2013

ZRT1 harbors an excess of nonsynonymous polymorphism and shows evidence of balancing selection in Saccharomyces cerevisiae

Elizabeth K. Engle
Washington University School of Medicine in St. Louis

Justin C. Fay
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation
https://digitalcommons.wustl.edu/open_access_pubs/4513
ZRT1 Harbors an Excess of Nonsynonymous Polymorphism and Shows Evidence of Balancing Selection in Saccharomyces cerevisiae

Elizabeth K. Engle* and Justin C. Fay†,1
*Molecular Genetics and Genomics Program and †Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University, St. Louis, Missouri 63108

ABSTRACT Estimates of the fraction of nucleotide substitutions driven by positive selection vary widely across different species. Accounting for different estimates of positive selection has been difficult, in part because selection on polymorphism within a species is known to obscure a signal of positive selection among species. While methods have been developed to control for the confounding effects of negative selection against deleterious polymorphism, the impact of balancing selection on estimates of positive selection has not been assessed. In Saccharomyces cerevisiae, there is no signal of positive selection within protein coding sequences as the ratio of nonsynonymous to synonymous polymorphism is higher than that of divergence. To investigate the impact of balancing selection on estimates of positive selection, we examined five genes with high rates of nonsynonymous polymorphism in S. cerevisiae relative to divergence from S. paradoxus. One of the genes, the high-affinity zinc transporter ZRT1 showed an elevated rate of synonymous polymorphism indicative of balancing selection. The high rate of synonymous polymorphism coincided with nonsynonymous divergence among three haplotype groups, among which we found no detectable differences in ZRT1 function. Our results implicate balancing selection in one of five genes exhibiting a large excess of nonsynonymous polymorphism in yeast. We conclude that balancing selection is a potentially important factor in estimating the frequency of positive selection across the yeast genome.

KEYWORDS yeast balancing selection zinc McDonald-Kreitman test

The frequency of adaptive substitutions driven by positive selection is central to our understanding of molecular evolution and divergence among species. The neutral theory assumes that most substitutions are effectively neutral and generates predictions that can be tested based on patterns of molecular evolution (Fay and Wu 2003). While many individual genes have been found to deviate from neutral patterns of evolution, the overall impact of positive selection across the genome remains a contentious issue (Hahn 2008; Sella et al. 2009; Nei et al. 2010; Fay 2011).

Genome-wide comparisons of polymorphism vs. divergence have been the primary means of estimating the frequency of positive selection among species. The McDonald-Kreitman (MK) test (McDonald and Kreitman 1991) has been used to estimate the frequency of positive selection within protein coding sequences based on an elevated ratio of nonsynonymous to synonymous divergence relative to that of polymorphism. However, applications of the MK test to plant, animal, and microbial genomes have revealed substantial differences in estimates of positive selection among species, ranging from zero to over half of all amino acid substitutions (Fay 2011). While the frequency of positive selection may differ due to a species’ effective population size and species-specific selective pressures (Bachtrog 2008; Gossmann et al. 2010; Siol et al. 2010; Slotte et al. 2010; Gossmann et al. 2012), estimating the frequency of positive selection during divergence among species depends on controlling for the effects of selection on polymorphism within species (Fay and Wu 2001; Bierne and Eyre-Walker 2004; Hughes et al. 2008).

Estimates of the frequency of positive selection can be influenced by a number of factors that can make it difficult to detect adaptation when it is present. Slightly deleterious polymorphisms segregate at low
frequencies due to weak negative selection and can increase the nonsynonymous-to-synonymous polymorphism ratio to a greater extent than that of divergence. As a consequence, deleterious polymorphism can obscure evidence of positive selection (Fay et al. 2002; Bierne and Eyre-Walker 2004; Charlesworth and Eyre-Walker 2008; Eyre-Walker and Keightley 2009). Methods have been developed to account for the effects of low-frequency deleterious polymorphism, but, even so, there are still some species with little or no evidence of positive selection (Fay 2011).

