicable), animal expert (if applicable), human gene therapy expertise or *ad hoc* consultant (if applicable); and (ii) biographical sketches of all Institutional Biosafety Committee members (including community members).

Section IV-B-2-a-(4). No member of an Institutional Biosafety Committee may be involved (except to provide information requested by the Institutional Biosafety Committee) in the review or approval of a project in which he/she has been or expects to be engaged or has a direct financial interest.

Section IV-B-2-a-(5). The institution, that is ultimately responsible for the effectiveness of the Institutional Biosafety Committee, may establish procedures that the Institutional Biosafety Committee shall follow in its initial and continuing review and approval of applications, proposals, and activities.

Section IV-B-2-a-(6). When possible and consistent with protection of privacy and proprietary interests, the institution is encouraged to open its Institutional Biosafety Committee meetings to the public.

Section IV-B-2-a-(7). Upon request, the institution shall make available to the public all Institutional Biosafety Committee meeting minutes and any documents submitted to or received from funding agencies which the latter are required to make available to the public. If public comments are made on Institutional Biosafety Committee actions, the institution shall forward both the public comments and the Institutional Biosafety Committee's response to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section IV-B-2-b. Functions

On behalf of the institution, the Institutional Biosafety Committee is responsible for:

Section IV-B-2-b-(1). Reviewing recombinant or synthetic nucleic acid molecule research conducted at or sponsored by the institution for compliance with the *NIH Guidelines* as specified in Section III, *Experiments Covered by the NIH Guidelines*, and approving those research projects that are found to conform with the *NIH Guidelines*. This review shall include: (i) independent assessment of the containment levels required by the *NIH Guidelines* for the proposed research; (ii) assessment of the facilities, procedures, practices, and training and expertise of personnel involved in recombinant or synthetic nucleic acid molecule research; (iii) ensuring that all aspects of Appendix M have been appropriately addressed by the Principal Investigator; (iv) ensuring that no research participant is enrolled (see definition of enrollment in Section I-E-7) in a human gene transfer experiment until the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*), Institutional Biosafety Committee approval (from the clinical trial site) has been obtained, Institutional Review Board approval has been obtained, and all applicable regulatory authorizations have been obtained; (v) for human gene transfer protocols selected for public RAC review and discussion, consideration of the issues raised and recommendations made as a result of this review and consideration of the Principal Investigator's response to the recommendations; (vi) ensuring that final IBC approval is granted only after the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*); and (vii) ensuring compliance with all surveillance, data reporting, and adverse event reporting requirements set forth in the *NIH Guidelines*.

Section IV-B-2-b-(2). Notifying the Principal Investigator of the results of the Institutional Biosafety Committee's review and approval.

Section IV-B-2-b-(3). Lowering containment levels for certain experiments as specified in Section III-D-2-a, *Experiments in which DNA from Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents is Cloned into Nonpathogenic Prokaryotic or Lower Eukaryotic Host-Vector Systems*.

Section IV-B-2-b-(4). Setting containment levels as specified in Sections III-D-4-b, *Experiments Involving Whole Animals*, and III-D-5, *Experiments Involving Whole Plants*.

Section IV-B-2-b-(5). Periodically reviewing recombinant or synthetic nucleic acid molecule research conducted at the institution to ensure compliance with the *NIH Guidelines*.

Section IV-B-2-b-(6). Adopting emergency plans covering accidental spills and personnel contamination resulting from recombinant or synthetic nucleic acid molecule research.

Note: The *Laboratory Safety Monograph* describes basic elements for developing specific procedures dealing with major spills of potentially hazardous materials in the laboratory, including information and references about decontamination and emergency plans. The NIH and the CDC are available to provide consultation and direct assistance, if necessary, as posted in the *Laboratory Safety Monograph*. The institution shall cooperate with the state and local public health departments by reporting any significant research-related illness or accident that may be hazardous to the public health.

Section IV-B-2-b-(7). Reporting any significant problems with or violations of the *NIH Guidelines* and any significant research-related accidents or illnesses to the appropriate institutional official and NIH OSP within 30 days, unless the Institutional Biosafety Committee determines that a report has already been filed by the Principal Investigator. Reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section IV-B-2-b-(8). The Institutional Biosafety Committee may not authorize initiation of experiments which are not explicitly covered by the *NIH Guidelines* until NIH (with the advice of the RAC when required) establishes the containment requirement.

Section IV-B-2-b-(9). Performing such other functions as may be delegated to the Institutional Biosafety Committee under Section IV-B-2, *Institutional Biosafety Committee*.

Section IV-B-3. Biological Safety Officer (BSO)

Section IV-B-3-a. The institution shall appoint a Biological Safety Officer if it engages in large-scale research or production activities involving viable organisms containing recombinant or synthetic nucleic acid molecules.

Section IV-B-3-b. The institution shall appoint a Biological Safety Officer if it engages in recombinant or synthetic nucleic acid molecule research at BL3 or BL4. The Biological Safety Officer shall be a member of the Institutional Biosafety Committee.

Section IV-B-3-c. The Biological Safety Officer's duties include, but are not be limited to:

Section IV-B-3-c-(1). Periodic inspections to ensure that laboratory standards are rigorously followed;

Section IV-B-3-c-(2). Reporting to the Institutional Biosafety Committee and the institution any significant problems, violations of the *NIH Guidelines*, and any significant research-related accidents or illnesses of which the Biological Safety Officer becomes aware unless the Biological Safety Officer determines that a report has already been filed by the Principal Investigator;

Section IV-B-3-c-(3). Developing emergency plans for handling accidental spills and personnel contamination and investigating laboratory accidents involving recombinant or synthetic nucleic acid molecule research;

Section IV-B-3-c-(4). Providing advice on laboratory security;

Section IV-B-3-c-(5). Providing technical advice to Principal Investigators and the Institutional Biosafety Committee on research safety procedures.

Note: See the *Laboratory Safety Monograph* for additional information on the duties of the Biological Safety Officer.

Section IV-B-4. Plant, Plant Pathogen, or Plant Pest Containment Expert

When the institution conducts recombinant or synthetic nucleic acid molecule research that requires Institutional Biosafety Committee approval in accordance with Appendix P, *Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Plants*, the institution shall appoint at least one individual with expertise in plant, plant pathogen, or plant pest containment principles (who is a member of the Institutional Biosafety Committee).

Section IV-B-5. Animal Containment Expert

When the institution conducts recombinant or synthetic nucleic acid molecule research that requires Institutional Biosafety Committee approval in accordance with Appendix Q, *Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Animals*, the institution shall appoint at least one individual with expertise in animal containment principles (who is a member of the Institutional Biosafety Committee).

Section IV-B-6. Human Gene Therapy Expertise

When the institution participates in or sponsors recombinant or synthetic nucleic acid molecule research involving human subjects, the institution must ensure that: (i) the Institutional Biosafety Committee has adequate expertise and training (using *ad hoc* consultants as deemed necessary) and (ii) all aspects of Appendix M, *Points to Consider in the Design and Submission of Protocols for the Transfer of Recombinant or Synthetic Nucleic Acid Molecules into One or More Human Subjects (Points to Consider)* , have been appropriately addressed by the Principal Investigator prior to its approval.

Section IV-B-7. Principal Investigator (PI)

On behalf of the institution, the Principal Investigator is responsible for full compliance with the *NIH Guidelines* in the conduct of recombinant or synthetic nucleic acid molecule research. A Principal Investigator engaged in human gene transfer research may delegate to another party, such as a corporate sponsor, the reporting

functions set forth in Appendix M, with written notification to the NIH OSP of the delegation and of the name(s), address, telephone, and fax numbers of the contact. The Principal Investigator is responsible for ensuring that the reporting requirements are fulfilled and will be held accountable for any reporting lapses.

Section IV-B-7-a. General Responsibilities

As part of this general responsibility, the Principal Investigator shall:

Section IV-B-7-a-(1). Initiate or modify no recombinant or synthetic nucleic acid molecule research which requires Institutional Biosafety Committee approval prior to initiation (see Sections III-A, III-B, III-C, III-D, and III-E, *Experiments Covered by the NIH Guidelines*) until that research or the proposed modification thereof has been approved by the Institutional Biosafety Committee and has met all other requirements of the *NIH Guidelines*;

Section IV-B-7-a-(2). Determine whether experiments are covered by Section III-E, *Experiments that Require Institutional Biosafety Committee Notice Simultaneous with Initiation*, and ensure that the appropriate procedures are followed;

Section IV-B-7-a-(3). Report any significant problems, violations of the *NIH Guidelines*, or any significant research-related accidents and illnesses to the Biological Safety Officer (where applicable), Greenhouse/Animal Facility Director (where applicable), Institutional Biosafety Committee, NIH OSP, and other appropriate authorities (if applicable) within 30 days. Reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section IV-B-7-a-(4). Report any new information bearing on the *NIH Guidelines* to the Institutional Biosafety Committee and to NIH OSP (reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section IV-B-7-a-(5). Be adequately trained in good microbiological techniques;

Section IV-B-7-a-(6). Adhere to Institutional Biosafety Committee approved emergency plans for handling accidental spills and personnel contamination; and

Section IV-B-7-a-(7). Comply with shipping requirements for recombinant or synthetic nucleic acid molecules (see Appendix H, *Shipment*, for shipping requirements and the *Laboratory Safety Monograph* for technical recommendations).

Section IV-B-7-b. Information to Be Submitted by the Principal Investigator to NIH OSP

The Principal Investigator shall:

Section IV-B-7-b-(1). Submit information to NIH OSP for certification of new host-vector systems;

Section IV-B-7-b-(2). Petition NIH OSP, with notice to the Institutional Biosafety Committee, for proposed exemptions to the *NIH Guidelines*;

Section IV-B-7-b-(3). Petition NIH OSP, with concurrence of the Institutional Biosafety Committee, for approval to conduct experiments specified in Sections III-A-1, *Major Actions Under the NIH Guidelines*, and III-B, *Experiments that Require NIH OSP and Institutional Biosafety Committee Approval Before Initiation*;

Section IV-B-7-b-(4). Petition NIH OSP for determination of containment for experiments requiring case-by-case review; and

Section IV-B-7-b-(5). Petition NIH OSP for determination of containment for experiments not covered by the *NIH Guidelines*.

Section IV-B-7-b-(6). Ensure that all aspects of Appendix M have been appropriately addressed prior to submission. No research participant shall be enrolled (see definition of enrollment in Section I-E-7) in a human

gene transfer experiment until the NIH protocol registration process has been completed (see Appendix M-I-B, *Selection of Individual Protocols for Public RAC Review and Discussion*); IBC approval (from the clinical trial site) has been obtained; Institutional Review Board (IRB) approval has been obtained; and all applicable regulatory authorization(s) have been obtained.

For a clinical trial site that is added after the completion of the NIH protocol registration process, no research participant shall be enrolled (see definition of enrollment in Section I-E-7) at the clinical trial site until IBC approval and IRB approval from that site have been obtained. Within 30 days of enrollment (see definition of enrollment in Section I-E-7) at a clinical trial site, the following documentation shall be submitted to NIH OSP: (1) Institutional Biosafety Committee approval (from the clinical trial site); (2) Institutional Review Board approval; (3) Institutional Review Board-approved informed consent document(s); and (4) NIH grant number(s) if applicable.

Section IV-B-7-c. Submissions by the Principal Investigator to the Institutional Biosafety Committee

The Principal Investigator shall:

Section IV-B-7-c-(1). Make an initial determination of the required levels of physical and biological containment in accordance with the *NIH Guidelines*;

Section IV-B-7-c-(2). Select appropriate microbiological practices and laboratory techniques to be used for the research;

Section IV-B-7-c-(3). Submit the initial research protocol and any subsequent changes (e.g., changes in the source of DNA or host-vector system), if covered under Sections III-A, III-B, III-C, III-D, or III-E (*Experiments Covered by the NIH Guidelines*), to the Institutional Biosafety Committee for review and approval or disapproval; and

Section IV-B-7-c-(4). Remain in communication with the Institutional Biosafety Committee throughout the conduct of the project.

Section IV-B-7-d. Responsibilities of the Principal Investigator Prior to Initiating Research

The Principal Investigator shall:

Section IV-B-7-d-(1). Make available to all laboratory staff the protocols that describe the potential biohazards and the precautions to be taken;

Section IV-B-7-d-(2). Instruct and train laboratory staff in: (i) the practices and techniques required to ensure safety, and (ii) the procedures for dealing with accidents; and

Section IV-B-7-d-(3). Inform the laboratory staff of the reasons and provisions for any precautionary medical practices advised or requested (e.g., vaccinations or serum collection).

Section IV-B-7-e. Responsibilities of the Principal Investigator During the Conduct of the Research

The Principal Investigator shall:

Section IV-B-7-e-(1). Supervise the safety performance of the laboratory staff to ensure that the required safety practices and techniques are employed;

Section IV-B-7-e-(2). Investigate and report any significant problems pertaining to the operation and implementation of containment practices and procedures in writing to the Biological Safety Officer (where applicable), Greenhouse/Animal Facility Director (where applicable), Institutional Biosafety Committee, NIH OSP, and other appropriate authorities (if applicable) (reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Section IV-B-7-e-(3). Correct work errors and conditions that may result in the release of recombinant or synthetic nucleic acid molecule materials; and

Section IV-B-7-e-(4). Ensure the integrity of the physical containment (e.g., biological safety cabinets) and the biological containment (e.g., purity and genotypic and phenotypic characteristics).

Section IV-B-7-e-(5). Comply with reporting requirements for human gene transfer experiments conducted in compliance with the *NIH Guidelines* (see Appendix M-I-C, *Reporting Requirements*).

Section IV-C. Responsibilities of the National Institutes of Health (NIH)

Section IV-C-1. NIH Director

The NIH Director is responsible for: (i) establishing the *NIH Guidelines*, (ii) overseeing their implementation, and (iii) their final interpretation. The NIH Director has responsibilities under the *NIH Guidelines* that involve OSP and RAC. OSP's responsibilities under the *NIH Guidelines* are administrative. Advice from RAC is primarily scientific, technical, and ethical. In certain circumstances, there is specific opportunity for public comment with published response prior to final action.

Section IV-C-1-a. General Responsibilities

The NIH Director is responsible for:

Section IV-C-1-a-(1). Promulgating requirements as necessary to implement the *NIH Guidelines*;

Section IV-C-1-a-(2). Establishing and maintaining RAC to carry out the responsibilities set forth in Section IV-C-2, *Recombinant DNA Advisory Committee* (RAC membership is specified in its charter and in Section IV-C-2);

Section IV-C-1-a-(3). Establishing and maintaining NIH OSP to carry out the responsibilities defined in Section IV-C-3, *Office of Science Policy*;

Section IV-C-1-a-(4). Conducting and supporting training programs in laboratory safety for Institutional Biosafety Committee members, Biological Safety Officers and other institutional experts (if applicable), Principal Investigators, and laboratory staff.

Section IV-C-1-a-(5). Establishing and convening Gene Therapy Policy Conferences as described in Appendix L, *Gene Therapy Policy Conferences.*

Section IV-C-1-b. Specific Responsibilities

In carrying out the responsibilities set forth in this section, the NIH Director, or a designee shall weigh each proposed action through appropriate analysis and consultation to determine whether it complies with the *NIH Guidelines* and presents no significant risk to health or the environment.

Section IV-C-1-b-(1). Major Actions

To execute *Major Actions*, the NIH Director shall seek the advice of RAC and provide an opportunity for public and Federal agency comment. Specifically, the Notice of Meeting and *Proposed Actions* shall be published in the *Federal Register* at least 15 days before the RAC meeting. The NIH Director's decision/recommendation (at his/her discretion) may be published in the *Federal Register* for 15 days of comment before final action is taken. The NIH Director's final decision/recommendation, along with responses to public comments, shall be published in the *Federal Register*. The RAC and Institutional Biosafety Committee Chairs shall be notified of the following decisions:

Section IV-C-1-b-(1)-(a). Changing containment levels for types of experiments that are specified in the *NIH Guidelines* when a *Major Action* is involved;

Section IV-C-1-b-(1)-(b). Assigning containment levels for types of experiments that are not explicitly considered in the *NIH Guidelines* when a *Major Action* is involved;

Section IV-C-1-b-(1)-(c). Promulgating and amending a list of classes of recombinant or synthetic nucleic acid molecules to be exempt from the *NIH Guidelines* because they consist entirely of DNA segments from species

that exchange DNA by known physiological processes or otherwise do not present a significant risk to health or the environment;

Section IV-C-1-b-(1)-(d). Permitting experiments specified by Section III-A, *Experiments that Require Institutional Biosafety Committee Approval, RAC Review, and NIH Director Approval Before Initiation*;

Section IV-C-1-b-(1)-(e). Certifying new host-vector systems with the exception of minor modifications of already certified systems (the standards and procedures for certification are described in Appendix I-II, *Certification of Host-Vector Systems*). Minor modifications constitute (e.g., those of minimal or no consequence to the properties relevant to containment); and

Section IV-C-1-b-(1)-(f). Adopting other changes in the *NIH Guidelines*.

Section IV-C-1-b-(2). Minor Actions

NIH OSP shall carry out certain functions as delegated to it by the NIH Director (see Section IV-C-3, *Office of Science Policy*). *Minor Actions* (as determined by NIH OSP in consultation with the RAC Chair and one or more RAC members, as necessary) will be transmitted to RAC and Institutional Biosafety Committee Chairs:

Section IV-C-1-b-(2)-(a). Changing containment levels for experiments that are specified in Section III, *Experiments Covered by the NIH Guidelines* (except when a *Major Action* is involved);

Section IV-C-1-b-(2)-(b). Assigning containment levels for experiments not explicitly considered in the *NIH Guidelines*;

Section IV-C-1-b-(2)-(c). Revising the *Classification of Etiologic Agents* for the purpose of these *NIH Guidelines* (see Section V-A, *Footnotes and References of Sections I-IV*).

Section IV-C-1-b-(2)-(d). Interpreting the *NIH Guidelines* for experiments to which the *NIH Guidelines* do not specifically assign containment levels;

Section IV-C-1-b-(2)-(e). Setting containment under Sections III-D-1-d, *Experiments Using Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents as Host-Vector Systems*, and III-D-2-b, *Experiments in which DNA from Risk Group 2, Risk Group 3, Risk Group 4, or Restricted Agents is Cloned into Nonpathogenic Prokaryotic or Lower Eukaryotic Host-Vector Systems*;

Section IV-C-1-b-(2)-(f). Approving minor modifications of already certified host-vector systems (the standards and procedures for such modifications are described in Appendix I-II, *Certification of Host-Vector Systems*);

Section IV-C-1-b-(2)-(g). Decertifying already certified host-vector systems;

Section IV-C-1-b-(2)-(h). Adding new entries to the list of molecules toxic for vertebrates (see Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*); and

Section IV-C-1-b-(2)-(i). Determining appropriate containment conditions for experiments according to case precedents developed under Section IV-C-1-b-(2)-(c).

Section IV-C-2. Recombinant DNA Advisory Committee (RAC)

The RAC is responsible for carrying out the functions specified in the *NIH Guidelines*, as well as others specified in its charter or assigned by the Secretary of Health and Human Services or the NIH Director. The RAC membership and procedures, in addition to those set forth in the *NIH Guidelines*, are specified in the charter for the RAC which is filed as provided in the General Services Administration Federal Advisory Committee Management regulations, 41 CFR part 101-6, and is available on the OSP web site, http://www.osp.od.nih.gov/sites/default/files/resources/RAC_2015-2017_Charter_Updated.pdf. In the event of a conflict between the *NIH Guidelines* and the charter, the charter shall control.

The RAC will consist of not less than 15 voting members, including the Chair, appointed under the procedures of the NIH and the Department of Health and Human Services. The maximum number of voting members will

be established in the charter of the RAC. At least a majority of the voting members must be knowledgeable in relevant scientific fields, e.g., molecular genetics, molecular biology, recombinant or synthetic nucleic acid molecule research, including clinical gene transfer research. At least 4 members of the RAC must be knowledgeable in fields such as public health, laboratory safety, occupational health, protection of human subjects of research, the environment, ethics, law, public attitudes or related fields. Representatives of the Federal agencies listed in the charter shall serve as non-voting members. Nominations for RAC members may be submitted to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

All meetings of the RAC shall be announced in the *Federal Register*, including tentative agenda items, 15 days before the meeting. Final agendas, if modified, shall be available at least 72 hours before the meeting. No item defined as a *Major Action* under Section IV-C-1-b-(1) may be added to an agenda following *Federal Register* publication.

RAC shall be responsible for:

Section IV-C-2-a. Advising the NIH Director on the following actions: (1) Adopting changes in the *NIH Guidelines*. (2) Assigning containment levels, changing containment levels, and approving experiments considered as *Major Actions* under the *NIH Guidelines*, i.e., the deliberate transfer of a drug resistance trait to microorganisms that are not known to acquire the trait naturally, if such acquisition could compromise the use of the drug to control disease agents in humans, veterinary medicine, or agriculture. (3) Promulgating and amending lists of classes of recombinant or synthetic nucleic acid molecules to be exempt from the *NIH Guidelines* because they consist entirely of DNA segments from species that exchange DNA by known physiological processes or otherwise do not present a significant risk to health or the environment. (4) Certifying new host-vector systems.

Section IV-C-2-b. Transmitting to the NIH Director specific comments/ recommendations about: (i) a specific human gene transfer experiment, or (ii) a category of human gene transfer experiments;

Section IV-C-2-c. Publicly reviewing human gene transfer clinical trial data and relevant information evaluated and summarized by NIH OSP in accordance with the annual data reporting requirements;

Section IV-C-2-d. Identifying broad scientific, safety, social, and ethical issues relevant to gene therapy research as potential Gene Therapy Policy Conference topics;

Section IV-C-2-e. Identifying novel social and ethical issues relevant to human applications of gene transfer and recommending appropriate guidance in the preparation of Informed Consent documents; and

Section IV-C-2-f. Identifying novel scientific and safety issues relevant to human applications of gene transfer and recommending appropriate guidance in the design of human gene transfer clinical trials.

Section IV-C-3. Office of Science Policy (OSP)

OSP shall serve as a focal point for information on recombinant or synthetic nucleic acid molecule activities and provide advice to all within and outside NIH including institutions, Biological Safety Officers, Principal Investigators, Federal agencies, state and local governments, and institutions in the private sector. OSP shall carry out such other functions as may be delegated to it by the NIH Director. OSP's responsibilities include (but are not limited to) the following:

Section IV-C-3-a. Serving as the focal point for public access to summary information pertaining to human gene transfer experiments;

Section IV-C-3-b. Serving as the focal point for data management of human gene transfer experiments;

Section IV-C-3-c. Administering the annual data reporting requirements (and subsequent review) for human gene transfer experiments (see Appendix M-I-C, *Reporting Requirements*);

Section IV-C-3-d. Transmitting comments/recommendations arising from public RAC discussion of a novel human gene transfer experiment to the NIH Director. RAC recommendations shall be forwarded to the Principal

Investigator, the sponsoring institution, and other DHHS components, as appropriate.

Section IV-C-3-e. Collaborating with Principal Investigators, Institutional Biosafety Committees, Institutional Review Boards, and other DHHS components (including FDA and the Office for Human Research Protections), to ensure human gene transfer experiment registration compliance in accordance with Appendix M-I, *Requirements for Protocol Submission, Review, and Reporting-Human Gene Transfer Experiments* of the *NIH Guidelines*.

Section IV-C-3-f. Administering Gene Therapy Policy Conferences as deemed appropriate by the NIH Director (see Appendix L, *Gene Therapy Policy Conferences*).

Section IV-C-3-g. Reviewing and approving experiments in conjunction with *ad hoc* experts involving the cloning of genes encoding for toxin molecules that are lethal for vertebrates at an LD_{50} of less than or equal to 100 nanograms per kilogram body weight in organisms other than *Escherichia coli* K-12 (see Section III-B-1, *Experiments Involving the Cloning of Toxin Molecules with LD_{50} of Less than 100 Nanograms Per Kilogram Body Weight*, Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*);

Section IV-C-3-h. Serving as the executive secretary of RAC;

Section IV-C-3-i. Publishing in the *Federal Register*:

Section IV-C-3-i-(1). Announcements of RAC meetings and tentative agendas at least 15 days in advance (Note: If the agenda for a RAC meeting is modified, OSP shall make the revised agenda available to anyone upon request in advance of the meeting);

Section IV-C-3-i-(2). Announcements of Gene Therapy Policy Conferences and tentative agendas at least 15 days in advance;

Section IV-C-3-i-(3). Proposed *Major Actions* (see Section IV-C-1-b-(1), *Major Actions*) at least 15 days prior to the RAC meeting; and

Section IV-C-3-j. Reviewing and approving the membership of an institution's Institutional Biosafety Committee, and where it finds the Institutional Biosafety Committee meets the requirements set forth in Section IV-B-2, *Institutional Biosafety Committee (IBC)*, giving its approval to the Institutional Biosafety Committee membership.

Section IV-C-4. Other NIH Components

Other NIH components shall be responsible for certifying maximum containment (BL4) facilities, inspecting them periodically, and inspecting other recombinant or synthetic nucleic acid molecule facilities as deemed necessary.

Section IV-D. Voluntary Compliance

Section IV-D-1. Basic Policy - Voluntary Compliance

Individuals, corporations, and institutions not otherwise covered by the *NIH Guidelines* are encouraged to follow the standards and procedures set forth in Sections I through IV. In order to simplify discussion, references hereafter to “institutions” are intended to encompass corporations and individuals who have no organizational affiliation. For purposes of complying with the *NIH Guidelines*, an individual intending to carry out research involving recombinant or synthetic nucleic acid molecules is encouraged to affiliate with an institution that has an Institutional Biosafety Committee approved under the *NIH Guidelines*.

Since commercial organizations have special concerns, such as protection of proprietary data, some modifications and explanations of the procedures are provided in Sections IV-D-2 through IV-D-5-b, *Voluntary Compliance*, in order to address these concerns.

Section IV-D-2. Institutional Biosafety Committee Approval - Voluntary Compliance

It should be emphasized that employment of an Institutional Biosafety Committee member solely for purposes of membership on the Institutional Biosafety Committee does not itself make the member an institutionally affiliated member. Except for the unaffiliated members, a member of an Institutional Biosafety Committee for an institution not otherwise covered by the *NIH Guidelines* may participate in the review and approval of a project in which the member has a direct financial interest so long as the member has not been, and does not expect to be, engaged in the project. Section IV-B-2-a-(4), *Institutional Biosafety Committee*, is modified to that extent for purposes of these institutions.

Section IV-D-3. Certification of Host-Vector Systems - Voluntary Compliance

A host-vector system may be proposed for certification by the NIH Director in accordance with the procedures set forth in Appendix I-II, *Certification of Host-Vector Systems*. In order to ensure protection for proprietary data, any public notice regarding a host-vector system which is designated by the institution as proprietary under Section IV-D, *Voluntary Compliance*, will be issued only after consultation with the institution as to the content of the notice.

Section IV-D-4. Requests for Exemptions and Approvals - Voluntary Compliance

Requests for exemptions or other approvals as required by the *NIH Guidelines* should be submitted based on the procedures set forth in Sections I through IV. In order to ensure protection for proprietary data, any public notice regarding a request for an exemption or other approval which is designated by the institution as proprietary under Section IV-D-5-a, *Voluntary Compliance*, will be issued only after consultation with the institution as to the content of the notice.

Section IV-D-5. Protection of Proprietary Data - Voluntary Compliance

Section IV-D-5-a. General

In general, the Freedom of Information Act requires Federal agencies to make their records available to the public upon request. However, this requirement does not apply to, among other things, “trade secrets and commercial or financial information that is obtained from a person and that is privileged or confidential.” Under 18 U.S.C. 1905, it is a criminal offense for an officer or employee of the U.S. or any Federal department or agency to publish, divulge, disclose, or make known “in any manner or to any extent not authorized by law any information coming to him in the course of his employment or official duties or by reason of any examination or investigation made by, or return, report or record made to or filed with, such department or agency or officer or employee thereof, which information concerns or relates to the trade secrets, (or) processes...of any person, firm, partnership, corporation, or association.” This provision applies to all employees of the Federal Government, including special Government employees. Members of RAC are “special Government employees.”

In submitting to NIH for purposes of voluntary compliance with the *NIH Guidelines*, an institution may designate those items of information which the institution believes constitute trade secrets, privileged, confidential, commercial, or financial information. If NIH receives a request under the Freedom of Information Act for information so designated, NIH will promptly contact the institution to secure its views as to whether the information (or some portion) should be released. If NIH decides to release this information (or some portion) in response to a Freedom of Information request or otherwise, the institution will be advised and the actual release will be delayed in accordance with 45 Code of Federal Regulations, Section 5.65(d) and (e).

Section IV-D-5-b. Pre-submission Review

Any institution not otherwise covered by the *NIH Guidelines*, which is considering submission of data or information voluntarily to NIH, may request pre-submission review of the records involved to determine if NIH will make all or part of the records available upon request under the Freedom of Information Act.

A request for pre-submission review should be submitted to NIH OSP along with the records involved to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: HGTprotocols@mail.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). These records

shall be clearly marked as being the property of the institution on loan to NIH solely for the purpose of making a determination under the Freedom of Information Act. NIH OSP will seek a determination from the responsible official under DHHS regulations (45 CFR Part 5) as to whether the records involved, (or some portion) will be made available to members of the public under the Freedom of Information Act. Pending such a determination, the records will be kept separate from NIH OSP files, will be considered records of the institution and not NIH OSP, and will not be received as part of NIH OSP files. No copies will be made of such records.

NIH OSP will inform the institution of the NIH Freedom of Information Officer's determination and follow the institution's instructions as to whether some or all of the records involved are to be returned to the institution or to become a part of NIH OSP files. If the institution instructs NIH OSP to return the records, no copies or summaries of the records will be made or retained by DHHS, NIH, or OSP. The NIH Freedom of Information Officer's determination will represent that official's judgment at the time of the determination as to whether the records involved (or some portion) would be exempt from disclosure under the Freedom of Information Act if at the time of the determination the records were in NIH OSP files and a request was received for such files under the Freedom of Information Act.

****

SECTION V. FOOTNOTES AND REFERENCES OF SECTIONS I THROUGH IV

Section V-A. The NIH Director, with advice of the RAC, may revise the classification for the purposes of the *NIH Guidelines* (see Section IV-C-1-b-(2)-(e), *Minor Actions*). The revised list of organisms in each Risk Group is reprinted in Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*.

Section V-B. Section III, *Experiments Covered by the NIH Guidelines*, describes a number of places where judgments are to be made. In all these cases, the Principal Investigator shall make the judgment on these matters as part of his/her responsibility to "make the initial determination of the required levels of physical and biological containment in accordance with the *NIH Guidelines*" (see Section IV-B-7-c-(1)). For cases falling under Sections III-A through III-E, *Experiments Covered by the NIH Guidelines*, this judgment is to be reviewed and approved by the Institutional Biosafety Committee as part of its responsibility to make an "independent assessment of the containment levels required by the *NIH Guidelines* for the proposed research" (see Section IV-B-2-b-(1), *Institutional Biosafety Committee*). The Institutional Biosafety Committee may refer specific cases to NIH OSP as part of NIH OSP's functions to "provide advice to all within and outside NIH" (see Section IV-C-3). NIH OSP may request advice from the RAC as part of the RAC's responsibility for "interpreting the *NIH Guidelines* for experiments to which the *NIH Guidelines* do not specifically assign containment levels" (see Section IV-C-1-b-(2)-(f), *Minor Actions*).

Section V-C. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention and the National Institutes of Health. *Biosafety in Microbiological and Biomedical Laboratories*, 5th Edition, 2007. Copies are available from: Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20401-0001, Phone (202) 512-1800 [http://www.gpo.gov/].

Section V-D. *Classification of Etiologic Agents on the Basis of Hazard*, 4th Edition, July 1974, U.S. Department of Health, Education, and Welfare, Public Health Service, Centers for Disease Control, Office of Biosafety, Atlanta, Georgia 30333.

Section V-E. Chin, James ed., *Control of Communicable Diseases Manual*, 17th Edition, 2000. ISBN: 087553-242-X, American Public Health Association, 800 I Street, N.W., Washington, D.C. Phone: (202) 777-2742.

Section V-F. *World Health Organization Laboratory Biosafety Manual*, 2nd edition. 1993. WHO Albany, NY. Copies are available from: WHO Publication Centre, USA, (Q Corp) 49 Sheridan Avenue, Albany, New York 12210; Phone: (518) 436-9686 (Order # 1152213).

Section V-G. A U.S. Department of Agriculture permit, required for import and interstate transport of plant and animal pathogens, may be obtained from the U.S. Department of Agriculture, ATTN: Animal and Plant Health Inspection Service (APHIS), Veterinary Services, National Center for Import-Export, Products Program, 4700 River Road, Unit 40, Riverdale, Maryland 20737. Phone: (301) 734-8499; Fax: (301) 734-8226.

Section V-H. American Type Culture Collection Catalogues of plant viruses, animal viruses, cells, bacteria,

fungi, etc. are available from American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209. Phone: (703) 365-2700.

Section V-I. U.S. Department of Labor, Occupational Safety and Health Administration, 29 CFR 1910.1030, *Bloodborne Pathogens*. See also*, Exposure to Bloodborne Pathogens,* OSHA 3127, 1996 (Revised).

Section V-J. As classified in the *Virus Taxonomy: The Classification and Nomenclature of Viruses. The Seventh Report of the International Committee on Taxonomy of Viruses*, Academic Press, 2000 (0123702003) San Diego, CA.

Section V-K. *i.e.*, the total of all genomes within a family shall not exceed two-thirds of the genome.

Section V-L. Organisms including alastrim, smallpox (variola) and whitepox may not be studied in the United States except at specified facilities. All activities, including storage of variola and whitepox, are restricted to the single national facility (World Health Organization Collaborating Center for Smallpox Research, Centers for Disease Control and Prevention, Atlanta, Georgia).

Section V-M. In accordance with accepted scientific and regulatory practices of the discipline of plant pathology, an exotic plant pathogen (e.g., virus, bacteria, or fungus) is one that is unknown to occur within the U.S. (see Section V-G, *Footnotes and References of Sections I-IV*). Determination of whether a pathogen has a potential for serious detrimental impact on managed (agricultural, forest, grassland) or natural ecosystems should be made by the Principal Investigator and the Institutional Biosafety Committee, in consultation with scientists knowledgeable of plant diseases, crops, and ecosystems in the geographic area of the research.

****

APPENDIX A. EXEMPTIONS UNDER SECTION III-F-6--SUBLISTS OF NATURAL EXCHANGERS

Certain specified recombinant or synthetic nucleic acid molecules that consist entirely of DNA segments from different species that exchange DNA by known physiological processes, though one or more of the segments may be a synthetic equivalent are exempt from these *NIH Guidelines* (see Section III-F-6, *Exempt Experiments*). Institutional Biosafety Committee registration is not required for these exempt experiments. A list of such exchangers will be prepared and periodically revised by the NIH Director with advice from the RAC after appropriate notice and opportunity for public comment (see Section IV-C-1-b-(1)-(c), *NIH Director--Specific Responsibilities*). For a list of natural exchangers that are exempt from the *NIH Guidelines*, see Appendices A-I through A-VI, *Exemptions under Section III-F-6 Sublists of Natural Exchangers*. Section III-F-6, *Exempt Experiments*, describes recombinant or synthetic nucleic acid molecules that are: (1) composed entirely of DNA segments from one or more of the organisms within a sublist, and (2) to be propagated in any of the organisms within a sublist (see *Classification of Bergey's Manual of Determinative Bacteriology*; 8th edition, R. E. Buchanan and N. E. Gibbons, editors, Williams and Wilkins Company; Baltimore, Maryland 1984). Although these experiments are exempt, it is recommended that they be performed at the appropriate biosafety level for the host or recombinant/synthetic organism (see *Biosafety in Microbiological and Biomedical Laboratories*, 5th edition, 2007, U.S. DHHS, Public Health Service, Centers for Disease Control and Prevention, Atlanta, Georgia, and NIH Office of Biosafety, Bethesda, Maryland).

Appendix A-I. Sublist A

Genus Escherichia
Genus *Shigella*
Genus *Salmonella* - including *Arizona*
Genus *Enterobacter*
Genus *Citrobacter* - including *Levinea*
Genus *Klebsiella* - including *oxytoca*
Genus *Erwinia*
Pseudomonas aeruginosa, Pseudomonas putida, Pseudomonas fluorescens, and Pseudomonas mendocina
Serratia marcescens
Yersinia enterocolitica

Appendix A-II. Sublist B

Bacillus subtilis
Bacillus licheniformis
Bacillus pumilus
Bacillus globigii
Bacillus niger
Bacillus nato
Bacillus amyloliquefaciens
Bacillus aterrimus

Appendix A-III. Sublist C

Streptomyces aureofaciens
Streptomyces rimosus
Streptomyces coelicolor

Appendix A-IV. Sublist D

Streptomyces griseus
Streptomyces cyaneus
Streptomyces venezuelae

Appendix A-V. Sublist E

One way transfer of *Streptococcus mutans* or *Streptococcus lactis* DNA into *Streptococcus sanguis*

Appendix A-VI. Sublist F

Streptococcus sanguis
Streptococcus pneumoniae
Streptococcus faecalis
Streptococcus pyogenes
Streptococcus mutans

****

APPENDIX B. CLASSIFICATION OF HUMAN ETIOLOGIC AGENTS ON THE BASIS OF HAZARD

This appendix includes those biological agents known to infect humans as well as selected animal agents that may pose theoretical risks if inoculated into humans. Included are lists of representative genera and species known to be pathogenic; mutated, recombined, and non-pathogenic species and strains are not considered. Non-infectious life cycle stages of parasites are excluded.

This appendix reflects the current state of knowledge and should be considered a resource document. Included are the more commonly encountered agents and is not meant to be all-inclusive. Information on agent risk assessment may be found in the *Agent Summary Statements* of the CDC/NIH publication, *Biosafety in Microbiological and Biomedical Laboratories* (see Sections V-C, V-D, V-E, and V-F, *Footnotes and References of Sections I through IV*. Further guidance on agents not listed in Appendix B may be obtained through: Centers for Disease Control and Prevention, Biosafety Branch, Atlanta, Georgia 30333, Phone: (404) 639-3883, Fax: (404) 639-2294; National Institutes of Health, Division of Safety, Bethesda, Maryland 20892, Phone: (301) 496-1357; Biosafety Manager, National Animal Disease Center, U.S. Department of Agriculture - ARS, Ames, Iowa 50010, Phone: (515) 337-7772.

A special committee of the American Society for Microbiology will conduct an annual review of this appendix and its recommendation for changes will be presented to the Recombinant DNA Advisory Committee as proposed amendments to the *NIH Guidelines*.

Appendix B - Table 1. Basis for the Classification of Biohazardous Agents by Risk Group (RG)

Risk Group 1 (RG1)	Agents that are not associated with disease in healthy adult humans
Risk Group 2 (RG2)	Agents that are associated with human disease which is rarely serious and for which preventive or therapeutic interventions are *often* available
Risk Group 3 (RG3)	Agents that are associated with serious or lethal human disease for which preventive or therapeutic interventions *may be* available (high individual risk but low community risk)
Risk Group 4 (RG4)	Agents that are likely to cause serious or lethal human disease for which preventive or therapeutic interventions are *not usually* available (high individual risk and high community risk)

Appendix B-I. Risk Group 1 (RG1) Agents

RG1 agents are not associated with disease in healthy adult humans. Examples of RG1 agents include asporogenic *Bacillus subtilis* or *Bacillus licheniformis* (see Appendix C-IV-A, *Bacillus subtilis* or *Bacillus licheniformis* Host-Vector Systems, Exceptions); adeno- associated virus (AAV – all serotypes); and recombinant or synthetic AAV constructs, in which the transgene does not encode either a potentially tumorigenic gene product or a toxin molecule and are produced in the absence of a helper virus. A strain of *Escherichia coli* (see Appendix C-II-A, *Escherichia coli* K-12 Host Vector Systems, Exceptions) is an RG1 agent if it (1) does not possess a complete lipopolysaccharide (*i.e.*, lacks the O antigen); and (2) does not carry any active virulence factor (*e.g.*, toxins) or colonization factors and does not carry any genes encoding these factors.

Those agents not listed in Risk Groups (RGs) 2, 3 and 4 are not automatically or implicitly classified in RG1; a risk assessment must be conducted based on the known and potential properties of the agents and their relationship to agents that are listed.

Appendix B-II. Risk Group 2 (RG2) Agents

RG2 agents are associated with human disease which is rarely serious and for which preventive or therapeutic interventions are *often* available.

Appendix B-II-A. Risk Group 2 (RG2) - Bacterial Agents Including Chlamydia

--*Acinetobacter baumannii* (formerly *Acinetobacter calcoaceticus*)
--*Actinobacillus*
--*Actinomyces pyogenes* (formerly *Corynebacterium pyogenes*)
--*Aeromonas hydrophila*
--*Amycolata autotrophica*
--*Archanobacterium haemolyticum* (formerly *Corynebacterium haemolyticum*)
--*Arizona hinshawii* - all serotypes
--*Bacillus anthracis*
--*Bartonella henselae, B. quintana, B. vinsonii*
--*Bordetella* including *B. pertussis*
--*Borrelia recurrentis, B. burgdorferi*
--*Burkholderia* (formerly *Pseudomonas* species) except those listed in Appendix B-III-A (RG3))
--*Campylobacter coli, C. fetus, C. jejuni*
--*Chlamydia psittaci*, *C. trachomatis*, *C. pneumoniae*
--*Clostridium botulinum*, *C. chauvoei, C. haemolyticum, C. histolyticum, C. novyi, C. septicum, C. tetani*
--*Coxiella burnetii* – specifically the Phase II, Nine Mile strain, plaque purified, clone 4
--*Corynebacterium diphtheriae, C. pseudotuberculosis, C. renale*
--*Dermatophilus congolensis*
--*Edwardsiella tarda*
--*Erysipelothrix rhusiopathiae*

--*Escherichia coli* - all enteropathogenic, enterotoxigenic, enteroinvasive and strains bearing K1 antigen, including *E. coli* O157:H7
--**Francisella tularensis* specifically **F. tularensis* subspecies *novicida* [aka *F. novicida*], strain Utah 112; **F. tularensis* subspecies *holarctica* LVS; **F. tularensis* biovar *tularensis* strain ATCC 6223 (aka strain B38)
*For research involving high concentrations, BL3 practices should be considered (see Appendix G-II-C-2. Special Practices (BL3)).
--*Haemophilus ducreyi, H. influenzae*
--*Helicobacter pylori*
--*Klebsiella* - all species except *K. oxytoca* (RG1)
--*Legionella* including *L. pneumophila*
--*Leptospira interrogans* - all serotypes
--*Listeria*
--*Moraxella*
--*Mycobacterium* (except those listed in Appendix B-III-A (RG3)) including *M. avium* complex, *M. asiaticum, M. bovis* BCG vaccine strain, *M. chelonae, M. fortuitum, M. kansasii, M. leprae, M. malmoense, M. marinum, M. paratuberculosis, M. scrofulaceum, M. simiae, M. szulgai, M. ulcerans, M. xenopi*
--*Mycoplasma*, except *M. mycoides* and *M. agalactiae* which are restricted animal pathogens
--*Neisseria gonorrhoeae*, *N. meningitidis*
--*Nocardia asteroides, N. brasiliensis, N. otitidiscaviarum, N. transvalensis*
--*Pseudomonas aeruginosa*
--*Rhodococcus equi*
--*Salmonella* including *S. arizonae, S. choleraesuis, S. enteritidis, S. gallinarum-pullorum, S. meleagridis, S. paratyphi,* A, B, C, *S. typhi*, *S. typhimurium*
--*Shigella* including *S. boydii, S. dysenteriae,* type 1, *S. flexneri, S. sonnei*
--*Sphaerophorus necrophorus*
--*Staphylococcus aureus*
--*Streptobacillus moniliformis*

--*Streptococcus* including *S. pneumoniae, S. pyogenes*
--*Treponema pallidum, T. carateum*
--*Vibrio cholerae, V. parahaemolyticus, V. vulnificus*
--*Yersinia enterocolitica*
--*Yersinia pestis* specifically *pgm*$^{(-)}$ strains (lacking the 102 kb pigmentation locus) and *lcr*$^{(-)}$ strains (lacking the LCR plasmid)

Appendix B-II-B. Risk Group 2 (RG2) - Fungal Agents

--*Blastomyces dermatitidis*
--*Cladosporium bantianum, C. (Xylohypha) trichoides*
--*Cryptococcus neoformans*
--*Dactylaria galopava (Ochroconis gallopavum)*
--*Epidermophyton*
--*Exophiala (Wangiella) dermatitidis*
--*Fonsecaea pedrosoi*
--*Microsporum*
--*Paracoccidioides braziliensis*
--*Penicillium marneffei*
--*Sporothrix schenckii*
--*Trichophyton*

Appendix B-II-C. Risk Group 2 (RG2) - Parasitic Agents

--*Ancylostoma* human hookworms including *A. duodenale, A. ceylanicum*
--*Ascaris* including *Ascaris lumbricoides suum*
--*Babesia* including *B. divergens, B. microti*
--*Brugia* filaria worms including *B. malayi, B. timori*
--*Coccidia*
--*Cryptosporidium* including *C. parvum*
--*Cysticercus cellulosae* (hydatid cyst, larva of *T. solium*)
--*Echinococcus* including *E. granulosis, E. multilocularis, E. vogeli*

--*Entamoeba histolytica*
--*Enterobius*
--*Fasciola* including *F. gigantica, F. hepatica*
--*Giardia* including *G. lamblia*
--*Heterophyes*
--*Hymenolepis* including *H. diminuta, H. nana*
--*Isospora*
--*Leishmania* including *L. braziliensis, L. donovani, L. ethiopia, L. major, L. mexicana, L. peruviana, L. tropica*
--*Loa loa* filaria worms
--*Microsporidium*
--*Naegleria fowleri*
--*Necator* human hookworms including *N. americanus*
--*Onchocerca* filaria worms including, *O. volvulus*
--*Plasmodium* including simian species, *P. cynomolgi, P. falciparum, P. malariae, P. ovale, P. vivax*
--*Sarcocystis* including *S. sui hominis*
--*Schistosoma* including *S. haematobium, S. intercalatum, S. japonicum, S. mansoni, S. mekongi*
--*Strongyloides* including *S. stercoralis*
--*Taenia solium*
--*Toxocara* including *T. canis*
--*Toxoplasma* including *T. gondii*
--*Trichinella spiralis*
--*Trypanosoma* including *T. brucei brucei, T. brucei gambiense, T. brucei rhodesiense, T. cruzi*
--*Wuchereria bancrofti* filaria worms

Appendix B-II-D. Risk Group 2 (RG2) - Viruses

Adenoviruses, human - all types

Alphaviruses (Togaviruses) - Group A Arboviruses
--Chikungunya vaccine strain 181/25
--Eastern equine encephalomyelitis virus
--Venezuelan equine encephalomyelitis vaccine strains TC-83 and V3526
--Western equine encephalomyelitis virus

Arenaviruses
--Junin virus candid #1 vaccine strain
--Lymphocytic choriomeningitis virus (non-neurotropic strains)
--Tacaribe virus complex
--Other viruses as listed in the reference source (see Section V-C, *Footnotes and References of Sections I through IV*)

Bunyaviruses
--Bunyamwera virus
--Rift Valley fever virus vaccine strain MP-12
--Other viruses as listed in the reference source (see Section V-C, *Footnotes and References of Sections I through IV*)

Caliciviruses

Coronaviruses

Flaviviruses - Group B Arboviruses
--Dengue virus serotypes 1, 2, 3, and 4
--Japanese encephalitis virus strain SA 14-14-2
--Yellow fever virus vaccine strain 17D
--Other viruses as listed in the reference source (see Section V-C, *Footnotes and References of Sections I through IV*)

Hepatitis A, B, C, D, and E viruses

Herpesviruses - except Herpesvirus simiae (Monkey B virus) (see Appendix B-IV-D, *Risk Group 4 (RG4) - Viral Agents*)
--Cytomegalovirus
--Epstein Barr virus
--*Herpes simplex* types 1 and 2
--*Herpes zoster*
--Human herpesvirus types 6 and 7

Orthomyxoviruses
--Influenza viruses types A, B, and C (except those listed in Appendix B-III-D, *Risk Group 3 (RG3) - Viruses and Prions*)
--Tick-borne orthomyxoviruses

Papilloma viruses
--All human papilloma viruses

Paramyxoviruses
--Newcastle disease virus
--Measles virus
--Mumps virus
--Parainfluenza viruses types 1, 2, 3, and 4
--Respiratory syncytial virus

Parvoviruses
--Human parvovirus (B19)

Picornaviruses
--Coxsackie viruses types A and B
--Echoviruses - all types
--Polioviruses - all types, wild and attenuated
--Rhinoviruses - all types

Poxviruses - all types except Monkeypox virus (see Appendix B-III-D, *Risk Group 3 (RG3) - Viruses and Prions*) and restricted poxviruses including Alastrim, Smallpox, and Whitepox (see Section V-L, *Footnotes and References of Sections I through IV*)

Reoviruses - all types including Coltivirus, human Rotavirus, and Orbivirus (Colorado tick fever virus)

Rhabdoviruses
--Rabies virus - all strains
--Vesicular stomatitis virus non exotic strains: VSV-Indiana 1 serotype strains (*e.g.* Glasgow, Mudd-Summers, Orsay, San Juan) and VSV-New Jersey serotype strains (*e.g.* Ogden, Hazelhurst)

Rubivirus (Togaviruses)
--Rubella virus

Appendix B-III. Risk Group 3 (RG3) Agents

RG3 agents are associated with serious or lethal human disease for which preventive or therapeutic interventions *may be* available.

Appendix B-III-A. Risk Group 3 (RG3) - Bacterial Agents Including Rickettsia

--*Bartonella*
--*Brucella* including *B. abortus, B. canis*, *B. suis*
--*Burkholderia (Pseudomonas) mallei, B. pseudomallei*
--*Coxiella burnetii (*except the Phase II, Nine Mile strain listed in Appendix B-II-A, *Risk Group 2 (RG2) - Bacterial Agents Including Chlamydia)*
--*Francisella tularensis (*except those strains listed in Appendix B-II-A, *Risk Group 2 (RG2) - Bacterial Agents*

Including Chlamydia)
--*Mycobacterium bovis* (except BCG strain, see Appendix B-II-A, *Risk Group 2 (RG2) - Bacterial Agents Including Chlamydia*), *M. tuberculosis*
--*Orientia tsutsugamushi* (was *R. tsutsugamushi*)
--*Pasteurella multocida* type B -"buffalo" and other virulent strains
--*Rickettsia akari*, *R. australis*, *R. canada*, *R. conorii*, *R. prowazekii*, *R. rickettsii*, *R, siberica*, *R. typhi* (*R. mooseri*)
--*Yersinia pestis (*except those strains listed in Appendix B-II-A, *Risk Group 2 (RG2) - Bacterial Agents Including Chlamydia)*

Appendix B-III-B. Risk Group 3 (RG3) - Fungal Agents

--*Coccidioides immitis* (sporulating cultures; contaminated soil)
--*Histoplasma capsulatum, H. capsulatum var. duboisii*

Appendix B-III-C. Risk Group 3 (RG3) - Parasitic Agents

None

Appendix B-III-D. Risk Group 3 (RG3) - Viruses and Prions

Alphaviruses (Togaviruses) - Group A Arboviruses
--Chikungunya virus (except the vaccine strain 181/25 listed in Appendix B-II-D Risk Group2 (RG2) – Viruses)
--Semliki Forest virus
--St. Louis encephalitis virus
--Venezuelan equine encephalomyelitis virus (except the vaccine strains TC-83 and V3526, see Appendix B-II-D (RG2) – Viruses)
--Other viruses as listed in the reference source (see Section V-C, *Footnotes and References of Sections I through IV*)

Arenaviruses
--Flexal
--Lymphocytic choriomeningitis virus (LCM) (neurotropic strains)

Bunyaviruses
--Hantaviruses including Hantaan virus
--Rift Valley fever virus

Coronaviruses
--SARS-associated coronavirus (SARS-CoV)
--Middle East respiratory syndrome coronavirus (MERS-CoV)

Flaviviruses - Group B Arboviruses
--Japanese encephalitis virus (except those strains listed in Appendix B-II-D Risk Group2 (RG2) - Viruses)
--West Nile virus (WNV)
--Yellow fever virus
--Other viruses as listed in the reference source (see Section V-C, *Footnotes and References of Sections I through IV*)

Orthomyxoviruses
-- Influenza viruses 1918-1919 H1N1 (1918 H1N1), human H2N2 (1957-1968), and highly pathogenic avian influenza H5N1 strains within the Goose/Guangdong/96-like H5 lineage (HPAI H5N1).

Poxviruses
--Monkeypox virus

Prions
--Transmissible spongiform encephalopathies (TSE) agents (Creutzfeldt-Jacob disease and kuru agents)(see Section V-C, *Footnotes and References of Sections I through IV*, for containment instruction)

Retroviruses
--Human immunodeficiency virus (HIV) types 1 and 2
--Human T cell lymphotropic virus (HTLV) types 1 and 2
--Simian immunodeficiency virus (SIV)

Rhabdoviruses
--Vesicular stomatitis virus (except those strains listed in Appendix B-II-D Risk Group2 (RG2) - Viruses)

Appendix B-IV. Risk Group 4 (RG4) Agents

RG4 agents are likely to cause serious or lethal human disease for which preventive or therapeutic interventions are *not usually* available.

Appendix B-IV-A. Risk Group 4 (RG4) - Bacterial Agents

None

Appendix B-IV-B. Risk Group 4 (RG4) - Fungal Agents

None

Appendix B-IV-C. Risk Group 4 (RG4) - Parasitic Agents

None

Appendix B-IV-D. Risk Group 4 (RG4) - Viral Agents

Arenaviruses
--Guanarito virus
--Lassa virus

--Junin virus (except the candid #1 vaccine strain listed in Appendix B-II-D Risk Group2 (RG2) – Viruses)
--Machupo virus
--Sabia

Bunyaviruses (Nairovirus)
--Crimean-Congo hemorrhagic fever virus

Filoviruses
--Ebola virus
--Marburg virus

Flaviruses - Group B Arboviruses
--Tick-borne encephalitis virus complex including Absetterov, Central European encephalitis, Hanzalova, Hypr, Kumlinge, Kyasanur Forest disease, Omsk hemorrhagic fever, and Russian spring-summer encephalitis viruses

Herpesviruses (alpha)
--Herpesvirus simiae (Herpes B or Monkey B virus)

Paramyxoviruses
--Equine Morbillivirus (Hendra virus)

Hemorrhagic fever agents and viruses as yet undefined

Appendix B-V. Animal Viral Etiologic Agents in Common Use

The following list of animal etiologic agents is appended to the list of human etiologic agents. None of these agents is associated with disease in healthy adult humans; they are commonly used in laboratory experimental work.

A containment level appropriate for RG1 human agents is recommended for their use. For agents that are infectious to human cells, e.g., amphotropic and xenotropic strains of murine leukemia virus, a containment level appropriate for RG2 human agents is recommended.

Baculoviruses

Herpesviruses
--Herpesvirus ateles
--Herpesvirus saimiri
--Marek's disease virus
--Murine cytomegalovirus

Papilloma viruses
--Bovine papilloma virus
--Shope papilloma virus

Polyoma viruses
--Polyoma virus
--Simian virus 40 (SV40)

Retroviruses
--Avian leukosis virus
--Avian sarcoma virus
--Bovine leukemia virus
--Feline leukemia virus
--Feline sarcoma virus
--Gibbon leukemia virus
--Mason-Pfizer monkey virus
--Mouse mammary tumor virus
--Murine leukemia virus

--Murine sarcoma virus
--Rat leukemia virus

Appendix B-V-1. Murine Retroviral Vectors

Murine retroviral vectors to be used for human transfer experiments (less than 10 liters) that contain less than 50% of their respective parental viral genome and that have been demonstrated to be free of detectable replication competent retrovirus can be maintained, handled, and administered, under BL1 containment.

****

APPENDIX C. EXEMPTIONS UNDER SECTION III-F-8

Section III-F-8 states that exempt from these *NIH Guidelines* are "those that do not present a significant risk to health or the environment (see Section IV-C-1-b-(1)-(c), *NIH Director--Specific Responsibilities*), as determined by the NIH Director, with the advice of the RAC, and following appropriate notice and opportunity for public comment. See Appendix C, *Exemptions under Sections III-F-8*, for other classes of experiments which are exempt from the *NIH Guidelines*." The following classes of experiments are exempt under Section III-F-8:

Appendix C-I. Recombinant or Synthetic Nucleic Acid Molecules in Tissue Culture

Recombinant or synthetic nucleic acid molecules containing less than one-half of any eukaryotic viral genome (all viruses from a single family being considered identical -- see Appendix C-IX-E, *Footnotes and References of Appendix C*), that are propagated and maintained in cells in tissue culture are exempt from these *NIH Guidelines* with the exceptions listed in Appendix C-I-A.

Appendix C-I-A. Exceptions

The following categories are not exempt from the *NIH Guidelines*: (i) experiments described in Section III-B which require NIH OSP and Institutional Biosafety Committee approval before initiation, (ii) experiments involving DNA from Risk Groups 3, 4, or restricted organisms (see Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, and Sections V-G and V-L, *Footnotes and References of Sections I through IV*) or cells known to be infected with these agents, (iii) experiments involving the deliberate introduction of genes coding for the biosynthesis of molecules that are toxic for vertebrates (see Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*), and (iv) whole plants regenerated from plant cells and tissue cultures are covered by the exemption provided they remain axenic cultures even though they differentiate into embryonic tissue and regenerate into plantlets.

Appendix C-II. *Escherichia coli K-12* Host-Vector Systems

Experiments which use *Escherichia coli* K-12 host-vector systems, with the exception of those experiments listed in Appendix C-II-A, are exempt from the *NIH Guidelines* provided that: (i) the *Escherichia coli* host does not contain conjugation proficient plasmids or generalized transducing phages; or (ii) lambda or lambdoid or Ff bacteriophages or non-conjugative plasmids (see Appendix C-IX-B, *Footnotes and References of Appendix C*) shall be used as vectors. However, experiments involving the insertion into *Escherichia coli* K-12 of DNA from prokaryotes that exchange genetic information (see Appendix C-IX-C, *Footnotes and References of Appendix C*) with *Escherichia coli* may be performed with any *Escherichia coli* K-12 vector (e.g., conjugative plasmid). When a non-conjugative vector is used, the *Escherichia coli* K-12 host may contain conjugation-proficient plasmids either autonomous or integrated, or generalized transducing phages. For these exempt laboratory experiments, Biosafety Level (BL) 1 physical containment conditions are recommended. For large-scale fermentation experiments, the appropriate physical containment conditions need be no greater than those for the host organism unmodified by recombinant or synthetic nucleic acid molecule techniques; the Institutional Biosafety Committee can specify higher containment if deemed necessary.

Appendix C-II-A. Exceptions

The following categories are not exempt from the *NIH Guidelines*: (i) experiments described in Section III-B which require NIH OSP and Institutional Biosafety Committee approval before initiation, (ii) experiments involving DNA from Risk Groups 3, 4, or restricted organisms (see Appendix B, *Classification of Human*

Etiologic Agents on the Basis of Hazard, and Sections V-G and V-L, *Footnotes and References of Sections I through IV*) or cells known to be infected with these agents may be conducted under containment conditions specified in Section III-D-2 with prior Institutional Biosafety Committee review and approval, (iii) large-scale experiments (e.g., more than 10 liters of culture), and (iv) experiments involving the cloning of toxin molecule genes coding for the biosynthesis of molecules toxic for vertebrates (see Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*).

Appendix C-III. *Saccharomyces* Host-Vector Systems

Experiments involving *Saccharomyces cerevisiae* and *Saccharomyces uvarum* host-vector systems, with the exception of experiments listed in Appendix C-III-A, are exempt from the *NIH Guidelines*. For these exempt experiments, BL1 physical containment is recommended. For large-scale fermentation experiments, the appropriate physical containment conditions need be no greater than those for the unmodified host organism; the Institutional Biosafety Committee can specify higher containment if deemed necessary.

Appendix C-III-A. Exceptions

The following categories are not exempt from the *NIH Guidelines*: (i) experiments described in Section III-B which require NIH OSP and Institutional Biosafety Committee approval before initiation, (ii) experiments involving DNA from Risk Groups 3, 4, or restricted organisms (see Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, and Sections V-G and V-L, *Footnotes and References of Sections I through IV*) or cells known to be infected with these agents may be conducted under containment conditions specified in Section III-D-2 with prior Institutional Biosafety Committee review and approval, (iii) large-scale experiments (e.g., more than 10 liters of culture), and (iv) experiments involving the deliberate cloning of genes coding for the biosynthesis of molecules toxic for vertebrates (see Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*).

Appendix C-IV. *Kluyveromyces* Host-Vector Systems

Experiments involving *Kluyveromyces lactis* host-vector systems, with the exception of experiments listed in Appendix C-IV-A, are exempt from the *NIH Guidelines* provided laboratory-adapted strains are used (i.e. strains that have been adapted to growth under optimal or defined laboratory conditions). For these exempt experiments, BL1 physical containment is recommended. For large-scale fermentation experiments, the appropriate physical containment conditions need be no greater than those for the unmodified host organism; the Institutional Biosafety Committee may specify higher containment if deemed necessary.

Appendix C-IV-A Exceptions

The following categories are not exempt from the *NIH Guidelines*: (i) experiments described in Section III-B, which require NIH OSP and Institutional Biosafety Committee approval before initiation; (ii) experiments involving DNA from Risk Groups 3, 4, or restricted organisms (see Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, and Sections V-G and V-L, *Footnotes and References of Sections I through IV*) or cells known to be infected with these agents may be conducted under containment conditions specified in Section III-D-2 with prior Institutional Biosafety Committee review and approval; (iii) large-scale experiments (e.g., more than 10 liters of culture), and (v) experiments involving the deliberate cloning of genes coding for the biosynthesis of molecules toxic for vertebrates (see Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*).

Appendix C-V. *Bacillus subtilis* or *Bacillus licheniformis* Host-Vector Systems

Any asporogenic *Bacillus subtilis* or asporogenic *Bacillus licheniformis* strain which does not revert to a spore-former with a frequency greater than 10^{-7} may be used for cloning DNA with the exception of those experiments listed in Appendix C-V-A, *Exceptions*. For these exempt laboratory experiments, BL1 physical containment conditions are recommended. For large-scale fermentation experiments, the appropriate physical containment conditions need be no greater than those for the unmodified host organism; the Institutional Biosafety Committee can specify higher containment if it deems necessary.

Appendix C-V-A. Exceptions

The following categories are not exempt from the *NIH Guidelines*: (i) experiments described in Section III-B which require NIH OSP and Institutional Biosafety Committee approval before initiation, (ii) experiments involving DNA from Risk Groups 3, 4, or restricted organisms (see Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, and Sections V-G and V-L, *Footnotes and References of Sections I through IV*) or cells known to be infected with these agents may be conducted under containment conditions specified in Section III-D-2 with prior Institutional Biosafety Committee review and approval, (iii) large-scale experiments (e.g., more than 10 liters of culture), and (iv) experiments involving the deliberate cloning of genes coding for the biosynthesis of molecules toxic for vertebrates (see Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*).

Appendix C-VI. Extrachromosomal Elements of Gram Positive Organisms

Recombinant or synthetic nucleic acid molecules derived entirely from extrachromosomal elements of the organisms listed below (including shuttle vectors constructed from vectors described in Appendix C), propagated and maintained in organisms listed below are exempt from these *NIH Guidelines*.

Bacillus amyloliquefaciens
Bacillus amylosacchariticus
Bacillus anthracis
Bacillus aterrimus
Bacillus brevis
Bacillus cereus
Bacillus globigii
Bacillus licheniformis
Bacillus megaterium
Bacillus natto

Bacillus niger
Bacillus pumilus
Bacillus sphaericus
Bacillus stearothermophilus
Bacillus subtilis
Bacillus thuringiensis
Clostridium acetobutylicum
Lactobacillus casei
Listeria grayi
Listeria monocytogenes
Listeria murrayi
Pediococcus acidilactici
Pediococcus damnosus
Pediococcus pentosaceus
Staphylococcus aureus
Staphylococcus carnosus
Staphylococcus epidermidis
Streptococcus agalactiae
Streptococcus anginosus
Streptococcus avium
Streptococcus cremoris
Streptococcus dorans
Streptococcus equisimilis
Streptococcus faecalis
Streptococcus ferus
Streptococcus lactis
Streptococcus ferns
Streptococcus mitior
Streptococcus mutans
Streptococcus pneumoniae
Streptococcus pyogenes

Streptococcus salivarius
Streptococcus sanguis
Streptococcus sobrinus
Streptococcus thermophilus

Appendix C-VI-A. Exceptions

The following categories are not exempt from the *NIH Guidelines*: (i) experiments described in Section III-B which require NIH OSP and Institutional Biosafety Committee approval before initiation, (ii) experiments involving DNA from Risk Groups 3, 4, or restricted organisms (see Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, and Sections V-G and V-L, *Footnotes and References of Sections I through IV*) or cells known to be infected with these agents may be conducted under containment conditions specified in Section III-D-2 with prior Institutional Biosafety Committee review and approval, (iii) large-scale experiments (e.g., more than 10 liters of culture), and (iv) experiments involving the deliberate cloning of genes coding for the biosynthesis of molecules toxic for vertebrates (see Appendix F, *Containment Conditions for Cloning of Genes Coding for the Biosynthesis of Molecules Toxic for Vertebrates*).

Appendix C-VII. The Purchase or Transfer of Transgenic Rodents

The purchase or transfer of transgenic rodents for experiments that require BL1 containment (See Appendix G-III-M, *Footnotes and References of Appendix G*) are exempt from the *NIH Guidelines*.

Appendix C-VIII. Generation of BL1 Transgenic Rodents via Breeding

The breeding of two different transgenic rodents or the breeding of a transgenic rodent and a non-transgenic rodent with the intent of creating a new strain of transgenic rodent that can be housed at BL1 containment will be exempt from the *NIH Guidelines* if:
(1) Both parental rodents can be housed under BL1 containment; and

(2) neither parental transgenic rodent contains the following genetic modifications: (i) incorporation of more than one-half of the genome of an exogenous eukaryotic virus from a single family of viruses; or (ii) incorporation of a transgene that is under the control of a gammaretroviral long terminal repeat (LTR); and
(3) the transgenic rodent that results from this breeding is not expected to contain more than one-half of an exogenous viral genome from a single family of viruses.

Appendix C-IX. Footnotes and References of Appendix C

Appendix C-IX-A. The NIH Director, with advice of the RAC, may revise the classification for the purposes of these *NIH Guidelines* (see Section IV-C-1-b-(2)-(b), *Minor Actions*). The revised list of organisms in each Risk Group is located in Appendix B.

Appendix C-IX-B. A subset of non-conjugative plasmid vectors are poorly mobilizable (e.g., pBR322, pBR313). Where practical, these vectors should be employed.

Appendix C-IX-C. Defined as observable under optimal laboratory conditions by transformation, transduction, phage infection, and/or conjugation with transfer of phage, plasmid, and/or chromosomal genetic information. Note that this definition of exchange may be less stringent than that applied to exempt organisms under Section III-F-6, *Exempt Experiments*.

Appendix C-IX-D. As classified in the *Third Report of the International Committee on Taxonomy of Viruses: Classification and Nomenclature of Viruses*, R. E. F. Matthews (ed.), Intervirology 12 (129-296), 1979.

Appendix C-IX-E. i.e., the total of all genomes within a Family shall not exceed one-half of the genome.

****

APPENDIX D. MAJOR ACTIONS TAKEN UNDER THE *NIH GUIDELINES*

As noted in the subsections of Section IV-C-1-b-(1), the Director, NIH, may take certain actions with regard to the *NIH Guidelines* after the issues have been considered by the RAC. Some of the actions taken to date include the following:

Appendix D-1. Permission is granted to clone foot and mouth disease virus in the EK1 host-vector system consisting of *E. coli* K-12 and the vector pBR322, all work to be done at the Plum Island Animal Disease Center.

Appendix D-2. Certain specified clones derived from segments of the foot and mouth disease virus may be transferred from Plum Island Animal Disease Center to the facilities of Genentech, Inc., of South San Francisco, California. Further development of the clones at Genentech, Inc., has been approved under BL1 + EK1 conditions.

Appendix D-3. The Rd strain of *Hemophilus influenzae* can be used as a host for the propagation of the cloned Tn 10 tet R gene derived from *E. coli* K-12 employing the non-conjugative *Hemophilus* plasmid, pRSF0885, under BL1 conditions.

Appendix D-4. Permission is granted to clone certain subgenomic segments of foot and mouth disease virus in HV1 *Bacillus subtilis* and *Saccharomyces cerevisiae* host-vector systems under BL1 conditions at Genentech, Inc., South San Francisco, California.

Appendix D-5. Permission is granted to Dr. Ronald Davis of Stanford University to field test corn plants modified by recombinant DNA techniques under specified containment conditions.

Appendix D-6. Permission is granted to clone in *E. coli* K-12 under BL1 physical containment conditions subgenomic segments of rift valley fever virus subject to conditions which have been set forth by the RAC.

Appendix D-7. Attenuated laboratory strains of *Salmonella typhimurium* may be used under BL1 physical containment conditions to screen for the *Saccharomyces cerevisiae* pseudouridine synthetase gene. The plasmid YEp13 will be employed as the vector.

Appendix D-8. Permission is granted to transfer certain clones of subgenomic segments of foot and mouth disease virus from Plum Island Animal Disease Center to the laboratories of Molecular Genetics, Inc., Minnetonka, Minnesota, and to work with these clones under BL1 containment conditions. Approval is contingent upon review of data on infectivity testing of the clones by a working group of the RAC.

Appendix D-9. Permission is granted to Dr. John Sanford of Cornell University to field test tomato and tobacco plants transformed with bacterial (*E.coli* K-12) and yeast DNA using pollen as a vector.

Appendix D-10. Permission is granted to Drs. Steven Lindow and Nickolas Panopoulos of the University of California, Berkeley, to release under specified conditions *Pseudomonas syringae*, pathovars (pv.) *syringae,* and *Erwinia herbicola* carrying *in vitro* generated deletions of all or part of the genes involved in ice nucleation.

Appendix D-11. Agracetus of Middleton, Wisconsin, may field test under specified conditions disease resistant tobacco plants prepared by recombinant DNA techniques.

Appendix D-12. Eli Lilly and Company of Indianapolis, Indiana, may conduct large-scale experiments and production involving *Cephalosporium acremonium* strain LU4-79-6 under less than Biosafety Level 1 - Large Scale (BL1-LS) conditions.

Appendix D-13. Drs. W. French Anderson, R. Michael Blaese, and Steven Rosenberg of the NIH, Bethesda, Maryland, can conduct experiments in which a bacterial gene coding for neomycin phosphotransferase will be inserted into a portion of the tumor infiltrating lymphocytes (TIL) of cancer patients using a retroviral vector, N2. The marked TIL then will be combined with unmarked TIL, and reinfused into the patients. This experiment is an addition to an ongoing adoptive immunotherapy protocol in which TIL are isolated from a patient's tumor, grown in culture in the presence of interleukin-2, and reinfused into the patient. The marker gene will be used to detect TIL at various time intervals following reinfusion.
Approval is based on the following four stipulations: (I) there will be no limitation of the number of patients in the continuing trial; (ii) the patients selected will have a life expectancy of about 90 days; (iii) the patients give fully informed consent to participate in the trial; and (iv) the investigators will provide additional data before inserting a gene for therapeutic purposes. (Protocol #8810-001)

Appendix D-14. U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID) may conduct certain experiments involving products of a yellow fever virus originating from the 17-D yellow fever clone at the Biosafety Level 3 containment level using HEPA filters and vaccination of laboratory personnel.

In addition, USAMRIID may conduct certain experiments involving vaccine studies of Venezuelan equine encephalitis virus at the Biosafety Level 3 containment level using HEPA filters and vaccination of laboratory personnel.

Appendix D-15. Drs. R. Michael Blaese and W. French Anderson of the NIH, Bethesda, Maryland, can conduct experiments in which a gene coding for adenosine deaminase (ADA) will be inserted into T lymphocytes of patients with severe combined immunodeficiency disease, using a retroviral vector, LASN. Following insertion of the gene, these T lymphocytes will be reinfused into the patients. The patients will then be followed for evidence of clinical improvement in the disease state, and measurement of multiple parameters of immune function by laboratory testing.

Approval is based on the following two stipulations: (I) that intraperitoneal administration of transduced T lymphocytes not be used before clearance by the Chair of the Recombinant DNA Advisory Committee; and (ii) that the number of research patients be limited to 10 at this time.

In addition to the conditions outlined in the initial approval, patients may be given a supplement of a CD-34+-enriched peripheral blood lymphocytes (PBL) which have been placed in culture conditions that favor progenitor cell growth. This enriched population of cells will be transduced with the retroviral vector, G1NaSvAd. G1NaSvAd is similar to LASN, yet distinguishable by PCR. LASN has been used to transduce peripheral blood T lymphocytes with the ADA gene. Lymphocytes and myeloid cells will be isolated from patients over time and assayed for the presence of the LASN or G1NaSvAd vectors. The primary objectives of this protocol are to transduce CD 34+ peripheral blood cells with the adenosine deaminase gene, administer these cells to patients, and determine if such cells can differentiate into lymphoid and myeloid cells *in vivo*. There is a potential for benefit to the patients in that these hematopoietic progenitor cells may survive longer, and divide to yield a broader range of gene-corrected cells. (Protocol #9007-002)

Appendix D-16. Dr. Steven A. Rosenberg of the National Institutes of Health, Bethesda, Maryland, can conduct experiments on patients with advanced melanoma who have failed all effective therapy. These patients will be treated with escalating doses of autologous tumor infiltrating lymphocytes (TIL) transduced with a gene coding for tumor necrosis factor (TNF). Escalating numbers of transduced TIL will be administered at three weekly intervals along with the administration of interleukin-2 (IL-2). The objective is to evaluate the toxicity and possible therapeutic efficacy of the administration of tumor infiltrating lymphocytes (TIL) transduced with the gene coding for TNF. (Protocol #9007-003)

Appendix D-17. Dr. Malcolm K. Brenner of St. Jude Children's Research Hospital of Memphis, Tennessee, can conduct experiments on patients with acute myelogenous leukemia (AML). Using the LNL6 retroviral vector, the autologous bone marrow cells will be transduced with the gene coding for neomycin resistance. The purpose of this gene marking experiment is to determine whether the source of relapse after autologous bone marrow transplantation for acute myelogenous leukemia is residual malignant cells in the harvested marrow or reoccurrence of tumor in the patient. Determining the source of relapse should indicate whether or not purging of the bone marrow is a necessary procedure. (Protocol #9102-004)

Appendix D-18. Dr. Malcolm K. Brenner of St. Jude Children's Research Hospital of Memphis, Tennessee, can conduct experiments on pediatric patients with Stage D (disseminated) neuroblastoma who are being treated with high-dose carboplatin and etoposide in either phase I/II or phase II trials. All the patients in these studies will be subjected to bone marrow transplantation since it will allow them to be exposed to chemoradiation that would be lethal were it not for the availability of stored autologous marrow for rescue. The bone marrow cells of these patients will be transduced with the gene coding for neomycin resistance using the LNL6 vector. The purpose of this gene marking study is to determine whether the source of relapse after autologous bone marrow transplantation is residual malignant cells in the harvested marrow or residual disease in the patient. Secondly, it is hoped to determine the contribution of marrow autographs to autologous reconstitution. (Protocol #9105-005/9105-006)

Appendix D-19. Dr. Albert B. Deisseroth of the MD Anderson Cancer Center of Houston, Texas, can conduct experiments on patients with chronic myelogenous leukemia who have been reinduced into a second chronic phase or blast cells. The patients in these studies will receive autologous bone marrow transplantation. Using the LNL6 vector, the bone marrow cells will be transduced with the gene coding for neomycin resistance. The purpose of these marking studies is to determine if the origin of relapse arises from residual leukemic cells in the patients or from viable leukemic cells remaining in the bone marrow used for autologous transplantation. (Protocol #9105-007)

Appendix D-20. Drs. Fred D. Ledley and Savio L. C. Woo of Baylor College of Medicine of Houston, Texas, can conduct experiments on pediatric patients with acute hepatic failure who are identified as candidates for hepatocellular transplantation. Using the LNL6 vector, the hepatocytes will be transduced with the gene coding for neomycin resistance. The purpose of using a genetic marker is to demonstrate the pattern of engraftment of transplanted hepatocytes and to help determine the success or failure of engraftment. (Protocol #9105-008)

Appendix D-21. Dr. Steven A. Rosenberg of the National Institutes of Health, Bethesda, Maryland, can conduct experiments on patients with advanced melanoma, renal cell cancer, and colon carcinoma who have failed all effective therapy. In an attempt to increase these patients' immune responses to the tumor, the tumor necrosis factor gene or the interleukin-2 gene will be introduced into a tumor cell line established from the patient. These gene-modified autologous tumor cells will then be injected into the thigh of the patient. To further utilize the immune system of the patient to fight the tumor, stimulated lymphocytes will be cultured from either the draining regional lymph nodes or the injected tumor itself. The patients will be evaluated for antitumor effects engendered by the injection of the gene modified tumor cells themselves as well as after the infusion of the cultured lymphocytes. (Protocol #9110-010/9110-011)

Appendix D-22. Dr. James M. Wilson of the University of Michigan Medical Center of Ann Arbor, Michigan, can conduct experiments on three patients with the homozygous form of familial hypercholesterolemia. Both children and adults will be eligible for this therapy. In an attempt to correct the basic genetic defect in this disease, the gene coding for the low-density lipoprotein (LDL) receptor will be introduced into liver cells taken from the patient. The gene-corrected hepatocytes will then be infused into the portal circulation of the patient through an indwelling catheter. The patients will be evaluated for engraftment of the these treated hepatocytes through a series of metabolic studies; three months after gene therapy, a liver biopsy will be taken and analyzed

for the presence of recombinant derived RNA and DNA to document the presence of the gene coding for the normal LDL receptor. (Protocol #9110-012)

Appendix D-23. Dr. Michael T. Lotze of the University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, can conduct experiments on 20 patients with metastatic melanoma who have failed conventional therapy. A gene transfer experiment will be performed, transducing the patients' tumor infiltrating lymphocytes (TILs) with the gene for neomycin resistance. Through the use of this gene marking technique, it is proposed to determine how long TIL cells can be detected *in vivo* in the peripheral blood of the patients, and how the administration of interleukin-2 and interleukin-4 affects localization and survival of TIL cells in tumor sites. (Protocol #9105-009)

Appendix D-24. Dr. Gary J. Nabel of the University of Michigan Medical School, Ann Arbor, Michigan, can conduct gene therapy experiments on twelve patients with melanoma or adenocarcinoma. Patient population will be limited to adults over the age of 18 and female patients must be postmenopausal or have undergone tubal ligation or orchiectomy. The patient's immune response will be stimulated by the introduction of a gene encoding for a Class I MHC protein, HLA-B7, in order to enhance tumor regression. DNA/liposome-mediated transfection techniques will be used to directly transfer this foreign gene into tumor cells. HLA-B7 expression will be confirmed *in vivo*, and the immune response stimulated by the expression of this antigen will be characterized. These experiments will be analyzed for their efficacy in treating cancer. (Protocol #9202-013)

Appendix D-25. Kenneth Cornetta of Indiana University, Indianapolis, Indiana, can conduct gene transfer experiments on up to 10 patients with acute myelogenous leukemia (AML) and up to 10 patients with acute lymphocytic leukemia (ALL). The patient population will be limited to persons between 18 and 65 years of age. Using the LNL-6 vector, autologous bone marrow cells will be marked with the neomycin resistance gene. Gene marked and untreated bone marrow cells will be reinfused at the time of bone marrow transplantation. Patients will then be monitored for evidence of the neomycin resistance gene in peripheral blood and bone marrow cells in order to determine whether relapse of their disease is a result of residual malignant cells remaining in the harvested marrow or inadequate ablation of the tumor cells by chemotherapeutic agents. Determining the source of relapse may indicate whether or not purging of the bone marrow is a necessary procedure for these leukemia patients. Further studies will be performed in order to determine the percentage of leukemic cells that contain the LNL-6 vector and the clonality of the marked cells. (Protocol #9202-014)

Appendix D-26. Dr. James S. Economou of the University of California, Los Angeles, can conduct gene transfer experiments on 20 patients with metastatic melanoma and 20 patients with renal cell carcinoma. These patients will be treated with various combinations of tumor-infiltrating lymphocytes and peripheral blood leukocytes, including CD8 and CD4 subsets of both types of cells. These effector cell populations will be given in combination with interleukin-2 (IL-2) in the melanoma patients and IL-2 plus alpha interferon in the renal cell carcinoma patients. The effector cells will be transduced with the neomycin resistance gene using either the LNL6 or G1N retroviral vectors. This "genetic marking" of the tumor-infiltrating lymphocytes and peripheral blood lymphocytes is designed to answer questions about the trafficking of these cells, their localization to tumors, and their *in vivo* life span. (Protocol #9202-015)

Appendix D-27. Drs. Philip Greenberg and Stanley R. Riddell of the Fred Hutchinson Cancer Research Center, Seattle, Washington, may conduct gene transfer experiments on 15 human immunodeficiency virus (HIV) seropositive patients (18-45 years old) undergoing allogeneic bone marrow transplantation for non-Hodgkin's lymphoma and 15 HIV-seropositive patients (18-50 years old) who do not have acquired immunodeficiency syndrome (AIDS)-related lymphoma and who are not undergoing bone marrow transplantation to evaluate the safety and efficacy of HIV-specific cytotoxic T lymphocyte (CTL) therapy. CTL will be transduced with a retroviral vector (HyTK) encoding a gene that is a fusion product of the hygromycin phosphotransferase gene (HPH) and the herpes simplex virus thymidine kinase (HSV-TK) gene. This vector will deliver both a marker gene and an ablatable gene in these T cell clones in the event that patients develop side effects as a consequence of CTL therapy. Data will be correlated over time, looking at multiple parameters of HIV disease activity. The objectives of these studies include evaluating the safety and toxicity of CTL therapy, determining the duration of *in vivo* survival of HIV-specific CTL clones, and determining if ganciclovir therapy can eradicate genetically modified, adoptively transferred CTL cells. (Protocol #9202-017)

Appendix D-28. Dr. Malcolm Brenner of St. Jude Children's Research Hospital, Memphis, Tennessee, can conduct gene therapy experiments on twelve patients with relapsed/refractory neuroblastoma who have relapsed after receiving autologous bone marrow transplant. In an attempt to stimulate the patient's immune response, the gene coding for Interleukin-2 (IL-2) will be used to transduce tumor cells, and these gene-modified

cells will be injected subcutaneously in a Phase 1 dose escalation trial. Patients will be evaluated for an anti-tumor response. (Protocol #9206-018)

Appendix D-29. Drs. Edward Oldfield, Kenneth Culver, Zvi Ram, and R. Michael Blaese of the National Institutes of Health, Bethesda, Maryland, can conduct gene therapy experiments on ten patients with primary malignant brain tumors and ten patients with lung cancer, breast cancer, malignant melanoma, or renal cell carcinoma who have brain metastases. The patient population will be limited to adults over the age of 18. Patients will be divided into two groups based on the surgical accessibility of their lesions. Both surgically accessible and surgically inaccessible lesions will receive intra-tumoral injections of the retroviral Herpes simplex thymidine kinase (HS-tk) vector-producer cell line, G1TkSvNa, using a guided stereotaxic approach. Surgically accessible lesions will be excised seven days after stereotaxic injection, and the tumor bed will be infiltrated with the HS-tk producer cells. The removed tumor will be evaluated for the efficiency of transduction. Ganciclovir (GCV) will be administered beginning on the fifth postoperative day. In the case of surgically inaccessible lesions, the patients will receive intravenous therapy with GCV seven days after receiving the intra-tumoral injections of the retroviral HS-tk vector-producer cells. (Protocol #9206-019)

Appendix D-30. Dr. Albert D. Deisseroth of MD Anderson Cancer Center, Houston, Texas, can conduct gene transfer experiments on ten patients who have developed blast crisis or accelerated phase chronic myelogenous leukemia (CML). The retroviral vectors G1N and LNL6 which code for neomycin resistance will be used to transduce autologous peripheral blood and bone marrow cells that have been removed and stored at the time of cytogenetic remission or re-induction of chronic phase in Philadelphia chromosome positive CML patients. Following reinduction of the chronic phase of CML and preparative chemotherapy, patients will be infused with the transduced autologous cells.

This protocol is designed to determine the cause of relapse of CML. If polyclonal CML neomycin marked blastic cells appear at the time of relapse, their presence will indicate that relapse arises from the leukemic CML blast cells present in the autologous cells infused following chemotherapy. If residual systemic disease contributes to relapse, the neomycin resistance gene will not be detected in the CML leukemic blasts at the time of relapse.

This study will compare the relative contributions of the peripheral blood and bone marrow to generate hematopoietic recovery after bone marrow transplantation and evaluate purging and selection of peripheral blood or bone marrow as a source of stem cells for transplant. The percentage of neomycin resistant CML cells which are leukemic will be determined by PCR analysis and detection of bcr-abl mRNA. (Protocol #9206-020)

Appendix D-31. Dr. Cynthia Dunbar of the National Institutes of Health, Bethesda, Maryland, can conduct gene transfer experiments on up to 48 patients with multiple myeloma, breast cancer, or chronic myelogenous leukemia. The retroviral vectors G1N and LNL6 will be used to transfer the neomycin resistance marker gene into autologous bone marrow and peripheral blood stem cells in the presence of growth factors to examine hematopoietic reconstitution after bone marrow transplantation. The efficiency of transduction of both short and long term autologous bone marrow reconstituting cells will be examined.

