Molecular Population Genetics and Evolution of a Prion-like Protein in *Saccharomyces cerevisiae*

Mark A. Jensen,* Heather L. True,† Yury O. Chernoff‡ and Susan Lindquist†

**Department of Ecology and Evolution,* †*Department of Molecular Genetics and Cell Biology and Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637 and* ‡*School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332-0230*

Manuscript received January 20, 2000
Accepted for publication August 2, 2001

ABSTRACT

The prion-like behavior of Sup35p, the eRF3 homolog in the yeast *Saccharomyces cerevisiae*, mediates the activity of the cytoplasmic nonsense suppressor known as [*PSI*$^+$]. Sup35p is divided into three regions of distinct function. The N-terminal and middle (M) regions are required for the induction and propagation of [*PSI*$^+$] but are not necessary for translation termination or cell viability. The C-terminal region encompasses the termination function. The existence of the N-terminal region in *SUP35* homologs of other fungi has led some to suggest that this region has an adaptive function separate from translation termination. To examine this hypothesis, we sequenced portions of *SUP35* in 21 strains of *S. cerevisiae*, including 13 clinical isolates. We analyzed nucleotide polymorphism within this species and compared it to sequence divergence from a sister species, *S. paradoxus*. The N domain of Sup35p is highly conserved in amino acid sequence and is highly biased in codon usage toward preferred codons. Amino acid changes are under weak purifying selection based on a quantitative analysis of polymorphism and divergence. We also conclude that the clinical strains of *S. cerevisiae* are not recently derived and that outcrossing between strains in *S. cerevisiae* may be relatively rare in nature.

PRIONS are infectious agents responsible for a wide variety of debilitating diseases in humans and other mammals (Horwich and Weissman 1997; Prusiner *et al.* 1998). While the dynamics of prion transmission are similar to that of infectious microorganisms, prions appear to consist solely of protein, having no associated nucleic acid to mediate replication. In the mammalian case, the prevailing hypothesis suggests that a prion is an ordinary host protein, PrP^C, that has assumed an alternative stable pathogenic structure, called PrP^{Sc}. PrP^{Sc} interacts with PrP^C, inducing PrP^C to change to PrP^{Sc} conformation. A cascading process of recruitment and conversion converts most of the cell's PrP^C protein to the pathogenic form, producing disease and new infectious material.

The discovery of prion-like phenomena in the yeast *Saccharomyces cerevisiae* has led to a new model system for prion study (Tuite and Lindquist 1996; Liebman and Derkatch 1999; Wickner and Chernoff 1999; True and Lindquist 2000). More importantly, it is representative of a novel mechanism for the inheritance of genetic information. A change in phenotype can be passed between generations via a self-perpetuating change in the conformational state of a protein, with no underlying changes in genotype. A series of investigations (Ter-Avanesyan *et al.* 1994; Chernoff *et al.* 1993, 1995; Derkatch *et al.* 1996; Patino *et al.* 1996; Glover *et al.* 1997; Paushkin *et al.* 1997) confirmed the model (Wickner 1994) that the cytoplasmic nonsense-suppression factor [*PSI*$^+$] (Cox 1965) is prion-like, consisting of the protein product of the gene *SUP35*. Sup35p is the yeast homolog of the translation termination factor eRF3. Sup35p in its prion-like conformation forms intracellular aggregates. This sequesters Sup35p, leading to reduced efficiency of polypeptide release and readthrough of all three nonsense codons. Yeast strains exhibiting this prion-mediated nonsense suppression are called [*PSI*$^+$]. [*PSI*$^+$] is a metastable phenomenon, which spontaneously arises and disappears at a rate less than once in 10^6 generations (Y. O. Chernoff, unpublished data).

The potentially deleterious effects of low-level translation termination disruption suggest that [*PSI*$^+$] could be a disease of yeast, by analogy with the mammalian case. However, several authors (Chernoff *et al.* 1992; Tuite and Lindquist 1996; True and Lindquist 2000) have suggested that the ability to switch to the [*PSI*$^+$] state may have adaptive value in *S. cerevisiae*. Their arguments depend in part upon the structure of Sup35p. Sup35p can be divided into three regions, N-terminal, M, and C-terminal (Figures 1 and 3). The N domain is extremely rich in glycine, glutamine, asparagine, and tyrosine, while the M domain harbors an unusual concentration of both positively and negatively charged amino acids. The C region is highly conserved among disparate taxa (Kushnirov *et al.* 1990; Gagny and Silar 1998; Wickner and Chernoff 1999) and can function

Corresponding author: Mark A. Jensen, Department of Microbiology, University of Washington, Seattle, WA 98195-8070.
E-mail: majensen@u.washington.edu

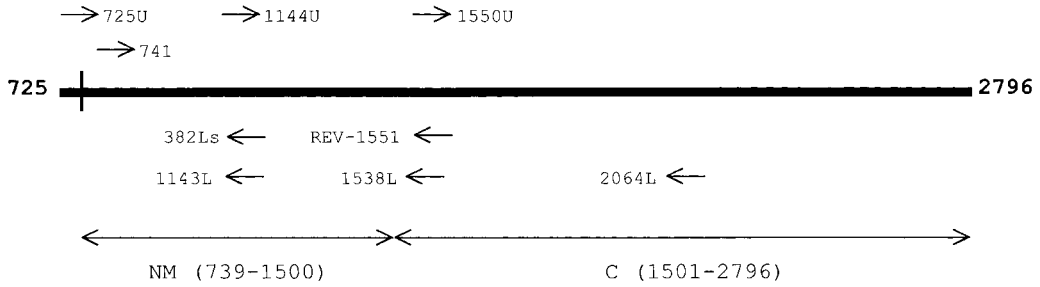

FIGURE 1.—Structure of *SUP35* coding region. Arrows indicate position and orientation of sequencing primers. Hash mark indicates start codon. Coordinates correspond to GenBank accession no. M21129. Primer sequences (5′–3′): 725U, GCCCACTAGCAAC AATGTCGG; 741, GTCGG ATTCAAACCAAGGCA; 1144U, GCCGCTCCCAAACCAAAGAAG; 1538L, ATTTACCGGCATCAACATGAC; 1550U, CCGGTAAAT CTACTATGGGTG; 1143L, ATACCGGAACTGGAGACAAGC; 382Ls, CTGCTTTTGTTGCTTTTGAAAGTC; REV-1551, GGCAT CAACATGACCCATGA; 2064L, CCGGATACTGGCATAAATACA.