A number of other factors can influence the detection of positive selection through their effects on slightly deleterious polymorphism. Such factors include mating system as well as population size and structure. For example, a decrease in population size can increase the abundance of slightly deleterious polymorphism within a species and obscure evidence of positive selection among species (Eyre-Walker 2002). In humans, there is little evidence for an excess of nonsynonymous divergence, yet it has been estimated that up to 40% of amino acid substitutions could have been driven by positive selection without being detected (Eyre-Walker and Keightley 2009). Controlling for these additional factors is often difficult as it requires specific knowledge of the species being examined and its population history.

Another factor that has received less attention but can also influence estimates of positive selection is balancing selection (Wright and Andolfatto 2008). The maintenance of multiple nonsynonymous polymorphisms within a species by balancing selection could increase the genome-wide ratio of nonsynonymous to synonymous polymorphism within a species above the ratio of nonsynonymous-to-synonymous divergence among species. Elevated rates of nonsynonymous polymorphism may also occur because of local adaptation (Charlesworth et al. 1997). If an appreciable number of genes are involved in adaptive divergence between different populations of the same species, genome-wide estimates of the frequency of positive selection among species could be substantially underestimated.

The yeast Saccharomyces cerevisiae is one species with little or no evidence of positive selection based on the MK test (Doniger et al. 2008; Liti et al. 2009; Elyashiv et al. 2010). In contrast to other species that lack evidence of positive selection (Foxe et al. 2008; Gossmann et al. 2010; Gossmann et al. 2012), its large effective population size ensures the efficient removal of weakly deleterious mutations and the ability to fix weakly advantageous mutations. However, S. cerevisiae also exhibits strong population structure, potentially facilitated by its low rate of outcrossing, i.e., mating between unrelated parents (Ruderfer et al. 2006), low rate of migration, or local adaptation to the diverse array of environments from which it has been isolated (Fay and Benavides 2005). Genome-wide patterns of population structure have revealed a number of genetically differentiated groups, including strains originating from sake in Japan, vineyards in Europe, and oak trees in North America (Liti et al. 2009; Schacherer et al. 2009). While these groups may have arisen as a result of geographic barriers, they might also have arisen as a consequence of domestication or adaptation to human-modified environments (Fay and Benavides 2005). However, even when these groups are taken into consideration and examined separately, the ratios of nonsynonymous to synonymous polymorphism within or between groups are higher than the ratio of nonsynonymous to synonymous divergence among species (Elyashiv et al. 2010).

In this study, we tested the hypothesis that genes with a large excess of nonsynonymous polymorphism are underbalancing selection. We reasoned that such genes have a disproportionate effect on estimates of positive selection and should be considered separately if underbalancing selection. We examined five genes that were previously shown to contain a large excess of nonsynonymous to synonymous polymorphism (Doniger et al. 2008; Liti et al. 2009). To distinguish between purifying selection and balancing selection on nonsynonymous polymorphism, we examined rates of synonymous polymorphism as negative selection is expected to decrease linked neutral variation, whereas balancing selection is expected to increase linked neutral variation (Charlesworth et al. 1997). We found one of the genes, ZRT1, showed a significantly elevated rate of synonymous polymorphism based on the Hudson-Kreitman-Aguade (HKA) test (Hudson et al. 1987), consistent with balancing selection. Our results show that a large number of amino acid polymorphisms can occur at certain loci, underbalancing selection.

**MATERIALS AND METHODS**

**Polymorphism and divergence data**

Data were collected for five genes that were previously found to exhibit an excess of nonsynonymous polymorphism in two studies (Doniger et al. 2008; Liti et al. 2009) and 30 randomly selected control genes, using 36 S. cerevisiae strains with genome sequence data (Supporting Information, Table S1). Twenty-seven of the genome sequences were accessed through a BLAST server (www.moseslab.csb.utoronto.ca/sgrp/), and the other 9 were accessed through the Saccharomyces Genome Database (www.yeastgenome.org). For each gene, sequences homologous to the coding region of the reference genome (S288C) were aligned using Clustal X version 2.0 software (Larkin et al. 2007). Strains with sequences that were <99% and <90% of the S288C sequence length for the neutral and selected genes, respectively, were removed. The cutoff of <90% was used to accommodate IRA2, for which few strains had BLAST hits covering the entire 9.2 kb of coding sequence present in the reference genome. Strains were removed from a gene analysis if a polymorphism led to an internal stop codon (4 cases), while unique single-base insertions were considered sequencing errors and the base was removed from the sequence (10 cases). A small number of heterozygous sites were present within the strains Vin13, VL3, and LalvinAQ23. At these sites, we randomly selected one of the two observed nucleotides to represent the position. Divergence was measured by comparison to the CBS432 strain of Saccharomyces paradoxus.