Autologous bone marrow and CD34+ peripheral blood stem cells will be enriched prior to transduction. Myeloma and CML patients will receive both autologous bone marrow and peripheral blood stem cell transplantation. These separate populations will be marked with both the G1N and LNL6 retroviral vectors. If short and long term marking experiments are successful, important information may be obtained regarding the biology of autologous reconstitution, the feasibility of retroviral gene transfer into hematopoietic cells, and the contribution of viable tumor cells within the autograft to disease relapse. (Protocol #9206-023/9206-024/9206-025)

Appendix D-32. Dr. Bernd Gansbacher of the Memorial Sloan-Kettering Cancer Center, New York, New York, can conduct gene therapy experiments on twelve patients over 18 years of age with metastatic melanoma who are HLA-A2 positive and who have failed conventional therapy. This is a phase I study to examine whether allogeneic HLA-A2 matched melanoma cells expressing recombinant human Interleukin-2 (IL-2) can be injected subcutaneously and used to create a potent tumor specific immune response without producing toxicity. By allowing the tumor cells to present the MHC Class I molecule as well as the secreted IL-2, a clonal expansion of tumor specific effector cells is expected. These effector populations may access residual tumor at distant sites via the systemic circulation. (Protocol #9206-021)

Appendix D-33. Dr. Bernd Gansbacher of the Memorial Sloan-Kettering Cancer Center, New York, New York, can conduct gene therapy experiments on twelve patients over 18 years of age with renal cell carcinoma who are HLA-A2 positive and who have failed conventional therapy. This Phase I study will examine whether allogeneic HLA-A2 matched renal cell carcinoma cells expressing recombinant human Interleukin-2 (IL-2) can be injected subcutaneously and used to create a potent tumor specific immune response without producing toxicity. By allowing the tumor cells to present the MHC Class I molecule as well as the secreted IL-2, a clonal expansion of tumor specific effector cells is expected. These effector populations may access residual tumor at distant sites via the systemic circulation. (Protocol #9206-022)

Appendix D-34. Dr. Michael T. Lotze, University of Pittsburgh, Pittsburgh, Pennsylvania, can conduct experiments on twenty patients with metastatic, and/or unresectable, locally advanced melanoma, renal cell carcinoma, breast cancer, or colon cancer who have failed standard therapy. Patients will receive multiple subcutaneous injections of autologous tumor cells combined with an autologous fibroblast cell line that has been transduced *in vitro* with the gene coding for Interleukin-4 (IL-4) to augment the *in vivo* antitumor effect. Patients will be monitored for antitumor effect by PCR analysis and multiple biopsy of the injection site. (Protocol #9209-033)

Appendix D-35. Dr. Friedrich G. Schuening, Fred Hutchinson Cancer Research Center, Seattle, Washington, can conduct human gene transfer experiments on patients ≥ 18 years of age with breast cancer, Hodgkin's disease, or non-Hodgkin's lymphoma. A total of 10 patients per year will be enrolled in the studies over a period of four years. Patients will undergo autologous bone marrow transplantation with a selected population of Interleukin-3 (IL-3) or granulocyte colony-stimulating factor (G-CSF) stimulated CD34(+) peripheral blood repopulating cells (PBRC) that have been transduced with the gene coding for neomycin resistance (neo^R) using the retroviral vector, LN. Patients will be continuously monitored for neo^R to determine the relative contribution

of autologous PBRCs to long-term hematopoietic reconstitution. Demonstration of long-term contribution of autologous PBRC to hematopoiesis will enable the use of PBRC alone for autologous transplants and suggest the use of PBRC as long-term carriers of therapeutically relevant genes. (Protocol #9209-027/9209-028)

Appendix D-36. Dr. Friedrich G. Schuening, Fred Hutchinson Cancer Research Center, Seattle, Washington, can conduct human gene transfer experiments on patients ≥ 18 years of age with breast cancer, Hodgkin's disease, or non-Hodgkin's lymphoma. A total of 5 patients per year will be enrolled in the study over a period of four years. Patients will undergo allogeneic bone marrow transplant with granulocyte colony-stimulating factor (G-CSF) stimulated CD34(+) PBRC harvested from an identical twin that have been transduced with neo^R using the retroviral vector, LN. Patients will be continuously monitored for neo^R to determine the relative contribution of G-CSF stimulated allogeneic PBRCs to long-term bone marrow engraftment. Demonstration of long-term contribution of allogeneic PBRC to hematopoiesis will enable the use of PBRC alone for allogeneic transplants and suggest the use of PBRC as long-term carriers of therapeutically relevant genes. (Protocol #9209-029)

Appendix D-37. Dr. Malcolm K. Brenner of St. Jude Children's Hospital, Memphis, Tennessee, and Dr. Bonnie J. Mills of Baxter Healthcare Corporation, Santa Ana, California, can conduct a multicenter uncontrolled human gene transfer experiment on 12 patients ≤ 21 years of age with Stage D Neuroblastoma in first or second marrow remission. Autologous bone marrow cells will be separated into two fractions, purged and unpurged. Each fraction will be transduced with the neo^R gene by either LNL6 or G1Na. Patients will be monitored by the polymerase chain reaction (PCR) for the presence of neo^R. The protocol is designed to evaluate the safety and efficacy of the Neuroblastoma Bone Marrow Purging System following high dose chemotherapy. (Protocol #9209-032)

Appendix D-38. Drs. Carolyn Keierleber and Ann Progulske-Fox of the University of Florida, Gainesville, Florida, can conduct experiments involving the introduction of a gene coding for tetracycline resistance into *Porphyromonas gingivalis* at a physical containment level of Biosafety Level-2 (BL-2).

Appendix D-39. Dr. Scott M. Freeman of Tulane University Medical Center, New Orleans, Louisiana, can conduct experiments on patients with epithelial ovarian carcinoma who have clinical evidence of recurrent, progressive, or residual disease who have no other therapy available to prolong survival. Patients will be injected intraperitoneally with the irradiated PA-1 ovarian carcinoma cell line which has been transduced with the herpes simplex thymidine kinase (HSV-TK) gene. The patients will then receive ganciclovir therapy. Previous, data indicates that HSV-TK+ tumor cells exhibit a killing effect on HSV-TK- cells when exposed to ganciclovir therapy. Patients will be evaluated for safety and side effects of this treatment. (Protocol #9206-016)

Appendix D-40. Dr. Michael J. Welsh, Howard Hughes Medical Institute Research Laboratories, University of Iowa College of Medicine, Iowa City, Iowa, may conduct experiments on 3 cystic fibrosis (CF) patients ≥ 18 years of age with mild to moderate disease. This Phase I study will determine the: (1) *in vivo* safety and efficacy of the administration of the replication-deficient type 2 adenovirus vector, Ad2/CFTR-1, to the nasal epithelium; (2) efficacy in correcting the CF chloride transport defect *in vivo*; and (3) effect of adenovirus vector dosage on safety and efficacy. (Protocol #9212-036)

Appendix D-41. Dr. Ronald G. Crystal, National Institutes of Health, Bethesda, Maryland, may conduct experiments on 10 cystic fibrosis (CF) patients ≥ 21 years of age. Patients will receive an initial administration of the replication-deficient type 5 adenovirus vector, AdCFTR, to their left nares. If no toxicity is observed from intranasal administration, patients will receive a single administration of AdCFTR to the respiratory epithelium of their left large bronchi. Five groups of patients (2 patients per group) will be studied based on increased dosage administration of AdCFTR. This study will determine the: (1) *in vivo* safety and efficacy of the administration of AdCFTR into the respiratory epithelium; (2) efficacy of the correction of the biologic abnormalities of CF in the respiratory epithelium; (3) duration of the biologic correction; (4) efficacy of the correction of the abnormal electrical potential difference of the airway epithelial sheet; (5) clinical parameters relevant to the disease process; and (6) if humoral immunity develops against AdCFTR sufficient to prevent repeat administration. (Protocol #9212-034)

Appendix D-42. Dr. Kenneth Culver, Iowa Methodist Medical Center, Des Moines, Iowa, and Dr. John Van Gilder, University of Iowa, Iowa City, Iowa, may conduct experiments on 15 patients ≥ 18 years of age with recurrent malignant primary brain tumors or lung, melanoma, renal cell carcinoma, or breast carcinoma brain metastases who have failed standard therapy for their disease. Patient eligibility will be limited to those patients who have measurable residual tumor immediately following the post-operative procedure as demonstrated by imaging studies. The number of patients treated will be equally divided between the Iowa Methodist Medical Center and the University of Iowa. If a positive response is observed in any of the first 15 patients, the investigators may submit a request to treat an additional 15 patients.
Following surgical debulking, patients will receive a maximum of 3 intralesional injections of the G1TkSvNa vector- producing cell line (VPC) to induce regression of residual tumor cells by ganciclovir (GCV) therapy. Patients who demonstrate stable disease for a minimum of 6 months following this treatment will be eligible for additional VPC injections and subsequent GCV therapy. (Protocol #9303-037)

Appendix D-43. Drs. Malcolm Brenner, Robert Krance, Helen E. Heslop, Victor Santana, and James Ihle, St. Jude Children's Research Hospital, Memphis, Tennessee, may conduct experiments on 35 patients ≥ 1 year and ≤ 21 years of age at the time of initial diagnosis of acute myelogenous leukemia (AML). The investigators will use the two retroviral vectors, LNL6 and G1Na, to determine the efficacy of the bone marrow purging techniques: 4-hydroxyperoxicyclophosphamide and interleukin-2 (IL-2) activation of endogenous cytotoxic effector cells, in preventing relapse from the reinfusion of autologous bone marrow cells. (Protocol #9303-039)

Appendix D-44. Drs. Helen E. Heslop, Malcolm Brenner, and Cliona Rooney, St Jude Children's Research Hospital, Memphis, Tennessee, may conduct experiments of 35 patients ≤ 21 years of age who will be recipients of mismatched-related or phenotypically similar unrelated donor marrow grafts for leukemia. In this Phase I dose escalation study, spontaneous lymphoblastoid cell lines will be established that express the same range of Epstein-Barr Virus (EBV) encoded proteins as the recipient. These EBV-specific cell lines will be transduced with LNL6 or G1Na and readministered at the time of bone marrow transplant. This study will determine: (1) survival and expansion of these EBV-specific cell lines *in vivo*, (2) the ability of these adoptively transferred cells to confer protection against EBV infection, and (3) appropriate dosage and administration schedules. (Protocol #9303-038)

Appendix D-45. Drs. Robert W. Wilmott and Jeffrey Whitsett, Children's Hospital Medical Center, Cincinnati, Ohio, and Dr. Bruce Trapnell, Genetic Therapy, Inc., Gaithersburg, Maryland, may conduct experiments on 15 cystic fibrosis (CF) patients who have mild to moderate disease ≥ 21 years of age. The replication-deficient type 5 adenovirus vector, Av1CF2, will be administered to the nasal and lobar bronchial respiratory tract of patients. This study will demonstrate the: (1) expression of normal cystic fibrosis transmembrane conductance regulator (CFTR) mRNA *in vivo*, (2) synthesis of CFTR protein, and (3) correction of epithelial cell cAMP dependent Cl^- permeability. The pharmacokinetics of CFTR expression and ability to re-infect the respiratory tract with AvCF2 will be determined. Systemic and local immunologic consequences of Av1CF2 infection, the time of viral survival, and potential for recombination or complementation of the virus will be monitored. (Protocol #9303-041)

Appendix D-46. Dr. James M. Wilson of the University of Pennsylvania Medical Center, Philadelphia, Pennsylvania, may conduct experiments on 20 adult patients with advanced cystic fibrosis lung disease. An isolated segment of the patients' lung will be transduced with the E1 deleted, replication-incompetent adenovirus vector, Ad.CB-CFTR using a bronchoscope for gene delivery. Ad.CB-CFTR contains the human gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Pulmonary biopsies will be obtained by bronchoscopy at 4 days, 6 weeks, and 3 months following treatment. Patients will be monitored for evidence of CFTR gene transfer and expression, immunological responses to CFTR or adenovirus proteins, and toxicity. (Protocol #9212-035)

Appendix D-47. Dr. Hilliard F. Seigler of Duke University Medical Center, Durham, North Carolina, may conduct experiments on 20 patients with disseminated malignant melanoma. Autologous tumor cells will be transduced with a retroviral vector, pHuγ-IFN, that contains the gene encoding human γ-IFN. Following lethal irradiation, the transduced cells will be readministered to patients for the purpose of generating cytotoxic T cells that are tumor specific along with the up-regulation of Class I major histocompatibility antigens. Patients will be monitored for clinical regression of tumors and generation of tumor-specific cytotoxic T lymphocytes. (Protocol #9306-043)

Appendix D-48. Drs. Stefan Karlsson and Cynthia Dunbar of the National Institutes of Health, Bethesda, Maryland, and Dr. Donald B. Kohn of the Children's Hospital of Los Angeles, Los Angeles, California, may conduct experiments on 10 patients with Gaucher disease. CD34(+) hematopoietic stem cells will be isolated from bone marrow or from peripheral blood treated with granulocyte-colony stimulating factor. CD34(+) cells will be transduced with a retrovirus vector, G1Gc, containing cDNA encoding human glucocerebrosidase and administered intravenously. Patients will be monitored for toxicity and glucocerebrosidase expression. (Protocol #9306-047)

Appendix D-49. Dr. Gary J. Nabel of the University of Michigan Medical Center, Ann Arbor, Michigan, may conduct experiments on 12 patients with AIDS to be divided into 4 experimental groups. CD4(+) lymphocytes will be isolated from peripheral blood and transduced with Rev M10, a transdominant inhibitory mutant of the *rev* gene of the human immunodeficiency virus (HIV). Transduction of the *rev* mutant will be mediated either by the retrovirus vector, PLJ-cREV M10, or by particle-mediated gene transfer of plasmid DNA. Patients will be monitored for survival of the transduced CD4(+) cells by polymerase chain reaction and whether Rev M10 can confer protection against HIV infection to CD4(+) cells. (Protocol #9306-049)

Appendix D-50. Dr. Gary J. Nabel of the University of Michigan Medical Center, Ann Arbor, Michigan, may conduct experiments on 24 patients with advanced cancer. Patients will undergo *in vivo* transduction with DNA/liposome complexes containing genes encoding the HLA-B7 histocompatibility antigen and beta-2 microglobulin in a non-viral plasmid. These DNA/liposome complexes will be administered either by intratumoral injection or catheter delivery. Patients will be monitored for enhanced immune responses against tumor cells, and safe and effective doses will be determined. (Protocol #9306-045)

Appendix D-51. Dr. John A. Barranger of the University of Pittsburgh, Pittsburgh, Pennsylvania, may conduct experiments on 5 patients with Gaucher disease. The CD34(+) hematopoietic stem cells will be isolated from peripheral blood and transduced *in vitro* with the retrovirus vector, N2-Sv-GC, encoding the glucocerebrosidase (GC) enzyme. Following reinfusion of the transduced cells, patients will be monitored by PCR analysis for GC expression in peripheral blood leukocytes. Patients currently receiving GC replacement therapy and who demonstrate clinical responsiveness will be withdrawn from exogenous GC therapy. Patients not previously treated with exogenous GC, will be monitored for clinical reversal of lipid storage symptoms. (Protocol #9306-046)

Appendix D-52. Dr. Robert Walker of the National Institutes of Health, Bethesda, Maryland, may conduct experiments on 12 HIV-infected patients who have a seronegative identical twin. CD4(+) and CD8(+) cells will be isolated from the seronegative twin and induced to polyclonal proliferation with anti-CD3 and interleukin-2. The enriched population of cells will be transduced with either LNL6 or G1Na, which contain the neo^R gene. The transduced cells will be expanded in tissue culture and administered to the HIV-infected twin. Patients will be monitored for immune function and the presence of marked cells. (Protocol #9209-026)

Appendix D-53. Dr. Corey Raffel of the Children's Hospital Los Angeles, Los Angeles, California, and Dr. Kenneth Culver of Iowa Methodist Medical Center, Des Moines, Iowa, may conduct experiments on 30 patients between 2 and 18 years of age with recurrent malignant astrocytoma. Fifteen patients will be accrued into this

study initially. If at least one patient responds to therapy, an additional 14 patients will be treated. Patients with either surgically accessible or non-accessible tumors will be treated with the vector producing cell line (PA317) carrying the retrovirus vector, G1TkSvNa. This vector will transduce tumor cells *in vivo* with the *Herpes simplex* thymidine kinase (HS-tk) gene that renders the cells sensitive to killing by ganciclovir. Surgically accessible patients will undergo surgical debulking of their tumor followed by repeated administration of the HS-tk vector producer cells into the tumor bed. Children with unresectable tumors will undergo stereotaxic injection of vector producer cells into tumors. (Protocol #9306-050)

Appendix D-54. Dr. Jeffrey E. Galpin of the University of Southern California, Los Angeles, California, and Dr. Dennis A. Casciato of the University of California, Los Angeles, California, may conduct experiments on 15 HIV(+) asymptomatic patients. Patients will receive 3 monthly intramuscular injections of the retrovirus vector (N2IIIBenv) encoding the HIV-1 IIIB envelope protein. Patients will be monitored for acute toxicity, CD4 levels, HIV-specific CTL responses, and viral burdens. (Protocol #9306-048)

Appendix D-55. Drs. Charles Hesdorffer and Karen Antman of Columbia University College of Physicians and Surgeons, New York, New York, may conduct experiments on 20 patients with advanced breast, ovary, and brain cancer. CD34(+) hematopoietic stem cells will be isolated from bone marrow, transduced with the retrovirus vector, PHaMDR1/A, and readministered to patients. Patients will be monitored for the presence and expression of the MDR-1 gene. The investigators will determine whether MDR-1 expression increases following chemotherapy. (Protocol #9306-051)

Appendix D-56. Dr. Enzo Paoletti of Virogenetics Corporation, Troy, New York, may conduct experiments with poxvirus vectors NYVAC, ALVAC, and TROVAC at Biosafety Level 1.

Appendix D-57. Drs. Richard C. Boucher and Michael R. Knowles of the University of North Carolina, Chapel Hill, North Carolina, may conduct experiments on 9 patients (18 years old or greater) with cystic fibrosis to test for the safety and efficacy of an E1-deleted recombinant adenovirus containing the cystic fibrosis transmembrane conductance regulator (CFTR) cDNA, Ad.CB-CFTR. A single dose of 10^8, 3×10^9 or 10^{11} pfu/ml will be administered to the nasal cavity of 3 patients in each dose group. Patients will be monitored by nasal lavage and biopsy to assess safety and restoration of normal epithelial function. (Protocol #9303-042)

Appendix D-58. Dr. Joyce A. O'Shaughnessy of the National Institutes of Health, Bethesda, Maryland, may conduct experiments on 18 patients (18-60 years old) with Stage IV breast cancer who have achieved a partial or complete response to induction chemotherapy. This study will determine the feasibility of obtaining engraftment of CD34(+) hematopoietic stem cells transduced by a retroviral vector, G1MD, and expressing a cDNA for the human multi-drug resistance-1 (MDR-1) gene following high dose chemotherapy, and whether the transduced MDR-1 gene confers drug resistance to hematopoietic cells and functions as an *in vivo* dominant selectable marker. Patients will be monitored for evidence of myeloprotection and presence of the transduced MDR-1 gene." (Protocol #9309-054)

Appendix D-59. Drs. Larry E. Kun, R. A. Sanford, Malcolm Brenner, and Richard L. Heideman of St. Jude Children's Research Hospital, Memphis, Tennessee, and Dr. Edward H. Oldfield of the National Institutes of Health, Bethesda, Maryland, may conduct experiments on 6 patients (3-21 years old) with progressive or recurrent malignant supratentorial tumors resistant to standard therapies. Mouse cells producing the retroviral vector containing the herpes simplex thymidine kinase gene (G1TKSVNa) will be instilled into the tumor areas via multiple stereotactically placed cannulas. Patients will be treated with ganciclovir to eliminate cells expressing the transduced gene. Patients will be monitored for central nervous system, hematologic, renal or other toxicities, and for anti-tumor responses by magnetic resonance imaging studies. (Protocol #9309-055)

Appendix D-60. The physical containment level may be reduced from Biosafety Level 3 to Biosafety Level 2 for a Semliki Forest Virus (SFV) vector expression system of Life Technologies, Inc., Gaithersburg, Maryland.

Appendix D-61. Dr. Albert B. Deisseroth of the University of Texas MD Anderson Cancer Center, Houston, Texas, may conduct experiments on 10 patients (≥ 16 to ≤ 60 years of age) with chronic lymphocytic leukemia. Autologous peripheral blood and bone marrow cells will be removed from patients following chemotherapy and marked by transduction with two distinguishable retroviral vectors, G1Na and LNL6, containing the neomycin resistance gene. The gene marked cells will be reinfused into patients to determine the efficiency of bone marrow purging and the origin of relapse following autologous bone marrow transplantation. (Protocol #9209-030)

Appendix D-62. Dr. Jonathan Simons of the Johns Hopkins Oncology Center, Baltimore, Maryland, may conduct experiments on 26 patients (≥ 18 years of age) with metastatic renal cell carcinoma to evaluate the safety and tolerability of intradermally injected autologous irradiated tumor cells transduced with the retrovirus vector, MFG, containing the human granulocyte-macrophage colony stimulating factor gene. Acute and long-term clinical toxicities and *in vitro* and *in vivo* induction of specific anti-tumor immune responses will be monitored. (Protocol #9303-040)

Appendix D-63. Dr. Albert B. Deisseroth of the University of Texas MD Anderson Cancer Center, Houston, Texas, may conduct experiments on 20 patients (≥ 18 and ≤ 60 years old) with ovarian cancer. A murine viral vector was constructed from the third generation of L series retroviruses with the insert of the human multi-drug resistance-1 (MDR-1) transduced gene. The investigators will assess the safety and feasibility of administering CD34 (+) autologous peripheral blood and bone marrow cells. Patients will be monitored for the presence of the MDR-1 gene and for the effect of gene transfer on hematopoietic function following the transplantation. (Protocol #9306-044)

Appendix D-64. Dr. Joseph Ilan of the Case Western Reserve University School of Medicine and University Hospital of Cleveland, Cleveland, Ohio, may conduct experiments on 12 patients (≥ 18 years of age) with advanced brain cancer. Human malignant glioma tumor cells will be cultured, transfected with Epstein-Barr virus (EBV)-based vector, anti-Insulin growth factor-I, lethally irradiated, and injected subcutaneously into patients in an attempt to express antisense Insulin growth factor-1. Patients will be monitored for toxicity and immunologic responses to the vector. (Protocol #9306-052)

Appendix D-65. Drs. James S. Economou and John Glaspy of the University of California, Los Angeles, California, may conduct experiments on 30 patients (≥ 18 to ≤ 70 years of age) with metastatic melanoma. A human melanoma cell line (M-24) will be transduced with the retroviral vector, G1NaCvi2, expressing the human interleukin-2 (IL-2) gene. The IL-2 producing cells will be mixed with the patient's autologous tumor cells, irradiated, and injected subcutaneously in an attempt to enhance the tumor-specific immunologic response. Patients will be monitored for toxicity, *in vitro* and *in vivo* immunologic responses, and clinical anti-tumor effects. (Protocol #9309-058)

Appendix D-66. Drs. Peter Cassileth, Eckhard R. Podack, and Kasi Sridhar of the University of Miami, and Niramol Savaraj of the Miami Veterans Administration Hospital, Miami, Florida, may conduct experiments on 12 patients (≥ 18 years of age) with limited stage small cell lung cancer. Autologous tumor cells will be removed, expanded in culture, and transduced by lipofection with the BMG-Neo-hIL2 vector (derived from bovine papilloma virus). The objective of this protocol is to demonstrate the safety and efficacy of administering IL-2 transduced autologous tumor cells in an attempt to stimulate a tumor-specific cytotoxic T lymphocyte (CTL) response, and to determine the quantity and characteristics of the CTL that have been generated. (Protocol #9309-053)

Appendix D-67. Drs. Edward H. Oldfield and Zvi Ram of the National Institutes of Health, Bethesda, Maryland, may conduct experiments on 20 patients (≥ 18 years of age) with leptomeningeal carcinomatosis. The patients will receive intraventricular or subarachnoid injection of murine vector producing cells containing the retroviral vector, G1Tk1SvNa. Tumor cells expressing the herpes simplex thymidine kinase gene will be rendered sensitive to killing by subsequent administration of ganciclovir. Patients will be monitored for safety and anti-tumor response by magnetic resonance imaging (MRI) and cerebral spinal fluid cytological analysis. (Protocol #9312-059)

Appendix D-68. Drs. Tapas K. Das Gupta and Edward P. Cohen of the University of Illinois College of Medicine, Chicago, Illinois, may conduct experiments on 12 subjects who differ in at least 3 out of 6 alleles at the Class I histocompatibility locus (≥ 18 years of age) with Stage IV malignant melanoma. The subjects will be immunized with a lethally irradiated allogeneic human melanoma cell line transduced with the human interleukin-2 expressing retroviral vector, pZipNeoSv-IL-2. Subjects will be monitored for toxicity, induction of B and T cell antitumor responses *in vitro* and *in vivo,* and any clinical evidence of antitumor effect. (Protocol #9309-056)

Appendix D-69A. Dr. Michael J. Welsh of the Howard Hughes Medical Institute, Iowa City, Iowa, may conduct experiments on 20 patients (≥ 18 years of age) with cystic fibrosis. The investigator will administer increasing doses of either of the two adenovirus vectors, AD2/CFTR-1 or AD2-ORF6/PGK-CFTR, to the nasal epithelium of 10 patients (1 nostril) or maxillary sinus epithelium of 10 patients (1 maxillary sinus). Patients will be isolated for a period of 24 hours following vector administration; however, if 1 patient demonstrates secreted virus at 24

hours, the investigator will notify the Recombinant DNA Advisory Committee for reconsideration of the isolated period. Patients will be assessed for the safety and efficacy of multiply administration of adenovirus vectors encoding the cystic fibrosis transmembrane conductance regulator (CFTR) gene. (Protocol #9312-067)

Appendix D-69B. Dr. Richard Haubrich of the University of California at San Diego Treatment Center, San Diego, California, may conduct experiments on 25 human immunodeficiency virus (HIV)-infected, seropositive, asymptomatic subjects (≥ 18 to ≤ 65 years of age). Subjects will receive 3 monthly intramuscular injections of the retroviral vector, N2/IIIB *env/rev*, which encodes for HIV-1 IIIB *env/rev* proteins. The objective of the study is to induce HIV-1- specific CD8(+) cytotoxic T lymphocyte and antibody responses in order to eliminate HIV-infected cells and residual virus. This Phase I/II study will evaluate acute toxicity, identify long-term treatment effects, and evaluate the disease progression. (Protocol #9312-062)

Appendix D-70. Dr. Mario Sznol of the National Institutes of Health, Frederick, Maryland, may conduct experiments on 50 subjects (≥ 18 years of age) with advanced stage melanoma. Subjects will receive subcutaneous injections of lethally irradiated allogeneic melanoma cells that have been transduced by lipofection with the plasmid DNA vector, CMV-B7, derived from bovine papilloma virus to express the human B7 antigen. The B7 antigen, which binds to the CD28 receptor of T cells, will serve as a co-stimulatory signal to elicit an antitumor immune response. Subjects will be monitored for induction of cytotoxic T lymphocyte, antitumor responses *in vitro* and *in vivo* and any clinical evidence of antitumor effect. (Protocol #9312-063)

Appendix D-71. Dr. Joseph Rubin of the Mayo Clinic, Rochester, Minnesota, may conduct experiments on 15 subjects with hepatic metastases from advanced colorectal cancer (≥ 18 years of age). Subjects will receive intratumoral hepatic injections of the plasmid DNA/lipid complex, pHLA-B7/β-2 microglobulin, expressing a heterodimeric cell surface protein consisting of the HLA-B7 histocompatibility antigen and β-2 microglobulin. Subjects must be HLA-B7 negative. The objective of this study is to determine a safe and effective dose of the DNA/lipid complex. Subjects will be monitored for antigen-specific immune responses and *in vivo* HLA-B7 expression. (Protocol #9312-064)

Appendix D-72. Dr. Nicholas J. Vogelzang of the University of Chicago Medical Center, Chicago, Illinois, may conduct experiments on 15 subjects with metastatic renal cell carcinoma ≥ 18 years of age. Subjects will receive intratumoral injections of the plasmid DNA/liposome vector pHLA-B7/β-2 microglobulin, expressing a heterodimeric cell surface protein consisting of the HLA-B7 histocompatibility antigen and β-2 microglobulin. Subjects must be HLA-B7 negative. Subjects will be monitored for antigen-specific immune responses and *in vivo* HLA-B7 expression. (Protocol #9403-071)

Appendix D-73. Dr. Evan M. Hersh of the Arizona Cancer Center and Drs. Emmanuel Akporiaye, David Harris, Alison T. Stopeck, Evan C. Unger, and James A. Warneke of the University of Arizona, Tucson, Arizona, may conduct experiments on 15 subjects with metastatic malignant melanoma ≥ 18 years of age. Subjects will receive intratumoral injections of the plasmid DNA/liposome vector, pHLA-B7/β-2 microglobulin, expressing a heterodimeric cell surface protein consisting of the HLA-B7 histocompatibility antigen and β-2 microglobulin. Subjects must be HLA-B7 negative. Subjects will be monitored for antigen-specific immune responses and *in vivo* HLA-B7 expression. (Protocol #9403-072)

Appendix D-74. Dr. Ralph Freedman of MD Anderson Cancer Center, Houston, Texas, may conduct gene marking experiments on 9 subjects with ovarian carcinoma or peritoneal carcinomatosis (≥ 16 years of age). Autologous CD3(+)/CD8(+) tumor infiltrating lymphocyte derived T cells will be transduced with the retroviral vector G1Na that encodes for neo^R. Subjects will receive intraperitoneal administration of bulk expanded transduced and nontransduced T cells and recombinant interleukin-2. Previously documented tumor sites and normal tissues will be monitored for neo^R and the proportion of CD3(+)/CD8(+) T cells will be determined. (Protocol #9406-075)

Appendix D-75. Drs. Helen Heslop, Malcolm Brenner, and Robert Krance of St. Jude Children's Research Hospital, Memphis, Tennessee, may conduct gene marking experiments on 20 subjects undergoing autologous bone marrow transplantation for therapy of leukemia or solid tumor (< 21 years of age). The distinguishable retroviral vectors, LNL6 and G1Na (both encoding neo^R), will be used to determine the rate of reconstitution of untreated versus cytokine expanded CD34(+) selected autologous bone marrow cells. (Protocol #9406-076)

Appendix D-76. Drs. Albert Deisseroth, Gabriel Hortobagyi, Richard Champlin, and Frankie Holmes of MD Anderson Cancer Center, Houston, Texas, may conduct experiments on 10 fully evaluable subjects (maximum of 20 entered) with Stage III or IV breast cancer (≥ 18 and ≤ 60 years of age). Subjects will receive autologous

CD34(+) peripheral blood cells that have been transduced with the retroviral vector, pVMDR-1, which encodes the multi-drug resistance gene. The objective of this study is to evaluate the safety and feasibility of transducing early hematopoietic progenitor cells with pVMDR-1 and to determine *in vivo* selection of chemotherapy resistant hematopoietic cells. (Protocol #9406-077)

Appendix D-77. Drs. Johnson M. Liu and Neal S. Young of the National Institutes of Health, Bethesda, Maryland, may conduct experiments on 6 patients with Fanconi anemia (≥ 5 years of age). Subjects will receive autologous CD34(+) cells that have been transduced with the retroviral vector, FACC, which encodes the normal Fanconi anemia complementation group C gene. The objective of this study is to determine whether autologous FACC transduced hematopoietic progenitor cells can be safely administered to subjects, the extent of engraftment, and correction of cell phenotype. (Protocol #9406-078)

Appendix D-78. Drs. Robert E. Sobol and Ivor Royston of the San Diego Regional Cancer Center, San Diego, California, may conduct experiments on 15 subjects with recurrent residual glioblastoma multi-forme (≥ 18 years of age). Subjects will receive subcutaneous injections of autologous tumor cells that have been lethally irradiated and transduced with the retroviral vector, G1NaCvi2.23, which encodes for interleukin-2. Subjects will be monitored *in vitro* for cellular and humoral antitumor responses and *in vivo* for antitumor activity. (Protocol #9406-080)

Appendix D-79. Dr. Alfred E. Chang of the University of Michigan Medical Center, Ann Arbor Michigan, may conduct gene marking experiments on 15 subjects with metastatic melanoma (≥ 18 years of age). Subjects will undergo adoptive immunotherapy of anti-CD3/interleukin-2 activated lymph node cells that have been primed *in vivo* with tumor cells that have been transduced with the retrovirus vector, GBAH4, encoding the gene for interleukin-4. The investigator will evaluate the antitumor efficacy and *in vivo* immunological reactivity in subjects receiving adoptively transferred T cells, and the *in vitro* immunological reactivities of the activated T cells that might correlate with their *in vivo* antitumor function. (Protocol #9312-065)

Appendix D-80. Dr. Robert Walker of the National Institutes of Health, Bethesda, Maryland, may conduct gene marking experiments on 40 HIV(+) subjects (≥ 18 years of age). The investigator may also enter an additional number of subjects (to be determined by the investigator) who will receive a single administration of 1×10^7 HIV-specific CD8(+) cells. The investigator will: (1) Assess the safety and tolerance of the adoptive transfer of anti-HIV cytotoxic, syngeneic, CD8(+) peripheral blood lymphocytes that have been transduced with the retrovirus vector, r*kat*4svgF3e-, that encodes for a universal chimeric T cell receptor. (2) Determine the longevity of the genetically marked CD8(+) lymphocytes in the subject's peripheral blood. (Protocol #9403-069)

Appendix D-81. Dr. Joseph Rosenblatt of the University of California, Los Angeles, California, and Dr. Robert Seeger of Children's Hospital, Los Angeles, California, may conduct gene transfer experiments on 18 subjects with neuroblastoma (≤ 21 years of age). Patients at high risk of relapse with minimal or no detectable disease following myeloablative therapy and autologous bone marrow transplantation, or patients with progressive or persistent disease despite conventional therapy will be serially immunized with autologous or allogeneic neuroblastoma cells transduced to express γ interferon. Cells will be transduced with the retroviral vector, pHuγ-IFN, encoding the human gene for γ interferon and lethally irradiated prior to use as an immunogen. The objectives of the study are: (1) to determine the maximum tolerable dose of transduced cells; (2) to determine the local, regional, and systemic toxicities of injected cells; and (3) to determine the antitumor response *in vivo* as measured by standard clinical tests and immunocytologic evaluation of marrow metastases. (Protocol #9403-068)

Appendix D-82. Dr. Kenneth L. Brigham of Vanderbilt University, Nashville, Tennessee, may conduct gene transfer experiments on 10 subjects (≤ 21 years of age) in two different patient protocols (5 for each protocol). Both protocols will use the same DNA/liposome preparations to deliver a plasmid DNA construct, pCMV4-AAT, encoding human alpha-1 antitrypsin gene driven by a cytomegalovirus promoter. In patients scheduled for elective pulmonary resection, the DNA/liposome complexes will be instilled through a fiber optic bronchoscope into a subsegment of the lung. Tissues of the lung will be obtained at the time of surgery. Transgene expression will be assessed by immunohistochemistry, *in situ* hybridization, and Western and Northern blot analyses. The effect of DNA/liposome complex administration on the histological appearance of the lung will also be evaluated. In patients with alpha-1 antitrypsin deficiency, the DNA/liposome complexes will be instilled into the nostril. Transgene expression will be determined in cells obtained by nasal lavage and nasal scraping, and the time course of transgene expression will be measured. The secretion of the alpha-1 antitrypsin protein in nasal fluid will be determined. Histological appearance of nasal mucosa will also be examined. The study will assess safety and feasibility of gene delivery to the human respiratory tract. (Protocol #9403-070)

Appendix D-83. Dr. H. Kim Lyerly of Duke University Medical Center, Durham, North Carolina, may conduct gene transfer experiments on 20 subjects with refractory or recurrent metastatic breast cancer (≥ 18 years of age). Autologous breast cancer cells will be transduced with the DNA/liposome complex, pMP6-IL2, containing a plasmid DNA vector derived from adeno-associated virus (AAV) that expresses the gene for human interleukin-2. Subjects will receive 4 subcutaneous injections of lethally irradiated tumor cells transduced with the DNA/liposome complex prior to injection. The objective of this study is to: (1) evaluate the safety and toxicity of the treatment, (2) determine the immunological effects, (3) determine the duration of clinical responses, and (4) measure patient survival. (Protocol #9409-086)

Appendix D-84. Drs. Flossie Wong-Staal, Eric Poeschla, and David Looney of the University of California at San Diego, La Jolla, California, may conduct gene transfer experiments on 6 subjects (≥ 18 and ≤ 65 years of age) infected with human immunodeficiency virus-1 (HIV-1). Autologous CD4(+) T lymphocytes will be transduced *ex vivo* with the retroviral vector, pMJT, expressing a hairpin ribozyme that cleaves the HIV-1 RNA in the 5' leader sequence. The transduced cells will be expanded and reinfused into the patients. The objectives of the study are: (1) to evaluate safety of reinfusing the transduced lymphocytes, (2) to compare (*in vivo*) the kinetics and survival of ribozyme-transduced cells with that of cells transduced with a control vector, (3) to determine *in vivo* expression of the ribozyme sequences in transduced lymphocytes, (4) to determine whether host immune responses directed against the transduced cells will occur *in vivo*, and (5) to obtain preliminary data on the effects of ribozyme gene therapy on *in vivo* HIV mRNA expression, viral burden and CD4(+) lymphocyte levels. (Protocol #9309-057)

Appendix D-85. Dr. Friedrich Schuening of the Fred Hutchinson Cancer Research Center, Seattle, Washington, may conduct gene transfer experiments on 10 subjects (≥ 18 years of age) with Type I Gaucher's disease. The peripheral blood repopulating cells (mobilized by patient pretreatment with recombinant granulocyte colony-stimulating factor) will be harvested and CD34(+) cells selected. CD34(+) cells will be transduced *ex vivo* with the retroviral vector, LgGC, that encodes human glucocerebrosidase cDNA. Following transduction, the transduced cells will be infused into the patient without myeloablative treatment. The primary endpoint of this study is to examine the safety of infusing CD34(+) cells transduced with the human glucocerebrosidase cDNA. Patients will be monitored for persistence and expression of the glucocerebrosidase gene in hematopoietic cells. (Protocol #9312-061)

Appendix D-86. Dr. Terence R. Flotte of the Johns Hopkins Children's Center, Baltimore, Maryland, may conduct gene transfer experiments on 16 subjects (≥ 18 years of age) with mild cystic fibrosis (CF). An adeno-associated virus (AAV) derived vector, encoding cystic fibrosis transmembrane conductance regulator (CFTR) gene, (tgAAVCF), will be administered to nasal (direct) and airway (bronchoscope) epithelial cells. This is a dose escalation study involving 8 cohorts. Each subject will receive both intranasal and bronchial administration of the adenoviral vector at 4 escalating doses. Nasal doses will range between 1×10^6 and 1×10^9 pfu. Lung administration will range between 1×10^7 and 1×10^{10} pfu. The primary goal of the study is to assess the safety of vector administration. Respiratory and nasal epithelial cells will be evaluated for gene transfer, expression, and physiologic correction. (Protocol #9409-083)

Appendix D-87. Drs. Jeffrey M. Isner and Kenneth Walsh of St. Elizabeth's Medical Center, Tufts University, Boston, Massachusetts, may conduct gene transfer experiments on 12 subjects (≥ 40 years of age) with peripheral artery disease (PAD). A plasmid DNA vector, phVEGF165, encoding the human gene for vascular endothelial growth factor (VEGF) will be used to express VEGF to induce collateral neovascularization. Percutaneous arterial gene transfer will be achieved using an angioplasty catheter with a hydrogel coated balloon to deliver the plasmid DNA vector to the artery. The objectives of the study are: (1) to determine the efficacy of arterial gene therapy to relieve rest pain and/or heal ischemic ulcers of the lower extremities in patients with PAD; and (2) to document the safety of the phVEGF arterial gene therapy for therapeutic angiogenesis. Subjects will undergo anatomic and physiologic examination to determine the extent of collateral artery development following phVEGF arterial gene therapy. (Protocol #9409-088)

Appendix D-88A. Dr. Ronald G. Crystal of New York Hospital-Cornell Medical Center, New York, New York, may conduct gene transfer experiments on 26 patients (≥ 15 years of age) with cystic fibrosis (CF). A replication deficient recombinant adenovirus vector will be used to transduce epithelial cells of the large bronchi with the E1/E3 deleted type 5 adenovirus vector, Ad_{GV}CFTR.10, which encodes the human cystic fibrosis transmembrane conductance regulator (CFTR) gene. The objective of this study is to define the safety and pharmacodynamics of CFTR gene expression in airway epithelial cells following single administration of

escalating doses to the vector. If single administration is determined to be safe, subjects will undergo repeat administration to localized areas of the bronchi. (Protocol #9409-085)

Appendix D-88B. Drs. Eric J. Sorscher and James L. Logan of the University of Alabama, Birmingham, Alabama, may conduct gene transfer experiments on 9 subjects (≥18 years of age) with cystic fibrosis (CF). The normal human cystic fibrosis transmembrane conductance regulator (CFTR) gene will be expressed by a plasmid DNA vector, pKCTR, driven by the simian virus-40 (SV40) early gene promoter. The CFTR DNA construct will be delivered by cationic liposome-based gene transfer to nasal epithelial cells. The objectives of the study are to: (1) evaluate the safety of lipid-mediated gene transfer to nasal epithelial cells (including local inflammation and mucosal tissue); and (2) evaluate efficacy as determined by correction of the chloride ion transport defect, and wild-type CFTR mRNA and protein expression. (Protocol #9312-066)

Appendix D-89. Dr. Steven M. Albelda of the University of Pennsylvania Medical Center, Philadelphia, Pennsylvania, may conduct gene transfer experiments on 12 subjects with advanced mesothelioma. The adenovirus vector encoding the *Herpes simplex* virus thymidine kinase (HSV-TK) gene, H5.020RSVTK, will be administered through a chest tube to the pleural cavity. Tumor biopsies will be assayed for gene transfer and expression. Subjects will be monitored for immunological responses to the adenovirus vector. Ganciclovir will be administered intravenously 14 days following vector administration. The primary objective of this Phase I study is to evaluate the safety of direct adenovirus vector gene delivery to the pleural cavity of patients with malignant melanoma. (Protocol #9409-090)

Appendix D-90. Drs. Jeffrey Holt and Carlos B. Arteaga of the Vanderbilt University, Nashville, Tennessee, may conduct gene transfer experiments on 10 female patients (over 18 years of age) with metastatic breast cancer. Patient effusions from pleura or peritoneum will be drained and the fluid will be replaced with supernatant containing the retroviral vectors, XM6:antimyc or XM6:antifos, which express c-*myc* and c-*fos* antisense sequences, respectively, under the control of a mouse mammary tumor virus promoter. The objectives of this study are to: (1) assess uptake and expression of the vector sequences in breast cancer cells present in pleural and peritoneal fluids, and determine if this expression is tumor specific, (2) assess the safety of localized administration of antisense retroviruses, and (3) monitor subjects for clinical evidence of antitumor response. (Protocol #9409-084)

Appendix D-91. Dr. Jack A. Roth of MD Anderson Cancer Center, Houston, Texas, may conduct gene transfer experiments on 14 non-small cell lung cancer subjects (≥ 18 and ≤ 80 years of age) who have failed conventional therapy and who have bronchial obstruction. LNSX-based retroviral vectors containing the β-actin promoter will be used to express: (1) the antisense RNA of the K-*ras* oncogene (LN-K-*ras*B), and (2) the wildtype *p53* tumor suppressor gene (LN*p53*B). Tumor biopsies will be obtained to characterized K-*ras* and *p53* mutations. Relative to their specific mutation, subjects will undergo partial endoscopic resection of the tumor bed followed by bronchoscopic administration of the appropriate retrovirus construct. The objective of this study is to evaluate the safety and efficacy of intralesional administration of LN-K-*ras*B and LN*p53* retrovirus constructs. (Protocol #9403-031)

Appendix D-92. Drs. Robert E. Sobol and Ivor Royston of the San Diego Regional Cancer Center, San Diego, California, may conduct gene transfer experiments on 12 subjects (≥ 18 years of age) with metastatic colon carcinoma. The autologous skin fibroblasts will be transduced with the retroviral vector, LNCX/IL-2, which encodes the gene for human interleukin-2 (IL-2). In this dose-escalation study, subjects will receive subcutaneous injections of lethally irradiated autologous tumor cells. The objectives of the study are to: (1) evaluate the safety of subcutaneous administration of LNCX/IL-2 transduced fibroblasts, (2) determine *in vivo* antitumor activity, and (3) monitor cellular and humoral antitumor responses. (Protocol #9312-060)

Appendix D-93. Dr. Michael Lotze of the University of Pittsburgh, Pittsburgh, Pennsylvania, may conduct gene transfer experiments on 18 subjects (≥ 18 years of age) with advanced melanoma, 6 with T-cell lymphoma, breast cancer, or head and neck cancer. Subjects should have accessible cutaneous tumors, and have failed standard therapy. Over 4 weeks, subjects will receive a total of 4 intratumoral injections of autologous fibroblasts transduced with the retrovirus vector, TFG-hIL-12-Neo. This vector, which consists of the murine MFG backbone, expresses both the p35 and p40 subunits of interleukin-12 (IL-12) and the *neo*^{R} selection marker. The objectives of the study are to: (1) define the local and systemic toxicity associated with peritumoral injections of gene-modified fibroblasts, (2) examine the local and systemic immunomodulatory effects of these injections, and (3) evaluate clinical antitumor efficacy. (Protocol #9406-081)

Appendix D-94. Drs. Evan Hersh, Emmanuel Akporiaye, David Harris, Alison Stopeck, Evan Unger, James Warneke, of the Arizona Cancer Center, Tucson, Arizona, may conduct gene transfer experiments on 25 subjects (≥ 18 years of age) with solid malignant tumors or lymphomas. A plasmid DNA/lipid complex designated as VCL-1102 (IL-2 Plasmid DNA/DMRIE/DOPE) will be used to transduce the human gene for interleukin-2 (IL-2). Patients with advanced cancer who have failed conventional therapy will undergo a procedure in which VCL-1102 is injected directly into the tumor mass to induce tumor-specific immunity. The objectives of the study are to: (1) determine safety and toxicity associated with escalating doses of VCL-1102; (2) confirm IL-2 expression in target cells; (3) determine biological activity and pharmacokinetics; and (4) determine whether IL-2 expression stimulates tumor regression in subjects with metastatic malignancies. (Protocol #9412-095)

Appendix D-95. Drs. Richard Morgan and Robert Walker of the National Institutes of Health, Bethesda, Maryland, may conduct gene transfer experiments on 48 human immunodeficiency virus (HIV) seropositive subjects (≥ 18 years of age). This Phase I/II study involves identical twins (one HIV seropositive and the other HIV seronegative). CD4(+) T cells will be enriched following apheresis of the HIV seronegative twin, induced to polyclonal proliferation with anti-CD3 and recombinant IL-2, transduced with either the LNL6/Neo^{R} or G1Na/Neo^{R}, and transduced with up to 2 additional retroviral vectors (G1RevTdSN and/or GCRTdSN(TAR)) containing potentially therapeutic genes (antisense TAR and/or transdominant Rev). These T cell populations will be expanded 10 to 1,000 fold in culture for 1 to 2 weeks and reinfused into the HIV seropositive twin. Subjects will receive up to 4 cycles of treatment using identical or different combinations of control and anti-HIV retrovirus vectors. The relative survival of these transduced T cell populations will be monitored by vector-specific polymerase chain reaction, while the subjects' functional immune status is monitored by standard *in vitro* and *in vivo* assays. (Protocol #9503-103)

Appendix D-96. Dr. Harry L. Malech of the National Institutes of Health, Bethesda, Maryland, may conduct gene transfer experiments on 2 subjects ≥ 18 years of age (with or without concurrent serious infection), and 3 subjects ≥ 18 years of age (with or without concurrent serious infection) or minors 13-17 years of age who have concurrent serious infection who have chronic granulomatous disease (CGD). CGD is an inherited immune deficiency disorder in which blood neutrophils and monocytes fail to produce antimicrobial oxidants (p47^{phox} mutation) resulting in recurrent life-threatening infections. Subjects will undergo CD34(+) mobilization with granulocyte colony stimulating factor (G-CSF). These CD34(+) cells will be transduced with the retrovirus vector, MFG-S-p47^{phox}, which encodes the gene for normal p47^{phox}. The objectives of this study are to: (1) determine the safety of administering MFG-S-p47^{phox} transduced CD34(+) cells, and (2) demonstrate increased functional oxidase activity in circulating neutrophils. (Protocol #9503-104)

Appendix D-97. Drs. Chris Evans and Paul Robbins of the University of Pittsburgh, Pittsburgh, Pennsylvania, may conduct gene transfer experiments on 6 subjects (≥ 18 and ≤ 76 years of age) with rheumatoid arthritis. Rheumatoid arthritis is a chronic, progressive disease thought to be of autoimmune origin. A gene encoding an interleukin-1 receptor antagonist protein (IRAP) will be delivered to the rheumatoid metacarpal-phalangeal joints to determine the autoimmune reactions can be interrupted. The vector construct, DFG-IRAP, is based on the MFG murine retrovirus vector backbone, and encodes the human IRAP gene. Synovial fibroblasts will be generated from the rheumatoid arthritic joint tissue obtained from patients who are scheduled to undergo surgery. The fibroblasts will be transduced with the DFG-IRAP vector, and the transduced cells injected into the synovial space. The synovial fluid and joint material will be collected 7 days later to determine the presence and location of the transduced synovial fibroblasts and the level of IRAP in the joint fluid. (Protocol 9406-074)

Appendix D-98. Dr. R. Scott McIvor of the University of Minnesota, Minneapolis, Minnesota, may conduct gene transfer experiments on 2 children with purine nucleoside phosphorylase (PNP) deficiency. PNP deficiency results in severe T-cell immunodeficiency, an autosomal recessive inherited disease which is usually fatal in the first decade of life. Autologous peripheral blood lymphocytes will be cultured in an artificial capillary cartridge in the presence of anti-CD3 monoclonal antibody and interleukin-2 and transduced with the retroviral vector, LPNSN-2, encoding human PNP. Subjects will undergo bimonthly intravenous administration of transduced T cells for a maximum of 1 year. The objectives of the study are to determine: (1) the safety of intravenous administration of transduced T cells in children with PNP deficiency, (2) the efficiency of PNP gene transfer and duration of gene expression *in vivo*, and (3) the effect of PNP gene transfer on immune function. (Protocol #9506-110)

Appendix D-99. Drs. Nikhil C. Munshi and Bart Barlogie of the University of Arkansas School for Medical Sciences, Little Rock, Arkansas, may conduct gene transfer experiments on 21 subjects (>18 and <65 years of age) with relapsed or persistent multiple myeloma who are undergoing T cell depleted allogeneic bone marrow transplantation. Donor peripheral blood lymphocytes will be cultured in vitro with interleukin-2 and anti-CD3 monoclonal antibody. T cell depleted lymphocytes will be transduced with the retroviral construct, G1Tk1SvNa.7, which encodes the Herpes simplex virus thymidine kinase (HSV-TK) gene. The transduced cells will be reinfused. In this dose escalation study, 3 subjects will undergo cell-mediated gene transfer per cohort (maximum of 5 cohorts) until Grade III or IV Graft versus Host Disease (GVHD) is observed. A maximum of 6 additional patients may be entered at that maximum tolerated dose. The objectives of this study are to determine the: (1) safety of transduced donor cell infusions, (2) effectiveness of donor cell infusions in decreasing the effects of severe GVHD, (3) effectiveness of donor cell infusions in prolonging multiple myeloma remission, and (4) effectiveness of ganciclovir in eliminating donor cells for the purpose of preventing the depletion of erythrocytes. (Protocol #9506-107)

Appendix D-100. Dr. Wayne A. Marasco of Dana-Farber Cancer Institute, Boston, Massachusetts, may conduct gene transfer experiments on 6 subjects (≥ 18 and ≤ 65 years of age) with human immunodeficiency virus type-1 (HIV-1). Autologous lymphocytes from asymptomatic subjects will be transduced *ex vivo* with a retroviral vector, LNCs105, encoding the sFv105 antibody specific for the HIV-1 envelope protein. An identical aliquot will be simultaneously transduced with a control retroviral vector lacking the sFv105 cassette. Transduced cells will be reinfused into patients and the differential survival of both populations of CD4+ lymphocytes compared. The objective of the study is to determine whether the intracellular expression of a human single chain antibody against HIV-1 envelope glycoprotein gp160 that blocks gp160 processing and the production of infectious virions can safely prolong the survival of CD4(+) lymphocytes in HIV-1-infected subjects. (Protocol #9506-111)

Appendix D-101. Dr. Henry Dorkin of the New England Medical Center, Boston, Massachusetts, and Dr. Allen Lapey of Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, propose to conduct gene transfer experiments on 16 subjects (≥ 18 years of age). An E1/partial E4-deleted, replication-deficient, type 2 adenovirus vector, AD2/CFTR-2, will be used to deliver the human cystic fibrosis transmembrane conductance regulator (CFTR) gene by aerosol administration (nebulization) to the lung of CF patients. Aerosol administration will be initiated only after initial safety data has been obtained from the lobar administration protocol (#9409-091). This is a single administration dose-escalation study in which subjects will receive between 8×10^6 and 2.5×10^{10} pfu. Subjects will be assessed for evidence of adverse, systemic, immune, inflammatory, or respiratory effects in response to AD2/CFTR-2. Subjects will be monitored for virus shedding and transgene expression. Health care workers present in the facility will be required to sign an Informed Consent document regarding the possibility of virus transmission. (Protocol #9412-074)

Appendix D-102. Drs. Charles J. Link and Donald Moorman of the Human Gene Therapy Research Institute, Des Moines, Iowa, may conduct gene transfer experiments on 24 female subjects (≥ 18 years of age) with refractory or recurrent ovarian cancer. Subjects will undergo intraperitoneal delivery (via Tenkhoff catheter) of the vector producing cells (VPC), PA317/LTKOSN.2. These VPC express the *Herpes simplex* virus thymidine kinase (HSV-TK) gene which confers sensitivity to killing by the antiviral drug, ganciclovir (GCV). The LTKOSN.2 retrovirus vector is based on the LXSN backbone. Two weeks following intraperitoneal delivery of the VPC, subjects will receive 5 mg/kg intravenous GCV twice daily for 14 days. Subjects will receive between 1×10^5 and 1×10^8 VPC/kg in this dose escalation study. Subjects will be evaluated by X-ray and peritoneoscopy of the abdomen for evidence of clinical response. The objectives of this study are to determine the safety of intraperitoneal VPC administration. (Protocol #9503-100)

Appendix D-103. Dr. David T. Curiel of the University of Alabama, Birmingham, Alabama, may conduct gene transfer experiment of 15 subjects (≥ 18 years of age) with metastatic colorectal cancer. Subjects will receive intramuscular injection of the polynucleotide vaccine, pGT63, which is a plasmid DNA vector expressing carcinoembryonic antigen (CEA) and hepatitis B surface antigen (HBsAg). The objectives of the study are to: (1) characterize the immune response to CEA and HBsAg following a single intramuscular injection and following 3 consecutive intramuscular injections, and (2) determine the safety of intramuscular injection of the plasmid DNA vector at doses ranging between 0.1 to 1.0 milligrams (single dose) and 0.9 to 3.0 milligrams (total multidose). (Protocol #9506-073)

Appendix D-104. Dr. Chester B. Whitley of the University of Minnesota, Minneapolis, Minnesota, may conduct gene transfer experiments on two adult subjects (18 years of age or older) with mild Hunter syndrome (Mucopolysaccharidosis Type II). The autologous peripheral blood lymphocytes will be transduced *ex vivo* with the retroviral vector, L2SN, encoding the human cDNA for iduronate-2-sulfatase (IDS). The transduced lymphocytes will be reinfused into the patients on a monthly basis. The study will determine the frequency of peripheral blood lymphocyte transduction and the half-life of the infused cells. Evaluation of patients will include measurement of blood levels of IDS enzyme, assessment of metabolic correction by urinary glycosaminoglycan levels, clinical response of the disease, and monitoring for potential toxicity. This Phase I study is to demonstrate the safety of the L2SN-mediated gene therapy and to provide a preliminary evaluation of clinical efficacy. (Protocol #9409-087)

Appendix D-105. Drs. James Economou, John Glaspy, and William McBride of the University of California, Los Angeles, California, may conduct gene transfer experiments on 25 subjects (≥ 18 years of age) with metastatic melanoma. The protocol is an open label, Phase I trial to evaluate the safety and immunological effects of administering lethally irradiated allogeneic and autologous melanoma cells transduced with the retroviral vector, IL-7/HyTK, which encodes the gene for human interleukin-7 (IL-7). Subjects will receive 1 x 10^7 irradiated unmodified autologous tumor cells in combination with escalating doses of IL-7/HyTK transduced allogeneic melanoma cells (M24 cell line). The number of M24 cells administered will be adjusted based on the level of IL-7 expression. Subjects will receive 3 biweekly subcutaneous injections of M24 cells expressing 10, 100, or 1000 nanograms of IL-7/hour *in vivo*. A final cohort of 5 subjects will receive IL-7/HyTK transduced autologous cells. Subjects will be monitored for antitumor activity by skin tests, biopsy analysis, tumor-specific antibody activity, and cytotoxic T lymphocyte precursor evaluation. Non-immunologic parameters will also be monitored. (Protocol #9503-101)

Appendix D-106. Dr. Jack A. Roth, MD Anderson Cancer Center, may conduct gene transfer experiments on 42 subjects (≥ 18 years of age) with refractory non-small cell lung cancer (NSCLC). Subjects will receive direct intratumoral injection of a replication-defective type 5 adenovirus vector, AD5CMV-p53, to deliver the normal human *p53* tumor suppressor gene. The E1 region of AD5CMV-p53 has been replaced with a *p53* expression cassette containing the human cytomegalovirus promoter (CMV). Subjects will be divided into 2 treatment groups: (1) 21 subjects will receive Ad5CMV-p53 alone, and (2) 21 subjects will receive Ad5CMV-p53 in combination with cisplatin. Following vector administration, subjects will be isolated for 96 hours during which time, assays will be conducted to demonstrate the lack of shedding of adenovirus vector. The objectives of this study are determine: (1) the maximum tolerated dose of AD5CMV-p53, (2) qualitative and quantitative toxicity related to vector administration, and (3) biologic activity.

Prior to administration, adenovirus vector stocks will be screened for p53 mutants using the SAOS osteosarcoma cell assay that was submitted by Dr. Roth on June 23, 1995. This biologic assay compares the activity of a standard stock of Adp53 vector to the activity of newly produced stocks. The standard stock of Adp53 will be defined as mediating cell death in 100% of SAOS cells (human osteosarcoma cell line with homozygous p53 deletion) at an MOI of 50:1 (titer > 5 x 10^{10}) on day 5 of culture. The sensitivity of the assay for detecting inactive (presumed mutant) Adp53 vector will be determined by adding increasing amounts of Adluc (control adenovirus vector containing the luciferase gene) to the Adp53 stock to determine the percentage of inactive vector required to decrease growth inhibition of SAOS cells mediated by Adp53. The test lot of Adp53 will be tested for its ability to inhibit SAOS in a 5 day assay. Significant loss of inhibitory activity compared with the standard would indicate an unacceptable level of inactive (presumed mutant) vector. (Protocol #9406-079)

Appendix D-107A. Dr. Gary Clayman. M.D. Anderson Cancer Center, Houston, Texas, may conduct gene transfer experiments on 21 subjects (≥ 18 years of age) with refractory squamous cell carcinoma of the head and neck. Subjects will receive direct intratumoral injection of a replication-defective type 5 adenovirus vector, AD5CMV-p53, to deliver the normal human *p53* tumor suppressor gene. The E1 region of AD5CMV-p53 has been replaced with a *p53* expression cassette containing the human cytomegalovirus promoter (CMV). Subjects will be divided into 2 treatment groups: (1) those with non-resectable tumors, and (2) those with surgically accessible tumors. Subjects will receive multiple injections of vector in each dose-escalation cohort. Following vector administration, subjects will be isolated for 48 hours during which time, assays will be conducted to demonstrate the lack of shedding of adenovirus vector. The objectives of the study are to determine: (1) the maximum tolerated dose of AD5CMV-p53, (2) qualitative and quantitative toxicity related to vector administration, and (3) biologic activity.

Prior to administration, adenovirus vector stocks will be screened for p53 mutants using the SAOS osteosarcoma cell assay that was submitted by Dr. Roth on June 23, 1995. This biologic assay compares the activity of a standard stock of Adp53 vector to the activity of newly produced stocks. The standard stock of Adp53 will be defined as mediating cell death in 100% of SAOS cells (human osteosarcoma cell line with homozygous p53 deletion) at an MOI of 50:1 (titer > $5x10^{10}$) on day 5 of culture. The sensitivity of the assay for detecting inactive (presumed mutant) Adp53 vector will be determined by adding increasing amounts of Adluc to the Adp53 stock to determine the percentage of inactive vector required to decrease growth inhibition of SAOS cells mediated by Adp53. The test lot of Adp53 will be tested for its ability to inhibit SAOS in a 5 day assay. Significant loss of inhibitory activity compared with the standard would indicate an unacceptable level of inactive (presumed mutant) vector. (Protocol #9412-096)

Appendix D-107B. Drs. Bernard A. Fox and Walter J. Urba of Earle A. Chiles Research Institute, Providence Medical Center, Portland, Oregon, may conduct gene transfer experiments on 18 subjects (≥ 18 years of age) with metastatic renal cell carcinoma or melanoma. Autologous tumor cells will be surgically removed, transduced *in vitro* with the cationic liposome plasmid vector, VCL-1005, which encodes human leukocyte antigen (HLA)-B7 and beta-2 microglobulin. Subjects will receive subcutaneous injection of lethally irradiated transduced cells in one limb. The contralateral limb will be injected with lethally irradiated untransduced tumor cells in combination with Bacille Calmette-Guerin (BCG). Approximately 21 days following tumor cell injection, subjects will undergo lymphadenectomy for subsequent in vitro expansion of anti-CD3 activated lymphocytes. Activated lymphocytes will be adoptively transferred on approximately day 35 in combination with a 5-day course of interleukin-2 (IL-2). On approximately day 45, subjects will receive a second cycle of IL-2. The objectives of this study are to determine: (1) the safety of administering anti-CD3 activated antitumor effector T cells in draining lymph nodes, and (2) whether HLA-B7/β-2 gene transfer augments the sensitization of anti-tumor effector T-cells in draining lymph nodes. (Protocol 9506-108)

Appendix D-108. Dr. Mitchell S. Steiner, University of Tennessee, Memphis, Tennessee, and Dr. Jeffrey T. Holt, Vanderbilt University School of Medicine, Nashville, Tennessee, may conduct gene transfer experiments on 15 male subjects (35 to 75 years of age) with metastatic prostate cancer. Malignant cells obtained from advanced prostate cancer subjects have been demonstrated to express high levels of the protooncogene c-myc *in vivo*. The mouse mammary tumor virus (MMTV) long terminal repeat (LTR) is expressed at high levels in prostate tissue. Following removal of malignant cells via biopsy, subjects will receive a single transrectal ultrasound-guided intraprostate quadrant injection of the retrovirus vector, XM6:MMTV-antisense c-myc, for 4 consecutive days at the site of the original biopsy. The objectives of this Phase I study are to: (1) quantitatively assess the uptake and expression of XM6:MMTV-antisense c-myc by prostate cancer cells *in vivo*, (2) determine whether c-myc gene expression is prostate tumor-specific, (3) assess safety of intraprostate injection of XM6:MMTV-antisense c-myc, and (4) biologic efficacy (antisense inhibition of tumor growth). (Protocol #9509-123)

Appendix D-109. Drs. Ronald G. Crystal, Edward Hershowitz, and Michael Lieberman, New York Hospital-Cornell Medical Center, New York, New York, may conduct gene transfer experiments on 18 subjects (18 to 70 years of age) with metastatic colon carcinoma with liver metastases. In this Phase I dose-escalation study, subjects will receive computed tomography (CT)-guided intratumoral injections of the adenovirus vector, Ad_{GV}CD.10, into the same hepatic metastasis in 4 equal volumes (100 microliters), each with a separate entry into the liver. This dosage schedule will be performed on Days 1 and 7. 5-fluorocytosine (200 milligrams/ kilogram/24 hours) will be administered orally in 4 equal doses starting on day 2 and continuing through the time of laparotomy. The objectives of this study are to: (1) determine the dose-dependent toxicity of direct administration of Ad_{GV}CD.10 to hepatic metastases combined with oral administration of 5-fluorocytosine, (2) quantitatively assess transfer and expression of the cytosine deaminase gene in target cells, and (3) determine the biologic effects of direct AD_{GV}CD.10 administration to hepatic metastases. (Protocol #9509-125)

Appendix D-110. Drs. Andres Berchuck and H. Kim Lyerly of Duke University Medical Center, Durham, North Carolina, may conduct gene transfer experiments on 18 subjects (≥ 18 years of age) with refractory metastatic ovarian cancer. Autologous tumor cells obtained from ascites or surgically removed tumor will be transduced with the cationic liposome vector, PMP6A-IL2, that contains an adeno-associated virus derived plasmid DNA, a cytomegalovirus (CMV) promoter, and interleukin-2 (IL-2) complementary DNA (cDNA). In this dose-escalation study, subjects will undergo 4 cycles of intradermal injections (thigh or abdomen) of *ex vivo* transduced, lethally irradiated tumor cells in an attempt to induce an antitumor response. The objectives of the study are to evaluate: (1) the safety of intradermally injected transduced cells, and (2) antitumor response following therapy. (Protocol #9506-110)

Appendix D-111. Drs. Stephen L. Eck and Jane B. Alavi of the University of Pennsylvania Medical Center, Philadelphia, Pennsylvania, may conduct gene transfer experiments on 18 subjects (>18 years of age) with malignant glioma. The adenovirus vector encoding the *Herpes simplex* virus thymidine kinase (HSV-TK) gene, H5.020RSVTK, will be injected by a stereotactic guided technique into brain tumors. Afterwards, the patients will receive systemic ganciclovir (GCV) treatment. Patients eligible to undergo a palliative debulking procedure will receive the same treatment followed by resection on day 7, and a second dose of the vector intra-operatively. Brain tissues removed by resection will be analyzed for adenovirus infection, transgene expression, and signs of inflammation. The size and metabolic activity of tumors will be monitored by scanning with magnetic resonance imaging and positron emission tomography. The objective of the study is to evaluate the overall safety of this treatment and to gain insight into the parameters that may limit the general applicability of this approach. (Protocol #9409-089)

Appendix D-112. Drs. Robert Grossman and Savio Woo of the Baylor College of Medicine & Methodist Hospital, Houston, Texas, may conduct gene transfer experiments on 20 subjects (≥ 18 years of age) with refractive central nervous system malignancies. Subjects will receive stereotaxic injections of a replication-defective, type 5 E1/E3-deleted adenovirus vector, ADV/RSV-tk, to deliver the *Herpes simplex* virus thymidine kinase (HSV-TK) gene to tumor cells. Expression of the HSV-TK gene is driven by a Rous sarcoma virus long terminal repeat (RSV-LTR). Subjects will receive a single time-course of intravenous ganciclovir (GCV) (14 consecutive days) following vector administration. Following demonstration of safety with the initial starting dose of 1×10^8 particles in 5 subjects, additional cohorts will receive between 5×10^8 and 1.5×10^9 particles. Each cohort will be monitored for toxicity for one month before administration of the next higher dose to subsequent cohorts. Subjects will be monitored for evidence of clinical efficacy by magnetic resonance imaging and/or computer tomography scans. The primary objective of this Phase I study is to determine the safety of vector administration. (Protocol #9412-098)

Appendix D-113. Drs. Gabriel N. Hortobagyi, Gabriel Lopez-Berestein, and Mien-Chie Hung, of the University of Texas MD Anderson Cancer Center, Houston, Texas, may conduct gene transfer experiments on a maximum of 24 adult patients (12 for each cancer) with metastatic breast or ovarian carcinoma. Overexpression of the HER-2/*neu* oncogene occurs in 30% of ovarian and breast cancers, and it is associated with enhanced metastatic potential, drug resistance, and poor survival. The E1A gene of the adenovirus type 5 functions as a tumor suppressor gene when transfected into cancer cells which overexpress the HER-2/*neu* oncogene. E1A expression induces down regulation of the level of the HER-2/*neu* oncoprotein by a transcriptional control mechanism. A plasmid, pE1A, encoding the adenovirus E1A gene with its own promoter will be administered as a DNA/lipid complex via the intraperitoneal or intrapleural route. The objectives of the study are: (1) to determine E1A gene transduction into malignant cells after the administration of E1A/lipid complex by intrapleural or intraperitoneal administration, (2) to determine whether E1A gene therapy can down-regulate HER-2/*neu* expression after intrapleural or intraperitoneal administration, (3) to determine the maximum biologically active dose (MBAD), or the maximum tolerated dose (MTD) of E1A/lipid complex, (4) to determine the toxicity and tolerance of E1A/lipid complex administered into the pleural or peritoneal space, and to assess the reversibility of such toxicity, and (5) to evaluate tumor response. (Protocol #9512-137)

Appendix D-114. Drs. Keith L. Black and Habib Fakhrai of the University of California, Los Angeles, California, may conduct gene transfer experiments on 12 subjects (≥ 18 years of age) with glioblastoma multiform. An Epstein-Barr virus (EBV) based plasmid vector, pCEP-4/TGF- β2 antisense, encoding antisense RNA will be used to inhibit TGF- β2 production. Tumor samples obtained from the patients at the time of clinically indicated surgery will be grown in culture to establish a cell line for each patient. The patients' tumor cells will be genetically altered with the pCEP-4/TGF-β2 vector to inhibit their secretion of TGF-β. Following completion of the traditional post surgical radiation therapy, the first cohort of patients will receive, at 3 week intervals, 4 injections of 5×10^6 irradiated gene modified autologous tumor cells. Subsequently, in dose escalation studies, the second cohort will receive 1×10^7 cells, and the third cohort, 2×10^7 cells. The results of this Phase I trial will be used to assess the safety of this form of gene therapy and may provide preliminary data to evaluate the potential utility of TGF- β2 antisense gene therapy in the management of gliomas. (Protocol #9512-138)

Appendix D-115. Dr. Ronald G. Crystal of New York Hospital-Cornell Medical Center, New York, New York, may conduct gene transfer experiments on a total of 21 (with an option for an additional 5) normal males and female subjects, age ≥ 18 years. Replication-deficient adenovirus (Ad) vector previously has been used in a number of human gene therapy strategies to transfer genes *in vivo* for therapeutic purposes. The purpose of this protocol is to characterize the local (skin), systemic (blood), and distant compartment (lung) immunity in normal individuals after intradermal administration of a replication deficient Ad5-based vector, named Ad_{GV}CD.10, carrying the gene coding for the *E. coli* enzyme, cytosine deaminase (CD). Following intradermal

administration of the vector to normal individuals, the skin, blood, and lung immune responses to the Ad vector and CD transgene will be evaluated over time. This vector has been safety administered intrahepatically ten times to five individuals with colon carcinoma. No adverse effects in Protocol #9509-125 have been observed. The present protocol will yield insights into normal human immune responses to both the Ad vector, as well as to a heterologous (i.e., non-human) gene product (CD). Note: This study is designed to answer basic biological questions regarding characterization of the immune responses to such vectors that have been previously documented. (Protocol #9701-171)

Appendix D-116. Dr. Daniel Rockey at Oregon State University and Dr. Walter Stamm at the University of Washington may conduct experiments to deliberately transfer a gene encoding tetracycline resistance from *Chlamydia suis* (a swine pathogen) into *C. trachomatis* (a human pathogen). This approval is specific to Drs. Rockey and Stamm and research with these resistant organisms may only occur under the conditions as specified by the NIH Director (72 FR 61661). This approval was effective as of September 24, 2007.

Appendix D-117. Dr. David Walker at the University of Texas Medical Branch may conduct experiments to deliberately introduce a gene encoding chloramphenicol resistance into *Rickettsia conorii*. This approval is specific to Dr. Walker and research with these resistant organisms may only occur under the conditions as specified by the NIH Director (73 FR 32719). This approval was effective as of April 7, 2008.

Appendix D-118. Dr. Harlan Caldwell at the Rocky Mountain Laboratories may conduct experiments to deliberately introduce a gene encoding tetracycline resistance into *Chlamydia trachomatis* serovar L2. This approval is specific to Dr. Caldwell and research with this resistant organism may only occur under the conditions as specified by the NIH Director (76 FR 27653). This approval was effective as of April 26, 2010.

****

APPENDIX E. CERTIFIED HOST-VECTOR SYSTEMS (See Appendix I, Biological Containment)

While many experiments using *Escherichia coli* K-12, *Saccharomyces cerevisiae*, and *Bacillus subtilis* are currently exempt from the *NIH Guidelines* under Section III-F, *Exempt Experiments*, some derivatives of these host-vector systems were previously classified as Host-Vector 1 Systems or Host-Vector 2 Systems. A listing of those systems follows:

Appendix E-I. *Bacillus subtilis*

Appendix E-I-A. *Bacillus subtilis* Host-Vector 1 Systems

The following plasmids are accepted as the vector components of certified *B. subtilis* systems: pUB110, pC194, pS194, pSA2100, pE194, pT127, pUB112, pC221, pC223, and pAB124. *B. subtilis* strains RUB 331 and BGSC 1S53 have been certified as the host component of Host-Vector 1 systems based on these plasmids.

Appendix E-I-B. *Bacillus subtilis* Host-Vector 2 Systems

The asporogenic mutant derivative of *Bacillus subtilis*, ASB 298, with the following plasmids as the vector component: pUB110, pC194, pS194, pSA2100, pE194, pT127, pUB112, pC221, pC223, and pAB124.

Appendix E-II. Saccharomyces cerevisiae

Appendix E-II-A. *Saccharomyces cerevisiae* Host-Vector 2 Systems

The following sterile strains of *Saccharomyces cerevisiae*, all of which have the ste-VC9 mutation, SHY1, SHY2, SHY3, and SHY4. The following plasmids are certified for use: YIp1, YEp2, YEp4, YIp5, YEp6, YRp7, YEp20, YEp21, YEp24, YIp25, YIp26, YIp27, YIp28, YIp29, YIp30, YIp31, YIp32, and YIp33.

Appendix E-III. *Escherichia coli*

Appendix E-III-A. *Escherichia coli* (EK2) Plasmid Systems

The *Escherichia coli* K-12 strain chi-1776. The following plasmids are certified for use: pSC101, pMB9, pBR313, pBR322, pDH24, pBR325, pBR327, pGL101, and pHB1. The following *Escherichia coli*/S. cerevisiae hybrid plasmids are certified as EK2 vectors when used in *Escherichia coli* chi-1776 or in the sterile yeast strains, SHY1, SHY2, SHY3, and SHY4: YIpI, YEp2, YEp4, YIp5, YEp6, YRp7, YEp20, YEp21, YEP24, YIp25, YIp26, YIp27, YIp28, YIp29, YIp30, YIp31, YIp32, and YIp33.

Appendix E-III-B. *Escherichia coli* (EK2) Bacteriophage Systems

The following are certified EK2 systems based on bacteriophage lambda:

Vector	Host
λgt *WES*λB'	DP50*sup*F
λgt *WES*λB*	DP50*sup*F
λgt Z*J vir*λB'	*Escherichia coli* K-12
λgtALO·λB	DP50*sup*F
Charon 3A	DP50 or DP50*sup*F
Charon 4A	DP50 or DP50*sup*F
Charon 16A	DP50 or DP50*sup*F
Charon 21A	DP50*sup*F
Charon 23A	DP50 or DP50*sup*F
Charon 24A	DP50 or DP50*sup*F

Escherichia coli K-12 strains chi-2447 and chi-2281 are certified for use with lambda vectors that are certified for use with strain DP50 or DP50*sup*F provided that the *su*-strain not be used as a propagation host.

Appendix E-IV. *Neurospora crassa*

Appendix E-IV-A. *Neurospora crassa* Host-Vector 1 Systems

The following specified strains of *Neurospora crassa* which have been modified to prevent aerial dispersion:

In1 (inositol-less) strains 37102, 37401, 46316, 64001, and 89601. Csp-1 strain UCLA37 and csp-2 strains FS 590, UCLA101 (these are conidial separation mutants).

Eas strain UCLA191 (an "easily wettable" mutant).

Appendix E-V. *Streptomyces*

Appendix E-V-A. *Streptomyces* Host-Vector 1 Systems

The following *Streptomyces* species: *Streptomyces coelicolor, S. lividans, S. parvulus*, and *S. griseus*. The following are accepted as vector components of certified *Streptomyces* Host-Vector 1 systems: *Streptomyces* plasmids SCP2, SLP1.2, pIJ101, actinophage phi C31, and their derivatives.

Appendix E-VI. *Pseudomonas putida*

Appendix E-VI-A. *Pseudomonas putida* Host-Vector 1 Systems

Pseudomonas putida strains KT2440 with plasmid vectors pKT262, pKT263, and pKT264.

****

APPENDIX F. CONTAINMENT CONDITIONS FOR CLONING OF GENES CODING FOR THE BIOSYNTHESIS OF MOLECULES TOXIC FOR VERTEBRATES

Appendix F-I. General Information

Appendix F specifies the containment to be used for the deliberate cloning of genes coding for the biosynthesis of molecules toxic for vertebrates. The cloning of genes coding for molecules toxic for vertebrates that have an LD_{50} of < 100 nanograms per kilograms body weight (e.g., microbial toxins such as the botulinum toxins, tetanus toxin, diphtheria toxin, *Shigella dysenteriae* neurotoxin) are covered under Section III-B-1 (*Experiments Involving the Cloning of Toxin Molecules with LD_{50} of Less than 100 Nanograms Per Kilogram Body Weight*) and require Institutional Biosafety Committee and NIH OSP approval before initiation. No specific restrictions shall apply to the cloning of genes if the protein specified by the gene has an LD_{50} ≥ 100 micrograms per kilograms of body weight. Experiments involving genes coding for toxin molecules with an LD_{50} of < 100 micrograms per kilograms and > 100 nanograms per kilograms body weight require Institutional Biosafety Committee approval and registration with NIH OSP prior to initiating the experiments. A list of toxin molecules classified as to LD_{50} is available from NIH OSP. Testing procedures for determining toxicity of toxin molecules not on the list are available from the Office of Science Policy, National Institutes of Health, preferably by submitting a request for this information to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). The results of such tests shall be forwarded to NIH OSP, which will consult with *ad hoc* experts, prior to inclusion of the molecules on the list (see Section IV-C-1-b-(2)-(c), *Minor Actions*).

Appendix F-II. Cloning of Toxin Molecule Genes in *Escherichia coli* K-12

Appendix F-II-A. Cloning of genes coding for molecules toxic for vertebrates that have an LD_{50} of >100 nanograms per kilograms and <1000 nanograms per kilograms body weight (e.g., abrin, *Clostridium perfringens* epsilon toxin) may proceed under Biosafety Level (BL) 2 + EK2 or BL3 + EK1 containment conditions.

Appendix F-II-B. Cloning of genes for the biosynthesis of molecules toxic for vertebrates that have an LD_{50} of >1 microgram per kilogram and <100 microgram per kilogram body weight may proceed under BL1 + EK1 containment conditions (e.g., *Staphylococcus aureus* alpha toxin, *Staphylococcus aureus* beta toxin, ricin, *Pseudomonas aeruginosa* exotoxin A, *Bordetella pertussis* toxin, the lethal factor of *Bacillus anthracis*, the *Pasteurella pestis* murine toxins, the oxygen-labile hemolysins such as streptolysin O, and certain neurotoxins present in snake venoms and other venoms).

Appendix F-II-C. Some enterotoxins are substantially more toxic when administered enterally than parenterally. The following enterotoxins shall be subject to BL1 + EK1 containment conditions: cholera toxin, the heat labile toxins of *Escherichia coli*, *Klebsiella*, and other related proteins that may be identified by neutralization with an antiserum monospecific for cholera toxin, and the heat stable toxins of *Escherichia coli* and of *Yersinia enterocolitica*.

Appendix F-III. Cloning of Toxic Molecule Genes in Organisms Other Than *Escherichia coli* K-12

Requests involving the cloning of genes coding for toxin molecules for vertebrates at an LD_{50} of <100 nanograms per kilogram body weight in host-vector systems other than *Escherichia coli* K-12 will be evaluated by NIH OSP in consultation with *ad hoc* toxin experts (see Sections III-B-1, *Experiments Involving the Cloning of Toxin Molecules with LD_{50} of Less than 100 Nanograms Per Kilogram Body Weight*, and IV-C-1-b-(2)-(c), *Minor Actions*).

Appendix F-IV. Specific Approvals

An updated list of experiments involving the deliberate formation of recombinant or synthetic nucleic acid molecules containing genes coding for toxins lethal for vertebrates at an LD_{50} of <100 nanograms per kilogram body weight is available from the Office of Science Policy, National Institutes of Health, preferably by submitting a request for this information to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

****

APPENDIX G. PHYSICAL CONTAINMENT

Appendix G specifies physical containment for standard laboratory experiments and defines Biosafety Level 1 through Biosafety Level 4. For large-scale (over 10 liters) research or production, Appendix K (*Physical Containment for Large Scale Uses of Organisms Containing Recombinant or Synthetic Nucleic Acid Molecules*) supersedes Appendix G. Appendix K defines Good Large Scale Practice through Biosafety Level 3 - Large Scale. For certain work with plants, Appendix P (*Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Plants*) supersedes Appendix G. Appendix P defines Biosafety Levels 1 through 4 - Plants. For certain work with animals, Appendix Q (*Physical and Biological Containment for Recombinant or Synthetic Nucleic Acid Molecule Research Involving Animals*) supersedes Appendix G. Appendix Q defines Biosafety Levels 1 through 4 - Animals.

Appendix G-I. Standard Practices and Training

The first principle of containment is strict adherence to good microbiological practices (see Appendices G-III-A through G-III-J, *Footnotes and References of Appendix G*). Consequently, all personnel directly or indirectly involved in experiments using recombinant or synthetic nucleic acid molecules shall receive adequate instruction (see Sections IV-B-1-h, *Responsibilities of the Institution--General Information*, and IV-B-7-d, *Responsibilities of the Principal Investigator Prior to Initiating Research*). At a minimum, these instructions include training in aseptic techniques and in the biology of the organisms used in the experiments so that the potential biohazards can be understood and appreciated.

Any research group working with agents that are known or potential biohazards shall have an emergency plan that describes the procedures to be followed if an accident contaminates personnel or the environment. The Principal Investigator shall ensure that everyone in the laboratory is familiar with both the potential hazards of the work and the emergency plan (see Sections IV-B-7-d, *Responsibilities of the Principal Investigator Prior to Initiating Research* and IV-B-7-e, *Responsibilities of the Principal Investigator During the Conduct of the Research*). If a research group is working with a known pathogen for which there is an effective vaccine, the vaccine should be made available to all workers. Serological monitoring, when clearly appropriate, will be provided (see Section IV-B-1-f, *Responsibilities of the Institution--General Information*).

The *Laboratory Safety Monograph* (see Appendix G-III-O, *Footnotes and References of Appendix G*) and *Biosafety in Microbiological and Biomedical Laboratories* (see Appendix G-III-A, *Footnotes and References of Appendix G*) describe practices, equipment, and facilities in detail.

Appendix G-II. Physical Containment Levels

The objective of physical containment is to confine organisms containing recombinant or synthetic nucleic acid molecules and to reduce the potential for exposure of the laboratory worker, persons outside of the laboratory, and the environment to organisms containing recombinant or synthetic nucleic acid molecules. Physical containment is achieved through the use of laboratory practices, containment equipment, and special laboratory design. Emphasis is placed on primary means of physical containment which are provided by laboratory practices and containment equipment. Special laboratory design provides a secondary means of protection against the accidental release of organisms outside the laboratory or to the environment. Special laboratory design is used primarily in facilities in which experiments of moderate to high potential hazard are performed.

Combinations of laboratory practices, containment equipment, and special laboratory design can be made to achieve different levels of physical containment. Four levels of physical containment, which are designated as BL1, BL2, BL3, and BL4 are described. It should be emphasized that the descriptions and assignments of physical containment detailed below are based on existing approaches to containment of pathogenic organisms (see Appendix G-III-A, *Footnotes and References of Appendix G*). The National Cancer Institute describes three levels for research on oncogenic viruses which roughly correspond to our BL2, BL3, and BL4 levels (see Appendix G-III-C, *Footnotes and References of Appendix G*).

It is recognized that several different combinations of laboratory practices, containment equipment, and special laboratory design may be appropriate for containment of specific research activities. The *NIH Guidelines*, therefore, allow alternative selections of primary containment equipment within facilities that have been designed to provide BL3 and BL4 levels of physical containment. The selection of alternative methods of primary containment is dependent, however, on the level of biological containment provided by the host-vector system used in the experiment. Consideration will be given to other combinations which achieve an equivalent

level of containment (see Sections IV-C-1-b-(1), *Major Actions* and IV-C-1-b-(2), *Minor Actions*).

Appendix G-II-A. Biosafety Level 1 (BL1) (See Appendix G-III-M, *Footnotes and References of Appendix G*)

Appendix G-II-A-1. Standard Microbiological Practices (BL1)

Appendix G-II-A-1-a. Access to the laboratory is limited or restricted at the discretion of the Principal Investigator when experiments are in progress.

Appendix G-II-A-1-b. Work surfaces are decontaminated once a day and after any spill of viable material.

Appendix G-II-A-1-c. All contaminated liquid or solid wastes are decontaminated before disposal.

Appendix G-II-A-1-d. Mechanical pipetting devices are used; mouth pipetting is prohibited.

Appendix G-II-A-1-e. Eating, drinking, smoking, and applying cosmetics are not permitted in the work area. Food may be stored in cabinets or refrigerators designated and used for this purpose only.

Appendix G-II-A-1-f. Persons wash their hands: (i) after they handle materials involving organisms containing recombinant or synthetic nucleic acid molecules and animals, and (ii) before exiting the laboratory.

Appendix G-II-A-1-g. All procedures are performed carefully to minimize the creation of aerosols.

Appendix G-II-A-1-h. In the interest of good personal hygiene, facilities (e.g., hand washing sink, shower, changing room) and protective clothing (e.g., uniforms, laboratory coats) shall be provided that are appropriate for the risk of exposure to viable organisms containing recombinant or synthetic nucleic acid molecules.

Appendix G-II-A-2. Special Practices (BL1)

Appendix G-II-A-2-a. Contaminated materials that are to be decontaminated at a site away from the laboratory are placed in a durable leak-proof container which is closed before being removed from the laboratory.