alone in the cell to allow wild-type translation termination efficiency (TER-AVANESYAN *et al.* 1993). When the NM or N regions are deleted, however, strains normally able to propagate [*PSI*$^+$] become unable to do so and are referred to as "[*PSI*$^+$] no more" (PNM; TER-AVANESYAN *et al.* 1994). On the other hand, the formation of the [*PSI*$^+$] prion can be induced by transiently overexpressing just the NM or N region on a plasmid (DERKATCH *et al.* 1996). Thus, the prion-forming function associated with the NM domain appears to be separable from the translation termination function associated with the C domain. The persistence of the NM region in *S. cerevisiae* in spite of potential deleterious effects on polypeptide termination, the presence in other yeasts and fungi of a flexible glutamine-rich region in the N terminus of their eRF3 homologs (KUSHNIROV *et al.* 1990; GAGNY and SILAR 1998), and the fact that the ability to acquire and propagate [*PSI*$^+$] is dependent upon the presence of NM and an appropriate balance of the chaperones hsp104 and hsp70 (CHERNOFF *et al.* 1995, 1999; NEWNAM *et al.* 1999; and S. LINDQUIST and J.-J. LIU, unpublished results) make it tempting to speculate that [*PSI*$^+$] itself could serve an adaptive purpose.

As a first step toward understanding the adaptive and other evolutionary properties of *SUP35*, we examined nucleotide polymorphism and divergence within a DNA fragment spanning the N and M domains among 23 *S. cerevisiae* alleles and a longer fragment containing N, M, and a portion of C for a set of 10 alleles (Figure 2). Clinical *S. cerevisiae* strains isolated from different immunocompromised patients worldwide (MCCUSKER *et al.* 1994; CLEMONS *et al.* 1997) provided 13 of the NM alleles. Lab strains of different origins, baking strains, and brewery strains were also included in this analysis. We compared *S. cerevisiae* sequence to homologous sequence from the sibling species *S. paradoxus* (MONTROCHER *et al.* 1998) to examine between-species divergence. The sequence of this species can be aligned base-for-base with the entire *S. cerevisiae* sequence. Other *SUP35* divergence studies (*e.g.*, KUSHNIROV *et al.* 1990) have used yeasts and other taxa too distantly related to allow an unambiguous alignment of the prion-associated NM region.

Our analysis provides evidence that the prion-determining domain of Sup35p, if not an adaptation, is at worst only weakly deleterious. We find that the amino acid sequence of all three regions is constrained, presumably by purifying selection against nonsynonymous nucleotide variants. *SUP35* is also biased in codon usage toward preferred codons, suggesting that selection can influence the frequency of weakly deleterious variants in this gene. It is possible that this purifying selection has relaxed in the NM region in the *S. cerevisiae* lineage. Purifying selection in the C region appears to be ongoing. Using recombination estimates based on linkage disequilibrium, we calculate a crude lower bound for the rate of outcrossing between clonal *S. cerevisiae* lineages.

MATERIALS AND METHODS

Strains: We sequenced alleles from 13 clinical isolates, seven laboratory strains, two commercial strains, and one wild vineyard strain of *S. cerevisiae* and one allele from a wild-caught *S. paradoxus* isolate. Details of the provenance of the strains are in Figure 2. We also included the GenBank sequence of KUSHNIROV *et al.* (1988) in our analyses.

Most of the NM sequences were derived from *SUP35* genes cloned during the course of ongoing studies of phenotypic variation in yeast strains carrying [*PSI*$^+$] in the Lindquist lab. Sequences that include a portion of the C region were obtained as part of a preliminary study of polymorphism in a subset of these strains; industrial strain sequences and part of the *S. paradoxus* sequence were provided by the Chernoff lab.

The sample, although not intentionally assembled for a population genetics study, nevertheless represents a random worldwide sample of *S. cerevisiae* and as such provides a reasonable first approximation of species-wide nucleotide variation at this locus.

Gene regions: For 23 alleles, we analyzed a 754-bp region spanning the N and M region of *SUP35* [corresponding to coordinates 744–1497 of GenBank accession no. M21129 (KUSHNIROV *et al.* 1988)]. For 10 of these, we also analyzed a region that extended 534 bp into the C domain (coordinates 1501 to 2034 of accession no. M21129).

DNA sequencing: Strains were sequenced by one of two methods. For sequencing performed at Chicago, the desired region was amplified from genomic template using PCR and cloned into either pZERO-2.1 (Invitrogen, San Diego) or pmCNMG (Lindquist lab yeast vector, sequence available on request). Two or three clones were sequenced in both direc-

strain	derivation	location[e]	N[f]							M									C
			0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
			8	8	8	9	9	0	0	1	1	1	2	2	2	3	3	3	4	6
			1	2	7	3	5	0	6	4	7	8	1	2	4	4	5	9	1	6
			3	7	9	6	7	8	4	9	7	9	8	3	5	1	5	2	1	9
SL1010-1A[a]	lab		T	A	A	C	T	G	A	T	G	G	G	G	T	G	C	T	C	T
W303	lab		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
74D694	lab		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
S288C[b]	lab		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y55[c]	lab		.	.	G	T	C	.	G	.	.	.	A	A	.	A	C	G	.	.
YJM145	clinical	Kansas City, MO	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YJM210	clinical	Romania	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YJM273	clinical	San Francisco,CA	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
YJM311	clinical	San Francisco,CA	.	.	.	.	.	.	.	.	.	.	A	.	T	A	C	G	.	.
YJM434	clinical		C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
YJM436	clinical	United Kingdom	.	.	.	T	.	.	G	.	.	.	A	.	.	A	.	G	.	.
YJM440	clinical	San Francisco,CA	.	.	.	T	.	.	G	.	.	.	A	.	.	.	.	.	.	.
YJM454	clinical	San Francisco,CA	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
YJM455	clinical	San Francisco,CA	.	.	.	T	.	.	G	.	.	.	A	A	T	A	C	G	.	.
KUSH88	lab		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SK1	lab		.	.	G	T	C	.	G	.	.	.	A	A	A	.	A	C	G	.
BASS	industrial		.	.	.	.	.	.	G	.	T	.	.	A	.	.	.	.	.	.
REDSTAR	industrial		.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.
A155	vineyard	France	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C
YJM128	clinical	Kansas City, MO	.	.	.	T	.	.	G	C	.	.	.	A	A	.	A	C	G	.
YJM308	clinical	S. California	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.
YJM310	clinical	San Francisco,CA	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.
YJM309	clinical	San Francisco,CA	.	.	.	.	.	C	.	.	.	T	.	A	.	.	.	.	.	.
syn/non[d]			s	n	s	s	s	n	n	s	n	n	s	n	n	n	n	s	n	n