The final dataset included an average of 29.9 strain alleles per gene, ranging from 23 to 36, and the five-gene set included an average of 21.4 strain alleles per gene, ranging from 18 to 24. Eight of the control genes were removed from analysis because they had few strains with sufficient sequence coverage or multiple strains with framehifts. One gene, RPS28B, was removed because it showed evidence of introgression between species (Doniger et al. 2008). The final dataset is available in File S1.

**Population genetic analysis**

MK tests were conducted using the number of nonsynonymous and synonymous polymorphic sites and fixed differences calculated using DnaSP version 5.10.01 software (Librado and Rozas 2009). The weighted neutrality index (Stoletzki and Eyre-Walker 2011) was estimated by the equation:

\[
NI = \frac{\sum D_{ni} P_{ni}}{(P_n + D_n)} - \frac{\sum P_{ni} D_{ni}}{(P_n + D_n)}
\]

where P and D are the number of polymorphic sites and fixed differences, respectively; subscripts s and n indicate synonymous
and nonsynonymous changes, respectively; and \( i \) indicates the \( i \)th gene.

HKA tests were conducted using maximum likelihood HKA test (MLHKA version 2.0) software (Wright and Charlesworth 2004) with rates of synonymous polymorphism and divergence obtained from DnaSP (Table S2). Each of the five genes found to be significant by the MK test were compared to 21 control genes, covering 22,974 sites, using the MLHKA test. The program was run with a chain length of 100,000 interactions for all analyses.

For analysis of the region surrounding \( ZRT1 \), from \( MNT2 \) through \( FZ1 \) (~11 kb), we downloaded \( S. \) cervisiae and \( S. \) paradoxus strain sequences from the Saccharomyces Genome Resequencing Project (Liti et al. 2009). A sliding window analysis of polymorphism and divergence was calculated with DnaSP using 36 \( S. \) cervisiae strains and 1 \( S. \) paradoxus strain, CBS432, with gaps in the alignment excluded. Because of difficulty in aligning the 4.67-kb noncoding region between \( A D H4 \) and \( ZRT1 \), we used only ~200 bases downstream of \( A D H4 \) and 800 bases upstream of \( ZRT1 \), where we were confident of alignment.

Bootstraped neighbor-joining trees for \( ZRT1 \) and the concatenated control gene set were constructed using MEGA5 and pairwise gap removal (Tamura et al. 2011).

**Strain construction and phenotype analysis**

\( ZRT1 \) was deleted in YFI186 (YPS163 background, \( M A T \ a \), \( H O:: \) dsdAMX4, \( u r a3-140 \)) by using the kanMX deletion cassette (Wach et al. 1994). Three \( ZRT1 \) alleles were integrated into this strain at the \( u r a3 \) locus by amplifying the entire \( u r a3 \) gene region, including 878 bases of the 5’ noncoding region and the entire 195 bases of the 3’ noncoding region, as well as 186 bases of the 3’ gene \( FZ1 \), using primers with homology to pRS306, and transforming the product along with the yeast integrative plasmid pRS306 (Sikorski and Hieter 1989). Integration of these constructs at the \( u r a3 \) locus was achieved by selection on plates lacking uracil, and each transformant was confirmed by PCR. The \( ZRT1 \) alleles were found to have between 1 and 3 mutations. However, most alleles had only single synonymous changes or changes within the 5’ or 3’ regions, and no mutations were shared among the alleles, including the replicated transformants. These mutations were considered not functional because of the lack of any phenotypic effects. Wild-type (YFI186) and \( ZRT1 \) deletion strains were integrated with the empty plasmid pRS306 as a control.