Appendix G-II-A-2-b. An insect and rodent control program is in effect.

Appendix G-II-A-3. Containment Equipment (BL1)

Appendix G-II-A-3-a. Special containment equipment is generally not required for manipulations of agents assigned to BL1.

Appendix G-II-A-4. Laboratory Facilities (BL1)

Appendix G-II-A-4-a. The laboratory is designed so that it can be easily cleaned.

Appendix G-II-A-4-b. Bench tops are impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat.

Appendix G-II-A-4-c. Laboratory furniture is sturdy. Spaces between benches, cabinets, and equipment are accessible for cleaning.

Appendix G-II-A-4-d. Each laboratory contains a sink for hand washing.

Appendix G-II-A-4-e. If the laboratory has windows that open, they are fitted with fly screens.

Appendix G-II-B. Biosafety Level 2 (BL2) (See Appendix G-III-N, *Footnotes and References of Appendix G*)

Appendix G-II-B-1. Standard Microbiological Practices (BL2)

Appendix G-II-B-1-a. Access to the laboratory is limited or restricted by the Principal Investigator when work with organisms containing recombinant or synthetic nucleic acid molecules is in progress.

Appendix G-II-B-1-b. Work surfaces are decontaminated at least once a day and after any spill of viable material.

Appendix G-II-B-1-c. All contaminated liquid or solid wastes are decontaminated before disposal.

Appendix G-II-B-1-d. Mechanical pipetting devices are used; mouth pipetting is prohibited.

Appendix G-II-B-1-e. Eating, drinking, smoking, and applying cosmetics are not permitted in the work area. Food may be stored in cabinets or refrigerators designated and used for this purpose only.

Appendix G-II-B-1-f. Persons wash their hands: (i) after handling materials involving organisms containing recombinant or synthetic nucleic acid molecules and animals, and (ii) when exiting the laboratory.

Appendix G-II-B-1-g. All procedures are performed carefully to minimize the creation of aerosols.

Appendix G-II-B-1-h. Experiments of lesser biohazard potential can be conducted concurrently in carefully demarcated areas of the same laboratory.

Appendix G-II-B-2. Special Practices (BL2)

Appendix G-II-B-2-a. Contaminated materials that are to be decontaminated at a site away from the laboratory are placed in a durable leak-proof container which is closed before being removed from the laboratory.

Appendix G-II-B-2-b. The Principal Investigator limits access to the laboratory. The Principal Investigator has the final responsibility for assessing each circumstance and determining who may enter or work in the laboratory.

Appendix G-II-B-2-c. The Principal Investigator establishes policies and procedures whereby only persons who have been advised of the potential hazard and meet any specific entry requirements (e.g., immunization) may enter the laboratory or animal rooms.

Appendix G-II-B-2-d. When the organisms containing recombinant or synthetic nucleic acid molecules in use in the laboratory require special provisions for entry (e.g., vaccination), a hazard warning sign incorporating the universal biosafety symbol is posted on the access door to the laboratory work area. The hazard warning sign identifies the agent, lists the name and telephone number of the Principal Investigator or other responsible person(s), and indicates the special requirement(s) for entering the laboratory.

Appendix G-II-B-2-e. An insect and rodent control program is in effect.

Appendix G-II-B-2-f. Laboratory coats, gowns, smocks, or uniforms are worn while in the laboratory. Before exiting the laboratory for non-laboratory areas (e.g., cafeteria, library, administrative offices), this protective clothing is removed and left in the laboratory or covered with a clean coat not used in the laboratory.

Appendix G-II-B-2-g. Animals not involved in the work being performed are not permitted in the laboratory.

Appendix G-II-B-2-h. Special care is taken to avoid skin contamination with organisms containing recombinant or synthetic nucleic acid molecules; gloves should be worn when handling experimental animals and when skin contact with the agent is unavoidable.

Appendix G-II-B-2-i. All wastes from laboratories and animal rooms are appropriately decontaminated before disposal.

Appendix G-II-B-2-j. Hypodermic needles and syringes are used only for parenteral injection and aspiration of fluids from laboratory animals and diaphragm bottles. Only needle-locking syringes or disposable syringe-needle units (i.e., needle is integral to the syringe) are used for the injection or aspiration of fluids containing organisms that contain recombinant or synthetic nucleic acid molecules. Extreme caution should be used when handling needles and syringes to avoid autoinoculation and the generation of aerosols during use and disposal. Needles should not be bent, sheared, replaced in the needle sheath or guard, or removed from the syringe following use. The needle and syringe should be promptly placed in a puncture-resistant container and decontaminated, preferably autoclaved, before discard or reuse.

Appendix G-II-B-2-k. Spills and accidents which result in overt exposures to organisms containing recombinant or synthetic nucleic acid molecules are immediately reported to the Institutional Biosafety Committee and NIH OSP. Reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Medical evaluation, surveillance, and treatment are provided as appropriate and written records are maintained.

Appendix G-II-B-2-l. When appropriate, considering the agent(s) handled, baseline serum samples for laboratory and other at-risk personnel are collected and stored. Additional serum specimens may be collected periodically depending on the agents handled or the function of the facility.

Appendix G-II-B-2-m. A biosafety manual is prepared or adopted. Personnel are advised of special hazards and are required to read and follow instructions on practices and procedures.

Appendix G-II-B-3. Containment Equipment (BL2)

Appendix G-II-B-3-a. Biological safety cabinets (Class I or II) (see Appendix G-III-L, *Footnotes and References of Appendix G*) or other appropriate personal protective or physical containment devices are used whenever:

Appendix G-II-B-3-a-(1). Procedures with a high potential for creating aerosols are conducted (see Appendix G-III-O, *Footnotes and References of Appendix G*). These may include centrifuging, grinding, blending, vigorous shaking or mixing, sonic disruption, opening containers of materials whose internal pressures may be different from ambient pressures, intranasal inoculation of animals, and harvesting infected tissues from animals or eggs.

Appendix G-II-B-3-a-(2). High concentrations or large volumes of organisms containing recombinant or synthetic nucleic acid molecules are used. Such materials may be centrifuged in the open laboratory if sealed beads or centrifuge safety cups are used and if they are opened only in a biological safety cabinet.

Appendix G-II-B-4. Laboratory Facilities (BL2)

Appendix G-II-B-4-a. The laboratory is designed so that it can be easily cleaned.

Appendix G-II-B-4-b. Bench tops are impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat.

Appendix G-II-B-4-c. Laboratory furniture is sturdy and spaces between benches, cabinets, and equipment are accessible for cleaning.

Appendix G-II-B-4-d. Each laboratory contains a sink for hand washing.

Appendix G-II-B-4-e. If the laboratory has windows that open, they are fitted with fly screens.

Appendix G-II-B-4-f. An autoclave for decontaminating laboratory wastes is available.

Appendix G-II-C. Biosafety Level 3 (BL3) (See Appendix G-III-P, *Footnotes and References of Appendix G*)

Appendix G-II-C-1. Standard Microbiological Practices (BL3)

Appendix G-II-C-1-a. Work surfaces are decontaminated at least once a day and after any spill of viable material.

Appendix G-II-C-1-b. All contaminated liquid or solid wastes are decontaminated before disposal.

Appendix G-II-C-1-c. Mechanical pipetting devices are used; mouth pipetting is prohibited.

Appendix G-II-C-1-d. Eating, drinking, smoking, storing food, and applying cosmetics are not permitted in the work area.

Appendix G-II-C-1-e. Persons wash their hands: (i) after handling materials involving organisms containing recombinant or synthetic nucleic acid molecules, and handling animals, and (ii) when exiting the laboratory.

Appendix G-II-C-1-f. All procedures are performed carefully to minimize the creation of aerosols.

Appendix G-II-C-1-g. Persons under 16 years of age shall not enter the laboratory.

Appendix G-II-C-1-h. If experiments involving other organisms which require lower levels of containment are to be conducted in the same laboratory concurrently with experiments requiring BL3 level physical containment, they shall be conducted in accordance with all BL3 level laboratory practices.

Appendix G-II-C-2. Special Practices (BL3)

Appendix G-II-C-2-a. Laboratory doors are kept closed when experiments are in progress.

Appendix G-II-C-2-b. Contaminated materials that are to be decontaminated at a site away from the laboratory are placed in a durable leak-proof container which is closed before being removed from the laboratory.

Appendix G-II-C-2-c. The Principal Investigator controls access to the laboratory and restricts access to persons whose presence is required for program or support purposes. The Principal Investigator has the final responsibility for assessing each circumstance and determining who may enter or work in the laboratory.

Appendix G-II-C-2-d. The Principal Investigator establishes policies and procedures whereby only persons who have been advised of the potential biohazard, who meet any specific entry requirements (e.g., immunization), and who comply with all entry and exit procedures entering the laboratory or animal rooms.

Appendix G-II-C-2-e. When organisms containing recombinant or synthetic nucleic acid molecules or experimental animals are present in the laboratory or containment module, a hazard warning sign incorporating the universal biosafety symbol is posted on all laboratory and animal room access doors. The hazard warning sign identifies the agent, lists the name and telephone number of the Principal Investigator or other responsible person(s), and indicates any special requirements for entering the laboratory such as the need for immunizations, respirators, or other personal protective measures.

Appendix G-II-C-2-f. All activities involving organisms containing recombinant or synthetic nucleic acid molecules are conducted in biological safety cabinets or other physical containment devices within the containment module. No work in open vessels is conducted on the open bench.

Appendix G-II-C-2-g. The work surfaces of biological safety cabinets and other containment equipment are decontaminated when work with organisms containing recombinant or synthetic nucleic acid molecules is finished. Plastic-backed paper toweling used on non-perforated work surfaces within biological safety cabinets facilitates clean-up.

Appendix G-II-C-2-h. An insect and rodent program is in effect.

Appendix G-II-C-2-i. Laboratory clothing that protects street clothing (e.g., solid front or wrap-around gowns, scrub suits, coveralls) is worn in the laboratory. Laboratory clothing is not worn outside the laboratory, and it is decontaminated prior to laundering or disposal.

Appendix G-II-C-2-j. Special care is taken to avoid skin contamination with contaminated materials; gloves should be worn when handling infected animals and when skin contact with infectious materials is unavoidable.

Appendix G-II-C-2-k. Molded surgical masks or respirators are worn in rooms containing experimental animals.

Appendix G-II-C-2-l. Animals and plants not related to the work being conducted are not permitted in the laboratory.

Appendix G-II-C-2-m. Laboratory animals held in a BL3 area shall be housed in partial-containment caging systems, such as Horsfall units (see Appendix G-III-K, *Footnotes and References of Appendix G*), open cages

placed in ventilated enclosures, solid-wall and -bottom cages covered by filter bonnets or solid-wall and -bottom cages placed on holding racks equipped with ultraviolet in radiation lamps and reflectors.

Note: Conventional caging systems may be used provided that all personnel wear appropriate personal protective devices. These protective devices shall include at a minimum wrap-around gowns, head covers, gloves, shoe covers, and respirators. All personnel shall shower on exit from areas where these devices are required.

Appendix G-II-C-2-n. All wastes from laboratories and animal rooms are appropriately decontaminated before disposal. For research involving mammalian-transmissible HPAI H5N1 virus, liquid effluents should be chemically disinfected or heat-treated, or collected and processed in a central effluent decontamination system. Decontamination of shower and toilet effluents is not required, provided appropriate practices and procedures are in place for primary containment of mammalian-transmissible HPAI H5N1 virus. Animal tissues, carcasses, and bedding originating from the animal room must be decontaminated by an effective and validated method (e.g., use of an autoclave) preferably before leaving the containment barrier. If waste must be transported, special practices should be developed for transport of infectious materials to designated alternate location(s) within the facility.

Appendix G-II-C-2-o. Vacuum lines are protected with high efficiency particulate air/HEPA filters and liquid disinfectant traps.

Appendix G-II-C-2-p. Hypodermic needles and syringes are used only for parenteral injection and aspiration of fluids from laboratory animals and diaphragm bottles. Only needle locking syringes or disposable syringe-needle units (i.e., needle is integral to the syringe) are used for the injection or aspiration of fluids containing organisms that contain recombinant or synthetic nucleic acid molecules. Extreme caution should be used when handling needles and syringes to avoid autoinoculation and the generation of aerosols during use and disposal. Needles should not be bent, sheared, replaced in the needle sheath or guard, or removed from the syringe following use. The needle and syringe should be promptly placed in a puncture-resistant container and decontaminated, preferably by autoclaving, before discard or reuse.

Appendix G-II-C-2-q. Spills and accidents which result in overt or potential exposures to organisms containing recombinant or synthetic nucleic acid molecules are immediately reported to the Biological Safety Officer, Institutional Biosafety Committee, and NIH OSP. Reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Appropriate medical evaluation, surveillance, and treatment are provided and written records are maintained.

Appendix G-II-C-2-r. Baseline serum samples for all laboratory and other at-risk personnel should be collected and stored in accordance with institutional policy and at least for the time period in which the personnel continues to work with the agent at biosafety level 3 containment. Such samples must be collected and stored for laboratory and other at-risk personnel who will work with mammalian-transmissible HPAI H5N1 virus. Additional serum specimens may be collected periodically depending on the agents handled or the function of the laboratory.

Appendix G-II-C-2-s. A biosafety manual is prepared or adopted. Personnel are advised of special hazards and are required to read and follow the instructions on practices and procedures.

Appendix G-II-C-2-t. Alternative Selection of Containment Equipment (BL3)

Experimental procedures involving a host-vector system that provides a one-step higher level of biological containment than that specified may be conducted in the BL3 laboratory using containment equipment specified for the BL2 level of physical containment. Experimental procedures involving a host-vector system that provides a one-step lower level of biological containment than that specified may be conducted in the BL3 laboratory using containment equipment specified for the BL4 level of physical containment. Alternative combination of containment safeguards are shown in Appendix G-Table 1, *Possible Alternate Combinations of Physical and Biological Containment Safeguards*.

Appendix G-II-C-3. Containment Equipment (BL3)

Appendix G-II-C-3-a. Biological safety cabinets (Class I, II, or III) (see Appendix G-III-L, *Footnotes and*

References of Appendix G) or other appropriate combinations of personal protective or physical containment devices (e.g., special protective clothing, masks, gloves, respirators, centrifuge safety cups, sealed centrifuge rotors, and containment caging for animals) are used for all activities with organisms containing recombinant or synthetic nucleic acid molecules which pose a threat of aerosol exposure. These include: manipulation of cultures and of those clinical or environmental materials which may be a source of aerosols; the aerosol challenge of experimental animals; the harvesting of infected tissues or fluids from experimental animals and embryonate eggs; and the necropsy of experimental animals.

Appendix G-II-C-4. Laboratory Facilities (BL3)

Appendix G-II-C-4-a. The laboratory is separated from areas which are open to unrestricted traffic flow within the building. Passage through two sets of doors is the basic requirement for entry into the laboratory from access corridors or other contiguous areas. Physical separation of the high containment laboratory from access corridors or other laboratories or activities may be provided by a double-door clothes change room (showers may be included), airlock, or other access facility which requires passage through two sets of doors before entering the laboratory.

Appendix G-II-C-4-b. The interior surfaces of walls, floors, and ceilings are water resistant so that they can be easily cleaned. Penetrations in these surfaces are sealed or capable of being sealed to facilitate decontaminating the area.

Appendix G-II-C-4-c. Bench tops are impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat.

Appendix G-II-C-4-d. Laboratory furniture is sturdy and spaces between benches, cabinets, and equipment are accessible for cleaning.

Appendix G-II-C-4-e. Each laboratory contains a sink for hand washing. The sink is foot, elbow, or automatically operated and is located near the laboratory exit door.

Appendix G-II-C-4-f. Windows in the laboratory are closed and sealed.

Appendix G-II-C-4-g. Access doors to the laboratory or containment module are self-closing.

Appendix G-II-C-4-h. An autoclave for decontaminating laboratory wastes is available preferably within the laboratory.

Appendix G-II-C-4-i. A ducted exhaust air ventilation system is provided. This system creates directional airflow that draws air into the laboratory from uncontaminated spaces surrounding the laboratory. The exhaust air is not recirculated to any other area of the building, is discharged to the outside, and is dispersed away from the occupied areas and air intakes. Personnel shall verify that the direction of the airflow (into the laboratory) is proper. The exhaust air from the laboratory room may be discharged to the outside without being filtered or otherwise treated unless research is being conducted with mammaliantransmissible HPAI H5N1 virus. For research with mammalian-transmissible HPAI H5N1 virus, exhaust air must be HEPA filtered and there must be sealed ductwork from the containment barrier to the filter. In addition, the air handling system shall be designed such that under failure conditions, the airflow will not be reversed and periodic verification, with annual verification of the HEPA filters, shall be performed. Finally, backup power shall be available for critical controls and instrumentation necessary to maintain containment.

Appendix G-II-C-4-j. The high efficiency particulate air/HEPA filtered exhaust air from Class I or Class II biological safety cabinets is discharged directly to the outside or through the building exhaust system. Exhaust air from Class I or II biological safety cabinets may be recirculated within the laboratory if the cabinet is tested and certified at least every twelve months. If the HEPA-filtered exhaust air from Class I or II biological safety cabinets is to be discharged to the outside through the building exhaust air system, it is connected to this system in a manner (e.g., thimble unit connection (see Appendix G-III-L, *Footnotes and References of Appendix G*)) that avoids any interference with the air balance of the cabinets or building exhaust system.

Appendix G-II-C-5. Biosafety Level 3 Enhanced for Research Involving Risk Group 3 Influenza Viruses. (See Appendices G-II-C-2-n, G-II-C-2-r, and G-II-C-4-i for additional guidance for facilities, waste

handling, and serum collection for research involving mammaliantransmissible HPAI H5N1 virus.

Appendix G-II-C-5-a. Containment, Practices, and Training for Research with Risk Group 3 Influenza Viruses (BL3 Enhanced).

Appendix G-II-C-5-a-(1). In addition to standard BL3 practices, the following additional personal protective equipment and practices shall be used: (1) Powered Air-purifying Respirators (PAPR) are worn. (2) Street clothes are changed to protective suit (e.g., wrap-back disposable gown, olefin protective suit). (3) Double gloves (disposable) are worn. For research with mammalian-transmissible HPAI H5N1 viruses, protective sleeves shall be worn over the gown while working in a biosafety cabinet. (4) Appropriate shoe coverings are worn (e.g., double disposable shoe coverings, single disposable shoe coverings if worn with footwear dedicated to BL3 enhanced laboratory use, or impervious boots or shoes of rubber or other suitable material that can be decontaminated). (5) Showers prior to exiting the laboratory should be considered depending on risk assessment of research activities, with the exception that showers prior to exiting the laboratory are required for all research with mammalian-transmissible HPAI H5N1 virus, including care of animals infected with mammalian-transmissible HPAI H5N1 virus. (6) For research with mammalian-transmissible HPAI H5N1 virus, prior to leaving containment, personal protective equipment shall be sprayed or wiped down with a disinfectant that has activity against influenza viruses. (7) In order to promote adherence to proper practices, including proper removal of personal protective equipment, and reporting of any loss of containment or exposures, at least two individuals should be in the laboratory at all times when research with mammalian-transmissible HPAI H5N1 virus involves experimental procedures with animals or sharps, or when procedures are being conducted whereby the generation of aerosols is reasonably anticipated. Removal of personal protective equipment should be observed.

Appendix G-II-C-5-a-(2). As proper training of laboratory workers is an essential component of biosafety, retraining and periodic reassessments (at least annually) in BL3 enhanced practices, especially the proper use of respiratory equipment, such as PAPRs, and clothing changes are required. For research with mammalian-transmissible HPAI H5N1 virus, laboratory workers shall be required to sign a document acknowledging their understanding of and intent to adhere to biosafety, biosecurity, and occupational health requirements. This document shall include a statement that the laboratory worker agrees to report any exposures or accidents, including those by other individuals in the lab.

Appendix G-II-C-5-a-(3). Reporting of all spills and accidents, even if relatively minor, is required as described in Appendix G-II-C-2-q.

Appendix G-II-C-5-a-(4). To avoid inadvertent cross contamination of 1918 H1N1, HPAI H5N1 or human H2N2 (1957-1968): (1) Containment facilities and practices appropriate for highest Risk Group virus shall be used at all times with lower Risk Group viruses, when studied in the same laboratory room. (2) Tissue cultures with these viruses shall be conducted at separate times (temporal spacing) in the same room. (3) Separate reagents shall be used to minimize risk of cross contamination. (4) A laboratory worker shall not perform concurrent influenza virus experiments that carry the risk of unintended reassortment among 1918 H1N1, human H2N2 (1957-1968), HPAI H5N1 and other human influenza viruses. (5) Two or more laboratory workers shall not perform within the same work area simultaneous influenza virus experiments that carry the risk of unintended segment reassortment between 1918 H1N1, or HPAI H5N1, or human H2N2 (1957-1968) and other human influenza viruses. (6) Between experiments good biosafety decontamination practices (e.g., surface and biosafety cabinet surface decontamination according to standard BL3 procedures) shall be used and there shall be a thirty minute wait period after decontamination before equipment is used for experiments with any other influenza A viruses. (7) Between experiments, in addition to decontamination of the work area, clothing changes and PAPR disinfection shall be performed prior to handling a different influenza virus in the same work area. (Shower-out capability may be required by USDA/APHIS for certain experiments with HPAI H5N1.)

Appendix G-II-C-5-a-(5). Continued susceptibility of the reassortant influenza viruses containing genes and/or segments from 1918 H1N1, HPAI H5N1, and human H2N2 (1957-1968) to antiviral agents shall be established by sequence analysis or suitable biological assays. After manipulation of genes that influence sensitivity to antiviral agents, susceptibility to these agents shall be reconfirmed. If susceptibility to neuraminidase inhibitors or other effective antiviral agents is lost as a result of genetic modification or serial passage of a mammalian-transmissible HPAI H5N1 virus, then any research with this antiviral agent resistant virus shall be stopped and research shall only proceed after review by the NIH (as outlined in Section III-A-I-a) or the appropriate federal regulatory agency.

Appendix G-II-C-5-b. Containment for Animal Research.
Guidance provided in Appendix G-II-C and Appendix Q-II-C is applicable with the following emphasis on standard BL3 or BL3-N containment or additional enhancements.

Appendix G-II-C-5-b-(1). Research with small animals shall be conducted in a class II biosafety cabinet. Small animals such as rodents (e.g. mice, hamsters, rats, guinea pigs) can be housed within a negative pressure BL3 animal suite using high-density individually vented caging (IVC) systems that independently supply high efficiency particulate air/HEPA-filtered and directional air circulation. Other animals (e.g. rabbits, ferrets) that are of a size or have growth or caging requirements that preclude the use of high-density IVC systems are to be housed in negative pressure bioisolators.

Appendix G-II-C-5-b-(2). Large animals such as non-human primates shall be housed in primary barrier environments according to BL3-N containment requirements (see Appendix Q-II-C).

Appendix G-II-C-5-b-(3). Specialized training and proven competency in all assigned practices and procedures shall be required for laboratory staff, including staff involved in animal care.

Appendix G-II-C-5-b-(4). For HPAI H5N1 research, the *NIH Guidelines* defer to USDA/APHIS recommendations for biocontainment practices for loose housed animals.

Appendix G-II-C-5-c. Occupational Health
A detailed occupational health plan shall be developed in advance of working with these agents in consultation, as needed, with individuals with the appropriate clinical expertise. In addition, the appropriate public health authority shall be consulted (e.g. local public health officials) on the plan and a mock drill of this plan shall be undertaken periodically. The plan shall include a description of the incident reporting system in place for incidents, which include any loss of containment, spills, accidents, or potential exposures. The plan must specify that all incidents must be reported immediately to the appropriate institutional authorities, and no later than 24 hours to the appropriate public health authorities (e.g., the USDA, the CDC, NIH, local and state health authorities).

Appendix G-II-C-5-c-(1). Laboratory workers shall be provided with medical cards which include, at a minimum, the following information: characterization of the influenza virus to which they have been potentially exposed, and 24-hour contact numbers for the Principal Investigator and institution's occupational health care provider(s).

Appendix G-II-C-5-c-(2). A detailed occupational health plan shall include: (1) Unless there is a medical contraindication to vaccination (e.g. severe egg allergy) annual seasonal influenza vaccination as prerequisite for research to reduce risk of influenza like illness that would require isolation and testing to rule out infection with experimental viruses and raise the risk for possible co-infection with circulating influenza strains. (2) Virus specific vaccination, if available, should be offered and if a licensed HPAI H5N1 vaccine is available, and there are no medical contraindications, laboratory workers performing research with mammalian-transmissible HPAI H5N1 virus should be vaccinated. A post-vaccination serum sample shall be collected, assessed for immune response, and stored in accordance with institutional policy, at least for the time in which the laboratory worker continues to conduct HPAI H5N1 virus research. (3) Reporting of all respiratory symptoms and/or fever (i.e. influenza-like illnesses). For research involving mammalian-transmissible HPAI H5N1 virus, laboratory workers shall be actively monitored for influenza-like illness (i.e., fever and respiratory symptoms). (4) 24-hour access to a medical facility that is prepared to implement appropriate respiratory isolation to prevent transmission and is able to provide appropriate antiviral agents. Real-time reverse transcription-polymerase chain reaction (RT-PCR) assays should be used for virus detection and to discriminate these viruses from currently circulating human influenza viruses. For exposures to viruses containing genes from 1918 H1N1 or the HA gene from human H2N2 (1957-1968), specimens shall be sent to the CDC for testing (RT-PCR and confirmatory sequencing).

Appendix G-II-C-5-c-(3). In preparing to perform research with 1918 H1N1, human H2N2 (1957-1968), or HPAI H5N1, Principal Investigators should develop a clear plan specifying who will be contacted in the event of a potential exposure (during and after work hours) to conduct a risk assessment and make decisions as to the required response, including the need for and extent of isolation of the exposed worker. After any kind of potential exposure, a rapid risk assessment shall be performed by the Principal Investigator, health and biosafety officials and subsequent actions should depend on the appraised level of risk of respiratory infection

for the individual and potential for transmission to others. A laboratory worker performing research with either an influenza virus containing the HA gene from human H2N2 or an influenza virus containing genes and/or segments from 1918 H1N1 or mammalian-transmissible HPAI H5N1 viruses, shall be informed in advance that, in the case of a *known* laboratory exposure with a high risk for infection, e.g., involving the upper or lower respiratory tract or mucous membranes, the laboratory worker will need to be isolated in a predetermined facility, rather than home isolation, until infection can be ruled out by testing (e.g., negative RT-PCR for 1918 H1N1 or human H2N2 (1957-1968), or HPAI H5N1) of appropriately timed specimens. Laboratory workers with a known laboratory exposure with high risk for infection during research with HPAI H5N1 virus strains that are not transmissible among mammals should be prepared to self-isolate (for example at home) until infection can be ruled out by testing (e.g., negative RT-PCR for HPAI H5N1) of appropriately timed specimens. The action taken for other types of exposures should be based on the risk assessment. In addition, based on the risk assessment: (1) treatment with appropriate antiviral agents shall be initiated, and (2) the appropriate public health authorities shall be notified.

Appendix G-II-C-5-c-(4). Influenza-like illness. If an individual has entered (within ten days) a laboratory conducting research with influenza viruses containing the human H2N2 HA gene or any gene from the 1918 H1N1 or HPAI H5N1 viruses, or housing animals exposed to such viruses, and the individual demonstrates symptoms and/or signs of influenza infection (e.g., fever/chills, cough, myalgias, headache), then he/she shall report by phone to the supervisor/Principal Investigator and other individuals identified in the occupational health plan. If needed, the person with influenza-like illness shall be transported, under the appropriate isolation conditions, to a healthcare facility that can provide adequate respiratory isolation, appropriate medical therapy, and testing to determine whether the infection is due to a recombinant or synthetic influenza virus. The appropriate public health authorities shall be informed whenever a suspected case is isolated.

Appendix G-II-C-5-c-(5). For 1918 H1N1 research, the use of antiviral agents (e.g., oseltamivir) for pre-exposure prophylaxis shall be discussed with laboratory workers in advance including a discussion of the data on the safety of long term exposure to these agents and their ability to reduce the risk of clinical disease and the limits of the data regarding protection of close contacts and the community.

Appendix G-II-C-5-c-(6). Antiviral agents for postexposure prophylaxis shall be provided only after medical evaluation. Home supplies shall not be provided in advance for research with 1918 H1N1, mammalian-transmissible HPAI H5N1 or influenza viruses containing the HA gene from human H2N2.

Appendix G-II-D. Biosafety Level 4 (BL4)

Appendix G-II-D-1. Standard Microbiological Practices (BL4)

Appendix G-II-D-1-a. Work surfaces are decontaminated at least once a day and immediately after any spill of viable material.

Appendix G-II-D-1-b. Only mechanical pipetting devices are used.

Appendix G-II-D-1-c. Eating, drinking, smoking, storing food, and applying cosmetics are not permitted in the laboratory.

Appendix G-II-D-1-d. All procedures are performed carefully to minimize the creation of aerosols.

Appendix G-II-D-2. Special Practices (BL4)

Appendix G-II-D-2-a. Biological materials to be removed from the Class III cabinets or from the maximum containment laboratory in a viable or intact state are transferred to a non-breakable, sealed primary container and then enclosed in a non-breakable, sealed secondary container which is removed from the facility through a disinfectant dunk tank, fumigation chamber, or an airlock designed for this purpose.

Appendix G-II-D-2-b. No materials, except for biological materials that are to remain in a viable or intact state, are removed from the maximum containment laboratory unless they have been autoclaved or decontaminated before exiting the facility. Equipment or material which might be damaged by high temperatures or steam is decontaminated by gaseous or vapor methods in an airlock or chamber designed for this purpose.

Appendix G-II-D-2-c. Only persons whose presence in the facility or individual laboratory rooms is required for program or support purposes are authorized to enter. The supervisor has the final responsibility for assessing each circumstance and determining who may enter or work in the laboratory. Access to the facility is limited by means of secure, locked doors; accessibility is managed by the Principal Investigator, Biological Safety Officer, or other person responsible for the physical security of the facility. Before entering, persons are advised of the potential biohazards and instructed as to appropriate safeguards for ensuring their safety. Authorized persons comply with the instructions and all other applicable entry and exit procedures. A logbook signed by all personnel indicates the date and time of each entry and exit. Practical and effective protocols for emergency situations are established.

Appendix G-II-D-2-d. Personnel enter and exit the facility only through the clothing change and shower rooms. Personnel shower each time they exit the facility. Personnel use the air locks to enter or exit the laboratory only in an emergency.

Appendix G-II-D-2-e. Street clothing is removed in the outer clothing change room and kept there. Complete laboratory clothing (may be disposable), including undergarments, pants and shirts or jump suits, shoes, and gloves, is provided and used by all personnel entering the facility. Head covers are provided for personnel who do not wash their hair during the exit shower. When exiting the laboratory and before proceeding into the shower area, personnel remove their laboratory clothing and store it in a locker or hamper in the inner change room. Protective clothing shall be decontaminated prior to laundering or disposal.

Appendix G-II-D-2-f. When materials that contain organisms containing recombinant or synthetic nucleic acid molecules or experimental animals are present in the laboratory or animal rooms, a hazard warning sign incorporating the universal biosafety symbol is posted on all access doors. The sign identifies the agent, lists the name of the Principal Investigator or other responsible person(s), and indicates any special requirements for entering the area (e.g., the need for immunizations or respirators).

Appendix G-II-D-2-g. Supplies and materials needed in the facility are brought in by way of the double-doored autoclave, fumigation chamber, or airlock which is appropriately decontaminated between each use. After securing the outer doors, personnel within the facility retrieve the materials by opening the interior doors or the autoclave, fumigation chamber, or airlock. These doors are secured after materials are brought into the facility.

Appendix G-II-D-2-h. An insect and rodent control program is in effect.

Appendix G-II-D-2-i. Materials (e.g., plants, animals, and clothing) not related to the experiment being conducted are not permitted in the facility.

Appendix G-II-D-2-j. Hypodermic needles and syringes are used only for parenteral injection and aspiration of fluids from laboratory animals and diaphragm bottles. Only needle-locking syringes or disposable syringe-needle units (i.e., needle is integral part of unit) are used for the injection or aspiration of fluids containing organisms that contain recombinant or synthetic nucleic acid molecules. Needles should not be bent, sheared, replaced in the needle sheath or guard, or removed from the syringe following use. The needle and syringe should be placed in a puncture-resistant container and decontaminated, preferably by autoclaving before discard or reuse. Whenever possible, cannulas are used instead of sharp needles (e.g., gavage).

Appendix G-II-D-2-k. A system is set up for reporting laboratory accidents, exposures, employee absenteeism, and for the medical surveillance of potential laboratory-associated illnesses. Spills and accidents which result in overt exposures to organisms containing recombinant or synthetic nucleic acid molecules are immediately reported to the Biological Safety Officer, Institutional Biosafety Committee, and NIH OSP. Reports to the NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Written records are prepared and maintained. An essential adjunct to such a reporting-surveillance system is the availability of a facility for quarantine, isolation, and medical care of personnel with potential or known laboratory associated illnesses.

Appendix G-II-D-2-l. Containment for Animal Research

Appendix G-II-D-2-l-(1). Laboratory animals involved in experiments requiring BL4 level physical containment shall be housed either in cages contained in Class III cabinets or in partial containment caging systems, such as

Horsfall units (see Appendix G-III-K, *Footnotes and References of Appendix G*), open cages placed in ventilated enclosures, or solid-wall and -bottom cages placed on holding racks equipped with ultraviolet irradiation lamps and reflectors that are located in a specially designed area in which all personnel are required to wear one-piece positive pressure suits.

Appendix G-II-D-2-l-(2). Non-human primates (NHP) may be housed (1) under the containment conditions described in Appendix G-II-D-2-l-(1) above, or (2) in open cages within a dedicated animal holding room that serves as the primary barrier and in which all personnel are required to wear one-piece positive pressure suits. A room serving as a primary barrier must be air-tight and capable of being decontaminated using fumigation. If NHPs are to be contained in a dedicated animal holding room serving as the primary barrier, the following conditions shall be met:
(i) Access to the animal holding room from service corridors outside of the BL4 containment space shall require passage through two sets of doors, and the inner most door must be an air pressure resistant (APR) door;
(ii) For any animal holding room considered to be a primary barrier, APR door(s) providing direct ingress from the exterior service corridor shall be fitted with appropriate and redundant lock-out mechanisms to prevent access when the animal holding room is contaminated and in use. There should be more than one mechanism to ensure that this primary barrier door cannot be opened when the animal room is contaminated and the APR door shall not serve as an emergency exit from the BL4 laboratory. The APR door shall be appropriately tested to demonstrate that in the closed, locked-out mode, the door provides an air-tight barrier proven by pressure decay testing or other equivalent method;
(iii) Any door(s) allowing access into a corridor from which there is direct ingress to an animal holding room must be fitted with either (1) an APR door or (2) a non-APR door, provided directional airflow is maintained from the laboratory corridor space into the animal room. For the purpose of fumigation, animal rooms equipped with non-APR doors opening into the adjacent interior corridors shall be considered one space (i.e., areas between air-tight doors shall be fumigated together);
(iv) Any door(s) used for access to the service corridor (the secondary barrier) shall be self-closing and of solid construction, designed not to corrode, split or warp;
(v) Access to the service corridor inside the secondary barrier shall be restricted and strictly controlled when animal rooms are in use. Whenever possible, the secondary barrier door(s) should be fitted with safety interlock switches designed to prevent it from opening when an animal holding room door (the primary barrier) is opened following room decontamination; if interlock devices cannot be used, specific administrative procedures shall be implemented to control access to the service corridor;
(vi) The service corridor shall maintain a negative pressure (inward directional airflow) relative to adjoining traffic corridors;
(vii) Prior to fumigation of the animal holding room, cages should be removed for autoclaving or chemical decontamination;
(viii) Caging should be chosen to reduce the amount of animal detritus that can be thrown out of the cage and onto the floor of the animal holding room;
(ix) The flow of personnel, material and equipment should be directed in order to minimize the spread of contamination from the animal holding room into adjacent areas of the laboratory;
(x) Following animal room decontamination, safeguards involving the use of personal protective equipment and appropriate administrative controls shall be implemented for the safe retrieval of biological indicators in order to prevent the spread of infectious agents in the event of a decontamination failure.

Appendix G-II-D-2-m. Alternative Selection of Containment Equipment (BL4)

Experimental procedures involving a host-vector system that provides a one-step higher level of biological containment than that specified may be conducted in the BL4 facility using containment equipment requirements specified for the BL3 level of physical containment. Alternative combinations of containment safeguards are shown in Appendix G-Table 1, *Possible Alternate Combinations of Physical and Biological Containment Safeguards*.

Appendix G-II-D-3. Containment Equipment (BL4)

Appendix G-II-D-3-a. All procedures within the facility with agents assigned to Biosafety Level 4 are conducted in the Class III biological safety cabinet or in Class I or II biological safety cabinets used in conjunction with one-piece positive pressure personnel suits ventilated by a life-support system.

Appendix G-II-D-4. Laboratory Facilities (BL4)

Appendix G-II-D-4-a. The maximum containment facility consists of either a separate building or a clearly demarcated and isolated zone within a building. Outer and inner change rooms separated by a shower are provided for personnel entering and exiting the facility. A double-doored autoclave, fumigation chamber, or ventilated airlock is provided for passage of those materials, supplies, or equipment which are not brought into the facility through the change room.

Appendix G-II-D-4-b. Walls, floors, and ceilings of the facility are constructed to form a sealed internal shell which facilitates fumigation and is animal and insect proof. The internal surfaces of this shell are resistant to liquids and chemicals, thus facilitating cleaning and decontamination of the area. All penetrations in these structures and surfaces are sealed. Any drains in the floors contain traps filled with a chemical disinfectant of demonstrated efficacy against the target agent, and they are connected directly to the liquid waste decontamination system. Sewer and other ventilation lines contain high efficiency particulate air/HEPA filters.

Appendix G-II-D-4-c. Internal facility appurtenances, such as light fixtures, air ducts, and utility pipes, are arranged to minimize the horizontal surface area on which dust can settle.

Appendix G-II-D-4-d. Bench tops have seamless surfaces which are impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat.

Appendix G-II-D-4-e. Laboratory furniture is simple and of sturdy construction; and spaces between benches, cabinets, and equipment are accessible for cleaning.

Appendix G-II-D-4-f. A foot, elbow, or automatically operated hand washing sink is provided near the door of each laboratory room in the facility.

Appendix G-II-D-4-g. If there is a central vacuum system, it does not serve areas outside the facility. In-line high efficiency particulate air/HEPA filters are placed as near as practicable to each use point or service cock. Filters are installed to permit in-place decontamination and replacement. Other liquid and gas services to the facility are protected by devices that prevent back-flow.

Appendix G-II-D-4-h. If water fountains are provided, they are foot operated and are located in the facility corridors outside the laboratory. The water service to the fountain is not connected to the back-flow protected distribution system supplying water to the laboratory areas.
Appendix G-II-D-4-i. Access doors to the laboratory are self-closing and locking.

Appendix G-II-D-4-j. Any windows are breakage resistant.

Appendix G-II-D-4-k. A double-door autoclave is provided for decontaminating materials passing out of the facility. The autoclave door which opens to the area external to the facility is sealed to the outer wall and automatically controlled so that the outside door can only be opened after the autoclave "sterilization" cycle has been completed.

Appendix G-II-D-4-l. A pass-through dunk tank, fumigation chamber, or an equivalent decontamination method is provided so that materials and equipment that cannot be decontaminated in the autoclave can be safely removed from the facility.

Appendix G-II-D-4-m. Liquid effluent from laboratory sinks, biological safety cabinets, floors, and autoclave chambers are decontaminated by heat treatment before being released from the maximum containment facility. Liquid wastes from shower rooms and toilets may be decontaminated with chemical disinfectants or by heat in the liquid waste decontamination system. The procedure used for heat decontamination of liquid wastes is evaluated mechanically and biologically by using a recording thermometer and an indicator microorganism with a defined heat susceptibility pattern. If liquid wastes from the shower room are decontaminated with chemical disinfectants, the chemical used is of demonstrated efficacy against the target or indicator microorganisms.

Appendix G-II-D-4-n. An individual supply and exhaust air ventilation system is provided. The system maintains pressure differentials and directional airflow as required to assure flows inward from areas outside of the facility toward areas of highest potential risk within the facility. Manometers are used to sense pressure

differentials between adjacent areas maintained at different pressure levels. If a system malfunctions, the manometers sound an alarm. The supply and exhaust airflow is interlocked to assure inward (or zero) airflow at all times.

Appendix G-II-D-4-o. The exhaust air from the facility is filtered through high efficiency particulate air/HEPA filters and discharged to the outside so that it is dispersed away from occupied buildings and air intakes. Within the facility, the filters are located as near the laboratories as practicable in order to reduce the length of potentially contaminated air ducts. The filter chambers are designed to allow *in situ* decontamination before filters are removed and to facilitate certification testing after they are replaced. Coarse filters and HEPA filters are provided to treat air supplied to the facility in order to increase the lifetime of the exhaust HEPA filters and to protect the supply air system should air pressures become unbalanced in the laboratory.

Appendix G-II-D-4-p. The treated exhaust air from Class I and II biological safety cabinets may be discharged into the laboratory room environment or the outside through the facility air exhaust system. If exhaust air from Class I or II biological safety cabinets is discharged into the laboratory the cabinets are tested and certified at minimum on a yearly basis. More frequent testing and certification, based on the amount of use or other safety factors, shall be left to the discretion of the IBC. The exhaust air from Class III biological safety cabinets is discharged, without recirculation through two sets of high efficiency particulate air/HEPA filters in series, via the facility exhaust air system. If the treated exhaust air from any of these cabinets is discharged to the outside through the facility exhaust air system, it is connected to this system in a manner (e.g., thimble unit connection (see Appendix G-III-L, *Footnotes and References of Appendix G*)) that avoids any interference with the air balance of the cabinets or the facility exhaust air system.

Appendix G-II-D-4-q. A specially designed suit area may be provided in the facility. Personnel who enter this area shall wear a one-piece positive pressure suit that is ventilated by a life-support system. The life-support system includes alarms and emergency backup breathing air tanks. Entry to this area is through an airlock fitted with airtight doors. A chemical shower is provided to decontaminate the surface of the suit before the worker exits the area. The exhaust air from the suit area is filtered by two sets of high efficiency particulate air/HEPA filters installed in series. A duplicate filtration unit, exhaust fan, and an automatically starting emergency power source are provided. The air pressure within the suit area is greater than that of any adjacent area. Emergency lighting and communication systems are provided. All penetrations into the internal shell of the suit are sealed. A double-doored autoclave is provided for decontaminating waste materials to be removed from the suit areas.

Appendix G - Table 1. Possible Alternate Combinations Of Physical And Biological Containment Safeguards

Classification of Physical & Biological Containment	Alternate Physical Containment			Alternate Biological Containment
	Laboratory Facilities	Laboratory Practices	Laboratory Equipment	
BL3/HV2	BL3	BL3	BL3	HV2
	BL3	BL3	BL4	HV1
BL3/HV1	BL3	BL3	BL3	HV1
	BL3	BL3	BL2	HV2
BL4/HV1	BL4	BL4	BL4	HV1
	BL4	BL4	BL3	HV2

BL - Biosafety Level
HV - Host-Vector System

Appendix G-III. Footnotes and References of Appendix G

Appendix G-III-A. *Biosafety in Microbiological and Biomedical Laboratories,* 5th edition, 2007, DHHS, Public Health Service, Centers for Disease Control and Prevention, Atlanta, Georgia, and National Institutes of Health, Bethesda, Maryland.

Appendix G-III-B. *Biosafety in Microbiological and Biomedical Laboratories*, 3rd edition, May 1993, U.S. DHHS, Public Health Service, Centers for Disease Control and Prevention, Atlanta, Georgia, and NIH, Bethesda, Maryland.

Appendix G-III-C. *National Cancer Institute Safety Standards for Research Involving Oncogenic Viruses*, U.S. Department of Health, Education, and Welfare Publication No. (NIH) 75-790, October 1974.

Appendix G-III-D. *National Institutes of Health Biohazards Safety Guide*, U.S. Department of Health, Education, and Welfare, Public Health Service, NIH, U.S. Government Printing Office, Stock No. 1740-00383, 1974.

Appendix G-III-E. A. Hellman, M. N. Oxman, and R. Pollack (eds.), *Biohazards in Biological Research*, Cold Spring Harbor Laboratory 1973.

Appendix G-III-F. N. V. Steere (ed.), *Handbook of Laboratory Safety*, 2nd edition, The Chemical Rubber Co., Cleveland, Ohio, 1971.

Appendix G-III-G. Bodily, J. L, "General Administration of the Laboratory," H. L. Bodily, E. L. Updyke, and J. O. Mason (eds.), *Diagnostic Procedures for Bacterial, Mycotic, and Parasitic Infections*, American Public Health Association, New York, 1970, pp. 11-28.

Appendix G-III-H. Darlow, H. M. (1969). "Safety in the Microbiological Laboratory," in J. R. Norris and D. W. Robbins (eds.), *Methods in Microbiology*, Academic Press, Inc., New York, pp. 169-204.

Appendix G-III-I. *The Prevention of Laboratory Acquired Infection*, C. H. Collins, E. G. Hartley, and R. Pilsworth, Public Health Laboratory Service, Monograph Series No. 6, 1974.

Appendix G-III-J. Chatigny, M. A., "Protection Against Infection in the Microbiological Laboratory: Devices and Procedures," in W. W. Umbreit (ed.), *Advances in Applied Microbiology*, Academic Press, New York, New York, 1961, 3:131-192.

Appendix G-III-K. Horsfall, F. L. Jr., and J. H. Baner, *Individual Isolation of Infected Animals in a Single Room*, J. Bact., 1940, 40, 569-580.

Appendix G-III-L. Biological safety cabinets referred to in this section are classified as Class I, Class II, or Class III cabinets. A Class I is a ventilated cabinet for personnel protection having an inward flow of air away from the operator. The exhaust air from this cabinet is filtered through a high efficiency particulate air/HEPA filter. This cabinet is used in three operational modes: (i) with a full-width open front, (ii) with an installed front closure panel (having four 6-inch diameter openings) without gloves, and (iii) with an installed front closure panel equipped with arm-length rubber gloves. The face velocity of the inward flow of air through the full-width open front is 75 feet per minute or greater. A Class II cabinet is a ventilated cabinet for personnel and product protection having an open front with inward air flow for personnel protection, and HEPA filtered mass recirculated air flow for product protection. The cabinet exhaust air is filtered through a HEPA filter. The face velocity of the inward flow of air through the full-width open front is 75 feet per minute or greater. Design and performance specifications for Class II cabinets have been adopted by the National Sanitation Foundation, Ann Arbor, Michigan. A Class III cabinet is a closed-front ventilated cabinet of gas tight construction which provides the highest level of personnel protection of all biosafety safety cabinets. The interior of the cabinet is protected from contaminants exterior to the cabinet. The cabinet is fitted with arm-length rubber gloves and is operated under a negative pressure of at least 0.5 inches water gauge. All supply air is filtered through HEPA filters. Exhaust air is filtered through two HEPA filters or one HEPA filter and incinerator before being discharged to the outside environment. National Sanitation Foundation Standard 49. 1976. Class II (Laminar Flow) Biohazard Cabinetry, Ann Arbor, Michigan.

Appendix G-III-M. Biosafety Level 1 is suitable for work involving agents of unknown or minimal potential hazard to laboratory personnel and the environment. The laboratory is not separated from the general traffic patterns in the building. Work is generally conducted on open bench tops. Special containment equipment is not required or generally used. Laboratory personnel have specific training in the procedures conducted in the laboratory and are supervised by a scientist with general training in microbiology or a related science (see Appendix G-III-A, *Footnotes and References of Appendix G*).

Appendix G-III-N. Biosafety Level 2 is similar to Level 1 and is suitable for work involving agents of moderate potential hazard to personnel and the environment. It differs in that: (1) laboratory personnel have specific training in handling pathogenic agents and are directed by competent scientists; (2) access to the laboratory is limited when work is being conducted; and (3) certain procedures in which infectious aerosols are created are conducted in biological safety cabinets or other physical containment equipment (see Appendix G-III-A, *Footnotes and References of Appendix G*).

Appendix G-III-O. Office of Research Safety, National Cancer Institute, and the Special Committee of Safety and Health Experts, *Laboratory Safety Monograph: A Supplement to the NIH Guidelines for Recombinant DNA Research*, NIH, Bethesda, Maryland 1978.

Appendix G-III-P. Biosafety Level 3 is applicable to clinical, diagnostic, teaching, research, or production facilities in which work is conducted with indigenous or exotic agents which may cause serious or potentially lethal disease as a result of exposure by the inhalation route. Laboratory personnel have specific training in handling pathogenic and potentially lethal agents and are supervised by competent scientists who are experienced in working with these agents. All procedures involving the manipulation of infectious material are conducted within biological safety cabinets or other physical containment devices or by personnel wearing appropriate personal protective clothing and devices. The laboratory has special engineering and design features. It is recognized, however, that many existing facilities may not have all the facility safeguards recommended for BL3 (e.g., access zone, sealed penetrations, and directional airflow, etc.). In these circumstances, acceptable safety may be achieved for routine or repetitive operations (e.g., diagnostic procedures involving the propagation of an agent for identification, typing, and susceptibility testing) in laboratories where facility features satisfy BL2 recommendations provided the recommended "Standard Microbiological Practices," "Special Practices," and "Containment Equipment" for BL3 are rigorously followed. The decision to implement this modification of BL3 recommendations should be made only by the Principal Investigator.

****

APPENDIX H. SHIPMENT

Recombinant or synthetic nucleic acid molecules contained in an organism or in a viral genome shall be shipped under the applicable regulations of the U.S. Postal Service (39 Code of Federal Regulations, Part 3); the Public Health Service (42 Code of Federal Regulations, Part 72); the U.S. Department of Agriculture (9 Code of Federal Regulations, Subchapters D and E; 7 CFR, Part 340); and/or the U.S. Department of Transportation (49 Code of Federal Regulations, Parts 171-179).

Note. A host-vector system may be proposed for certification by the NIH Director in accordance with the procedures set forth in Appendix I-II, *Certification of Host-Vector Systems*. In order to ensure protection for proprietary data, any public notice regarding a host-vector system which is designated by the institution as proprietary under Section IV-D, *Voluntary Compliance*, will be issued only after consultation with the institution as to the content of the notice (see Section IV-D-3, *Certification of Host-Vector Systems - Voluntary Compliance)*.

Appendix H-I. Host organisms or viruses will be shipped as etiologic agents, regardless of whether they contain recombinant or synthetic nucleic acid molecules, if they are regulated as human pathogens by the Public Health Service (42 Code of Federal Regulations, Part 72) or as animal pathogens or plant pests under the U.S. Department of Agriculture, Animal and Plant Health Inspection Service (Titles 9 and 7 Code of Federal Regulations, respectively).

Appendix H-II. Host organisms and viruses will be shipped as etiologic agents if they contain recombinant or synthetic nucleic acid molecules when: (i) the recombinant or synthetic nucleic acid molecule includes the

complete genome of a host organism or virus regulated as a human or animal pathogen or a plant pest; or (ii) the recombinant or synthetic nucleic acid molecule codes for a toxin or other factor directly involved in eliciting human, animal, or plant disease or inhibiting plant growth, and is carried on an expression vector or within the host chromosome and/or when the host organism contains a conjugation proficient plasmid or a generalized transducing phage; or (iii) the recombinant or synthetic nucleic acid molecule comes from a host organism or virus regulated as a human or animal pathogen or as a plant pest and has not been adequately characterized to demonstrate that it does not code for a factor involved in eliciting human, animal, or plant disease.

Appendix H-III. Footnotes and References of Appendix H

For further information on shipping etiologic agents contact: (i) The Centers for Disease Control and Prevention, ATTN: Biohazards Control Office, 1600 Clifton Road, Atlanta, Georgia 30333, (404) 639-3883, FTS 236-3883; (ii) The U.S. Department of Transportation, ATTN: Office of Hazardous Materials Transportation, 400 7th Street, S.W., Washington, DC 20590, (202) 366-4545; or (iii) U.S. Department of Agriculture, ATTN: Animal and Plant Health Inspection Service (APHIS), Veterinary Services, National Center for Import-Export, Products Program, 4700 River Road, Unit 40, Riverdale, Maryland 20737. Phone: (301) 734-8499; Fax: (301) 734-8226.

****

APPENDIX I. BIOLOGICAL CONTAINMENT (See Appendix E, *Certified Host-Vector Systems*)

Appendix I-I. Levels of Biological Containment

In consideration of biological containment, the vector (plasmid, organelle, or virus) for the recombinant or synthetic nucleic acid molecule and the host (bacterial, plant, or animal cell) in which the vector is propagated in the laboratory will be considered together. Any combination of vector and host which is to provide biological containment shall be chosen or constructed so that the following types of "escape" are minimized: (i) survival of the vector in its host outside the laboratory, and (ii) transmission of the vector from the propagation host to other non-laboratory hosts. The following levels of biological containment (host-vector systems) for prokaryotes are established. Appendices I-I-A through I-II-B describe levels of biological containment (host-vector systems) for prokaryotes. Specific criteria will depend on the organisms to be used.

Appendix I-I-A. Host-Vector 1 Systems

Host-Vector 1 systems provide a moderate level of containment. Specific Host-Vector 1 systems are:

Appendix I-I-A-1. *Escherichia coli* K-12 Host-Vector 1 Systems (EK1)

The host is always *Escherichia coli* K-12 or a derivative thereof, and the vectors include non-conjugative plasmids (e.g., pSC101, Co1E1, or derivatives thereof (see Appendices I-III-A through G, *Footnotes and References of Appendix I*) and variants of bacteriophage, such as lambda (see Appendices I-III-H through O, *Footnotes and References of Appendix I*). The *Escherichia coli* K-12 hosts shall not contain conjugation-proficient plasmids, whether autonomous or integrated, or generalized transducing phages.

Appendix I-I-A-2. Other Host-Vector 1 Systems

At a minimum, hosts and vectors shall be comparable in containment to *Escherichia coli* K-12 with a non-conjugative plasmid or bacteriophage vector. Appendix I-II, *Certification of Host-Vector Systems*, describes the data to be considered and mechanism for approval of Host-Vector 1 systems.

Appendix I-I-B. Host-Vector 2 Systems (EK2)

Host-Vector 2 Systems provide a high level of biological containment as demonstrated by data from suitable tests performed in the laboratory. Escape of the recombinant or synthetic nucleic acid molecule either via survival of the organisms or via transmission of the recombinant or synthetic nucleic acid molecule to other organisms should be < $1/10^8$ under specified conditions. Specific Host-Vector 2 systems are:

Appendix I-I-B-1. For *Escherichia coli* K-12 Host-Vector 2 systems (EK2) in which the vector is a plasmid, no

more than $1/10^8$ host cells shall perpetuate a cloned DNA fragment under the specified non-permissive laboratory conditions designed to represent the natural environment, either by survival of the original host or as a consequence of transmission of the cloned DNA fragment.

Appendix I-I-B-2. For *Escherichia coli* K-12 Host-Vector 2 systems (EK2) in which the vector is a phage, no more than $1/10^8$ phage particles shall perpetuate a cloned DNA fragment under the specified non-permissive laboratory conditions designed to represent the natural environment, either as a prophage (in the inserted or plasmid form) in the laboratory host used for phage propagation, or survival in natural environments and transferring a cloned DNA fragment to other hosts (or their resident prophages).

Appendix I-II. Certification of Host-Vector Systems

Appendix I-II-A. Responsibility

Host-Vector 1 systems (other than *Escherichia coli* K-12) and Host-Vector 2 systems may not be designated as such until they have been certified by the NIH Director. Requests for certification of host-vector systems may be submitted to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Proposed host-vector systems will be reviewed by the RAC (see Section IV-C-1-b-(1)-(f), *Major Actions*). Initial review will based on the construction, properties, and testing of the proposed host-vector system by a subcommittee composed of one or more RAC members and/or *ad hoc* experts. The RAC will evaluate the subcommittee's report and any other available information at the next scheduled RAC meeting. The NIH Director is responsible for certification of host-vector systems, following advice of the RAC. Minor modifications to existing host-vector systems (i.e., those that are of minimal or no consequence to the properties relevant to containment) may be certified by the NIH Director without prior RAC review (see Section IV-C-1-b-(2)-(f), *Minor Actions*). Once a host-vector system has been certified by the NIH Director, a notice of certification will be sent by NIH OSP to the applicant and to the Institutional Biosafety Committee Chairs. A list of all currently certified host-vector systems is available from the Office of Science Policy, National Institutes of Health, preferably by submitting a request for this information to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). The NIH Director may rescind the certification of a host-vector system (see Section IV-C-1-b-(2)-(g), *Minor Actions*). If certification is rescinded, NIH will instruct investigators to transfer cloned DNA into a different system or use the clones at a higher level of physical containment level, unless NIH determines that the already constructed clones incorporate adequate biological containment. Certification of a host-vector system does not extend to modifications of either the host or vector component of that system. Such modified systems shall be independently certified by the NIH Director. If modifications are minor, it may only be necessary for the investigator to submit data showing that the modifications have either improved or not impaired the major phenotypic traits on which the containment of the system depends. Substantial modifications to a certified host-vector system requires submission of complete testing data.

Appendix I-II-B. Data to be Submitted for Certification

Appendix I-II-B-1. Host-Vector 1 Systems Other than *Escherichia coli* K-12

The following types of data shall be submitted, modified as appropriate for the particular system under consideration: (i) a description of the organism and vector; the strain's natural habitat and growth requirements; its physiological properties, particularly those related to its reproduction, survival, and the mechanisms by which it exchanges genetic information; the range of organisms with which this organism normally exchanges genetic information and the type of information is exchanged; and any relevant information about its pathogenicity or toxicity; (ii) a description of the history of the particular strains and vectors to be used, including data on any mutations which render this organism less able to survive or transmit genetic information; and (iii) a general description of the range of experiments contemplated with emphasis on the need for developing such an Host-Vector 1 system.

Appendix I-II-B-2. Host-Vector 2 Systems

Investigators planning to request Host-Vector 2 systems certification may obtain instructions from NIH OSP concerning data to be submitted (see Appendices I-III-N and O, *Footnotes and References of Appendix I*). In general, the following types of data are required: (i) description of construction steps with indication of source, properties, and manner of introduction of genetic traits; (ii) quantitative data on the stability of genetic traits that

contribute to the containment of the system; (iii) data on the survival of the host-vector system under non-permissive laboratory conditions designed to represent the relevant natural environment; (iv) data on transmissibility of the vector and/or a cloned DNA fragment under both permissive and non-permissive conditions; (v) data on all other properties of the system which affect containment and utility, including information on yields of phage or plasmid molecules, ease of DNA isolation, and ease of transfection or transformation; and (vi) in some cases, the investigator may be asked to submit data on survival and vector transmissibility from experiments in which the host-vector is fed to laboratory animals or one or more human subjects. Such *in vivo* data may be required to confirm the validity of predicting *in vivo* survival on the basis of *in vitro* experiments. Data shall be submitted 12 weeks prior to the RAC meeting at which such data will be considered by the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Investigators are encouraged to publish their data on the construction, properties, and testing of proposed Host Vector 2 systems prior to consideration of the system by the RAC and its subcommittee. Specific instructions concerning the submission of data for proposed *Escherichia coli* K-12 Host-Vector 2 system (EK2) involving either plasmids or bacteriophage in *Escherichia coli* K-12, are available from the Office of Science Policy, National Institutes of Health, preferably by submitting a request for this information to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov).

Appendix I-III. Footnotes and References of Appendix I

Appendix I-III-A. Hersfield, V., H. W. Boyer, C. Yanofsky, M. A. Lovett, and D. R. Helinski, *Plasmid Co1E1 as a Molecular Vehicle for Cloning and Amplification of DNA*. Proc. Nat. Acad. Sci., 1974, 71, pp. 3455-3459.

Appendix I-III-B. Wensink, P. C., D. J. Finnegan, J. E. Donelson, and D. S. Hogness, *A System for Mapping DNA Sequences in the Chromosomes of Drosophila Melanogaster*. Cell, 1974, 3, pp. 315-335.

Appendix I-III-C. Tanaka, T., and B. Weisblum, *Construction of a Colicin El-R Factor Composite Plasmid in Vitro: Means for Amplification of Deoxyribonucleic Acid*. J. Bacteriol., 1975, *121*, pp. 354-362.

Appendix I-III-D. Armstrong, K. A., V. Hershfield, and D. R. Helinski, *Gene Cloning and Containment Properties of Plasmid Col E1 and Its Derivatives*, Science, 1977, *196*, pp. 172-174.

Appendix I-III-E. Bolivar, F., R. L. Rodriguez, M. C. Betlack, and H. W. Boyer, *Construction and Characterization of New Cloning Vehicles: I. Ampicillin-Resistant Derivative of PMB9*, Gene, 1977, 2, pp. 75-93.

Appendix I-III-F. Cohen, S. N., A. C. W. Chang, H. Boyer, and R. Helling. *Construction of Biologically Functional Bacterial Plasmids in Vitro*. Proc. Natl. Acad, Sci., 1973, 70, pp. 3240-3244.

Appendix I-III-G. Bolivar, F., R. L. Rodriguez, R. J. Greene, M. C.Batlack, H. L. Reyneker, H. W. Boyer, J. H. Cross, and S. Falkow, 1977, *Construction and Characterization of New Cloning Vehicles II. A Multi-Purpose Cloning System*, Gene, 1977, 2, pp. 95-113.

Appendix I-III-H. Thomas, M., J. R. Cameron, and R. W. Davis (1974). *Viable Molecular Hybrids of Bacteriophage Lambda and Eukaryotic DNA*. Proc. Nat. Acad. Sci., 1974, 71, pp. 4579-4583.

Appendix I-III-I. Murray, N. E., and K. Murray, *Manipulation of Restriction Targets in Phage Lambda to Form Receptor Chromosomes for DNA Fragments*. Nature, 1974, 51, pp. 476-481.

Appendix I-III-J. Ramback, A., and P. Tiollais (1974). *Bacteriophage Having EcoRI Endonuclease Sites Only in the Non-Essential Region of the Genome*. Proc. Nat. Acad. Sci., 1974, 71, pp. 3927-3820.

Appendix I-III-K. Blattner, F. R., B. G. Williams, A. E. Bleche, K. Denniston-Thompson, H. E. Faber, L. A. Furlong, D. J. Gunwald, D. O. Kiefer, D. D. Moore, J. W. Shumm, E. L. Sheldon, and O. Smithies, *Charon Phages: Safer Derivatives of Bacteriophage Lambda for DNA Cloning*, Science 1977, 196, pp. 163-169.

Appendix I-III-L. Donoghue, D. J., and P. A. Sharp, *An Improved Lambda Vector: Construction of Model Recombinants Coding for Kanamycin Resistance*, Gene, 1977, 1, pp. 209-227.

Appendix I-III-M. Leder, P., D. Tiemeier and L. Enquist (1977), *EK2 Derivatives of Bacteriophage Lambda Useful in the Cloning of DNA from Higher Organisms: The λgt WES System*, Science, 1977, 196, pp. 175-177.

Appendix I-III-N. Skalka, A., *Current Status of Coliphage AEK2 Vectors*, Gene, 1978, 3, pp. 29-35.

Appendix I-III-O. Szybalski, W., A. Skalka, S. Gottesman, A. Campbell, and D. Botstein, *Standardized Laboratory Tests for EK2 Certification*, Gene, 1978, 3, pp. 36-38.

****

APPENDIX J. BIOTECHNOLOGY RESEARCH SUBCOMMITTEE

The National Science and Technology Council's Committee on Fundamental Science determined that a subcommittee should be continued to identify and coordinate Federal research efforts, identify research needs, stimulating international cooperation, and assess national and international policy issues concerning biotechnology sciences. The primary emphasis will be on scientific issues to increase the overall effectiveness and productivity of the Federal investment in biotechnology sciences, especially regarding issues which cut across agency boundaries. This subcommittee is called the Biotechnology Research Subcommittee.

Membership of the Biotechnology Research Subcommittee will include Federal agencies that support biotechnology research. Agencies represented are: U.S. Department of Agriculture, Department of Commerce, Department of Defense, Department of Energy, Department of Health and Human Services, Department of Interior, Department of Justice, Department of State, Department of Veterans Affairs, Agency for International Development, Environmental Protection Agency, National Aeronautics and Space Administration, and National Science Foundation. The Biotechnology Research Subcommittee will function in an advisory capacity to the Committee on Fundamental Science, the Director of the Office of Science and Technology Policy, and the Executive Office of the President. The Biotechnology Research Subcommittee will review the scientific aspects of proposed regulations and guidelines as they are developed.

The primary responsibilities of the Biotechnology Research Subcommittee are to: (i) describe and review current Federal efforts in biotechnology research; (ii) identify and define the priority areas for future Federal biotechnology research, including areas needing greater emphasis, describing the role of each agency in those areas, and delineate where interagency cooperation would enhance progress in the biotechnology sciences, with an emphasis on integrated research efforts, where appropriate; (iii) assess major international efforts in the biotechnology sciences and develop mechanisms for international collaboration. For example, activities of the U.S.-European Community Task Force on Biotechnology have been coordinated through the Biotechnology Research Subcommittee; (iv) identify and review national and international policy issues (such as public education) associated with biotechnology; and (v) provide reviews, analyses, and recommendations to the Chairs of the Committee on Fundamental Science on scientific issues related to regulations and the applications of biotechnology research and biotechnology policies and issues.

In 1990, the Biotechnology Research Subcommittee replaced the Biotechnology Sciences Coordinating Committee. Both the Biotechnology Research Subcommittee and the Biotechnology Sciences Coordinating Committee previously functioned under the Federal Coordinating Council on Science, Engineering, and Technology (FCCSET). While regulatory issues became the primary focus of the Biotechnology Sciences Coordinating Committee, the Biotechnology Research Subcommittee focuses on scientific issues, although it will still provide scientific support for regulatory responsibilities.

****

APPENDIX K. PHYSICAL CONTAINMENT FOR LARGE SCALE USES OF ORGANISMS CONTAINING RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULES

Appendix K specifies physical containment guidelines for large-scale (greater than 10 liters of culture) research or production involving viable organisms containing recombinant or synthetic nucleic acid molecules. It shall apply to large-scale research or production activities as specified in Section III-D-6, *Experiments Involving More than 10 Liters of Culture*. It is important to note that this appendix addresses only the biological hazard associated with organisms containing recombinant or synthetic nucleic acid molecules. Other hazards

accompanying the large-scale cultivation of such organisms (e.g., toxic properties of products; physical, mechanical, and chemical aspects of downstream processing) are not addressed and shall be considered separately, albeit in conjunction with this appendix.

All provisions shall apply to large-scale research or production activities with the following modifications: (i) Appendix K shall supersede Appendix G, *Physical Containment*, when quantities in excess of 10 liters of culture are involved in research or production. Appendix K-II applies to Good Large Scale Practice; (ii) the institution shall appoint a Biological Safety Officer if it engages in large-scale research or production activities involving viable organisms containing recombinant or synthetic nucleic acid molecules. The duties of the Biological Safety Officer shall include those specified in Section IV-B-3, *Biological Safety Officer*; (iii) the institution shall establish and maintain a health surveillance program for personnel engaged in large-scale research or production activities involving viable organisms containing recombinant or synthetic nucleic acid molecules which require Biosafety Level (BL) 3 containment at the laboratory scale. The program shall include: pre assignment and periodic physical and medical examinations; collection, maintenance, and analysis of serum specimens for monitoring serologic changes that may result from the employee's work experience; and provisions for the investigation of any serious, unusual, or extended illnesses of employees to determine possible occupational origin.

Appendix K-I. Selection of Physical Containment Levels

The selection of the physical containment level required for recombinant or synthetic nucleic acid molecule research or production involving more than 10 liters of culture is based on the containment guidelines established in Section III, *Experiments Covered by the NIH Guidelines*. For purposes of large-scale research or production, four physical containment levels are established. The four levels set containment conditions at those appropriate for the degree of hazard to health or the environment posed by the organism, judged by experience with similar organisms unmodified by recombinant or synthetic nucleic acid molecule techniques and consistent with Good Large Scale Practice. The four biosafety levels of large-scale physical containment are referred to as Good Large Scale Practice, BL1-Large Scale, BL2-Large Scale, and BL3-Large Scale. Good Large Scale Practice is recommended for large-scale research or production involving viable, non-pathogenic, and non-toxigenic recombinant or synthetic strains derived from host organisms that have an extended history of safe large-scale use. Good Large Scale Practice is recommended for organisms such as those included in Appendix C, *Exemptions under Section III-F-8*, which have built-in environmental limitations that permit optimum growth in the large-scale setting but limited survival without adverse consequences in the environment. BL1-Large Scale is recommended for large-scale research or production of viable organisms containing recombinant or synthetic nucleic acid molecules that require BL1 containment at the laboratory scale and that do not qualify for Good Large Scale Practice. BL2-Large Scale is recommended for large-scale research or production of viable organisms containing recombinant or synthetic nucleic acid molecules that require BL2 containment at the laboratory scale. BL3-Large Scale is recommended for large-scale research or production of viable organisms containing recombinant or synthetic nucleic acid molecules that require BL3 containment at the laboratory scale. No provisions are made for large-scale research or production of viable organisms containing recombinant or synthetic nucleic acid molecules that require BL4 containment at the laboratory scale. If necessary, these requirements will be established by NIH on an individual basis.

Appendix K-II. Good Large Scale Practice (GLSP)

Appendix K-II-A. Institutional codes of practice shall be formulated and implemented to assure adequate control of health and safety matters.

Appendix K-II-B. Written instructions and training of personnel shall be provided to assure that cultures of viable organisms containing recombinant or synthetic nucleic acid molecules are handled prudently and that the work place is kept clean and orderly.

Appendix K-II-C. In the interest of good personal hygiene, facilities (e.g., hand washing sink, shower, changing room) and protective clothing (e.g., uniforms, laboratory coats) shall be provided that are appropriate for the risk of exposure to viable organisms containing recombinant or synthetic nucleic acid molecules. Eating, drinking, smoking, applying cosmetics, and mouth pipetting shall be prohibited in the work area.

Appendix K-II-D. Cultures of viable organisms containing recombinant or synthetic nucleic acid molecules shall be handled in facilities intended to safeguard health during work with microorganisms that do not require containment.

Appendix K-II-E. Discharges containing viable recombinant or synthetic organisms shall be handled in accordance with applicable governmental environmental regulations.

Appendix K-II-F. Addition of materials to a system, sample collection, transfer of culture fluids within/between systems, and processing of culture fluids shall be conducted in a manner that maintains employee's exposure to viable organisms containing recombinant or synthetic nucleic acid molecules at a level that does not adversely affect the health and safety of employees.

Appendix K-II-G. The facility's emergency response plan shall include provisions for handling spills.

Appendix K-III. Biosafety Level 1 (BL1) - Large Scale

Appendix K-III-A. Spills and accidents which result in overt exposures to organisms containing recombinant or synthetic nucleic acid molecules are immediately reported to the Laboratory Director. Medical evaluation, surveillance, and treatment are provided as appropriate and written records are maintained.

Appendix K-III-B. Cultures of viable organisms containing recombinant or synthetic nucleic acid molecules shall be handled in a closed system (e.g., closed vessel used for the propagation and growth of cultures) or other primary containment equipment (e.g., biological safety cabinet containing a centrifuge used to process culture fluids) which is designed to reduce the potential for escape of viable organisms. Volumes less than 10 liters may be handled outside of a closed system or other primary containment equipment provided all physical containment requirements specified in Appendix G-II-A, *Physical Containment Levels--Biosafety Level 1*, are met.

Appendix K-III-C. Culture fluids (except as allowed in Appendix K-III-D) shall not be removed from a closed system or other primary containment equipment unless the viable organisms containing recombinant or synthetic nucleic acid molecules have been inactivated by a validated inactivation procedure. A validated inactivation procedure is one which has been demonstrated to be effective using the organism that will serve as the host for propagating the recombinant or synthetic nucleic acid molecules. Culture fluids that contain viable organisms or viral vectors intended as final product may be removed from the primary containment equipment by way of closed systems for sample analysis, further processing or final fill.

Appendix K-III-D. Sample collection from a closed system, the addition of materials to a closed system, and the transfer of culture fluids from one closed system to another shall be conducted in a manner which minimizes the release of aerosols or contamination of exposed surfaces.

Appendix K-III-E. Exhaust gases removed from a closed system or other primary containment equipment shall be treated by filters which have efficiencies equivalent to high efficiency particulate air/HEPA filters or by other equivalent procedures (e.g., incineration) to minimize the release of viable organisms containing recombinant or synthetic nucleic acid molecules to the environment.

Appendix K-III-F. A closed system or other primary containment equipment that has contained viable organisms containing recombinant or synthetic nucleic acid molecules shall not be opened for maintenance or other purposes unless it has been sterilized by a validated sterilization procedure except when the culture fluids contain viable organisms or vectors intended as final product as described in Appendix K-III-C above. A validated sterilization procedure is one which has been demonstrated to be effective using the organism that will serve as the host for propagating the recombinant or synthetic nucleic acid molecules.

Appendix K-III-G. Emergency plans required by Sections IV-B-2-b-(6), *Institutional Biosafety Committee*, and IV-B-3-c-(3), *Biological Safety Officer*, shall include methods and procedures for handling large losses of culture on an emergency basis.

Appendix K-IV. Biosafety Level 2 (BL2) - Large Scale

Appendix K-IV-A. Spills and accidents which result in overt exposures to organisms containing recombinant or synthetic nucleic acid molecules are immediately reported to the Biological Safety Officer, Institutional Biosafety Committee, NIH OSP, and other appropriate authorities (if applicable). Reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Medical

evaluation, surveillance, and treatment are provided as appropriate and written records are maintained.

Appendix K-IV-B. Cultures of viable organisms containing recombinant or synthetic nucleic acid molecules shall be handled in a closed system (e.g., closed vessel used for the propagation and growth of cultures) or other primary containment equipment (e.g., Class III biological safety cabinet containing a centrifuge used to process culture fluids) which is designed to prevent the escape of viable organisms. Volumes less than 10 liters may be handled outside of a closed system or other primary containment equipment provided all physical containment requirements specified in Appendix G-II-B, *Physical Containment Levels--Biosafety Level 2*, are met.

Appendix K-IV-C. Culture fluids (except as allowed in Appendix K-IV-D) shall not be removed from a closed system or other primary containment equipment unless the viable organisms containing recombinant or synthetic nucleic acid molecules have been inactivated by a validated inactivation procedure. A validated inactivation procedure is one which has been demonstrated to be effective using the organism that will serve as the host for propagating the recombinant or synthetic nucleic acid molecules. Culture fluids that contain viable organisms or viral vectors intended as final product may be removed from the primary containment equipment by way of closed systems for sample analysis, further processing or final fill.

Appendix K-IV-D. Sample collection from a closed system, the addition of materials to a closed system, and the transfer of cultures fluids from one closed system to another shall be conducted in a manner which prevents the release of aerosols or contamination of exposed surfaces.

Appendix K-IV-E. Exhaust gases removed from a closed system or other primary containment equipment shall be treated by filters which have efficiencies equivalent to high efficiency particulate air/HEPA filters or by other equivalent procedures (e.g., incineration) to prevent the release of viable organisms containing recombinant or synthetic nucleic acid molecules to the environment.

Appendix K-IV-F. A closed system or other primary containment equipment that has contained viable organisms containing recombinant or synthetic nucleic acid molecules shall not be opened for maintenance or other purposes unless it has been sterilized by a validated sterilization procedure except when the culture fluids contain viable organisms or vectors intended as final product as described in Appendix K-IV-C above. A validated sterilization procedure is one which has been demonstrated to be effective using the organisms that will serve as the host for propagating the recombinant or synthetic nucleic acid molecules.

Appendix K-IV-G. Rotating seals and other mechanical devices directly associated with a closed system used for the propagation and growth of viable organisms containing recombinant or synthetic nucleic acid molecules shall be designed to prevent leakage or shall be fully enclosed in ventilated housings that are exhausted through filters which have efficiencies equivalent to high efficiency particulate air/HEPA filters or through other equivalent treatment devices.

Appendix K-IV-H. A closed system used for the propagation and growth of viable organisms containing recombinant or synthetic nucleic acid molecules and other primary containment equipment used to contain operations involving viable organisms containing sensing devices that monitor the integrity of containment during operations.

Appendix K-IV-I. A closed system used for the propagation and growth of viable organisms containing the recombinant or synthetic nucleic acid molecules shall be tested for integrity of the containment features using the organism that will serve as the host for propagating recombinant or synthetic nucleic acid molecules. Testing shall be accomplished prior to the introduction of viable organisms containing recombinant or synthetic nucleic acid molecules and following modification or replacement of essential containment features. Procedures and methods used in the testing shall be appropriate for the equipment design and for recovery and demonstration of the test organism. Records of tests and results shall be maintained on file.

Appendix K-IV-J. A closed system used for the propagation and growth of viable organisms containing recombinant or synthetic nucleic acid molecules shall be permanently identified. This identification shall be used in all records reflecting testing, operation, and maintenance and in all documentation relating to use of this equipment for research or production activities involving viable organisms containing recombinant or synthetic nucleic acid molecules.

Appendix K-IV-K. The universal biosafety sign shall be posted on each closed system and primary

containment equipment when used to contain viable organisms containing recombinant or synthetic nucleic acid molecules.

Appendix K-IV-L. Emergency plans required by Sections IV-B-2-b-(6), *Institutional Biosafety Committee*, and IV-B-3-c-(3), *Biological Safety Officer*, shall include methods and procedures for handling large losses of culture on an emergency basis.

Appendix K-V. Biosafety Level 3 (BL3) - Large Scale

Appendix K-V-A. Spills and accidents which result in overt exposures to organisms containing recombinant or synthetic nucleic acid molecules are immediately reported to the Biological Safety Officer, Institutional Biosafety Committee, NIH OSP, and other appropriate authorities (if applicable). Reports to NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Medical evaluation, surveillance, and treatment are provided as appropriate and written records are maintained.

Appendix K-V-B. Cultures of viable organisms containing recombinant or synthetic nucleic acid molecules shall be handled in a closed system (e.g., closed vessels used for the propagation and growth of cultures) or other primary containment equipment (e.g., Class III biological safety cabinet containing a centrifuge used to process culture fluids) which is designed to prevent the escape of viable organisms. Volumes less than 10 liters may be handled outside of a closed system provided all physical containment requirements specified in Appendix G-II-C, *Physical Containment Levels--Biosafety Level 3*, are met.

Appendix K-V-C. Culture fluids (except as allowed in Appendix K-V-D) shall not be removed from a closed system or other primary containment equipment unless the viable organisms containing recombinant or synthetic nucleic acid molecules have been inactivated by a validated inactivation procedure. A validated inactivation procedure is one which has been demonstrated to be effective using the organisms that will serve as the host for propagating the recombinant or synthetic nucleic acid molecules. Culture fluids that contain viable organisms or viral vectors intended as final product may be removed from the primary containment equipment by way of closed systems for sample analysis, further processing or final fill.

Appendix K-V-D. Sample collection from a closed system, the addition of materials to a closed system, and the transfer of culture fluids from one closed system to another shall be conducted in a manner which prevents the release of aerosols or contamination of exposed surfaces.

Appendix K-V-E. Exhaust gases removed from a closed system or other primary containment equipment shall be treated by filters which have efficiencies equivalent to high efficiency particulate air/HEPA filters or by other equivalent procedures (e.g., incineration) to prevent the release of viable organisms containing recombinant or synthetic nucleic acid molecules to the environment.

Appendix K-V-F. A closed system or other primary containment equipment that has contained viable organisms containing recombinant or synthetic nucleic acid molecules shall not be opened for maintenance or other purposes unless it has been sterilized by a validated sterilization procedure except when the culture fluids contain viable organisms or vectors intended as final product as described in Appendix K-V-C above. A validated sterilization procedure is one which has been demonstrated to be effective using the organisms that will serve as the host for propagating the recombinant or synthetic nucleic acid molecules.

Appendix K-V-G. A closed system used for the propagation and growth of viable organisms containing recombinant or synthetic nucleic acid molecules shall be operated so that the space above the culture level will be maintained at a pressure as low as possible, consistent with equipment design, in order to maintain the integrity of containment features.

Appendix K-V-H. Rotating seals and other mechanical devices directly associated with a closed system used to contain viable organisms containing recombinant or synthetic nucleic acid molecules shall be designed to prevent leakage or shall be fully enclosed in ventilated housings that are exhausted through filters which have efficiencies equivalent to high efficiency particulate air/HEPA filters or through other equivalent treatment devices.

Appendix K-V-I. A closed system used for the propagation and growth of viable organisms containing recombinant or synthetic nucleic acid molecules and other primary containment equipment used to contain

operations involving viable organisms containing recombinant or synthetic nucleic acid molecules shall include monitoring or sensing devices that monitor the integrity of containment during operations.

Appendix K-V-J. A closed system used for the propagation and growth of viable organisms containing recombinant or synthetic nucleic acid molecules shall be tested for integrity of the containment features using the organisms that will serve as the host for propagating the recombinant or synthetic nucleic acid molecules. Testing shall be accomplished prior to the introduction of viable organisms containing recombinant or synthetic nucleic acid molecules and following modification or replacement of essential containment features. Procedures and methods used in the testing shall be appropriate for the equipment design and for recovery and demonstration of the test organism. Records of tests and results shall be maintained on file.

Appendix K-V-K. A closed system used for the propagation and growth of viable organisms containing recombinant or synthetic nucleic acid molecules shall be permanently identified. This identification shall be used in all records reflecting testing, operation, maintenance, and use of this equipment for research production activities involving viable organisms containing recombinant or synthetic nucleic acid molecules.

Appendix K-V-L. The universal biosafety sign shall be posted on each closed system and primary containment equipment when used to contain viable organisms containing recombinant or synthetic nucleic acid molecules.

Appendix K-V-M. Emergency plans required by Sections IV-B-2-b-(6), *Institutional Biosafety Committee*, and IV-B-3-c-(3), *Biological Safety Officer*, shall include methods and procedures for handling large losses of culture on an emergency basis.

Appendix K-V-N. Closed systems and other primary containment equipment used in handling cultures of viable organisms containing recombinant or synthetic nucleic acid molecules shall be located within a controlled area which meets the following requirements:

Appendix K-V-N-1. The controlled area shall have a separate entry area. The entry area shall be a double-doored space such as an air lock, anteroom, or change room that separates the controlled area from the balance of the facility.

Appendix K-V-N-2. The surfaces of walls, ceilings, and floors in the controlled area shall be such as to permit ready cleaning and decontamination.

Appendix K-V-N-3. Penetrations into the controlled area shall be sealed to permit liquid or vapor phase space decontamination.

Appendix K-V-N-4. All utilities and service or process piping and wiring entering the controlled area shall be protected against contamination.

Appendix K-V-N-5. Hand washing facilities equipped with foot, elbow, or automatically operated valves shall be located at each major work area and near each primary exit.

Appendix K-V-N-6. A shower facility shall be provided. This facility shall be located in close proximity to the controlled area.

Appendix K-V-N-7. The controlled area shall be designed to preclude release of culture fluids outside the controlled area in the event of an accidental spill or release from the closed systems or other primary containment equipment.

Appendix K-V-N-8. The controlled area shall have a ventilation system that is capable of controlling air movement. The movement of air shall be from areas of lower contamination potential to areas of higher contamination potential. If the ventilation system provides positive pressure supply air, the system shall operate in a manner that prevents the reversal of the direction of air movement or shall be equipped with an alarm that would be actuated in the event that reversal in the direction of air movement were to occur. The exhaust air from the controlled area shall not be recirculated to other areas of the facility. The exhaust air from the controlled area may not be discharged to the outdoors without being high efficiency particulate air/HEPA filtered, subjected to thermal oxidation, or otherwise treated to prevent the release of viable organisms.

Appendix K-V-O. The following personnel and operational practices shall be required:

Appendix K-V-O-1. Personnel entry into the controlled area shall be through the entry area specified in Appendix K-V-N-1.

Appendix K-V-O-2. Persons entering the controlled area shall exchange or cover their personal clothing with work garments such as jump suits, laboratory coats, pants and shirts, head cover, and shoes or shoe covers. On exit from the controlled area the work clothing may be stored in a locker separate from that used for personal clothing or discarded for laundering. Clothing shall be decontaminated before laundering.

Appendix K-V-O-3. Entry into the controlled area during periods when work is in progress shall be restricted to those persons required to meet program or support needs. Prior to entry, all persons shall be informed of the operating practices, emergency procedures, and the nature of the work conducted.

Appendix K-V-O-4. Persons under 18 years of age shall not be permitted to enter the controlled area.

Appendix K-V-O-5. The universal biosafety sign shall be posted on entry doors to the controlled area and all internal doors when any work involving the organism is in progress. This includes periods when decontamination procedures are in progress. The sign posted on the entry doors to the controlled area shall include a statement of agents in use and personnel authorized to enter the controlled area.

Appendix K-V-O-6. The controlled area shall be kept neat and clean.

Appendix K-V-O-7. Eating, drinking, smoking, and storage of food are prohibited in the controlled area.

Appendix K-V-O-8. Animals and plants shall be excluded from the controlled area.

Appendix K-V-O-9. An effective insect and rodent control program shall be maintained.

Appendix K-V-O-10. Access doors to the controlled area shall be kept closed, except as necessary for access, while work is in progress. Service doors leading directly outdoors shall be sealed and locked while work is in progress.

Appendix K-V-O-11. Persons shall wash their hands when exiting the controlled area.

Appendix K-V-O-12. Persons working in the controlled area shall be trained in emergency procedures.

Appendix K-V-O-13. Equipment and materials required for the management of accidents involving viable organisms containing recombinant or synthetic nucleic acid molecules shall be available in the controlled area.

Appendix K-V-O-14. The controlled area shall be decontaminated in accordance with established procedures following spills or other accidental release of viable organisms containing recombinant or synthetic nucleic acid molecules.