FIGURE 2.—*SUP35* polymorphic sites. Strains received from (a) S. Liebman, (b) C. Holm, and (c) J. McCusker (and YJM strains). (d) Site variant is a nonsynonymous or synonymous change. (e) Location of first sampling, if known. (f) *SUP35* region.

tions, with the consensus base among clones accepted. Sequences were run by the Genotyping and Sequencing Center of the Department of Ecology and Evolution at the University of Chicago. For sequencing performed at Georgia Tech, genomic PCR product was used directly in a cycle-sequencing reaction. Sequences were run by the Molecular Genetics Instrumentation Facility of the University System of Georgia, Athens. The differences reflect differences in standard laboratory protocol; no particular difficulty in sequencing any strain was encountered by either group.

For nine of the longer fragments, we obtained sequence by first amplifying genomic DNA using primers SUP35C725U and SUP35C2064L (see Figure 1 legend for primer sequences). PCR products were cloned into a derivative of plasmid pZERO-2.1 (Invitrogen). Clones were sequenced in both directions, using universal M13 primers M13-20 and M13rev and internal primers SUP35C1143L, SUP35C1144U, SUP35C1538L, and SUP35C1550U.

For the strains Red Star, Bass Ale, and SK-1, genomic DNA was amplified with primers SUP35C725U and SUP35C2064L and was gel purified and sequenced directly, using these primers plus internal primers as above.

The NM regions of the remaining *S. cerevisiae* alleles were cloned into Lindquist lab yeast vector pmCNMG at *Bam*HI/*Sac*I sites from genomic PCR products. Plasmid DNA was sequenced directly in both directions using 5CNMSEQ (5′-GTG CAATATCATATAGAAGTCATCG-3′), 3CNMSEQ (5′-CAA CAAGAATTGGGACAACTCCAG-3′), and internal primers SUP35C1144U and SUP35-382Ls. 5CNMSEQ and 3CNMSEQ were designed from vector sequence, with 3′ ends lying 51 and 26 bp from the ends of the insert, respectively.

The NM region of *S. paradoxus* was amplified with primers SUP35-741 and SUP35REV-1551, gel purified, and sequenced directly by using the same primers and internal primers SUP35C1143L and SUP35-741. To amplify the C region of *S. paradoxus*, we performed a first round amplification on genomic DNA with primers SUP35C725U and SUP35C2064L. This product was then amplified in second round reactions containing SUP35C2064L and either SUP35C725U or SUP35C1144U. Both long and short products were cloned and sequenced using universal primers.

Sequences were managed with Sequencher 3.0 software (Gene Codes, Ann Arbor, MI). *S. cerevisiae* sequences have been deposited in GenBank under accession nos. AY052599–AY052611 and AY052613–AY052621. *S. paradoxus* sequence is at accession no. AY052612.

Polymorphism and divergence analysis: Sequence alignments were performed using Sequencher and MegAlign (DNASTAR, Madison, WI). Nucleotide polymorphism and divergence were analyzed using programs DNAsp 3.0 (ROZAS and ROZAS 1999) and SITES 1.1 (WAKELEY and HEY 1997). Simulated distributions for Tajima's test were generated using a web-based program of Y.-X. Fu (http://www.hgc.sph.uth.tmc.edu/fu/; FU 1997). Recombination was estimated using the method of HUDSON (1987) as implemented in SITES. Estimates of the average selection against replacement substitutions were obtained using the method of SAWYER and HARTL (1992), as extended by AKASHI (1995), using a C program (written by M. A. Jensen). SAWYER's (1989) gene conversion test was performed using a C program (written by M. A. Jensen). Source code and other details are available upon request.

Protein constraint and codon bias analysis: Simulation estimates of the substitution rates K_a, K_s, and their distributions were obtained using a program of COMERON (1999). We calculated the empirical distribution of the codon adaptation index (CAI; SHARP and LI 1987) for all named genes >300