Experiments comparing growth under low zinc conditions were conducted using low zinc media (LZM) composed of 0.17% yeast nitrogen base without amino acids, (NH\(_4\))\(_2\)SO\(_4\), or zinc (MP Biomedicals); 0.5% (NH\(_4\))\(_2\)SO\(_4\); 20 mM trisodium citrate, pH 4.2; 2% glucose; 1 mM Na\(_2\)EDTA; 25 \( \mu \)M MnCl\(_2\) and 10 \( \mu \)M FeCl\(_3\), as previously described (Gitan et al. 1998; Gitan et al. 2003). Strains were grown overnight in LZM, washed, and diluted to a starting 

### Table 1 McDonald-Kreitman (MK) test results

<table>
<thead>
<tr>
<th>Gene(s)</th>
<th>Number of Sites</th>
<th>( P_n )</th>
<th>( P_s )</th>
<th>( D_n )</th>
<th>( D_s )</th>
<th>MK P Value</th>
<th>Neutrality Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRA2</td>
<td>9117</td>
<td>69</td>
<td>108</td>
<td>131</td>
<td>735</td>
<td>0.000000</td>
<td>3.6</td>
</tr>
<tr>
<td>OPT2</td>
<td>2436</td>
<td>35</td>
<td>20</td>
<td>18</td>
<td>187</td>
<td>0.000000</td>
<td>18.2</td>
</tr>
<tr>
<td>PEP1</td>
<td>4293</td>
<td>48</td>
<td>45</td>
<td>168</td>
<td>320</td>
<td>0.002240</td>
<td>2.0</td>
</tr>
<tr>
<td>SAS10</td>
<td>1680</td>
<td>12</td>
<td>7</td>
<td>41</td>
<td>120</td>
<td>0.002186</td>
<td>5.0</td>
</tr>
<tr>
<td>ZRT1</td>
<td>1113</td>
<td>55</td>
<td>45</td>
<td>12</td>
<td>62</td>
<td>0.000000</td>
<td>6.3</td>
</tr>
<tr>
<td>5 Genes</td>
<td>18,639</td>
<td>219</td>
<td>225</td>
<td>370</td>
<td>1424</td>
<td>0.000000</td>
<td>3.8</td>
</tr>
<tr>
<td>21 Genes</td>
<td>22,815</td>
<td>158</td>
<td>260</td>
<td>711</td>
<td>1623</td>
<td>0.000000</td>
<td>1.3</td>
</tr>
</tbody>
</table>

\( P_n \) and \( P_s \) are the number of nonsynonymous (\( P_n \)) and synonymous (\( P_s \)) polymorphic sites. \( D_n \) and \( D_s \) are the number of nonsynonymous (\( D_n \)) and synonymous (\( D_s \)) fixed differences. The 5 genes and 21 control genes are described in the text.

### RESULTS

**Identification of genes exhibiting an excess of amino acid polymorphism**

From two previous independent genome-wide screens based on the McDonald-Kreitman (MK) test, we identified five genes that were significant (\( P < 0.001 \)) in both studies (Doniger et al. 2008; Liti et al. 2009). The five genes were \( IRA2, OPT2, PEP1, SAS10, \) and \( ZRT1 \). All of the genes showed a ratio of nonsynonymous to synonymous polymorphism (\( P_n/P_s \)) that was more than twofold greater than that of divergence (\( D_n/D_s \)). We repeated the MK test using a strain set composed of 36 strains for which genome sequences are publicly available (see Materials and Methods and Table S1), of which 29 were not included in previous screens. All five genes retained significance according to the MK test results (\( P < 0.05 \), Bonferroni corrected) (Table 1). As a control, we randomly selected 21 genes from those not significant in either previous study. Only two of the genes were significant according to the MK test (\( P < 0.05 \), Bonferroni corrected) (Table S2), both were characterized as having a \( P_n/P_s \) ratio greater than that of \( D_n/D_s \).