Appendix K - Table 1. Comparison of Good Large Scale Practice (GLSP) and Biosafety Level (BL) - Large Scale (LS) Practice (See Appendix K-VI-A, *Footnotes Of Appendix K*)

CRITERION [See Appendix K-VI-B, *Footnotes of Appendix K*]		GLSP	BL1-LS	BL2-LS	BL3-LS
1.	Formulate and implement institutional codes of practice for safety of personnel and adequate control of hygiene and safety measures.	K-II-A	G-I		
2.	Provide adequate written instructions and training of personnel to keep work place clean and tidy and to keep exposure to biological, chemical or physical agents at a level that does not adversely affect health and safety of employees.	K-II-B	G-I		
3.	Provide changing and hand washing facilities as well as protective clothing, appropriate to the risk, to be worn during work.	K-II-C	G-II-A-1-h	G-II-B-2-f	G-II-C-2-i
4.	Prohibit eating, drinking, smoking, mouth pipetting, and applying cosmetics in the work place.	K-II-C	G-II-A-1-d G-II-A-1-e	G-II-B-1-d G-II-B-1-e	G-II-C-1-c G-II-C-1-d
5.	Internal accident reporting.	K-II-G	K-III-A	K-IV-A	K-V-A
6.	Medical surveillance.	NR	NR		
7.	Viable organisms should be handled in a system that physically separates the process from the external environment (closed system or other primary containment).	NR	K-III-B	K-IV-B	K-V-B
8.	Culture fluids not removed from a system until organisms are inactivated.	NR	K-III-C	K-IV-C	K-V-C
9.	Inactivation of waste solutions and materials with respect to their biohazard potential.	K-II-E	K-III-C	K-IV-C	K-V-C
10.	Control of aerosols by engineering or procedural controls to prevent or minimize release of organisms during sampling from a system, addition of materials to a system, transfer of cultivated cells, and removal of material, products, and effluent from a system.	Minimize *Procedure* K-II-F	Minimize *Engineer* K-III-B K-III-D	Prevent *Engineer* K-IV-B K-IV-D	Prevent *Engineer* K-V-B K-V-D
11.	Treatment of exhaust gases from a closed system to minimize or prevent release of viable organisms.	NR	Minimize K-III-E	Prevent K-IV-E	Prevent K-V-E
12.	Closed system that has contained viable organisms not to be opened until sterilized by a validated procedure.	NR	K-III-F	K-IV-F	K-V-F
13.	Closed system to be maintained at as a low pressure as possible to maintain integrity of containment features.	NR	NR	NR	K-V-G
14.	Rotating seals and other penetrations into closed system designed to prevent or minimize leakage.	NR	NR	Prevent K-IV-G	Prevent K-V-H
15.	Closed system shall incorporate monitoring or sensing devices to monitor the integrity of containment.	NR	NR	K-IV-H	K-V-I
16.	Validated integrity testing of closed containment system.	NR	NR	K-IV-I	K-V-J
17.	Closed system to be permanently identified for record keeping purposes.	NR	NR	K-IV-J	K-V-K
18.	Universal biosafety sign to be posted on each closed system.	NR	NR	K-IV-K	K-V-L
19.	Emergency plans required for handling large losses of cultures.	K-II-G	K-III-G	K-IV-L	K-V-M
20.	Access to the work place.	NR	G-II-A-1-a	G-II-B-1-a	K-V-N
21.	Requirements for controlled access area.	NR	NR	NR	K-V-N&O

NR = not required

Appendix K-VI. Footnotes of Appendix K

Appendix K-VI-A. This table is derived from the text in Appendices G (*Physical Containment*) and K and is not to be used in lieu of Appendices G and K.

Appendix K-VI-B. The criteria in this grid address only the biological hazards associated with organisms containing recombinant or synthetic nucleic acid. Other hazards accompanying the large-scale cultivation of such organisms (e.g., toxic properties of products; physical, mechanical, and chemical aspects of downstream processing) are not addressed and shall be considered separately, albeit in conjunction with this grid.

Appendix K-VII. Definitions to Accompany Containment Grid and Appendix K

Appendix K-VII-A. Accidental Release. An accidental release is the unintentional discharge of a microbiological agent (i.e., microorganism or virus) or eukaryotic cell due to a failure in the containment system.

Appendix K-VII-B. Biological Barrier. A biological barrier is an impediment (naturally occurring or introduced) to the infectivity and/or survival of a microbiological agent or eukaryotic cell once it has been released into the environment.

Appendix K-VII-C. Closed System. A closed system is one in which by its design and proper operation, prevents release of a microbiological agent or eukaryotic cell contained therein.

Appendix K-VII-D. Containment. Containment is the confinement of a microbiological agent or eukaryotic cell that is being cultured, stored, manipulated, transported, or destroyed in order to prevent or limit its contact with people and/or the environment. Methods used to achieve this include: physical and biological barriers and inactivation using physical or chemical means.

Appendix K-VII-E. *De minimis* Release. *De minimis* release is the release of: (i) viable microbiological agents or eukaryotic cells that does not result in the establishment of disease in healthy people, plants, or animals; or (ii) in uncontrolled proliferation of any microbiological agents or eukaryotic cells.

Appendix K-VII-F. Disinfection. Disinfection is a process by which viable microbiological agents or eukaryotic cells are reduced to a level unlikely to produce disease in healthy people, plants, or animals.

Appendix K-VII-G. Good Large Scale Practice Organism. For an organism to qualify for Good Large Scale Practice consideration, it must meet the following criteria [Reference: Organization for Economic Cooperation and Development, *Recombinant DNA Safety Considerations*, 1987, p. 34-35]: (i) the host organism should be non-pathogenic, should not contain adventitious agents and should have an extended history of safe large-scale use or have built-in environmental limitations that permit optimum growth in the large-scale setting but limited survival without adverse consequences in the environment; (ii) the recombinant or synthetic nucleic acid molecule-engineered organism should be non-pathogenic, should be as safe in the large-scale setting as the host organism, and without adverse consequences in the environment; and (iii) the vector/insert should be well characterized and free from known harmful sequences; should be limited in size as much as possible to the DNA required to perform the intended function; should not increase the stability of the construct in the environment unless that is a requirement of the intended function; should be poorly mobilizable; and should not transfer any resistance markers to microorganisms unknown to acquire them naturally if such acquisition could compromise the use of a drug to control disease agents in human or veterinary medicine or agriculture.

Appendix K-VII-H. Inactivation. Inactivation is any process that destroys the ability of a specific microbiological agent or eukaryotic cell to self-replicate.

Appendix K-VII-I. Incidental Release. An incidental release is the discharge of a microbiological agent or eukaryotic cell from a containment system that is expected when the system is appropriately designed and properly operated and maintained.

Appendix K-VII-J. Minimization. Minimization is the design and operation of containment systems in order that any incidental release is a *de minimis* release.

Appendix K-VII-K. Pathogen. A pathogen is any microbiological agent or eukaryotic cell containing sufficient

genetic information, which upon expression of such information, is capable of producing disease in healthy people, plants, or animals.

Appendix K-VII-L. Physical Barrier. A physical barrier is considered any equipment, facilities, or devices (e.g., fermentors, factories, filters, thermal oxidizers) which are designed to achieve containment.

Appendix K-VII-M. Release. Release is the discharge of a microbiological agent or eukaryotic cell from a containment system. Discharges can be incidental or accidental. Incidental releases are *de minimis* in nature; accidental releases may be *de minimis* in nature.

****

APPENDIX L. GENE THERAPY POLICY CONFERENCES (GTPCS)

In order to enhance the depth and value of public discussion relevant to scientific, safety, social, and ethical implications of gene therapy research, the NIH Director will convene GTPCs at regular intervals. As appropriate, the NIH Director may convene a GTPC in conjunction with a RAC meeting. GTPCs will be administered by NIH OSP. Conference participation will not involve a standing committee membership but rather will offer the unique advantage of assembling numerous participants who possess significant scientific, ethical, and legal expertise and/or interest that is directly applicable to a specific gene therapy research issue. At least one member of RAC will serve as Co-chair of each GTPC and report the findings of each GTPC to RAC at its next scheduled meeting. The RAC representative for each GTPC will be chosen based on the participant's area of expertise relative to the specific gene therapy research issue to be discussed. All RAC members will be invited to attend GTPCs. GTPCs will have representation from other Federal agencies, including FDA and OHRP. GTPCs will focus on broad overarching policy and scientific issues related to gene therapy research. Proposals for GTPC topics may be submitted by members of RAC, representatives of academia, industry, patient and consumer advocacy organizations, other Federal agencies, professional scientific societies, and the general public. GTPC topics will not be limited to discussion of human applications of gene therapy research, i.e., they may include basic research on the use of novel gene delivery vehicles, or novel applications of human gene transfer. The RAC, with the Director's approval, will have the primary responsibility for planning GTPC agendas. GTPC findings will be transmitted to the NIH Director and will be made publicly available. The NIH Director anticipates that this public policy forum will serve as a model for interagency communication and collaboration, concentrated expert discussion of novel scientific issues and their potential societal implications, and enhanced opportunity for public discussion of specific issues and potential impact of such applications on human health and the environment.

****

APPENDIX M. POINTS TO CONSIDER IN THE DESIGN AND SUBMISSION OF PROTOCOLS FOR THE TRANSFER OF RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULES INTO ONE OR MORE HUMAN RESEARCH PARTICIPANTS (POINTS TO CONSIDER)

Appendix M applies to research conducted at or sponsored by an institution that receives any support for recombinant or synthetic nucleic acid molecule research from NIH. Researchers not covered by the *NIH Guidelines* are encouraged to use Appendix M (see Section I-C, *General Applicability*).

The acceptability of human somatic cell gene transfer has been addressed in several public documents as well as in numerous academic studies. In November 1982, the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research published a report, *Splicing Life*, which resulted from a two-year process of public deliberation and hearings. Upon release of that report, a U.S. House of Representatives subcommittee held three days of public hearings with witnesses from a wide range of fields from the biomedical and social sciences to theology, philosophy, and law. In December 1984, the Office of Technology Assessment released a background paper, *Human Gene Therapy*, which concluded that civic, religious, scientific, and medical groups have all accepted, in principle, the appropriateness of gene transfer of somatic cells in humans for specific genetic diseases. Somatic cell gene transfer is seen as an extension of present methods that might be preferable to other technologies. In light of this public support, the NIH is prepared to consider proposals for somatic cell gene transfer.

The NIH will not at present entertain proposals for germ line alterations but will consider proposals involving somatic cell gene transfer. The purpose of somatic cell gene transfer is to treat an individual patient, e.g., by inserting a properly functioning gene into the subject's somatic cells. Germ line alteration involves a specific attempt to introduce genetic changes into the germ (reproductive) cells of an individual, with the aim of changing the set of genes passed on to the individual's offspring.

The NIH continues to explore the issues raised by the potential of *in utero* gene transfer clinical research. However, the NIH concludes that, at present, it is premature to undertake any *in utero* gene transfer clinical trial. Significant additional preclinical and clinical studies addressing vector transduction efficacy, biodistribution, and toxicity are required before a human *in utero* gene transfer protocol can proceed. In addition, a more thorough understanding of the development of human organ systems, such as the immune and nervous systems, is needed to better define the potential efficacy and risks of human *in utero* gene transfer. Prerequisites for considering any specific human *in utero* gene transfer procedure include an understanding of the pathophysiology of the candidate disease and a demonstrable advantage to the *in utero* approach. Once the above criteria are met, the NIH would be willing to consider well rationalized human *in utero* gene transfer clinical trials.

Research proposals involving the deliberate transfer of recombinant or synthetic nucleic acid molecules, or DNA or RNA derived from such nucleic acid molecules, into one or more human subjects (human gene transfer) will be considered through a registration process involving the NIH, oversight bodies involved in the review at an initial site(s), and regulatory authorities, when appropriate. Investigators shall submit the relevant information on the proposed human gene transfer experiment to the oversight bodies involved in the review at an initial site(s) and then to the NIH. The format of the submission is described in Appendix M-I-A, *Requirements for Protocol Submission*. Submission to the NIH OSP shall be for registration purposes and will ensure continued public access to relevant human gene transfer information conducted in compliance with the *NIH Guidelines*.

Public RAC review and discussion of a human gene transfer experiment will be initiated in two exceptional circumstances: (1) Following a request for public RAC review from one or more oversight bodies involved in the review at an initial site(s), the NIH concurs that (a) the individual protocol would significantly benefit from RAC review and (b) that the submission meets one or more of the following NIH RAC review criteria: i) the protocol uses a new vector, genetic material, or delivery methodology that represents a first-in-human experience, thus presenting an unknown risk; ii) the protocol relies on preclinical safety data that were obtained using a new preclinical model system of unknown and unconfirmed value; or iii) the proposed vector, gene construct, or method of delivery is associated with possible toxicities that are not widely known and that may render it difficult for oversight bodies involved in the review at an initial site(s) to evaluate the protocol rigorously. However, if one or more oversight bodies involved in the review at an initial site(s) requests public RAC review, but the NIH does not concur that (a) the individual protocol would significantly benefit from RAC review and (b) that the submission meets one or more of the RAC review criteria (listed in i, ii, or iii), then the NIH OSP will inform, within 10 working days, the requesting and other oversight bodies involved in the review at an initial site(s) that public RAC review is not warranted. (2) The NIH Director, in consultation (if needed) with appropriate regulatory authorities, determines that the submission: (a) meets one or more of the NIH RAC review criteria (listed in i, ii, or iii) and that public RAC review and discussion would provide a clear and obvious benefit to the scientific community or the public; or (b) raises significant scientific, societal, or ethical concerns.

If it is determined that a human gene transfer trial will undergo public RAC review, the NIH will immediately notify the Principal Investigator. Recommendations following public review on a specific human gene transfer experiment shall be forwarded to the Principal Investigator, oversight bodies involved in the review at an initial site(s), and regulatory authorities, as appropriate. Relevant documentation will be included in the material for the RAC meeting at which the human gene transfer trial is scheduled to be discussed. RAC meetings will be open to the public except where trade secrets and proprietary information are reviewed (see Section IV-D-5, *Protection of Proprietary Data – Voluntary Compliance*). Information provided in response to Appendix M should not contain any proprietary data or trade secrets, enabling all aspects of the review to be open to the public.

Appendix M-I. Requirements for Protocol Submission, Review, and Reporting – Human Gene Transfer Experiments

Appendix M-I-A. Requirements for Protocol Submission

The following documentation must be submitted according to institutional policy, to the appropriate oversight

bodies involved in the review at an initial site(s) and subsequently in electronic form to the NIH OSP:

1. A scientific abstract.
2. The proposed clinical protocol, including tables, figures, and any relevant publications.
3. Summary of preclinical studies conducted in support of the proposed clinical trial or reference to the specific section of the protocol providing this information.
4. A description of the product:
    a. Describe the derivation of the delivery vector system including the source (e.g., viral, bacterial, or plasmid vector); and modifications (e.g., deletions to attenuate or self-inactivate, encapsulation in any synthetic complex, changes to tropisms, etc.). Please reference any previous clinical experience with this vector or similar vectors.
    b. Describe the genetic content of the transgene or nucleic acid delivered including the species source of the sequence and whether any modifications have been made (e.g. mutations, deletions, and truncations). What are the regulatory elements contained in the construct?
    c. Describe any other material to be used in preparation of the agent (vector and transgene) that will be administered to the human research subject (e.g., helper virus, packaging cell line, carrier particles).
    d. Describe the methods for replication-competent virus testing, if applicable.
    e. Describe the intended *ex vivo* or *in vivo* target cells and transduction efficiency.
    f. Describe the gene transfer agent delivery method.
5. The proposed informed consent document(s).
6. Specifically for submission to the NIH OSP, the Principal Investigator shall provide additional documentation originating from oversight bodies involved in the review at an initial site(s) regarding their assessment of whether public RAC review is warranted. In the event that review is requested, a justification that one or more of the NIH RAC review criteria (see Section III-C-1) are met shall be included.

Note: Any application submitted shall not contain any document that is designated as 'confidential' in its entirety. In the event that a determination has been made that a specific portion of a document should be considered proprietary or trade secret, each specific portion shall be clearly identified as such. In the event that a specific portion of the submission is identified to be proprietary or trade secret, the submission to the NIH OSP must contain a letter that: (1) clearly indicates what select portions of the application contain information considered as proprietary or trade secret, and (2) provides justification as to why this information is considered to be proprietary or trade secret. The justification must be able to demonstrate with specificity how release of that information will reveal a trade secret or will result in substantial competitive harm.

Appendix M-I-B. Selection of Individual Protocols for Public RAC Review and Discussion

As part of the NIH protocol registration process, documentation originating from all oversight bodies involved in the review at an initial site(s) regarding their assessment of whether public RAC review is warranted must accompany the Principal Investigator's submission to the NIH. If no oversight body involved in the review at an initial site(s) requests public RAC review, then the required documentation to register the protocol (see Appendix M-I-A) shall be submitted to the NIH OSP at any time, but not less than 10 working days prior to the anticipated date of enrollment of the first subject (see definition of enrollment in Section I-E-7). This information shall be provided in electronic form to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: HGTprotocols@mail.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). An acknowledgement that the protocol registration process is complete will occur within the 10 working days period prior to the anticipated date of enrollment. Final IBC approval may then be granted.

If one or more oversight bodies involved in the review at an initial site(s) requests public RAC review, but the NIH does not concur that (a) the individual protocol would significantly benefit from RAC review and (b) that the submission meets one or more of the RAC review criteria, the NIH OSP will notify the Principal Investigator, oversight bodies involved in the review at an initial site(s), and regulatory authorities, as appropriate, that public RAC review is not warranted. An acknowledgement that the protocol registration process is complete will accompany this decision. Final IBC approval may then be granted.

If an oversight body involved in the review at an initial site(s) determines that: (1) a protocol submission would significantly benefit from public RAC review and discussion and (2) that one or more of the following NIH RAC review criteria are met: (i) the protocol uses a new vector, genetic material, or delivery methodology that represents a first-in-human experience, thus presenting an unknown risk; or (ii) the protocol relies on preclinical

safety data that were obtained using a new preclinical model system of unknown and unconfirmed value; or (iii) the proposed vector, gene construct, or method of delivery is associated with possible toxicities that are not widely known and that may render it difficult for oversight and federal regulatory bodies to evaluate the protocol rigorously, and is therefore requesting RAC review and public discussion. The Principal Investigator shall then submit the documentation as outlined in Appendix M-I-A at least 8 weeks prior to the next scheduled meeting in order to be reviewed at that RAC meeting. The submission shall include documentation originating from oversight bodies involved in the review at an initial site(s) regarding their assessment of whether public RAC review is warranted and that one or both have justified their request according the NIH RAC review criteria listed above. The submission shall be provided to the NIH in electronic form to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: HGTprotocols@mail.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). If the NIH concurs that (a) the individual protocol would significantly benefit from RAC review and (b) that the submission meets one or more of the NIH RAC review criteria, the protocol will undergo public RAC review and discussion.

Even if an oversight body involved in the review at an initial site(s) does not request public RAC review, the NIH Director, after consultation (if needed) with appropriate regulatory authorities, may initiate public RAC review if (a) the protocol has one or more of the characteristics listed above (i, ii, or iii) and public RAC review and discussion would provide a clear and obvious benefit to the scientific community or public; or (b) the protocol otherwise raises significant scientific, societal, or ethical concerns. If a protocol is to undergo RAC public discussion a complete human gene transfer protocol package must be submitted at least 8 weeks before a scheduled RAC meeting to be reviewed at that upcoming meeting.

After a human gene transfer experiment is publicly reviewed by the full RAC at a regularly scheduled meeting, the NIH OSP will send a letter summarizing the RAC's comments and recommendations (if any) regarding the protocol to the Principal Investigator(s), oversight bodies involved in the review at an initial site(s), and regulatory authorities as appropriate. Unless the NIH determines that there are exceptional circumstances, the NIH will send this letter to the Principal Investigator within 10 working days after the completion of the RAC meeting at which the experiment was reviewed. Receipt of this letter concludes the protocol registration process. Final IBC approval may then be granted.

RAC meetings will be open to the public except where trade secrets or confidential commercial information are reviewed. To enable all aspects of the protocol review process to be open to the public, information provided in response to Appendix M-I-A should not contain trade secrets or confidential commercial or financial information. Documentation submitted to the NIH OSP shall not be designated as 'confidential' in its entirety. In the event that a determination has been made that a specific portion of a document submitted should be considered as proprietary or trade secret, each specific portion should be clearly identified as such. The cover letter (attached to the submitted material) shall: (1) clearly indicate what select portions contain information considered as proprietary or a trade secret; and (2) provide justification as to why this information is considered to be proprietary or trade secret. This justification must be able to demonstrate with specificity how release of that information will reveal a trade secret or will result in substantial competitive harm.

Appendix M-I-C. Reporting Requirements

Appendix M-I-C-1. Initiation of the Clinical Investigation

No later than 30 days after enrollment (see definition of enrollment in Section I-E-7) of the first research participant in a human gene transfer experiment, the Principal Investigator(s) shall submit the following documentation to NIH OSP: (1) a copy of the informed consent document(s) approved by the Institutional Review Board (IRB); (2) a copy of the protocol approved by the Institutional Biosafety Committee (IBC) and IRB; (3) a copy of the final IBC approval from the clinical trial site; (4) a copy of the final IRB approval; (5) a brief written report that includes the following information: (a) how the investigator(s) responded to each of the RAC's recommendations on the protocol (if applicable); and (b) any modifications to the protocol as required by FDA; (6) applicable NIH grant number(s); (7) the FDA Investigational New Drug Application (IND) number; and (8) the date of the initiation of the trial. The purpose of requesting the FDA IND number is for facilitating interagency collaboration in the Federal oversight of human gene transfer research.

Appendix M-I-C-2. Additional Clinical Trial Sites

For a clinical trial site that is added after the completion of the NIH protocol registration process, no research participant shall be enrolled (see definition of enrollment in Section I-E-7) at the clinical trial site until IBC

approval and IRB approval from that site have been obtained. Within 30 days of enrollment (see definition of enrollment in Section I-E-7) at a clinical trial site, the following documentation shall be submitted to NIH OSP: (1) Institutional Biosafety Committee approval (from the clinical trial site); (2) Institutional Review Board approval; (3) Institutional Review Board-approved informed consent document(s); and (4) NIH grant number(s) if applicable.

Appendix M-I-C-3. Annual Reports

Within 60 days after the one-year anniversary of the date on which the investigational new drug (IND) application went into effect, and after each subsequent anniversary until the trial is completed, the Principal Investigator (or delegate) shall submit the information set forth in (a), (b), and (c). When multiple studies are conducted under the single IND, the Principal Investigator (or delegate) may choose to submit a single annual report covering all studies, provided that each study is identified by its NIH protocol number.

(a) Clinical Trial Information. A brief summary of the status of each trial in progress and each trial completed during the previous year. The summary is required to include the following information for each trial: (1) the title and purpose of the trial; (2) clinical site; (3) the Principal Investigator; (4) clinical protocol identifiers, including the NIH protocol number, NIH grant number(s) (if applicable), and the FDA IND application number; (5) participant population (such as disease indication and general age group, e.g., adult or pediatric); (6) the total number of participants planned for inclusion in the trial; the number entered into the trial to date; the number

whose participation in the trial was completed; and the number who dropped out of the trial with a brief description of the reasons; (7) the status of the trial, e.g., open to accrual of subjects, closed but data collection ongoing, or fully completed, and (8) if the trial has been completed, a brief description of any study results.

(b) Progress Report and Data Analysis. Information obtained during the previous year's clinical and non-clinical investigations, including: (1) a narrative or tabular summary showing the most frequent and most serious adverse experiences by body system; (2) a summary of all serious adverse events submitted during the past year; (3) a summary of serious adverse events that were expected or considered to have causes not associated with the use of the gene transfer product such as disease progression or concurrent medications; (4) if any deaths have occurred, the number of participants who died during participation in the investigation and causes of death; and (5) a brief description of any information obtained that is pertinent to an understanding of the gene transfer product's actions, including, for example, information about dose-response, information from controlled trials, and information about bioavailability.

(c) A copy of the updated clinical protocol including a technical abstract.

Appendix M-I-C-4. Safety Reporting

Principal Investigators must submit, in accordance with this section, Appendix M-I-C-4-a and Appendix M-I-C-4-b, a written report on: (1) any serious adverse event that is both unexpected and associated with the use of the gene transfer product (i.e., there is reasonable possibility that the event may have been caused by the use of the product; investigators should not await definitive proof of association before reporting such events); and (2) any finding from tests in laboratory animals that suggests a significant risk for human research participants including reports of mutagenicity, teratogenicity, or carcinogenicity. The report must be clearly labeled as a "Safety Report" and must be submitted to the NIH Office of Science Policy (NIH OSP) and to the local Institutional Biosafety Committee within the timeframes set forth in Appendix M-I-C-4-b.

Principal Investigators should adhere to any other serious adverse event reporting requirements in accordance with federal regulations, state laws, and local institutional policies and procedures, as applicable.

Principal Investigators may delegate to another party, such as a corporate sponsor, the reporting functions set forth in Appendix M, with written notification to the NIH OSP of the delegation and of the name(s), address, telephone and fax numbers of the contact(s). The Principal Investigator is responsible for ensuring that the reporting requirements are fulfilled and will be held accountable for any reporting lapses.

The three alternative mechanisms for reporting serious adverse events to the NIH OSP are: by e-mail to: HGTprotocols@mail.nih.gov; by fax to 301-496-9839; or by mail to the Office of Science Policy, National Institutes of Health, MSC 7985, 6705 Rockledge Drive, Suite 750, Bethesda, Maryland 20892-7985 (20817 for non-USPS mail).

Appendix M-I-C-4-a. Safety Reporting: Content and Format

The serious adverse event report must include, but need not be limited to: (1) the date of the event; (2) designation of the report as an initial report or a follow-up report, identification of all safety reports previously filed for the clinical protocol concerning a similar adverse event, and an analysis of the significance of the adverse event in light of previous similar reports; (3) clinical site; (4) the Principal Investigator; (5) NIH Protocol number; (6) FDA's Investigational New Drug (IND) Application number; (7) vector type , e.g., adenovirus; (8) vector subtype, e.g., type 5, relevant deletions; (9) gene delivery method, e.g., *in vivo, ex vivo* transduction; (10) route of administration, e.g., intratumoral, intravenous; (11) dosing schedule; (12) a complete description of the event; (13) relevant clinical observations; (14) relevant clinical history; (15) relevant tests that were or are planned to be conducted; (16) date of any treatment of the event; and (17) the suspected cause of the event. These items may be reported by using the recommended Adverse Event Reporting Template available on NIH OSP's web site at: http://osp.od.nih.gov/office-biotechnology-activities/biomedical-technology-assessment/hgt/gemcris, the FDA MedWatch forms, or other means provided that all of the above elements are specifically included.

Reports from laboratory animal studies as delineated in Appendix M-I-C-4 must be submitted in a narrative format.

Appendix M-I-C-4-b. Safety Reporting: Time frames for Expedited Reports

Any serious adverse event that is fatal or life-threatening, that is unexpected, and associated with the use of the gene transfer product must be reported to the NIH OSP as soon as possible, but not later than 7 calendar days after the sponsor's initial receipt of the information (i.e., at the same time the event must be reported to the FDA).

Serious adverse events that are unexpected and associated with the use of the gene transfer product, but are not fatal or life-threatening, must be reported to the NIH OSP as soon as possible, but not later than 15 calendar days after the sponsor's initial receipt of the information (i.e., at the same time the event must be reported to the FDA).

Changes in this schedule are permitted only where, under the FDA IND regulations [21 CFR 312(c)(3)], changes in this reporting schedule have been approved by the FDA and are reflected in the protocol.

If, after further evaluation, an adverse event initially considered not to be associated with the use of the gene transfer product is subsequently determined to be associated, then the event must be reported to the NIH OSP within 15 days of the determination.

Relevant additional clinical and laboratory data may become available following the initial serious adverse event report. Any follow-up information relevant to a serious adverse event must be reported within 15 calendar days of the sponsor's receipt of the information. If a serious adverse event occurs after the end of a clinical trial and is determined to be associated with the use of the gene transfer product, that event shall be reported to the NIH OSP within 15 calendar days of the determination.

Any finding from tests in laboratory animals that suggests a significant risk for human research participants including reports of mutagenicity, teratogenicity, or carcinogenicity must be reported as soon as possible, but not later than 15 calendar days after the sponsor's initial receipt of the information (i.e., at the same time the event must be reported to the FDA).

Appendix M-I-C-5. Confidentiality

Data submitted in accordance with Appendix M-I-C that are claimed to be confidential commercial or trade secret information must be clearly labeled as such. Prior to making its determination about the confidentiality of data labeled confidential commercial or trade secret, the NIH will contact the Principal Investigator or delegate to ascertain the basis for the claim and subsequently will notify the Principal Investigator or delegate of its final determination regarding the claim.

If NIH determines that the data so labeled are confidential commercial or trade secret and that their public disclosure would promote an understanding of key scientific or safety issues, the NIH will seek agreement from

the appropriate party to release such data. Public discussion of scientific and safety issues raised by data submitted in accordance with Appendix M-I-C is vital to informing both investigators and human subjects about the safety of gene transfer research.

To protect the privacy of participants in gene transfer research, any serious adverse event or annual reports submitted to NIH OSP must not contain any information that would identify the human research participants.

Appendix M-I-D. Safety Assessment in Human Gene Transfer Research

A working group of the RAC, the NIH Gene Transfer Safety Assessment Board, with staff support from the NIH OSP, will: 1) review in closed session as appropriate safety information from gene transfer trials for the purpose of assessing toxicity and safety data across gene transfer trials; 2) identify significant trends or significant single events; and 3) report significant findings and aggregated trend data to the RAC. It is expected that this process will enhance review of new protocols, improve the development, design, and conduct of human gene transfer trials, promote public understanding and awareness of the safety of human gene transfer research studies, and inform the decision-making of potential trial participants.

Appendix M-II. Long-Term Follow-Up

To permit evaluation of long-term safety and efficacy of gene transfer, prospective subjects should be informed that they are expected to cooperate in long-term follow-up that extends beyond the active phase of the study. A list of persons who can be contacted in the event that questions arise during the follow-up period should be provided to the investigator. In addition, the investigator should request that subjects continue to provide a current address and telephone number.

The subjects should be informed that any significant findings resulting from the study will be made known in a timely manner to them and/or their parent or guardian including new information about the experimental procedure, the harms and benefits experienced by other individuals involved in the study, and any long-term effects that have been observed.

Additional guidance is available in the FDA Guidance for Industry: Gene Therapy Clinical Trials - Observing Subjects for Delayed Adverse Events (available at the following URL: http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/CellularandGeneTherapy/default.htm).

Appendix M-III. Footnotes of Appendix M

Appendix M-III-A. Human studies in which induction or enhancement of an immune response to a vector-encoded microbial immunogen is the major goal, such an immune response has been demonstrated in model systems, and the persistence of the vector-encoded immunogen is not expected, are exempt from Appendix M-I, *Requirements for Protocol Submission*, *Review and Reporting – Human Gene Transfer Experiments*.

****

APPENDIX P. PHYSICAL AND BIOLOGICAL CONTAINMENT FOR RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULE RESEARCH INVOLVING PLANTS

Appendix P specifies physical and biological containment conditions and practices suitable to the greenhouse conduct of experiments involving recombinant or synthetic nucleic acid molecule-containing plants, plant-associated microorganisms, and small animals. All provisions of the *NIH Guidelines* apply to plant research activities with the following modifications:

Appendix P shall supersede Appendix G (*Physical Containment*) when the research plants are of a size, number, or have growth requirements that preclude the use of containment conditions described in Appendix G. The plants covered in Appendix P include but are not limited to mosses, liverworts, macroscopic algae, and vascular plants including terrestrial crops, forest, and ornamental species.

Plant-associated microorganisms include viroids, virusoids, viruses, bacteria, fungi, protozoans, certain small algae, and microorganisms that have a benign or beneficial association with plants, such as certain *Rhizobium*

species and microorganisms known to cause plant diseases. The appendix applies to microorganisms which are being modified with the objective of fostering an association with plants.

Plant-associated small animals include those arthropods that: (i) are in obligate association with plants, (ii) are plant pests, (iii) are plant pollinators, or (iv) transmit plant disease agents, as well as other small animals such as nematodes for which tests of biological properties necessitate the use of plants. Microorganisms associated with such small animals (e.g., pathogens or symbionts) are included.

The Institutional Biosafety Committee shall include at least one individual with expertise in plant, plant pathogen, or plant pest containment principles when experiments utilizing Appendix P require prior approval by the Institutional Biosafety Committee.

Appendix P-I. General Plant Biosafety Levels

Appendix P-I-A. The principal purpose of plant containment is to avoid the unintentional transmission of a recombinant or synthetic nucleic acid molecule-containing plant genome, including nuclear or organelle hereditary material or release of recombinant or synthetic nucleic acid molecule-derived organisms associated with plants.

Appendix P-I-B. The containment principles are based on the recognition that the organisms that are used pose no health threat to humans or higher animals (unless deliberately modified for that purpose), and that the containment conditions minimize the possibility of an unanticipated deleterious effect on organisms and ecosystems outside of the experimental facility, e.g., the inadvertent spread of a serious pathogen from a greenhouse to a local agricultural crop or the unintentional introduction and establishment of an organism in a new ecosystem.

Appendix P-I-C. Four biosafety levels, referred to as Biosafety Level (BL) 1 - Plants (P), BL2-P, BL3-P, and BL4-P, are established in Appendix P-II, *Physical Containment Levels*. The selection of containment levels required for research involving recombinant or synthetic nucleic acid molecules in plants or associated with plants is specified in Appendix P-III, *Biological Containment Practices*. These biosafety levels are described in Appendix P-II, *Physical Containment Levels*. This appendix describes greenhouse practices and special greenhouse facilities for physical containment.

Appendix P-I-D. BL1-P through BL4-P are designed to provide differential levels of biosafety for plants in the absence or presence of other experimental organisms that contain recombinant or synthetic nucleic acid molecules. These biosafety levels, in conjunction with biological containment conditions described in Appendix P-III, *Biological Containment Practices*, provide flexible approaches to ensure the safe conduct of research.

Appendix P-I-E. For experiments in which plants are grown at the BL1 through BL4 laboratory settings, containment practices shall be followed as described in Appendix G, *Physical Containment*. These containment practices include the use of plant tissue culture rooms, growth chambers within laboratory facilities, or experiments performed on open benches. Additional biological containment practices should be added by the Greenhouse Director or Institutional Biosafety Committee as necessary (see Appendix P-III, *Biological Containment Practices*), if botanical reproductive structures are produced that have the potential of being released.

Appendix P-II. Physical Containment Levels

Appendix P-II-A. Biosafety Level 1 - Plants (BL1-P)

Appendix P-II-A-1. Standard Practices (BL1-P)

Appendix P-II-A-1-a. Greenhouse Access (BL1-P)

Appendix P-II-A-1-a-(1). Access to the greenhouse shall be limited or restricted, at the discretion of the Greenhouse Director, when experiments are in progress.

Appendix P-II-A-1-a-(2). Prior to entering the greenhouse, personnel shall be required to read and follow instructions on BL1-P greenhouse practices and procedures. All procedures shall be performed in accordance with accepted greenhouse practices that are appropriate to the experimental organism.

Appendix P-II-A-1-b. Records (BL1-P)

Appendix P-II-A-1-b-(1). A record shall be kept of experiments currently in progress in the greenhouse facility.

Appendix P-II-A-1-c. Decontamination and Inactivation (BL1-P)

Appendix P-II-A-1-c-(1). Experimental organisms shall be rendered biologically inactive by appropriate methods before disposal outside of the greenhouse facility.

Appendix P-II-A-1-d. Control of Undesired Species and Motile Macroorganisms (BL1-P)

Appendix P-II-A-1-d-(1). A program shall be implemented to control undesired species (e.g., weed, rodent, or arthropod pests and pathogens), by methods appropriate to the organisms and in accordance with applicable state and Federal laws.

Appendix P-II-A-1-d-(2). Arthropods and other motile macroorganisms shall be housed in appropriate cages. If macroorganisms (e.g., flying arthropods or nematodes) are released within the greenhouse, precautions shall be taken to minimize escape from the greenhouse facility.

Appendix P-II-A-1-e. Concurrent Experiments Conducted in the Greenhouse (BL1-P)

Appendix P-II-A-1-e-(1). Experiments involving other organisms that require a containment level lower than BL1-P may be conducted in the greenhouse concurrently with experiments that require BL1-P containment, provided that all work is conducted in accordance with BL1-P greenhouse practices.

Appendix P-II-A-2. Facilities (BL1-P)

Appendix P-II-A-2-a. Definitions (BL1-P)

Appendix P-II-A-2-a-(1). The term "greenhouse" refers to a structure with walls, a roof, and a floor designed and used principally for growing plants in a controlled and protected environment. The walls and roof are usually constructed of transparent or translucent material to allow passage of sunlight for plant growth.

Appendix P-II-A-2-a-(2). The term "greenhouse facility" includes the actual greenhouse rooms or compartments for growing plants, including all immediately contiguous hallways and head-house areas, and is considered part of the confinement area.

Appendix P-II-A-2-b. Greenhouse Design (BL1-P)

Appendix P-II-A-2-b-(1). The greenhouse floor may be composed of gravel or other porous material. At a minimum, impervious (e.g., concrete) walkways are recommended.

Appendix P-II-A-2-b-(2). Windows and other openings in the walls and roof of the greenhouse facility may be open for ventilation as needed for proper operation and do not require any special barrier to contain or exclude pollen, microorganisms, or small flying animals (e.g., arthropods and birds); however, screens are recommended.

Appendix P-II-B. Biosafety Level 2 - Plants (BL2-P)

Appendix P-II-B-1. Standard Practices (BL2-P)

Appendix P-II-B-1-a. Greenhouse Access (BL2-P)

Appendix P-II-B-1-a-(1). Access to the greenhouse shall be limited or restricted, at the discretion of the Greenhouse Director, to individuals directly involved with the experiments when they are in progress.

Appendix P-II-B-1-a-(2). Personnel shall be required to read and follow instructions on BL2-P practices and procedures. All procedures shall be conducted in accordance with accepted greenhouse practices that are appropriate to the experimental organisms.

Appendix P-II-B-1-b. Records (BL2-P)

Appendix P-II-B-1-b-(1). A record shall be kept of experimental plants, microorganisms, or small animals that are brought into or removed from the greenhouse facility.

Appendix P-II-B-1-b-(2). A record shall be kept of experiments currently in progress in the greenhouse facility.

Appendix P-II-B-1-b-(3). The Principal Investigator shall report any greenhouse accident involving the inadvertent release or spill of microorganisms to the Greenhouse Director, Institutional Biosafety Committee, NIH OSP and other appropriate authorities immediately (if applicable). Reports to the NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Documentation of any such accident shall be prepared and maintained.

Appendix P-II-B-1-c. Decontamination and Inactivation (BL2-P)

Appendix P-II-B-1-c-(1). Experimental organisms shall be rendered biologically inactive by appropriate methods before disposal outside of the greenhouse facility.

Appendix P-II-B-1-c-(2). Decontamination of run-off water is not necessarily required. If part of the greenhouse is composed of gravel or similar material, appropriate treatments should be made periodically to eliminate, or render inactive, any organisms potentially entrapped by the gravel.

Appendix P-II-B-1-d. Control of Undesired Species and Motile Macroorganisms (BL2-P)

Appendix P-II-B-1-d-(1). A program shall be implemented to control undesired species (e.g., weed, rodent, or arthropod pests and pathogens) by methods appropriate to the organisms and in accordance with applicable state and Federal laws.

Appendix P-II-B-1-d-(2). Arthropods and other motile macroorganisms shall be housed in appropriate cages. If macroorganisms (e.g., flying arthropods or nematodes) are released within the greenhouse, precautions shall be taken to minimize escape from the greenhouse facility.

Appendix P-II-B-1-e. Concurrent Experiments Conducted in the Greenhouse (BL2-P)

Appendix P-II-B-1-e-(1). Experiments involving other organisms that require a containment level lower than BL2-P may be conducted in the greenhouse concurrently with experiments that require BL2-P containment provided that all work is conducted in accordance with BL2-P greenhouse practices.

Appendix P-II-B-1-f. Signs (BL2-P)

Appendix P-II-B-1-f-(1). A sign shall be posted indicating that a restricted experiment is in progress. The sign shall indicate the following: (i) the name of the responsible individual, (ii) the plants in use, and (iii) any special requirements for using the area.

Appendix P-II-B-1-f-(2). If organisms are used that have a recognized potential for causing serious detrimental impacts on managed or natural ecosystems, their presence shall be indicated on a sign posted on the greenhouse access doors.

Appendix P-II-B-1-f-(3). If there is a risk to human health, a sign shall be posted incorporating the universal biosafety symbol.

Appendix P-II-B-1-g. Transfer of Materials (BL2-P)

Appendix P-II-B-1-g-(1). Materials containing experimental microorganisms, which are brought into or removed from the greenhouse facility in a viable or intact state, shall be transferred in a closed non-breakable container.

Appendix P-II-B-1-h. Greenhouse Practices Manual (BL2-P)

Appendix P-II-B-1-h-(1). A greenhouse practices manual shall be prepared or adopted. This manual shall: (i) advise personnel of the potential consequences if such practices are not followed, and (ii) outline contingency plans to be implemented in the event of the unintentional release of organisms.

Appendix P-II-B-2. Facilities (BL2-P)

Appendix P-II-B-2-a. Definitions (BL2-P)

Appendix P-II-B-2-a-(1). The term "greenhouse" refers to a structure with walls, a roof, and a floor designed and used principally for growing plants in a controlled and protected environment. The walls and roof are usually constructed of transparent or translucent material to allow passage of sunlight for plant growth.

Appendix P-II-B-2-a-(2). The term "greenhouse facility" includes the actual greenhouse rooms or compartments for growing plants, including all immediately contiguous hallways and head-house areas and is considered part of the confinement area.

Appendix P-II-B-2-b. Greenhouse Design (BL2-P)

Appendix P-II-B-2-b-(1). A greenhouse floor composed of an impervious material. Concrete is recommended, but gravel or other porous material under benches is acceptable unless propagules of experimental organisms are readily disseminated through soil. Soil beds are acceptable unless propagules of experimental organisms are readily disseminated through soil.

Appendix P-II-B-2-b-(2). Windows and other openings in the walls and roof of the greenhouse facility may be open for ventilation as needed for proper operation and do not require any special barrier to exclude pollen or microorganisms; however, screens are required to exclude small flying animals (e.g., arthropods and birds).

Appendix P-II-B-2-c. Autoclaves (BL2-P)

Appendix P-II-B-2-c-(1). An autoclave shall be available for the treatment of contaminated greenhouse materials.

Appendix P-II-B-2-d. Supply and Exhaust Air Ventilation Systems (BL2-P)

Appendix P-II-B-2-d-(1). If intake fans are used, measures shall be taken to minimize the ingress of arthropods. Louvers or fans shall be constructed such that they can only be opened when the fan is in operation.

Appendix P-II-B-2-e. Other (BL2-P)

Appendix P-II-B-2-e-(1). BL2-P greenhouse containment requirements may be satisfied by using a growth chamber or growth room within a building provided that the external physical structure limits access and escape of microorganisms and macroorganisms in a manner that satisfies the intent of the foregoing clauses.

Appendix P-II-C. Biosafety Level 3 - Plants (BL3-P)

Appendix P-II-C-1. Standard Practices (BL3-P)

Appendix P-II-C-1-a. Greenhouse Access (BL3-P)

Appendix P-II-C-1-a-(1). Authorized entry into the greenhouse shall be restricted to individuals who are required for program or support purposes. The Greenhouse Director shall be responsible for assessing each circumstance and determining those individuals who are authorized to enter the greenhouse facility.

Appendix P-II-C-1-a-(2). Prior to entering the greenhouse, personnel shall be required to read and follow instructions on BL3-P practices and procedures. All procedures shall be conducted in accordance with accepted greenhouse practices that are appropriate to the experimental organisms.

Appendix P-II-C-1-b. Records (BL3-P)

Appendix P-II-C-1-b-(1). A record shall be kept of experimental plants, microorganisms, or small animals that are brought into or removed from the greenhouse facility.

Appendix P-II-C-1-b-(2). A record shall be kept of experiments currently in progress in the greenhouse facility.

Appendix P-II-C-1-b-(3). The Principal Investigator shall report any greenhouse accident involving the inadvertent release or spill of microorganisms to the Biological Safety Officer, Greenhouse Director, Institutional Biosafety Committee, NIH OSP, and other appropriate authorities immediately (if applicable). Reports to the NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Documentation of any such accident shall be prepared and maintained.

Appendix P-II-C-1-c. Decontamination and Inactivation (BL3-P)

Appendix P-II-C-1-c-(1). All experimental materials shall be sterilized in an autoclave or rendered biologically inactive by appropriate methods before disposal, except those that are to remain in a viable or intact state for experimental purposes; including water that comes in contact with experimental microorganisms or with material exposed to such microorganisms, and contaminated equipment and supplies.

Appendix P-II-C-1-d. Control of Undesired Species and Motile Macroorganisms (BL3-P)

Appendix P-II-C-1-d-(1). A program shall be implemented to control undesired species (e.g., weed, rodent, or arthropod pests and pathogens) by methods appropriate to the organisms and in accordance with applicable state and Federal laws.

Appendix P-II-C-1-d-(2). Arthropods and other motile macroorganisms shall be housed in appropriate cages. When appropriate to the organism, experiments shall be conducted within cages designed to contain the motile organisms.

Appendix P-II-C-1-e. Concurrent Experiments Conducted in the Greenhouse (BL3-P)

Appendix P-II-C-1-e-(1). Experiments involving organisms that require a containment level lower than BL3-P may be conducted in the greenhouse concurrently with experiments that require BL3-P containment provided that all work is conducted in accordance with BL3-P greenhouse practices.

Appendix P-II-C-1-f. Signs (BL3-P)

Appendix P-II-C-1-f-(1). A sign shall be posted indicating that a restricted experiment is in progress. The sign shall indicate the following: (i) the name of the responsible individual, (ii) the plants in use, and (iii) any special requirements for using the area.

Appendix P-II-C-1-f-(2). If organisms are used that have a recognized potential for causing serious detrimental impacts on managed or natural ecosystems, their presence should be indicated on a sign posted on the greenhouse access doors.

Appendix P-II-C-1-f-(3). If there is a risk to human health, a sign shall be posted incorporating the universal biosafety symbol.

Appendix P-II-C-1-g. Transfer of Materials (BL3-P)

Appendix P-II-C-1-g-(1). Experimental materials that are brought into or removed from the greenhouse facility in a viable or intact state shall be transferred to a non-breakable sealed secondary container. At the time of transfer, if the same plant species, host, or vector are present within the effective dissemination distance of propagules of the experimental organism, the surface of the secondary container shall be decontaminated. Decontamination may be accomplished by passage through a chemical disinfectant or fumigation chamber or by an alternative procedure that has demonstrated effective inactivation of the experimental organism.

Appendix P-II-C-1-h. Greenhouse Practices Manual (BL3-P)

Appendix P-II-C-1-h-(1). A greenhouse practices manual shall be prepared or adopted. This manual shall: (i) advise personnel of the potential consequences if such practices are not followed, and (ii) outline contingency plans to be implemented in the event of the unintentional release of organisms with recognized potential for serious detrimental impact.

Appendix P-II-C-1-i. Protective Clothing (BL3-P)

Appendix P-II-C-1-i-(1). Disposable clothing (e.g., solid front or wrap-around gowns, scrub suits, or other appropriate clothing) shall be worn in the greenhouse if deemed necessary by the Greenhouse Director because of potential dissemination of the experimental microorganisms.

Appendix P-II-C-1-i-(2). Protective clothing shall be removed before exiting the greenhouse and decontaminated prior to laundering or disposal.

Appendix P-II-C-1-j. Other (BL3-P)

Appendix P-II-C-1-j-(1). Personnel are required to thoroughly wash their hands upon exiting the greenhouse.

Appendix P-II-C-1-j-(2). All procedures shall be performed carefully to minimize the creation of aerosols and excessive splashing of potting material/soil during watering, transplanting, and all experimental manipulations.

Appendix P-II-C-2. Facilities (BL3-P)

Appendix P-II-C-2-a. Definitions (BL3-P)

Appendix P-II-C-2-a-(1). The term "greenhouse" refers to a structure with walls, roof, and floor designed and used principally for growing plants in a controlled and protected environment. The walls and roof are usually constructed of transparent or translucent material to allow passage of sunlight for plant growth.

Appendix P-II-C-2-a-(2). The term "greenhouse facility" includes the actual greenhouse rooms or compartments for growing plants, including all immediately contiguous hallways and head-house areas, and is considered part of the confinement area. The need to maintain negative pressure should be considered when constructing or renovating the greenhouse.

Appendix P-II-C-2-b. Greenhouse Design (BL3-P)

Appendix P-II-C-2-b-(1). The greenhouse floor shall be composed of concrete or other impervious material with provision for collection and decontamination of liquid run-off.

Appendix P-II-C-2-b-(2). Windows shall be closed and sealed. All glazing shall be resistant to breakage (e.g., double-pane tempered glass or equivalent).

Appendix P-II-C-2-b-(3). The greenhouse shall be a closed self-contained structure with a continuous covering that is separated from areas that are open to unrestricted traffic flow. The minimum requirement for greenhouse entry shall be passage through two sets of self-closing locking doors.

Appendix P-II-C-2-b-(4). The greenhouse facility shall be surrounded by a security fence or protected by equivalent security measures.

Appendix P-II-C-2-b-(5). Internal walls, ceilings, and floors shall be resistant to penetration by liquids and chemicals to facilitate cleaning and decontamination of the area. All penetrations into these structures and surfaces (e.g., plumbing and utilities) shall be sealed.

Appendix P-II-C-2-b-(6). Bench tops and other work surfaces should have seamless surfaces that are impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat.

Appendix P-II-C-2-b-(7). The greenhouse contains a foot, elbow, or automatically operated sink, which is located near the exit door for hand washing.

Appendix P-II-C-2-c. Autoclaves (BL3-P)

Appendix P-II-C-2-c-(1). An autoclave shall be available for decontaminating materials within the greenhouse facility. A double-door autoclave is recommended (not required) for the decontamination of materials passing out of the greenhouse facility.

Appendix P-II-C-2-d. Supply and Exhaust Air Ventilation Systems (BL3-P)

Appendix P-II-C-2-d-(1). An individual supply and exhaust air ventilation system shall be provided. The system maintains pressure differentials and directional airflow, as required, to assure inward (or zero) airflow from areas outside of the greenhouse.

Appendix P-II-C-2-d-(2). The exhaust air from the greenhouse facility shall be filtered through high efficiency particulate air-HEPA filters and discharged to the outside. The filter chambers shall be designed to allow *in situ* decontamination before filters are removed and to facilitate certification testing after they are replaced. Air filters shall be 80-85% average efficiency by the American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) Standard 52-68 test method using atmosphere dust. Air supply fans shall be equipped with a back-flow damper that closes when the air supply fan is off. Alternatively, a HEPA filter may be used on the air supply system instead of the filters and damper. The supply and exhaust airflow shall be interlocked to assure inward (or zero) airflow at all times.

Appendix P-II-C-2-e. Other (BL3-P)

Appendix P-II-C-2-e-(1). BL3-P greenhouse containment requirements may be satisfied using a growth chamber or growth room within a building provided that the location, access, airflow patterns, and provisions for decontamination of experimental materials and supplies meet the intent of the foregoing clauses.

Appendix P-II-C-2-e-(2). Vacuum lines shall be protected with high efficiency particulate air/HEPA or equivalent filters and liquid disinfectant traps.

Appendix P-II-D. Biosafety Level 4 - Plants (BL4-P)

Appendix P-II-D-1. Standard Practices (BL4-P)

Appendix P-II-D-1-a. Greenhouse Access (BL4-P)

Appendix P-II-D-1-a-(1). Authorized entry into the greenhouse shall be restricted to individuals who are required for program or support purposes. The Greenhouse Director shall be responsible for assessing each circumstance and determining those individuals who are authorized to enter the greenhouse facility or work in the greenhouse during experiments.

Appendix P-II-D-1-a-(2). Access shall be managed by the Greenhouse Director, Biological Safety Officer, or other individual responsible for physical security of the greenhouse facility; and access limited by means of secure, locked doors.

Appendix P-II-D-1-a-(3). Prior to entering, individuals shall be advised of the potential environmental hazards and instructed on appropriate safeguards for ensuring environmental safety. Individuals authorized to enter the greenhouse facility shall comply with the instructions and all other applicable entry/exit procedures.

Appendix P-II-D-1-a-(4). Personnel shall enter and exit the greenhouse facility only through the clothing change and shower rooms and shall shower each time they exit the greenhouse facility. Personnel shall use the airlocks to enter or exit the laboratory only in an emergency. In the event of an emergency, every reasonable effort should be made to prevent the possible transport of viable propagules from containment.

Appendix P-II-D-1-a-(5). Prior to entering the greenhouse, personnel shall be required to read and follow instructions on BL4-P practices and procedures.

Appendix P-II-D-1-b. Records (BL4-P)

Appendix P-II-D-1-b-(1). A record shall be kept of all experimental materials brought into or removed from the greenhouse.

Appendix P-II-D-1-b-(2). A record shall be kept of experiments currently in progress in the greenhouse facility.

Appendix P-II-D-1-b-(3). A record shall be kept of all personnel entering and exiting the greenhouse facility, including the date and time of each entry.

Appendix P-II-D-1-b-(4). The Principal Investigator shall report any greenhouse accident involving the inadvertent release or spill of microorganisms to the Biological Safety Officer, Greenhouse Director, Institutional Biosafety Committee, NIH OSP, and other appropriate authorities immediately (if applicable). Reports to the NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Documentation of any such accident shall be prepared and maintained.

Appendix P-II-D-1-c. Decontamination and Inactivation (BL4-P)

Appendix P-II-D-1-c-(1). All materials, except for those that are to remain in a viable or intact state for experimental purposes, shall be autoclaved prior to removal from the maximum containment greenhouse. Equipment or material that could be damaged by high temperatures or steam shall be decontaminated by alternative methods (e.g., gas or vapor sterilization) in an airlock or chamber designed for this purpose.

Appendix P-II-D-1-c-(2). Water that comes in contact with experimental microorganisms or with material exposed to such microorganisms (e.g., run-off from watering plants) shall be collected and decontaminated before disposal.

Appendix P-II-D-1-c-(3). Standard microbiological procedures shall be followed for decontamination of equipment and materials. Spray or liquid waste or rinse water from containers used to apply the experimental microorganisms shall be decontaminated before disposal.

Appendix P-II-D-1-d. Control of Undesired Species and Motile Macroorganisms (BL4-P)

Appendix P-II-D-1-d-(1). A chemical control program shall be implemented to eliminate undesired pests and pathogens in accordance with applicable state and Federal laws.

Appendix P-II-D-1-d-(2). Arthropods and other motile macroorganisms used in conjunction with experiments requiring BL4-P level physical containment shall be housed in appropriate cages. When appropriate to the organism, experiments shall be conducted within cages designed to contain the motile organisms.

Appendix P-II-D-1-e. Concurrent Experiments Conducted in the Greenhouse (BL4-P)

Appendix P-II-D-1-e-(1). Experiments involving organisms that require a containment level lower than BL4-P may be conducted in the greenhouse concurrently with experiments that require BL4-P containment provided that all work is conducted in accordance with BL4-P greenhouse practices. When the experimental microorganisms in use require a containment level lower than BL4-P, greenhouse practices reflect the level of containment required by the highest containment level microorganisms being tested.

Appendix P-II-D-1-f. Signs (BL4-P)

Appendix P-II-D-1-f-(1). A sign shall be posted indicating that a restricted experiment is in progress. The sign shall indicate the following: (i) the name of the responsible individual, (ii) the plants in use, and (iii) any special requirements for using the area.

Appendix P-II-D-1-f-(2). If organisms are used that have a recognized potential for causing serious detrimental impacts on managed or natural ecosystems, their presence shall be indicated by a sign posted on the greenhouse access doors.

Appendix P-II-D-1-f-(3). If there is a risk to human health, a sign shall be posted incorporating the universal biosafety symbol.

Appendix P-II-D-1-g. Transfer of Materials (BL4-P)

Appendix P-II-D-1-g-(1). Experimental materials that are brought into or removed from the greenhouse in a viable or intact state shall be transferred to a non-breakable, sealed, primary container then enclosed in a non-breakable, sealed secondary container. These containers shall be removed from the greenhouse facility through a chemical disinfectant, fumigation chamber, or an airlock designed for this purpose.

Appendix P-II-D-1-g-(2). Supplies and materials shall be brought into the greenhouse facility through a double-door autoclave, fumigation chamber, or airlock that is appropriately decontaminated between each use. After securing the outer doors, personnel within the greenhouse facility shall retrieve the materials by opening the interior door of the autoclave, fumigation chamber, or airlock. These doors shall be secured after the materials are brought into the greenhouse facility.

Appendix P-II-D-1-h. Greenhouse Practices Manual (BL4-P)

Appendix P-II-D-1-h-(1). A greenhouse practices manual shall be prepared or adopted. This manual shall include contingency plans to be implemented in the event of the unintentional release of experimental organisms.

Appendix P-II-D-1-i. Protective Clothing (BL4-P)

Appendix P-II-D-1-i-(1). Street clothing shall be removed in the outer clothing change room. Complete laboratory clothing (may be disposable) including undergarments, pants, and shirts, jump suits, shoes, and hats shall be provided and worn by all personnel entering the greenhouse facility.

Appendix P-II-D-1-i-(2). Personnel shall remove laboratory clothing when exiting the greenhouse facility and before entering the shower area. This clothing shall be stored in a locker or hamper in the inner change room.

Appendix P-II-D-1-i-(3). All laboratory clothing shall be autoclaved before laundering.

Appendix P-II-D-2. Facilities (BL4-P)

Appendix P-II-D-2-a. Greenhouse Design (BL4-P)

Appendix P-II-D-2-a-(1). The maximum containment greenhouse facility shall consist of a separate building or a clearly demarcated and isolated area within a building. The need to maintain negative pressure should be considered when constructing or renovating the greenhouse facility.

Appendix P-II-D-2-a-(2). Outer and inner change rooms, separated by a shower, shall be provided for personnel entering and exiting the greenhouse facility.

Appendix P-II-D-2-a-(3). Windows shall be closed and sealed. All glazing shall be resistant to breakage (e.g., double-pane tempered glass or equivalent).

Appendix P-II-D-2-a-(4). Access doors to the greenhouse shall be self-closing and locking.

Appendix P-II-D-2-a-(5). The greenhouse facility shall be surrounded by a security fence or protected by equivalent security measures.

Appendix P-II-D-2-a-(6). The walls, floors, and ceilings of the greenhouse shall be constructed to form a sealed internal shell that facilitates fumigation and is animal and arthropod-proof. These internal surfaces shall be resistant to penetration and degradation by liquids and chemicals to facilitate cleaning and decontamination of the area. All penetrations into these structures and surfaces (e.g., plumbing and utilities) shall be sealed.

Appendix P-II-D-2-a-(7). Bench tops and other work surfaces shall have seamless surfaces impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat.

Appendix P-II-D-2-a-(8). A double-door autoclave, fumigation chamber, or ventilated airlock shall be provided for passage of all materials, supplies, or equipment that are not brought into the greenhouse facility through the change room.

Appendix P-II-D-2-b. Autoclaves (BL4-P)

Appendix P-II-D-2-b-(1). A double-door autoclave shall be provided for the decontamination of materials removed from the greenhouse facility. The autoclave door, which opens to the area external to the greenhouse facility, shall be sealed to the outer wall and automatically controlled so that it can only be opened upon completion of the sterilization cycle.

Appendix P-II-D-2-c. Supply and Exhaust Air Ventilation Systems (BL4-P)

Appendix P-II-D-2-c-(1). An individual supply and exhaust air ventilation system shall be provided. The system shall maintain pressure differentials and directional airflow as required to assure inward (or zero) airflow from areas outside of the greenhouse. Differential pressure transducers shall be used to sense pressure levels. If a system malfunctions, the transducers shall sound an alarm. A backup source of power should be considered. The supply and exhaust airflow shall be interlocked to assure inward (or zero) airflow at all times. The integrity of the greenhouse shall have an air leak rate (decay rate) not to exceed 7 percent per minute (logarithm of pressure against time) over a 20-minute period at 2 inches of water gauge pressure. Nominally, this is 0.05 inches of water gauge pressure loss in 1 minute at 2 inches water gauge pressure.

Appendix P-II-D-2-c-(2). Exhaust air from the greenhouse facility shall be filtered through high efficiency particulate air/HEPA filters and discharged to the outside and dispersed away from occupied buildings and air intakes. Filter chambers shall be designed to allow *in situ* decontamination before filters are removed and to facilitate certification testing after they are replaced. HEPA filters shall be provided to treat air supplied to the greenhouse facility. HEPA filters shall be certified annually.

Appendix P-II-D-2-d. Other (BL4-P)

Appendix P-II-D-2-d-(1). Sewer vents and other ventilation lines contain high efficiency particulate air/HEPA filters. HEPA filters shall be certified annually.

Appendix P-II-D-2-d-(2). A pass-through dunk tank, fumigation chamber, or an equivalent method of decontamination shall be provided to ensure decontamination of materials and equipment that cannot be decontaminated in the autoclave.

Appendix P-II-D-2-d-(3). Liquid effluent from sinks, floors, and autoclave chambers shall be decontaminated by heat or chemical treatment before being released from the maximum containment greenhouse facility. Liquid wastes from shower rooms and toilets may be decontaminated by heat or chemical treatment. Autoclave and chemical decontamination of liquid wastes shall be evaluated by appropriate standard procedures for autoclaved wastes. Decontamination shall be evaluated mechanically and biologically using a recording thermometer and an indicator microorganism with a defined heat susceptibility pattern. If liquid wastes are decontaminated with chemical disinfectants, the chemicals used must have demonstrated efficacy against the target or indicator microorganisms.

Appendix P-II-D-2-d-(4). If there is a central vacuum system, it shall not serve areas outside of the greenhouse facility. In-line high efficiency particulate air/HEPA filters shall be placed as near as practicable to each use point or vacuum service cock. Other liquid and gas services to the greenhouse facility shall be protected by devices that prevent back-flow. HEPA filters shall be certified annually.

Appendix P-III. Biological Containment Practices

Appropriate selection of the following biological containment practices may be used to meet the containment requirements for a given organism. The present list is not exhaustive; there may be other ways of preventing effective dissemination that could possibly lead to the establishment of the organism or its genetic material in the environment resulting in deleterious consequences to managed or natural ecosystems.

Appendix P-III-A. Biological Containment Practices (Plants)

Appendix P-III-A-1. Effective dissemination of plants by pollen or seed can be prevented by one or more of the following procedures: (i) cover the reproductive structures to prevent pollen dissemination at flowering and seed dissemination at maturity; (ii) remove reproductive structures by employing male sterile strains, or harvest the

plant material prior to the reproductive stage; (iii) ensure that experimental plants flower at a time of year when cross-fertile plants are not flowering within the normal pollen dispersal range of the experimental plant; or (iv) ensure that cross-fertile plants are not growing within the known pollen dispersal range of the experimental plant.

Appendix P-III-B. Biological Containment Practices (Microorganisms)

Appendix P-III-B-1. Effective dissemination of microorganisms beyond the confines of the greenhouse can be prevented by one or more of the following procedures: (i) confine all operations to injections of microorganisms or other biological procedures (including genetic manipulation) that limit replication or reproduction of viruses and microorganisms or sequences derived from microorganisms, and confine these injections to internal plant parts or adherent plant surfaces; (ii) ensure that organisms, which can serve as hosts or promote the transmission of the virus or microorganism, are not present within the farthest distance that the airborne virus or microorganism may be expected to be effectively disseminated; (iii) conduct experiments at a time of year when plants that can serve as hosts are either not growing or are not susceptible to productive infection; (iv) use viruses and other microorganisms or their genomes that have known arthropod or animal vectors, in the absence of such vectors; (v) use microorganisms that have an obligate association with the plant; or (vi) use microorganisms that are genetically disabled to minimize survival outside of the research facility and whose natural mode of transmission requires injury of the target organism, or assures that inadvertent release is unlikely to initiate productive infection of organisms outside of the experimental facility.

Appendix P-III-C. Biological Containment Practices (Macroorganisms)

Appendix P-III-C-1. Effective dissemination of arthropods and other small animals can be prevented by using one or more of the following procedures: (i) use non-flying, flight-impaired, or sterile arthropods; (ii) use non-motile or sterile strains of small animals; (iii) conduct experiments at a time of year that precludes the survival of escaping organisms; (iv) use animals that have an obligate association with a plant that is not present within the dispersal range of the organism; or (v) prevent the escape of organisms present in run-off water by chemical treatment or evaporation of run-off water.

****

APPENDIX Q. PHYSICAL AND BIOLOGICAL CONTAINMENT FOR RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULE RESEARCH INVOLVING ANIMALS

Appendix Q specifies containment and confinement practices for research involving whole animals, both those in which the animal's genome has been altered by stable introduction of recombinant or synthetic nucleic acid molecules, or DNA derived therefrom, into the germ-line (transgenic animals) and experiments involving viable recombinant or synthetic nucleic acid molecule-modified microorganisms tested on whole animals. The appendix applies to animal research activities with the following modifications:

Appendix Q shall supersede Appendix G (*Physical Containment*) when research animals are of a size or have growth requirements that preclude the use of containment for laboratory animals. Some animals may require other types of containment (see Appendix Q-III-D, *Footnotes and References for Appendix Q*). The animals covered in Appendix Q are those species normally categorized as animals including but not limited to cattle, swine, sheep, goats, horses, and poultry.

The Institutional Biosafety Committee shall include at least one scientist with expertise in animal containment principles when experiments utilizing Appendix Q require Institutional Biosafety Committee prior approval.

The institution shall establish and maintain a health surveillance program for personnel engaged in animal research involving viable recombinant or synthetic nucleic acid molecule-containing microorganisms that require Biosafety Level (BL) 3 or greater containment in the laboratory.

Appendix Q-I. General Considerations

Appendix Q-I-A. Containment Levels

The containment levels required for research involving recombinant or synthetic nucleic acid molecules associated with or in animals is based on classification of experiments in Section III, *Experiments Covered by the NIH Guidelines*. For the purpose of animal research, four levels of containment are established. These are referred to as BL1-Animals (N), BL2-N, BL3-N, and BL4-N and are described in the following appendices of Appendix Q. The descriptions include: (i) standard practices for physical and biological containment, and (ii) animal facilities.

Appendix Q-I-B. Disposal of Animals (BL1-N through BL4-N)

Appendix Q-I-B-1. When an animal covered by Appendix Q containing recombinant or synthetic nucleic acid molecules or a recombinant or synthetic nucleic acid molecule-derived organism is euthanized or dies, the carcass shall be disposed of to avoid its use as food for human beings or animals unless food use is specifically authorized by an appropriate Federal agency.

Appendix Q-I-B-2. A permanent record shall be maintained of the experimental use and disposal of each animal or group of animals.

Appendix Q-II. Physical and Biological Containment Levels

Appendix Q-II-A. Biosafety Level 1 - Animals (BL1-N)

Appendix Q-II-A-1. Standard Practices (BL1-N)

Appendix Q-II-A-1-a. Animal Facility Access (BL1-N)

Appendix Q-II-A-1-a-(1). The containment area shall be locked.

Appendix Q-II-A-1-a-(2). Access to the containment area shall be limited or restricted when experimental animals are being held.

Appendix Q-II-A-1-a-(3). The containment area shall be patrolled or monitored at frequent intervals.

Appendix Q-II-A-1-b. Other (BL1-N)

Appendix Q-II-A-1-b-(1). All genetically engineered neonates shall be permanently marked within 72 hours after birth, if their size permits. If their size does not permit marking, their containers should be marked. In addition, transgenic animals should contain distinct and biochemically assayable DNA sequences that allow identification of transgenic animals from among non-transgenic animals.

Appendix Q-II-A-1-b-(2) A double barrier shall be provided to separate male and female animals unless reproductive studies are part of the experiment or other measures are taken to avoid reproductive transmission. Reproductive incapacitation may be used.

Appendix Q-II-A-1-b-(3). The containment area shall be in accordance with state and Federal laws and animal care requirements.

Appendix Q-II-A-2. Animal Facilities (BL1-N)

Appendix Q-II-A-2-a. Animals shall be confined to securely fenced areas or be in enclosed structures (animal rooms) to minimize the possibility of theft or unintentional release.

Appendix Q-II-B. Biosafety Level 2 - Animals (BL2-N) (See Appendix Q-III-A, *Footnotes and References for Appendix Q*)

Appendix Q-II-B-1. Standard Practices (BL2-N)

Appendix Q-II-B-1-a. Animal Facility Access (BL2-N)

Appendix Q-II-B-1-a-(1). The containment area shall be locked.

Appendix Q-II-B-1-a-(2). The containment area shall be patrolled or monitored at frequent intervals.

Appendix Q-II-B-1-a-(3). The containment building shall be controlled and have a locking access.

Appendix Q-II-B-1-a-(4). The Animal Facility Director shall establish policies and procedures whereby only persons who have been advised of the potential hazard and who meet any specific entry requirements (e.g., vaccination) may enter the laboratory or animal rooms.

Appendix Q-II-B-1-a-(5). Animals of the same or different species, which are not involved in the work being performed, shall not be permitted in the animal area.

Appendix Q-II-B-1-b. Decontamination and Inactivation (BL2-N)

Appendix Q-II-B-1-b-(1). Contaminated materials that are decontaminated at a site away from the laboratory shall be placed in a closed durable leak-proof container prior to removal from the laboratory.

Appendix Q-II-B-1-b-(2). Needles and syringes shall be promptly placed in a puncture-resistant container and decontaminated, preferably by autoclaving, before discard or reuse.

Appendix Q-II-B-1-c. Signs (BL2-N)

Appendix Q-II-B-1-c-(1). When the animal research requires special provisions for entry (e.g., vaccination), a warning sign incorporating the universal biosafety symbol shall be posted on all access doors to the animal work area. The sign shall indicate: (i) the agent, (ii) the animal species, (iii) the name and telephone number of the Animal Facility Director or other responsible individual, and (iv) any special requirements for entering the laboratory.

Appendix Q-II-B-1-d. Protective Clothing (BL2-N)

Appendix Q-II-B-1-d-(1). Laboratory coats, gowns, smocks, or uniforms shall be worn while in the animal area or attached laboratory. Before entering non-laboratory areas (e.g., cafeteria, library, administrative offices), protective clothing shall be removed and kept in the work entrance area.

Appendix Q-II-B-1-d-(2). Special care shall be taken to avoid skin contamination with microorganisms containing recombinant or synthetic nucleic acid molecules. Impervious and/or protective gloves shall be worn when handling experimental animals and when skin contact with an infectious agent is unavoidable.

Appendix Q-II-B-1-e. Records (BL2-N)

Appendix Q-II-B-1-e-(1). Any incident involving spills and accidents that result in environmental release or exposures of animals or laboratory workers to organisms containing recombinant or synthetic nucleic acid molecules shall be reported immediately to the Animal Facility Director, Institutional Biosafety Committee, NIH OSP, and other appropriate authorities (if applicable). Reports to the NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Medical evaluation, surveillance, and treatment shall be provided as appropriate and written records maintained. If necessary, the area shall be appropriately decontaminated.

Appendix Q-II-B-1-e-(2). When appropriate and giving consideration to the agent handled, baseline serum samples shall be collected and stored for animal care and other at-risk personnel. Additional serum specimens may be collected periodically depending on the agent handled and the function of the animal facility.

Appendix Q-II-B-1-f. Transfer of Materials (BL2-N)

Appendix Q-II-B-1-f-(1). Biological materials removed from the animal containment area in a viable or intact state shall be transferred to a non-breakable sealed primary container and then enclosed in a non-breakable sealed secondary container. All containers, primary and secondary, shall be disinfected before removal from the animal facility. Advance approval for transfer of material shall be obtained from the Animal Facility Director. Packages containing viable agents may only be opened in a facility having an equivalent or higher level of physical containment unless the agent is biologically inactivated or incapable of reproduction.

Appendix Q-II-B-1-g. Other (BL2-N)

Appendix Q-II-B-1-g-(1). All genetically engineered neonates shall be permanently marked within 72 hours after birth, if their size permits. If their size does not permit marking, their containers should be marked. In addition, transgenic animals should contain distinct and biochemically assayable DNA sequences that allow identification of transgenic animals from among non-transgenic animals.

Appendix Q-II-B-1-g-(2). Needles and syringes shall be used only for parenteral injection and aspiration of fluids from laboratory animals and diaphragm bottles. Only needle-locking syringes or disposable syringe-needle units (i.e., needle is integral to the syringe) shall be used for the injection or aspiration of fluids containing organisms that contain recombinant or synthetic nucleic acid molecules. Extreme caution shall be used when handling needles and syringes to avoid autoinoculation and the generation of aerosols during use and disposal. Following use, needles shall not be bent, sheared, replaced in the needle sheath or guard, or removed from the syringe. Needles and syringes shall be promptly placed in a puncture-resistant container and decontaminated, preferably by autoclaving, before discard or reuse.

Appendix Q-II-B-1-g-(3). Appropriate steps should be taken to prevent horizontal transmission or exposure of laboratory personnel. If the agent used as a vector is known to be transmitted by a particular route (e.g., arthropods), special attention should be given to preventing spread by that route. In the absence of specific knowledge of a particular route of transmission, all potential means of horizontal transmission (e.g., arthropods, contaminated bedding, or animal waste, etc.) should be prevented.

Appendix Q-II-B-1-g-(4). Eating, drinking, smoking, and applying cosmetics shall not be permitted in the work area.

Appendix Q-II-B-1-g-(5). Individuals who handle materials and animals containing recombinant or synthetic nucleic acid molecules shall be required to wash their hands before exiting the containment area.

Appendix Q-II-B-1-g-(6). A double barrier shall be provided to separate male and female animals unless reproductive studies are part of the experiment or other measures are taken to avoid reproductive transmission. Reproductive incapacitation may be used.

Appendix Q-II-B-1-g-(7). The containment area shall be in accordance with state and Federal laws and animal care requirements.

Appendix Q-II-B-1-g-(8). A biosafety manual shall be prepared or adopted. Personnel shall be advised of special hazards and required to read and follow instructions on practices and procedures.

Appendix Q-II-B-2. Animal Facilities (BL2-N)

Appendix Q-II-B-2-a. Animals shall be contained within an enclosed structure (animal room or equivalent) to minimize the possibility of theft or unintentional release and to avoid arthropod access. The special provision to avoid the entry or escape of arthropods from the animal areas may be waived if the agent in use is not known to be transmitted by arthropods.

Appendix Q-II-B-2-b. Surfaces shall be impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat.

Appendix Q-II-B-2-c. The animal containment area shall be designed so that it can be easily cleaned.

Appendix Q-II-B-2-d. Windows that open shall be fitted with fly screens.

Appendix Q-II-B-2-e. An autoclave shall be available for decontamination of laboratory wastes.

Appendix Q-II-B-2-f. If arthropods are used in the experiment or the agent under study can be transmitted by an arthropod, interior work areas shall be appropriately screened (52 mesh). All perimeter joints and openings shall be sealed and additional arthropod control mechanisms used to minimize arthropod entry and propagation, including appropriate screening of access doors or the equivalent.

Appendix Q-II-C. Biosafety Level 3 - Animals (BL3-N) (See Appendix Q-III-B, *Footnotes and References for Appendix Q*)

Appendix Q-II-C-1. Standard Practices (BL3-N)

Appendix Q-II-C-1-a. Animal Facility Access (BL3-N)

Appendix Q-II-C-1-a-(1). The containment area shall be locked.

Appendix Q-II-C-1-a-(2). The containment area shall be patrolled or monitored at frequent intervals.

Appendix Q-II-C-1-a-(3). The containment building shall be controlled and have a locking access.

Appendix Q-II-C-1-a-(4). The Animal Facility Director shall establish policies and procedures whereby only persons who have been advised of the potential hazard and who meet any specific entry requirements (e.g., vaccination) shall enter the laboratory or animal rooms.

Appendix Q-II-C-1-a-(5). Animal room doors, gates, or other closures shall be kept closed when experiments are in progress.

Appendix Q-II-C-1-b. Decontamination and Inactivation (BL3-N)

Appendix Q-II-C-1-b-(1). The work surfaces of containment equipment shall be decontaminated when work with organisms containing recombinant or synthetic nucleic acid molecules is finished. Where feasible, plastic-backed paper toweling shall be used on nonporous work surfaces to facilitate clean-up.

Appendix Q-II-C-1-b-(2). All animals shall be euthanized at the end of their experimental usefulness and the carcasses decontaminated before disposal in an approved manner.

Appendix Q-II-C-1-b-(3). Needles and syringes shall be promptly placed in a puncture-resistant container and decontaminated, preferably by autoclaving, before discard or reuse.

Appendix Q-II-C-1-b-(4). Special safety testing, decontamination procedures, and Institutional Biosafety Committee approval shall be required to transfer agents or tissue/organ specimens from a BL3-N animal facility to a facility with a lower containment classification.

Appendix Q-II-C-1-b-(5). Liquid effluent from containment equipment, sinks, biological safety cabinets, animal rooms, primary barriers, floor drains, and sterilizers shall be decontaminated by heat treatment before being released into the sanitary system. The procedure used for heat decontamination of liquid wastes shall be monitored with a recording thermometer. The effectiveness of the heat decontamination process system shall be revalidated at minimum on a yearly basis with an indicator organism. More frequent validation, based on the amount of use or other safety factors, shall be left to the discretion of the IBC.

Appendix Q-II-C-1-c. Signs (BL3-N)

Appendix Q-II-C-1-c-(1). When the animal research requires special provisions for entry (e.g., vaccination), a warning sign incorporating the universal biosafety symbol shall be posted on all access doors to the animal work area. The sign shall indicate: (i) the agent, (ii) the animal species, (iii) the name and telephone number of the Animal Facility Director or other responsible individual, and (iv) any special requirements for entering the laboratory.

Appendix Q-II-C-1-d. Protective Clothing (BL3-N)

Appendix Q-II-C-1-d-(1). Full protective clothing that protects the individual (e.g., scrub suits, coveralls, uniforms) shall be worn in the animal area. Clothing shall not be worn outside the animal containment area and shall be decontaminated before laundering or disposal. Personnel shall be required to shower before exiting the BL3-N area and wearing of personal clothing.

Appendix Q-II-C-1-d-(2). Special care shall be taken to avoid skin contamination with microorganisms containing recombinant or synthetic nucleic acid molecules. Impervious and/or protective gloves shall be worn when handling experimental animals and when skin contact with an infectious agent is unavoidable.

Appendix Q-II-C-1-d-(3). Appropriate respiratory protection shall be worn in rooms containing experimental animals.

Appendix Q-II-C-1-e. Records (BL3-N)

Appendix Q-II-C-1-e-(1). Documents regarding experimental animal use and disposal shall be maintained in a permanent record book.

Appendix Q-II-C-1-e-(2). Any incident involving spills and accidents that result in environmental release or exposure of animals or laboratory workers to organisms containing recombinant or synthetic nucleic acid molecules shall be reported immediately to the Biological Safety Office, Animal Facility Director, Institutional Biosafety Committee, NIH OSP, and other appropriate authorities (if applicable). Reports to the NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Medical evaluation, surveillance, and treatment shall be provided as appropriate and written records maintained. If necessary, the area shall be appropriately decontaminated.

Appendix Q-II-C-1-e-(3). When appropriate and giving consideration to the agent handled, baseline serum samples shall be collected and stored for animal care and other at-risk personnel. Additional serum specimens may be collected periodically depending on the agent handled or the function of the facility.

Appendix Q-II-C-1-f. Transfer of Materials (BL3-N)

Appendix Q-II-C-1-f-(1). Biological materials removed from the animal containment laboratory in a viable or intact state shall be transferred to a non-breakable sealed primary container and then enclosed in a non-breakable sealed secondary container. All containers, primary and secondary, shall be disinfected before removal from the animal facility. Advance approval for transfer of material shall be obtained from the Animal Facility Director. Packages containing viable agents may be opened only in a facility having an equivalent or higher level of physical containment unless the agent is biologically inactivated or incapable of reproduction.

Appendix Q-II-C-1-f-(2). Special safety testing, decontamination procedures, and Institutional Biosafety Committee approval shall be required to transfer agents or tissue/organ specimens from a BL3-N animal facility to a facility with a lower containment classification.

Appendix Q-II-C-1-g. Other (BL3-N)

Appendix Q-II-C-1-g-(1). All genetically engineered neonates shall be permanently marked within 72 hours after birth, if their size permits. If their size does not permit marking, their containers should be marked. In addition, transgenic animals should contain distinct and biochemically assayable DNA sequences that allow identification of transgenic animals from among non-transgenic animals.

Appendix Q-II-C-1-g-(2). Appropriate steps should be taken to prevent horizontal transmission or exposure of laboratory personnel. If the agent used as the vector is known to be transmitted by a particular route (e.g., arthropods), special attention should be given to preventing spread by that route. In the absence of specific knowledge of a particular route of transmission, all potential means of horizontal transmission (e.g., arthropods, contaminated bedding, or animal waste) should be prevented.

Appendix Q-II-C-1-g-(3). Eating, drinking, smoking, and applying cosmetics shall not be permitted in the work area.

Appendix Q-II-C-1-g-(4). Individuals who handle materials and animals containing recombinant or synthetic nucleic acid molecules shall be required to wash their hands before exiting the containment area.

Appendix Q-II-C-1-g-(5). Experiments involving other organisms that require containment levels lower than BL3-N may be conducted in the same area concurrently with experiments requiring BL3-N containment provided that they are conducted in accordance with BL3-N practices.

Appendix Q-II-C-1-g-(6). Animal holding areas shall be cleaned at least once a day and decontaminated immediately following any spill of viable materials.

Appendix Q-II-C-1-g-(7). All procedures shall be performed carefully to minimize the creation of aerosols.

Appendix Q-II-C-1-g-(8). A double barrier shall be provided to separate male and female animals unless reproductive studies are part of the experiment or other measures are taken to avoid reproductive transmission. Reproductive incapacitation may be used.

Appendix Q-II-C-1-g-(9). The containment area shall be in accordance with state and Federal laws and animal care requirements.

Appendix Q-II-C-1-g-(10). All animals shall be euthanized at the end of their experimental usefulness and the carcasses decontaminated before disposal in an approved manner.

Appendix Q-II-C-1-g-(11). Personnel shall be required to shower before exiting the BL3-N area and wearing personal clothing.

Appendix Q-II-C-1-g-(12). Animals of the same or different species, which are not involved in the work being performed, shall not be permitted in the animal area.

Appendix Q-II-C-1-g-(13). Needles and syringes shall be used only for parenteral injection and aspiration of fluids from laboratory animals and diaphragm bottles. Only needle-locking syringes or disposable syringe-needle units (i.e., needle is integral to the syringe) shall be used for the injection or aspiration of fluids containing organisms that contain recombinant or synthetic nucleic acid molecules. Extreme caution shall be used when handling needles and syringes to avoid autoinoculation and the generation of aerosols during use and disposal. Following use, needles shall not be bent, sheared, replaced in the needle sheath or guard or removed from the syringe. The needles and syringes shall be promptly placed in a puncture-resistant container and decontaminated, preferably by autoclaving, before discard or reuse.

Appendix Q-II-C-1-g-(14). A biosafety manual shall be prepared or adopted. Personnel shall be advised of special hazards and required to read and follow instructions on practices and procedures.

Appendix Q-II-C-2. Animal Facilities (BL3-N)

Appendix Q-II-C-2-a. Animals shall be contained within an enclosed structure (animal room or equivalent) to minimize the possibility of theft or unintentional release and avoid arthropod access. The special provision to avoid the entry or escape of arthropods from the animal areas may be waived if the agent in use is not known to be transmitted by arthropods.

Appendix Q-II-C-2-b. The interior walls, floors, and ceilings shall be impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat, to facilitate cleaning. Penetrations in these structures and surfaces (e.g., plumbing and utilities) shall be sealed.

Appendix Q-II-C-2-c. Windows in the animal facility shall be closed, sealed, and breakage resistant (e.g., double-pane tempered glass or equivalent). The need to maintain negative pressure should be considered when constructing or renovating the animal facility.

Appendix Q-II-C-2-d. An autoclave, incinerator, or other effective means to decontaminate animals and waste shall be available, preferably within the containment area. If feasible, a double-door autoclave is preferred and should be positioned to allow removal of material from the containment area.

Appendix Q-II-C-2-e. If arthropods are used in the experiment or the agent under study can be transmitted by an arthropod, the interior work area shall be appropriately screened (52 mesh). All perimeter joints and openings shall be sealed, and additional arthropod control mechanisms used to minimize arthropod entry and propagation, including appropriate screening, or the equivalent of access doors.

Appendix Q-II-C-2-f. Access doors to the containment area shall be self-closing.

Appendix Q-II-C-2-g. The animal area shall be separated from all other areas. Passage through two sets of doors shall be the basic requirement for entry into the animal area from access corridors or other contiguous areas. The animal containment area shall be physically separated from access corridors and other laboratories or areas by a double-door clothes change room, equipped with integral showers and airlock.

Appendix Q-II-C-2-h. Liquid effluent from containment equipment, sinks, biological safety cabinets, animal rooms, primary barriers, floor drains, and sterilizers shall be decontaminated by heat treatment before being released into the sanitary system. The procedure used for heat decontamination of liquid wastes shall be monitored with a recording thermometer. The effectiveness of the heat decontamination process system shall be revalidated at minimum on a yearly basis with an indicator organism. More frequent validation, based on the amount of use or other safety factors, shall be left to the discretion of the IBC.

Appendix Q-II-C-2-i. An exhaust air ventilation system shall be provided. This system shall create directional airflow that draws air into the animal room through the entry area. The building exhaust, or the exhaust from primary containment units, may be used for this purpose if the exhaust air is discharged to the outside and shall be dispersed away from occupied areas and air intakes. Personnel shall verify that the direction of the airflow (into the animal room) is proper.

Appendix Q-II-C-2-j. If the agent is transmitted by aerosol, then the exhaust air shall pass through a high efficiency particulate air/HEPA filter.

Appendix Q-II-C-2-k. Vacuum lines shall be protected with high efficiency particulate air/HEPA filters and liquid disinfectant traps.