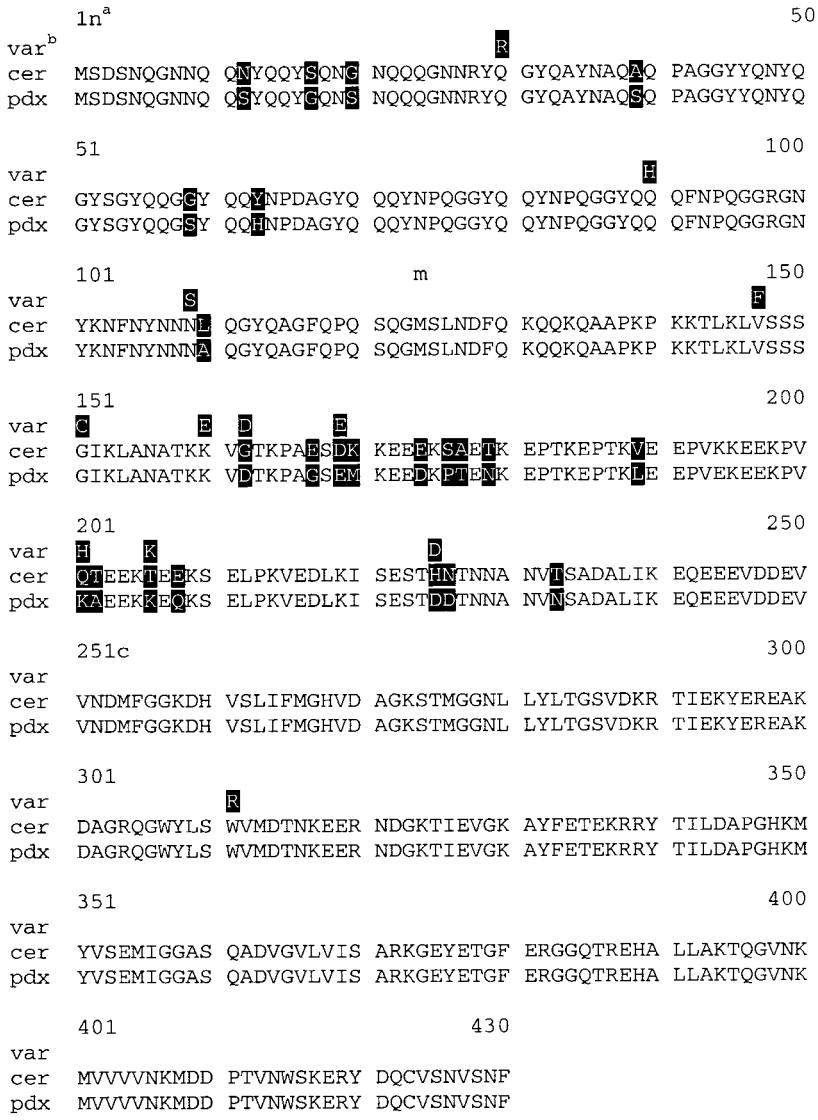

FIGURE 3.—Amino acid comparison of *S. cerevisiae* and *S. paradoxus*. (a) n, m, and c denote MET codons delimiting the N, M, and C domains. (b) For polymorphic amino acids, the rare amino acid is given in this line.

bp in length (obtained October 1999 from the Saccharomyces Genome Database, http://genome-www.stanford.edu/Saccharomyces/) and compared the CAI for *SUP35* to this distribution.

RESULTS

Polymorphism and divergence: We did not observe statistically significant differences between regions N and M for any of our parameter estimates. We therefore combine the two, referring to the NM region in the following analysis. For the NM region of 754 bp, we found 17 segregating sites in 23 alleles. For 534 bp of the C region, we found 1 singleton segregating site in 10 alleles. Polymorphic sites are displayed in Figure 2, and amino acid sequences between *S. cerevisiae* and *S. paradoxus* are compared in Figure 3. Evolutionary parameters are summarized in Table 1. Tajima's *D*, a measure of the departure of the allelic frequency spectrum from neutrality (TAJIMA 1989), is not significant for either region.

The overall divergence in the region between *S. cerevisiae* and *S. paradoxus* is 0.10, using KIMURA's (1981) two-parameter model. The two species' sequences can be unambiguously aligned, without gaps, over the regions analyzed. Table 2 contains numbers of polymorphic sites and fixed differences for each of the regions, separated into synonymous and nonsynonymous changes. Independence tests performed on the 2×2 comparisons for each region constitute the McDonald-Kreitman (MK) test (MCDONALD and KREITMAN 1991) of evolutionary neutrality; *P* values (two-tailed, by Fisher's exact test) are provided in Table 2. Under neutrality, the number of polymorphisms relative to fixed differences should be the same for both synonymous and nonsynonymous sites; the MK test determines whether the data depart significantly from this hypothesis. The MK test within NM shows no significant departure from neutrality. The C region exhibits a departure significant at the 3% level, with the number of synonymous fixed differences exceeding nonsynonymous differences. The

TABLE 1

Polymorphism statistics

Region	n	bp	SS	S_{tot}	S_{SS}	π_{tot} (%)	π_{SS} (%)	θ_{tot} (%)	θ_{SS} (%)	D
NM	23	754	157	17	7	0.52	0.86	0.65	1.20	−0.65
Clinical isolates[a]	13	754	157	13	4	0.51	0.73	0.56	0.82	−0.33
C	10	534	119	1	0	0.03	0.00	0.06	0.00	−0.97
NMC	10	1288	271	16	6	0.34	0.57	0.44	0.76	−0.93

n, number of alleles; bp, length of region in base pairs; SS, available synonymous sites; S_{tot}, number of segregating sites in sample; S_{SS}, number of segregating synonymous sites; π_{tot}, nucleotide diversity per base; θ_{tot}, scaled mutation rate estimate (Watterson 1975) per base; D, Tajima's (1989) D statistic.

[a] Set of clinical isolates only.

test is significant for the entire sequenced region at the 0.5% level.

Protein constraint: We measure evolutionary constraint against amino acid change in *SUP35*, using the rates of nonsynonymous nucleotide substitution, K_a, and synonymous substitution, K_s. The program of Comeron (1999) uses nucleotide divergence between the sister species to calculate observed K_a and K_s for the two protein regions. It then determines confidence intervals for those rates by simulation, generating a distribution by repeatedly applying Poisson mutation processes to the sequence data. In Comeron's method, the means of the synonymous and nonsynonymous processes are the observed rates, corrected for any bias imparted by the multiple-hit correction used to obtain those rates. Mutations to stop codons are not accepted in the simulations. Simulation results for *SUP35* and two other genes are given in Table 3.

Strength of selection: To assess the strength of purifying selection against amino acid changes, we applied the method of (Sawyer and Hartl 1992; Akashi 1995) to our polymorphism and divergence data. The Sawyer-Hartl procedure can be interpreted as a quantitative MK test that calculates the average selection on nonsynonymous changes required to give the observed departure from neutral expectation. The quantity estimated is Ns, the product of the effective population size and the average selection coefficient.