The two gene sets exhibited marked differences in their neutrality indexes (Table 1). The weighted neutrality index (Stoletzki and Eyre-Walker 2011) is 3.80 for the five selected genes and 1.33 for the 21 control genes. The high neutrality index of the five genes, indicating an excess of nonsynonymous polymorphism, is unlikely a consequence of selective pressure on synonymous sites as average codon bias is similar between the two groups, and the five genes have nearly equal numbers of changes to preferred and unpreferred codons for both polymorphism, 63 and 58 respectively, and divergence, 414 and 397, respectively. Thus, the five-gene set is highly enriched for genes with an excess of nonsynonymous polymorphism relative to that of divergence.
shown for the
cleotide diversity

Balancing selection in ZRT1
A high ratio of nonsynonymous to synonymous polymorphism can result from slightly deleterious nonsynonymous mutations that contribute to polymorphism but not divergence or from a recent loss of functional constraint. In either scenario, the rate of synonymous polymorphism should not be affected. Alternatively, a high rate of nonsynonymous polymorphism can result from balancing selection on multiple nonsynonymous alleles. If balancing selection is responsible for the elevated rate of nonsynonymous polymorphism, rates of linked synonymous polymorphism should also be elevated (Charlesworth et al. 1997).

To test whether the rate of synonymous polymorphism was elevated in any of the five genes, we used the HKA test (Hudson et al. 1987). Using a maximum likelihood MLHKA test, we found only ZRT1 showed a significantly elevated rate of synonymous polymorphism in comparison to that of the control gene set (Table S3). Figure 1 shows that of all the genes we tested, ZRT1 was characterized by an exceptionally high rate of synonymous polymorphism. One gene in the control gene set, TRX2, appeared to be an outlier characterized by both a high rate of synonymous polymorphism and a low rate of synonymous divergence. TRX2 is nominally significant for an excess synonymous polymorphism relative to the remaining neutral gene set, by MLHKA test results ($P = 0.0132$). However, removal of TRX2 from the neutral gene set only increased the significance of ZRT1. Thus, of the five genes, only ZRT1 exhibits evidence of balancing selection.

Balancing selection of ZRT1 is also supported by patterns of nonsynonymous and synonymous divergence among strains. A neighbor-joining tree of ZRT1 shows that all of the ZRT1 alleles, except for the EC1118 allele, cluster into three groups distinguished by multiple nonsynonymous and synonymous differences (Figure 2). With the exception of the wine strain group, the groups showed no close correspondence to the source from which each strain was obtained or to a neighbor-joining tree generated from the concatenated control gene set (Figure 2 and Figure S1). Of the 111 polymorphic sites used to generate the tree, 86 can be placed on a single branch without homoplasious traits, of which 24 nonsynonymous and 17 synonymous changes occurred on one of the four main internal branches (Figure 2). In comparison, 62 synonymous but only 12 nonsynonymous differences separate S. cerevisiae from S. paradoxus. Thus, the high ratio of nonsynonymous to synonymous polymorphism is not limited to external branches, as would be expected to occur if most nonsynonymous polymorphisms were deleterious.

The presence of intermediate frequency alleles, many of which contribute to the unique grouping of ZRT1 alleles, also supports balancing selection. For synonymous sites, results of Tajima’s D test were positive across all strains ($D = 0.718$, $P > 0.10$) but negative within each of the three ZRT1 strain groups (M22 group $D = -1.265$; YPS163 group $D = -0.59894$; and S288C group $D = -0.61182$; all $P > 0.10$). In comparison, the average D result of the control gene set was $0.294$, and only four of the genes had positive D values greater than 0.3. These results further highlight the unique pattern of variation present in ZRT1.