Appendix Q-II-C-2-l. In lieu of open housing in the special animal room, animals held in a BL3-N area may be housed in partial-containment caging systems (e.g., Horsfall units or gnotobiotic systems, or other special containment primary barriers). Prudent judgment must be exercised to implement this ventilation system (e.g., animal species) and its discharge location.

Appendix Q-II-C-2-m. Each animal area shall contain a foot, elbow, or automatically operated sink for hand washing. The sink shall be located near the exit door.

Appendix Q-II-C-2-n. Restraining devices for animals may be required to avoid damage to the integrity of the animal containment facility.

Appendix Q-II-D. Biosafety Level 4 - Animals (BL4-N) (See Appendix Q-III-C, *Footnotes and References for Appendix Q*)

Appendix Q-II-D-1. Standard Practices (BL4-N)

Appendix Q-II-D-1-a. Animal Facility Access (BL4-N)

Appendix Q-II-D-1-a-(1). Individuals under 16 years of age shall not be permitted to enter the animal area.

Appendix Q-II-D-1-a-(2). The containment area shall be locked.

Appendix Q-II-D-1-a-(3). The containment area shall be patrolled or monitored at frequent intervals.

Appendix Q-II-D-1-a-(4). The containment building shall be controlled and have a locking access.

Appendix Q-II-D-1-a-(5). The Animal Facility Director shall establish policies and procedures whereby only persons who have been advised of the potential hazard and who meet any specific entry requirements (e.g., vaccination) may enter the laboratory or animal room.

Appendix Q-II-D-1-a-(6). Individuals shall enter and exit the animal facility only through the clothing change and shower rooms.

Appendix Q-II-D-1-a-(7). Personnel shall use the airlocks to enter or exit the laboratory only in an emergency.

Appendix Q-II-D-1-a-(8). Animal room doors, gates, and other closures shall be kept closed when experiments are in progress.

Appendix Q-II-D-1-b. Decontamination and Inactivation (BL4-N)

Appendix Q-II-D-1-b-(1). All contaminated liquid or solid wastes shall be decontaminated before disposal.

Appendix Q-II-D-1-b-(2). The work surfaces and containment equipment shall be decontaminated when work with organisms containing recombinant or synthetic nucleic acid molecules is finished. Where feasible, plastic-backed paper toweling shall be used on nonporous work surfaces to facilitate clean-up.

Appendix Q-II-D-1-b-(3). All wastes from animal rooms and laboratories shall be appropriately decontaminated before disposal in an approved manner.

Appendix Q-II-D-1-b-(4). No materials, except for biological materials that are to remain in a viable or intact state, shall be removed from the maximum containment laboratory unless they have been autoclaved or decontaminated. Equipment or material that might be damaged by high temperatures or steam shall be decontaminated by gaseous or vapor methods in an airlock or chamber designed for this purpose.

Appendix Q-II-D-1-b-(5). When ventilated suits are required, the animal personnel shower entrance/exit area shall be equipped with a chemical disinfectant shower to decontaminate the surface of the suit before exiting the area. A neutralization or water dilution device shall be integral with the chemical disinfectant discharge piping before entering the heat sterilization system. Entry to this area shall be through an airlock fitted with airtight doors.

Appendix Q-II-D-1-b-(6). Needles and syringes shall be promptly placed in a puncture-resistant container and decontaminated, preferably by autoclaving, before discard or reuse.

Appendix Q-II-D-1-b-(7). Supplies and materials needed in the animal facility shall be brought in by way of the double-door autoclave, fumigation chamber, or airlock that shall be appropriately decontaminated between each use.

Appendix Q-II-D-1-b-(8). An autoclave, incinerator, or other effective means to decontaminate animals and wastes shall be available, preferably within the containment area. If feasible, a double-door autoclave is preferred and should be positioned to allow removal of material from the containment area.

Appendix Q-II-D-1-b-(9). Liquid effluent from containment equipment, sinks, biological safety cabinets, animal rooms, primary barriers, floor drains, and sterilizers shall be decontaminated by heat treatment before being released into the sanitary system. If required by design, regulation, local ordinance or policy, liquid wastes from shower rooms and toilets shall be decontaminated with chemical disinfectants or heat by methods demonstrated to be effective. The procedure used for heat decontamination of liquid wastes shall be monitored with a recording thermometer. The effectiveness of the heat decontamination process system shall be revalidated at minimum on a yearly basis with an indicator organism. More frequent validation, based on the amount of use or other safety factors, shall be left to the discretion of the IBC. If required by design, regulation, local ordinance or policy, liquid wastes from the shower shall be chemically decontaminated using an Environmental Protection Agency-approved germicide. The efficacy of the chemical treatment process shall be validated with an indicator organism. Chemical disinfectants shall be neutralized or diluted before release into general effluent waste systems.

Appendix Q-II-D-1-c. Signs (BL4-N)

Appendix Q-II-D-1-c-(1). When the animal research requires special provisions for entry (e.g., vaccination), a warning sign incorporating the universal biosafety symbol shall be posted on all access doors to the animal work area. The sign shall indicate: (i) the agent, (ii) the animal species, (iii) the name and telephone number of the

Animal Facility Director, or other responsible individual, and (iv) any special requirements for entering the laboratory.

Appendix Q-II-D-1-d. Protective Clothing (BL4-N)

Appendix Q-II-D-1-d-(1). Individuals shall enter and exit the animal facility only through the clothing change and shower rooms. Street clothing shall be removed and kept in the outer clothing change room. Complete laboratory clothing (may be disposable), including undergarments, pants, shirts, jump suits, and shoes shall be provided for all personnel entering the animal facility. When exiting the BL4-N area and before proceeding into the shower area, personnel shall remove their laboratory clothing in the inner change room. All laboratory clothing shall be autoclaved before laundering. Personnel shall shower each time they exit the animal facility.

Appendix Q-II-D-1-d-(2). A ventilated head-hood or a one-piece positive pressure suit, which is ventilated by a life-support system, shall be worn by all personnel entering rooms that contain experimental animals when appropriate. When ventilated suits are required, the animal personnel shower entrance/exit area shall be equipped with a chemical disinfectant shower to decontaminate the surface of the suit before exiting the area. A neutralization or water dilution device shall be integral with the chemical disinfectant discharge piping before entering the heat sterilization system. Entry to this area shall be through an airlock fitted with airtight doors.

Appendix Q-II-D-1-d-(3). Appropriate respiratory protection shall be worn in rooms containing experimental animals.

Appendix Q-II-D-1-e. Records (BL4-N)

Appendix Q-II-D-1-e-(1). Documents regarding experimental animal use and disposal shall be maintained in a permanent record book.

Appendix Q-II-D-1-e-(2). A system shall be established for: (i) reporting laboratory accidents and exposures that are a result of overt exposures to organisms containing recombinant or synthetic nucleic acid molecules, (ii) employee absenteeism, and (iii) medical surveillance of potential laboratory-associated illnesses. Permanent records shall be prepared and maintained. Any incident involving spills and accidents that results in environmental release or exposures of animals or laboratory workers to organisms containing recombinant or synthetic nucleic acid molecules shall be reported immediately to the Biological Safety Officer, Animal Facility Director, Institutional Biosafety Committee, NIH OSP, and other appropriate authorities (if applicable). Reports to the NIH OSP shall be sent to the Office of Science Policy, National Institutes of Health, preferably by e-mail to: NIHGuidelines@od.nih.gov; additional contact information is also available here and on the OSP website (www.osp.od.nih.gov). Medical evaluation, surveillance, and treatment shall be provided as appropriate and written records maintained. If necessary, the area shall be appropriately decontaminated.

Appendix Q-II-D-1-e-(3). When appropriate and giving consideration to the agents handled, baseline serum samples shall be collected and stored for animal care and other at-risk personnel. Additional serum specimens may be collected periodically depending on the agents handled or the function of the facility.

Appendix Q-II-D-1-e-(4). A permanent record book indicating the date and time of each entry and exit shall be signed by all personnel.

Appendix Q-II-D-1-f. Transfer of Materials (BL4-N)

Appendix Q-II-D-1-f-(1). No materials, except for biological materials that are to remain in a viable or intact state, shall be removed from the maximum containment laboratory unless they have been autoclaved or decontaminated. Equipment or material that might be damaged by high temperatures or steam shall be decontaminated by gaseous or vapor methods in an airlock or chamber designed for this purpose.

Appendix Q-II-D-1-f-(2). Biological materials removed from the animal maximum containment laboratory in a viable or intact state shall be transferred to a non-breakable sealed primary container and then enclosed in a non-breakable sealed secondary container that shall be removed from the animal facility through a disinfectant dunk tank, fumigation chamber, or an airlock designed for this purpose. Advance approval for transfer of material shall be obtained from the Animal Facility Director. Such packages containing viable agents can only be opened in another BL4-N animal facility if the agent is biologically inactivated or incapable of reproduction. Special safety testing, decontamination procedures, and Institutional Biosafety Committee approval shall be

required to transfer agents or tissue/organ specimens from a BL4-N animal facility to one with a lower containment classification.

Appendix Q-II-D-1-f-(3). Supplies and materials needed in the animal facility shall be brought in by way of the double-door autoclave, fumigation chamber, or airlock that shall be appropriately decontaminated between each use. After securing the outer doors, personnel within the animal facility retrieve the materials by opening the interior doors of the autoclave, fumigation chamber, or airlock. These doors shall be secured after materials are brought into the animal facility.

Appendix Q-II-D-1-g. Other (BL4-N)

Appendix Q-II-D-1-g-(1). All genetically engineered neonates shall be permanently marked within 72 hours after birth, if their size permits. If their size does not permit marking, their containers should be marked. In addition, transgenic animals should contain distinct and biochemically assayable DNA sequences that allow identification of transgenic animals from among non-transgenic animals.

Appendix Q-II-D-1-g-(2). Eating, drinking, smoking, and applying cosmetics shall not be permitted in the work area.

Appendix Q-II-D-1-g-(3). Individuals who handle materials and animals containing recombinant or synthetic nucleic acid molecules shall be required to wash their hands before exiting the containment area.

Appendix Q-II-D-1-g-(4). Experiments involving other organisms that require containment levels lower than BL4-N may be conducted in the same area concurrently with experiments requiring BL4-N containment provided that they are conducted in accordance with BL4-N practices.

Appendix Q-II-D-1-g-(5). Animal holding areas shall be cleaned at least once a day and decontaminated immediately following any spill of viable materials.

Appendix Q-II-D-1-g-(6). All procedures shall be performed carefully to minimize the creation of aerosols.

Appendix Q-II-D-1-g-(7). A double barrier shall be provided to separate male and female animals. Animal isolation barriers shall be sturdy and accessible for cleaning. Reproductive incapacitation may be used.

Appendix Q-II-D-1-g-(8). The containment area shall be in accordance with state and Federal laws and animal care requirements.

Appendix Q-II-D-1-g-(9). The life support system for the ventilated suit or head hood is equipped with alarms and emergency back-up air tanks. The exhaust air from the suit area shall be filtered by two sets of high efficiency particulate air/HEPA filters installed in series or incinerated. A duplicate filtration unit, exhaust fan, and an automatically starting emergency power source shall be provided. The air pressure within the suit shall be greater than that of any adjacent area. Emergency lighting and communication systems shall be provided. A double-door autoclave shall be provided for decontamination of waste materials to be removed from the suit area.

Appendix Q-II-D-1-g-(10). Needles and syringes shall be used only for parenteral injection and aspiration of fluids from laboratory animals and diaphragm bottles. Only needle-locking syringes or disposable syringe-needle units (i.e., needle is integral to the syringe) shall be used for the injection or aspiration of fluids containing organisms that contain recombinant or synthetic nucleic acid molecules. Extreme caution shall be used when handling needles and syringes to avoid autoinoculation and the generation of aerosols during use and disposal. Following use, needles shall not be bent, sheared, replaced in the needle sheath or guard, or removed from the syringe. The needles and syringes shall be promptly placed in a puncture-resistant container and decontaminated, preferably by autoclaving, before discard or reuse.

Appendix Q-II-D-1-g-(11). An essential adjunct to the reporting-surveillance system is the availability of a facility for quarantine, isolation, and medical care of personnel with potential or known laboratory-associated illnesses.

Appendix Q-II-D-1-g-(12). A biosafety manual shall be prepared or adopted. Personnel shall be advised of special hazards and required to read and follow instructions on practices and procedures.

Appendix Q-II-D-1-g-(13). Vacuum lines shall be protected with high efficiency particulate air/HEPA filters and liquid disinfectant traps.

Appendix Q-II-D-2. Animal Facilities (BL4-N)

Appendix Q-II-D-2-a. Animals shall be contained within an enclosed structure (animal room or equivalent) to minimize the possibility of theft or unintentional release and avoid arthropod access.

Appendix Q-II-D-2-b. The interior walls, floors, and ceilings shall be impervious to water and resistant to acids, alkalis, organic solvents, and moderate heat, to facilitate cleaning. Penetrations in these structures and surfaces (e.g., plumbing and utilities) shall be sealed.

Appendix Q-II-D-2-c. Windows in the animal facility shall be closed, sealed, and breakage resistant (e.g., double-pane tempered glass or equivalent).

Appendix Q-II-D-2-d. An autoclave, incinerator, or other effective means to decontaminate animals and wastes shall be available, preferably within the containment area. If feasible, a double-door autoclave is preferred and should be positioned to allow removal of material from the containment area.

Appendix Q-II-D-2-e. Access doors to the containment area shall be self-closing.

Appendix Q-II-D-2-f. All perimeter joints and openings shall be sealed to form an arthropod-proof structure.

Appendix Q-II-D-2-g. The BL4-N laboratory provides a double barrier to prevent the release of recombinant or synthetic nucleic acid molecule containing microorganisms into the environment. Design of the animal facility shall be such that if the barrier of the inner facility is breached, the outer barrier will prevent release into the environment. The animal area shall be separated from all other areas. Passage through two sets of doors shall be the basic requirement for entry into the animal area from access corridors or other contiguous areas. Physical separation of the animal containment area from access corridors or other laboratories or activities shall be provided by a double-door clothes change room equipped with integral showers and airlock.

Appendix Q-II-D-2-h. A necropsy room shall be provided within the BL4-N containment area.

Appendix Q-II-D-2-i. Liquid effluent from containment equipment, sinks, biological safety cabinets, animal rooms, primary barriers, floor drains, and sterilizers shall be decontaminated by heat treatment before being released into the sanitary system. If required by design, regulation, local ordinance or policy, liquid wastes from shower rooms and toilets shall be decontaminated with chemical disinfectants or heat by methods demonstrated to be effective. The procedure used for heat decontamination of liquid wastes shall be monitored with a recording thermometer. The effectiveness of the heat decontamination process system shall be revalidated at minimum on a yearly basis with an indicator organism. More frequent validation, based on the amount of use or other safety factors, shall be left to the discretion of the IBC. If required by design, regulation, local ordinance or policy, liquid wastes from the shower shall be chemically decontaminated using an Environmental Protection Agency-approved germicide. The efficacy of the chemical treatment process shall be validated with an indicator organism. Chemical disinfectants shall be neutralized or diluted before release into general effluent waste systems.

Appendix Q-II-D-2-j. A ducted exhaust air ventilation system shall be provided that creates directional airflow that draws air into the laboratory through the entry area. The exhaust air, which is not recirculated to any other area of the building, shall be discharged to the outside and dispersed away from the occupied areas and air intakes. Personnel shall verify that the direction of the airflow (into the animal room) is proper.

Appendix Q-II-D-2-k. Exhaust air from BL4-N containment area shall be double high efficiency particulate air/HEPA filtered or treated by passing through a certified HEPA filter and an air incinerator before release to the atmosphere. Double HEPA filters shall be required for the supply air system in a BL4-N containment area.

Appendix Q-II-D-2-l. All high efficiency particulate air/HEPA filters' frames and housings shall be certified to have no detectable smoke [dioctyl phthalate] leaks when the exit face (direction of flow) of the filter is scanned above 0.01 percent when measured by a linear or logarithmic photometer. The instrument must demonstrate a threshold sensitivity of at least $1x10^{-3}$ micrograms per liter for 0.3 micrometer diameter dioctyl phthalate particles

and a challenge concentration of 80-120 micrograms per liter. The air sampling rate should be at least 1 cfm (28.3 liters per minute).

Appendix Q-II-D-2-m. If an air incinerator is used in lieu of the second high efficiency particulate air/HEPA filter, it shall be biologically challenged to prove all viable test agents are sterilized. The biological challenge must be minimally $1x10^8$ organisms per cubic foot of airflow through the incinerator. It is universally accepted if bacterial spores are used to challenge and verify that the equipment is capable of killing spores, then assurance is provided that all other known agents are inactivated by the parameters established to operate the equipment. Test spores meeting this criterion are *Bacillus subtilis* var. *niger* or *Bacillus stearothermophilus*. The operating temperature of the incinerator shall be continuously monitored and recorded during use.

Appendix Q-II-D-2-n. All equipment and floor drains shall be equipped with deep traps (minimally 5 inches). Floor drains shall be fitted with isolation plugs or fitted with automatic water fill devices.

Appendix Q-II-D-2-o. Each animal area shall contain a foot, elbow, or automatically operated sink for hand washing. The sink shall be located near the exit door.

Appendix Q-II-D-2-p. Restraining devices for animals may be required to avoid damage to the integrity of the containment animal facility.

Appendix Q-II-D-2-q. The supply water distribution system shall be fitted with a back-flow preventer or break tank.

Appendix Q-II-D-2-r. All utilities, liquid and gas services, shall be protected with devices that avoid back-flow.

Appendix Q-II-D-2-s. Sewer and other atmospheric ventilation lines shall be equipped minimally with a single high efficiency particulate/HEPA filter. Condensate drains from these type housings shall be appropriately connected to a contaminated or sanitary drain system. The drain position in the housing dictates the appropriate system to be used.

Appendix Q-III. Footnotes and References for Appendix Q

Appendix Q-III-A. If a recombinant or synthetic nucleic acid molecule is derived from a Class 2 organism requiring BL2 containment, personnel shall be required to have specific training in handling pathogenic agents and directed by knowledgeable scientists.

Appendix Q-III-B. Personnel who handle pathogenic and potentially lethal agents shall be required to have specific training and be supervised by knowledgeable scientists who are experienced in working with these agents. BL3-N containment also minimizes escape of recombinant or synthetic nucleic acid molecule-containing organisms from exhaust air or waste material from the containment area.

Appendix Q-III-C. Risk Group 4 and restricted microorganisms (see Appendix B, *Classification of Human Etiologic Agents on the Basis of Hazard*, and Sections V-G and V-L, *Footnotes and References of Sections I through IV*) pose a high level of individual risk for acquiring life-threatening diseases to personnel and/or animals. To import animal or plant pathogens, special approval must be obtained from U.S. Department of Agriculture, Animal and Plant Health Inspection Service (APHIS), Veterinary Services, National Center for Import-Export, Products Program, 4700 River Road, Unit 40, Riverdale, MD 20737. Phone: (301) 734-8499; Fax: (301) 734-8226.

Laboratory staff shall be required to have specific and thorough training in handling extremely hazardous infectious agents, primary and secondary containment, standard and special practices, and laboratory design characteristics. The laboratory staff shall be supervised by knowledgeable scientists who are trained and experienced in working with these agents and in the special containment facilities.

Within work areas of the animal facility, all activities shall be confined to the specially equipped animal rooms or support areas. The maximum animal containment area and support areas shall have special engineering and design features to prevent the dissemination of microorganisms into the environment via exhaust air or waste disposal.

Appendix Q-III-D. Other research with non-laboratory animals, which may not appropriately be conducted under conditions described in Appendix Q, may be conducted safely by applying practices routinely used for controlled culture of these biota. In aquatic systems, for example, BL1 equivalent conditions could be met by utilizing growth tanks that provide adequate physical means to avoid the escape of the aquatic species, its gametes, and introduced exogenous genetic material. A mechanism shall be provided to ensure that neither the organisms nor their gametes can escape into the supply or discharge system of the rearing container (e.g., tank, aquarium, etc.) Acceptable barriers include appropriate filtration, irradiation, heat treatment, chemical treatment, etc. Moreover, the top of the rearing container shall be covered to avoid escape of the organism and its gametes. In the event of tank rupture, leakage, or overflow, the construction of the room containing these tanks should prevent the organisms and gametes from entering the building's drains before the organism and its gametes have been inactivated.

Other types of non-laboratory animals (e.g., nematodes, arthropods, and certain forms of smaller animals) may be accommodated by using the appropriate BL1 through BL4 or BL1-P through BL4-P containment practices and procedures as specified in Appendices G and P.

****

ATTACHMENT I--FINAL RISK ASSESSMENT OF SACCHAROMYCES CEREVISIAE

(February 1997)

I. INTRODUCTION

Saccharomyces cerevisiae has an extensive history of use in the area of food processing. Also known as Baker's Yeast or Brewer's Yeast, this organism has been used for centuries as leavening for bread and as a fermenter of alcoholic beverages. With a prolonged history of industrial applications, this yeast has been either the subject of or model for various studies in the principles of microbiology. Jacob Henle based his theories of disease transmission on studies of strains of Brewer's Yeast. Currently, S. cerevisiae is the subject of a major international effort to characterize a eucaryotic genome (Anderson, 1992).

History of Commercial Use and Products Subject to TSCA Jurisdiction

Saccharomyces cerevisiae, in addition to its use in food processing, is widely used for the production of macromolecular cellular components such as lipids, proteins including enzymes, and vitamins (Bigelis, 1985; Stewart and Russell, 1985).

The Food and Drug Administration rates Brewer's Yeast extract as Generally Recognized as Safe (FDA, 1986). Furthermore, the National Institutes of Health in its Guidelines for Research Involving Recombinant DNA Molecules (DHHS, 1986) considers S. cerevisiae a safe organism. Most experiments involving S. cerevisiae have been exempted from the NIH Guidelines based on an analysis of safety (see Appendix C-II of the NIH Guidelines). While alcoholic beverages, vitamins, and bread leavening are covered under the Federal Food, Drug and Cosmetic Act, the production of enzymes and other macromolecules may be subject to TSCA regulation. The abundance of information on S. cerevisiae, derived from its role in industry, has positioned it as a primary model for genetic studies and, by extension, as a strong candidate for genetic manipulation for TSCA applications (Dynamac, 1990).

II. IDENTIFICATION AND CLASSIFICATION

A. Taxonomy and Characterization

Saccharomyces cerevisiae is a yeast. The organism can exist either as a single-celled organism or as pseudomycelia. The cells reproduce by multilateral budding. It produces from one to four ellipsoidal, smooth-walled ascospores. S. cerevisiae can be differentiated from other yeasts based on growth characteristics and physiological traits: principally the ability to ferment individual sugars. Clinical identification of yeast is conducted using commercially available diagnostic kits which classify the organism through analysis of the ability of the yeast to utilize distinct carbohydrates as sole sources of carbon (Buesching et al., 1979; Rosini et al., 1982). More recently, developments in systematics have led to the design of sophisticated techniques for classification, including gas-liquid chromatography of lysed whole cells (Brondz and Olsen, 1979).

As a result of the application of newer techniques arising from innovative approaches, the taxonomy of Saccharomyces is subject to greater scrutiny. The initial classification was based principally on morphological characteristics with specific physiological and biochemical traits used to differentiate between isolates with similar morphological traits. Using these criteria, there are as many as 18 species listed in the literature. In addition, what had been classified as one large heterogeneous species, S. cerevisiae, may, in the future, be divided into four distinct species based on DNA homology studies. The four species are S. cerevisiae, S. bayanus (also known as S. uvarum), S. pasteurianus (also known as S. carlsbergensis), and S. paradoxus. All four represent industrially important species. None of these organisms or other closely related species has been associated with pathogenicity toward humans or has been shown to have adverse effects on the environment.

Any assessment of Saccharomyces must take into consideration the malleability of the current classification. For this assessment of S. cerevisiae the reviews of the organism are based on the classification proposed by Van der Walt (1971).

B. Related Species of Concern

None of the above strains or other closely related species has been associated with pathogenicity toward humans or has been shown to have adverse effects on the environment.

III. HAZARD ASSESSMENT

A. Human Health Hazards

1. Colonization and Pathogenicity

S. cerevisiae is a commonly used industrial microorganism and is ubiquitous in nature, being present on fruits and

vegetables. Industrial workers and the general public come into contact with S. cerevisiae on a daily basis through both inhalation and ingestion (see section IV). Saccharomyces spp. are frequently recovered from the stools and throats of normally healthy individuals. This indicates that humans are in constant contact with these yeasts.

There are individuals who may ingest large quantities of S. cerevisiae every day, for example, people who take the yeast as part of a "health food" regimen. Therefore, studies were conducted to ascertain whether the ingestion of large numbers of these yeasts might result in either colonization, or colonization and secondary spread to other organs of the body. It was found that the installation of very large numbers of S. cerevisiae into the colons of animals would result in both colonization and passage of the yeasts to draining lymph nodes. It required up to 10^{10} S. cerevisiae in a single oral treatment to rats to achieve a detectable passage from the intestine to the lymph nodes (Wolochow et al., 1961). The concentrations of S. cerevisiae required were well beyond those that would be encountered through normal human daily exposure.

S. cerevisiae is not considered a pathogenic microorganism, but has been reported rarely as a cause of opportunistic infections. Eng et al. (1984) described five cases of such infections and reviewed the literature on eight other S. cerevisiae infections (also briefly reviewed by Walsh and Pizzo, 1988). All of the patients in the cases had underlying disease. Some of them had also received antibiotic therapy, thereby suppressing normal bacterial flora and allowing mycotic organisms to become established.

A low concern for the pathogenicity of S. cerevisiae is also illustrated by a series of surveys conducted at hospitals over the last several years. S. cerevisiae accounted for less than 1% of all yeast infections isolated at a cancer hospital and in most of the cases the organism was isolated from the respiratory system (Kiehn et al., 1980). At Yale-New Haven Hospital over the past five years, there have been 50 isolates of S. cerevisiae recovered from patients; however, most of the isolates were considered contaminants (Dynamac, 1991).

2. Toxin Production

There have been no reports of isolates of S. cerevisiae that produce toxins against either humans or animals. However, S. cerevisiae has been shown to produce toxins against other yeasts. These toxins, termed "killer toxins", are proteins or glycoproteins produced by a range of yeasts. The yeasts have been genetically modified to alter activity and are used in industrial settings as a means of controlling contamination of fermentation systems by other yeasts (Sid et al., 1988).

3. Measure of the Degree of Virulence

A number of individual virulence factors have been identified as being associated with the ability of yeasts to cause disease. The principal virulence factors associated with yeasts appear to be phospholipase A and lysophospholipase. It is believed that these enzymes enhance the ability of the yeast to adhere to the cell-wall surface and result in colonization as a first step in the infectious process. Nonpathogenic yeast had considerably lower phospholipase activities. Of a wide range of fungi assayed for phospholipase production, S. cerevisiae was found to have the lowest level of activity (Barrett-Bee et al., 1985). Therefore, based on the phospholipase virulence factor S. cerevisiae is considered a nonpathogenic yeast.

A second factor associated with virulence in yeast is the ability of a fungus to impair the host's immune capabilities. The cell walls of most fungi have the capacity to impede the immune response of the host. In a study to determine the overall pathogenicity of a number of yeasts used in industrial processes, animals exposed to both high levels of S. cerevisiae and cortisone demonstrated a greater ability of the fungus to colonize compared with those animals treated with only the yeast. However, the animals suffered no ill-effects from exposure to S. cerevisiae (Holzschu et al., 1979). Therefore, this study suggests that even with the addition of high levels of an immunosuppressant agent, S. cerevisiae appears to be nonpathogenic.

4. Ability to Transfer Virulence Factor Genes

S. cerevisiae does not carry virulence factors to humans or animals. However, the species does carry linear, double-stranded plasmids which can be transmitted to other Saccharomyces. These plasmids carry genes that encode the "killer toxins" discussed above can be transferred from one Saccharomyces to another. Therefore, gene constructs involving the incorporation of traits using these linear plasmids should be considered to be nonstable.

5. Summary

In conclusion, S. cerevisiae is a organism which has an extensive history of safe use. Despite considerable use of the organism in research and the presence of S. cerevisiae in food, there are limited reports in the literature of its pathogenicity to humans or animals, and only in those cases where the human had a debilitating condition. Factors associated with the virulence of yeasts (i.e., phospholipases) indicate that this organism is nonpathogenic. The organism has not been shown to produce toxins to humans.

B. Environmental Hazards

S. cerevisiae is ubiquitous in nature. It has been recovered from a variety of sites under varying ecological conditions. The organism is used in a variety of industrial scenarios. S. cerevisiae is commonly recovered from a variety of fresh fruits and vegetables, generally those fruits with high levels of fermentable sugars. However, it is not listed as the causative agent of food spoilage for fruits and vegetables (Phaff et al., 1966). The only adverse effect to the environment noted in the literature is the presence of the "killer toxins" which is active against other strains of Saccharomyces.

IV. EXPOSURE ASSESSMENT

A. Worker Exposure

S. cerevisiae is considered a Class 1 Containment Agent under the National Institute of Health (NIH) Guidelines for Recombinant DNA Molecules (U.S. Department of Health and Human Services, 1986).

No data were available for assessing the release and survival specifically for fermentation facilities using S. cerevisiae. Therefore, the potential worker exposures and routine releases to the environment from large-scale, conventional fermentation processes were estimated on information available from eight premanufacture notices submitted to EPA under TSCA Section 5 and from published information collected from non-engineered microorganisms (Reilly, 1991). These values are based on reasonable worst-case scenarios and typical ranges or values are given for comparison.

During fermentation processes, worker exposure is possible during laboratory pipetting, inoculation, sampling, harvesting, extraction, processing and decontamination procedures. A typical site employs less than 10 workers/shift and operates 24 hours/day throughout the year. NIOSH has conducted walk-through surveys of several fermentation facilities in the enzyme industry and monitored for microbial air contamination. These particular facilities were not using recombinant microorganisms, but the processes were considered typical of fermentation process technology. Area samples were taken in locations where the potential for worker exposure was considered to be potentially greatest, i.e., near the fermentor, the seed fermentor, sampling ports, and separation processes (either filter press or rotary drum filter). The workers with the highest potential average exposures at the three facilities visited were those involved in air sampling. Area samples near the sampling port revealed average airborne concentrations ranging from 350 to 648 cfu/m^3. Typically, the Chemical Engineering Branch would not use area

monitoring data to estimate occupational exposure levels since the correlation between area concentrations and worker exposure is highly uncertain. Personal sampling data are not available at the present time. Thus, area sampling data have been the only means of assessing exposures for previous PMN biotechnology submissions. Assuming that 20 samples per day are drawn and that each sample takes up to 5 minutes to collect, the duration of exposure for a single worker will be about 1.5 hours/day. Assuming that the concentration of microorganisms in the worker's breathing zone is equivalent to the levels found in the area sampling, the worst-case daily inhalation exposure is estimated to range up to 650 to 1200 cfu/day. The uncertainty associated with this estimated exposure value is not known (Reilly, 1991).

B. Environmental and General Exposure

1. Fate of the Organism

S. cerevisiae is a normal inhabitant of soils and is widespread in nature. S. cerevisiae is able to take up a wide variety of sugars and amino acids. These traits enhance the organism's ability for long term survival. S. cerevisiae can be isolated from fruits and grains and other materials with a high concentration of carbohydrates (LaVeck, 1991).

2. Releases

Estimates of the number of S. cerevisiae organisms released during production are tabulated in Table 1 (Reilly, 1991). The uncontrolled/untreated scenario assumes no control features for the fermentor off-gases, and no inactivation of the fermentation broth for the liquid and solid waste releases. The containment criteria required for the full exemption scenario assume the use of features or equipment that minimize the number of viable cells in the fermentor off-gases. They also assume inactivation procedures resulting in a validated 6-log reduction of the number of viable microorganisms in the liquid and solid wastes relative to the maximum cell density of the fermentation broth.

TABLE 1. Estimated Number of Viable Saccharomyces cerevisiae Organisms Released During Production

Release Media	Uncontrolled/ Untreated (cfu/day)	Full Exemption (cfu/day)	Release (days/year)
Air Vents	$2x10^{8}$ - 1x1011	$<2x10^{8}$ - 1x1011	350
Rotary Drum Filter	250	250	350
Surface Water	$7x10^{12}$	$7x10^{6}$	90
Soil/Landfill	$7x10^{14}$	$7x10^{8}$	90

Source: Reilly, 1991

These are "worst-case" estimates which assume that the maximum cell density in the fermentation broth for fungi is 10^{7} cfu/ml, with a fermentor size of 70,000 liters, and the separation efficiency for the rotary drum filter is 99 percent.

3. Air

Specific data which indicate the survivability of S. cerevisiae in the atmosphere after release are currently unavailable. Survival of vegetative cells during aerosolization is typically limited due to stresses such as shear forces, desiccation, temperature, and UV light exposure. As with naturally-occurring strains, human exposure may occur via inhalation as the organisms are dispersed in the atmosphere attached to dust particles, or lofted through mechanical or air disturbance.

Air releases from fermentor off-gas could potentially result in nonoccupational inhalation exposures due to point source releases. To estimate exposures from this source, the sector averaging form of the Gaussian algorithm described in Turner (1970) was used. For purposes of this assessment, a release height of 3 meters and downward contact at a distance of 100 meters were assumed. Assuming that there is no removal of organisms by controls/equipment for off-gases, potential human inhalation dose rates are estimated to range from 3.0×10^{3} to 1.5×10^{6} cfu/year for the uncontrolled/untreated scenario and less than that for systems with full exemptions. It should be noted that these estimates represent hypothetical exposures under reasonable worst case conditions (Versar, 1992).

4. Water

The concentrations of S. cerevisiae in surface water were estimated using stream flow values for water bodies receiving

process wastewater discharges from facilities within SIC Code 283 (drugs, medicinal chemicals, and pharmaceuticals). The surface water release data (cfu/day) tabulated in Table 1 were divided by the stream flow values to yield a surface water concentration of the organism (cfu/l). The stream flow values for SIC Code 283 were based on discharger location data retrieved from the Industrial Facilities Dischargers (IFD) database on December 5, 1991, and surface water flow data retrieved from the RXGAGE database. Flow values were obtained for water bodies receiving wastewater discharges from 154 indirect (facilities that send their waste to a POTW) and direct dischargers facilities that have a NPDES permit to discharge to surface water). Tenth percentile values indicate flows for smaller rivers within this distribution of 154 receiving water flows and 50th percentile values indicate flows for more average rivers. The flow value expressed as 7Q10 is the lowest flow observed over seven consecutive days during a 10-year period. The use of this methodology to estimate concentrations of S. cerevisiae in surface water assumes that all of the discharged organisms survive wastewater treatment and that growth is not enhanced by any component of the treatment process. Estimated concentrations of S. cerevisiae in surface water for the uncontrolled/untreated and the full exemption scenarios are tabulated in Table 2 (Versar, 1992).

TABLE 2. S. cerevisiae Concentrations in Surface Water

Flow	Receiving Stream Flow (MLD*)		Organisms (cfu/l)	
	Mean	7Q10	Mean	7Q10
Uncontrolled/Untreated				
10th Percentile	156	5.60	4.5×10^{4}	1.25×10^{6}
50th Percentile	768	68.13	9.11×10^{3}	1.03×10^{5}
Full Exemption				
10th Percentile	156	5.60	4.5×10^{-2}	1.25×100^{0}
50th Percentile	768	68.13	9.11×10^{-3}	1.03×10^{-1}

*MLD = million liters per day
Source: Versar, 1992

5. Soil

Since soil is a natural habitat for S. cerevisiae, it would be expected to survive well in soil. These releases could result in human and environmental exposure (Versar, 1992). It is

currently estimated that over one million tons of naturally-occurring yeast are produced annually during brewing and distilling practices (LaVeck, 1991).

6. Summary

Although direct monitoring data are unavailable, worst case estimates do not suggest high levels of exposure of S. cerevisiae to either workers or the public resulting from normal fermentation operations.

V. INTEGRATED RISK ASSESSMENT

A. Discussion

There is an extensive history of use of and exposure to S. cerevisiae with a very limited record of adverse effects to the environment or human health. Yeast has been used for centuries as a leavening for bread and fermenter of beer without records of virulence. S. cerevisiae is currently classified as a class 1 containment organism under the NIH Guidelines based largely on the extensive history of safe use.

Factors associated with the development of disease states in fungi have been reviewed. Data suggests that only with the ingestion of high levels of S. cerevisiae or with the use of immunosuppressants can S. cerevisiae colonize in the body. Even under those conditions, there were no noted adverse effects. In the few cases which S. cerevisiae was found in association with a disease state, the host was a debilitated individual, generally with an impaired immune system. In other cases the organism was recovered from an immunologically privileged site (i.e., respiratory tract). Many scientists believe that under appropriate conditions any microorganism could serve as an opportunistic pathogen. The cases noted in the above Human Health Assessment, where S. cerevisiae was found in association with a disease state, appear to be classic examples of opportunistic pathogenicity (see III.A.3).

The organism is not a plant or animal pathogen. Despite the fact that S. cerevisiae is ubiquitous in nature, it has not been found to be associated with disease conditions in plants or animals. The only adverse environmental condition that was noted is the production of "killer toxins" by some strains of the yeast. These toxins have a target range that is limited to susceptible yeasts. The toxins, proteins and glycoproteins, are not expected to have a broad environmental effect based largely on the anticipated short persistence of the toxins in soil or

water and by the limited target range. S. cerevisiae "killer toxin" has been used industrially to provide a level of protection against contamination by other yeasts in the fermentation beer.

The current taxonomy of Saccharomyces is under revision based on the development of alternative criteria. However, this should not have a major effect on the risk associated with closely related species. Saccharomyces, as a genus, present low risk to human health or the environment. Criteria used to differentiate between species are based on their ability to utilize specific carbohydrates without relevance to pathogenicity. Nonetheless, this risk assessment applies to those organisms that fall under the classical definition of S. cerevisiae as described by van der Walt (1971).

S. cerevisiae is a ubiquitous organism which, despite its broad exposure, has very limited reported incidence of adverse effects. The extensive history of use, the diversity of products currently produced by the organism, and the attention given this organism as a model for genetic studies collectively makes this organism a prime candidate for full exemption. The increased knowledge derived from the ongoing research should further enhance this organisms' biotechnological uses.

B. Recommendation

Saccharomyces cerevisiae is recommended for the tiered exemption.

VI. REFERENCES

21 CFR 184.1983.

Anderson, A. 1992. Yeast genome project: 300,000 and counting. Science 256:462.

Barrett-Bree, Y. Hayes, R.G. Wilson, and J.F. Ryley. 1985. A comparison of phospholipase activity, cellular adherence and pathogenicity of yeasts. J. Gen. Microbiol. 131:1217-1221.

Bigelis, R. 1985. Primary metabolism and industrial fermentations. in Gene Manipulations in Fungi. J.W. Bennet and L.L. Lasure (ed.) Academic Press, New York, NY. PP 357-

Brondz, I., and I. Olsen. 1990. Multivariate analyses of cellular carbohydrates and fatty acids of Candida albicans, Torulopsis glabrata, and Saccharomyces cerevisiae. J. Clin. Microbiol. 28:1854-1857.

Buesching, W.J., K. Kurek, and G.D. Roberts. 1979. Evaluation of the modified API 20C system for identification of clinically important yeasts. J. Clin. Microbiol. 9565-569.

Bussey, H., T. Vernet, and A.-M. Sdicu. 1988. Mutual antagonism among killer yeasts: competition between K1 and K2 killers and a novel cDNA-based K1-K2 killer strain of Saccharomyces cerevisiae. Can. J. Microbiol. 34:38-44.

Dynamac. 1990. Evaluation of microorganisms for possible exemption under TSCA section 5. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Dynamac. 1991. Human health assessment for Saccharomyces cerevisiae. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Holzschu, D.L., F.W. Chandler, L. Ajello, and D.G. Ahearn. 1979. Evaluation of industrial yeasts for pathogenicity. Sabouraudia 17:71-78.

Kiehn, T.E., F.F. Edwards, and D. Armstrong. 1980. The prevalence of yeasts in gastrointestinal inoculation in antibiotic treated mice. Sabouraudia 21:27-33.

Kough, J. 1990. Clarification of taxonomic questions on Saccharomyces cerevisiae. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

LaVeck, G. 1991. Exposure assessments of microorganisms considered for 5(h)(4) exemptions under the proposed biotech

rule. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Martini, A. and C. Kurtzman. 1985. Deoxyribonucleic acid relatedness among species of the genus Saccharomyces sensu stricto. Int. J. System. Bacteriol. 35:508-511.

Nobre, G.N., and A.F. Ferreira. 1986. Enhancement of Streptococcus faecalis infection and complement depletion in yeast-treated mice. J. Gen. Microbiol. 132:1277-1281.

Organization for Economic Cooperation and Development. 1986. Recombinant DNA safety considerations. Paris, France.

Phaff, H., M. Miller, and E. Mrak. 1966. The life of yeasts. Harvard University Press, Boston, MA.

Reilly, B. 1991. Analysis of environmental releases and occupational exposure in support of the proposed TSCA 5(h)(4) exemption. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Rosini, G., F. Federici, A.E. Vaughn, and A. Martini. 1982. Systematics of the species of the yeast genus Saccharomyces associated with the fermentation industry. Eur. J. Appl. Biotechnol. 15:188-193.

Sayre, P. 1991. Environmental hazard assessment of Saccharomyces cerevisiae for proposed 5(h)4 exemption. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Stewart, G.C. and I. Russell. 1985. The biology of Saccharomyces, pp. 511-536. In A.L. Demain and N.A. Solomon, (eds.), Biology of industrial organisms. Benjamin Cummins Publishers, Menlo Park.

Tao, J., I. Ginsberg, N. Banerjee, W. Held, Y. Koltin, and J.A. Bruenn. 1990. Ustilago maydis KP6 Killer toxin: structure, expression in Saccharomyces cerevisiae, and relationship to other cellular toxins. Mol. Cell. Biol. 10:1373-1381.

Thayer, D.W. 1990. Personnel communication with C. Felkner. (Cited in Dynamac, 1990).

U.S. Department of Health and Human Services, 1986. Guidelines for research involving recombinant DNA molecules; Notice. 51 FR 16958-16985.

Van der Walt, J. 1971. Saccharomyces, pp. 597-605. In J.Lodder, (ed.), The Yeasts, a taxonomic study. North Holland Publ. Co. Amsterdam.

Versar, 1992. Screening level exposure assessment of Saccharomyces uvarum for the 5(h)(4) exemption under the proposed biotech rule. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Wolochow, H., G.J. Hildegrand, and C. Lamanna. 1961. Translocation of microorganisms across the intestinal wall of the rat: effect of microbial size and concentration. J. Infect. Dis. 116:523-528.

RESEARCH ARTICLE

Remanence and survival of commercial yeast in different ecological niches of the vineyard

Gustavo Cordero-Bueso[1], Teresa Arroyo[1], Ana Serrano[1] & Eva Valero[2]

[1]Departamento de Agroalimentación, Instituto Madrileño de Investigación y Desarrollo Rural Agrario y Alimentario Autovía A2, Madrid, Spain; and [2]Departamento de Biología Molecular e Ingeniería Bioquímica, Universidad Pablo de Olavide, Sevilla, Spain

Correspondence: Eva Valero, Departamento de Biología Molecular e Ingeniería Bioquímica, Universidad Pablo de Olavide, Ctra. de Utrera Km 1, s/n. 41013 Sevilla, Spain. Tel.: +34 954 977 894; fax:+34 954 349 813; e-mail: evalero@upo.es

Received 3 November 2010; revised 7 April 2011; accepted 26 April 2011.

DOI:10.1111/j.1574-6941.2011.01124.x

Editor: Philippe Lemanceau

Keywords
dissemination; commercial yeast; vineyard; environment; *Saccharomyces* strains.

Abstract

The use of commercial wine yeast strains as starters has been grown extensively over the past three decades. Wine yeasts are annually released in winery environments; however, little is known about the fate of these strains in the vineyard. To evaluate the industrial starter yeasts' ability to survive in nature and become part of the natural microbiota of musts, commercial yeast was disseminated voluntarily in an experimental vineyard in the Madrid region (Spain). A large sampling plan was devised over 3 years, including samples of grapes, leaves, bark and soil. The disseminated yeast was well represented in the vineyard during the first 8 months. After 2 years, the commercial yeast strain had not survived in the sprayed plants, but a residual population was found in plants situated 50 m east of the sprayed area. After 3 years, commercial yeast disseminated was not found in the sampled vineyard. Grapes and soil showed the highest number of yeasts isolated in the vegetative period, the bark being the main natural reservoir during the resting stages. The result of analysis of population variations from year to year indicated that permanent implantation of commercial strain (K1M) in the vineyard did not occur and its presence was limited in time.

Introduction

The use of commercial wine yeasts strains as starters has grown extensively over the past three decades. Today, the majority of wine production is based on the use of active dried yeast, which ensures rapid and reliable fermentations, and reduces the risk of sluggish or stuck fermentations and of microbial contaminations. Most commercial wine yeast has been selected in the vineyard for oenological traits such as fermentation performance, ethanol tolerance, absence of off-flavours and the production of desirable metabolites. These and other technological developments have contributed towards improving wine quality, and have enhanced the ability of winemakers to control the fermentation process and achieve specific outcomes.

On the other hand, and as a result of modern winemaking practices and diversification of wine products, there is an increasing quest for specialized wine yeast strains. Recombinant-DNA technologies have been successfully applied to wine yeast, generating specialized wine yeast strains that have been engineered for specific traits, such as improved fermentation performance and process efficiency, wine sensory quality and health benefits for consumers. Recent advances in yeast selection tending towards genetic engineering have led to much discussion. Consumers' and governments' concern about public health and the environmental safety of microbial strains engineered by recombinant DNA technologies remains a hurdle to the commercial use of these yeasts (Schuller & Casal, 2005; Verstrepen *et al.* 2006; Fleet, 2008). Wine yeast strain development is an important source of new genetic diversity to increase the options available to winemakers. Consumer demands for newer styles of wines and increasing concerns about the environmental consequences of wine production are providing new challenges for innovation in wine fermentation technology (Pretorius & Hoj, 2005; Fleet, 2008). A list of commercially available strains of wine yeasts has been compiled by Henschke (2007). Nevertheless, an important question is whether the use of these microorganisms could have a real environmental impact.

Wineries are open systems and commercial yeasts are used without any special control and could therefore be

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

FEMS MICROBIOLOGY ECOLOGY

dispersed into the environment in large quantities. The behaviour of these yeasts in the grape berries was studied by several authors (Schuller *et al.* 2005; Valero *et al.* 2005, 2007; Comitini & Ciani, 2006; Goddard *et al.* 2010). In these studies, commercial yeasts were disseminated into the environment year by year and it was not possible to determine their real permanence in the vineyard. This study aims to complete these data by determining the infiltration and permanence in time of the disseminated yeast population. Similarly, as the studies mentioned analyse the populations present in grapes, we do not know whether the soil or any other part of the plant could act as a natural reservoir for commercial yeasts. Therefore, the permeating of commercial yeast in other vineyard niches, such as bark, leaf and soil, is totally unknown, as well as their potential impact on the autochthonous microbiota. In particular, it is not known whether commercial dry yeasts are able to survive in non-grape berry niches, nor whether they will become members of the must microbiota in the following years.

On the other hand, the long-term survival on a spatial scale of a community of commercial strains can be influenced by numerous factors: self-fertilization (Cubillos *et al.* 2009), nutrient quantities in different parts of a vineyard (Palková & Váchová, 2006), dissemination vectors such as insects, small mammal and human activities (Goddard *et al.* 2010), soil type and water run-off (Valero *et al.*, 2005; Marques *et al.*, 2010), the age of the vineyard (Pretorius, 2000), environmental adaptation to a new geographic area (Salinas *et al.* 2010) and the continuous change of climatic conditions such as rainfall, wind direction, temperature, etc. (Pretorius, 2000; Schuller *et al.*, 2005; Valero *et al.* 2005, 2007; Francesca *et al.* 2010).

This study aims to evaluate the behaviour of commercial yeast in the environment via the voluntary dissemination of *Saccharomyces cerevisiae* K1M (Lallemand, France) into different ecological niches of the vineyard in order to answer various questions. Can commercial yeasts remain in the vineyard permanently and form part of the autochthonous microbiota? Is there an ecological niche that acts as the natural reservoir of these yeasts? This information will be highly useful for the continuity of the oenological practice of the use of selected yeasts and even for the possible use of yeasts obtained via new technologies.

Materials and methods

Sampling plan and fermentation procedure

This study was carried out over 3 years (2006–2009), in an experimental vineyard with Mazuelo, Cabernet Sauvignon, and Prieto Picudo (*Vitis vinifera* L.) grapevine varieties located in Alcalá de Henares in the Madrid winegrowing region, Spain (40°31′N, 3°17′W, 610 m altitude). The climatological data were obtained from the vineyard weather station. For these 4 years, the data were fairly homogeneous, and read as follows: for 2006, 2007, 2008 and 2009, the annual mean air temperature was 15.4, 13.4, 14.1 and 15.2 °C, respectively. Regarding the annual mean for precipitation, the data obtained were 414.01, 429.71, 484.90 and 468.11 mm. Under high pressure, the dominant wind direction in the sampling area is from north-east (NE) to south-west (SW). The grapes were harvested in an experimental vineyard with vertical trellises facing in the direction of the gradient with Guyot pruning and bare soil by tillage. Irrigation was performed through a drip system, placing a drip every 75 cm, with a water flow of 2.2 L h^{-1}, resulting in 150 h $year^{-1}$. The soil is a typical Henares river terrace. It is characterized by nonsaturated water at any time of the year and by the presence of large cracks in dry seasons. According to Soil Survey Staff (1999), the soil was classified as follows: Calcic Haploxeralf Alfisol [FAO (2007) equivalence; Calcic Luvisol] with a variable depth of 0, 14, 32 and 60 cm. The soil composition was 44% sand, 35% silt, 20% clay and 1% organic matter. The pH obtained was 8.0.

In order to evaluate the evolution of commercial yeast in the environment, a cell suspension of the dry active commercial yeast K1M (Lallemand, France) was prepared to obtain a concentration of approximately 1×10^6 cells mL^{-1} and each stump or plant and the soil around it were sprayed with 1L of this cell suspension, at the beginning of harvest in the region. The study was carried out over a period of 36 consecutive months starting with the 2006 harvest until the 2009 campaign. Three grapevine plants were sampled within the sprayed area (25 m radius). The subsequent samples were always collected from the same vine. With the present experimental design, samples of grape, leaf, soil and bark were collected at the following times: day 0 (on the same day of the dissemination), 7 days, and 1, 3, 6, 8, 12, 18, 24 and 36 months after dissemination. A control sampling was performed before the dissemination. In order to determine whether sprayed yeasts were disseminated to other points in the vineyard, additional samples were collected over 25 m around the dissemination point, for the control sample, and from 0 to 18 months. After 24 and 36 months, additional samples were collected at specific distances of 25, 50 and 100 m to the north, west, east and south of the point of dissemination. This sampling plan is shown in Fig. 1. Approximately 2 kg of grapes, stems included, and 5 g of leaf, soil and bark were collected under aseptic conditions from each sampling point and placed directly into sterile bags, which were transported to the laboratory under refrigeration conditions and processed within the following 2 h.

At the laboratory, grapes were squeezed by hand in the plastic bags. These bags were opened in the laminar airflow bench, and 80 mL of juice was poured into 100-mL sterile fermenters. Another 50 mL of the must was centrifuged for

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

YM95NC Attachment -- Page 299

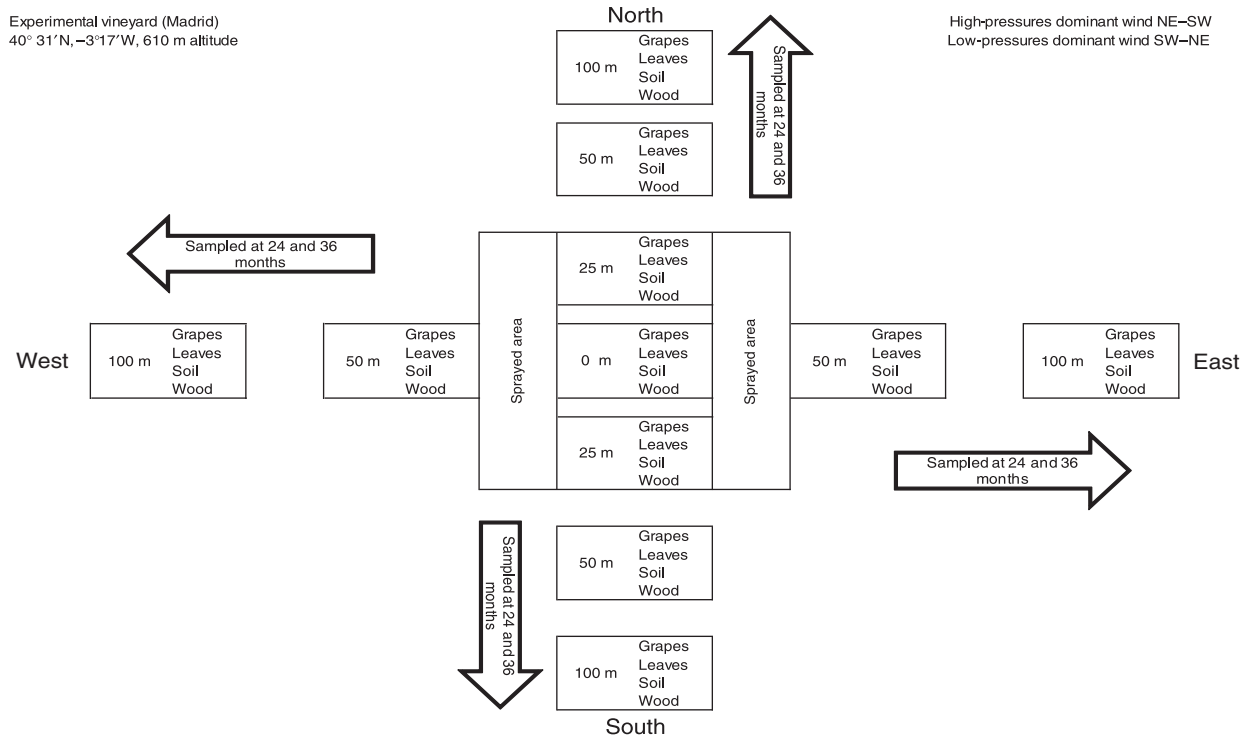


Fig. 1. Map of the experimental vineyard located in Alcalá de Henares, in the Madrid winegrowing region.

5 min at 17 734 ***g***, the supernatant was taken to measure its pH with a pH meter (Crison GLP21, Barcelona, Spain) and Brix degree by refractometry (Atago digital refractometer model Co. Ltd, Tokyo, Japan).

In order to dislodge the yeasts from the leaf, soil and bark samples, 1 g of each were transferred into an assay tube (30 × 115 mm) containing 25 mL of a sterile 0.9% sodium chloride solution. The tubes were shaken on a Mixer Vortex vigorously at a high speed for 60 s to mix the sodium chloride solution with the yeasts. The mix samples were poured into 100-mL sterile fermenters containing 55 mL of synthetic must, MS medium, which mimics a standard grape must (Bely *et al*, 1990); the pH of the MS medium is adjusted to 3.3 by NaOH.

The fermenters with 80 mL of must were placed in a controlled temperature chamber at 20 °C with mechanical agitation (150 r.p.m.). Fermentation progress was monitored daily by weight determination.

Yeast isolation

The yeast community present in the fermentation was evaluated when the weight of the must was reduced by 70 g L^{-1}, corresponding to the consumption of about two-thirds of the sugar content. Tenfold dilutions of 100 μL of must were spread on plates with YPD medium (yeast extract 1% w/v, meat peptone 1% w/v, glucose 2% w/v and agar 2% w/v). The plates were incubated at 28 °C for 24–48 h to allow their growth. After that, colonies were counted and 30 colonies were randomly selected from each fermentation sample. Microbial density was expressed as CFU mL^{-1}.

L-lysine agar (Barnett *et al*., 2000), which is unable to support the growth of *S. cerevisiae*, was used to assess the non-*Saccharomyces* yeasts. All isolates that were not able to grow on the YNB medium with L-lysine as the sole nitrogen source, but grew on the control medium YNB with ammonium sulphate, were considered as *S. cerevisiae* and selected for further steps.

Saccharomyces cerevisiae (K1M) detection

The commercial yeast strain of *S. cerevisiae* K1M (Lallemand, France) was used to monitor the persistence and the evolution of commercial yeast in the vineyard. This yeast has two resistance markers: to the herbicide Diuron and to the antibiotic erythromycin. Although it was isolated in the south east of France, it is not autochthonous of the area under study and it was not used either in the area or within a radius of 60–80 km, as there are no wineries in the vicinity. Furthermore, to the best of our knowledge, this yeast has never been used in wineries in the Madrid region.

In order to select K1M strains, the isolates were grown in a selective medium for these strains, named N+E+D. N medium contains yeast extract 1% w/v, meat peptone 1% w/v, glycerol 2% w/v and agar 2% w/v diluted in 1 L of Sörensen

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved
YM95NC Attachment -- Page 300

buffer 0.05 M, pH 6.25, plus Diuron, minimum 98% (Sigma-Aldrich Chemie GmbH, Germany) 0.08% w/v diluted in 2.5 mL of acetone (Merck KGaA, Germany) and erythromycin (Sigma-Aldrich Chemie GmbH) 0.1% w/v diluted in 4 mL of absolute ethanol (Merck KGaA). A medium without Diuron and erythromycin was used as a negative control.

Molecular identification

A current molecular biology technique, microsatellite multiplex PCR (Vaudano & Garcia-Moruno, 2008), was used in order to confirm the results obtained with the traditional plate-count technique using the selective medium. Colonies isolated from N+E+D medium were assayed to verify whether they had the same microsatellite pattern as *S. cerevisiae* strain K1M (Lallemand).

DNA extraction from yeast isolates was carried out using a commercial kit (ArchivePure DNA Purification System, 5 Prime, Germany), following the instructions provided by the manufacturer, but centrifuging at 17 734 ***g***. The DNA was then stored at $-20\,^{\circ}$C. A UV-Vis spectrophotometer (Nanodrop 1000, Thermo Fisher Scientific Inc.) was used to calculate the quantity of DNA extracted, covering a spectral range from 220 to 750 nm.

Strains were genotyped with three microsatellite loci: SC8132X, YOR267C and SCPTSY7. They were used because of their high degree of polymorphism (Field and Wills, 1998; González-Techera *et al.* 2001; Vaudano & Garcia-Moruno, 2008). The PCR reaction mix and the amplification protocols followed were the same as those used by Vaudano & Garcia-Moruno (2008). Amplified products were separated on an agarose gel (2.5% w/v) with 5 μL mL^{-1} of ethidium bromide (Applichem), in 1 × TBE buffer (Sigma-Aldrich Chemie GmbH, Germany), at 100 V for 120 min. DNA fragment sizes were determined by comparison with a molecular marker (100-pb ladder, Promega, Madison, WI). Moreover, fragment differentiation and allele size determination were performed by single capillary automatic electrophoresis in an ABI 3130 Genetic Analyzer (Applied Biosystem). In order to determine whether the patterns obtained from isolated yeast were identical to sprayed yeast, they were compared with the pattern of the commercial yeast strains K1M (Lallemand).

Results

The aim of this work was to determine the presence and permanence in the vineyard of a population of commercial wine yeast disseminated in the vineyard to determine their impact on the vineyard microbiota. The climatological conditions and other agronomic parameters, such as pruning, irrigation, soil type, etc. are indicated in Materials and methods.

In order to obtain more detailed information about the persistence and implantation of commercial yeast in the environment, K1M strains (Lallemand) were sprayed on an experimental vineyard in the Madrid winegrowing region (Spain). This commercial yeast was isolated in the Languedoc region of France and was never found as autochthonous yeast in Spain. The vineyard sprayed was sampled over the next 3 years, in the area of dissemination and at different distances from this area.

A total of 172 samples were collected over 36 months, 37 of which were taken from grapes, 37 from leaves, 49 from bark and 49 from soil. A sample in triplicate of grapes, leaves, bark and soil, respectively, was taken before commercial yeast dissemination and was used as a control. In the first year of the study, 12 samples, three for each ecological niche (grape, leaf, bark and soil), were collected after 0, 7 days, 1 and 12 months. Also during the period when there are no grapes or leaves on the vine, 3, 6, 8 and 18 months, a total of six samples (three from the bark and three from the soil) were collected. Finally, to evaluate the distribution and evolution of K1M strains and because several chemical, physical and biotic factors could influence the dissemination, at 24 and 36 months, the sampling plan was also extended to 25, 50 and 100 m from the initial area of dissemination (Fig. 1). Hence, at both times, a total of 11 samples were collected from each part of the vine (grape, leaf, soil and bark).

The musts obtained from samples of grape berries were in optimal conditions to carry out the spontaneous fermentations in the presence of fermentative microorganisms. The pH and the Brix degree of the musts obtained are shown in Table 1. Fermentation of the collected samples from the soil and other parts of the vine was carried out in synthetic must, mimicking a standard grape must, as indicated in Materials and methods. Because it is based on an enrichment procedure, the counts obtained with this method do not strictly reflect the initial population in the vineyard, which will always be less than after fermentation. This method was used in order to ensure the detection of fermentative yeast present in very low numbers in the vineyard (Mortimer & Polsinelli, 1999; Pretorius 2000; Martini, 2003).

Table 1. Brix grade and pH for 23 musts obtained from the experimental vineyard during 36 months of the study (mean ± SD)

	Control	Time 0	1 Week	1 Month	3 Months	6 Months	8 Months	12 Months	18 Months	24 Months	36 Months
Brix	24.6 ± 0.4	24.9 ± 0.6	24.9 ± 0.6	25.2 ± 1.3	–	–	–	26.2 ± 0.1	–	22.0 ± 2.8	26.9 ± 2.1
pH	3.4 ± 0.0	3.8 ± 0.0	3.7 ± 0.1	3.5 ± 0.0	–	–	–	3.6 ± 0.0	–	3.5 ± 0.1	3.7 ± 0.2

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

YM95NC Attachment -- Page 301

Table 2. Global distribution and frequency of the commercial strains of *Saccharomyces cerevisiae* (K1M) isolated after spontaneous fermentation from the experimental vineyard over the 36 months studied

	Control	Time 0	1 Week	1 Month	3 Months	6 Months	8 Months	12 Months	18 Months	24 Months	36 Months	Total
Grape												
Samples	3	3	3	3	–	–	–	3	–	11	11	37
Spontaneous Fermentations	3	3	3	3	–	–	–	2	–	3	6	23
Isolates of non-*Saccharomyces*	90	5	18	47	–	–	–	40	–	55	180	435
Isolates of *Saccharomyces* non-K1M	0	0	0	0	–	–	–	0	–	30	0	30
Isolates of *Saccharomyces* K1M	0	85	72	43	–	–	–	20	–	5	0	225
% of K1M	0	94	80	48	–	–	–	33	–	6	0	30
Leaf												
Samples	3	3	3	3	–	–	–	3	–	11	11	37
Spontaneous Fermentations	0	3	2	3	–	–	–	2	–	0	1	11
Isolates of non-*Saccharomyces*	0	60	60	8	–	–	–	30	–	0	30	188
Isolates of *Saccharomyces* non-K1M	0	0	0	0	–	–	–	30	–	0	0	30
Isolates of *Saccharomyces* K1M	0	30	0	82	–	–	–	0	–	0	0	112
% of K1M	0	33	0	91	–	–	–	0	–	0	0	15
Wood												
Samples	3	3	3	3	3	3	3	3	3	11	11	49
Spontaneous Fermentations	0	2	2	3	1	3	2	1	1	0	3	18
Isolates of non-*Saccharomyces*	0	60	60	60	30	0	30	30	0	0	90	360
Isolates of *Saccharomyces* non-K1M	0	0	0	0	0	0	0	0	17	0	0	17
Isolates of *Saccharomyces* K1M	0	0	0	30	0	90	30	0	13	0	0	163
% of K1M	0	0	0	33	0	100	50	0	43	0	0	22
Soil												
Samples	3	3	3	3	3	3	3	3	3	11	11	49
Spontaneous Fermentations	0	2	1	3	3	3	3	3	0	0	2	20
Isolates of non-*Saccharomyces*	0	0	0	48	58	82	46	59	0	0	60	353
Isolates of *Saccharomyces* non-K1M	0	0	0	0	0	0	0	1	0	0	0	1
Isolates of *Saccharomyces* K1M	0	60	30	42	32	8	44	30	0	0	0	246
% of K1M	0	100	100	47	36	9	49	33	0	0	0	33

From the total of 172 samples collected, 72 completed spontaneous fermentations under laboratory conditions, of which 23 were from grapes, 11 from leaves, 18 from the bark and 20 from soil. Thirty colonies were randomly selected from each of the 72 samples that reached spontaneous fermentations. Thus, a total of 2160 yeast colonies were isolated. Based on the L-lysine utilization method (Barnett *et al.* 2000), 1336 were non-*Saccharomyces* yeasts (435, 188, 360 and 353 from grapes, leaves, bark and soil, respectively) and 824 were *Saccharomyces*, of which 78 strains were autochthonous fermentative yeasts. They were isolated from different niches (30 from grapes, 30 from leaves, 17 from the bark and one from the soil). According to the K1M detection test, 746 were commercial *S. cerevisiae* (K1M). The global distribution and frequency of fermentative yeasts population isolated from grapes, leaves, bark and soil, after spontaneous fermentation from the experimental vineyard over the 36 months studied, are shown in Table 2.

The dissemination of K1M strain was efficient; high populations of yeast were found at time 0 (on the same day as dissemination) in grapes, leaves and soil (94%, 33% and 100%, respectively). All isolates in bark corresponded to noncommercial *S. cerevisiae*. After a week, the K1M strain was only isolated in grapes and soil in very high percentages (80% and 100%). The absence of yeast in the bark could be explained by the fact that this part of the plant is dryer than others; furthermore, water and nutrients are concentrated mainly in grapes. One month later, close to senescence, the distribution of commercial yeast in the sprayed area was well represented in all parts of the vines (48% on grapes, 91% on leaves, 33% in the bark and 47% in the soil around the stump). According to Valero *et al.* (2007), during the period of ripeness, higher quantities of nutrients are available for yeasts; this favours their proliferation. Moreover, sugars on leaves are transferred to bark as reservoirs in cold periods. At 3, 6 and 10 months during the resting stages of plants, the soil showed the maximum level of yeasts and the K1M strain was well represented in the bark, although a decreasing tendency was perceived. Twelve months after the voluntary dissemination, the percentage of K1M strain in the grapes and soil (both 33%) was lower than the values observed in the first week. At 18 months (March), only the bark

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved
YM95NC Attachment -- Page 302

contained the sprayed yeast (43%). At 24 months, the K1M strain had not survived in the sprayed plants, but a residual population (6%) was found in grapes situated 50 m east of the dissemination area (Table 3). In the last sampling, at 36 months, commercial yeasts were not found in the vineyard. The dispersion and the number of isolates of K1M according to the time are shown in Fig. 2.

The microsatellite multiplex PCR technique was used to confirm the results obtained. Colonies isolated from N+E+D medium were assayed to verify whether they had identical microsatellite patterns as commercial *S. cerevisiae* K1M (Lallemand), which was used as a control pattern. The 78 indigenous fermentative strains isolated from grapes, leaves, bark and soil were also analysed and compared with the profiles in the IMIDRA yeast collection database showing a unique profile different from K1M and belonging to native yeast of the Madrid winegrowing region (Table 4).

Discussion

Winemakers have used *S. cerevisiae* to make alcoholic beverages for thousands of years. Nowadays, this super-model research organism is central to advances in our biological understanding (Goddard *et al.* 2010). Fermentative yeast populations comprise distinct domesticated and natural groups as well as mosaic strains, but in-depth knowledge of the colonization capacities and environmental risks of commercial yeasts is needed. Wineries are open systems and commercial yeasts have been used without any special control and could therefore be dispersed into the environment in large quantities via the sewage.

The importance of the colonization and persistence of commercial strains of *S. cerevisiae* on the vineyard has also been highlighted in several studies (Schuller *et al.* 2005; Valero *et al.* 2005, 2007; Comitini & Ciani, 2006; Francesca *et al.* 2010). Nevertheless, no data are available regarding the length of permanence of these yeasts in the vineyard or their presence in other vineyard niches. For this reason, this study, carried out in the Madrid winegrowing region, includes aspects that have not been considered in previous works, such as the long-term survival of disseminated commercial yeast in grapes and must, or their permanence in different parts of the vine (leaves and bark) and the soil around the vine, which may act as the natural reservoir of these yeasts during the resting stages of the plants. In this way, it is possible to obtain a more complete picture of the impact of commercial yeast on the natural microbiota of the vineyard.

In the present study, 2160 strains were isolated, of which 1336 were non-*Saccharomyces* yeasts, 78 strains were indigenous *S. cerevisiae* and, according to the K1M detection test, 746 were commercial *S. cerevisiae* (K1M), over a period of 36 months. Yeast biodiversity studies in the vineyard carried out by members of this research team in the Languedoc winegrowing region in the south of France (Valero *et al.* 2007) indicate that the methodology used showed an acceptable reflection of the initial biodiversity. This previous study showed a large proportion of non-*Saccharomyces* yeast species (66%), on the same level as our present study, where these strains represented 63% of the total yeasts isolated over the 36 months. These new data

Table 3. Global distribution and frequency of the commercial strains of *Saccharomyces cerevisiae* (K1M) collected at different distances (0–25, 50 and 100 m) and directions (DP, disseminated point; N, north; S, south; E, east; and W, west) during 2006–2009

	Distance (m)									
	0–25	50				100				
	DP	N	S	E	W	N	S	E	W	Total
Samples	108	8	8	8	8	8	8	8	8	172
Spontaneous Fermentations	62	0	2	2	0	0	2	2	2	72
Isolates of non-*Saccharomyces*	1071	0	60	55	0	30	0	30	90	1336
Isolates of *Saccharomyces* non-K1M	48	0	0	0	0	0	0	30	0	78
Isolates of *Saccharomyces* K1M	741	0	0	5	0	0	0	0	0	746
% of K1M	40	0	0	8	0	0	0	0	0	34

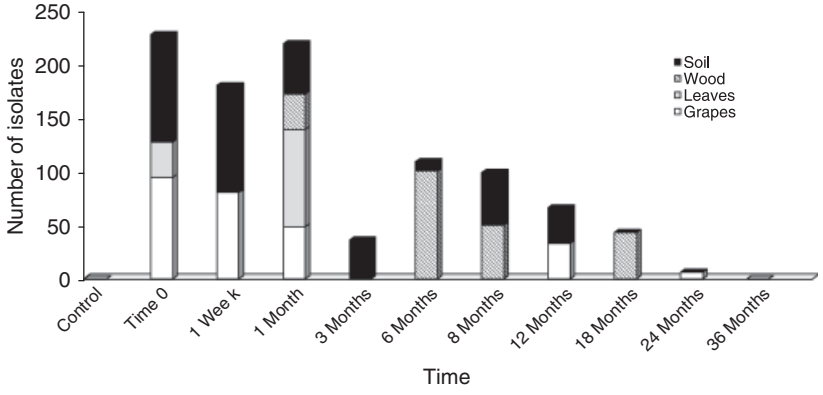


Fig. 2. Overall distribution during 3 years of commercial yeast strain K1M.

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

YM95NC Attachment -- Page 303

Table 4. Microsatellite patterns of *Saccharomyces cerevisiae* strains isolated in the disseminated vineyard

	Allele size (bp)					
Genotype	SCPTSY7-1	SCPTSY7-2	SC8132X-1	SC8132X-2	YOR267C-1	YOR267C-2
K1M*	286	286	193	193	389	389
K1M	286	286	193	193	389	389
A	269	269	193	193	421	421

K1M*, commercial yeast pattern as control; K1M, disseminated commercial yeast; A, indigenous yeast.

confirm previous reports indicating that *S. cerevisiae* is not present in vineyards in large numbers (Pretorius, 2000).

Several authors considered that the origin of wine yeast is still controversial (Mortimer & Polsinelli, 1999; Pretorius 2000; Martini, 2003); in our study, a total of 78 strains of autochthonous *S. cerevisiae* were isolated. These results clearly indicate that indigenous fermentative yeasts occur in a very low percentage (4%) in vineyard ecosystems belonging to the Madrid winegrowing region. In their natural environment, yeasts have to cope with changing temperature, humidity (excess of water or drying) and the effects of various toxic compounds arising either from the environment (e.g. drugs) or produced by other organisms in their immediate surroundings (Palková & Váchová, 2006). Spontaneous fermentations were achieved in order to obtain a significant population of these yeasts, owing to lower availability in the vineyard and with the aim of ascertaining whether *S. cerevisiae* (K1M) was present in each sample. These samples yielded 78 colonies identified as indigenous *S. cerevisiae,* among which we found a unique genotype different from the K1M strain genotype (Table 4). We believe that this provides strong evidence for a discrete population of fermentative yeasts residing in the sampled vineyard. Overall, these communities were found on the grape berries and leaves (30 isolated in both niches) and 17 strains in bark. From soil, only one strain was isolated, but it was the same genotype as that found in the different niches. This therefore appears to indicate that the native yeast found could be a possible 'terroir' yeast candidate. In this way, by preserving and encouraging the autochthonous microbiota, their expression in wines could be asserted (Renouf *et al.* 2005; Francesca *et al.* 2010).

Regarding disseminated K1M, 34% of the total number of isolates from grapes, leaves, bark and soil were the commercial yeast in the proposed sampling plan during 2006, 2007, 2008 and 2009. The highest proportion was found during the 2006 harvest, while a notable decrease was seen in 2007. In 2008, a residual population was found in grapes situated 50 m east of the dissemination area; many factors could have influenced their transportation, such as geography, rainfall and wind direction (this year was the wettest; under low pressures, the predominant winds usually blow from SW to NE in the vineyard area), birds, insects, small mammals and human-aided vectors. After 36 months, in 2009, no disseminated yeasts were found in the epicentre of the vineyard and within a radius of 100 m around the disseminated area. This distance was determined by Valero *et al.* (2005) as the perimeter in which the dissemination is at a maximum.

Regarding the presence of K1M strains in the different niches during the 36 months, grapes and soil were their main reservoirs, containing 30% and 33%, respectively, of the total of the isolates. The other 22% was found in the vine bark and 15% of the commercial yeast corresponded to leaves. With this, soil and bark were the potential containers of these yeasts during the latency period, given that they showed the highest values during the resting stages of the plants.

Sensu strictu species of the genus *Saccharomyces*, as their scientific name implies, usually need high concentrations of sugar and humidity during their biological cycle (Fay & Benavides, 2005). Nevertheless, grapes are sugar-rich environments related to the ripeness having a seasonal occurrence and consequently constitute one of the yeasts' habitats. K1M strains were sprayed over the vine (including grape berries) during the grape development. In this season, higher sugar concentrations are present in grapes, making K1M yeasts' survival possible.

With respect to the soil, a high percentage of isolates of the K1M commercial yeast was found. This could be due to the vine building up organic matter by means of the leaves and old grapes during senescence periods. Thus, yeasts are provided with humidity and sugar and carbon sources. Moreover, in some yeasts of the genus *Saccharomyces*, sexual reproduction is triggered by adverse environmental conditions. A starved diploid cell enters meiosis and produces resistant haploid spores. This form of sexual reproduction, however, usually results in the maintenance of the yeasts during the resting periods or very dry seasons.

According to Barnett *et al.* (2000), *Saccharomyces* species exhibit a strong preference for hexoses such as glucose, fructose, mannose and galactose or simple oligosaccharides, especially maltose, sucrose and raffinose. Sampaio & Gonçalves (2008) analysed via HPLC the presence of these sugars in many tree bark samples from which fermentative yeasts were isolated. It was possible to detect the presence of at least one of the sugars in the bark. However, in the vine before

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

YM95NC Attachment -- Page 304

senescence, high concentrations of sugar are transported from the leaves to the bark in order to have a nutrient reservoir during the resting stages. Our results showed a high percentage of isolates of the disseminated *Saccharomyces* K1M between the third month and the eighth month of the study, when grapes and leaves were not present. This result suggests the availability of hexoses on the vine bark, where there was yeast growth.

The occurrence, and especially domination, of *Saccharomyces* in soils, plant leaves and decaying plant debris is extremely rare. The amount of exudates available to epiphytic microorganisms depends on a number of factors, such as cuticle thickness, the number of stoma and other anatomical features of the leaves (the presence of trichomes, glandular fuzz and extrafloral nectaries), which determine the availability of water and nutrients for epiphytic yeasts, which vary in plants of different ecological and taxonomic groups and necessarily affect yeast numbers (Glushakova *et al.* 2007). In earlier studies, epiphytic yeasts were counted mainly in the autumn, when their population and species diversity were maximal (Bab'eva *et al.* 1995; Glushakova & Chernov, 2007) or, rarely, three or four times a year (Inácio *et al.* 2002). The data obtained by Glushakova *et al.* (2007) indicate that *Saccharomyces* yeasts, contrary to the existing ideas about their predilection for sugar-rich substrates, are typical epiphytic species and can form a major part of the plant-associated yeast population.

Sláviková *et al.* (2007) demonstrated that yeasts were isolated from leaf surfaces of five species of fruit trees located in southwest Slovakia. Leaves are exposed to rapidly fluctuating temperature and relative humidity, which may have an impact on the yeast population. Large fluxes of UV radiation are also one of the most prominent features of the leaf surface environment to which microorganisms have presumably had to adapt (Lindow & Brandl 2003; Sláviková *et al.* 2007). To our knowledge, this is the first time that yeasts have been isolated from *V. vinifera* L. leaves in the Madrid winegrowing region. A total of 188 non-*Saccharomyces* strains were isolated over the 36 months. The highest proportion was found at times 0 and 1 week, and a significant decrease was observed after 1 month; nevertheless, after 12 and 36 months, an increase in the yeast population was observed. After 24 months, no yeasts were isolated. Indigenous *Saccharomyces* were also found at 12 months in vine leaves. Regarding disseminated commercial yeasts K1M, 30 and 82 strains were collected only at times 0 and 1 month owing to an initial presence. However, no survival was found in subsequent samplings.

Summarizing, our results show that the permanent establishment of commercial yeast (K1M) in the vineyard did not occur and is restricted to short distances and a limited period of time. This result reinforces our previous findings (Valero *et al.*, 2005, 2007), which showed that other factors were more important than commercial yeast utilization for the biodiversity of the vineyard. Furthermore, no implantation was produced in the fermentation, as the presence of indigenous strains was not subsequently affected, principally after the first 2 years. It would appear that the environmental risks of the use of commercial yeasts strains were very limited. Our study demonstrates that commercial yeast (K1M) does not displace autochthonous *Saccharomyces*, and they are not sufficiently capable of colonization and adaptation to new vineyard environments on a permanent basis. Nevertheless, more long-term sampling could be advisable to confirm these results and the presence of commercial yeast must be monitored. As far as possible, wineries must avoid the massive release of commercial yeasts in the environment in order to preserve the biodiversity of autochthonous microbiota.

Acknowledgements

This work was supported by the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) and the Ministerio de Ciencia e Innovación, Spain (projects RM 2006-00012-00-00 and RM2007-00008-C02-00). The authors wish to thank J. Tello and I. Aporta for their support during K1M testing repetitions.

References

Bab'eva IP & Chernov IY (1995) Geographical aspects of yeast ecology. *Physiol Genet Biol Rev* **9**: 1–54.

Barnett JA, Payne RW & Yarrow D (2000) *Yeast: Characteristics and Identification*, 3rd edn. Cambridge University Press, Cambridge.

Bely M, Sablayrolles JM & Barre P (1990) Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions. *J Ferment Bioeng* **70**: 246–252.

Comitini F & Ciani M (2006) Survival of inoculated *Saccharomyces cerevisiae* strain on wine grapes during two vintages. *Lett Appl Microbiol* **42**: 248–253.

Cubillos FA, Vásquez C, Faugerón S, Ganga A & Martínez C (2009) Self-fertilization is the main sexual reproduction mechanism in native wine yeast populations. *FEMS Microbiol Ecol* **67**: 162–170.

Fay JC & Benavides JA (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PloS Genet* **1**: e5.

Field D & Wills C (1998) Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellite in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressure and a variety of selective forces. *P Natl Acad Sci USA* **95**: 1647–1652.

Fleet GH (2008) Wine yeast for the future. *FEMS Yeast Res* **8**: 979–995.

Francesca N, Chiurazzi M, Romano R, Aponte M, Settani L & Moschetti G (2010) Indigenous yeast communities in the

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

YM95NC Attachment -- Page 305

environment of '*Rovello bianco*' grape variety and their use in commercial white wine fermentation. *World J Microb Biot* **26**: 337–351.

Glushakova AM & Chernov IY (2007) Seasonal dynamic of the numbers of epiphytic yeasts. *Microbiology* **76**: 590–595.

Glushakova AM, Ivannikova VY, Naumova ES, Chernov IY & Naumov GI (2007) Massive isolation and identification of *Saccharomyces* paradoxus yeast from plant phyllosphere. *Microbiology* **76**: 205–210.

Goddard MR, Anfang N, Tang R, Gardner RC & Jun C (2010) A distinct population of *Saccharomyces cerevisiae* in New Zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Environ Microbiol* **12**: 63–73.

González-Techera A, Jubany S, Carrau FM & Caggero C (2001) Differentiation of industrial wine yeast strains using microsatellite markers. *Lett Appl Microbiol* **33**: 71–75.

Henschke P (2007) Yeast strains available for winemaking 2007–2008. *Aust Wine Res Inst Technol Rev* **171**: 9–29.

Inácio J, Pereira P, de Carvalho M, Fonseca A, Amaral-Collaco MT & Spencer-Martins I (2002) Estimation and diversity of phylloplane mycrobiota on selected plants in a Mediterranean-type ecosystem in Portugal. *Microb Ecol* **44**: 344–353.

IUSS Working Group WRB 2007 *World Reference Base for Soil Resources. World Soil Resources Reports No 103*. FAO, Rome.

Lindow SE & Brandl MT (2003) Microbiology of the phyllosphere. *Appl Environ Microbiol* **69**: 1875–1883.

Marques MJ, García-Muñoz S, Muñoz-Organero G & Bienes R (2010) Soil conservation beneath grass cover in hillside vineyards under mediterranean climatic conditions (Madrid, Spain). *Land Degrad Dev* **21**: 122–131.

Martini A (2003) Biotechnology of natural and winery-associated strains of *Saccharomyces cerevisiae*. *Int Microbiol* **6**: 207–209.

Mortimer R & Polsinelli M (1999) On the origins of wine yeast. *Res Microbiol* **150**: 199–204.

Palková Z & Váchová L (2006) Life within a community: benefit to yeast long-term survival. *FEMS Microbiol Rev* **30**: 806–824.

Pretorius IS (2000) Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* **16**: 675–729.

Pretorius IS & Hoj PB (2005) Grape and wine technology: challenges, opportunities and potential benefits. *Aust J Grape Wine R* **11**: 83–108.

Renouf V, Claisse O & Lonvaud-Funel A (2005) Understanding the microbial ecosystem on the grape berry surface through numeration and identification of yeast and bacteria. *Aust J Grape Wine R* **11**: 316–327.

Salinas F, Mandakovic D, Urzua U, Massera A, Miras S, Combina M, Ganga MA & Marínez C (2010) Genomic and phenotypic comparison between similar wine yeast strains of *Saccharomyces cerevisiae* from different geographic origins. *J Appl Microbiol* **108**: 1850–1858.

Sampaio JP & Gonçalves P (2008) Natural populations of *Saccharomyces kudriavzeii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl Environ Microb* **74**: 2144–2152.

Schuller D & Casal M (2005) The use of genetically modified *Saccharomyces cerevisiae* strains in the wine industry. *Appl Microbiol Biot* **68**: 292–304.

Schuller D, Alves H, Dequin S & Casal M (2005) Ecological survey of *Saccharomyces* strains from vineyards in the Vinho Verde region of Portugal. *FEMS Microbiol Ecol* **51**: 167–177.

Sláviková E, Vadkertiová R & Vránová D (2007) Yeasts colonizing the leaf surfaces. *J Basic Microb* **47**: 344–350.

(1999) *Soil Taxonomy: a Basic System of Soil Classification for Making and Interpreting Soil Surveys, 2nd edn. USDA, Agriculture Handbook. Natural Resources Conservation Service No 436*. Soil Survey Staff, Pittsburgh, PA. pp. 163–260.

Valero E, Schuller D, Cambon B, Casal M & Dequin S (2005) Dissemination and survival of comercial wine yeast in the vineyard: a large scale, three-year study. *FEMS Yeast Res* **5**: 959–969.

Valero E, Cambon B, Schuller D, Casal M & Dequin S (2007) Biodiversity of *Saccharomyces* yeast strains from grape berries of wine-producing areas using starter commercial yeasts. *FEMS Yeast Res* **7**: 317–329.

Vaudano E & Garcia-Moruno E (2008) Discrimination of *Saccharomyces cerevisiae* wine strains using microsatellite multiplex PCR and band pattern analysis. *Food Microbiol* **25**: 56–64.

Verstrepen KJ, Chambers PJ & Pretorius IS (2006) The development of superior yeast strains for the food and beverage industries: challenges, opportunities and potential benefits. *Yeasts in Food and Beverages* (Querol A & Fleet GH, eds), pp. 399–444. Springer, Berlin.