Assuming free recombination among sites in a coding region in which replacement changes are under selection with constant selection coefficient s and synonymous changes are neutral, Sawyer and Hartl (1992) use a diffusion approximation with selection to obtain an expression involving Ns for the expected ratio of the number of nonsynonymous polymorphisms to the number of nonsynonymous fixed differences between samples from two divergent but closely related species. This expression requires an estimate of the divergence time between the two species, which is obtained using the ratio of synonymous polymorphisms to synonymous differences. These expressions can be found conveniently in Akashi (1995). Estimates of Ns are obtained by equating the observed nonsynonymous polymorphism/fixed difference ratio to the expected ratio and solving numerically for Ns. To obtain confidence intervals, following Akashi (1995), we generate a distribution of Ns values by drawing Poisson random variates for each class of site (nonsynonymous polymorphic, nonsynonymous fixed, synonymous polymorphic, and synonymous fixed), from a distribution with mean equal to the observed number in that class, and calculate Ns as above. Source code in C written to perform this analysis is available upon request.

TABLE 2

Polymorphism and divergence

	NM (n = 23)[a]		C (n = 10)		NMC (n = 10)	
	Poly	Div	Poly	Div	Poly	Div
Syn	7	40	0	28	6	68
Non	10	21	1	0	10	21
P	0.09		0.03		0.005	
Ns	−1.16		−2.90		−1.79	
	[−2.25, 0.20]		[−3.77, −1.62]		[−2.87, −0.68]	

Poly, number of polymorphic sites in *S. cerevisiae*; Div, number of fixed sites differing from homologue in *S. paradoxus*; Syn, synonymous sites; Non, replacement sites; P, P value, Fisher's exact test, two-tailed; Ns, estimate of scaled selection coefficient against average replacement, 95% confidence intervals obtained by simulation in brackets.

[a] Number of alleles.

Table 2 contains the median estimates and 95% confidence intervals for Ns, the product of the effective population size and the average selection coefficient. Limits of confidence intervals are the Ns values at the 2.5th and 97.5th percentiles of the simulated distribution. The average amino acid replacement is under weak purifying selection across the two regions, with stronger selection against changes evident in the C-terminal region.

Codon bias: The level of codon bias complements protein sequence constraint as a measure of the long-term importance of a gene to organism fitness. Amino

acid conservation implies functional importance of the particular protein secondary structure. If the choice of codons encoding those constrained amino acids is highly biased, then selection for translational accuracy and efficiency is great enough to significantly constrain the nucleotide sequence of the gene at synonymous sites (see Sharp *et al.* 1995 for review). This provides further support for long-term functional importance of the gene under study.

We calculated the empirical distribution of the CAI (Sharp and Li 1987) for all *S. cerevisiae* named genes of >300 bp in length, as in Moriyama and Powell (1998). This is a set of 3106 open reading frames (October 1999), obtained from the Saccharomyces Genome Database (http://genome-www.stanford.edu/Saccharomyces/). The CAI measures the extent to which codon usage in a particular gene is skewed toward use of the most frequently occurring codon for each amino acid in a set of highly expressed genes. It takes values between 0.0 and 1.0; a gene with a CAI of 1.0 always uses the most frequent codon to encode each amino acid. Our calculation gives a mean CAI of 0.213, comparable to 0.22 obtained for 2169 genes by Moriyama and Powell (1998). CAI values are as follows: *SUP35* (entire gene), 0.333 (88.8th percentile); NM alone, 0.262 (83.7th percentile); C alone, 0.372 (90.3rd percentile). On the basis of the distribution of simulated coding sequences whose amino acid composition and GC content were equivalent to that of the yeast genome, Coghlan and Wolfe (2000) suggest a gene may be considered biased if its CAI value is >0.141, two standard deviations above the mean of their simulated CAI distribution. *SUP35* clearly meets this criterion. These values are consistent with the high level of expression of *SUP35*, which is in the 93rd percentile of genes, as determined by microarray hybridization (Wodicka *et al.* 1997).

Estimates of genetic exchange: Polymorphism data allow us to make inferences concerning the rates of recombination and gene conversion, scaled to the population's effective size (Hudson and Kaplan 1985; Hudson 1987; Hey and Wakeley 1997). The data for NM show evidence of recombination, with the four-gamete test of Hudson and Kaplan revealing three apparent recombination events. Hudson's (1987) estimator of the scaled recombination rate $C = 4NR$, where R is the recombination fraction and N is the effective population size, is 11.25, or 8.7×10^{-3} per base. The value of C/θ_{SS} is 1.24; this is comparable to that found in Drosophila (median for 24 *Drosophila melanogaster* loci, ~1.5; Andolfatto and Przeworski 2000) and humans (~1.3; Hey and Wakeley 1997). This value is much lower than expected given the map length in this region and suggests an excess of linkage disequilibrium among variants (see discussion). Sawyer's (1989) runs test for gene conversion is not significant ($P = 0.45$; number of permutations, 10,000).

TABLE 3

Protein constraint

Gene	K_a	K_s
SUP35NM	0.048	0.224
95.0[a]	[0.029, 0.070]	[0.151, 0.306]
99.9[b]	[0.019, 0.083]	[0.112, 0.368]
SUP35C	0.000	0.233
95.0	N/A	[0.147, 0.332]
99.9	N/A	[0.104, 0.418]
REC104	0.106	0.372
95.0	[0.070, 0.146]	[0.255, 0.522]
99.9	[0.049, 0.181]	[0.182, 0.705]
REC114	0.152	0.364
95.0	[0.124, 0.185]	[0.287, 0.450]
99.9	[0.106, 0.216]	[0.238, 0.517]

K_a and K_s are rates of nonsynonymous and synonymous nucleotide change between the sister species.

[a] 95% confidence interval, based on simulation.

[b] 99.9% confidence interval.

DISCUSSION

Amino acid constraint in Sup35p: We can draw several conclusions of interest. First, the termination function is clearly very highly constrained. There are no nonsynonymous changes between species among the 178 codons of this region that we have analyzed. This is in spite of the fact that the synonymous rates of the two regions are the same; the distributions of K_s for the two regions overlap almost completely (Table 3). We can therefore assume that the intrinsic mutation rates for the two regions are the same. Although we consider only a portion of the termination function, the lack of divergence is still significant. Consider the following comparison with the NM region. Out of 534 sites, 415 are nonsynonymous in the C region (Table 2). Using the lowest rate of nonsynonymous substitution for the NM region, we expect to see at least $(415)(0.019) = 7$ nonsynonymous substitutions in the portion of the C region we examine, were the two regions evolving at the same rate. This suggests that the two regions are under different selective regimes and complements the functional analysis that distinguishes the two regions.