Regional variation around ZRT1
The elevated rate of synonymous polymorphism in ZRT1 could be a consequence of balancing selection on ZRT1 amino acid polymorphism but also could be caused by selection on its promoter or on adjacent genes. To determine whether the signal of balancing selection extends into adjacent genes and gene regions, we applied the MLHKA test to ADH4 and EZF1, the two genes adjacent to ZRT1. Only ADH4

Exceptional diversity in ZRT1

Figure 1 shows that of all the genes we tested, ZRT1 was characterized by an exceptionally high rate of synonymous polymorphism. One gene in the control gene set, TRX2, appeared to be an outlier characterized by both a high rate of synonymous polymorphism and a low rate of synonymous divergence. TRX2 is nominally significant for an excess synonymous polymorphism relative to the remaining neutral gene set, by MLHKA test results ($P = 0.0132$). However, removal of TRX2 from the neutral gene set only increased the significance of ZRT1. Thus, of the five genes, only ZRT1 exhibits evidence of balancing selection.

Balancing selection of ZRT1 is also supported by patterns of nonsynonymous and synonymous divergence among strains. A neighbor-joining tree of ZRT1 shows that all of the ZRT1 alleles, except for the EC1118 allele, cluster into three groups distinguished by multiple nonsynonymous and synonymous differences (Figure 2). With the exception of the wine strain group, the groups showed no close correspondence to the source from which each strain was obtained or to a neighbor-joining tree generated from the concatenated control gene set (Figure 2 and Figure S1). Of the 111 polymorphic sites used to generate the tree, 86 can be placed on a single branch without homoplasious traits, of which 24 nonsynonymous and 17 synonymous changes occurred on one of the four main internal branches (Figure 2). In comparison, 62 synonymous but only 12 nonsynonymous differences separate S. cerevisiae from S. paradoxus. Thus, the high ratio of nonsynonymous to synonymous polymorphism is not limited to external branches, as would be expected to occur if most nonsynonymous polymorphisms were deleterious.

The presence of intermediate frequency alleles, many of which contribute to the unique grouping of ZRT1 alleles, also supports balancing selection. For synonymous sites, results of Tajima’s D test were positive across all strains ($D = 0.718$, $P > 0.10$) but negative within each of the three ZRT1 strain groups (M22 group $D = -1.265$; YPS163 group $D = -0.59894$; and S288C group $D = -0.61182$; all $P > 0.10$). In comparison, the average D result of the control gene set was $0.294$, and only four of the genes had positive D values greater than 0.3. These results further highlight the unique pattern of variation present in ZRT1.

Regional variation around ZRT1
The elevated rate of synonymous polymorphism in ZRT1 could be a consequence of balancing selection on ZRT1 amino acid polymorphism but also could be caused by selection on its promoter or on adjacent genes. To determine whether the signal of balancing selection extends into adjacent genes and gene regions, we applied the MLHKA test to ADH4 and EZF1, the two genes adjacent to ZRT1. Only ADH4
was significant in comparison to the control gene set (MLHKA test, 
\(P = 0.0007\)) (Table S3) and was characterized by both high rates of 
polymorphism and also low rates of divergence at synonymous sites 
(Figure 1). Hence, we also tested MNT2, the next gene adjacent to 
\(ADH4\), and found no significant departure from neutrality as mea-
sured by the MLHKA test. Additionally, none of the three adjacent 
genes that were examined showed a significant excess of amino acid 
polymorphism as measured by the MK test (Table S2).

To more precisely track the signal of balancing selection within 
and around \(ZRT1\), we used a sliding window analysis of polymor-
phism to divergence, including both coding and noncoding regions. 
Figure 3 shows that the highest rate of polymorphism occurred within 
the coding region of \(ZRT1\) and extended into its 5’ noncoding region. 
The overall rate of polymorphism is much lower in the two adjacent 
genes \(ADH4\) and \(FZF1\). In \(ADH4\), the rate of divergence is also quite 
low and likely contributes to the significance of the MLHKA test. 
Interestingly, a portion of MNT2 had a very low rate of polymor-
phism, whereas its more distal portion had another peak of polymor-
phism. Based on the sliding window analysis and the MLHKA test 
results, the signature of balancing selection appears to be concentrated 
at the \(ZRT1\) locus.