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

YM95NC Attachment -- Page 306

ELSEVIER

FEMS Yeast Research 5 (2005) 959–969

FEMS Yeast Research

www.fems-microbiology.org

Dissemination and survival of commercial wine yeast in the vineyard: A large-scale, three-years study

Eva Valero [a,1], Dorit Schuller [b,1], Brigitte Cambon [a], Margarida Casal [b], Sylvie Dequin [a,*]

[a] *UMR Sciences pour l'Oenologie, Microbiologie, INRA, 34060 Montpellier Cedex 1, France*
[b] *Centro de Biologia, Departamento de Biologia, Universidade do Minho, Braga, Portugal*

Received 24 December 2004; received in revised form 5 April 2005; accepted 19 April 2005

First published online 13 June 2005

Abstract

The use of commercial wine yeast strains as starters has been extensively generalised over the past two decades. In this study, a large-scale sampling plan was devised over a period of three years in six different vineyards to evaluate the dynamics and survival of industrial yeast strains in the vineyard. A total of 198 grape samples were collected at various distances from the wineries, before and after harvest, and yeast strains isolated after spontaneous fermentation were subsequently identified by molecular methods. Among 3780 yeast strains identified, 296 isolates had a genetic profile identical to that of commercial yeast strains. For a large majority (94%), these strains were recovered at very close proximity to the winery (10–200 m). Commercial strains were mostly found in the post-harvest samples, reflecting immediate dissemination. Analysis of population variations from year to year indicated that permanent implantation of commercial strains in the vineyard did not occur, but instead that these strains were subject to natural fluctuations of periodical appearance/disappearance like autochthonous strains. Our data show that dissemination of commercial yeast in the vineyard is restricted to short distances and limited periods of time and is largely favoured by the presence of water run-off.
© 2005 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Wine yeast dissemination; Vineyard ecosystems; Molecular identification methods

1. Introduction

The predominant yeast species used in the production of wine is *Saccharomyces cerevisiae*, universally known as "wine yeast". Under selective conditions of grape must fermentation, yeasts efficiently compete with other microorganisms present in musts, such as moulds and lactic- and acetic-acid bacteria. A succession of various yeast species – the apiculate yeasts *Hanseniaspora uvarum* (=*Kloeckera apiculata*) and other yeasts of the genera *Metschnikowia*, *Candida* or *Pichia* – is found in the early stages of fermentation [1]. As the concentration of ethanol increases, these species are rapidly outgrown by *S. cerevisiae* and related species, which invariably dominate the later stages of the process.

Since the beginning of the 1980s, the use of active dried *S. cerevisiae* yeast starters has been extensively generalised. Today, the majority of wine production is based on the use of active dried yeast, which ensures rapid and reliable fermentation, and reduces the risk of sluggish or stuck fermentation and of microbial contamination. Most commercial wine yeast strains available today have been selected in the vineyard for enological traits such as fermentation performance, ethanol tolerance, absence of off-flavors and production of desirable metabolites. These and other technological developments have contributed to an improvement in the

* Corresponding author. Tel.: +33 4 9961 2528; fax: +33 4 9961 2857.
E-mail address: dequin@ensam.inra.fr (S. Dequin).
[1] Contributed equally to this work.

1567-1356/$22.00 © 2005 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.femsyr.2005.04.007

quality of wine, and have enhanced the ability of winemakers to control the fermentation process and achieve specific outcomes.

As a result of modern winemaking practices and diversification of wine products, there is an increasing quest for specialised wine yeast strains. During the last two decades a considerable knowledge of *S. cerevisiae* genetics and physiology has been generated as well as numerous genetics tools. Recombinant-DNA technologies have been successfully applied to wine yeast, generating specialized wine yeast strains which have been engineered for specific traits, such as improved fermentation performance and process efficiency, wine sensory quality and health benefits for consumers [2–8].

In view of a possible future use of genetically modified wine yeasts, a sound evaluation of the potential environmental impact of genetically modified wine yeast is absolutely required. In this context, industrial yeasts used as fermentation starters are a good study model to evaluate the competition and the influence of inoculated strains on the fermentations of the following years, especially those performed according to traditional practices which rely on spontaneous fermentations. Commercial yeasts are classically used in winemaking without any special containment and are annually released in large quantities, together with liquid and solid wine-making residues, in the environment around the winery. The behaviour of these yeasts in the ecosystem of the vineyard is totally unknown, as is their potential impact on the natural microbiota. In particular, it is not known if commercial strains are able to survive in nature and to become members of the vineyard microbiota.

Limited data are available that could contribute to an evaluation of the importance of starter yeast dissemination and permanence in the vineyard [9–11]. Recently, a large-scale biogeographical study in South-African vineyards has been carried out over four years. In five areas, situated in the coastal region vineyards of the Western Cape, 13 samples were collected and commercial yeasts were recovered from three samples [12,13]. These studies have made it necessary to carry out this type of study on a larger scale, with the aim of increasing the statistic significance of the results obtained.

The present large-scale study, carried out at different geographical localizations in France and Portugal, aims at evaluating the industrial starter yeasts' ability to spread and survive in nature.

2. Materials and methods

2.1. Sampling plan and selection of wineries

Grapes were harvested during three consecutive years (2001–2003) in six vineyards, three of which were located in the south of France and three in the northwest of Portugal, as shown in Fig. 1. In France, the wineries were located in the Languedoc-Roussillon Region, around the Mediterranean city of Montpellier, and the vineyards were situated at a distance of 30 and 80 km from each other. In Portugal, the three wineries were located in the north, centre and south of the Região Demarcada dos Vinhos Verdes, the distance between each being approximately 50 km. In each vineyard, six sampling points were defined according to the predominating wind direction at a distance between 20 and 1000 m from the winery, as shown in Fig. 1.

In order to evaluate the survival of commercial yeast over years, a first sampling campaign was performed before the winery started wine production with the use of commercial yeast strains (pre-harvest samples). In a second post-harvest sampling campaign, the grapes were collected after the onset of wine production, in order to evaluate the immediate commercial yeast dissemination from the winery. The gap between the pre- and post-harvest campaigns was ten days, during which time waste water was released from the wineries. In the consecutive years, samples were always collected from the same area at a radius of 5 m. With the present experimental design, 72 grape samples were collected each year. The selected wineries have used one or more commercial yeast strains consecutively during at least the last five years. Tables 1 and 2 show the commercial yeasts used in each winery during the period studied (2001–2003) and their geographic origin, respectively.

2.2. Sample collection and yeast isolation

From each sampling point, approximately 2 kg of grapes were collected aseptically and directly placed into sterile plastic bags, which were transported to the laboratory in cool bags. At the laboratory, grapes were crushed by hand in the plastic bags, which were opened and 180 ml of juice was poured into sterile 250-ml fermentors. The fermentors were placed in a temperature-controlled room at 20 °C with mechanical agitation. Daily weight determinations allowed the monitoring of the fermentation progress. The yeast community was analysed when the must weight was reduced by 70 g l^{-1}, corresponding to the consumption of about 2/3 of the sugar content. Must samples were diluted and spread on plates with YEPD medium (yeast extract 1% w/v, peptone 1% w/v, glucose 2% w/v, agar 2% w/v), and after 2 days of incubation 30 randomly selected colonies were collected from each spontaneous fermentation.

2.3. Selection of Saccharomyces

To rapidly discriminate between *Saccharomyces* and non-*Saccharomyces* yeast, every isolate was evaluated according to its ability to grow on L-lysine [14]. All

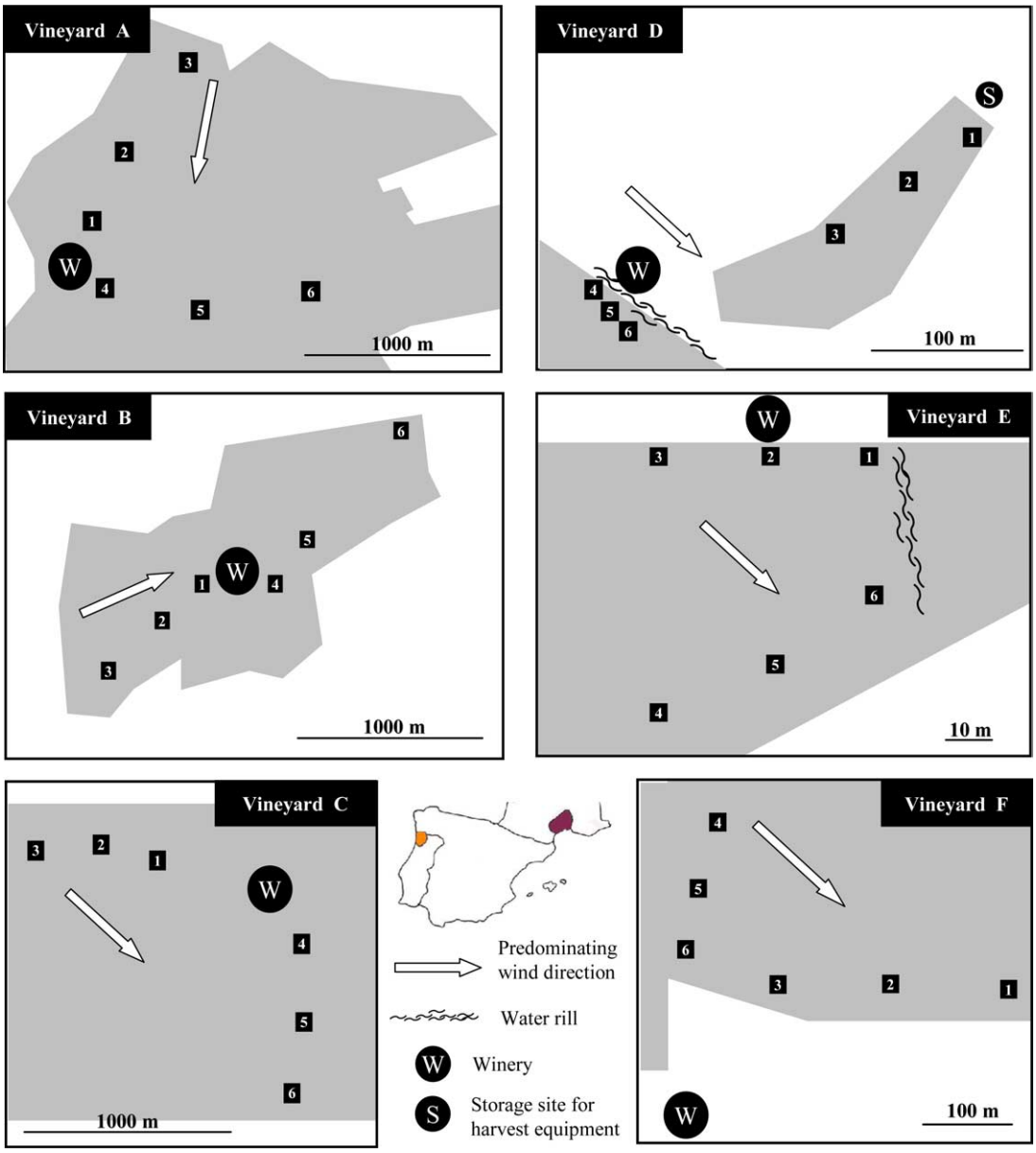


Fig. 1. Geographic localization of the vineyards belonging to the Languedoc (A–C) and Vinho Verde (D–F) wine regions, with an indication of the sampling sites in each of the six vineyards. In each site, two samples (pre- and post-harvest campaign) were collected. Factors that may influence the dissemination of the yeasts are indicated in the figure.

isolates that were not able to grow on the YNB medium with L-lysine as the sole nitrogen source, but grew on the control medium YNB with ammonium sulphate were considered as *Saccharomyces* and selected for further molecular identification.

2.4. Molecular identification methods

DNA was extracted from yeast cells cultivated in 1 ml YEPD medium (36 h, 28 °C, 160 rpm) as described previously [15], with a modified cell lysis procedure, using 25 U of Zymolase (Sigma, St. Louis, MO). Cell lysis was dependent on the strain and lasted between 20 min and 1 h (37 °C).

Mitochondrial-DNA restriction profiles were established as described previously [16]. Digestions (*Hinf*I) were performed overnight at 37 °C in a final volume of 20 μl [17].

Microsatellite analysis was performed using six loci (ScAAT1–ScAAT6), previously described by Pérez et al. [18], that were amplified (iCycler thermal cycler Bio-Rad, Marnes-La Coquette, France) in two multiplex reactions. The samples were denatured and separated by capillary electrophoresis in an ABI Prism 310 DNA

Table 1
Commercial yeast strains used in each winery during 2001–2003

Year	Winery A	Winery B	Winery C	Winery D	Winery E	Winery F
2001	**K1M ICV-INRA**[a] **ICV D254** **Enolevure K34** **Lalvin QA23** **ICV D47**	**K1M ICV-INRA** **ICV D254** ICV D80 Uvaline BL Lalvin BM45 Maurivin AWRI2	**K1M ICV-INRA** **Zymaflore VL3** **Maurivin PDM** ICV D254 ICV D47 Uvaline arôme Vitilevure-Chardonnay Anchor VIN 13	**ZymafloreVL1** **ZymafloreVL3** **Zymaflore F10** Zymaflore F15 Uvaferm L2056 Lalvin CY 3079 Uvaferm ALB Uvaferm 2287	**Zymaflore VL1** Lalvin EC 1118	**Zymaflore VL1**
2002	**K1M ICV-INRA** **ICV D254** **Lalvin QA23**	**K1M ICV-INRA** ICV D80 Uvaline BL Lalvin BM45 Maurivin AWRI2 Uvaline CVR	**K1M ICV-INRA** **Maurivin PDM** ICV D47 Anchor VIN 13 Zymaflore VL3 Anchor NT 116 Vitilevure-Sauvignon	**ZymafloreVL1** **ZymafloreVL3** **Zymaflore F10** Zymaflore F15 Uvaferm ALB Uvaferm 228 Uvaferm CS2	**Zymaflore VL1** Lalvin EC 1118 Levuline BRG Fermichamp	**Zymaflore VL1** Lalvin EC 1118
2003	**K1M ICV-INRA** ICV D254 Enolevure K34	**K1M ICV-INRA** Uvaline BL Lalvin BM45 Anchor NT 45 Anchor NT 50 ICV D80 Uvaline CVR Enolevure K34 Maurivin PDM	**K1M ICV-INRA** **Zymaflore VL3** **Maurivin PDM** Vitilevure-Chardonnay Vitilevure-Sauvignon	**ZymafloreVL1** **ZymafloreVL3** **Zymaflore F10** Zymaflore F15 Zymaflore VL2 Uvaferm ALB	**Zymaflore VL1** Fermafine Fermafruit IOC 18-2007 Lalvin CY 3079	**Zymaflore VL1** Lalvin CY 3079

All strains belong to *S. cerevisiae*.
[a] The strains shown in bold were used for at least the last five years prior to the study.

sequencer (Applied Biosystems, Foster City, CA) and analysed using Genescan (Metairie, LA) software. The complete method was described by Schuller et al. [17].

Chromosomal profiles were established by pulsed-field gel electrophoresis (PFGE) using yeast chromosomal DNA prepared in plugs and the TAFE (transverse alternating field electrophoresis) system (Geneline, Beckman, Villepinte, France) as previously described [19]. The gels were run for 6 h at 250 V with a 35-s pulse time, followed by 20 h at 275 V with a 55-s pulse time, at 14 °C.

3. Results and discussion

3.1. Sampling sites and isolation of Saccharomyces

A large sampling plan was followed: a total of 198 samples were collected during three consecutive campaigns (2001–2003), 108 of which were taken in France and 90 in Portugal. It is to be noted, as can be observed in Fig. 1, that due to geographical constraints the samples in Portugal were collected much closer to the winery than those in France. In the French wineries (A–C), the sample sites were located at a distance of between 100 and 1000 m from the winery, whereas in the Portuguese wineries (D–F) half of the sampling sites were located at a distance of less than 70 m from the winery and none was located further than 400 m.

Table 3 shows the global data in each country, broken down into years. Of the 198 samples, 126 musts (64%) produced spontaneous fermentations, 20% and 44% by musts from pre-harvest and post-harvest campaigns, respectively. The percentages of spontaneous fermentations in both countries were similar, 66% in France and 60% in Portugal. A total of 3780 colonies were isolated from these fermentations (2160 and 1620 in France and Portugal, respectively).

Discrimination between *Saccharomyces* and non-*Saccharomyces* strains isolated in Languedoc was performed using a selective medium with L-lysine as the sole nitrogen source [14]. By this method only two species of the genus *Saccharomyces* (i.e., *S. kluyveri* and *S. unisporus*), which do not occur in enological environments, are capable of growing with L-lysine. From this we concluded that yeasts, isolated after fermentation, that can utilize L-lysine, do not belong to the genus *Saccharomyces*. To confirm this hypothesis, isolates from the fastest fermentations that grew in L-lysine medium were identified by PCR-RFLP analysis of the rDNA ITS region [20]. The results confirmed that they were non-*Saccharomyces* strains, belonging mainly to the genus *Kloeckera* (data not shown). All isolates not able to grow on the L-lysine medium were therefore selected for molecular identification. In Portugal, all isolates were assigned to different groups according to their mtDNA RFLP pattern. One representative strain from each group was ran-

Table 2
Geographic origin of the commercial yeast strains used in the wineries studied

Strains	Origin
Anchor NT 45	South Africa
Anchor NT 50	South Africa
Anchor NT 116	South Africa
Anchor VIN 13	Stellenbosch, South Africa
Enolevure K34	Valencia, Spain
Fermafine	Not known
Fermafruit	Not known
Fermichamp	Alsace, France
ICV D 47	Rhône, France
ICV D 80	Rhône, France
ICV D 254	Languedoc, France
IOC 18-2007	Not known
K1M ICV-INRA	Languedoc, France
Lalvin BM 45	Sangiovese, Italy
Lalvin EC1118	Champagne, France
Lalvin QA23	Portugal
Lalvin Cy 3079	Bourgogne, France
Levuline BRG	Not known
Maurivin AWR12	Bordelais, France
Maurivin PDM	Champagne, France
Uvaferm 228	France
Uvaferm ALB	Not known
Uvaferm CS2	Alsace, France
Uvaferm L 2056	Rhône, France
Uvaline arôme	Loire, France
Uvaline BL	Champagne, France
Uvaline CVR	Not known
Vitilevure Chardonnay	Languedoc, France
Vitilevure Sauvignon	Sauvignon, France
Zymaflore F10	Bordelais, France
Zymaflore F15	Gironde, France
Zymaflore VL1	Gironde, France
Zymaflore VL2	Burgundy, France
Zymaflore VL3	Gironde, France

domly withdrawn, and its ability to grow on L-lysine was tested. Based on these methods, 2355 *Saccharomyces* strains were selected from the 3780 isolates collected during the three years.

3.2. Geographic distribution of recovered commercial yeast strains

The global composition of the yeast population isolated after fermentation from the six wineries over the three years studied, in pre- and post-harvest campaigns, is shown in Fig. 2. Table 4 shows the distribution and frequency of commercial yeasts in each vineyard.

Identification of *Saccharomyces* strains was performed by different molecular-typing methods, depending on the specific resources of each laboratory. An example of genetic profiles, both of natural isolates and commercial yeast strains from France and Portugal, is shown in Fig. 3. Chromosomal-pattern analysis of 735 *Saccharomyces* isolates from France (wineries A–C) was performed, and compared with that of the 19 commercial yeasts used in the three wineries. In Portugal, all 1620 isolates were analysed by mtDNA RFLP (*Hinf*I), and their patterns compared to those of a strain collection including all strains used by the three wineries. At least one representative isolate of each group of strains showing identical mtDNA RFLP patterns identical to those of commercial strains was further confirmed by microsatellite analysis. In order to evaluate the discriminatory power of these three methods, we had previously performed a survey of the genetic polymorphisms generated by distinct methods on a total of 23 commercial yeast strains used in the wineries of the two countries [17]. The results had shown that the discriminatory power of microsatellite typing using these six different loci, and that of mtDNA RFLP patterns generated by the enzyme *Hinf*I, was the same and similar to that of karyotype analysis. Among the 23 commercial yeast strains analysed, 21 different patterns were obtained using the first two methods and 22 using the last. Due to the verified similarity of the discriminatory power of these methods, any of them can be used for our study, with comparable results.

The analysis of genetic profiles of 2355 out of 3780 *Saccharomyces* isolates resulted in the identification of 296 commercial yeasts, representing 7.8% of the fermentative yeast community (Table 4), the majority of which (5.8%) were recovered in post-harvest campaigns (Fig. 2). It should be noted that in this study fermentation is used as an enrichment tool for *Saccharomyces* strains. Therefore, the present results do not allow conclusions about the number of strains occurring on the surface of the grape, which is in fact very low, but reflect only those strains that could possibly have some enological use. Instead, the number of fermentations with at least one commercial yeast strain gives a better picture of the situation as it occurs in vineyards; commercial yeast

Table 3
Distribution of global data by country and year

	2001		2002		2003		Total
	France	Portugal	France	Portugal	France	Portugal	
Samples	36	36	36	18	36	36	198
Spontaneous fermentations	24	19	33	12	15	23	126
Isolates	720	570	990	360	450	690	3780
Saccharomyces strains	406	570	120	360	209	690	2355

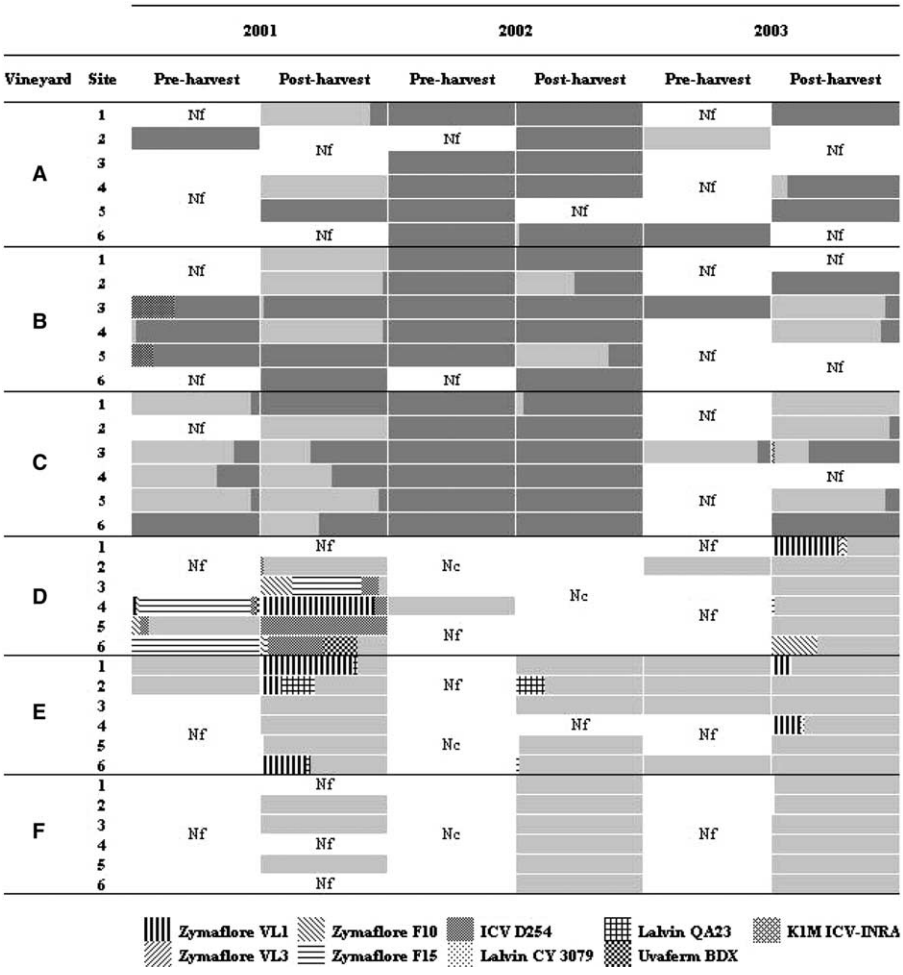


Fig. 2. Global composition of the yeast communities isolated from each site at the six wineries during the pre- and post-harvest sampling campaigns over the three years. The motifs show the presence of commercial yeasts, light grey indicates other *Saccharomyces* strains and dark grey the non-*Saccharomyces* strains. Nf: no fermentation; Nc: not collected.

Table 4
Commercial yeast strains recovered in each vineyard over the three years studied

Vineyards	A	B	C	D	E	F	Total
Spontaneous fermentations	19	24	29	16	23	15	126
Spontaneous fermentations with $\geqslant$ 1 commercial yeast strains	0	2	1	11	9	2	25
Isolates	570	720	870	480	690	450	3780
Commercial yeast strains	0	15[a]	1	206	54 + 18[a]	2	296
% Commercial yeast/number of isolates	0	2	0.1	43	10	0.5	7.8

[a] Commercial yeasts initially isolated in the same region.

strains were recovered from 12% of the samples (Tables 3 and 4).

These global data reflect very different situations. In the vineyards where the sampling sites were placed at a greater distance from the winery, i.e., vineyard F in Portugal and the three French vineyards (A–C), the occurrence of commercial yeasts was very low, representing between 0% and 2% of the fermentative community, and these strains were isolated from only five samples (Table 4). In France, the genetic profile of 16 clones out of 735 *Saccharomyces* isolates (2%) was identical to that of commercial yeasts (Fig. 2). These strains corresponded to 0.8% of the yeast strains isolated after fermentation. With only one exception, these strains (fifteen isolates) had a profile identical to that of the autochthonous strain ICV D254 and were found at the same site (winery B), in pre-harvest samples taken in 2001 (Fig. 2). This fact could indicate previous dissemination, but this cannot be confirmed, since the commercial yeast strain ICV D254 was initially isolated from the same region in the south of France (Table 2) where the study was carried out. No commercial yeasts were found

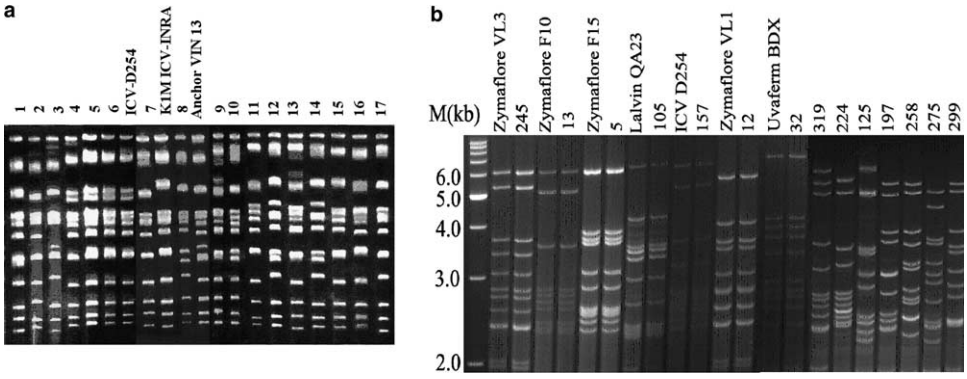


Fig. 3. Examples of molecular fingerprinting of commercial yeasts and natural isolates (indicated by numbers). (a) Chromosomal profiles of commercial yeasts and natural isolates from France. Profiles 4–6 are identical to ICV-D254. (b) mtDNA RFLP profiles of commercial yeasts and natural isolates from Portugal. Profiles 245, 13, 5, 105, 157, 12 and 32 are identical to Zymaflore VL3, F10, F15, Lalvin QA23, ICV-D254, Zymaflore VL1 and Uvaferm BDX, respectively.

from winery A and one colony, isolated in 2003 from winery C (site 3), had the same profile as K1M ICV-INRA, used in the three French wineries for the last 5–15 years. It is noteworthy that this yeast, which has been used extensively for a considerable length of time, has never been found in the vineyard, except in this case. The same situation occurs in the Portuguese winery F, where only two isolates were found with the same profile as the extensively used commercial yeast, Zymaflore VL1, in use for five years. Since strain ICV D254 was initially isolated in the region from which it was recovered, dissemination in these four vineyards was proved only by the presence of three isolates (0.1% of the fermentative flora), one of K1M ICV-INRA and two of Zymaflore VL1. Their presence might be considered to be due to immediate dissemination, probably mediated by insects or another occasional dissemination vector. It is, in any case, evident that the presence of the most-widely used commercial yeasts for the last 5–10 years in French (i.e., K1M ICV-INRA) and in Portuguese wineries (i.e., Zymaflora VL1) was incidental and that these yeasts hardly ever dominated the microflora of any of these four vineyards. These results, in accordance with those obtained previously in South-African vineyards [12,13], indicate a very low level of dissemination/implantation of commercial yeasts in the vineyard ecosystem.

The results were very different in the Portuguese wineries D and E, for which a high number of commercial strains was isolated from 20 spontaneous fermentations, representing 43% and 10% of the fermentative yeast community, respectively. Indeed, the large majority (94%) of the commercial strains isolated within the six vineyards were recovered from these two vineyards only, and 70% solely from vineyard D. It can be observed from Figs. 1 and 2 that the majority of the commercial strains in these two vineyards were isolated from sites closest to the winery, namely sites 4–6 in vineyard D and sites 1, 2 and 6 in vineyard E. The major difference between these two vineyards and the four others is that the sample sites in the first two were located in close proximity to the winery (Fig. 1). In addition, the presence of water run-off at these sites indicates that dissemination is probably largely favoured by liquid effluents. In vineyard D, due to the ground inclination, water run-off flowing from the winery to the vine may contribute to the frequent occurrence of commercial strains at these sites. It is also noteworthy that site 1 of winery E, where the highest number of VL1 strains was recovered, is located close to a rill that transports run-off water from the winery, emphasizing the importance of water as a vehicle for yeast strain dissemination. Furthermore, the dumping site of macerated grape skins also is adjacent to site 1, constituting a fermenting sugary substrate harbouring large amounts of yeasts that can be distributed throughout the vineyard.

An overview of the dissemination of commercial strains in relation to their distance from the winery is shown in Fig. 4. Of the commercial strains, 94% were found in a radius of around 10–200 m from the winery and a large majority (78%) was recovered at sites very close (10–50 m) to the wineries (vineyards D and E). A major proportion (73%) was collected in post-harvest campaigns, indicating immediate dissemination. With the exception of the autochthonous ICV D254 strain collected in French winery B, commercial yeasts in pre-harvest campaigns were only collected at sites very close to winery D (10–50 m) and the strain found in the greatest quantity (87%) was Zymaflore F15. In the post-harvest samples, strain VL1 represented 49% of the commercial strains recovered. This strain was derived from sites close to the area where macerated grape skin was deposited or water run-off occurred, and never further than 10–20 m from the winery (Figs. 1 and 2). A lower percentage of the other predominant strains: Zymaflore F10 and F15, the formerly used minority strains Uvaferm BDX and ICV D254, and the autochthonous strain Lalvin QA23, were found at sites closer

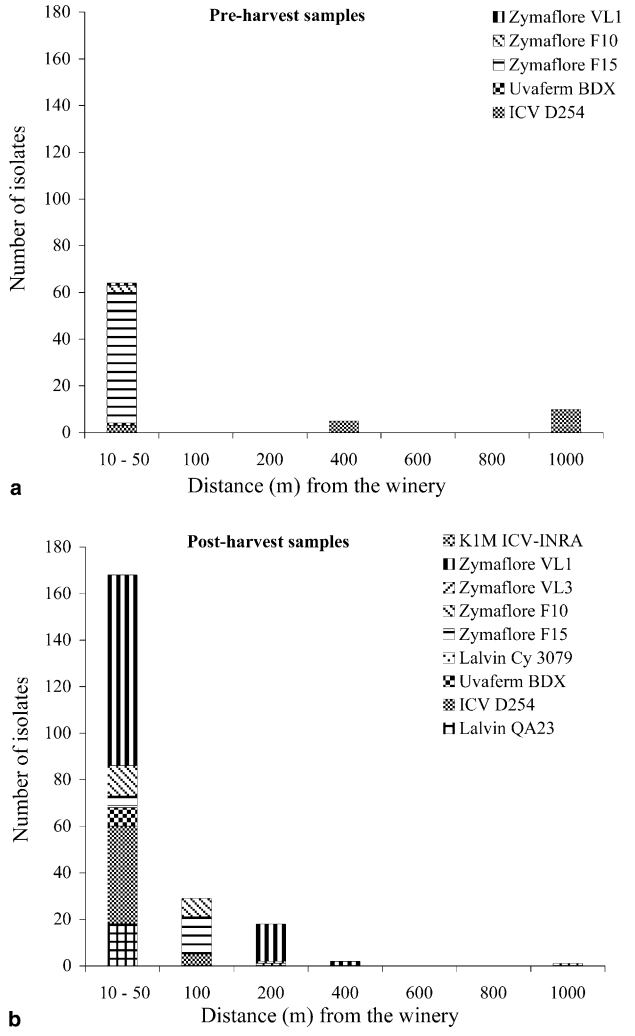


Fig. 4. Overall distribution during three years of commercial yeast strains according to the distance from the wineries in pre-harvest (a) and in post-harvest (b) campaigns.

to the winery (10–50 m). Zymaflore F15, F10 and ICV D254 were also found at about 100 m from the winery. The occurrence of several isolates found at 200 m (site 1, winery D) can be attributed to the presence of a small building for storage of harvest transport equipment. Two samples taken in France at a distance of 400 and 1000 m contained yeasts with a karyotype identical to that of the indigenous strain ICV D254. In very rare cases, dissemination to sites located further away from the winery (i.e., two isolates at 400 m and one at 1000 m from a total of 3780 strains) was revealed and may be attributable to other factors, such as insects or wind.

3.3. Dissemination of commercial strains as a function of their utilisation

As shown in Table 5, the 296 strains collected had a genetic profile identical to that of only nine commercial yeast strains from a total of 34 strains used in the six wineries. In most instances, the strains with a profile similar to that of a commercial strain were recovered from a vineyard in which the same commercial yeast was used, except for ICV D254 and Uvaferm BDX, which were collected in vineyard D and not used there during the study. However, these strains had been used previously (1998–2000) in the same vineyard. The other exception was strain Lalvin QA23, which was used in vineyard A and collected only from vineyard E. Since this strain had initially been isolated in this Portuguese region, the most likely explanation is that the strain isolated in vineyard E is not the result of dissemination, but that it is a member of the indigenous yeast community.

The industrial yeasts most commonly used in the wineries were usually collected in great abundance in the vineyard. However, this was not always the case, because the strain K1M ICV-INRA was the most widely used in the three French wineries and only one isolate out of 2160 isolates collected in France had a genotypic pattern identical to that of this strain. In Portuguese wineries, Zymaflore VL1 was predominantly and continuously used for more than 10 years, followed by Zymaflore F10 and VL3 (Table 1). The strains VL1 and F10 were frequently recovered, but this could be due to the fact that the sites where they had been collected were

Table 5
Origin of 296 strains with genetic patterns identical to commercial yeast strains used in the wineries

Commercial wine strains	Number of strains with identical genetic pattern	Wineries were these strains were used	Vineyard were these strains were collected	Utilization level during the 3 years
Zymaflore VL1	99	D, E, F	D, E, F	+++
Zymaflore F15	74	D	D	+
ICV D254	68	A, B, C, D[a]	B[b], D	++
Zymaflore F10	24	D	D	++
Lalvin QA23	19	A	E[b]	+
Uvaferm BDX	9	D[a]	D	–
K1M ICV-INRA	1	A, B, C	C	+++
Zymaflore VL3	1	C, D	D	+
Lalvin CY 3079	1	D, E, F	E	+

[a] Used before the study
[b] Isolated in the same region.

located close to the winery (i.e., vineyards D or E, Fig. 2). The strain Zymaflore F15, although frequently collected in the same vineyard D, was used to a lesser extent. Over the period of the study, Zymaflore VL3 was also widely used in Portuguese wineries, and the genetic profile of only one isolate was identical to this commercial yeast. As a whole, these data indicate that there is no strict correlation between the utilisation level and the frequency of dissemination.

3.4. Evolution of the fermentative yeast community

The evolution of the total yeast community isolated after fermentation in the different wineries of France and Portugal during the three years studied is shown in Fig. 5. From a total of 296 commercial yeasts recovered during this period in the six vineyards, 76% were found in 2001, in pre- and post-harvest samples collected in vineyard D and in post-harvest samples collected in vineyard E. In the following two years commercial yeasts were detected only in certain post-harvest but not in pre-harvest samples. As can be observed in Fig. 5, five different commercial yeast strains were found in the pre-harvest campaign of winery D in 2001, namely the predominantly used strains VL1, F10 and F15, and in much smaller quantities the strains Uvaferm BDX and ICV D254, used from 1998 to 2000, thus showing their survival in the vineyard from one

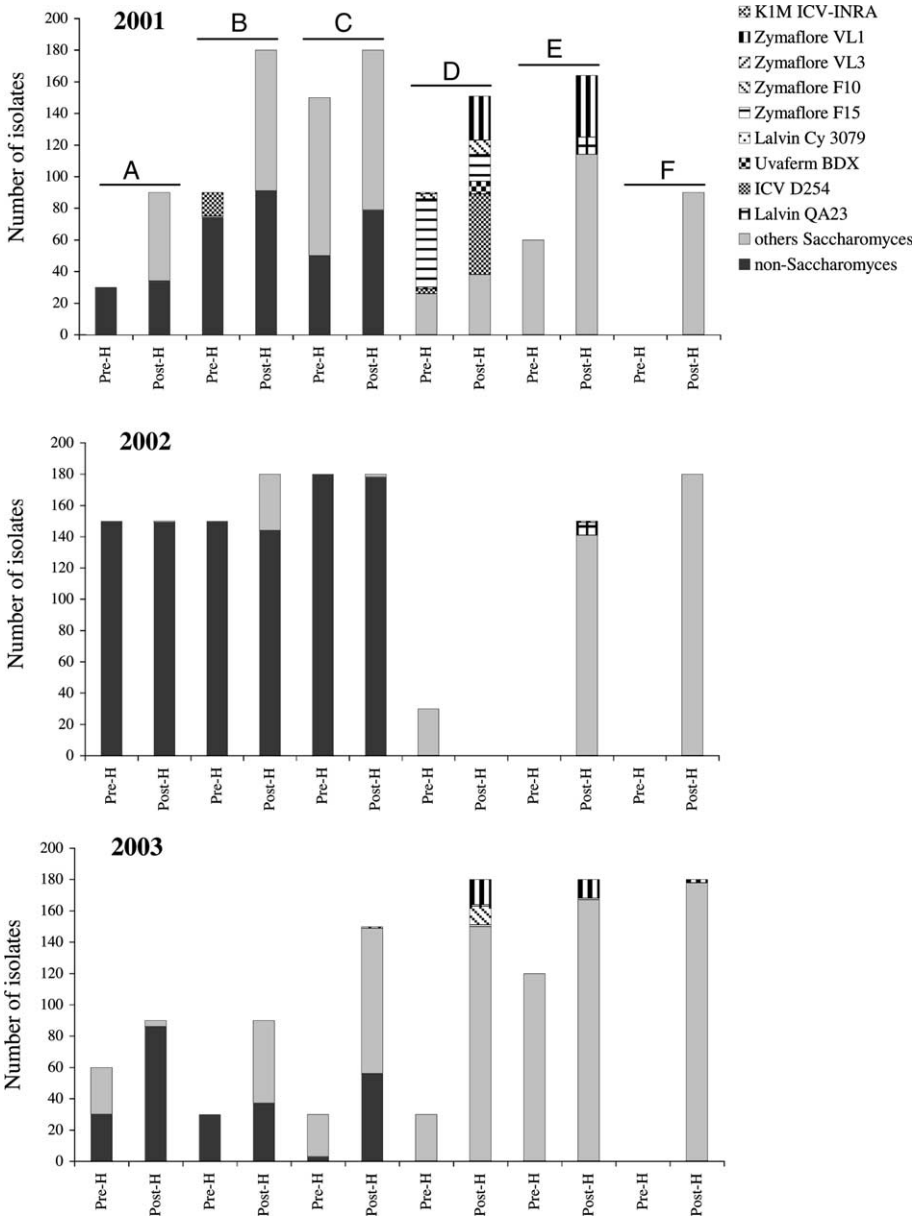


Fig. 5. Evolution of the total fermentative yeast communities from each of the wineries (A–F) during the three years in pre-harvest and in post-harvest campaigns.

year to another. However, given that the two latter strains appeared in 2001 only, their permanence was limited.

The commercial yeasts collected at each site were nevertheless different. The highest number of Zymaflore VL1 isolates was obtained from grapes collected after harvest at site 4, whereas in samples collected before harvest, VL1 and F10 rarely occurred. This contrasts to the abundance of the strain F15 in the pre-harvest campaign of 2001 at sites 4 and 6 (Fig. 2). These data suggest a better ability of strain Zymaflore F15 to remain in the vineyard, although no isolate of strain F15 was found in 2002 and only one in 2003 (Figs. 2 and 4). In this winery no samples were collected in the post-harvest campaign of 2002, and a lower quantity of commercial yeasts was found in 2003. In addition, the presence of one isolate of Zymaflore VL3, not present in 2001, was detected. In the post-harvest campaign of 2001, two commercial yeast strains, Zymaflore VL1 and the autochthonous yeast Lalvin QA23, were isolated in winery E. This last strain was the only commercial yeast found in the same winery in 2002, but it was not present in 2003. On the other hand, Zymaflore VL1 was not found in this winery in 2002, but was present in 2003, although in lower proportions. The situation observed in Portuguese winery F, as described previously, was similar to that in French wineries. No commercial yeasts were detected in 2001 and 2002, and only two isolates of Zymaflora VL1 were found in 2003. In winery B, the autochthonous strain ICV D254 was found in the pre-harvest campaign in 2001, but did not occur in the following years. Only one isolate of K1M ICV-INRA was found in 2003 in winery C. As a whole, the evolution of the fermentative yeast communities over the three years studied showed that the same strains were not found at the same sites from one year to another. This indicates that, if some of these strains are able to remain in the ecosystem, as suggested by the presence of commercial yeasts in pre-harvest samples taken in 2001 in Portugal, they are not capable of dominating the natural yeast community of the vineyard.

In conclusion, this systematic study has provided new insights in relation to the impact of commercial yeasts on the communities of fermentative yeasts that inhabit surrounding vineyards. The methodology used, based on analysis of the yeast community after spontaneous fermentation, permitted the isolation of a very large number of *Saccharomyces* wine yeasts, which are poorly found on the grapes. A significant number of non-*Saccharomyces* strains was also found in the spontaneous fermentations from the French samples, but not from the Portuguese grape musts (Table 3). Climatic factors and differences in phytosanitary treatment may be the cause of these discrepancies. In future studies, the occurrence of non-*Saccharomyces* strains during fermentation could be reduced by adding SO_2 to the grape musts prior to fermentation. It is important to mention that among the 30 colonies analyzed per fermentation, the number of different genetic profiles varied from 1 to 21, with an average of about five different *Saccharomyces* biotypes per sample ([22] and unpublished data). This reflects great differences in the samples regarding the presence of *Saccharomyces*. These data also demonstrate that the number of colonies analysed per sample was high enough to show the initial biodiversity. For future studies, increasing the initial amount of grapes collected may increase the number of spontaneous fermentations and therefore of *S. cerevisiae* strains isolated.

Data obtained in the present study show that dissemination of commercial yeasts in the vineyard is restricted to short distances and limited periods of time. More than 90% of commercial yeasts were found at a radius between 10 and 200 m from the winery and did not become implanted in the ecosystem in a systematic way. Dispersal of commercial strains seems to be mainly mediated by water run-off and may also derive from macerated grape skin at dumping sites. This situation was observed during the habitual functioning of a winery, where commercial strains are used without any containment. Avoiding grape skin deposition and canalisation of water run-off are low-cost measures, which can significantly reduce the population size of commercial yeast strains around the winery.

Given that they are used in large quantities, commercial strains tend to out-compete autochthonous strains inside the winery [21]. In contrast, they do not seem to settle in the vineyard. Rather, they show natural fluctuations of periodical appearance and disappearance, just like autochthonous strains do. Moreover, vine-associated autochthonous *Saccharomyces* biodiversity is not affected by long-term use of commercial yeasts [22]. Considering commercial yeast strains as an appropriate model system for genetically modified yeast strains, our data can contribute to the in-depth environmental-risk assessment concerning the use of such strains in the wine industry.

Acknowledgements

This study was supported by the project ENOSAFE (No. 762, Programa AGRO, medida 8.1) and the grant No. 657 C2 from the cooperation agreement between the Portuguese Institute for International Scientific and Technological Cooperation (ICCTI) and the French Embassy in Lisbon and the Marie Curie Fellowship of the European Community programme of Quality of Life under Contract QLK4-CT-2001-51873. The authors wish to thank A. Ribeiro, H. Alves, L. Quintas, C. Camarasa and C. Machado for their support during grape collection and sample processing. M.S. Graça is gratefully acknowledged for the operation of the DNA

sequencer. We would like to express our sincere gratitude to D. Delteil from ICV-Montpellier for his help in the selection of sample sites in France. We also appreciate the kind assistance of the enologists of the Portuguese wineries (R. Cunha, A. Mendes, E. Rodrigues and J. Domingues) as well as the enologists (B. Agay, E. Feneuil and E. Bru) and directors the (J.L. Refle, J. Combette and R. Bruno) of the French wineries, facilitating sampling campaigns in the vineyards.

References

[1] Fleet, G.H. and Heard, G.M. (1993) Yeast-growth during fermentation. In: Wine Microbiology and Biotechnology (Fleet, G.H., Ed.), pp. 27–54. Harwood Academic Publishers, Singapore.
[2] Dequin, S., Salmon, J.M., Nguyen, H.V. and Blondin, B. (2003) Wine yeasts. In: Yeasts in Food, Beneficial and Detrimental Aspects (Boekhout, T. and Robert, V., Eds.), pp. 389–412. B. Berhr's Verlag GmbH and Co,, Hamburg, Germany.
[3] Butzke, C.E. and Bisson, L.F. (1996) Genetic engineering of yeast for wine production. Agro Food Industry Hi-Tech. 7, 26–30.
[4] Dequin, S. (2001) The potential of genetic engineering for improving brewing, wine-making and baking yeasts. Appl. Microbiol. Biotechnol. 56, 577–588.
[5] Henschke, P.A. (1997) Wine yeast. In: Yeast Sugar Metabolism (Zimmermann, F.K. and Entian, K.D., Eds.), pp. 527–560. Technomic Publishing Co., Lancaster, PA.
[6] Pretorius, I.S. (2000) Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. Yeast 16, 675–729.
[7] Pretorius, I.S. and Bauer, F.F. (2002) Meeting the consumer challenge through genetically customized wine-yeast strains. Trends Biotechnol. 20, 426–432.
[8] Querol, A. and Ramon, D. (1996) The application of molecular techniques in wine microbiology. Trends Food Sci. Technol. 7, 73–78.
[9] Frezier, V. and Dubourdieu, D. (1992) Ecology of yeast strain *Saccharomyces cerevisiae* during spontaneous fermentation in a Bordeaux winery. Am. J. Enol. Vitic. 43, 375–380.
[10] Guillamón, J.M., Barrio, E. and Querol, A. (1996) Characterization of wine yeast strains of the *Saccharomyces* genus on the basis of molecular markers; relationships between genetic distance and geographic or ecological origin. System Appl. Microbiol. 19, 122–132.
[11] Vezinhet, F., Hallet, J.N., Valade, M. and Poulard, A. (1992) Ecological survey of wine yeast strains by molecular methods of identification. Am. J. Enol. Vitic. 43, 83–86.
[12] Van der Westhuizen, T.J., Augustyn, O.P.H. and Pretorius, I.S. (2000) Geographical distribution of indigenous *Saccharomyces cerevisiae* strains isolated from vineyards in the coastal regions of the Western Cape in South Africa. S. Afr. J. Enol. Vitic. 21, 3–9.
[13] Van der Westhuizen, T.J., Augustyn, O.P.H., Khan, W. and Pretorius, I.S. (2000) Seasonal variation of indigenous *Saccharomyces cerevisiae* strains isolated from vineyards of the Western Cape in South Africa. S. Afr. J. Enol. Vitic. 21, 10–16.
[14] Barnett, J.A., Payne, R.W. and Yarrow, D. (1990) Yeasts: Characteristics and Identification, second ed. Cambridge University Press, Cambridge, UK.
[15] López, V., Querol, A., Ramón, D. and Fernández-Espinar, M.T. (2001) A simplified procedure to analyse mitochondrial DNA from industrial yeasts. Int. J. Food Microbiol. 68, 75–81.
[16] Querol, A., Barrio, E. and Ramon, D. (1992) A comparative study of different methods of yeast strain characterization. Syst. Appl. Microbiol. 15, 439–446.
[17] Schuller, D., Valero, E., Dequin, S. and Casal, M. (2004) Survey of molecular methods for the typing of wine yeast strains. FEMS Microbiol. Lett. 231, 19–26.
[18] Pérez, M.A., Gallego, F.J., Martinez, I. and Hidalgo, P. (2001) Detection, distribution and selection of microsatellites (SSRs) in the genome of the yeast *Saccharomyces cerevisiae* as molecular markers. Lett. Appl. Microbiol. 33, 461–466.
[19] Blondin, B. and Vezinhet, F. (1988) Identification de souches de levures oenologiques par leurs caryotypes obtenus en électrophorèse en champs pulsées. Rev. Fr. Oenol. 28, 7–11.
[20] Granchi, L., Bosco, M., Messini, A. and Vicenzini, M. (1999) Rapid detection and quantification of yeast species during spontaneous wine fermentation by PCR-RFLP analysis of the rDNA ITS region. J. Appl. Microbiol. 87, 949–956.
[21] Beltran, G., Torija, M.J., Novo, M., Ferrer, N., Poblet, M., Guillamón, J.M., Rozes, N. and Mas, A. (2002) Analysis of yeast populations during alcoholic fermentation: A six year follow-up study. Syst. Appl. Microbiol. 25, 287–293.
[22] Schuller, D., Alves, H., Dequin, S. and Casal, M. (2005) Ecological survey of *Saccharomyces* strains from vineyards in the Vinho Verde Region of Portugal. FEMS Microbiol. Ecol. 51, 167–177.

Biomolecular Detection and Quantification 7 (2016) 27–33

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

journal homepage: www.elsevier.com/locate/bdq

Research paper

Evaluation of microbial qPCR workflows using engineered *Saccharomyces cerevisiae*

S.M. Da Silva[a,*,1], L.K. Vang[a,1], N.D. Olson[a,1], S.P. Lund[b], A.S. Downey[a,2], Z. Kelman[c,d], M.L. Salit[a,e], N.J. Lin[a,*], J.B. Morrow[a]

[a] *Biosystems and Biomaterials Division, National Institute of Standards and Technology, Gaithersburg, MD, USA*
[b] *Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, USA*
[c] *Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, MD, USA*
[d] *Institute for Bioscience and Biotechnology Research, Rockville, MD, USA*
[e] *Department of Bioengineering, Stanford University, Stanford, CA, USA*

ARTICLE INFO

Article history:
Received 13 August 2015
Received in revised form 15 January 2016
Accepted 25 January 2016
Available online 19 February 2016

Keywords:
ERCC
Evaluation material
Interlaboratory study
Saccharomyces cerevisiae
qPCR
Reference material

ABSTRACT

Aims: We describe the development and interlaboratory study of modified *Saccharomyces cerevisiae* as a candidate material to evaluate a full detection workflow including DNA extraction and quantitative polymerase chain reaction (qPCR).
Methods and results: S. cerevisiae NE095 was prepared by stable insertion of DNA sequence External RNA Control Consortium-00095 into *S. cerevisiae* BY4739 to convey selectivity. For the interlaboratory study, a binomial regression model was used to select three cell concentrations, high (4×10^7 cells ml^{-1}), intermediate (4×10^5 cells ml^{-1}) and low (4×10^3 cells ml^{-1}), and the number of samples per concentration. Seven participants, including potential end users, had combined rates of positive qPCR detection (quantification cycle <37) of 100%, 40%, and 0% for high, intermediate, and low concentrations, respectively.
Conclusions: The NE095 strain was successfully detected by all participants, with the high concentration indicating a potential target concentration for a reference material.
Significance and impact of the study: The engineered yeast has potential to support measurement assurance for the analytical process of qPCR, encompassing the method, equipment, and operator, to increase confidence in results and better inform decision-making in areas of applied microbiology. This material can also support process assessment for other DNA-based detection technologies.
Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Data of high quality is essential in microbial detection, identification, and quantification because of the impact these organisms (pathogenic and beneficial) have on human life in areas including environmental monitoring, food safety, biothreat detection, and clinical outbreaks [1–3]. Despite the importance, analysis of microbial samples remains a practical and technological challenge when it comes to confidence in the measurements, especially for measurements made at the point of need or point of care where results are used to inform critical decision-making. In the biodefense field, for example, there are over 300 technologies marketed for use in biological detection claiming the ability to detect pathogens (biothreats) in suspicious materials [4–6]. However, there remains a scarcity of standards, reference materials, and third-party testing to demonstrate the reliability of these technologies in the hands of end users, despite considerable efforts by the stakeholder community and Federal government. Only a few commercially available biodetection technologies have been submitted to third-party validation, including the Razor^{TM} EX BioDetection System [7], a qPCR-based assay.

qPCR is a well-established technique that provides selectivity and sensitivity in detecting nucleic acid markers [8,9]. However, obtaining reliable data can be challenging because of factors that compromise nucleic acid amplification such as residue from the crude sample (matrix effect) [10]. In addition, measurement inaccuracy as a result of sample collection, processing, and nucleic acid extraction is often observed [10–13]. The absence of method validation brings into question the reliability of the generated data

* Corresponding authors at: National Institute of Standards and Technology, 100 Bureau Dr. Gaithersburg, MD 20899-8543, USA.
E-mail addresses: sdasilva@nist.gov (S.M. Da Silva), nancy.lin@nist.gov (N.J. Lin).
[1] These authors contributed equally to this work.
[2] Current affiliation: Institute of Medicine, National Academy of Sciences, Washington, DC, USA.

http://dx.doi.org/10.1016/j.bdq.2016.01.001
2214-7535/Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

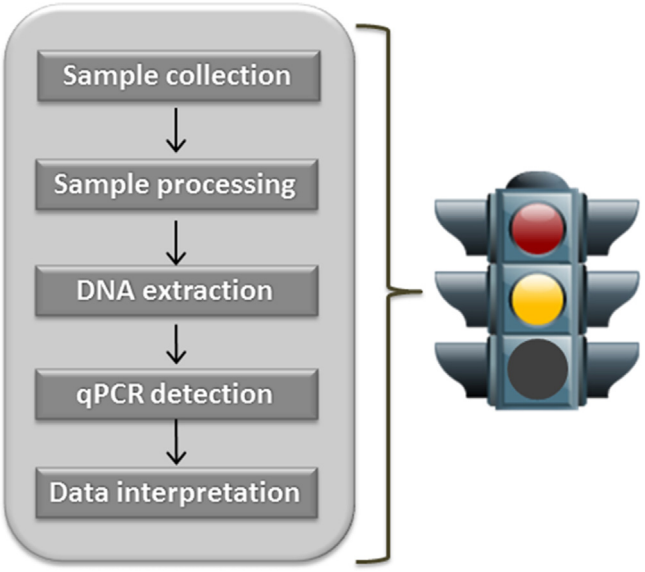


Fig. 1. Workflow steps for a downstream qPCR sample analysis. Analogous to a traffic light, the unsuccessful detection of a reference material provides a red light indicating the analytical process is unsuccessful and potentially helping identify the source of the problem. Successful detection of the material provides a yellow light indicating that the analytical process is working properly, that is the methods are appropriate, the equipment is functioning, and the user is proficient in the required skills. There is no green light since the material cannot be used to validate a specific detection assay for an organism of interest.

[14]. Further, even with the use of a successful detection assay, such as the Biothreat Panel multiplexed PCR-based assay for the detection of *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis* or the rapid, high-throughput, culture-based PCR methods to analyze samples for viable spores of *B. anthracis* and its surrogates [15], the final result in the field still depends upon the entire analytical process. The process includes the methods and protocols, the measurement workflow which encompasses all steps applied to obtain the final result, the performance of the equipment at the point of need, operator capabilities and skills, and proficiency testing. Without all of these components in place, results cannot be used with confidence to support decision-making.

Control materials can serve to evaluate or validate analytical processes, establish reliable and comparable analytical results among laboratories or analysts through proficiency testing or competence assessment, and verify accuracy of measurement performance on a daily basis [16]. For example, human DNA Standard Reference Materials (SRMs) are used in the forensics community to reduce variability within and among laboratories [17,18]. Control materials are typically thoroughly characterized using measurement methods with well understood biases and variability. Measurement controls such as reference materials can help provide confidence in the application of qPCR to microbial measurements, with the format of the material enabling performance evaluation at various workflow steps (Fig. 1). For instance, a reference material can be mixed into a matrix of interest to demonstrate a successful DNA extraction step, which is highly susceptible to matrix effects. The use of a reference material to demonstrate a successful qPCR analytical process can increase confidence when the measurement capability is applied in a real-case scenario, such as in a clinical setting, environmental monitoring, or biothreat scenario.

One promising organism to meet these needs for a qPCR reference material is *Saccharomyces cerevisiae* (*S. cerevisiae*), a eukaryotic model system widely used in biology fields such as bioengineering [19]. Yeast can be genetically modified to enable specific detection and offers a low DNA extraction efficiency to challenge extraction protocols [11]. Its physiological resilience under low nutrient conditions and stability under various environmental conditions make yeast suitable for different formats, including a liquid or powder, for broader applicability. Moreover, *S. cerevisiae* has minimal health and security risks and can be handled without special precautions or training. It can therefore serve as a surrogate material for routine training and process evaluation in applications where the true agent of interest is a pathogen or threat agent and not easily used. Further, use of the yeast, and not the agent of interest, can essentially eliminate false positives during real microbial detection situations due to residual material on equipment.

The presence of a non-native target DNA sequence in the yeast genome can eliminate false positives from near-neighbor organisms. For specificity, the target sequence should be rare and not typically found in the environment of interest. NIST SRM 2374 contains a series of nucleic acid sequences selected by the External RNA Control Consortium (ERCC) as control sequences that are rarely (if at all) found in normal environmental conditions (temperature and pressure). One such sequence is ERCC-00095, derived from *Methanocaldococcus jannaschii*, a deep-sea vent archaeon found only in extremely harsh conditions (an extremophile).

The objective of this study was to develop a *S. cerevisiae* strain containing a non-yeast target DNA insert and evaluate the strain via interlaboratory study as a potential material for assessing the qPCR analytical process, in efforts toward a reference material for qPCR. Yeast cells were transformed by inserting the ERCC-00095 DNA sequence into the yeast genome. The modified yeast strain, termed NE095, was prepared at three different cell concentrations and evaluated using qPCR in an interlaboratory pilot study involving five public health laboratories, one mobile laboratory and one in-house laboratory. The NE095 was detectable at the expected concentrations in multiple laboratories and is suitable for a reference material.

2. Materials and methods

2.1. *Preparation and characterization of engineered* S. cerevisiae

2.1.1. *cerevisiae transformation*

The parent organism was a URA3 deficient yeast strain: *Saccharomyces cerevisiae* BY4739 (MATalpha leu2Δ0 lys2Δ0 ura3Δ0) (procured from Open Biosystems) derived from *S. cerevisiae* S288C. The URA3 gene encodes for orotidine-5′-phosphate decarboxylase from *S. cerevisiae* S288C. A target DNA sequence and full-length URA3 gene were inserted via homologous recombination into chromosome IV of the yeast (Fig. 2, Supplemental material: *S. cerevisiae* NE095 transformation, Figs. S1 and S2). The DNA insert was prepared by ligating the target sequence, ERCC-00095 (from NIST SRM 2374, Genbank accession KC702204, without the polyA tail found in the SRM), to the URA3 gene by overlapping PCR. The full-length URA3 gene and URA3 promoter sequence were PCR amplified from a pYES2 vector (Life Technologies, part # V825-20). See Supplemental material for detailed methods describing the transformation process.

2.1.2. *NE095 characterization*

The presence of the insert was demonstrated by PCR, and the sequence of the amplified insert, including the ERCC-00095 and URA3 gene, was confirmed by Sanger sequencing (analyzed by Eurofins MWG Operon, Alabama, USA) (Supplemental material: NE095 insert sequence confirmation, Figs. S3, S4, and S5). Once the insert was demonstrated, the yeast cells were cultured in SD/-Ura broth for subsequent experiments.

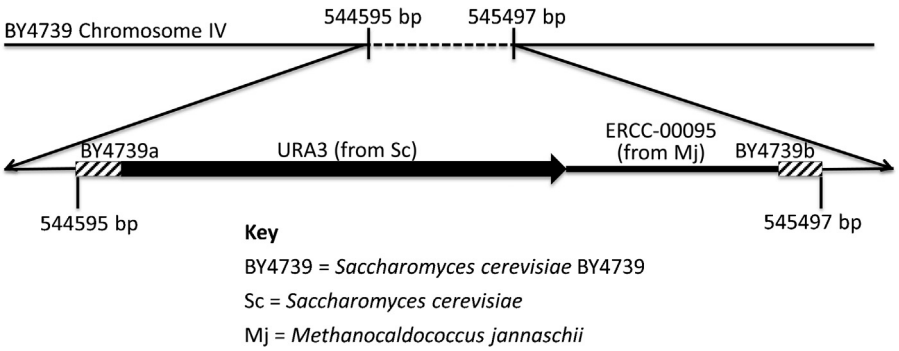


Fig. 2. Map of final linear construct inserted into BY4739 chromosome IV. BY4739a is a 50 bp region homologous to Open Biosystems strain *S. cerevisiae* BY4739 and corresponding to the non-coding region on chromosome IV between 544595 bp and 544644 bp. The full annotation of chromosome IV can be found at http://browse.yeastgenome.org/fgb2/gbrowse/scgenome/. URA3 is a 1106 bp sequence that encodes *S. cerevisiae* S288C orotidine-5′-phosphate decarboxylase, an enzyme that catalyzes the synthesis of pyrimidine ribonucleotides. The sequence was PCR amplified from a pYES2 vector (Life Technologies, part # V825-20) and includes a working copy of the 804 bp URA3 gene (Gene ID: 856692) as well as the URA3 promoter sequence. ERCC-00095 represents a 438 bp sequence from NIST SRM 2374: DNA Sequence Library for External RNA Controls. ERCC-00095 (GenBank accession number KC702204.1) is derived from donor microorganism *M. jannaschii*. The sequence was amplified from the ERCC-00095 plasmid using primers 95F and 95R (Table S1). BY4739b is 51 bp homologous to the non-coding region on chromosome IV between 545447 bp and 545497 bp in Open Biosystems strain *S. cerevisiae* BY4739.

2.1.3. NE095 growth conditions and quantification via microscopy

NE095 cells were cultured in SD/-Ura broth overnight. Each culture was generated from one NE095 colony in 5 ml of broth (30 °C, ~16 h,) followed by a 25-fold dilution in broth and subsequent growth (30 °C, ~8 h). The number of replicate yeast cultures used for each set of experiments was as follows: binomial regression analysis ($n = 8$), stability study ($n = 5$) and interlaboratory study ($n = 1$). Samples for the interlaboratory study were collected from a single culture to assure that all participants obtained cells from the same batch. Variability in number of cells between batches under the selected experimental conditions was determined by cell counting while live/dead staining experiments were applied to estimate cell viability under different storage condition (see Supplemental material: *S. cerevisiae* NE095 material stability, Figs. S8, S9, and S10). Cells were counted using a cellometer (disposable hemocytometer) following the manufacturer's instructions (Nexcelom, USA). Cellometers were loaded with yeast and imaged, and counts were performed using Image J software (version 1.47 v) [20]. Throughout this work, data are represented as the mean value ± one standard deviation, where the standard deviation serves as the estimate for the standard uncertainty.

2.1.4. DNA extraction and qPCR detection assay

The following protocol was used to analyze the genomic DNA (gDNA) for the yeast stability study and interlaboratory study, and to generate the regression model. gDNA from NE095 cells was extracted using 200 μl (800 cells, 8×10^4 cells, or 8×10^6 cells for low, intermediate, and high concentrations, respectively) aliquots of cells and a QIAamp DNA Blood Mini Kit (cat# 51104 QIAGEN, USA). The spin protocol for DNA purification from blood or body fluids was followed according to the manufacturer's directions with a few modifications. DNA was incubated at 56 °C for 1 h following the addition of protease and lysis buffer. Extracted DNA was eluted in 200 μl elution buffer provided in the DNA extraction kit and stored at 4 °C prior to downstream analysis.

qPCR analysis was performed on a 7900HT PCR system (Life Technologies, Carlsbad, CA), except in the interlaboratory study where participants used their own qPCR technology. A TaqMan FAM-MGB primer and probe set was custom designed and synthesized by Life Technologies (Life Technologies, Carlsbad, CA) to amplify a 105 bp region of the ERCC-00095 sequence (Figs. S6 and S7). The reaction mixture contained 10 μl of Environmental MasterMix (Life Technologies, Carlsbad, CA), 4 μl of extracted genomic DNA, 1 μl of amplification primers and probe with the following sequences: CAGTCATCTTTAACCTCATCCCACAA (forward primer), CATTTGGCCCAAGAATTCATGGAAT (reverse primer), and CCCACTCAACAATCTT (probe), and 5 μl of sterilized water. Reaction conditions were: 5 min at 50 °C, 10 min at 95 °C for Taq polymerase activation, and 50 cycles of 15 s at 95 °C, 1 min at 60 °C for DNA amplification. A gDNA standard curve was prepared via 10-fold serial dilutions of the transformed yeast genomic DNA. The initial concentration was estimated by Nanodrop (Thermo Scientific, USA) and Qubit (Invitrogen, USA). The efficiency of amplification (E) for the transformed yeast using this customized qPCR assay (primers and probes) was 98.4% (Fig. S7). All chemicals were purchased from Fisher Scientific unless otherwise stated.

For the qPCR assay, primer and probe specificity was verified using Primer BLAST (results available at http://dx.doi.org/10.6084/m9.figshare.875419). Three primer combinations were compared to the GenBank nr database using Primer BLAST: forward-reverse, forward-probe, and probe-reverse. The only matches to the database for the primer-probe set were to *M. jannaschii* DSM 2661 (GenBank Accession: CP001781.1) and *Methanocaldococcus vulcanius* M7 complete genomes (L77117.1) as well as the ERCC-00095 (KC702204.1) and *M. jannaschii* spike-in control MJ-500-42 (DQ516759.1). The probe and reverse primer sequences had no mismatches to the *Methanocaldococcus* sequences or ERCC-00095, whereas the forward primer had two mismatches at the 5′ end (positions 1 and 3) to the *M. vulcanius* M7 genome, and three mismatches at the 5′ end (positions 1–3) for the *M. jannaschii* spike-in control MJ-500-42, ERCC-00095, and *M. jannaschii* DSM 2661 genome.

2.2. Interlaboratory study

2.2.1. Experimental design and statistical analysis approach

A logistic regression model was used to develop the experimental design for the interlaboratory study [21], specifically the cell concentrations and number of samples per concentration to include in the sample panel. To generate the regression model, four independent cultures were grown on two separate days with an average cell concentration of $(6.3 \pm 0.8) \times 10^7$ cells ml^{-1} ($n = 8$). A ten-fold cell dilution series was generated for each culture with cell concentrations ranging from 10^3 cells ml^{-1} to 10^7 cells ml^{-1}. gDNA extracted from the cells was analyzed using the qPCR protocol described above. Samples were scored as positive if the quantification cycle (C_q) for detecting the ERCC-00095 insert was less than

37 for at least one of the triplicate reactions. The same C_q value criterion was applied to evaluate the raw data from the interlaboratory study. The qPCR data were analyzed using the statistical programming language R [22]. Source code and raw data for the binomial regression can be found at the following location: (http://dx.doi.org/10.6084/m9.figshare.875419).

2.2.2. Participants

Participants consisted of five public health laboratories, one mobile laboratory, and one in-house laboratory; six of these participants are laboratories currently active in the biosurveillance community. Each participant received the yeast sample panel, DNA extraction kit, primers and probes, and a detailed protocol (Supplemental material: Interlaboratory SOP). Participants used either the Applied Biosystems Instruments (ABI) 7500 or the Joint Biological Agent Identification and Diagnostic System (JBAIDS) platform for qPCR. To assure an even execution of the experiments, participants reported to the organizer laboratory the arrival date and condition of the sample panel, the platform and software used to process the samples and collect data, and any deviation from the recommended protocol. All participants submitted their results to the organizer laboratory within one week.

2.2.3. Yeast sample panel

Based upon the logistic regression model results, each sample panel was composed of ten randomized blinded samples: a 5% ethanol no template-control (blank), and nine yeast samples as follows: low (4.0×10^3 cells ml^{-1}, $n = 2$), intermediate (4.0×10^5 cells ml^{-1}, $n = 5$), and high (4.0×10^7 cells ml^{-1}, $n = 2$). All samples in the panel came from a single overnight cell culture batch and were quantified by cell counting using a cellometer. Samples were shipped in 5% ethanol, a condition selected to maintain cell and DNA integrity for short term (2 weeks), liquid storage at room temperature. Stability of cells under the conditions used in the interlaboratory study was demonstrated by live/dead staining and qPCR as detailed in supporting information.

2.2.4. Statistical analysis for intermediate samples

The experimental design generated 35 intermediate concentration samples, which were prepared and distributed among seven labs (five samples to each lab; samples are nested within lab). In most cases, three replicate C_q measurements were taken for each sample. Measurements with undetermined runs (no detection at all) were modeled as being right-censored at 40 (i.e., the corresponding observed C_q was assumed to be greater than 40, but otherwise unknown). Let $C_{q_{ijk}}$ denote the natural log of the C_q value observed in replicate measurement k from sample j at lab$i$. We analyzed the observed data with the following random effects or hierarchical model:

$$\ln(C_{q_{ijk}}) = \mu + \alpha_i + \beta_j \varepsilon_{ijk}$$

where μ denotes a global average $\ln(C_q)$ across all labs and samples, $\alpha_i \sim \text{Normal}\left(0, \sigma^2_{\text{lab}}\right)$ is a random effect for lab i ($i = 1, \ldots, 5, 7$), $\beta_j \sim \text{Normal}\left(0, \sigma^2_{\text{sample}}\right)$ is a random effect for sample j ($j = 1, \ldots, 25, 31, \ldots, 35$), and $\varepsilon_{ijk} \sim \text{Normal}\left(0, \sigma^2_{\text{error}}\right)$ is a random error corresponding to replicate measurement k from sample j at lab i. Each variance component is assigned a prior distribution such that σ_{lab}, σ_{sample} and σ_{error} follow a uniform distribution between 0 and 0.5. The overall mean μ was assigned a uniform prior distribution from 3 to 3.9, corresponding to a range of mean C_q values given by (20.1, 49.4). The posterior distributions of the model parameters μ, σ_{lab}, σ_{sample} and σ_{error} given the observed data were evaluated using Markov chain Monte Carlo simulation via the R package *rjags*. The simulation included 5 chains, each recording 10,000 observations following a 50,000 iteration burn in.

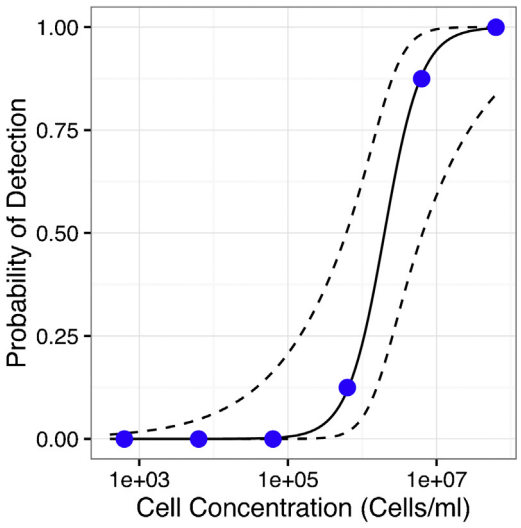


Fig. 3. Logistic regression model. The model was generated based on the number of positive qPCR reactions (at least 1 of 3 reactions with a C_q value <37) using a cell dilution series from 8 independent cultures. The filled circles represent the experimental data. The solid and dashed lines indicate the regression model and the 95% pointwise confidence interval, respectively.

3. Results

3.1. NE095 characterization

The DNA insert was successfully prepared and incorporated into chromosome IV to produce the NE095 strain. NE095 has a functional URA3 gene and otherwise is not anticipated to have any other changes in phenotype or traits. The modified yeast strain grows under typical conditions in uracil-deficient medium, as expected. Sequencing confirmed the presence of the expected 1,645 bp insert, including the functional URA3 gene and the ERCC-00095 sequence (Fig. S5). In addition, the insert was successfully amplified by the selected qPCR primers and probes. An alignment of the qPCR primers with the de-novo assembly of the insert sequence revealed insertions 4 bp from the 5′ end of the forward and reverse primers (Fig. S6). Additionally, positions 1–3 from the 5′ end of the forward primer that were in disagreement with the GenBank sequences, due to the previously identified insertions, as identified with Primer BLAST agree with the Sanger sequence data. Primer mismatches can decrease assay efficiency; however the mismatches are at the 5′ end and less likely to impact the qPCR assay efficiency [23]. Even with the mismatches the qPCR assay efficiency was over 98%, which is suitable for quantitative assays.

The stability of the NE095 was monitored at room temperature in three different storage conditions, where it was determined that 5% ethanol was a suitable storage condition for the time frame required in the interlaboratory study (see supporting information). In 5% ethanol, no change in cell concentration was observed after two weeks of storage ($p > 0.05$) (Table S4). The stability of the yeast (membrane integrity) in the medium was also confirmed by live/dead staining while the integrity of the DNA was confirmed using qPCR (Figs. S8, S9, S10).

3.2. Experimental design for the interlaboratory study

The appropriate sample concentrations for the interlaboratory study were determined by fitting the qPCR results from a dilution series of yeast cells to a binomial regression model (Fig. 3)

The probability of detection for a sample with 10^7 cells ml^{-1} was 99.7% with a 95% confidence interval of 83.6–99.9%. The probability of detection for a sample with a concentration of 10^3 cells ml^{-1} was 4.2×10^{-3}% with a 95% confidence interval of 3.6×10^{-6}% to 4.7%. Based on these probabilities of detection, 10^3 cells ml^{-1} and 10^7 cells ml^{-1} were chosen as the negative and positive control samples, respectively. A concentration of 1.9×10^6 cells ml^{-1} had a 50% probability of detection with a 95% confidence interval for detection of 18–81% based on the generated regression model. However, because of an observed increase in DNA recovery with storage time (Fig. S10), a lower sample concentration (10^5 cells ml^{-1}) was chosen as the intermediate value to represent samples near the limit of detection. That concentration had a 0.0052% probability of detection with a 95% confidence interval of 0.0001–21%. Five samples were distributed to the participants for the intermediate concentration, compared to 2 samples for the high and low concentration samples, to increase the statistical power to assess probability of detection at the intermediate concentration.

3.3. Interlaboratory study

A total of 190 qPCR data points were collected from the interlaboratory study (Fig. 4). Laboratory 4 ran single qPCR reactions instead of the triplicate reactions requested in the provided protocol. Despite the single run, the data were otherwise suitable to be included in the analysis. Laboratory 6 extracted DNA from 400 μl of cells, as opposed to 200 μl as described in the protocol, and obtained the lowest quantification cycle (C_q) values for the high concentration samples. Thus, data from laboratory 6 were excluded from further analysis because of deviations from the provided protocol. Potential contamination was monitored by blank samples (data not shown) and no detection was observed as expected. High concentration (10^7 cells ml^{-1}) samples were positive (detection) across all laboratories while low samples (10^3 cells ml^{-1}) scored negative since any detectable C_q values (43.1 and 38.5) were above the threshold of 37. For intermediate samples (10^5 cells ml^{-1}), 40% of the samples (12 out of 30) were scored as positive, suggesting that this concentration is approaching the yeast concentration required to produce a positive qPCR signal in essentially all measurements.

According to the posterior distributions of their respective variance parameters, the probability that the variability due to differences between labs exceeds the variability due to differences between samples is 80%. That is, in 80% of the simulated iterations the standard deviation parameter for the labs, σ_{lab}, had a larger value than the standard deviation parameter for the samples, σ_{sample}. The posterior median of $\sigma_{lab}/\sigma_{sample}$ is 1.54, with a 95% posterior credible interval given by (0.50, 5.44) (Fig. S11).

4. Discussion

In this study we aimed to demonstrate the use of yeast as a candidate reference material in a workflow composed of various steps that are not necessarily measurable individually, and we envision the yeast having broader application, such as to assess performance parameters related to quantitative methods. We successfully developed *S. cerevisiae* NE095, a yeast strain with a stable DNA insertion for qPCR measurements, and demonstrated its detection in the hands of potential users via interlaboratory study. This work demonstrates the potential of the yeast material to be further developed into a reference material to evaluate processes associated with nucleic acid-based detection technologies. The insert incorporated into the yeast genome attributes specificity by eliminating the risk of false positive amplification from environmental background, near-neighbor organisms, or contaminants. Because the yeast is a versatile organism that can be manipulated in a broad range of environments, the authors foresee its use as a reference material in multiple areas of microbial detection including environmental monitoring, food safety, biothreat detection, and clinical outbreaks.

The modified yeast strain was developed to evaluate analytical processes used for microbial measurements and to instill confidence in measurement capabilities, particularly for applications where critical decision-making depends upon the measurement results. NE095 can challenge nucleic acid detection technologies and processes through using a pre-determined amount of cells in a typical qPCR workflow (e.g., sample collection, DNA extraction, qPCR with a custom assay for the target DNA sequence) with the expectation of positive detection. Qualitative assessment would provide a yes/no answer to detection, whereas lack of detection may indicate error in sample processing or defective equipment. When used for quantitative assessment, NE095 can evaluate the efficiency of the analytical process, track changes in efficiency over time, or determine the effects of modifications in the process or operator skills on efficiency. Additionally, NE095 can be used as a confidence checker (to assure correct equipment function), in proficiency testing, training, and other applications. It is noteworthy to mention that NE095 is not designed to validate detection assays/technologies for a specific organism of interest (e.g., pathogen or biothreat) but rather the analytical process in place to assure that it is working properly.

The probability of detection at a given concentration was modeled using binomial regression with a logistic link function (Fig. 3). Binomial regression has been used traditionally in toxicology studies [24] but has gained application in microbiology in the last 10 years. For instance, Janse et al. [25] used binomial regression analysis to support the development of a rapid qPCR assay to distinguish *Burkholderia mallei* and *Burkholderia pseudomallei* strains, two closely related, highly virulent bacteria. Results from the regression model were used to select the concentration levels, including a positive control (high concentration), negative control (low concentration), and intermediate concentration samples, for the interlaboratory study. We believe the logistic regression model was adequate for this study, where the main goal was to determine the probability of detection at various concentration levels, e.g., probit [26].

The interlaboratory study revealed that all participants succeeded in detecting the yeast as predicted by the organizer laboratory. Laboratories 4 and 5 observed C_q's for samples at the low concentration level (10^3 cells ml^{-1}); however those values did not meet the established criteria for positive scoring ($C_q < 37$) and were therefore scored as negative in agreement with predictions. Although a common cycle threshold for qPCR analysis is 40, a threshold of 37 was selected to reduce the chance of false positives due to low levels of contamination or instrument noise. Statistical analysis also showed that variability between samples (within labs) is about the same scale as the variability between repeated measurements made on a single sample (within sample) (Fig. S11). The estimated variability among laboratories is larger than the estimated variability between samples and between repeated measurements for 80% of the iterations. Variations in Cq values observed across all laboratories might be due to differences in arrival date, sample processing date, calibration of the qPCR instrument, adherence to the distributed protocol, or other factors.

This work is the initial phase of a larger project with the end goal of providing a material for process evaluation in qPCR and other nucleic acid-based techniques as a means to improve confidence in the analytical process and ultimately the measurement results. This study established cell concentrations that can be expected to produce a positive or negative qPCR result using this protocol and gives insight into an appropriate cell concentration for the development of a reference material. The next step is to preserve the

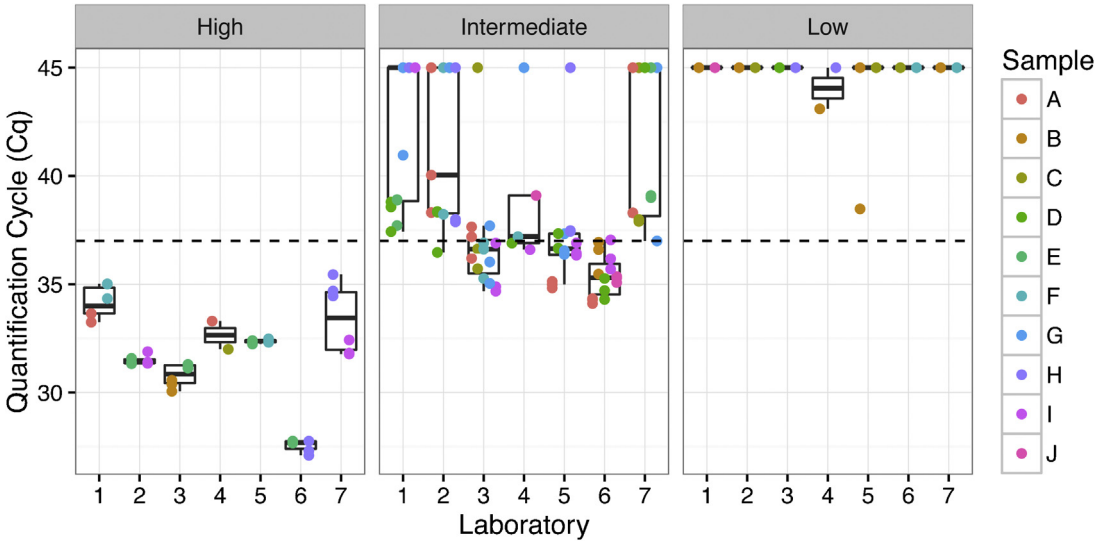


Fig. 4. Boxplots describing the quantification cycle (C_q) as a function of laboratory for the 10 samples analyzed in the interlaboratory study. The concentration of samples was individually randomized and blinded for each participant. The circles (●) in the plots represent individual datapoints, with different colors indicating the randomization of the samples. A value of 45 was assigned for undetectable C_q values in order to visualize the generated data. The dashed line indicates the threshold (C_q = 37) used to indicate positive detection. Note that data from laboratory 6 were not included in the statistical analysis because of deviation from the provided protocol.

yeast in a format that provides stability over longer time periods, such as a dry format (e.g., via lyophilization). Using the system described herein, the results indicate that a yeast concentration of 10^7 cells ml^{-1} should result consistently in positive detection. Dry yeast cells could be rehydrated to challenge the qPCR workflow or crushed into a powder to challenge additional steps involved in field microbial risk assessment, such as sampling [27]. In addition to challenging the qPCR analytical process, a quantitative reference material composed of intact yeast cells has potential application for a broad range of other processes where microbial cells or cell components are measured.

5. Conclusions

The need for stable, well-characterized biological materials to aid in reproducible and reliable analytical processes and improve confidence in microbial measurements is relevant to nearly any activity involving detection or quantification of microorganisms. In this study, we describe the development and assessment of a candidate material based on *S. cerevisiae* to evaluate nucleic-acid based detection technologies (e.g., qPCR). We successfully incorporated the ERCC-00095 DNA sequence into the *S. cerevisiae* genome to confer selectivity, and demonstrated via interlaboratory study the detection of NE095 in blinded samples in the hands of potential future users. In addition to a qualitative (yes/no) detection of the yeast, quantitative detection was also achieved by challenging the capability of qPCR to detect predetermined cell concentrations. These results support the further development of NE095 as a reference material for qPCR technologies involved in microbial detection and risk assessment.

Disclaimers

Opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of the Department of Homeland and Security (DHS), National Institute of Standards and Technology (NIST), or affiliated venues. Certain commercial equipment, instruments, or materials are identified in this paper only to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose. Official contribution of NIST; not subject to copyrights in USA.

Conflicts of interest

None declared.

Acknowledgements

The Department of Homeland Security (DHS) Science and Technology Directorate sponsored the production of this material under Interagency Agreement HSHQPM-12-X-00078 with the National Institute of Standards and Technology (NIST). We thank the Florida Department of Health, Michigan Department of Community Health, Minnesota Department of Health, New York Department of Health, Washington Department of Health and 4th Civil Support Team, Georgia Army National Guard for participating in the interlaboratory study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bdq.2016.01.001.

References

[1] T. Butler, Plague gives surprises in the first decade of the 21st century in the United States and worldwide, Am. J. Trop. Med. Hyg. 89 (4) (2013) 788–793.
[2] M. Gandhi, M.L. Chikindas, Listeria: a foodborne pathogen that knows how to survive, Int. J. Food Microbiol. 113 (1) (2007) 1–15.
[3] M. Gueimonde, et al., Antibiotic resistance in probiotic bacteria, Front. Microbiol. 4 (2013) 202.
[4] P. Emanuel, M. Caples, Chemical, Biological, Radiological Technology Survey, US Army RDECOM, 2011.
[5] U.S. Department of Homeland Security, Framework for a Biothreat Field Response Mission Capability. Washington, D.C., 2011.
[6] U.S. Department of Homeland Security, Guide for the Selection of Biological Agent Detection Equipment for Emergency First Responders, 2nd ed. Whashington, D.C., 2007.
[7] T. Hadfield, et al., RAZOR EX anthrax air detection system for detection of *Bacillus anthracis* spores from aerosol collection samples: collaborative study, J. AOAC Int. 96 (2) (2013) 392–398.
[8] J. Kim, J. Lim, C. Lee, Quantitative real-time PCR approaches for microbial community studies in wastewater treatment systems: applications and considerations, Biotechnol. Adv. 31 (8) (2013) 1358–1373.

[9] M. Botes, M. de Kwaadsteniet, T.E. Cloete, Application of quantitative PCR for the detection of microorganisms in water, Anal. Bioanal. Chem. 405 (1) (2013) 91–108.
[10] I.G. Wilson, Inhibition and facilitation of nucleic acid amplification, Appl. Environ. Microbiol. 63 (10) (1997) 3741–3751.
[11] N.D. Olson, J.B. Morrow, DNA extract characterization process for microbial detection methods development and validation, BMC Res. Notes (2012) 668.
[12] L. Rossen, et al., Inhibition of PCR by components of food samples: microbial diagnostic assays and DNA-extraction solutions, Int. J. Food Microbiol. 17 (1) (1992) 37–45.
[13] S.A. Bustin, et al., The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments, Clin. Chem. 55 (4) (2009) 611–622.
[14] U.S. Government Accoutability Office, Anthrax Detection: Agencies need to validate sampling activities in order to increase confidence in negative results. GAO: Washington, D.C., 2005.
[15] S.R. Kane, et al., Rapid: high-throughput, culture-based PCR methods to analyze samples for viable spores of *Bacillus anthracis* and its surrogates, J. Microbiol. Methods 76 (3) (2009) 278–284.
[16] Bipm, International vocabulary of metrology and general concepts and associated terms (VIM), JCGM 200 (2008) 2008.
[17] M.C. Kline, et al., Production and certification of NIST Standard Reference Material 2372 human DNA Quantitation Standard, Anal. Bioanal. Chem. 394 (4) (2009) 1183–1192.
[18] M.C. Kline, et al., Results from the NIST 2004 DNA quantitation study, J. Forensic Sci. 50 (3) (2005) 571–578.
[19] R. Akada, Genetically modified industrial yeast ready for application, J. Biosci. Bioeng. 94 (6) (2002) 536–544.
[20] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (7) (2012) 671–675.
[21] M. Smieja, et al., Replicate PCR testing and probit analysis for detection and quantitation of *Chlamydia pneumoniae* in clinical specimens, J. Clin. Microbiol. 39 (5) (2001) 1796–1801.
[22] Team, R.C., R: A language and environment for statistical computing. R foundation for Statistical computing, Vienna, Austria. http://www.R-project.org. 2013.
[23] R. Stadhouders, et al., The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay, J. Mol. Diagn. 12 (1) (2010) 109–117.
[24] B. Altshuler, Modeling of dose-response relationships, Environ. Health Perspect. 42 (1981) 23–27.
[25] I. Janse, et al., Multiplex qPCR for reliable detection and differentiation of *Burkholderia mallei* and *Burkholderia pseudomallei*, BMC Infect. Dis. 13 (2013) 86.
[26] F. Gardini, et al., Suitability of log-linear models to evaluate the microbiological quality of baby clams (*Chamelea gallina* L.) harvested in the Adriatic Sea, Int. J. Food Microbiol. 54 (1–2) (2000) 63–74.
[27] ASTM, Standard Guide for Operational Guidelines for Initial Response to a Suspected Biothreat Agent, ASTM International, West Conshohocken, PA, USA, 2010.