The NM region clearly has been under constraint, though it is more labile than the C region. We see this in two ways. First, K_a is less than K_s, and the distributions do not overlap (Table 3). If K_s is taken as an estimate of the neutral mutation rate, and selection is assumed to be the agent that causes substitution rates to vary, significantly slower replacement site evolution is evidence of purifying selection against replacement changes or "selective constraint" on the NM amino acid sequence (Kimura 1983, p. 44).

Note that this conclusion is conservative. *SUP35* is highly codon biased, so that synonymous changes are

also constrained; the actual neutral rate is likely to be higher than K_s. In other words, if synonymous changes are under purifying selection, then nonsynonymous changes must be as well, according to these results.

Second, we can compare the evolutionary rates at NM with rates at other loci. Ideally, we would compare the K_a of NM with the K_a values of a large set of genes having similar rates of synonymous substitution. We could then ask whether NM is relatively highly constrained, with respect not to a neutral model but to the average expressed coding sequence. At this time, only two *S. paradoxus* coding sequences encoding nonduplicated nuclear genes are currently available for such a comparison. The genes *REC104* and *REC114* are so-called "early exchange" genes, required for meiotic recombination and accurate meiotic division (Jiao *et al.* 1999), and their products are 180 amino acids (aa) and 428 aa long, respectively. *REC104* and *REC114* are under strong constraint between the species, as their respective K_a and K_s distributions are distinct at the 0.1% confidence level (Table 3). While the K_s 95% confidence interval for NM overlaps that of both genes, K_a for NM is less than K_a for the two other genes, its distribution is distinct from the others at the 5% confidence level, and the K_a/K_s ratio of NM is less than that of either gene. The NM region therefore appears to be at least as strongly constrained in amino acid sequence as these important meiotic genes.

Evolutionary conservation of the prion-determining domain: Prions are proteins that can switch to different stable conformations that alter their function in a self-perpetuating way. If the protein in the prion state disrupts the function of the protein in wild-type conformation, as in the case of [*PSI*$^+$], the presence of the prion may be deleterious. In such a case, one may expect natural selection to lead to elimination of the capacity of the protein to adopt the alternate conformation, if constraints on the wild-type function of the protein permit this. In this light, the results of this study are somewhat counterintuitive. In spite of its ability to foster prion formation and its apparent dispensability with respect to the termination function, the prion-determining NM region of Sup35p is significantly conserved in amino acid sequence. In addition, *SUP35* is biased in codon usage across its length, which is evidence that translational accuracy and efficiency are important for both regions.

The biological function of the NM region is unknown. Deletion of the homologous region of *eRF3* in the fungus *Podospora anserina* reduces reproductive efficiency (Gagny and Silar 1998), while, in yeast, the NM region is involved in interactions with a variety of other yeast proteins, including a cytoskeletal assembly component (Bailleul *et al.* 1999). Whatever function NM may have, loss of this function is evidently more deleterious than deleterious effects caused by a heritable reduction of termination efficiency upon periodic manifestation of [*PSI*$^+$].

How deleterious is the prion state in yeast? If a prion is sufficiently deleterious, there will be selective pressure to eliminate it while preserving other functions of the prion-determining domain, if this is possible. Our study suggests that selection against the prion state need not be very strong to affect the sequence of *SUP35*. Codon bias results show that synonymous changes away from preferred codons have been eliminated over time. Such changes are under very weak selection (Akashi 1995). Thus, the sequence of *SUP35* is sensitive to very weak changes in selective pressures; that is, its effective population size is quite high.

Single replacement changes in NM have been identified that prevent [*PSI*$^+$] propagation (PNM mutations) or inhibit nonsense suppression [anti-suppressor (ASU) mutations; Doel *et al.* 1994; DePace *et al.* 1998]. If the prion state were at least as deleterious as the average nonpreferred synonymous change, such mutations would confer an advantage that might cause at least one of them to reach high frequency. Nevertheless, we do not observe any of these mutations in any of our strains. This may be so because the prion state is only weakly deleterious. However, it is possible that in some of these strains the genetic background cannot support [*PSI*$^+$] or that some of the replacement variants observed in this study are PNM or ASU mutations not detected in DePace *et al.*'s (1998) comprehensive screen. It is also possible that these replacements cause deleterious pleiotropic effects on NM's primary function. Since NM can be deleted without affecting viability, and the average replacement change is only weakly deleterious, such a pleiotropic effect is probably weak.

Mode of selection on *SUP35*: The significant McDonald-Kreitman test over the entire fragment indicates a deficiency of amino acid replacements relative to polymorphism. That is, weak purifying selection against replacement changes appears to be acting in the region (see, *e.g.*, the discussion in Nachman 1998). The Sawyer-Hartl analysis indicates that *Ns* against the average replacement change is very likely negative but is relatively small in magnitude, on the order of unity. The distribution of polymorphism within NM is consistent with neutrality, according to the McDonald-Kreitman and Tajima's tests, while *Ns* for changes specifically in NM tends toward purifying selection but is not significantly different from zero. It is possible that a relaxation of purifying selection at NM has taken place in *S. cerevisiae*. Note that this result is not in conflict with our primary result that purifying selection has generally conserved amino acid and codon usage over the period of divergence of the two species. The Sawyer-Hartl analysis is based on polymorphism within the species. This means that it necessarily considers relatively recent evolution of gene frequencies. It also means that sites that are rarely polymorphic, *e.g.*, those sites at which mutation

leads to lethality, are not considered, though selection is clearly operating at such sites. That is, the Sawyer-Hartl results are limited to sites that are not subject to strong selection.

Origin of clinical isolates: The neutral distribution of polymorphism in clinical isolates (see Table 1) suggests that these are not highly "clonal," confirming a study of genomic restriction fragment length polymorphism variation in these strains (Clemons *et al.* 1997). That is, these virulent strains do not appear to be recently descended from a strain that spontaneously acquired virulence determinants. One caveat to this conclusion is that strong clonality (*i.e.*, a low number of haplotypes for a given number of segregating sites) might be expected only in regions rather closely linked to virulence genes, and there is no reason to imagine that *SUP35* would be linked to such genes. The data suggest, however, that recombination events that would break up long-range associations are infrequent; see discussion below.