We next examined the degree to which polymorphism within 
\(ZRT1\) is independent of polymorphism within adjacent genes. There 
is ample evidence for recombination within and around \(ZRT1\). Across 
the entire region, from MNT2 through FZF1, there have been a 
minimum of 26 recombination events based on the four-gamete 
test (Hudson and Kaplan 1985). As expected in the presence of 
recombination, the genealogies of \(ADH4\) and \(FZF1\) differed from 
that of \(ZRT1\) (Figure S2), although all three genes showed a similar 
grouping of wine strains. \(ADH4\) was the most similar to \(ZRT1\) but 
had less divergence. As measured by the HKA test, \(ZRT1\) showed 
significantly elevated rates of polymorphism compared to \(FZF1\) 
\((P = 0.0087)\) but not \(ADH4\) or MNT2 \((P > 0.05)\).

\textbf{ZRT1 alleles confer no detectable} 
\textbf{phenotype differences}

If selection has acted on \(ZRT1\), then different alleles of \(ZRT1\) should 
confer different phenotypes. \(ZRT1\) is a high-affinity zinc transporter 
that is activated only when zinc levels are very low and facilitates 
growth under limiting zinc conditions (Zhao and Eide 1996). We 
compared the effects of three \(ZRT1\) alleles integrated into a strain in 
which the endogenous \(ZRT1\) gene was deleted. The three alleles were 
from the S288C (laboratory), M22 (wine), and YPS163 (nature) 
strains and were selected as representatives from the three major 
groups of strains (Figure 2). While deletion of \(ZRT1\) resulted in 
a significant growth defect in zinc-limiting conditions and each of 
the three alleles rescued the growth deficiency, we found no signif-
icant difference among the three \(ZRT1\) alleles for maximum growth 
(Figure 4) or growth rate (not shown).

In addition to its requirements for growth, zinc is an essential 
cofactor for many enzymes, including alcohol dehydrogenase, and has 
been shown to influence rates of fermentation (De Nicola and Walker 
2011). To test whether the \(ZRT1\) alleles affect rates of fermentation, we 
measured CO\(_2\) release during fermentation of grape juice into wine. 
In the presence of metal chelators, deletion of \(ZRT1\) had a dramatic effect 
on the rate of fermentation, but no differences were found among the 
three \(ZRT1\) alleles tested (analysis of variance [ANOVA], \(P > 0.05\)) 
(Figure 5). No differences in rates of fermentation were found among 
any of the four strains in grape juice without chelators at any of the 
time points (ANOVA, \(P > 0.05\)).

\textbf{DISCUSSION}

Application of the MK test to a variety of species has revealed 
substantial differences in the estimated frequency of positive selection 
on protein coding sequences (Fay 2011). While differences in effective 
population size are capable of explaining some of the differences 
among species (Eyre-Walker and Keightley 2009; Gossmann et al. 
2010; Halligan et al. 2010; Siol et al. 2010; Slotte et al. 2010; Gossmann 
et al. 2012), a small effective population cannot explain the absence 
of evidence for positive selection in yeast. In this study, we examined 
whether balancing selection can explain the high rate of nonsynon-
ymous polymorphism observed in a small set of genes exhibiting 
a disproportionately large excess of nonsynonymous polymorphism, 
as this could obscure evidence of positive selection in \(S.\) \textit{cerevisiae}. 
We showed that one of the five genes tested exhibited a significantly 
elevated rate of synonymous polymorphism, indicative of balancing 
selection. While patterns of polymorphism and divergence around 
\(ZRT1\) suggest that nonsynonymous polymorphism within \(ZRT1\) itself 
is the most likely target of balancing selection, we found no functional 
differences among three alleles, using two different phenotype assays. 
Our results illustrate how balancing selection might obscure a signal of 
positive selection.