This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.

Designation: E2458 – 17

Standard Practices for Bulk Sample Collection and Swab Sample Collection of Visible Powders Suspected of Being Biological Agents and Toxins from Nonporous Surfaces[1]

This standard is issued under the fixed designation E2458; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 These practices address collection of visible powders that are suspected biological agents and toxins from solid nonporous surfaces using a bulk collection method, using a dry swab and laminated card, followed by a swab sampling method using a sterile moistened swab. Bulk powder samples are collected and packaged in a manner that permits the maximum amount of the sample to be safely transported to a reference laboratory within the Centers for Disease Control and Prevention (CDC) national Laboratory Response Network (LRN)[2] for confirmatory identification and safe storage. If the source of the powder is a letter or small package, that item is also packaged in a manner that permits it to be safely transported to an LRN reference laboratory. A sterile moistened swab may be used to collect residual powder from the nonporous surface and may be used to conduct on-site biological assessments for the purpose of testing for biological agents and toxins.

1.2 These practices are performed in coordination with the Federal Bureau of Investigation (FBI) as part of a risk assessment including hazard assessment and threat credibility evaluation as recommended and clarified in Guide E2770. The decision to implement these practices and collect a public safety sample will be made by members of the response community of the jurisdiction assuming responsibility through coordination with the FBI and the receiving LRN reference laboratory.

1.3 Sample Collection Method A covers the bulk collection and packaging of suspicious visible powders that are suspected biological agents and toxins from solid nonporous surfaces. All samples suspected to be biological agents and toxins on nonporous surfaces should be collected according to Sample Collection Method A and sent to an LRN reference laboratory for confirmatory testing.

1.4 Sample Collection Method B covers swab sampling of residual suspicious powders that are suspected biological agents and toxins from solid nonporous surfaces. Swab samples can be used for on-site biological assessment; however results from on-site biological assessments are not definitive; confirmatory testing by the LRN reference laboratory is necessary to make public health decisions.

1.5 These practices incorporate reference guidance for packaging and transport of suspicious visible powders to comply with all appropriate federal regulations regarding biosafety and biosecurity.

1.6 These practices should only be used to collect visible samples that are suspected biological agents and toxins and have been field screened according to reference guidance for explosive hazard, radiological hazard, and other acute chemical hazards.

1.7 The values stated in SI units are to be regarded as standard. The values given in parentheses are for information only.

1.8 *This standard does not purport to address all of the safety concerns associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

1.9 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[3]

[1] These practices are under the jurisdiction of ASTM Committee E54 on Homeland Security Applications and are the direct responsibility of Subcommittee E54.01 on CBRNE Sensors and Detectors.

Current edition approved May 15, 2017. Published June 2017. Originally approved in 2006. Last previous edition approved in 2010 as E2458 – 10. DOI: 10.1520/E2458-17.

[2] The CDC Laboratory Response Network is the network responsible for handling clinical specimens and environmental samples containing suspected biothreat agents.

[3] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

Copyright © ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959. United States

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

E2770 Guide for Operational Guidelines for Initial Response to Suspected Biological Agents and Toxins
F2412 Test Methods for Foot Protection
F2413 Specification for Performance Requirements for Protective (Safety) Toe Cap Footwear

2.2 *Federal Government Regulations:*[4]
18 USC 178 Prohibitions with respect to biological weapons
DOT 49 CFR, Parts 171-180 Hazardous Materials Regulations
DOT - 49 CFR 172 Subpart H, Transportation Training
DOT - 49 CFR 173 General Requirements for Shipments and Packagings
DOT - 49 CFR 178 Specifications for Packagings
EPA - 40 CFR 300 National Oil and Hazardous Substances Pollution Contingency Plan (NCP)
EPA - 40 CFR 311 Worker Protection
NRC - 10 CFR 20 Standards for Protection against Radiation
NIOSH - 42 CFR 84 Respiratory Protective Devices
OSHA - 29 CFR 1910 Subpart Z and 29 CFR 1926 Subpart Z Toxic and Hazardous Substances
OSHA - 29 1910.1096 and 29 CFR 1926.53 Ionizing Radiation
OSHA - 29 CFR 1910.120 Hazardous Waste Operations and Emergency Response (HAZWOPER) standard
OSHA - 29 CFR 1910 Subpart I (Sections 132 to 139) Personal Protective Equipment
OSHA - 29 CFR 1910.1200 Hazard Communication

2.3 *Federal Guidance:*
OSHA - CPL 02-02-073 Inspection Procedures for 29 CFR 1910.120 and 1926.65, Paragraph (q): Emergency Response to Hazardous Substance Releases
NIOSH Publication No. 2009-132 Recommendations for the Selection and Use of Respirators and Protective Clothing for Protection Against Biological Agents
FBI Laboratory Publication Handbook of Forensic Services 2013
FBI-DHS-HHS/CDC Coordinated Document Guidance on Initial Response to a Suspicious Letter/Container with a Potential Biological threat, November 2, 2004
CDC/NIOSH Surface Sampling Procedures for *Bacillus anthracis* Spores from Smooth, Non-porous Surfaces, April 26, 2012[5]
DHS Framework for a Biothreat Field Response Mission Capability, April 2011[6]
Sandia National Laboratories SAND2005-3237 (LBNL-54973 (II)) Guidelines to Improve Airport Preparedness Against Chemical and Biological Terrorism[7]

2.4 *NFPA Standards:*[8]
NFPA 472 Standard for Competence of Responders to Hazardous Materials/Weapons of Mass Destruction Incidents, 2013 Edition
NFPA 1994 Standard on Protective Ensembles for Chemical/Biological Terrorism Incidents

2.5 *IATA Standards:*[9]
IATA PI 602 Infectious Diseases (Infectious Substances)
IATA PI 650 Shipping of Diagnostic Samples
IATA DGR 46th Edition, 2005
IATA DGR Addendum I, January 2005
IATA DGR Addendum II, March 2005
IATA DGR Addendum III, July 2005

2.6 *ANSI Standards:*[10]
ANSI Z41-1999 American National Standard for Personal Protection - Protective Footwear
ANSI Z87.1-2003 American National Standard for Occupational and Educational Personal Eye and Face Protection Devices
ANSI Z88.2-1992 American National Standard Practices for Respiratory Protection
ANSI Z88.10-2001 American National Standard for Personal Protection - Respirator Fit Testing Methods
ANSI/ISEA Z89.1-2003 American National Standard for Personal Protection - Protective Headwear for Industrial Workers Requirements
ANSI/Compressed Gas Association, CGA G-7.1-1997 Commodity Specification for Air

2.7 *IAFC Guidance:*[11]
Model Procedures for Responding to a Package with Suspicion of a Biological Threat, October 2008

3. Terminology

3.1 *Definitions:*

3.1.1 *aseptic technique, n*—operation or performance of a procedure or method under carefully controlled conditions to reduce the risk of exposure and prevent the introduction of unwanted material/matter (contamination) into a sample.

3.1.2 *biological agent, n*—any microorganism (including but not limited to, bacteria, viruses, fungi, rickettsiae, or protozoa), or infectious substance, or any naturally occuring, bioengineered or synthesized component of any such microorganism or infectious substance, capable of causing: *(1)* death, disease or other biological malfunction in a human, an animal, a plant, or another living organism; *(2)* deterioration of food, water, equipment, supplies, or material of any kind; or *(3)* deleterious alteration of the environment. **18 USC 178**

3.1.3 *bulk powder, n*—a visible powder, at least approximately 5 mL (1 teaspoon) in volume amassed or dispersed over

[4] Available from U.S. Government Printing Office Superintendent of Documents, 732 N. Capitol St., NW, Mail Stop: SDE, Washington, DC 20401, and also available online from Occupational Safety and Health Administration (www.osha.gov).
[5] Available from http://www.cdc.gov/niosh/topics/emres/surface-sampling-bacillus-anthracis.html.
[6] Available from http://www.hsdl.org/?view&did=767721.
[7] Available from http://share-ng.sandia.gov/news/resources/releases/2005/images/unlsand-2005-3237.pdf.
[8] Available from National Fire Protection Association (NFPA), 1 Batterymarch Park, Quincy, MA 02269-9101.
[9] Available from the International Air Transport Association, 800 Place Victoria, PO Box 113, Montreal-H4Z 1M1, Quebec, Canada.
[10] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.
[11] Available from International Association of Fire Chiefs (IAFC), 4025 Fair Ridge Drive, Suite 300 Fairfax, VA 22033, http://www.iafc.org.

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

a limited area (optimally, area should be less than 20 by 20 cm (approximately 8 by 8 in.).

3.1.4 *chain of custody, n*—set of procedures and documents to account for the integrity of sample by tracking its handling and storage from point of sample collection to final disposition of the sample.

3.1.5 *cold zone, n*—the uncontaminated area where workers are unlikely to be exposed to hazardous substances or dangerous conditions; also known as Clean Zone or Support Zone. **CPL 02-02-071 Directive**

3.1.6 *confirmatory analysis, n*—a test or a series of assays that definitively identifies the presence of a suspected substance or agent.

3.1.6.1 *Discussion*—Confirmatory analysis of a biological agent for public health action can be performed only by an LRN national or reference laboratory.

3.1.7 *decontamination, n*—the physical or chemical process, or both, of reducing and preventing the spread of contaminants from people, animals, the environment, or equipment involved at hazardous materials/weapons of mass destruction (WMD) incidents. **NFPA**

3.1.8 *field screening, n*—field measurements utilized early in the response to define and characterize the potential hazards present, including corrosive, flammable, volatile, radioactive, or oxidizer hazards, and to support tactical decision making to address operational safety measures.

3.1.8.1 *Discussion*—Field screening does not include measurements of biological properties, which is termed on-site biological assessments (see 3.1.12).

3.1.9 *hazard, n*—something that is potentially dangerous or harmful, often the root cause of an unwanted outcome; a danger or peril. **NIMS**

3.1.10 *hot zone, n*—the area, located on the site where contamination is either known or expected and where potential for greatest exposure exists; also known as Exclusion Zone or ExZ. **CPL 02-02-071 Directive**

3.1.11 *incident commander (IC), n*—the individual responsible for all incident activities, including the development of strategies and tactics and the ordering and release of resources.

3.1.11.1 *Discussion*—The IC has overall authority and responsibility for conducting incident operations and is responsible for the management of all incident operations at the incident site. **NIMS**

3.1.12 *on-site biological assessment, n*—measurements of properties inherent to biological materials performed in the field using rapid, field-based procedures and assays.

3.1.13 *personal protective equipment (PPE), n*—equipment provided to shield or isolate a person from the chemical, biological, physical, and thermal hazards that can be encountered at hazardous materials/weapons of mass destruction (WMD) incidents. **NFPA**

3.1.14 *presumptive test, n*—non-definitive test used to evaluate a material for the presence of a substance or agent, or the presence of signatures of a substance or agent.

3.1.15 *risk, n*—the probability of suffering a loss or harm or injury; peril.

3.1.16 *toxin, n*—the toxic material or product of plants, animals, microorganisms (including but not limited to, bacteria, viruses, fungi, rickettsiae, or protozoa), or infectious substances, or a recombinant or synthesized molecule, whatever their origin and method of production, and includes: (*1*) any poisonous substance or biological product that may be engineered as a result of biotechnology produced by a living organism; or (*2*) any poisonous isomer or biological product, homolog, or derivative of such a substance (18 USC 178).

3.1.17 *threat, n*—an indication of possible violence, harm, or danger and may include an indication of intent and capability. **NIMS**

3.1.18 *warm zone, n*—the transition area between the Exclusion Zone (ExZ or hot zone) and the Support Zone (SZ or cold zone) used to reduce and limit the amount of contamination on people and equipment, and in the air, water, and soil that may be transferred into nonhazardous areas; the CRZ contains decontamination facilities, and functions as a buffer zone surrounding the ExZ; also known as the contamination reduction zone or CRZ. **CPL 02-02-071 Directive**

3.1.19 *weapon of mass destruction (WMD), n*—any weapon or device that is intended, or has the capability, to cause death or serious bodily injury to a significant number of people through the release, dissemination, or impact of (*1*) toxic or poisonous chemicals or their precursors; (*2*) a disease organism; or (*3*) radiation or radioactivity. **U.S. Code Title 50, Ch. 40, Sect. 2302, War and National Defense Definitions**

3.2 *Acronyms:*

3.2.1 *AHJ*—Authority Having Jurisdiction

3.2.2 *ANSI*—American National Standards Institute

3.2.3 *ASTM*—American Society for Testing and Materials

3.2.4 *CDC*—Centers for Disease Control and Prevention

3.2.5 *CFR*—Code of Federal Regulations

3.2.6 *CRZ*—Contamination Reduction Zone

3.2.7 *CST*—Civil Support Team

3.2.8 *DHS*—Department of Homeland Security

3.2.9 *DOT*—Department of Transportation

3.2.10 *EOC*—Emergency Operations Center

3.2.11 *EPA*—Environmental Protection Agency

3.2.12 *ExZ*—Exclusion Zone

3.2.13 *FBI*—Federal Bureau of Investigation

3.2.14 *FEMA*—Federal Emergency Management Agency

3.2.15 *HAZMAT*—Hazardous Materials

3.2.16 *HHS*—Health and Human Services

3.2.17 *IAFC*—International Association of Fire Chiefs

3.2.18 *IATA*—International Air Transport Association

3.2.19 *IC*—Incident Commander

3.2.20 *ICS*—Incident Command System

3.2.21 *IEC*—International Electrotechnical Commission

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

3.2.22 *ISEA*—International Safety Equipment Association

3.2.23 *ISO*—International Organization for Standardization

3.2.24 *LRN*—Laboratory Response Network

3.2.25 *MACS*—Multiagency Coordination System

3.2.26 *NFPA*—National Fire Protection Association

3.2.27 *NIMS*—National Incident Management System

3.2.28 *NIOSH*—National Institute for Occupational Safety and Health

3.2.29 *NRC*—Nuclear Regulatory Commission

3.2.30 *OSHA*—Occupational Safety and Health Administration

3.2.31 *PBS*—Phosphate Buffered Saline

3.2.32 *PPE*—Personal Protective Equipment

3.2.33 *SZ*—Support Zone

3.2.34 *US&R*—Urban Search and Rescue

3.2.35 *WMD*—Weapons of Mass Destruction

4. Significance and Use

4.1 These practices should be used only to collect visible samples that are suspected biological agents and toxins and have been field screened as defined by the FBI-DHS-HHS/CDC Coordinated Document for explosive hazard, radiological hazard, and other acute chemical hazards.

4.2 These practices provide standardized methods for collecting, packaging, and transporting suspicious visible powder samples that are suspected biological agents and toxins. Collection of a bulk powder material from a nonporous surface using a sterile swab and laminated card as the collection devices to move the material into a container will depend on several factors, including (but not limited to): *(1)* amount of visible powder present; *(2)* sample composition; *(3)* choice of collection device; *(4)* size and shape of the collection container; *(5)* ability of the powder to become aerosolized; *(6)* texture and porosity of the surface; *(7)* humidity; *(8)* air movement; and *(9)* electrostatic properties of powders and collection tools/containers.

4.3 Similarly, these practices standardize methods for sampling suspicious visible powders for on-site analysis, although wipe and swab sampling is often employed in the field for subsequent LRN reference laboratory analysis. The ability to collect suitable samples from nonporous surfaces using a sterile moistened swab will depend on the following factors: *(1)* swabbing procedure; *(2)* swab material; *(3)* sample composition; and *(4)* texture of the surface.

4.4 These practices standardize suspicious powder collection and packaging procedures and swab sampling procedures in order to reduce exposure risk, to reduce variability associated with sample handling and sample analysis, and to increase reliability of sampling visible powder samples from nonporous surfaces.

4.5 The bulk sample collection practice and the swab sampling practice are recommended for collecting amassed or dispersed powder samples from all nonporous surfaces on which the suspicious powder sample is clearly visible.

4.6 These practices are not recommended for samples on porous materials such as upholstery, carpeting, air filters, or ceiling tiles.

4.7 These practices are recommended for collecting visible powders where the bulk of the powder sample is amassed or dispersed over a limited area (optimally, area should be less than 20 by 20 cm (approximately 8 by 8 in.) or 400 cm^2 (approximately 64 $in.^2$).

4.8 These practices are to be performed by personnel who are adequately trained to work with hazardous materials in the hot zone (see NFPA 472, or OSHA - 29 CFR 1910.120). Personnel performing collection or screening under these practices shall be adequately trained in the use of sampling equipment, materials, and procedures. This includes personnel performing the prior initial chemical and radiological screening. Personnel should use the appropriate level of personal protective equipment (PPE) to mitigate hazards during collection and screening. Personnel performing collection or screening under these practices shall be aware of evidence preservation and sampling procedures (NFPA 472 section 6.5).

4.9 These standard practices should be used in accordance with Guide E2770 for best practices for planning, training and evaluation of competency.

SAMPLE COLLECTION METHOD A—BULK SAMPLE COLLECTION FOR LABORATORY ANALYSIS

5. Scope of Method A

5.1 This sample collection method applies to the bulk collection and packaging of suspicious visible powders that are suspected biological agents or toxins, or both, from solid nonporous surfaces.

5.2 These practices are performed in coordination with the FBI and receiving LRN reference laboratory after a risk assessment including a hazard assessment and threat credibility evaluation is conducted and the sample is deemed potentially to be a credible threat as recommended and clarified in Guide E2770.

5.3 This sample collection method applies to suspicious visible powders that are amassed or dispersed in a limited area where the bulk of the powder sample is in an area that is less than 20 by 20 cm (approximately 8 by 8 in.) or 400 cm^2 (approximately 64 $in.^2$).

5.4 These practices should be used only to collect samples that are suspected biological agent and toxin hazards that have been field screened as defined by the FBI-DHS-HHS/CDC Coordinated Document for explosive, radiological, and chemical hazards.

6. Summary of Sample Collection Method A

6.1 A visible powder sample that is a suspected biological agent or toxin, or both, and its source should be field screened for non-biological hazards as defined in the FBI-DHS-HHS/CDC Coordinated Document and according to the reference

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017

Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

guidance including appropriate NFPA documents. Non-biological hazards include explosive, radiological, and chemical hazards. The visible powder sample, amassed or dispersed in a limited area, is collected from a nonporous surface using a swab and a laminated card to move the sample into a sterile dry collection container that is held close to the surface.

6.2 The method provides guidance on performing these procedures in a manner that will minimize sample loss and aerosolization of the powder. The bulk powder sample and swab are sent to an LRN reference laboratory for confirmatory analysis.

7. Sampling and Packaging Equipment and Supplies

7.1 *Personal Protective Equipment*—Level A, B, or C personal protective equipment ensembles as indicated (see Section 2 for additional guidance, including OSHA - 29 CFR 1910.120 Appendix B and NFPA 1994).

7.2 Clean drop cloth to create an isolated work area.

7.3 *Sample Transport Container*—Bucket or large heavy duty plastic bag used to transport the samples and materials from the hot zone through to the decontamination line.

7.4 *Non-powdered Nitrile or Vinyl Examination Gloves.*

7.5 *Two Sterile Polypropylene Wide-mouth Screw-capped Sample Collection Containers (Sample Containers)*—Containers must possess a leak-resistant seal; Diameter of container mouth must be large enough to accomodate the 4 by 6.5 cm plastic cards (Section 7.10); each pre-labeled as "POWDER SAMPLE" with unique sample identifier numbers.

7.6 *Bleach Solution*—Fresh pH-adjusted bleach solution (household bleach diluted 1:9; pH-adjusted to 6.8-8.0) to be prepared outside of the hot zone just prior to use by (step 1) mixing one part household bleach (5.25 to 6.0 % sodium hypochlorite) with 5 parts water (volume/volume); (step 2) adding 1 part white vinegar; (step 3) adding 3 parts of additional water.[12]

7.7 *Labeling or Marking Tape.*

7.8 *Solvent-resistant Indelible Marker.*

7.9 *Sterile Culture Swabs,* (rayon, macrofoam, or polyester) individually packaged and sterile, self-contained in sealed plastic tubes, with absorbent material wrapped around one end of a plastic stick, unopened.

7.10 *Two at least 4 by 6.5 cm (approximately 1.5 by 2.5 in.) Sterile Plastic Laminated Cards.*

7.11 *Six 1-gal Sealable Plastic Bag(s)*—Both bags with sliding lock and Trademark Whirl-Pak[13] bags are not recommended for this purpose since they may not seal completely; the use of colored or opaque bags is discouraged because it makes viewing of the sample more difficult once transported back to the laboratory in the CDC Laboratory Response Network.

7.12 *Bucket,* or other container marked "BIOHAZARD WASTE".

7.13 *1-gal Sealable Plastic Bag(s),* pre-labeled as "BIOHAZARD WASTE".

7.14 *Field Screening Results Form*—See example in Appendix X1.

7.15 *Sample Collection Sheet*—See example in Appendix X1. Note that a single sample collection sheet can be used for all items collected at a single location including primary source, swab, and powder sample(s).

7.16 *Chain-of-Custody Form*—See example in Appendix X1.

7.17 *Two Plastic Transparent Document Pouches,* with adhesive on back.

7.18 *Durable Hard-sided Outer Container,* with lid or screw cap, for sample transport (such as metal can with lid, or plastic container with lid).

7.19 Attach documentation described in the Sample Collection and Submission section of Guide E2770. Documentation should include the contact information for local, state, and federal law enforcement, local health officials and local/national testing facilities.

8. Procedure

8.1 Practices employed herein preserve the integrity of the material in the event that it becomes evidence. Efforts should be made to minimize or reduce air currents that may re-aerosolize or disturb suspected biological agents, including ceasing the operation of heating, ventilation, and air conditioning (HVAC) and exhaust fans and other forms of air movement. The location of the work area should be considered prior to entry into the hot zone. When selecting work location, teams should consider proximity to sample location to reduce possible contamination of the work area, while also minimizing the distance necessary to move materials during sample collection. Additionally these actions should be supported by a written incident action plan and site health and safety plan developed under the direction of the Incident Commander. A coordinated risk assessment including a hazard and threat credibility evaluation should be performed prior to determination of a sampling mission. These practices assume that a sampling site/hot zone has already been defined. These practices are to be performed by personnel who are adequately trained to work with hazardous materials in the hot zone (NFPA 472, and 29 CFR, Part 1910.120).

8.2 Prior to allowing people to enter the hot zone for sampling purposes, proper site safety practices should be implemented, including establishing decontamination areas, and assuring that the appropriate PPE is selected based on a risk-assessment.

8.3 Prior to performing these practices, review the sampling plan and sampling procedures and assemble all necessary equipment before entering the hot zone. At minimum, a two-person team is required to perform these sampling procedures in the hot zone.

[12] Additional information about decontamination with pH-adjusted bleach solution is available at nepis.epa.gov.

[13] Whirl-Pak is a trademark of Nasco International, Inc.

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

8.4 Prior to entering the hot zone for sampling purposes, label the two unopened, sterile plastic sample collection containers. The sample collection containers are the primary containment for bulk dry powder sample and are labeled as "POWDER SAMPLE" with unique sample identifier numbers on each container. Properly label five 1-gal self-sealing bags in the following manner: *(1)* and *(2)* containment tubes for two dry swabs each labeled as "DRY SWAB" and with unique sample identification numbers; *(3)* if applicable, the primary containment for source of powder (letter or small package) labeled as "PRIMARY SOURCE" and with a unique sample identifier number; and *(4)* and *(5)* containment for two sample collection containers labeled as "POWDER SAMPLE" and with unique sample identifiers.

8.5 After entering the hot zone, ensure that the sample material is protected from wind currents and moisture until all sampling is completed.

8.6 Perform basic field screening on material prior to sample collection to assess explosive hazards, radiological hazards, and acute chemical hazards (NFPA 472). Use screening methods to minimize sample consumption, thereby conserving as much of the sample as possible. Ensure that the results of field screening are documented indicating which tests have been performed and their outcome (see example of field screening results form in Appendix X1). In situations where biological agents or toxins, or both, are suspected, the item(s) should be field screened for explosive devices and substances, radiological materials, corrosive materials and volatile organic compounds as defined in the joint FBI-DHS-HHS/CDC Coordinated Document guidelines for responders to suspicious letters and packages.

8.7 This instruction is provided for a two-person sampling team to conduct the sampling procedure in the hot zone; the procedure should not be performed by a single individual. The first team member (assistant sampler, also known as the facilitator) is responsible for communication, photography (FBI Laboratory Publication, Handbook of Forensic Services 2003), ensuring that the sample collection sheet is filled out, and facilitating collection (for example, opening and handing materials to the sampler as required, including sample collection containers, gloves, swab, laminated card, other sampling materials, and packaging materials). The second team member (sampler) is the person collecting the sample and should be the only person to come in contact with the suspicious material. The sampler is also responsible for signing the final chain-of-custody form outside of the hot zone.

8.8 Some jurisdictions may have standard operating procedures requiring the collection of negative controls. Sampling teams should refer to standard operating procedures regarding the collection of any negative controls (also referred to as field and media blanks). Negative controls include unopened sampling tools and any wetting solutions. Blanks must be submitted for each lot number used.

8.9 All team members must put on a new pair of non-powdered nitrile or vinyl examination gloves over the gloves that are part of standard PPE ensemble (that is, team members will have multiple layers of gloves on) for sample collection.

NOTE 1—The use of appropriate aseptic techniques including doffing of gloves between sample collections is done to minimize cross contamination.

8.10 *Assistant Sampler*—Lay down the clean drop cloth to create an isolated work area, and place materials on drop cloth.

8.11 *Assistant Sampler*—Ensure the following items are documented through radio communication between hot and cold zone personnel and finalization of paperwork in the cold zone prior to sending the sample to the laboratory (see example in Appendix X1).

8.11.1 Unique sample number or identifier,
8.11.2 Sample location and address,
8.11.3 Type of sample,
8.11.4 Type and texture of surface sampled,
8.11.5 Time and date of sample,
8.11.6 Names and signatures of persons collecting sample,
8.11.7 Measured size of the area sampled, and
8.11.8 Map of sample area and photograph is possible. Additional efforts to photograph the materials associated with the primary source (for example, envelope, package outer wrap, or letter, or combinations thereof) to discern writings and markings prior to collection are useful to support the law enforcement mission.

8.12 *Assistant Sampler*—If the source of the powder is present on the surface, and the source is a letter or other small package that can fit easily into a 1-gal sealable plastic bag, position the pre-labeled bag labeled as "PRIMARY SOURCE" above the surface next to the source, and hold the bag open.

8.13 *Sampler*—Place the source gently into the plastic bag taking care to make sure all writing and markings are visible through the bag. If there are both documents and packaging, collect them in separate bags to ensure visibility if needed. Limited handling of the primary source is recommended to preserve forensic attributes for additional testing and to minimize resuspension potential.

NOTE 2—Written and other markings on documents associated with the threat are valuable to law enforcement early in the investigation. Special care should be taken to ensure law enforcement entities at the end of the decontamination line can clearly see anything written or labels, or both, on packing associated with the primary source.

8.14 *Assistant Sampler*—Carefully dispel excess air, and seal the bag. Place the sealed bag containing the source into another transparent sealable bag.

8.15 *Assistant Sampler*—Carefully dispel excess air; seal the bag, and place into sample transport container for decontamination.

8.16 *Assistant Sampler*—Give laminated card to sampler. Loosen the cap of the tube containing the swab.

8.17 *Sampler*—Remove the swab from the tube while the assistant sampler holds the tube. Hold the laminated card at an angle on the surface next to the powder. If the surface is smooth, use the card to push the powder into a pile on the surface. Use the sterile swab to gently push the dry powder onto the laminated card. Be sure to use slow, deliberate motions while moving the powder.

8.18 *Assistant Sampler*—Firmly hold the tube for the sampler.

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

8.19 *Sampler*—Place swab firmly into tube, taking care not to touch the outside of the tube with the swab.

8.20 *Assistant Sampler*—Close the tube. Place the tube containing the swab into a transparent sealable bag labeled as "DRY SWAB". Carefully dispel excess air; seal the bag, and place into sample transport container for decontamination.

8.21 *Assistant Sampler*—Open the sample container labeled as "POWDER SAMPLE" and hand the open container to the sampler.

8.22 *Sampler*—Hold the sample collection container on its side parallel to the surface and place the laminated card (with the powder on top) into the sample collection container. Do this slowly and gently so as to minimize aerosolization of the powder.

8.23 *Sampler*—Take the lid from the assistant sampler and place the lid on the sample collection container containing the laminated card and dry bulk powder.

8.24 *Sampler*—Place the closed, pre-labeled container containing the dry bulk powder into a transparent sealable bag held open by the assistant sampler. Do not touch the outside of the bag.

8.25 *Assistant Sampler*—Carefully dispel excess air; seal the bag, and place into sample transport container for decontamination.

8.26 If the amount of powder is too large to collect it all using a single laminated card and swab, repeat steps described in 8.16 – 8.25 with a new card, swab, and pre-labeled sample collection container. In situations where significant material is present after the initial sample is collected for the laboratory, the sampling team should coordinate and discuss with the receiving laboratory and FBI the disposition of the residual material.

8.27 If there is residual powder and conditions allow for the team to remain in the hot zone, the team may perform Sample Collection Method B now prior to leaving the hot zone.

8.28 Sampling team transports all samples out of the hot zone and into the decontamination line in the warm zone.

8.29 *Assistant Sampler*—Rinse or wipe the outside of the sealed plastic bags containing the primary source, powder sample(s), and swab(s) with decontamination solution—do not dry the outside of the bags afterwards as this will allow appropriate decontamination solution contact time. The outer surface of the sealed plastic bag should be decontaminated using a 10 % bleach solution adjusted to a pH of 7. A contact time of 10 min is needed for treating the outer surface of bagged items. Treated outer surfaces of bagged items should remain wet when placed in additional containment bags to insure adequate contact times are achieved. Collect any bleach runoff in a container marked "BIOHAZARD WASTE".

8.30 *Assistant Sampler*—Place the rinsed, sealed bags containing the primary source, powder sample(s), and swab(s) into separate sealable plastic bags and seal.

8.31 Any unused supplies that were carried into the hot zone (plastic bags, sample containers) should also be placed into "BIOHAZARD WASTE" bag(s).

8.32 *Assistant Sampler and Sampler*—After following all proper personnel decontamination procedures (NFPA 472), move the samples to the cold zone. Place the rinsed, sealed bags into a durable hard-sided outer container. Responding personnel should contact the receiving LRN reference laboratory to ensure that the dimensions of the outer container will fit in the laboratory biosafety cabinet. Place the lid on the container.

8.33 *Assistant Sampler and Sampler*—Initial and date the sample package with the indelible marker.

8.34 *Assistant Sampler and Sampler*—Confirm radio communicated information is consistent or if only a 2 person team responded, transfer all sample and field screening information to a clean sample collection sheet and field screening results form in the cold zone.

8.35 *Assistant Sampler and Sampler*—Transfer or verify, or both, all unique sample numbers or identifiers and other pertinent information from the sample collection sheet onto a chain-of-custody form (see example in Appendix X1).[14]

8.36 *Assistant Sampler and Sampler*—Attach a self-sticking document pouch to the durable outer container. Place the clean sample collection sheet and field screening results form and any additional paperwork or documentation inside the document pouch. Place labeling tape over the opening of the document pouch and initial and date the tape with the indelible marker.

8.37 *Sampler*—Sign a chain-of-custody form at the receiving area. Hand the package to the person who is responsible for transporting the package (transporter).

8.38 *Transporter*—Examine the outer package to determine that it is properly packaged and sealed, and that the chain-of-custody form is completed. Sign the original chain-of-custody form. Give one copy of the form to the submitter (sampler) and retain the original form. Place the original signed chain-of-custody form into a second self-sticking document pouch and adhere to the outside of the package. Notify receiving LRN reference laboratory of package transportation and estimate time of delivery. Transport all samples under secure conditions to the predesignated, approved LRN reference laboratory. Any transfers of the materials from one person to another should be documented on the chain-of-custody form with signatures.

8.39 The primary source, powder sample(s), and swab(s) are transported in a manner that complies with all state and federal regulations (IATA PI 602, IATA PI 650, IATA DGR 46th Edition, IATA DGR Addendum I, IATA DGR Addendum II, IATA DGR Addendum III, and DOT 49 CFR, Parts 171-180). All materials are maintained under chain of custody in a secure and biosafe area at the LRN reference laboratory until testing is completed.

[14] Prior to shipping any specimen suspected to contain a biological agent, contact your state public health laboratory or nearest laboratory in the national response network (currently the CDC Laboratory Response Network) for specific guidance. Any materials that might be used as evidence in any investigation must be controlled by a chain of custody at all times.

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

SAMPLE COLLECTION METHOD B—SWAB SAMPLE COLLECTION FOR ON-SITE ANALYSIS

9. Scope of Method B

9.1 This sample collection method applies to swab sampling of suspected biological agents and toxins from nonporous surfaces to collect residual powder after Method A has been applied. The swab sampling procedure produces a sample that may be used for optional, on-site biological assessments. While the sample collected with Method B may be used for on-site biological assessments, it is recommended that assessment methods have been validated by nationally recognized consensus standards (for example, AOAC International, Stakeholders' Panel on Agent Detecting Assays (SPADA) performance specifications[15]) and supported as defined in Guide E2770.

9.2 These practices are performed only after collecting the bulk sample in the hot zone as described in Sample Collection Method A-Bulk Sample Collection.

9.3 These practices should be used only to collect samples that are suspected biological hazards and that have been field-screened for explosive hazards, radiological hazards, and acute chemical hazards.

10. Summary of Sample Collection Method B

10.1 A nonporous surface from which a suspicious visible bulk powder, a suspected biological agent, has previously been collected is swabbed using a moistened swab to collect any residual powder. The sample may be utilized in on-site biological assessment to support public safety operations.

11. Sampling Equipment and Supplies for Sample Collection Method B

11.1 *Personal Protective Equipment*—Level A, B, or C as necessary (OSHA - 29 CFR 1910.120 Appendix B and NFPA 1994).

11.2 Sample collection tools are the following: *(1)* sterile culture swab (rayon, macrofoam, or polyester), individually packaged and self-contained in sealed plastic tube, with absorbent material wrapped around one end of a plastic stick, and, if available, *(2)* sample collection device provided by manufacturer of the on-site biological assessment kit.

11.3 Sterile solutions for moistening the swab are as follows: *(1)* sterile vial with lid, pre-labeled as "BUFFER," containing minimally 0.5 mL of Phosphate Buffered Saline (PBS) solution with 0.1 % Tween-20,[16] or if available, *(2)* buffer solution from manufacturer of on-site biological assessment kit if required by manufacturer. Background control samples of collection materials including buffer solutions utilized in sampling of residual powder should be submitted to the receiving LRN reference laboratory. Any control buffers or materials must be clearly labeled and indicated in submission documentation.

11.4 *1-gal Sealable Bag*—Pre-labeled as "WET SWAB" with unique sample identifier number.

11.5 *Sample Collection Sheet.*

12. Procedure

12.1 *Sampler*—The residual powder that remains on the surface following bulk powder sample collection may be used for on-site biological assessments. Follow instructions provided by the manufacturer regarding sample collection for on-site biological assessments using a commercially available kit. Personnel performing on-site biological assessments should be adequately trained in the use of biological assessment technologies (NFPA 472).

12.2 *Assistant Sampler*—Record the methods and results of on-site biological assessments. If the residual powder sample is consumed for on-site analysis, a chain-of-custody form is not needed.

12.3 If there is significant amount of residual powder on the surface and no on-site biological assessment is performed, perform the steps described in 12.4 – 12.11 to collect residual powder for packaging and transport to the receiving LRN reference laboratory.

12.4 *Assistant Sampler*—Remove the lid from the vial labeled as "BUFFER" and hand the vial to the sampler. Loosen swab from tube (but do not remove swab) and hand tube with swab to sampler. Sampler holds the tube and "BUFFER" vial in the same hand.

12.5 *Sampler*—Remove swab from tube and place into "BUFFER" vial to moisten, using aseptic technique to prevent cross contamination. Press the swab against the side of the vial to remove excess liquid. Remove swab from "BUFFER" vial and drop "BUFFER" vial into biohazardous waste container.

12.6 *Sampler*—Wipe the swab over the surface where the powder was originally found, using closely spaced vertical S-strokes or Z-strokes over the entire sampling area.

12.7 *Sampler*—Roll the swab handle (end of the plastic stick furthest from absorbent material) between fingers to rotate the swab, thereby exposing a fresh surface. Wipe the swab over the entire area again, this time using horizontal S-strokes or Z-strokes over the surface. The swab area should preferably not exceed the maximum recommended area of 20 by 20 cm (8 by 8 in.) or 400 cm^2 (approximately 64 $in.^2$).

12.8 *Sampler*—Place swab into tube, and seal tube by pressing firmly.

12.9 *Sampler*—Place the closed tube containing the swab into a transparent sealable bag labeled as "WET SWAB" held open by the assistant sampler. Do not touch the outside of the bag.

12.10 *Assistant Sampler*—Carefully dispel excess air; seal the bag, and place into sample transport container for decontamination.

12.11 Continue with procedure in 8.28 through to the end of Method A as determined appropriate through correspondence with the receiving LRN reference laboratory.

13. Keywords

13.1 packaging; sample collection; suspicious powders; swab

[15] SPADA specifications are available at www.aoac.org.
[16] Tween-20 is a trademark of ICI Americas, Inc.

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

APPENDIX

(Nonmandatory Information)

These appendices provide example forms for the user. Use of these specific forms is not mandatory.

X1. EXAMPLE OF BEST PRACTICES FORMS

X1.1 See Table X1.1, Fig. X1.1, Table X1.2, Fig. X1.2, Fig. X1.3, and Fig. X1.4.

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017

Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

TABLE X1.1 Field Screening Results Form

Location:

Date:

Screening Team:

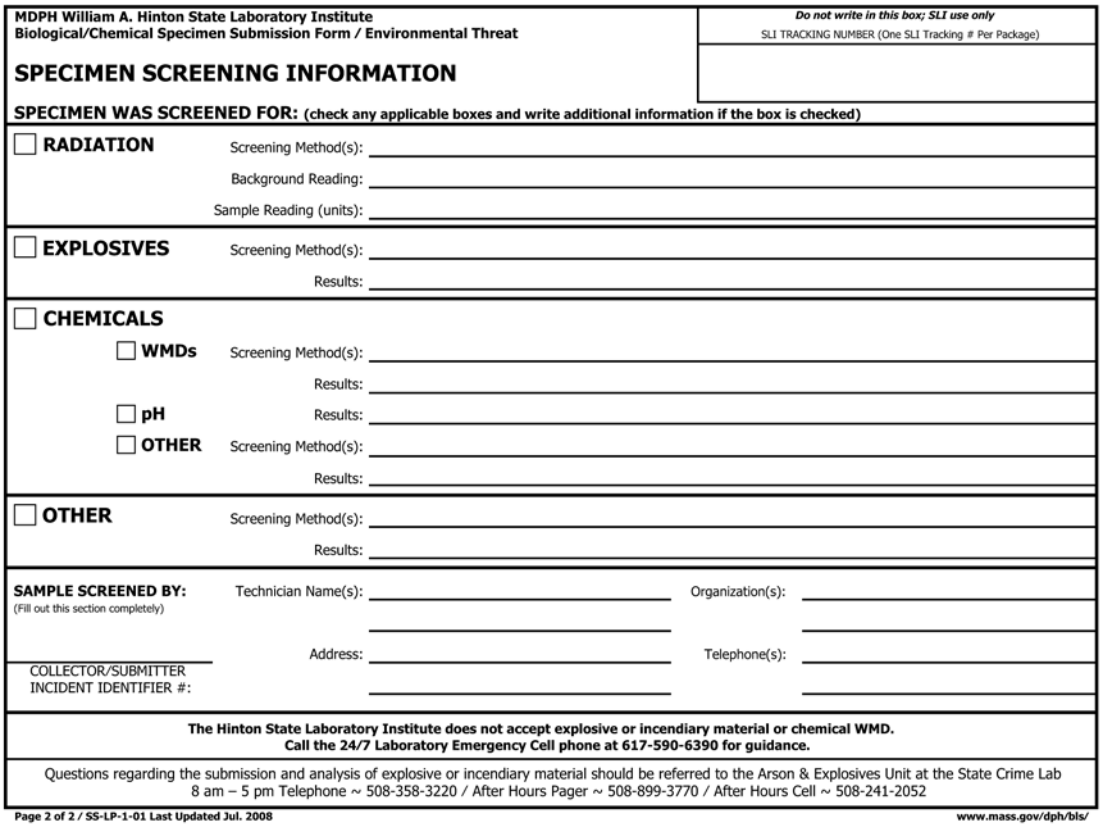
Time:

Sample #	Color	Rad (uR/h)	Flammable (% LEL)	PID/VOC (ppm)	O_2 (%)	Oxidizer or energetic material	Observations/Comments/Other Screening Tools Used

Monitoring Instruments Used:

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017

Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.



MDPH William A. Hinton State Laboratory Institute
Biological/Chemical Specimen Submission Form / Environmental Threat

Do not write in this box; SLI use only
SLI TRACKING NUMBER (One SLI Tracking # Per Package)

SPECIMEN SCREENING INFORMATION

SPECIMEN WAS SCREENED FOR: (check any applicable boxes and write additional information if the box is checked)

☐ **RADIATION**
Screening Method(s): _____
Background Reading: _____
Sample Reading (units): _____

☐ **EXPLOSIVES**
Screening Method(s): _____
Results: _____

☐ **CHEMICALS**
☐ **WMDs** Screening Method(s): _____
Results: _____
☐ **pH** Results: _____
☐ **OTHER** Screening Method(s): _____
Results: _____

☐ **OTHER**
Screening Method(s): _____
Results: _____

SAMPLE SCREENED BY:
(Fill out this section completely)
Technician Name(s): _____ Organization(s): _____
Address: _____ Telephone(s): _____
COLLECTOR/SUBMITTER INCIDENT IDENTIFIER #: _____

The Hinton State Laboratory Institute does not accept explosive or incendiary material or chemical WMD.
Call the 24/7 Laboratory Emergency Cell phone at 617-590-6390 for guidance.

Questions regarding the submission and analysis of explosive or incendiary material should be referred to the Arson & Explosives Unit at the State Crime Lab
8 am – 5 pm Telephone ~ 508-358-3220 / After Hours Pager ~ 508-899-3770 / After Hours Cell ~ 508-241-2052

Page 2 of 2 / SS-LP-1-01 Last Updated Jul. 2008 www.mass.gov/dph/bls/

FIG. X1.1 Example Specimen Screening Form

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017

YM95NC Attachment -- Page 335
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

TABLE X1.2 Example of Sample Collection Sheet

Sample Number or Sample Identifier: _____

Date/Time of Sample: _____

Type of Sample: _____

Type/texture of surface sampled: _____

Description of Material Sampled (e.g., color, texture, homogeneity etc.):

Name of Persons Collecting Sample:

Sampler

Printed Name: _____

Signature: _____

Phone Number: _____

Facilitator

Printed Name: _____

Signature: _____

Phone Number: _____

Measured Size of Area Sampled: _____

Sample Location (include agency, address, room number, description of sample location):

_____

_____

_____

_____

Map of Sample Area:

Other Comments:

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017

Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

1. NAME OF SAMPLE COLLECTOR		2. LOCATION OF SAMPLE COLLECTION ADDRESS (CITY, ST, ZIP)			
3. REASON OBTAINED		4. TIME/DATE OBTAINED			
5. ITEM #	6. QUANTITY	7. DESCRIPTION OF SAMPLE (Liquid, Solid, Color, etc)			
8. ITEM #	9. QUANTITY	10. DESCRIPTION OF SAMPLE PACKAGING			
12. ITEM #	13. DATE	14. RELEASED BY	15. RECEIVED BY	16. PURPOSE OF CHANGE OF CUSTODY	17. SHIPMENT DESCRIPTION
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		

FIG. X1.2 Example of Chain of Custody Form

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017

Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

14. ITEM #	15. DATE	16. RELEASED BY	17. RECEIVED BY	18. PURPOSE OF CHANGE OF CUSTODY	19. SHIPMENT DESCRIPTION
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		
		Signature	Signature		
		Print: Name, Grade, Title	Print: Name, Grade, Title		

20. BACKGROUND INFORMATION

a. Wind Speed	b. Wind Direction (from)	c. Temperature	d. Humidity	e. Visibility	f. Terrain	g. Other Remarks

21. FINAL DISPOSAL ACTION

RELEASE TO OWNER OR OTHER (Name/Unit) _____

DESTROY _____

OTHER (Specify) _____

22. FINAL DISPOSAL AUTHORITY

ITEM(S) _____ ON THIS DOCUMENT, PERTAINING TO THE INVESTIGATION INVOLVING

(Grade)

_____ (IS) (ARE) NO LONGER

(Name) (Organization)

REQUIRED AS EVIDENCE AND MAY BE DISPOSED OF AS INDICATED ABOVE. (If article(s) must be retained, do not sign, but explain in separate correspondence.)

(Typed/Printed Name, Grade, Title) (Signature) (Date)

23. WITNESS TO DESTRUCTION OF EVIDENCE

THE ARTICLE(S) LISTED AT ITEM NUMBER(S) _____ (WAS) (WERE) DESTROYED BY THE EVIDENCE CUSTODIAN, IN MY PRESENCE, ON THE DATE INDICATED ABOVE.

(Typed/Printed Name, Organization) (Signature)

FIG. X1.2 Example of Chain of Custody Form *(continued)*

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017

Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

New York State Department of Health
Wadsworth Center
Biodefense Laboratory
120 New Scotland Avenue
Albany, NY 12208 Phone (518) 474-4177

Laboratory Response Network
Biothreat Tracking Form

Specimen Information
Incident ID#_____ **Collection Date/Time**_____
Investigating agency and contact information_____
Collection County_____
Collection Site (address)_____
Targeted Individual's name (if any)_____
Specimen Description_____

Collection Site Information

Building evacuated	no ☐	yes ☐	details _____
Ventilation system shut down	no ☐	yes ☐	details _____
Site/building locked-down	no ☐	yes ☐	details _____
Media on-site	no ☐	yes ☐	details _____
Medical response initiated	no ☐	yes ☐	details _____

Credible Biohazard Assessment Criteria

Stated or implied Threat	no ☐	yes ☐	describe _____
Visible, testable Material	no ☐	yes ☐	describe _____
Uncertain or suspicious Origin	no ☐	yes ☐	describe _____
Exposure or illness			
Targeted individual	no ☐	yes ☐	illness ☐ _____
First responders	no ☐	yes ☐	illness ☐ _____
Sample collectors	no ☐	yes ☐	illness ☐ _____

Credible Biohazard Assessment performed by _____
Field Hazard Screens performed by_____
Explosive Device negative ☐ Instrument used_____
Chemical Hazard negative ☐ Instrument used_____
Rad/Nuc Hazard negative ☐ Instrument used_____
ALL samples must be NEGATIVE by ALL Screens to be accepted at testing laboratory

Sample Collected by_____
Sample Container Decontaminated by_____
UNYRIC notified (by whom)_____ **date/time** _____
NYSPIN BIO1 submitted (by whom)_____ **date/time** _____

Submitter Information
Name _____
Phone ( ) - _____ After hours ( ) - _____ Message OK? yes ☐ no ☐

Report Results To (if different than Submitter)
Name _____
Phone ( ) - _____ After hours ( ) - _____ Message OK? yes ☐ no ☐

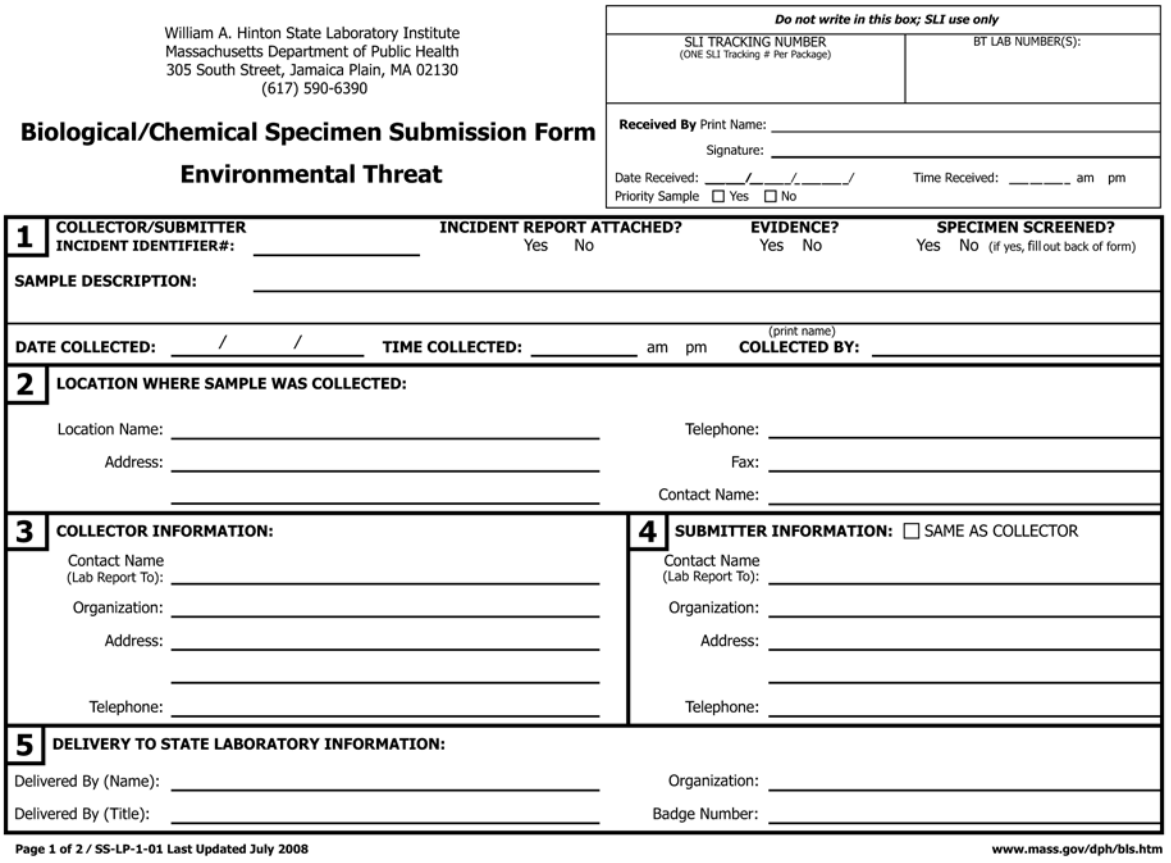
Incomplete information reporting on this tracking form or a failure to conduct hazard screening procedures as outlined will result in this sample being refused at the laboratory and/or returned to the submitting agency.

White and Yellow copies to lab *Pink copy to law enforcement* *Blue copy to local health department*

DOH 4348 (04/05)

FIG. X1.3 Example of Biological Agent Tracking Form

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.



William A. Hinton State Laboratory Institute
Massachusetts Department of Public Health
305 South Street, Jamaica Plain, MA 02130
(617) 590-6390

Biological/Chemical Specimen Submission Form

Environmental Threat

Do not write in this box; SLI use only

SLI TRACKING NUMBER (ONE SLI Tracking # Per Package) | BT LAB NUMBER(S):

Received By Print Name: _____
Signature: _____
Date Received: ____/____/____/____/ Time Received: _____ am pm
Priority Sample ☐ Yes ☐ No

1 COLLECTOR/SUBMITTER INCIDENT IDENTIFIER#: _____ **INCIDENT REPORT ATTACHED?** Yes No **EVIDENCE?** Yes No **SPECIMEN SCREENED?** Yes No (if yes, fill out back of form)

SAMPLE DESCRIPTION: _____

DATE COLLECTED: ____/____/____ **TIME COLLECTED:** _____ am pm **COLLECTED BY:** _____ (print name)

2 LOCATION WHERE SAMPLE WAS COLLECTED:

Location Name: _____ Telephone: _____
Address: _____ Fax: _____
_____ Contact Name: _____

3 COLLECTOR INFORMATION:

Contact Name (Lab Report To): _____
Organization: _____
Address: _____
_____
Telephone: _____

4 SUBMITTER INFORMATION: ☐ SAME AS COLLECTOR

Contact Name (Lab Report To): _____
Organization: _____
Address: _____
_____
Telephone: _____

5 DELIVERY TO STATE LABORATORY INFORMATION:

Delivered By (Name): _____ Organization: _____
Delivered By (Title): _____ Badge Number: _____

Page 1 of 2 / SS-LP-1-01 Last Updated July 2008 www.mass.gov/dph/bls.htm

FIG. X1.4 Example of Specimen Submission Form

ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.

This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or service@astm.org (e-mail); or through the ASTM website (www.astm.org). Permission rights to photocopy the standard may also be secured from the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Tel: (978) 646-2600; http://www.copyright.com/

Copyright by ASTM Int'l (all rights reserved); Wed Jun 21 16:43:00 EDT 2017
Downloaded/printed by
U.S. DEPT OF COMMERCE-NIST (U.S. DEPT OF COMMERCE-NIST) pursuant to License Agreement. No further reproductions authorized.

On this Page

- Macrofoam Swab Procedure
- Cellulose Sponge Procedure
- Gauze Procedure
- Blanks
- Decontamination
- Sample Shipment

Surface sampling procedures for *Bacillus anthracis* spores from smooth, non-porous surfaces

Revised April 26, 2012

General

These sampling procedures were prepared by the Centers for Disease Control and Prevention (CDC) to provide standardized methods for industrial hygienists, or other trained samplers under the direction of sampling experts, to use when sampling for *Bacillus anthracis* (*B. anthracis*) spores. These procedures supersede previous CDC procedures for collecting environmental samples for *B. anthracis*, including *Comprehensive Procedures for Collecting Environmental Samples for Culturing Bacillus anthracis*, which were developed during the 2001 anthrax terrorist events. As stated in that guidance, CDC planned to update the guidance as new information becomes available. In particular, one major change is the recent efforts by CDC to validate methods for the laboratory processing and analysis of *B. anthracis* spores.

These procedures are meant to be used for collection of samples on smooth, non-porous surfaces and can be used in both indoor and outdoor environments. Examples of non-porous surfaces are stainless steel, painted wall board, floor tile, or wood laminate. Each sampling method has its specific uses and advantages. Sampling methods must be coordinated with the laboratory to ensure that they are ready to accept and process all the samples. This is particularly important if deviating from the validated sampling procedures.

Swabs are appropriate for sampling small surfaces or hard to reach locations of less than 4 inches square (in^2), like crevices, corners, supply air diffusers, air return grills, and hard-to-reach places. Wipes are appropriate for sampling larger non-porous surfaces including walls, desks, and floors. An interagency effort known as the Validated Sampling Plan Working Group (VSPWG) is currently developing a document titled *Reference Guide for Developing and Executing Bacillus anthracis Sampling Plans in Indoor Settings*, which outlines approaches and methodologies to characterize and guide remediation of indoor sites with potential or actual contamination. The VSPWG document, once published, should be consulted for additional information about strategies and guidance for sampling *B. anthracis*. This document will also assist with the interpretation of results for samples collected with macrofoam swabs and cellulose sponges on smooth, non-porous surfaces following these collection procedures.

Note:

Additional information regarding Bacillus anthracis sampling, including recommendations for protecting investigators while sampling, are available at emergency.cdc.gov/agent/anthrax/environment/ (http://emergency.cdc.gov/agent/anthrax/environment/) and www.cdc.gov/niosh/docs/2009-132/default.html (http://www.cdc.gov/niosh/docs/2009-132/default.html).

The collection of samples associated with a crime scene or suspicion of a criminal event should be coordinated with law enforcement authorities. Sampling teams need to be aware that samples collected may be or become criminal evidence, and certain additional procedures associated with sample collection will be necessary.

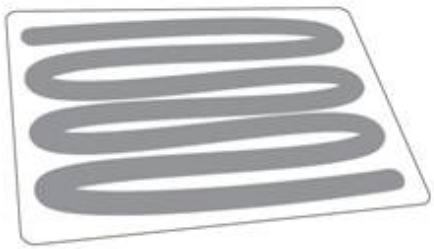
Macrofoam Swab Procedure

Swab Materials

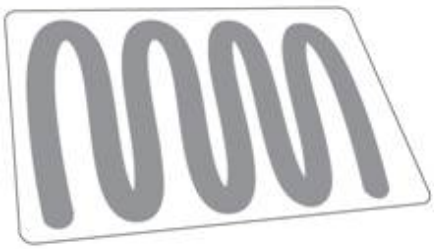
1. Gloves, nitrile
2. Ruler, disposable, and masking tape
   or
   Sample template, disposable, sample area size 4 in^2 (26 cm^2)
3. Macrofoam swab, sterile, 3/16 inch thick medical-grade polyurethane foam head, 100 pores per inch, thermally bonded to a polypropylene stick (such as the Sterile Foam Tipped Applicators Scored with Thumb Stop [Puritan, Guilford, Maine; catalog number 25-1607 1PF SC] or equivalent)
4. General neutralizing buffer that will inactivate halogen disinfectants and quaternary ammonium compounds, 10 milliliter (mL), sterile (such as the Neutralizing Buffer [Hardy Diagnostics, Santa Maria, California; catalog number K105] or equivalent)
5. Screw-cap centrifuge tubes, sterile, 15 mL (such as 15 mL High-Clarity Polypropylene Conical Centrifuge Tube [Becton Dickinson, Franklin Lakes, New Jersey; catalog number 352097] or equivalent)
6. Sample labels
   or
   Permanent marker
7. Re-sealable plastic bag, 1-quart or smaller
8. Re-sealable plastic bag, 1-gallon or larger

Swab Sampling Procedure

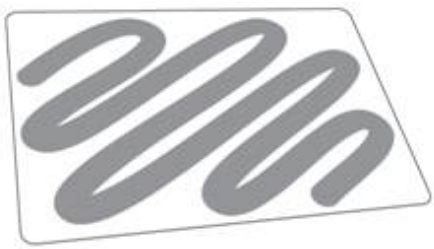
1. Wearing a clean pair of gloves over existing gloves, place the disposable template over the area to be sampled and secure it. If the template cannot be used, measure the sampling area with a disposable ruler, and delineate the area to be sampled with masking tape.
2. Remove the sterile swab from its package. Grasp the swab near the top of the handle. Do not handle below the thumb stop.
3. If the sterile swab is not pre-moistened, moisten the sterile swab by dipping it in the 10 mL container of neutralizing buffer solution. Remove any excess liquid by pressing the swab head on the inside surface of the neutralizing buffer solution container.
   Note: Once a sterile swab has been moistened, the remaining neutralizing buffer solution and container must be discarded.
4. Swab the surface to be sampled using the moistened sterile swab. Use an overlapping 'S' pattern to cover the entire surface with horizontal strokes.



   Note: Depending on the design of the swab, a rolling motion can be used when swabbing the surface to maximize swab contact with the surface.
5. Rotate the swab and swab the same area again using vertical 'S'-strokes.



6. Rotate the swab once more and swab the same area using diagonal 'S'-strokes.



7. Place the head of the swab directly into a sterile screw-capped centrifuge tube. Break off the head of the swab by bending the handle. The end of the swab handle, touched by the collector, should not touch the inside of the tube. Securely tighten the screw-cap and label the tube (e.g., unique sample identifier, sample location, initials of collectors and date and time sample was collected). Collection tubes and re-sealable bags may be pre-labeled to assist with sampling efficiency.

8. Place the sample container in a re-sealable 1-quart plastic bag. Securely seal and label the bag (e.g., sample location, date and time sample was collected, and name of individual collecting the sample).
   Note: Remove excessive air from the re-sealable plastic bags to increase the number of samples that can be shipped in one container.
9. Dispose of the template, if used.
10. Remove outer gloves and discard. Clean gloves must be worn for each new sample.

Cellulose Sponge Procedure

Cellulose Sponge Materials

1. Gloves, nitrile
2. Ruler, disposable, and masking tape
   or
   Sample template, disposable, sample area size 100 in^2 (645 cm^2)
3. Sponge, sterile, pre-moistened with 10 mL neutralizing buffer solution, 1.5 by 3 inches cellulose sponge folded over a handle (such as the 3M™ Sponge-Stick [3M, St. Paul, Minnesota; catalog number SSL-10NB] or equivalent)[1 (http://www.cdc.gov/niosh/topics/emres/surface-sampling-bacillus-anthracis.html#)]
   or
   Sponge, sterile, dry, 1.5 by 3 inches cellulose sponge folded over a handle (such as the 3M™ Sponge-Stick [3M, St. Paul, Minnesota; catalog number SSL-100] or equivalent)
   and
   General neutralizing buffer that will inactivate halogen disinfectants and quaternary ammonium compounds, sterile, 10 mL (such as the Neutralizing Buffer [Hardy Diagnostics, Santa Maria, California; catalog number K105] or equivalent)
4. Screw-cap specimen container, sterile, individually wrapped 4 ounce (such as General Purpose Specimen Container [Kendall Healthcare, Mansfield, Massachusetts; catalog number 8889-207026] or equivalent)
5. Sample labels
   or
   Permanent marker
6. Re-sealable plastic bag, 1-quart or smaller
7. Re-sealable plastic bag, 1-gallon or larger

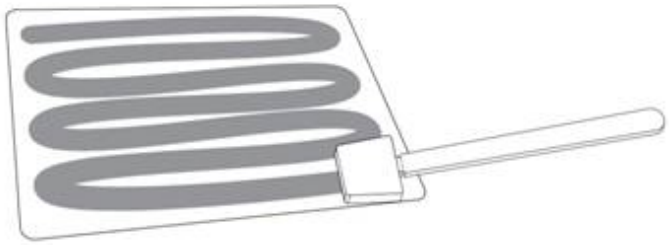
Cellulose Sponge Sampling Procedure

1. Wearing a clean pair of gloves over existing gloves, place the disposable template over the area to be sampled and secure it. If a template cannot be used, measure the sampling area with a disposable ruler, and delineate the area to be sampled with masking tape. The surface area sampled should be less than or equal to 100 in^2 (645 cm^2).
2. Remove the sterile sponge from its package. Grasp the sponge near the top of the handle. Do not handle below the thumb stop.
3. If the sterile sponge is not pre-moistened, moisten the sponge by pouring the 10 mL container of neutralizing buffer solution over the dry sponge.

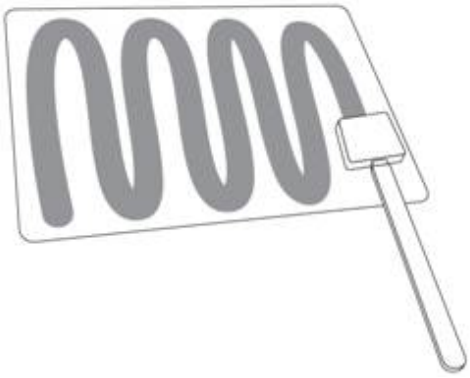
Note: The moistened sponge should not be dripping neutralizing buffer solution.

Note: Any unused neutralizing buffer solution **must** be discarded.

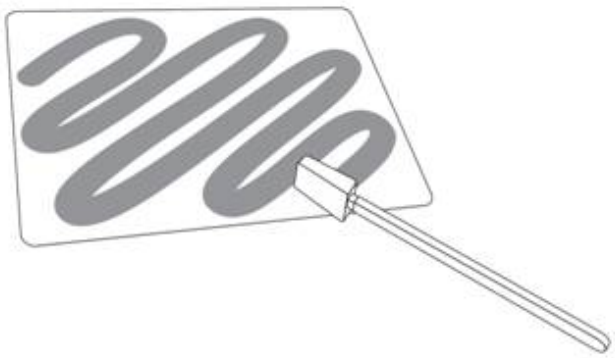
4. Wipe the surface to be sampled using the moistened sterile sponge by laying the widest part of the sponge on the surface, leaving the leading edge slightly lifted. Apply gentle but firm pressure and use an overlapping 'S' pattern to cover the entire surface with horizontal strokes.



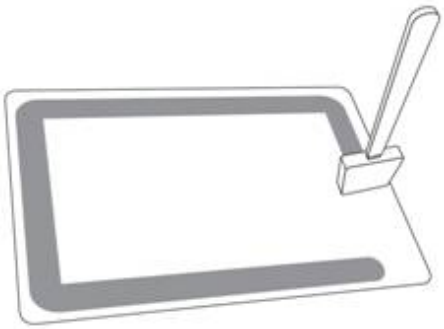
5. Turn the sponge over and wipe the same area again using vertical 'S'-strokes.



6. Use the edges of the sponge (narrow sides) to wipe the same area using diagonal 'S'-strokes.



7. Use the tip of the sponge to wipe the perimeter of the sampling area.



8. Place the head of the sponge directly into a sterile specimen container. Break off the head of the sponge by bending the handle. The end of the sponge handle, touched by the collector, should not touch the inside of the specimen container. Securely seal and label the container (e.g., unique sample identifier, sample location, initials of collector and date and time sample was collected).

9. Place the sample container in a re-sealable 1-quart plastic bag. Securely seal and label the bag (e.g., sample location, date and time sample was collected, and name of individual collecting the sample). Specimen containers and re-sealable bags may be pre-labeled to assist with sampling efficiency.
   Note: Remove excessive air from the re-sealable plastic bags to increase the number of samples that can be shipped in one container.
10. Dispose of the template, if used.
11. Remove outer gloves and discard. Clean gloves should be worn for each new sample.

Gauze Procedure

Gauze Materials

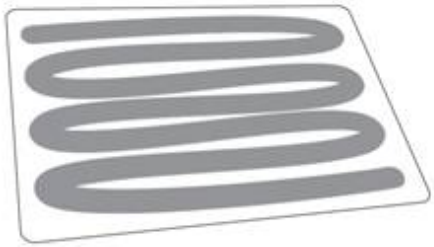
Note: This sampling and analytical method has not been validated by CDC. A standard sampling procedure is provided in the event that the macrofoam swab or cellulose sponge methods cannot be utilized.

1. Gloves, nitrile
2. Gloves, sterile, nitrile
3. Ruler, disposable, and masking tape
   or
   Sample template, disposable, sample area of 144 in^2 (929 cm^2)
4. Gauze, sterile, non-cotton, polyester blend sponge or rayon/polyester blend, 2 by 2 in (such as the Versalon All-Purpose Sponge [Kendall Healthcare, Mansfield, Massachusetts; catalog number 8042; includes two gauze squares/packet] or equivalent)
5. General neutralizing buffer that will inactivate halogen disinfectants and quaternary ammonium compounds solution, 10 mL, sterile (such as the Neutralizing Buffer [Hardy Diagnostics, Santa Maria, California; catalog number K105] or equivalent)
6. Pipette, 5 mL, sterile, individually wrapped (such as the Greenwood Products' Sterile 5mL Standard Transfer Pipette [Greenwood Products, Inc., Middlesex, New Jersey; catalog number GS137038] or equivalent)
7. Screw-cap specimen container, 4 ounce, sterile, individually wrapped (such as General Purpose Specimen Container [Kendall Healthcare, Mansfield, Massachusetts; catalog number 8889-207026] or equivalent)
8. Sample labels
   or
   Permanent marker
9. Re-sealable plastic bag, 1-quart or smaller
10. Re-sealable plastic bag, 1-gallon or larger

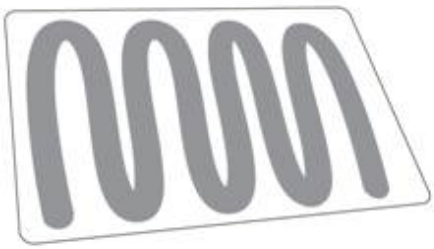
Gauze Sampling Procedure

1. Wearing a pair of gloves over existing gloves, place the disposable template over the area to be sampled and secure it. If the template cannot be used, measure the sampling area with a disposable ruler, and delineate the area (144 $in^2$to be sampled with masking tape.
2. Partially peel open the sterile gauze package carefully exposing the gauze.
   Note: The sterile gauze should not be touched without sterile gloves.

3. Measure 5 mL of neutralizing buffer solution from the 10 mL container using a disposable pipette and apply to sterile gauze in its original packaging. Remove outer gloves.
   Note: The moistened gauze should not be dripping neutralizing buffer solution.
   Note: Any unused neutralizing buffer solution and the pipette **must** be discarded.
4. Don a pair of sterile gloves.
   Note: Sterile gloves are required when sampling with gauze because of the direct contact with the sampling media.
5. Remove one of the sterile gauze (if two per package) and dispose of or retain the other gauze as a field blank (see section 4.1).
6. Completely unfold the remaining moistened sterile gauze, and then fold in half.
7. Wipe the surface to be sampled using the moistened sterile gauze, fingertips should be held together and apply gentle but firm pressure. Use an overlapping 'S' pattern to cover the entire surface with horizontal strokes.



8. Fold the exposed side of the gauze in and wipe the same area again using vertical 'S'-strokes.



9. Fold the exposed side of the gauze in once more and wipe the same area using diagonal 'S'-strokes.

10. Fold the gauze, exposed side in, and place it into a sterile screw-cap specimen container.
11. Securely tighten the srew-cap and label the container (e.g., unique sample identifier, sample location, initials of the collectors and date and time sample was collected).
12. Place the sample container into a re-sealable 1-quart plastic bag. Securely seal and label the bag (e.g., sample location, date and time sample was collected, and name of individual collecting the sample). Specimen containers and resealable bags may be pre-labeled to assist with sampling efficiency. Note: Remove excessive air from the re-sealable plastic bags to increase the number of samples that can be shipped in one container.
13. Dispose of the template, if used.
14. Remove outer gloves and discard. Clean sterile gloves should be worn for each new sample.

Blanks

Field Blanks

Field blanks are samples handled exactly the same as those used to collect field samples, except they are not allowed to come into contact with the contaminated surface. It is used to estimate contamination arising from preparation for sampling, sampler technique, and shipment and storage prior to analysis. The number of field blanks collected should be equal to at least 10% of the samples collected. Field blanks should be collected while in the contaminated area. While wearing clean (sterile gloves for handling gauze), the macrofoam swab, cellulose sponge, or gauze should be removed from its packaging, moistened (as needed, see sections 3.1, 3.2 and 3.3), and then placed in the appropriate container (either a centrifuge tube or specimen container). An aliquot of the unused portion of the opened neutralizing buffer solution should also be collected when using macrofoam swab, cellulose sponge, or gauze media that are not pre-moistened.

Media Blanks

Media blanks are unexposed samples used for background correction of sample readings or for recovery studies. Provide two unopened sample media (macrofoam swab, cellulose sponge, or gauze) per lot used and provide two unopened, unused samples of the neutralizing buffer solution (if not using pre-moistened media) as media blanks to the processing laboratory.

Decontamination

Sample Bag Decontamination

1. Place multiples of the re-sealable 1-quart plastic bags into a 1-gallon re-sealable plastic bag. Securely seal the 1-gallon re-sealable plastic bag and label the bag (e.g., identify samples contained in the re-sealable plastic bag, sample locations, date and time samples were collected, and name of individual collecting the samples).
2. Decontaminate the outer surface of the larger re-sealable plastic bag using a fresh pH-adjusted bleach solution (household bleach diluted 1:9; pH-adjusted to 6.8-8.0) with a 10-minute contact time before the re-sealable plastic bag leaves the contaminated area. This solution can be prepared by:
   Step 1: Mixing one part household bleach (5.25 to 6.0 % sodium hypochlorite) with 5 parts water (v/v);
   Step 2: Adding 1 part white vinegar; and
   Step 3: Adding 3 parts of additional water.
   Note: Additional information about decontamination is available at www.epa.gov/opp00001/factsheets/chemicals/bleachfactsheet.htm (http://www.epa.gov/opp00001/factsheets/chemicals/bleachfactsheet.htm) .
3. Thoroughly dry the outside of the re-sealable plastic bag.
4. Complete a chain of custody form.
   Note: Once the outer re-sealable plastic bag is decontaminated, it should not be opened outside of appropriate containment in a laboratory.
5. Place the larger re-sealable plastic bag into an appropriate container for shipping (See Sample Shipment section).

Decontamination of Bags Containing Documentation

1. Place sample sheets and other documentation in a separate re-sealable plastic bag. The sheets should be placed two to a bag with the face of each sheet facing out. Securely seal and label the bag (e.g., corresponding sample locations, date and time samples were collected, and name of individual collecting the samples).
   Note: It is important to only write on one side of the paper, the face, when collecting information.
2. Decontaminate the outer surface of the larger re-sealable plastic bag using a fresh pH-adjusted bleach solution (household bleach diluted 1:9; pH-adjusted to 6.8-8.0) with a 10-minute contact time before the re-sealable plastic bag leaves the contaminated area. This solution can be prepared by:
   Step 1: Mixing one part household bleach (5.25 to 6.0 % sodium hypochlorite) with 5 parts water (v/v);
   Step 2: Adding 1 part white vinegar; and
   Step 3: Adding 3 parts of additional water.
   Note: Additional information about decontamination is available at www.epa.gov/opp00001/factsheets/chemicals/bleachfactsheet.htm (http://www.epa.gov/opp00001/factsheets/chemicals/bleachfactsheet.htm) .
3. Thoroughly dry the outside of the re-sealable plastic bag.
4. Complete a chain of custody form.
   Note: Once the outer re-sealable plastic bag is decontaminated, it should not be opened outside of appropriate containment.
5. Place the larger re-sealable plastic bag into an appropriate container for shipping.

Sample Shipment

1. Transport all samples to the processing laboratory on wet ice or on cold packs.
   Note: Samples may be stored at 2°C–8°C prior to processing and should be processed within 48 hours of collection.
2. Appropriate chain of custody forms and analytical request forms should be included with each shipment sent to the processing laboratory. Containers used to transport the samples and accompanying contaminated documentation and equipment should be prepared and shipped according to the appropriate regulations for transporting infectious. The most current Code of Federal Regulations, International Air Transport Association guidelines, and other appropriate regulator or guidance publications should be consulted for compete instructions. The shipper is responsible for ensuring adherence to the most current and appropriate regulations.
   Note: Do not transport contaminated equipment/supplies in the same container as the samples.

Related Topics

Emergency Response Health Monitoring & Surveillance

Follow NIOSH

Facebook (http://www.facebook.com/NIOSH)

Flickr (http://www.flickr.com/photos/NIOSH)

Pinterest (http://www.pinterest.com/cdcgov/workplace-safety-and-health/)

Twitter (http://twitter.com/NIOSH)

YouTube (http://www.youtube.com/user/NIOSHSafetyVideos)

NIOSH Homepage (http://www.cdc.gov/niosh/)

NIOSH A-Z (http://www.cdc.gov/niosh/az/a.html)

Workplace Safety & Health Topics (http://www.cdc.gov/niosh/topics/)

Publications and Products (http://www.cdc.gov/niosh/pubs/)

Programs (http://www.cdc.gov/niosh/programs.html)

Contact NIOSH (http://www.cdc.gov/niosh/contact/)

Page last reviewed: January 30, 2012

Page last updated: October 12, 2016

Content source: National Institute for Occupational Safety and Health (https://www.cdc.gov/NIOSH/) Office of the Director

Error processing SSI file

Copyright © 1996 by the Genetics Society of America

Saccharomyces cerevisiae S288C Has a Mutation in *FLO8*, a Gene Required for Filamentous Growth

Haoping Liu,* Cora Ann Styles† and Gerald R. Fink†

**Department of Biological Chemistry, College of Medicine, University of California, Irvine, California 92717 and †Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142*

Manuscript received May 15, 1996
Accepted for publication August 8, 1996

ABSTRACT

Diploid strains of baker's yeast *Saccharomyces cerevisiae* can grow in a cellular yeast form or in filaments called pseudohyphae. This dimorphic transition from yeast to pseudohyphae is induced by starvation for nitrogen. Not all laboratory strains are capable of this dimorphic switch; many grow only in the yeast form and fail to form pseudohyphae when starved for nitrogen. Analysis of the standard laboratory strain S288C shows that this defect in dimorphism results from a nonsense mutation in the *FLO8* gene. This defect in *FLO8* blocks pseudohyphal growth in diploids, haploid invasive growth, and flocculation. Since feral strains of *S. cerevisiae* are dimorphic and have a functional *FLO8* gene, we suggest that the *flo8* mutation was selected during laboratory cultivation.

SACCHAROMYCES cerevisiae is a dimorphic fungus capable of interconverting between a unicellular yeast form and a multicellular filamentous form called a pseudohypha (HANSEN 1886; GIMENO *et al.* 1992). Pseudohyphae are induced in *MAT***a**/α diploid cells by starvation for nitrogen in the presence of a fermentable carbon source (GIMENO *et al.* 1992). Pseudohyphal cells exhibit at least five distinctive characteristics: They have an elongated cell shape, fail to separate after cell division, divide in a unipolar manner (the first and subsequent buds emerge from the free ends), invade beneath the surface of the agar, and have a distinct cell cycle that appears to lack the extended G1 phase typical of yeast form cells (KRON *et al.* 1994).

Elements of the yeast mating signal transduction pathway Ste20p(PAK), Ste11p(MEKK), Ste7p(MEK), and the transcription factor Ste12p are required for pseudohyphal growth as well as for mating (LIU *et al.* 1993). Mutations in the genes encoding these proteins of the mitogen-activated protein (MAP) kinase cascade block the formation of filaments. Neither of the MAP kinases, Fus3p or Kss1p, appears to be required for pseudohyphal growth because the *fus3kss1* double mutant forms normal pseudohyphae under conditions of nitrogen starvation. Recently, it has been shown that this filamentation kinase cascade is activated by the RAS2 gene (MOSCH *et al.* 1996). A number of other genes have been reported to affect pseudohyphal growth including *SHR3* (LJUNGDAHL *et al.* 1992), *GRR1* (BARRAL *et al.* 1995; BARRAL and MANN 1995), *PHD1-7* (GIMENO and FINK 1994), *ELM1* (BLACKETER *et al.* 1993), and *PPS1* (BLACKETER *et al.* 1994). However, the relationship of these genes to the kinase cascade has not yet been elucidated. The MAP kinase cascade appears to control development in other fungi as well; Cph1p, a Ste12p homologue has been implicated in the regulation of hyphal growth in *Candida albicans*, the most common fungal pathogen in humans (LIU *et al.* 1994). Similarly, Fuz7p, a Ste7p homologue, is involved in filamentation in *Ustilago maytis* (BANUETT and HERSKOWITZ 1994).

Haploid strains of Saccharomyces exhibit an invasive growth behavior with many similarities to pseudohyphal development, including filament formation and agar penetration. The same components of the MAP kinase cascade necessary for diploid pseudohyphal development (*STE20, STE11, STE7,* and *STE12*) are also required for haploid invasive growth (ROBERTS and FINK 1994). Despite the similarities between diploid pseudohyphae and haploid filamentous growth, there are several differences. Haploids invade well on rich YPD medium whereas diploids do not. Moreover, diploids but not haploids form pseudohyphae under conditions of nitrogen starvation.

Although many feral strains of Saccharomyces are capable of forming filaments, a number of laboratory strains dubbed "wild type" are incapable of forming pseudohyphae or of haploid invasion. In this report we examine the genetic basis for this variation between different laboratory strains. The genetic and molecular analysis of a strain that can form pseudohyphae (Σ1278b) with one that can not (S288C) shows that S288C contains a "naturally occurring mutation" in the *FLO8* gene, which prevents both pseudohyphal development and haploid invasion.

MATERIALS AND METHODS

Media and yeast strains: Standard yeast media were prepared and yeast genetic manipulations were performed as

Corresponding author: Gerald R. Fink, Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142.

YM95NC Attachment -- Page 351

TABLE 1

Yeast strains used in this study

Yeast strains	Genotype	Background and source
CGx68	**a**/α	Σ1278b, GIMENO *et al.* (1992)
CG25	**a** *ura3-52 shr3-102*	Σ1278b, GIMENO *et al.* (1992)
CG41	α *ura3-52 shr3-102*	Σ1278b, GIMENO *et al.* (1992)
CGx52	**a**/α *ura3-52/ura3-52 shr3Δ1::URA3/shr3Δ1::URA3*	Σ1278b, LJUNGDAHL *et al.* (1992)
L4973	α *ura3-52 ade2 lys2 shr3Δ1::URA3 flo8-1*	S288C, *shr3Δ1::URA3* transformant in AA255 RUDOLPH *et al.* (1989)
F762	**a** *trp1Δ1 ura3-52 flo8-1*	S288C, gift from JIM THOMAS
10362-6A	**a** *shr3Δ1::URA3 ura3-52 ade2 flo8-1*	S288C, ascospore of L4973 × F762
10362-3A	α *shr3Δ1::URA3 ura3-52 flo8-1*	S288C, ascospore of L4973 × F762
L5809	**a**/α *shr3Δ1::URA3/shr3Δ1::URA3 flo8-1/flo8-1 ura3-52/ura3-52 ade2/ADE2*	10362-6A × 10362-3A
X10375	**a**/α *shr3Δ1::URA3/shr3-102 ura3-52/ura3-52 ade2/ADE2 flo8-1/FLO8*	10362-6A × CG25
10375-6B	α *ura3-52 shr3-102 flo8-1*	S288C and Σ1278b, ascospore of X10375
10415-8D	**a** *ura3-52 shr3-102 flo8-1*	S288C and Σ1278b
L5281	**a** *ura3-52 shr3-102 flo8-1::URA3::FLO8*	10415-8D transformed with pHL11
X10397	**a**/α *ura3-52/ura3-52 shr3-102/shr3-102 flo8-1/FLO8*	10375-6B × CG25
10397-9B	**a** *flo8-1 shr3-102 ura3-52*	S288C and Σ1278b, ascospore of X10397
10397-7D	α *flo8-1 shr3-102 ura3-52*	S288C and Σ1278b, ascospore of X10397
L5107	**a**/α *flo8-1/flo8-1 shr3-102/shr3-102 ura3/ura3-52*	10397-9B × 10397-7D
L5810	**a**/α *flo8-1/flo8-1 shr3-102/shr3-102 ura3/ura3-52 (pRS316)*	L5107 + pRS316
L5811	**a**/α *flo8-1/flo8-1 shr3-102/shr3-102 ura3/ura3-52 (pHL1)*	L5107 + pHL1
L5812	**a**/α *flo8-1/flo8-1 shr3-102/shr3-102 ura3/ura3-52 (pHL36)*	L5107 + pHL36
FY3	**a** *ura3-52 flo8-1 GAL*$^+$	S288C, WINSTON *et al.* (1995)
FY2	α *ura3-52 flo8-1 GAL*$^+$	S288C, WINSTON *et al.* (1995)
L5302	**a** *flo8-1::URA3::FLO8 ura3-52 GAL+*	S288C
L5304	α *flo8-1::URA3::FLO8 ura3-52 GAL+*	S288C
L5813	**a**/α *ura3-52/ura3-52 flo8-1/flo8-1 GAL+*	FY3 × FY2
L5814	**a**/α *ura3-52/ura3-52 flo8-1/flo8-1 GAL+ (URA3)*	L5813+ pRS316
L5306	**a**/α *flo8-1::URA3::FLO8 ura3-52/flo8-1::URA3::FLO8 ura3-52*	L5302 × L5304
10459-4B	α *flo8-1::URA3::FLO8 ura3-52 his3Δ200*	S288C
16-1	**a** *spr6::HIS3::SPR6 his3 trp1 leu2 ura3 ilv2 met6 ade1(or ade2)*	KALLAL *et al.* (1990)
10480-5C	**a** *ura3-52*	Σ1278b
10480-5D	α *ura3-52*	Σ1278b
L5366	**a**/α *ura3-52/ura3-52*	Σ1278b, 10480-5C × 10480-5D
L5816	**a** *flo8-2 ura3-52*	Σ1278b
L5817	α *flo8-2 ura3-52*	Σ1278b
L5818	**a**/α *flo8-2/flo8-2 ura3-52/ura3-52*	L5816 × L5817
L5819	**a**/α *ura3-52/ura3-52 + (FLO8/URA3/2μ)*	L5366 + pHL135
DA1rho^{0}XJJM113	**a**/α *lys2/LYS2 ura3/URA3 ade2/ADE2* [*cox2-10*]	MULERO and FOX (1994), MULLER *et al.* (1984)
D273-10B/A1	α *met6*	SHERMAN; CORUZZI *et al.* (1978)
A364A	**a** *ade1 ade2 gal1 his7 lys2 tyr1 ura1*	HARTWELL.
L5682	**a**/α *ura3-1/ura3-1 can1-100/can1-100 eu2-3,112/LEU2 his3Δ200/HIS3 trp1-1/TRP1*	W303, F1575 × EY698, ELION *et al.* (1991a)
EY699	**a** *can1-100 ura3-1 leu2-3,112 his3-11.15 ade2-1 trp1-1*	W303, ELION *et al.* (1991a)
EM93	**a**/α	S288C progenitor, MORTIMER and JOHNSTON (1986)
Fleischmann's yeast	**a**/α	Specialty Brands, San Francisco, CA.
Red Star dry yeast	**a**/α	Universal Foods Corp., Milwaukee, WI.
Beer2	**a**/α	JAIME CONDE, Seville, Spain

described (SHERMAN *et al.* 1986). SLAD medium (synthetic low ammonia dextrose medium) was prepared as described (GIMENO *et al.* 1992).