Frequency of outcrossing: We used Hudson's (1987) method to estimate the scaled recombination rate $C = 4NR$ in the NM region, where R is distance in centimorgans. Since the scaled neutral mutation rate θ equals $4NU$, where U is the mutation rate for the region, the estimated value of C/θ_{SS} would closely estimate the biological parameter c/u in a panmictic sexual population, where c is the coefficient of exchange, or R divided by the number of bases, and u is the mutation rate per base pair per generation.

Taking c for the *SUP35* region to be 3.1×10^{-6} [estimated by comparing physical and genetic maps of chromosome IV (Cherry *et al.* 1997)] and $u = 2.2 \times 10^{-10}$ (the estimated rate of spontaneous mutation per base pair per replication in *S. cerevisiae*; Drake *et al.* 1998), c/u for this region is 1.41×10^4, over 4 orders of magnitude greater than our estimate of C/θ_{SS}, 1.24. This discrepancy indicates an excess of linkage disequilibrium among variants compared to a neutrally evolving panmictic population (Andolfatto and Przeworski 2000).

Several processes are known to increase linkage disequilibrium for a given rate of recombination, including population subdivision (see, *e.g.*, Strobeck 1987), inbreeding (Nordborg 2000), and various types of selection, such as transient directional selection (Hudson *et al.* 1997) or epistatic selection (Ohta 1982). In organisms like Saccharomyces for which sexual reproduction is facultative and infrequent, clonal reproduction will increase linkage disequilibrium as well.

If we make the simplifying assumption that population subdivision, inbreeding, and asexual reproduction are the principal causes of disequilibrium, we can make a crude estimate of the rate at which genetically divergent partially geographically isolated *S. cerevisiae* clones interbreed. Supposing that such outcrossing occurs once every m generations (meiosis or mitosis), we must replace c in the above ratio with c/m. Then C/θ_{SS} estimates c/mu; that is, the frequency of outcrossing is $1/m = (1.24)/(1.41 \times 10^4) = 8.79 \times 10^{-5}$/generation. Outcrossing in nature, therefore, may be rare. Ignoring selective forces that increase disequilibrium, however, makes this rough calculation a lower bound.

We are grateful to S. M. Belenkiy for help in some of the experiments. We thank C. Holm, J. H. McCusker, and S. W. Liebman for generous gifts of yeast strains, and T. Serio for the pmCNMG construct. We also thank P. Andolfatto, J. M. Comeron, M. Kreitman, M. Long, P. Sniegowski, and G. Wyckoff for insightful discussions and comments; B. Charlesworth and R. R. Hudson for their thorough critiques of the manuscript; and K. H. Jensen for support. This work was supported in part by a National Science Foundation Doctoral Dissertation Improvement Grant to M.A.J. and grants R21GM55091 and R01GM-58763 from the National Institute of General Medical Sciences to Y.O.C. H.L.T. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

LITERATURE CITED

Akashi, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics **139:** 1067–1076.

Andolfatto, P., and M. Przeworski, 2000 A genome-wide departure from the standard neutral model in natural populations of Drosophila. Genetics **156:** 257–268.

Bailleul, P. A., G. P. Newnam, J. N. Steenbergen and Y. O. Chernoff, 1999 Genetic study of interactions between the cytoskeletal assembly protein Sla1 and prion-forming domain of the release factor Sup35 (eRF3) in *Saccharomyces cerevisiae*. Genetics **153:** 81–94.

Chernoff, Y. O., M. V. Ptyushkina, M. G. Samsonova, G. I. Sizonencko, Y. I. Pavlov *et al.*, 1992 Conservative system for dosage-dependent modulation of translational fidelity in eukaryotes. Biochemie **74:** 455–461.

Chernoff, Y. O., I. L. Derkach and S. G. Inge-Vechtomov, 1993 Multicopy *SUP35* gene induces *de novo* appearance of *psi*-like factors in the yeast *Saccharomyces cerevisiae*. Curr. Genet. **24:** 286–270.

Chernoff, Y. O., S. L. Lindquist, B. Ono, S. G. Inge-Vechtomov and S. W. Liebman, 1995 Role of the chaperone protein hsp104 in propagation of the yeast prion-like factor [*PSI*$^+$]. Science **268:** 880–884.

Chernoff, Y. O., G. P. Newnam, J. Kumar, K. Allen and A. D. Zink, 1999 Evidence for a "protein mutator" in yeast: role of the Hsp70-related chaperone Ssb in formation, stability and toxicity of the [*PSI*] prion. Mol. Cell. Biol. **19:** 8103–8112.

Cherry, J. M., C. Ball, S. Weng, G. Juvik, R. Schmidt *et al.*, 1997 Genetic and physical maps of *Saccharomyces cerevisiae*. Nature **387:** 67–73.

Clemons, K. V., P.-S. Park, J. H. McCusker, M. J. McCullough, R. W. Davis *et al.*, 1997 Application of DNA typing methods and genetic analysis to epidemiology and taxonomy of Saccharomyces isolates. J. Clin. Microbiol. **35:** 1822–1828.

Coghlan, A., and K. H. Wolfe, 2000 Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. Yeast **16:** 1131–1145.

Comeron, J. M., 1999 K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics **15:** 763–764.

Cox, B. S., 1965 ψ, a cytoplasmic suppressor of super-suppressor in yeast. Heredity **20:** 505–521.

DePace, A. H., A. Santosos, P. Hillner and J. S. Weissman, 1998 A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. Cell **93:** 1241–1252.

Derkatch, I. L., Y. O. Chernoff, V. V. Kushnirov, S. G. Inge-Vechtomov and S. W. Liebman, 1996 Genesis and variability

of [*PSI*] prion factors in *Saccharomyces cerevisiae*. Genetics **144:** 1375–1386.

Doel, S. M., S. J. McCready, C. R. Nierras and B. S. Cox, 1994 The dominant *PNM*$^-$ mutation which eliminates the *psi* factor of *Saccharomyces cerevisiae* is the result of a missense mutation in the *SUP35* gene. Genetics **137:** 659–670.

Drake, J. W., B. Charlesworth, D. Charlesworth and J. Crow, 1998 Rates of spontaneous mutation. Genetics **148:** 1667–1686.

Fu, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147:** 915–925.

Gagny, B., and P. Silar, 1998 Identification of the genes encoding the cytosolic translation release factors from *Podospora anserina* and analysis of their role during the life cycle. Genetics **149:** 1763–1775.

Glover, J. R., A. S. Kowal, E. C. Schirmer, M. M. Patino, J.-J. Liu *et al.*, 1997 Self-seeded fibers formed by Sup35, the protein determinant of [*PSI*$^+$], a heritable prion-like factor of *S. cerevisiae*. Cell **89:** 811–819.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Horwich, A. L., and J. S. Weissman, 1997 Deadly conformations—protein misfolding in prion disease. Cell **89:** 499–510.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hudson, R. R., A. G. Saez and F. J. Ayala, 1997 DNA variation at the *Sod* locus of *Drosophila melanogaster*: an unfolding story of natural selection. Proc. Natl. Acad. Sci. USA **94:** 7725–7728.

Jiao, K., S. A. Bullard, L. Salem and R. E. Malone, 1999 Coordination of the initiation of recombination and reductional division in meiosis in *Saccharomyces cerevisiae*. Genetics **152:** 117–128.

Kimura, M., 1981 Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA **78:** 454–458.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kushnirov, V. V., M. D. ter-Avanesyan, M. V. Telckov, A. P. Surguchov, V. N. Smirnov *et al.*, 1988 Nucleotide sequence of the *SUP2* (*SUP35*) gene of *Saccharomyces cerevisiae*. Gene **66:** 45–54.

Kushnirov, V. V., M. D. ter-Avanesyan, S. A. Didichenko, V. N. Smirnov, Y. O. Chernoff *et al.*, 1990 Divergence and conservation of *SUP2* (*SUP35*) gene of yeasts *Pichia pinus* and *Saccharomyces cerevisiae*. Yeast **6:** 461–472.

Liebman, S. W., and I. L. Derkatch, 1999 The yeast [*PSI*$^+$] prion: making sense of nonsense. J. Biol. Chem. **274:** 1181–1184.

McCusker, J. H., K. V. Clemons, D. A. Stevens and R. W. Davis, 1994 Genetic characterization of pathogenic *Saccharomyces cerevisiae* isolates. Genetics **136:** 1261–1269.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution and the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

Montrocher, R., M.-C. Verner, J. Briolay, C. Gautier and R. Marmeisse, 1998 Phylogenetic analysis of the *Saccharomyces cerevisiae* group based on polymorphisms of rDNA spacer sequences. Int. J. Syst. Bacteriol. **48:** 295–303.

Moriyama, E. N., and J. R. Powell, 1998 Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae*, and *Escherichia coli*. Nucleic Acids Res. **26:** 3188–3193.

Nachman, M. W., 1998 Deleterious mutations in animal mitochondrial DNA. Genetica **102(103):** 61–69.

Newnam, G. P., R. D. Wegrzyn, S. L. Lindquist and Y. O. Chernoff, 1999 Antagonistic interactions between yeast chaperones Hsp-104 and Hsp70 in prion curing. Mol. Cell. Biol. **19:** 1325–1333.

Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics **154:** 923–929.

Ohta, T., 1982 Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc. Natl. Acad. Sci. USA **79:** 1940–1944.

Patino, M. M., J.-J. Liu, J. R. Glover and S. Lindquist, 1996 Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. Science **273:** 622–626.

Paushkin, S. V., V. V. Kushnirov, V. N. Smirnov and M. D. Ter-Avanesyan, 1997 In vitro propagation of the prion-like state of yeast Sup35 protein. Science **277:** 381–383.

Prusiner, S. B., M. R. Scott, S. J. DeArmond and F. E. Cohen, 1998 Prion protein biology. Cell **93:** 337–348.

Rozas, M., and R. Rozas, 1999 DNASP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Sawyer, S., 1989 Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:** 526–538.

Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics **132:** 1161–1176.

Sharp, P. M., and W.-H. Li, 1987 The codon adaptation index: a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:** 1281–1295.

Sharp, P. M., M. Averof, A. T. Lloyd, G. Matassi and J. Peden, 1995 DNA sequence evolution: the sounds of silence. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **349:** 241–247.

Strobeck, C., 1987 Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics **117:** 149–153.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **129:** 753–762.

Ter-Avanesyan, M. D., V. V. Kushnirov, A. R. Dagkesamanskaya, S. A. Didichenko, Y. O. Chernoff *et al.*, 1993 Deletion analysis of the *SUP35* gene of the yeast *Saccharomyces cerevisiae* reveals two non-overlapping functional regions in the encoded protein. Mol. Microbiol. **7:** 683–692.

Ter-Avanesyan, M. D., A. R. Dagkesamanskaya, V. V. Kushnirov and V. N. Smirnov, 1994 The *SUP35* omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [*PSI*$^+$] in the yeast *Saccharomyces cerevisiae*. Genetics **137:** 671–676.

True, H. L., and S. L. Lindquist, 2000 A yeast prion provides a mechanism for genetic variation and phenotypic diversity. Nature **407:** 477–483.

Tuite, M. F., and S. Lindquist, 1996 Maintenance and inheritance of yeast prions. Trends Genet. **12:** 467–471.

Wakeley, J., and J. Hey, 1997 Estimating ancestral population parameters. Genetics **145:** 847–855.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wickner, R. B., 1994 [*URE3*] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. Science **264:** 566–568.

Wickner, R., and Y. O. Chernoff, 1999 Prions of fungi: [*URE3*], [*PSI*], and [*Het-s*] discovered as heritable traits, pp. 229–272 in *Prion Biology and Diseases*, edited by S. B. Prusiner. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Wodicka, L., H. Dong, M. Mittman, M. H. Ho and D. J. Lockhart, 1997 Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat. Biotechnol. **15:** 1359–1367.

Communicating editor: W. Stephan