The yeast strains used in this study are listed in Table 1. All Σ strains were derived from Σ1278b (also known as MB1000) and MB758-5B (BRANDRISS and MAGASANIK 1979; SIDDIQUI and BRANDRISS 1988). As documented in RESULTS, the pseudohyphal defect in S288C results from a mutation in the *FLO8* gene. The gene in our lab was originally designated *PHD10* before recent identification of this sequence as that of *FLO8* in the database. To avoid confusion, the *FLO8* acronym is used in all gene designations. The scoring of pseudohyphal growth in crosses was complicated by the segregation of modifiers, which sometimes made it difficult to distinguish *FLO8* from *flo8*. To aid in the scoring of tetrads (in the cross between S288C and Σ1278b) we used strains that carried the *shr3* mutation, an enhancer of pseudohyphal growth [either a spontaneous mutation *shr3-102* or an *in vitro*-constructed deletion *shr3Δ1::URA3* (GIMENO *et al.* 1992; LJUNGDAHL *et al.* 1992)]. Two different S288C strain backgrounds were used in this study. In the initial crosses between Σ1278b and S288C, we used a S288C strain (10362-6A) that carries a *shr3* mutation. 10362-6A was one of the ascospore progeny from a cross between L4973 and F762 (a gift from JIM THOMAS). As described in the RESULTS, 10362-6A differs from Σ1278b by confounding mutations in addition to *flo8* and cannot be restored

TABLE 2

Plasmids used in this study

Plasmid	Description	Source or reference
B3258	Σ yeast genomic library on pRS316	This study
pHL1	4.3-kb insert with *FLO8* on pRS316	This study
pHL36	>6-kb insert with *FLO8* on pRS316	This study
pHL11	*FLO8/YIP/URA3*	This study
pHL129	*flo8::hisGURA3KanRhisG*	This study
pHL130	*FLO8::myc3/URA3/CEN*	This study
pHL138	*FLO8::myc3/URA3/2μ*	This study
pHL132	*flo8::LacZ/URA3/2μ*	This study
pHL135	*FLO8/URA3/2μ*	This study
pCT3	S288C yeast genomic library	THOMPSON *et al.* (1993)
YEp357R	*LacZ/URA3/2μ*	MYERS *et al.* (1986)
pKB241	Triple myc epitope in pUC19	DANIEL KORNITZER and STEVE KRON
pSE1076	*hisGURA3Kan^{R}hisG*	STEVE ELLEDGE
pRS316	*URA3/CEN*	SIKORSKI and HIETER (1989)
pRS306	*URA3/YIP*	SIKORSKI and HIETER (1989)
pRS202	*URA3/2μ*	PHIL HIETER, unpublished data

to pseudohyphal growth simply by transformation with *FLO8*. However, ascospore derivatives from a Σ1278b × 10362-6A, which segregate only *flo8*, can be complemented by *FLO8* on a plasmid. After we obtained the *FLO8* gene on a DNA plasmid, we used another set of S288C strains (FY3 and FY2), which do not seem to have these confounding mutations and can be transformed to Phd$^+$ by *FLO8*.

We also examined other laboratory strains, some industrial strains, and EM93, a progenitor of S288C (MORTIMER and JOHNSTON 1986) to determine the presence of a *flo8* mutation and the ability to form filaments.

Construction of a Σ yeast genomic library: Genomic DNA was prepared from a Σ strain CGx68 and partially digested with *Sau3A*. DNA fragments were then size-fractionated over a gel filtration column (a prepacked G50 column from BRL) to collect fragments larger than 4 kb. The *Sau3A* ends were partially filled in with dATP and dGTP and ligated to dCTP- and dTTP-filled *Sal*I ends in the *URA3/CEN* plasmid pRS316 (SIKORSKI and HIETER 1989). Over 75,000 independent *Escherichia coli* transformants were obtained. Almost all the plasmids in the library have inserts, two-thirds of which carry inserts of 4–8 kb, and one-third of which have inserts of 2–4 kb.

Cloning of *FLO8* by complementation in a *flo8/flo8* diploid strain: Our Σ yeast genomic library was transformed into a diploid *flo8/flo8* strain (L5107) yielding ~9000 Ura$^+$ transformants. These Ura$^+$ transformants were collected, titered, and replated onto SLAD plates at a colony density of ~100–200 per plate. After 4–5 days of growth at 30°, ~7500 colonies were visually screened for pseudohyphal growth under a dissecting microscope. One clone (pHL1) from the screen complemented the *flo8/flo8* pseudohyphal growth defect of L5107 very well. In a second experiment, ~100,000 transformants were obtained on SC −Ura plates. These were pooled and replated onto SLAD plates at ~1000 per plate. We visually screened about 700,000 total colonies on SLAD medium, and isolated one clone (pHL36) that fully complemented the Phd$^-$ phenotype in L5107. Nine transformants that failed to complement L5107 fully were also isolated. Restriction digestions indicate that pHL1 and pHL36 contain overlapping DNA fragments, and the other nine isolates all carry another gene.

Plasmid constructions: Plasmids constructed in this study are listed in Table 2. pHL1 and pHL36 are the two original *FLO8* isolates from our Σ yeast genomic library. pHL1 has 4.8-kb insert and pHL36 has an ~6-kb insert. The pHL1 and pHL36 have an overlap of ~3 kb.

The *FLO8::URA3*/YIP plasmid (pHL11) was constructed by placing the insert from pHL1 (a *Xho*I-*Xba*I fragment) into the *Xho*I-*Xba*I site in the polylinker of pRS306 (SIKORSKI and HIETER 1989). For integrative transformation at *FLO8* the plasmid was linearized with *Bgl*II enzyme.

The *FLO8::LacZ* plasmid (pHL132) was constructed by cloning the *Xba*I-*Kpn*I fragment from pHL1 into a *LacZ/URA3/2μ* plasmid, YEp357R (MYERS *et al.* 1986). The *Xba*I site is in the polylinker and the *Kpn*I site is in *FLO8* coding region. This fragment contains 5′ flanking sequence and part of the coding sequence of *FLO8* fused in frame to the *LacZ* gene at the *Kpn*I site.

The *FLO8/URA3/2μ* plasmid (pHL135) was constructed by cloning the insert from pHL1 (a *Sac*I-*Xho*I fragment) into an *URA3/2μ* plasmid pRS202 (at the *Sac*I-*Xho*I site of the polylinker).

The *FLO8::myc3/URA3/CEN* plasmid (pHL130) was constructed by inserting a *Bam*HI triple myc (myc3) fragment from B2768 (a gift from Drs. D. KORNITZER and S. KRON) at the *Bgl*II site of the *FLO8* in pHL1. The orientation of the insertion and in-frame fusion of the myc3 in *FLO8* was confirmed by sequencing. A *FLO8::myc3/URA3/2μ* plasmid (pHL138) was created by subcloning the insert from pHL130 (a *Sac*I-*Xho*I fragment) into pRS202, a *URA3/2μ* plasmid.

A deletion of *FLO8* called *flo8-2* on plasmid pHL129 was constructed in two steps. First, the *Bst*EII to *Sna*BI region of the *FLO8* in pHL1 plasmid was replaced by the *Bam*HI-*Bgl*II fragment of *hisGURA3KanRhisG* from pSE1076 (a gift from Dr. STEVE ELLEDGE) by blunt-end ligation and subsequent selection of transformants for both ampicillin and kanamycin resistance. After isolating the Amp^{r}Kan^r transformants and the corresponding plasmid, we cut out the entire *flo8::hisG-URA3KanRhisG::flo8* region with *Sac*I and *Kpn*I and cloned it into a bluescript plasmid to form plasmid pHL129.

DNA sequence analysis of *FLO8*: Nested deletions (from the *Sac*I direction only) of the insert in pHL36 were generated by digestion with *Exo*III (HENIKOFF 1984) and single-strand nucleotide sequence was determined by the dideoxy chain termination method (SANGER *et al.* 1977). Based on this sequence, the appropriate primers were synthesized to sequence the complementary strand. Protein homology searches were performed via the World Wide Web by using

the BLAST Search Page at National Center for Biotechnology Information (ALTSCHUL *et al.* 1990). The GenBank accession number for *FLO8* is U51431.

DNA manipulations of *FLO8*: To integrate *FLO8* at the homologous chromosomal site, we cut pHL11 with *Bgl*II to linearize the plasmid at the *FLO8* coding region and then transformed yeast Ura$^-$ strains. Ura$^+$ yeast transformants were selected and integration by gene replacement was confirmed by Southern analysis. A strain carrying a null allele of *FLO8* was constructed by digesting plasmid pHL129 carrying *flo8-2* with *Kpn*I and *Sac*I to cut out the *flo8::hisGURA3kan^{R}hisG* fragment and then transforming Σ *FLO8* strains with the digested plasmid. Ura$^+$ transformants were first screened by PCR reactions for loss of the *FLO8* band and the putative *flo8-2* strains were confirmed by Southern blot analysis.

The *flo8* mutant allele present in S288C was cloned by gap repair. The autonomously replicating plasmid pHL1, containing *FLO8*, was digested with *Bst*EII and the digested DNA fragments were separated on an agarose gel from which the larger DNA fragment (containing gapped *FLO8* on the vector) was eluted. *Bst*EII, which cuts twice in *FLO8* gene and has no sites in the vector, liberates a segment of DNA that spans the region of the presumed mutation in the S288C *flo8* gene (see Figure 2). The purified DNA fragment was used to transform an S288C *flo8-1/flo8-1* diploid strain (L5107). Ura$^+$ transformants were tested for their ability to form pseudohyphae on SLAD medium. All Ura$^+$ transformants with gapped pHL1 gave Phd$^-$ colonies. DNA was extracted from the Ura$^+$ Phd$^-$ transformants and transformed into *E. coli* to amplify the plasmid for DNA sequencing. Control experiments in which a plasmid with an intact *FLO8* gene on the plasmid was used to transform a *flo8/flo8* diploid gave Phd$^+$ transformants.

Photomicroscopy: Photographs of colonies were taken under bright field (GIMENO *et al.* 1992). Indirect immunofluorescence was carried out using standard procedures for yeast (PRINGLE *et al.* 1991). Formaldehyde was added to fresh early log cultures to a final concentration of 3.7%, cultures were incubated at room temperature for 30 min. Cells were then collected and resuspended in 3.7% formaldehyde, 0.1 M potassium phosphate buffer (pH 6.5), incubated for 1 hr at room temperature, washed three times with spheroplasting solution (1.2 M sorbitol, 0.1 M potassium phosphate pH 7.5) and resuspended in 1 ml of the same solution plus 25 mM β-mercaptoethanol and 25 μg/ml zymolyase. After 30 min of digestion at 30°, spheroplasts were collected and loaded onto poly-L-lysine coated multi-well slides. Excess cells were aspirated away. Spheroplasts on the slide were immersed in cold methanol for 5 min at −20° and rewetted with phosphate-buffered saline (PBS) solution containing 1 mg/ml bovine serum albumin (BSA). Cells were then covered with 15 μl primary anti-myc mouse monoclonal 9E10 (1:100 diluted into the PBS/BSA solution) and incubated in a moisture chamber for 1 hr. The second antibody was an DTAF-conjugated anti-mouse IgG (1:50 dilution). Antibodies were washed away after each incubation. 4′,6-diamidino-2-phenylindole (DAPI) (1 μg/ml) was incubated with the cells before loading the mounting solution.

RESULTS

S288C strains have a mutation that blocks pseudohyphal formation: Haploid strains (Mat**a** and Matα) from the S288C and Σ1278b (Σ) backgrounds were mated in all combinations to form diploids, and these were subsequently analyzed for their ability to make pseudohyphae on SLAD medium. Diploids obtained from Σ × Σ matings form abundant pseudohyphae (Phd$^+$), whereas those from a S288C × S288C mating fail to form pseudohyphae (Phd$^-$) (Figure 1, A and B). In all matings between a Σ haploid and a S288C haploid, the diploids made pseudohyphae (Phd$^+$, Figure 1C). One of the crosses (X10375) between a strain from the Σ background (CG25) and one from the S288C background (10362-6A) was examined by tetrad analysis to determine whether the difference between Σ and S288C resulted from a mutation in a single gene. In each tetrad the ability of an ascospore to form pseudohyphae was assayed by mating it to a strain from the S288C background as a tester (10362-6A, 10362-3A) and determining the Phd phenotype of the resulting diploid. Tetrads derived from the Σ × S288C cross (X10375) gave strictly 2:2 segregation for all known markers. Each tetrad segregated two Phd$^-$ spores, diploids formed between these and an S288C strain failed to form any pseudohyphae. However, the pseudohyphal phenotype of the other two spores varied considerably, occasionally, one of the two formed florid pseudohyphae; more often, they produced very few pseudohyphae. These observations indicate that this cross was segregating a mutation in a single gene that blocks pseudohyphal formation, but was also segregating additional mutations that modify the level of pseudohyphal

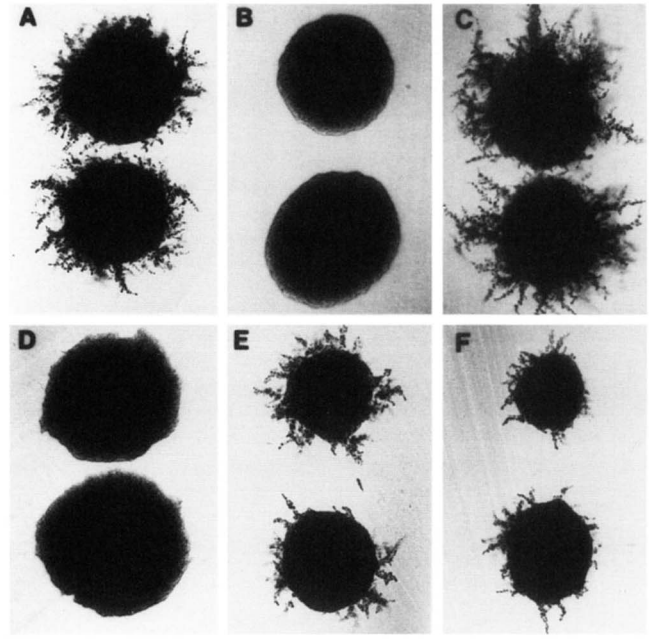


FIGURE 1.—S288C strains have a mutation (*flo8-1*) in *FLO8*, a gene necessary for filamentous growth. (A) A Σ *FLO8/FLO8* diploid strain (CGx52). (B) A S288C *flo8-1/flo8-1* diploid strain (10362-6A × 3A). (C) A diploid formed by a mating between Σ *FLO8* and S288C *flo8-1* (X10375). (D) A *flo8-1/flo8-1* diploid strain formed between progeny from the cross X10375 (L5810). (E) The *flo8-1/flo8-1* diploid strain in D transformed with *FLO8* in plasmid pHL1: (L5811). (F) The *flo8-1/flo8-1* diploid strain transformed with *FLO8* in plasmid pHL36: (L5812). Yeast cells were streaked out on SLAD medium and colonies were photographed after growth for 4 days at 30°.

growth. The effects of these additional mutations on pseudohyphal growth were subtle and variable, precluding their isolation and identification by genetic analysis. However, the mutation with the strongest Phd$^-$ phenotype could be separated from other mutations by backcrosses.

To obtain strains segregating the single strong pseudohyphal mutation we backcrossed a Phd$^-$ spore (10375-6B) from this cross (X10375) to a Σ strain (CG25) to produce diploid X10397. The cross X10397 yielded tetrads with a clear 2 Phd$^+$:2 Phd$^-$ segregation and little variation among the Phd$^+$ spores. The gene required for pseudohyphal formation that is segregating in this cross was provisionally designated *PHD10.*

Cross X10397 produced a convenient set of haploid *phd10* testers (10397-7D, 10397-9B). Matings between these *phd10* tester strains and any other *phd10* strains fail to produce pseudohyphae (Phd$^-$). Using these testers, we found that strains from the Σ1278b background are all *PHD10,* whereas those from the S288C background are all *phd10.* As noted previously and in MATERIALS AND METHODS, many S288C derivatives contain additional mutations that mute the ability of S288C *PHD10* strains to form pseudohyphae in diploids formed in matings with either Σ or S288C *PHD10* strains.

***PHD10* is *FLO8*:** A *URA3* centromere-based plasmid library was constructed from a Σ strain (see MATERIALS AND METHODS) and used to clone a functional copy of *PHD10.* The Σ genomic library was transformed into a diploid *phd10/phd10* strain L5107, and the transformants were screened visually on SLAD medium for clones that complemented the Phd$^-$ defect. Restriction digests of the 11 plasmids isolated by this visual assay indicated that they represent two genes. One of these, deemed the *PHD10* gene, represented by plasmids pHL1 and pHL36, complemented the Phd$^-$ phenotype fully (Figure 1, E and F). The other nine isolates represent another gene, which did not complement fully and only stimulated surface pseudohyphal growth in a *phd10/phd10* diploid. Subsequent gene integration and genetic analysis showed that the second gene is not linked to the *phd10* mutation, and therefore it was not studied further.

To determine whether the putative *PHD10* clones (pHL1 and pHL36) actually contained the *PHD10* gene, we first integrated the cloned DNA segment together with the *URA3* gene (pHL11), into a site of *PHD10* homology in a haploid strain (10415-8D) by transformation (MATERIALS AND METHODS). The *URA3* gene functions as a selectable marker that tags the integrated region. Southern analysis confirmed that the integrations occurred by homologous recombination. One Ura$^+$ transformant (L5281) was mated to a *phd10* strain (10397-7D) and the resulting diploid was fully capable of forming pseudohyphae. Therefore, the integrated DNA segment (pHL11) in single copy is capable of complementing the Phd$^-$ defect of *phd10* strains. The diploid strain was dissected and all 16 spores from four tetrads were analyzed for their ability to form pseudohyphae in diploids formed in matings with *phd10* tester strains 10397-7D or 10397-9B. In all four tetrads, pseudohyphal growth segregated 2 Phd$^+$:2 Phd$^-$ with Phd$^+$ completely linked to Ura$^+$. Thus, the sequence that conferred the Phd$^+$ phenotype segregated as a single gene and was linked to the *URA3* sequence that was introduced by integrative transformation. The Ura$^+$ transformant (L5281) was also mated to a *PHD10* strain (CG41). The resulting diploid was dissected and 12 tetrads were analyzed for their ability to form pseudohyphae in diploids formed by mating with *phd10* tester strains (10397-7D or 10397-9B). All the meiotic progeny from the diploid strain were Phd$^+$. This result shows that plasmid pHL11 integrated at the *PHD10* locus. From the ability of pHL11 to suppress the functional defect of *phd10* strains and its ability to integrate by homology at the *PHD10* locus we conclude that we have cloned the Σ1278b *PHD10* gene.

To map the *PHD10* gene, we first hybridized a *PHD10* probe to a panel of phage lambda clones that covers 82% of the yeast genome (RILES *et al.* 1993). *PHD10* hybridized with clones 2650 and 6052 from chromosome *V* on the blot, locating the gene to the region of overlap between these two clones. The positions of clone 2650 and 6052 suggested the following order for genes in this region: *TRP1, RAD51, PHD10, BEM2* and *SPR6.* This position of *PHD10* was confirmed by meiotic mapping. Tetrad analysis of a three point cross between yeast strains 16-1 and 10459-4B place *PHD10* between *met6* and *spr6* about 0.9 cM from *spr6.*

The DNA segment shared by both *PHD10* complementing plasmids pHL1 and pHL36 was sequenced. One large open reading frame (ORF) with 799 amino acid residues was identified. The putative amino acid sequence did not reveal any informative motifs. There is a stretch of glutamine residues that is present in many transcription factors, which accounts for the fact that several yeast transcription factors have the highest similarity to Phd10p from our BLAST search (Gal1p, Pho2p, TBP and Ssn6p). After submission of this paper, this sequence we were calling *PHD10* appeared in Genbank as *FLO8,* a putative transcriptional activator of *FLO1* (KOBAYASHI *et al.* 1996). Although previous work had mislocated *FLO8* on chromosome *VIII* (TEUNISSEN *et al.* 1995), *FLO8* and *PHD10* are the same gene, which is located on chromosome *V.* Therefore, we will refer to this gene as *FLO8* and its mutations as *flo8.* All the strain designations in the tables and figures follow this assignment.

The S288C *flo8-1* mutant allele: The SGD sequence of *FLO8,* derived from S288C, differs by 1 bp in the coding region from our sequence of *FLO8* obtained from Σ (Figure 2). The existence of this base-pair change in S288C strains was confirmed by cloning and

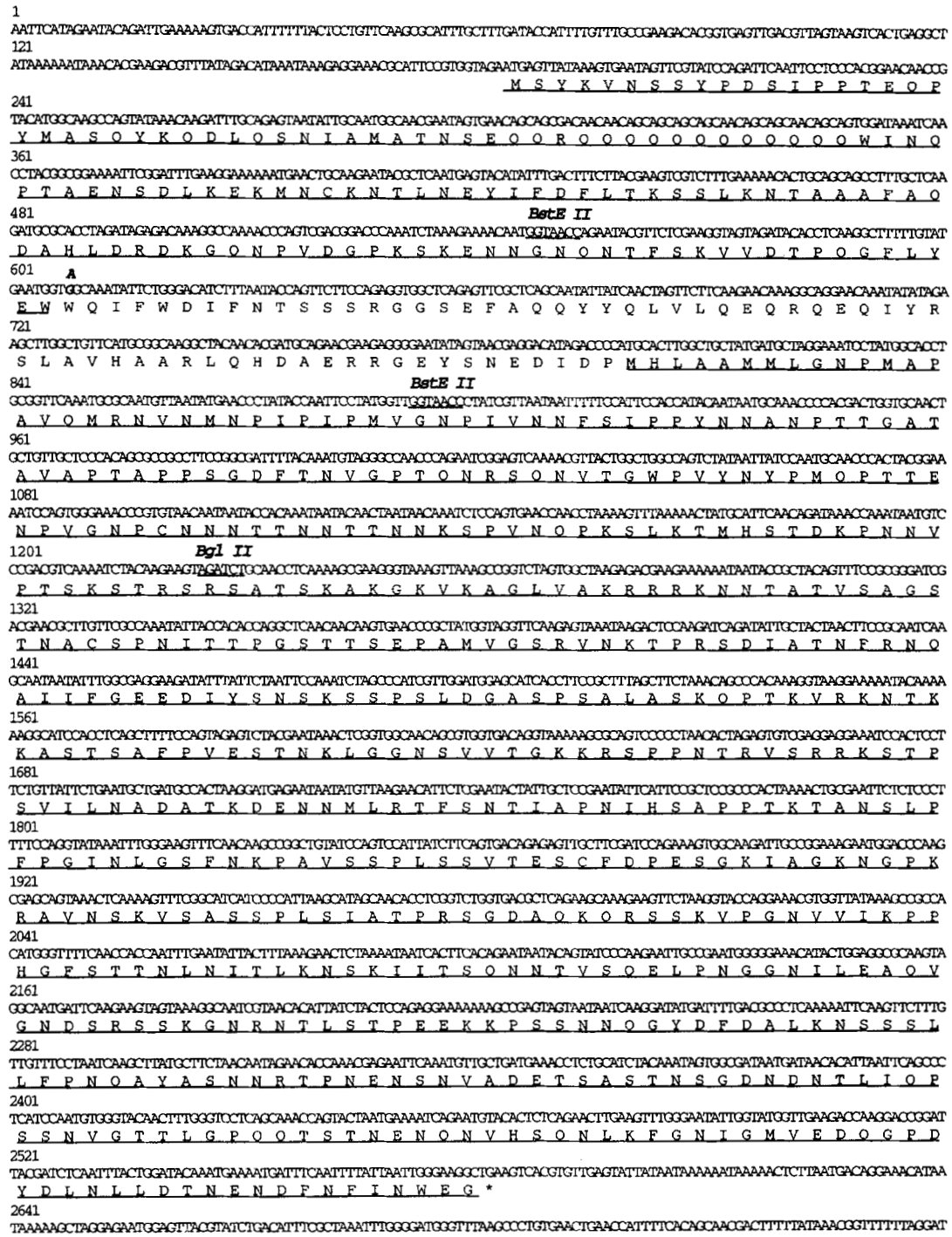


FIGURE 2.—DNA sequences of *FLO8* from Σ strains and *flo8-1* of S288C strains. The *FLO8* DNA sequence from Σ strains and its deduced protein sequence are shown. The change at position 608 from A to G in the S288C strains (*flo8-1* allele) is indicated. The underlined protein segments represent the expected ORFs from S288C strains. The region between two *Bst*EII sites was removed to clone the *flo8-1* mutation by gap repair. The myc-tag is inserted at the *Bgl*II site of *FLO8*.

sequencing the *flo8* region from S288C strains. In one approach, we used the Σ *FLO8* gene to probe a S288C yeast genomic library and cloned the S288C *flo8* gene (MATERIALS AND METHODS). In an alternative approach, we cloned the S288C mutation by gap repair. *S. cerevisiae* replaces a gap on a plasmid with its corresponding chromosomal DNA by homologous recombination. The Σ *FLO8* region corresponding to the mutation in plasmid pHL1 was deleted by *Bst*EII digestion (see Figure 2) and the gapped *FLO8* on this plasmid was transformed into an S288C strain identified as *flo8/flo8* (L5107). All transformants of the *Bst*EII-digested pHL1 showed a Phd$^-$ phenotype, whereas control transformants of undigested pHL1 gave a Phd$^+$ phenotype. This result indi-

cates that the gap-repaired plasmid had replaced the gapped region with the corresponding chromosomal DNA that carries a *flo8* mutation. Sequence comparison among the cloned S288C *flo8* gene, gap-repaired *flo8,* and the Σ *FLO8* confirmed that S288C *flo8* and Σ *FLO8* indeed differ by only 1 bp. S288C *flo8,* now called *flo8-1,* has one base change from G to A, resulting in a stop codon in the coding sequence (see Figure 2). Therefore, S288C strains probably produce a truncated inactive Flo8p (Figure 2). Our finding of a nonsense mutation in S288C agrees with our genetic data indicating that S288C has a mutation in the *FLO8* gene that blocks filamentous growth.

Northern analysis indicates that the *FLO8* gene is expressed in both haploid cells and diploid cells grown in SC medium (not shown). The same result was obtained with a *FLO8-LacZ* construct, which showed roughly equivalent activity in both haploid and diploid cells (not shown).

***FLO8* is required for pseudohyphal growth in both Σ and S288C strains:** The *FLO8* gene was deleted in Σ strains 10480-5C and 10480-5D using plasmid pHL129 to replace the resident *FLO8* gene with deletion *flo8-2* (creating strains L5816 and L5817, respectively). The presence of the deletion was confirmed by Southern analysis. The deletion *flo8-2* has 85% of the *FLO8* coding region replaced by a *hisGURA3KanRhisG* insertion, resulting in a null mutation. Σ strains carrying this null mutation have no obvious growth defect on rich media. On SLAD medium, diploid *flo8-2/flo8-2* strains are defective in pseudohyphal development (Figure 3, A and B); they fail to invade or form filaments.

Expression of *FLO8* from a high copy plasmid (pHL135) in Σ diploids enhances filament formation on SLAD medium (Figure 3C). This enhancement is limited to this nitrogen starvation medium. No enhancement was observed on media with standard levels of nitrogen. By comparison, *PHD1* and *CPH1,* genes whose overexpression enhances pseudohyphal growth, stimulate filament formation even on media containing standard nitrogen concentrations (GIMENO and FINK 1994; LIU *et al.* 1994).

We also introduced a single copy of the *FLO8* gene into an S288C *flo8-1* strain background and found that this functional *FLO8* gene introduced by transformation is sufficient to suppress the pseudohyphal defect. The experiment was executed by transforming *FLO8* (pHL11) into haploid *flo8-1* strains FY3 and FY2 yielding transformants L5302 and L5304, respectively. Diploids (*e.g.,* L5306) from the mating between the transformed strains L5302 and L5304 are Phd^+, whereas a diploid (L5813) from the mating between the untransformed FY3 and FY2 is Phd^- (Figure 3, E and F). Diploids formed by mating transformed strains by the Phd^- parents are also Phd^+. These results show that *flo8-1* is the mutation in the S288C strain responsible for the defect in pseudohyphal growth.

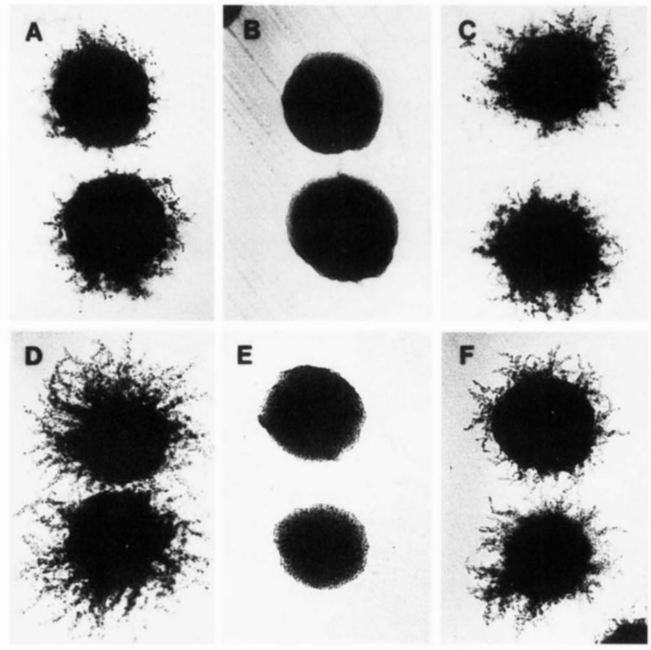


FIGURE 3.—*FLO8* is necessary for pseudohyphal development in diploids. (A) A Σ *FLO8/FLO8* strain (L5366). (B) A Σ *flo8-2/flo8-2* strain (L5818+pRS202). (C) A Σ *FLO8/FLO8* strain carrying *FLO8* on a high copy plasmid (L5366+pHL135). (D) EM93. (E) S288C *flo8-1/flo8-1* strain L5813. (F) S288C *FLO8/FLO8* diploid strain created by integrative transformation of *FLO8* (L5306). Yeast cells were streaked out on SLAD medium for single colonies. A–C were grown for 3 days before photography; D–F were grown for 4 days.

The *flo8* mutation also affects haploid invasion and flocculation: Previous studies have shown that haploids of Saccharomyces are capable of invading the agar (ROBERTS and FINK 1994). Though the haploid invasion and diploid pseudohyphal pathways have many features in common, haploid invasion is different from diploid invasion because haploid cells invade on rich YPD medium and poorly on SLAD, whereas diploids invade well on SLAD and poorly on YPD. We examined the Σ *FLO8* and *flo8* strains to determine whether *flo8* was required for haploid invasion. A haploid Σ *FLO8* strain invades well, whereas an isogenic Σ strain carrying the null allele of *FLO8* (*flo8*-2) is unable to invade (Figure 4). This phenotype is completely linked to *flo8-2* in crosses, suggesting that *FLO8* in Σ is required both for diploid pseudohyphal growth and haploid invasion.

Although haploid *flo8* strains from the S288C background also fail to invade on YPD, *FLO8* transformants of S288C do not completely restore a vigorous haploid invasion phenotype. A S288C haploid strain (FY3) and its *FLO8* transformants (L5302) were analyzed on YPD. The S288C::*FLO8* transformants were only partially restored for haploid invasive growth, despite the fact that the corresponding S288C::*FLO8* diploid (L5306) had completely normal filamentous growth on SLAD (Figure 3F). S288C must carry mutations in addition to *flo8* that affect haploid invasive growth, but not diploid pseudohyphal growth. This result further supports the

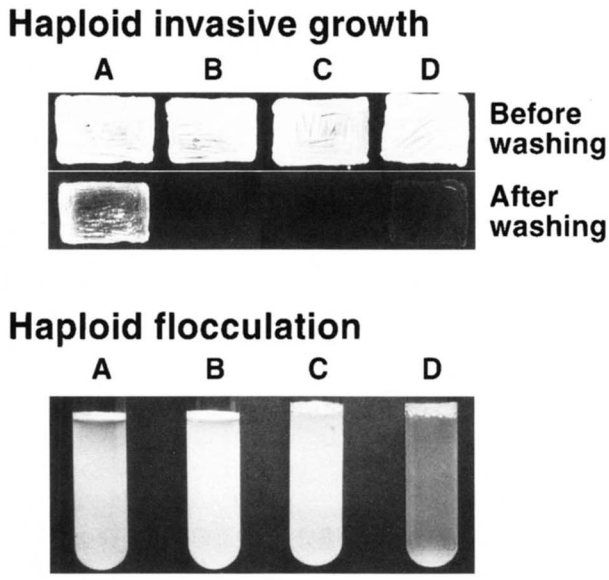


FIGURE 4.—*FLO8* is required for invasive growth and flocculation in haploids. (A) A Σ *FLO8* strain (10480-5C). (B) A Σ *flo8-2* strain (L5816). (C) A S288C *flo8-1* strain (FY3). (D) A S288C::*FLO8* strain (L5302). The top panel shows the invasive growth of haploid cells on solid YPD medium. Strains were patched and grown on YPD medium for 3 days at 30°. The plate was then washed with running tap water to rinse away cells on the agar surface. Pictures were taken before and after washing. The bottom panel shows the flocculation of different strains in liquid YPD medium. Cells were grown in YPD to saturation, agitated on a vortex mixer, allowed to settle for 10 min and then photographed. The S288C::*FLO8* culture (D) is more transparent because flocculent cells settle rapidly. The Σ *FLO8* strain is only slightly more flocculent than the Σ *flo8-2* strain, and this subtle difference is not apparent in the photograph.

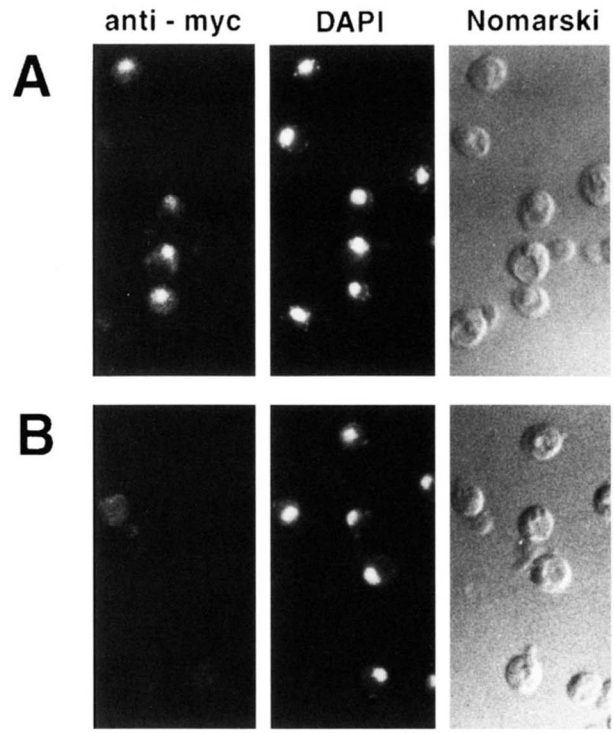


FIGURE 5.—Nuclear localization of the myc-tagged Flo8p. (A) A *flo8-2/flo8-2* diploid strain carrying the myc-tagged Flo8p (L5818+pHL138). (B) The same *flo8/flo8* diploid strain carrying untagged Flo8p (L5818+pHL135).

idea that although haploid invasive growth and diploid pseudohyphal growth share common features, they are not identical developmental events. Some mutations may affect one process without affecting the other.

FLO8 haploids in the S288C background flocculate and settle more rapidly in liquid media than *flo8* haploids (Figure 4, C and D). This flocculation phenotype cosegregates with *FLO8* in S288C strains. The flocculation in *FLO8* strains is a consequence of an especially avid aggregation of the cells because the aggregated cells can be dispersed by chelating out the calcium (STRATFORD 1989; TAYLOR and ORTON 1975); addition of 4 mM EDTA disperses the cell aggregates, and subsequent addition of Ca^{++} leads to reaggregation. However, this flocculation phenotype is not as pronounced in Σ*FLO8* haploid strains (Figure 4, A and B), indicating the existence of other differences between S288C and Σ that affect the flocculation phenotype. Thus, with respect to both haploid invasion and flocculation, the magnitude of the difference between *FLO8* and *flo8* strains depends upon the genetic background (S288C or Σ).

Flo8p is localized to the nucleus: An epitope-tagged version of Flo8p was localized by immunofluorescence microscopy. A triple myc epitope tag from pKB241 was inserted in frame with Flo8p (Figure 2) and *FLO8::myc3* was expressed from its own promoter on a high copy yeast vector (details in MATERIALS AND METHODS). This construct partially complements the Phd^- phenotypes: *FLO8::myc3* completely restores haploid invasive growth to *flo8* (L5816) cells; however, it only partially complements the pseudohyphal growth defect in diploid *flo8/flo8* (L5818). The diploid *flo8/flo8* strain carrying *FLO8::myc3* was subjected to indirect immunofluorescence microscopy. myc-tagged Flo8p colocalized with the DAPI staining, suggesting that Flo8p is localized to the nucleus (Figure 5A). This nuclear localization was present in cells observed at all phases of yeast cell cycle. A control strain overexpressing nontagged *FLO8* showed only a hazy background level of staining throughout the cell (Figure 5B).

Pseudohyphal growth in frequently used laboratory strains We tested frequently used haploid laboratory strains for their ability to form pseudohyphae in diploids formed by mating with the Σ haploid *flo8* tester strains 10397-7D and 10397-9B. Diploids formed between a haploid strain D273-10B/A1 obtained from Dr. F. SHERMAN (used frequently in mitochondrial studies) and our *flo8* tester form pseudohyphae on SLAD medium. This result suggests that the D273-10B/A1 strains have a functional copy of the *FLO8* gene. Moreover, a diploid from that strain background, DA1rho°XJJM113, forms pseudohyphae on SLAD showing that respira-

tion-deficient strains are fully competent to undergo the morphological switch. Other studies differ in their assessment of the filamentation behavior of D273-10B strains; one concludes that D273-10B haploid strains are capable of forming filaments (WRIGHT *et al.* 1993), whereas others found strains from this background incapable of forming pseudohyphae (BLACKETER *et al.* 1993, 1994). The discrepancy could mean that mutations such as *flo8* were introduced upon subculturing in one case and not in the other. Another haploid strain, A364A from L. HARTWELL (used extensively in cell division cycle studies) when mated to our *flo8* tester also produces diploids that filament under nitrogen starvation. On the other hand, W303 diploid strains (L5682) or diploids from a mating between a W303 haploid strain (EY699) and the *flo8* tester are all Phd$^-$. This result suggests that W303 like S288C contains a nonfunctional *FLO8* gene. Transformation of *FLO8* into a W303 diploid strain did not rescue the Phd$^-$ phenotype, suggesting that W303 has mutations in both the *FLO8* gene and at least one other gene essential for pseudohyphal growth (result not shown).

We also tested the correlation between diploid pseudohyphal growth and haploid invasive growth in these laboratory strains. As summarized in Table 3, Σ1278b strains and the A364A strains are competent in both haploid invasion and diploid pseudohyphal growth, whereas S288C strains and W303 strains are deficient in both. However, D273-10B/A1 fails to manifest haploid invasive growth on YPD medium, even though the diploid strain of the same background forms filaments on SLAD. This result emphasizes the difference between haploid invasive growth and filamentous growth, D273-10B/A1 haploids contain a mutation(s) that interferes with haploid invasive growth, but this mutation does not affect diploid filament formation.

TABLE 3

Comparison of haploid invasive growth and diploid filamentous growth in various strain backgrounds

Strains	Haploid invasion	Pseudohyphal growth
Haploids		
Σ1278b	+	+++
S288C	−	−
D273-10B/A1 (Sherman)	−	++
A364A (Hartwell)	+	+++
W303-1A	−	−
Diploids or polyploids		
EM93	NA	++++
Fleischmann's yeast	NA	−
Red Star dry yeast	NA	++++
Beer 2	NA	−

The +, pseudohyphal growth was observed. The number of +s indicates the relative extent of pseudohypal growth. −, no pseudohyphal growth was observed. The + for haploid invasion means that the strain invaded the agar as determined by the haploid invasion assay (RESULTS). NA, not applicable. For the haploid strains pseudohyphal growth was determined in diploids, made in each instance by mating the relevant haploid by a S288C *flo8* strain.

Pseudohyphal formation in feral and industrial strains: EM93 is a feral strain isolated in 1936 from a rotting fig in Merced, California (MORTIMER and JOHNSTON 1986). This strain is the ancestor of S288C and contributes most of the genome in S288C. EM93 is a heterothallic diploid and develops florid pseudohyphae on standard nitrogen starvation medium (Figure 3D). The spores from seven EM93 tetrads were mated to *flo8* tester strains to determine whether they could form pseudohyphae in diploids. In each tetrad all four spores were capable of forming pseudohyphae with the *flo8* tester, indicating that EM93 was homozygous for *FLO8* and was not the origin of the *flo8* mutation in S288C. In addition to EM93, many strains isolated from AIDS patients are dimorphic, making elaborate pseudohyphae upon nitrogen starvation. One of these extensively analyzed by MCCUSKER (MCCUSKER *et al.* 1994a,b) is a diploid.

Industrial strains are not strictly comparable either to laboratory or feral strains although both the laboratory and industrial strains are presumed to be derived from the feral strains. One difficulty in comparing these strains of different origin is that the industrial strains are often polyploid and of uncertain chromosome constitution. Nevertheless, several of them were analyzed for pseudohyphal formation on nitrogen starvation medium. Red Star baking yeast made dramatic filaments similar to those of EM93. However, Fleischmann's yeast, another baking strain, fails to make filaments when starved for nitrogen. A brewing yeast strain from the brewery in Seville, Spain also fails to make filaments.

DISCUSSION

***FLO8* is required for pseudohyphal formation in diploids, and invasive growth and flocculation in haploids:** Several lines of evidence indicate that a functional copy of *FLO8* is necessary for filamentous growth in diploids. First a *flo8* mutation in S288C strains blocks the pseudohyphal growth of S288C diploids. Introduction of a single copy of Σ-derived *FLO8* into a S288C *flo8* strain restores the Phd$^+$ phenotype. Second, deletion of the *FLO8* gene in Σ strains renders them Phd$^-$. Moreover, overproduction of *FLO8* in Σ diploids promotes more vigorous pseudohyphal growth.

FLO8 is also involved in two haploid specific processes: invasive growth and flocculation. Invasive growth is defined operationally by washing cells off the agar surface of petri plates; strains that are competent to invade leave a visible residue after the wash, whereas those that fail to invade do not (ROBERTS and FINK 1994). Although haploid invasive growth requires many of the same genes as diploid pseudohyphal growth, including elements of the pheromone response pathway

and *FLO8*, the two processes have a number of differences. First, haploid invasive growth is limited to rich media and does not occur on SLAD. By contrast, diploids invade much more poorly than haploids on rich medium. Second, the haploid cells on rich medium can be removed from the agar surface by rubbing with a finger, whereas diploid filaments on low nitrogen medium have penetrated far beneath the agar and cannot be rubbed off.

Both the invasive growth on rich medium and flocculation are under the control of *FLO8*. Since these are haploid phenotypes that are not manifest in diploids, both may be regulated by **a**1-α2 repression. Although a functional *FLO8* gene is required for both processes, a comparison of Σ *FLO8* and S288C::*FLO8* haploid strains indicates that the extent of invasive growth does not correlate with that of flocculation, the Σ *FLO8* strain is more invasive but less flocculent than S288C::*FLO8*.

The function of Flo8p: The *FLO8* gene was recently reported to be the putative transcriptional activator for *FLO1* (KOBAYASHI *et al.* 1996). *FLO1* and *FLO5* are dominant flocculation genes that encode cell wall-associated proteins that share 96% identity in protein sequence (TEUNISSEN *et al.* 1993, 1995a; BIDARD *et al.* 1995). Our observation that Flo8p is localized to the nucleus and is required for flocculation in S288C is consistent with its reported role as a transcriptional activator of *FLO1*. Furthermore, the finding that *FLO8* is also required for diploid filamentous growth and haploid invasion suggests the existence of other genes under Flo8p regulation. Ssn6p and Tup1p(Sp12p) unlike Flo8p are negative regulators of flocculation (TEUNISSEN *et al.* 1995b); haploid strains containing the *ssn6* or *tup1* mutations are more flocculent than *SSN6 TUP1* strains. Ssn6p-Tup1p, which acts as a complex in transcriptional repression (KELEHER *et al.* 1992), could play a direct role in repressing *FLO1* transcription.

Many laboratory and industrial strains are defective in pseudohyphal formation: A number of laboratory strains (S288C and W303) have a recessive mutation in the *FLO8* gene that blocks haploid invasive growth and diploid pseudohyphal growth. Where did the *flo8* mutation arise? Presumably, the *flo8* mutation was selected passively by yeast geneticists who found the properties of the *FLO8* haploids a nuisance for genetic and biochemical analysis. The diploid progenitor of S288C is EM93, which is homozygous for a functional *FLO8* gene. Its haploid progeny are extremely flocculent making them difficult to transfer upon replica-plating. Moreover, in liquid cultures, these flocculent cells do not stay in suspension, making it difficult to measure their growth rate or to obtain colonies that arise from a single cell upon plating on solid medium. It is likely that in the decades since yeast was first domesticated the *flo8* mutation was selected by yeast geneticists who picked nonflocculent colonies that replica-plate well on solid medium and stay in suspension in liquid.

Laboratory strains of Saccharomyces contain many naturally occurring mutations in addition to *flo8*. S288C is known to have a number of mutations including *gal2*, *suc2*, *mal*, and *flo1* (MORTIMER and JOHNSTON 1986). The *suc2* mutation is the result of an amber codon in the coding sequence of the *SUC2* gene (CARLSON *et al.* 1981; GOZALBO and HOHMANN 1989). The *mal* and *flo1* mutations are the consequence of deletions of repeated sequence close to telomeres (MORTIMER and JOHNSTON 1986). S288C also has a mutation in the *KSS1* gene (ELION *et al.* 1991a,b). Another example of naturally occurring mutations is *SSD1* in W303 strains (SUTTON *et al.* 1991; COSTIGAN *et al.* 1992; STETTLER *et al.* 1993; KIKUCHI *et al.* 1994). There is also variation in the copy number and location of the *RTM1* gene in different *S. cerevisiae* strains (NESS and AIGLE 1995). Some strains are completely devoid of the *RTM1* sequence. The difference between *RTM1* or *rtm1* strains appears to be a consequence of deletions caused by rearrangements between repeated telomeric sequences.

With the completion of the sequence of the entire yeast genome, the presence of other naturally occurring mutations may become recognized. One way to identify these mutations is to search for pairs of ORFs transcribed from the same DNA strand that are interrupted by a nonsense codon or short insertion sequence. Our survey of ORFs on yeast chromosome *III* have identified at least three pairs of ORFs that are separated by a very short DNA segment, including: YCL061c/YCL060c, YCR086w/YCR087w, and YCR099c/YCR100c. These short interruptions could represent genes that are nonfunctional in S288C, but functional in other strain backgrounds.

We also find that many industrial strains do not form pseudohyphae. Most industrial strains are diploid or of higher ploidy and should be capable of filament formation on SLAD medium. Yet, many do not form filaments. Presumably, the filament formation and aggregation properties of pseudohyphal cells were not desirable for the growth in a fermentor and were selected against when these brewing and baking strains were developed. It would be interesting to know whether these strains also have a defective *FLO8* gene.

Many feral strains of Saccharomyces are filamentous: A number of Saccharomyces strains that have been isolated directly from their natural habitat (fruits or vegetables) and analyzed without extensive laboratory culture appear to be dimorphic diploids. EM93 the progenitor of S288C was isolated from a rotting fig in Merced, California. Recently, strains of *S. cerevisiae* have been isolated from immunocompromised humans. Many of these strains are diploid and filamentous (MCCUSKER *et al.* 1994a,b; C. A. STYLES and G. R. FINK, unpublished results). We suspect that dimorphism, the ability of Saccharomyces to switch between its yeast and filamentous form, contributes to its ability to survive in both plant and animal tissue. The earliest recorded

observation of filamentous growth in feral strains of Saccharomyces was made by EMIL CHRISTIAN HANSEN (HANSEN 1886).

> . . . I have shown that "veil formation" commonly occurs with a variety of microorganisms, bacteria as well as fungi.
>
> We have observed this with the Saccharomyces as well, and learned that the cells in the veils of old cultures develop a more elongated form, and more complex colonies; filiform cells were formed, resembling bacteria.
>
> (translated by Dr. DANIEL KORNITZER)

It is probably significant that the unusual filaments observed by HANSEN formed in cultures that were incubated over long periods of time. Presumably, the cells in these cultures were starved for nutrients such as nitrogen.

We thank C. GIMENO and MARY CLANCY for sharing strains and information, J. LOEB for critical reading of this manuscript, and members of the FINK laboratory for helpful discussions. This work was supported by a Jane Coffin Childs Fellowship (H.L.), a Bristol-Myers Squibb Award in Infectious Disease, and National Institutes of Health grant GM402661 (G.R.F.). G.R.F. is an American Cancer Society Professor of Genetics.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

BANUETT, F., and I. HERSKOWITZ, 1994 Identification of Fuz7, a Ustilago maydis MEK/MAPKK homolog required for a-locus-dependent and -independent steps in the fungal life cycle. Genes Dev. **8:** 1367–1378.

BARRAL, Y., and C. MANN, 1995 G1 cyclin degradation and cell differentiation in *Saccharomyces cerevisiae.* C. R. Acad. Sci. **318:** 43–50.

BARRAL, Y., S. JENTSCH and C. MANN, 1995 G1 cyclin turnover and nutrient uptake are controlled by a common pathway in yeast. Genes Dev. **9:** 399–409.

BIDARD, F., M. BONY, B. BLONDIN, S. DEQUIN and P. BARRE, 1995 The *Saccharomyces cerevisiae* FLO1 flocculation gene encodes for a cell surface protein. Yeast **11:** 809–822.

BLACKETER, M. J., C. M. KOEHLER, S. G. COATS, A. M. MYERS and P. MADAULE, 1993 Regulation of dimorphism in *Saccharomyces cerevisiae:* involvement of the novel protein kinase homolog Elm1p and protein phosphatase 2A. Mol. Cell. Biol. **13:** 5567–5581.

BLACKETER, M. J., P. MADAULE and A. M. MYERS, 1994 The *Saccharomyces cerevisiae* mutation elm4-1 facilitates pseudohyphal differentiation and interacts with a deficiency in phosphoribosylpyrophosphate synthase activity to cause constitutive pseudohyphal growth. Mol. Cell. Biol. **14:** 4671–4681.

BRANDRISS, M. C., and B. MAGASANIK, 1979 Genetics and physiology of proline utilization in *Saccharomyces cerevisiae.* J. Bacteriol. **143:** 1403–1410.

CARLSON, M., B. C. OSMOND and D. BOTSTEIN, 1981 Genetic evidence for a SUC gene in yeast. Genetics **98:** 41–54.

CHRISTIANSON, T. W., R. S. SIKORSKI, M. DANTE, J. H. SHERO and P. HIETER, 1992 Multifunctional yeast high-copy-number shuttle vectors. Gene **110:** 119–122.

CORUZZI, G., M. K. TREMBATH and A. TZAGOLOFF, 1978 Assembly of the mitochondrial membrane system: mutations in the pho2 locus of the mitochondrial genome of *Saccharomyces cerevisiae.* Eur. J. Biochem. **92:** 279–287.

COSTIGAN, C., S. GEHRUNG and M. SNYDER, 1992 A synthetic lethal screen identifies SLK1, a novel protein kinase homolog implicated in yeast cell morphogenesis and cell growth. Mol. Cell. Biol. **12:** 1162–1178.

ELION, E. A., J. A. BRILL and G. R. FINK, 1991a Functional redundancy in the yeast cell cycle: FUS3 and KSS1 have both overlapping and unique functions. Cold Spring Harbor Symp. Quant. Biol. **56:** 41–49.

ELION, E. A., J. A. BRILL and G. R. FINK, 1991b FUS3 represses CLN1 and CLN2 and in concert with KSS1 promotes signal transduction. Proc. Natl. Acad. Sci. USA **88:** 9392–9396.

GIMENO, C. J., and G. R. FINK, 1994 Induction of pseudohyphal growth by overexpression of PHD1, a *Saccharomyces cerevisiae* gene related to transcriptional regulators of fungal development. Mol. Cell. Biol. **14:** 2100–2112.

GIMENO, C. J., P. O. LJUNGDAHL, C. A. STYLES and G. R. FINK, 1992 Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. Cell **68:** 1077–1090.

GOZALBO, D., and S. HOHMANN, 1989 The naturally occurring silent invertase structural gene suc2 zero contains an amber stop codon that is occasionally read through. Mol. Gen. Genet. **216:** 511–516.

HANSEN, E. C., 1886 *Recherches sur la physiologie et al morphologie des ferments alcooliques.* C. R. Trav. Lab.

HENIKOFF, S., 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:** 351–359.

KALLAL, L. A., M. BHATTACHARYYA, S. N. GROVE, R. F. IANNACONE, T. A. PUGH *et al.,* 1990 Functional analysis of the sporulation-specific SPR6 gene of *Saccharomyces cerevisiae.* Curr. Genet. **18:** 293–301.

KELEHER, C. A., M. J. REDD, J. SCHULTZ, M. CARLSON and A. D. JOHNSON, 1992 Ssn6-Tup1 is a general repressor of transcription in yeast. Cell **68:** 709–719.

KIKUCHI, Y., Y. OKA, M. KOBAYASHI, Y. UESONO, A. TOH-E *et al.,* 1994 A new yeast gene, HTR1, required for growth at high temperature, is needed for recovery from mating pheromone-induced G1 arrest. Mol. Gen. Genet. **245:** 107–116.

KOBAYASHI, O., H. SUDA, T. OHTANI and H. SONE, 1996 Molecular cloning and analysis of the dominant flocculation gene FLO8 from Saccharomyces cerevisiae. Mol. Gen. Genet. **251:** 707–715.

KRON, S. J., C. A. STYLES and G. R. FINK, 1994 Symmetric cell division in pseudohyphae of the yeast *Saccharomyces cerevisiae.* Mol. Biol. Cell **5:** 1003–1022.

LIU, H., C. A. STYLES and G. R. FINK, 1993 Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science **262:** 1741–1744.

LIU, H., J. KOHLER and G. R. FINK, 1994 Suppression of hyphal formation in Candida albicans by mutation of a STE12 homolog. Science **266:** 1723–1726.

LJUNGDAHL, P. O., C. J. GIMENO, C. A. STYLES and G. R. FINK, 1992 SHR3: a novel component of the secretory pathway specifically required for localization of amino acid permeases in yeast. Cell **71:** 463–478.

MCCUSKER, J. H., K. V. CLEMONS, D. A. STEVENS and R. W. DAVIS, 1994a Genetic characterization of pathogenic *Saccharomyces cerevisiae* isolates. Genetics **136:** 1261–1269.

MCCUSKER, J. H., K. V. CLEMONS, D. A. STEVENS and R. W. DAVIS, 1994b *Saccharomyces cerevisiae* virulence phenotype as determined with CD-1 mice is associated with the ability to grow at 42 degrees C and form pseudohyphae. Infect. Immun. **62:** 5447–5455.

MORTIMER, R. K., and J. R. JOHNSTON, 1986 Genealogy of principal strains of the yeast genetic stock center. Genetics **113:** 35–43.

MÖSCH, H. U., R. ROBERTS and G. R. FINK, 1996 Ras2 signals via the Cdc42/Ste20/MAPK module to induce filamentous growth in *Saccharomyces cerevisiae.* Proc. Natl. Acad. Sci. USA **93:** 5352–5356.

MULERO, J. J., and T. D. FOX, 1994 Reduced but accurate translation from a mutant AUA initiation codon in the mitochondrial COX2 mRNA of *Saccharomyces cerevisiae.* Mol. Gen. Genet. **242:** 383–390.

MULLER, P. P., M. K. REIF, S. ZONGHOU, C. SENGSTAG, T. L. MASON *et al.,* 1984 A nuclear mutation that post-transcriptionally blocks accumulation of a yeast mitochondrial gene product can be suppressed by a mitochondrial gene rearrangement. J. Mol. Biol. **175:** 431–452.

MYERS, A. M., A. TZAGOLOFF, D. M. KINNEY and C. J. LUSTY, 1986 Yeast shuttle and integrative vectors with multiple cloning sites suitable for construction of lacZ fusions. Gene **45:** 299–310.

NESS, F., and M. AIGLE, 1995 RTM1: a member of a new family of telomeric repeated genes in yeast. Genetics **140:** 945–956.

PRINGLE, J. R., A. E. M. ADAMS, D. G. DRUBIN and B. K. HAARER, 1991 *Immunofluorescence Methods for Yeast.* Academic Press, Inc., London.

RILES, L., J. E. DUTCHIK, A. BAKTHA, B. K. MCCAULEY, E. C. THAYER *et al.,* 1993 Physical maps of the six smallest chromosomes of *Saccharomyces cerevisiae* at a resolution of 2.6 kilobase pairs. Genetics **134:** 81–150.

ROBERTS, R. L., and G. R. FINK, 1994 Elements of a single MAP kinase cascade in *Saccharomyces cerevisiae* mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev. **8:** 2974–2985.

RUDOLPH, H. K., A. ANTEBI, G. R. FINK, C. M. BUCKLEY, T. E. DORMAN *et al.,* 1989 The yeast secretory pathway is perturbed by mutations in PMR1, a member of a Ca2+ ATPase family. Cell **58:** 133–145.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SHERMAN, F., G. R. FINK and J. B. HICKS, 1986 *Methods in Yeast Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

SIDDIQUI, A. H., and M. C. BRANDRISS, 1988 A regulatory region responsible for proline-specific induction of the yeast *PUT2* gene as adjacent to its TATA box. Mol. Cell. Biol. **8:** 4634–4641.

SIKORSKI, R. S., and P. HIETER, 1989 A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae.* Genetics **122:** 19–27.

STETTLER, S., N. CHIANNILKULCHAI, S. HERMANN-LE DENMAT, D. LALO, F. LACROUTE *et al.,* 1993 A general suppressor of RNA polymerase I, II and III mutations in *Saccharomyces cerevisiae.* Mol. Gen. Genet. **239:** 169–176.

STRATFORD, M., 1989 Yeast flocculation: calcium specificity. Yeast **5:** 487–496.

SUTTON, A., D. IMMANUEL and K. T. ARNDT, 1991 The SIT4 protein phosphatase functions in late G1 for progression into S phase. Mol. Cell. Biol. **11:** 2133–2148.

TAYLOR, N. W., and W. L. ORTON, 1975 Calcium in flocculence of *Saccharomyces cerevisiae.* J. Instit. Brew. **81:** 53–57.

TEUNISSEN, A. W., E. HOLUB, J. VAN DER HUCHT, J. A. VAN DEN BERG and H. Y. STEENSMA, 1993 Sequence of the open reading frame of the FLO1 gene from *Saccharomyces cerevisiae.* Yeast **9:** 423–427.

TEUNISSEN, A. W., J. A. VAN DEN BERG and H. Y. STEENSMA, 1995a Localization of the dominant flocculation genes FLO5 and FLO8 of *Saccharomyces cerevisiae.* Yeast **11:** 735–745.

TEUNISSEN, A. W., J. A. VAN DEN BERG and H. Y. STEENSMA, 1995b Transcriptional regulation of flocculation genes in *Saccharomyces cerevisiae.* Yeast **11:** 435–446.

THOMPSON, C. M., A. J. KOLESKE, D. M. CHAO and R. A. YOUNG, 1993 A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. Cell **73:** 1361–1375.

WINSTON, F., C. DOLLARD and S. L. RICUPERO-HOVASSE, 1995 Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. Yeast **11:** 53–55.

WRIGHT, R. M., T. REPINE and J. R. REPINE, 1993 Reversible pseudohyphal growth in haploid Saccharomyces cerevisiae is an aerobic process. Curr. Genet. **23:** 388–391.

Communicating editor: A. G. HINNEBUSCH

ATTACHMENT I--FINAL RISK ASSESSMENT OF SACCHAROMYCES CEREVISIAE

(February 1997)

I. INTRODUCTION

Saccharomyces cerevisiae has an extensive history of use in the area of food processing. Also known as Baker's Yeast or Brewer's Yeast, this organism has been used for centuries as leavening for bread and as a fermenter of alcoholic beverages. With a prolonged history of industrial applications, this yeast has been either the subject of or model for various studies in the principles of microbiology. Jacob Henle based his theories of disease transmission on studies of strains of Brewer's Yeast. Currently, S. cerevisiae is the subject of a major international effort to characterize a eucaryotic genome (Anderson, 1992).

History of Commercial Use and Products Subject to TSCA Jurisdiction

Saccharomyces cerevisiae, in addition to its use in food processing, is widely used for the production of macromolecular cellular components such as lipids, proteins including enzymes, and vitamins (Bigelis, 1985; Stewart and Russell, 1985).

The Food and Drug Administration rates Brewer's Yeast extract as Generally Recognized as Safe (FDA, 1986). Furthermore, the National Institutes of Health in its Guidelines for Research Involving Recombinant DNA Molecules (DHHS, 1986) considers S. cerevisiae a safe organism. Most experiments involving S. cerevisiae have been exempted from the NIH Guidelines based on an analysis of safety (see Appendix C-II of the NIH Guidelines). While alcoholic beverages, vitamins, and bread leavening are covered under the Federal Food, Drug and Cosmetic Act, the production of enzymes and other macromolecules may be subject to TSCA regulation. The abundance of information on S. cerevisiae, derived from its role in industry, has positioned it as a primary model for genetic studies and, by extension, as a strong candidate for genetic manipulation for TSCA applications (Dynamac, 1990).

II. IDENTIFICATION AND CLASSIFICATION

A. Taxonomy and Characterization

Saccharomyces cerevisiae is a yeast. The organism can exist either as a single-celled organism or as pseudomycelia. The cells reproduce by multilateral budding. It produces from one to four ellipsoidal, smooth-walled ascospores. S. cerevisiae can be differentiated from other yeasts based on growth characteristics and physiological traits: principally the ability to ferment individual sugars. Clinical identification of yeast is conducted using commercially available diagnostic kits which classify the organism through analysis of the ability of the yeast to utilize distinct carbohydrates as sole sources of carbon (Buesching et al., 1979; Rosini et al., 1982). More recently, developments in systematics have led to the design of sophisticated techniques for classification, including gas-liquid chromatography of lysed whole cells (Brondz and Olsen, 1979).

As a result of the application of newer techniques arising from innovative approaches, the taxonomy of Saccharomyces is subject to greater scrutiny. The initial classification was based principally on morphological characteristics with specific physiological and biochemical traits used to differentiate between isolates with similar morphological traits. Using these criteria, there are as many as 18 species listed in the literature. In addition, what had been classified as one large heterogeneous species, S. cerevisiae, may, in the future, be divided into four distinct species based on DNA homology studies. The four species are S. cerevisiae, S. bayanus (also known as S. uvarum), S. pasteurianus (also known as S. carlsbergensis), and S. paradoxus. All four represent industrially important species. None of these organisms or other closely related species has been associated with pathogenicity toward humans or has been shown to have adverse effects on the environment.

Any assessment of Saccharomyces must take into consideration the malleability of the current classification. For this assessment of S. cerevisiae the reviews of the organism are based on the classification proposed by Van der Walt (1971).

B. Related Species of Concern

None of the above strains or other closely related species has been associated with pathogenicity toward humans or has been shown to have adverse effects on the environment.

III. HAZARD ASSESSMENT

A. Human Health Hazards

1. Colonization and Pathogenicity

S. cerevisiae is a commonly used industrial microorganism and is ubiquitous in nature, being present on fruits and

vegetables. Industrial workers and the general public come into contact with S. cerevisiae on a daily basis through both inhalation and ingestion (see section IV). Saccharomyces spp. are frequently recovered from the stools and throats of normally healthy individuals. This indicates that humans are in constant contact with these yeasts.

There are individuals who may ingest large quantities of S. cerevisiae every day, for example, people who take the yeast as part of a "health food" regimen. Therefore, studies were conducted to ascertain whether the ingestion of large numbers of these yeasts might result in either colonization, or colonization and secondary spread to other organs of the body. It was found that the installation of very large numbers of S. cerevisiae into the colons of animals would result in both colonization and passage of the yeasts to draining lymph nodes. It required up to 10^{10} S. cerevisiae in a single oral treatment to rats to achieve a detectable passage from the intestine to the lymph nodes (Wolochow et al., 1961). The concentrations of S. cerevisiae required were well beyond those that would be encountered through normal human daily exposure.

S. cerevisiae is not considered a pathogenic microorganism, but has been reported rarely as a cause of opportunistic infections. Eng et al. (1984) described five cases of such infections and reviewed the literature on eight other S. cerevisiae infections (also briefly reviewed by Walsh and Pizzo, 1988). All of the patients in the cases had underlying disease. Some of them had also received antibiotic therapy, thereby suppressing normal bacterial flora and allowing mycotic organisms to become established.

A low concern for the pathogenicity of S. cerevisiae is also illustrated by a series of surveys conducted at hospitals over the last several years. S. cerevisiae accounted for less than 1% of all yeast infections isolated at a cancer hospital and in most of the cases the organism was isolated from the respiratory system (Kiehn et al., 1980). At Yale-New Haven Hospital over the past five years, there have been 50 isolates of S. cerevisiae recovered from patients; however, most of the isolates were considered contaminants (Dynamac, 1991).

2. Toxin Production

There have been no reports of isolates of S. cerevisiae that produce toxins against either humans or animals. However, S. cerevisiae has been shown to produce toxins against other yeasts. These toxins, termed "killer toxins", are proteins or glycoproteins produced by a range of yeasts. The yeasts have been genetically modified to alter activity and are used in industrial settings as a means of controlling contamination of fermentation systems by other yeasts (Sid et al., 1988).

3. Measure of the Degree of Virulence

A number of individual virulence factors have been identified as being associated with the ability of yeasts to cause disease. The principal virulence factors associated with yeasts appear to be phospholipase A and lysophospholipase. It is believed that these enzymes enhance the ability of the yeast to adhere to the cell-wall surface and result in colonization as a first step in the infectious process. Nonpathogenic yeast had considerably lower phospholipase activities. Of a wide range of fungi assayed for phospholipase production, S. cerevisiae was found to have the lowest level of activity (Barrett-Bee et al., 1985). Therefore, based on the phospholipase virulence factor S. cerevisiae is considered a nonpathogenic yeast.

A second factor associated with virulence in yeast is the ability of a fungus to impair the host's immune capabilities. The cell walls of most fungi have the capacity to impede the immune response of the host. In a study to determine the overall pathogenicity of a number of yeasts used in industrial processes, animals exposed to both high levels of S. cerevisiae and cortisone demonstrated a greater ability of the fungus to colonize compared with those animals treated with only the yeast. However, the animals suffered no ill-effects from exposure to S. cerevisiae (Holzschu et al., 1979). Therefore, this study suggests that even with the addition of high levels of an immunosuppressant agent, S. cerevisiae appears to be nonpathogenic.

4. Ability to Transfer Virulence Factor Genes

S. cerevisiae does not carry virulence factors to humans or animals. However, the species does carry linear, double-stranded plasmids which can be transmitted to other Saccharomyces. These plasmids carry genes that encode the "killer toxins" discussed above can be transferred from one Saccharomyces to another. Therefore, gene constructs involving the incorporation of traits using these linear plasmids should be considered to be nonstable.

5. Summary

In conclusion, S. cerevisiae is a organism which has an extensive history of safe use. Despite considerable use of the organism in research and the presence of S. cerevisiae in food, there are limited reports in the literature of its pathogenicity to humans or animals, and only in those cases where the human had a debilitating condition. Factors associated with the virulence of yeasts (i.e., phospholipases) indicate that this organism is nonpathogenic. The organism has not been shown to produce toxins to humans.

B. Environmental Hazards

S. cerevisiae is ubiquitous in nature. It has been recovered from a variety of sites under varying ecological conditions. The organism is used in a variety of industrial scenarios. S. cerevisiae is commonly recovered from a variety of fresh fruits and vegetables, generally those fruits with high levels of fermentable sugars. However, it is not listed as the causative agent of food spoilage for fruits and vegetables (Phaff et al., 1966). The only adverse effect to the environment noted in the literature is the presence of the "killer toxins" which is active against other strains of Saccharomyces.

IV. EXPOSURE ASSESSMENT

A. Worker Exposure

S. cerevisiae is considered a Class 1 Containment Agent under the National Institute of Health (NIH) Guidelines for Recombinant DNA Molecules (U.S. Department of Health and Human Services, 1986).

No data were available for assessing the release and survival specifically for fermentation facilities using S. cerevisiae. Therefore, the potential worker exposures and routine releases to the environment from large-scale, conventional fermentation processes were estimated on information available from eight premanufacture notices submitted to EPA under TSCA Section 5 and from published information collected from non-engineered microorganisms (Reilly, 1991). These values are based on reasonable worst-case scenarios and typical ranges or values are given for comparison.

During fermentation processes, worker exposure is possible during laboratory pipetting, inoculation, sampling, harvesting, extraction, processing and decontamination procedures. A typical site employs less than 10 workers/shift and operates 24 hours/day throughout the year. NIOSH has conducted walk-through surveys of several fermentation facilities in the enzyme industry and monitored for microbial air contamination. These particular facilities were not using recombinant microorganisms, but the processes were considered typical of fermentation process technology. Area samples were taken in locations where the potential for worker exposure was considered to be potentially greatest, i.e., near the fermentor, the seed fermentor, sampling ports, and separation processes (either filter press or rotary drum filter). The workers with the highest potential average exposures at the three facilities visited were those involved in air sampling. Area samples near the sampling port revealed average airborne concentrations ranging from 350 to 648 cfu/m^3. Typically, the Chemical Engineering Branch would not use area

monitoring data to estimate occupational exposure levels since the correlation between area concentrations and worker exposure is highly uncertain. Personal sampling data are not available at the present time. Thus, area sampling data have been the only means of assessing exposures for previous PMN biotechnology submissions. Assuming that 20 samples per day are drawn and that each sample takes up to 5 minutes to collect, the duration of exposure for a single worker will be about 1.5 hours/day. Assuming that the concentration of microorganisms in the worker's breathing zone is equivalent to the levels found in the area sampling, the worst-case daily inhalation exposure is estimated to range up to 650 to 1200 cfu/day. The uncertainty associated with this estimated exposure value is not known (Reilly, 1991).

B. Environmental and General Exposure

1. Fate of the Organism

S. cerevisiae is a normal inhabitant of soils and is widespread in nature. S. cerevisiae is able to take up a wide variety of sugars and amino acids. These traits enhance the organism's ability for long term survival. S. cerevisiae can be isolated from fruits and grains and other materials with a high concentration of carbohydrates (LaVeck, 1991).

2. Releases

Estimates of the number of S. cerevisiae organisms released during production are tabulated in Table 1 (Reilly, 1991). The uncontrolled/untreated scenario assumes no control features for the fermentor off-gases, and no inactivation of the fermentation broth for the liquid and solid waste releases. The containment criteria required for the full exemption scenario assume the use of features or equipment that minimize the number of viable cells in the fermentor off-gases. They also assume inactivation procedures resulting in a validated 6-log reduction of the number of viable microorganisms in the liquid and solid wastes relative to the maximum cell density of the fermentation broth.

TABLE 1. Estimated Number of Viable Saccharomyces cerevisiae Organisms Released During Production

Release Media	Uncontrolled/ Untreated (cfu/day)	Full Exemption (cfu/day)	Release (days/year)
Air Vents	$2x10^{8}$ - 1x1011	$<2x10^{8}$ - 1x1011	350
Rotary Drum Filter	250	250	350
Surface Water	$7x10^{12}$	$7x10^{6}$	90
Soil/Landfill	$7x10^{14}$	$7x10^{8}$	90

Source: Reilly, 1991

These are "worst-case" estimates which assume that the maximum cell density in the fermentation broth for fungi is 10^{7} cfu/ml, with a fermentor size of 70,000 liters, and the separation efficiency for the rotary drum filter is 99 percent.

3. Air

Specific data which indicate the survivability of S. cerevisiae in the atmosphere after release are currently unavailable. Survival of vegetative cells during aerosolization is typically limited due to stresses such as shear forces, desiccation, temperature, and UV light exposure. As with naturally-occurring strains, human exposure may occur via inhalation as the organisms are dispersed in the atmosphere attached to dust particles, or lofted through mechanical or air disturbance.

Air releases from fermentor off-gas could potentially result in nonoccupational inhalation exposures due to point source releases. To estimate exposures from this source, the sector averaging form of the Gaussian algorithm described in Turner (1970) was used. For purposes of this assessment, a release height of 3 meters and downward contact at a distance of 100 meters were assumed. Assuming that there is no removal of organisms by controls/equipment for off-gases, potential human inhalation dose rates are estimated to range from 3.0×10^{3} to 1.5×10^{6} cfu/year for the uncontrolled/untreated scenario and less than that for systems with full exemptions. It should be noted that these estimates represent hypothetical exposures under reasonable worst case conditions (Versar, 1992).

4. Water

The concentrations of S. cerevisiae in surface water were estimated using stream flow values for water bodies receiving

process wastewater discharges from facilities within SIC Code 283 (drugs, medicinal chemicals, and pharmaceuticals). The surface water release data (cfu/day) tabulated in Table 1 were divided by the stream flow values to yield a surface water concentration of the organism (cfu/l). The stream flow values for SIC Code 283 were based on discharger location data retrieved from the Industrial Facilities Dischargers (IFD) database on December 5, 1991, and surface water flow data retrieved from the RXGAGE database. Flow values were obtained for water bodies receiving wastewater discharges from 154 indirect (facilities that send their waste to a POTW) and direct dischargers facilities that have a NPDES permit to discharge to surface water). Tenth percentile values indicate flows for smaller rivers within this distribution of 154 receiving water flows and 50th percentile values indicate flows for more average rivers. The flow value expressed as 7Q10 is the lowest flow observed over seven consecutive days during a 10-year period. The use of this methodology to estimate concentrations of S. cerevisiae in surface water assumes that all of the discharged organisms survive wastewater treatment and that growth is not enhanced by any component of the treatment process. Estimated concentrations of S. cerevisiae in surface water for the uncontrolled/untreated and the full exemption scenarios are tabulated in Table 2 (Versar, 1992).

TABLE 2. S. cerevisiae Concentrations in Surface Water

Flow	Receiving Stream Flow (MLD*)		Organisms (cfu/l)	
	Mean	7Q10	Mean	7Q10
Uncontrolled/Untreated				
10th Percentile	156	5.60	4.5×10^{4}	1.25×10^{6}
50th Percentile	768	68.13	9.11×10^{3}	1.03×10^{5}
Full Exemption				
10th Percentile	156	5.60	4.5×10^{-2}	1.25×100^{0}
50th Percentile	768	68.13	9.11×10^{-3}	1.03×10^{-1}

*MLD = million liters per day
Source: Versar, 1992

5. Soil

Since soil is a natural habitat for S. cerevisiae, it would be expected to survive well in soil. These releases could result in human and environmental exposure (Versar, 1992). It is

currently estimated that over one million tons of naturally-occurring yeast are produced annually during brewing and distilling practices (LaVeck, 1991).

6. Summary

Although direct monitoring data are unavailable, worst case estimates do not suggest high levels of exposure of S. cerevisiae to either workers or the public resulting from normal fermentation operations.

V. INTEGRATED RISK ASSESSMENT

A. Discussion

There is an extensive history of use of and exposure to S. cerevisiae with a very limited record of adverse effects to the environment or human health. Yeast has been used for centuries as a leavening for bread and fermenter of beer without records of virulence. S. cerevisiae is currently classified as a class 1 containment organism under the NIH Guidelines based largely on the extensive history of safe use.

Factors associated with the development of disease states in fungi have been reviewed. Data suggests that only with the ingestion of high levels of S. cerevisiae or with the use of immunosuppressants can S. cerevisiae colonize in the body. Even under those conditions, there were no noted adverse effects. In the few cases which S. cerevisiae was found in association with a disease state, the host was a debilitated individual, generally with an impaired immune system. In other cases the organism was recovered from an immunologically privileged site (i.e., respiratory tract). Many scientists believe that under appropriate conditions any microorganism could serve as an opportunistic pathogen. The cases noted in the above Human Health Assessment, where S. cerevisiae was found in association with a disease state, appear to be classic examples of opportunistic pathogenicity (see III.A.3).

The organism is not a plant or animal pathogen. Despite the fact that S. cerevisiae is ubiquitous in nature, it has not been found to be associated with disease conditions in plants or animals. The only adverse environmental condition that was noted is the production of "killer toxins" by some strains of the yeast. These toxins have a target range that is limited to susceptible yeasts. The toxins, proteins and glycoproteins, are not expected to have a broad environmental effect based largely on the anticipated short persistence of the toxins in soil or

water and by the limited target range. S. cerevisiae "killer toxin" has been used industrially to provide a level of protection against contamination by other yeasts in the fermentation beer.

The current taxonomy of Saccharomyces is under revision based on the development of alternative criteria. However, this should not have a major effect on the risk associated with closely related species. Saccharomyces, as a genus, present low risk to human health or the environment. Criteria used to differentiate between species are based on their ability to utilize specific carbohydrates without relevance to pathogenicity. Nonetheless, this risk assessment applies to those organisms that fall under the classical definition of S. cerevisiae as described by van der Walt (1971).

S. cerevisiae is a ubiquitous organism which, despite its broad exposure, has very limited reported incidence of adverse effects. The extensive history of use, the diversity of products currently produced by the organism, and the attention given this organism as a model for genetic studies collectively makes this organism a prime candidate for full exemption. The increased knowledge derived from the ongoing research should further enhance this organisms' biotechnological uses.

B. Recommendation

Saccharomyces cerevisiae is recommended for the tiered exemption.

VI. REFERENCES

21 CFR 184.1983.

Anderson, A. 1992. Yeast genome project: 300,000 and counting. Science 256:462.

Barrett-Bree, Y. Hayes, R.G. Wilson, and J.F. Ryley. 1985. A comparison of phospholipase activity, cellular adherence and pathogenicity of yeasts. J. Gen. Microbiol. 131:1217-1221.

Bigelis, R. 1985. Primary metabolism and industrial fermentations. in Gene Manipulations in Fungi. J.W. Bennet and L.L. Lasure (ed.) Academic Press, New York, NY. PP 357-

Brondz, I., and I. Olsen. 1990. Multivariate analyses of cellular carbohydrates and fatty acids of Candida albicans, Torulopsis glabrata, and Saccharomyces cerevisiae. J. Clin. Microbiol. 28:1854-1857.

Buesching, W.J., K. Kurek, and G.D. Roberts. 1979. Evaluation of the modified API 20C system for identification of clinically important yeasts. J. Clin. Microbiol. 9565-569.

Bussey, H., T. Vernet, and A.-M. Sdicu. 1988. Mutual antagonism among killer yeasts: competition between K1 and K2 killers and a novel cDNA-based K1-K2 killer strain of Saccharomyces cerevisiae. Can. J. Microbiol. 34:38-44.

Dynamac. 1990. Evaluation of microorganisms for possible exemption under TSCA section 5. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Dynamac. 1991. Human health assessment for Saccharomyces cerevisiae. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Holzschu, D.L., F.W. Chandler, L. Ajello, and D.G. Ahearn. 1979. Evaluation of industrial yeasts for pathogenicity. Sabouraudia 17:71-78.

Kiehn, T.E., F.F. Edwards, and D. Armstrong. 1980. The prevalence of yeasts in gastrointestinal inoculation in antibiotic treated mice. Sabouraudia 21:27-33.

Kough, J. 1990. Clarification of taxonomic questions on Saccharomyces cerevisiae. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

LaVeck, G. 1991. Exposure assessments of microorganisms considered for 5(h)(4) exemptions under the proposed biotech

rule. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Martini, A. and C. Kurtzman. 1985. Deoxyribonucleic acid relatedness among species of the genus Saccharomyces sensu stricto. Int. J. System. Bacteriol. 35:508-511.

Nobre, G.N., and A.F. Ferreira. 1986. Enhancement of Streptococcus faecalis infection and complement depletion in yeast-treated mice. J. Gen. Microbiol. 132:1277-1281.

Organization for Economic Cooperation and Development. 1986. Recombinant DNA safety considerations. Paris, France.

Phaff, H., M. Miller, and E. Mrak. 1966. The life of yeasts. Harvard University Press, Boston, MA.

Reilly, B. 1991. Analysis of environmental releases and occupational exposure in support of the proposed TSCA 5(h)(4) exemption. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Rosini, G., F. Federici, A.E. Vaughn, and A. Martini. 1982. Systematics of the species of the yeast genus Saccharomyces associated with the fermentation industry. Eur. J. Appl. Biotechnol. 15:188-193.

Sayre, P. 1991. Environmental hazard assessment of Saccharomyces cerevisiae for proposed 5(h)4 exemption. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Stewart, G.C. and I. Russell. 1985. The biology of Saccharomyces, pp. 511-536. In A.L. Demain and N.A. Solomon, (eds.), Biology of industrial organisms. Benjamin Cummins Publishers, Menlo Park.

Tao, J., I. Ginsberg, N. Banerjee, W. Held, Y. Koltin, and J.A. Bruenn. 1990. Ustilago maydis KP6 Killer toxin: structure, expression in Saccharomyces cerevisiae, and relationship to other cellular toxins. Mol. Cell. Biol. 10:1373-1381.

Thayer, D.W. 1990. Personnel communication with C. Felkner. (Cited in Dynamac, 1990).

U.S. Department of Health and Human Services, 1986. Guidelines for research involving recombinant DNA molecules; Notice. 51 FR 16958-16985.

Van der Walt, J. 1971. Saccharomyces, pp. 597-605. In J.Lodder, (ed.), The Yeasts, a taxonomic study. North Holland Publ. Co. Amsterdam.

Versar, 1992. Screening level exposure assessment of Saccharomyces uvarum for the 5(h)(4) exemption under the proposed biotech rule. Unpublished, U.S. Environmental Protection Agency, Washington, D.C.

Wolochow, H., G.J. Hildegrand, and C. Lamanna. 1961. Translocation of microorganisms across the intestinal wall of the rat: effect of microbial size and concentration. J. Infect. Dis. 116:523-528.

Lyophilized or Dried Biological Material Preparations

SECTION 1: IDENTIFICATION

Lyophilized or Dried Biological Material Preparations

Microbiologics, Inc.
200 Cooper Avenue North
St. Cloud, Minnesota 56303 USA
320-253-1640

SECTION 2: HAZARDS INSTRUCTIONS

Microorganism preparations:

No known OSHA hazards

GHS Classification:	None
Signal word:	None
Pictogram:	None

This biological material is classified as either Risk Group 1 or Risk Group 2 by the World Health Organization (WHO). These microorganisms may cause human infection, may pose a hazard to laboratory personnel, but are unlikely to spread in the community. Exposure to these microorganisms in the laboratory rarely causes infection. Effective prevention and treatment is typically available.

SECTION 3: COMPOSITION/INFORMATION ON INGREDIENTS

Each lyophilized or dried pellet, or swab, contains a pure or mixed population of microorganisms. Each pellet or swab also contains excipients for the purpose of structure and/or stability.

SECTION 4: FIRST AID MEASURES

Eyes: Rinse cautiously with water for several minutes. Remove contact lenses, if present and easy to do. Continue Rinsing. If irritation persists: Get medical advice/attention.

Skin: Non-irritant. If skin contact occurs, wash with an appropriate biocidal solution.

Inhalation: Avoid the production of aerosols. If inhalation occurs, move to an area of fresh air and seek medical advice.

Ingestion: Avoid hand to mouth contact. If ingested, seek medical advice.

SECTION 5: FIRE-FIGHTING MEASURES

N/A

SECTION 6: ACCIDENTAL RELEASE MEASURES

In case of accidental spillage, please see LIT.115, Biohazard Cleanup on our website at www.microbiologics.com.

SECTION 7: HANDLING AND STORAGE

The viable biological material preparation must be stored at 2°C - 8°C in the original sealed container. The inactivated biological material preparations must be stored at 2°C - 25°C in the original sealed container. The preparations contain viable microorganisms that may, under certain circumstances, produce disease, or may contain inactivated biological material.

Proper techniques must be employed to avoid exposure and contact with microorganism growth, and rehydrated pellet suspensions. The microbiology laboratory must be equipped, and have the facilities to receive, process, maintain, store and dispose of biohazard material. The microbiology laboratory personnel using these devices must be trained, experienced, and demonstrate proficiency in processing, maintaining, storing and disposing of biohazard material.

SECTION 8: EXPOSURE CONTROLS/PERSONAL PROTECTIVE EQUIPMENT

Good laboratory practices must be observed and followed at all times. The use of a biological safety cabinet, the prevention of aerosols and the use of gloves, moisture impervious aprons, and other protective clothing must be dictated by the standard operational procedures of each individual laboratory.

SECTION 9: PHYSICAL AND CHEMICAL PROPERTIES

Inert, odorless, dry material.

SECTION 10: STABILITY AND REACTIVITY

When stored as directed, the biological material preparations are stable until the last day of the stated month of the expiration date. The length of storage does not affect the risk of infection.

SECTION 11: TOXICOLOGICAL INFORMATION

N/A

SECTION 12: ECOLOGICAL INFORMATION

N/A

SECTION 13: DISPOSAL CONSIDERATIONS

The materials, rehydrated suspensions, and subsequent growth of these microorganisms on culture media, are considered to be biohazard material. Agencies and statutes regulate the disposal of all biohazard materials. Each laboratory must be aware of, and comply with, the proper disposal of biohazard materials.

SECTION 14: TRANSPORT INFORMATION

Refer to national and international regulations regarding the shipment and transport of biohazard materials. UN Classification: UN3373 Biological Substance, Category B for Microbiologics biological material. Select Microbiologics catalog numbers may be shipped according to UN2814, Infectious Substances. Refer to TIB.2023 Strains Shipped UN2814 for the complete list of products shipped according to UN2814.

SECTION 15: REGULATORY INFORMATION

N/A

SECTION 16: OTHER INFORMATION

N/A

The statements contained herein are offered for informational purposes only and are intended to be followed only by persons having related technical skills and at their own discretion and risk. Since conditions and manner of use are outside of our control, we make no warranties, expressed or implied, and assume no liability in connection with any use of this information.

U.S. Department of Labor
Occupation Safety and Health Administration