\textbf{Balancing selection in \(ZRT1\)}

Evidence for balancing selection on \(ZRT1\) is based on an elevated rate 
of synonymous polymorphism as measured by the HKA test (Figure 1 
and Table S3), a high ratio of polymorphism to divergence that is 
centered on \(ZRT1\) (Figure 3), an increased frequency of intermediate 
frequency alleles, and the coincidence of multiple synonymous and 
nonsynonymous changes that distinguish three groups of strains (Fig-
ure 2). However, it is worth noting that balancing selection in the
that much of the nonsynonymous polymorphism may be weakly measured by the MK test. Genome-wide estimates in yeast suggest an excess of nonsynonymous polymorphism as ZRT1 for most of the control gene set but a positive Tajima's D value at ZRT1, consistent with balancing selection. However, our results are not able to distinguish between these different forms of selection, but rather distinguish them from patterns that can be explained by population structure, loss of selective constraint, and selection on adjacent genes.

In S. cerevisiae, there is extensive population structure related to both geographic origin and the ecological source from which each strain was isolated (Fay and Benavides 2005; Liti et al. 2009; Schacherer et al. 2009), which are frequently correlated with one another. Such groups include sake strains from Japan, oak tree strains from North America, and strains isolated from Europe or vineyards. The neighbor-joining tree of the 21 control genes generally recapitulates these previously defined groups. While the ZRT1 tree bears some resemblance to that of the control gene set, particularly the vineyard group, the three main groups of strains that differ at ZRT1 are not obviously related by either their geographic origin or the ecological source from which they were isolated. More importantly, population structure by itself does not explain the elevated rates of polymorphism at ZRT1 relative to the 21-control-gene set. In support of selection acting at ZRT1, we observed negative Tajima's D values for most of the control gene set but a positive Tajima's D value at ZRT1, consistent with balancing selection.

Loss of functional constraint or weak negative selection is another explanation for an excess of nonsynonymous polymorphism as measured by the MK test. Genome-wide estimates in yeast suggest that much of the nonsynonymous polymorphism may be weakly deleterious (Elyashiv et al. 2010). In the case of ZRT1, we cannot exclude the possibility that some of the nonsynonymous polymorphisms are neutral or slightly deleterious. However, two lines of evidence indicate that at least some of the nonsynonymous changes within ZRT1 have been underbalancing selection. First, many neutral and most deleterious polymorphisms are expected to be rare and present only in a small number of strains. While 78% of nonsynonymous alleles are at less than 10% frequency in the 21-control-gene set, only 38% of nonsynonymous polymorphism are at less than 10% frequency in ZRT1. In addition, of the nonsynonymous changes that are specific to one or more lineages, 55% are positioned along the four internal branches that distinguish the three major groups of ZRT1 alleles. Second, neither loss of constraint or negative selection on nonsynonymous polymorphism should increase variation at linked synonymous sites.

While patterns of variation within and around ZRT1 indicated that it is the most likely target of balancing selection, selection on linked sites could have influenced observed patterns of variation at ZRT1. Patterns of polymorphism within FZF1, the adjacent gene, indicated no excess of synonymous or nonsynonymous polymorphism. However, FZF1 may have experienced a recent selective sweep as there is evidence of positive selection during S. cerevisiae and S. paradoxus divergence, both within its coding region and between the intergenic region between ZRT1 and FZF1 (Sawyer et al. 2005; Engle and Fay 2012). Patterns of polymorphism within the adjacent genes ADH4 and MNT2 are more complex. Both genes show rates of synonymous polymorphism that are higher than those of the 21-control-gene set, except for the outlier gene TRX2. MNT2 shows regions with high and low polymorphism levels, but the region closest to ZRT1 has the lower rate of polymorphism (Figure 3). ADH4 shows a significantly elevated rate of synonymous polymorphism relative to divergence by
the genome (Smith and Kruglyak 2008), and alleles of ever, phism do not show a tendency toward an excess of nonsynonymous
balancing selection throughout the yeast genome (Cutter and Moses
nonsynonymous and synonymous polymorphism are correlated
genome is more dif
measured by the HKA test.
for numerous environment-specified constraint or slightly deleterious polymorphism. Interestingly, alleles of other four genes is most likely a consequence of loss of functional
morphism according to MK test results, only
fi
fected the rate of nonsynonymous polymorphism across many
genes without generating a strong effect on linked synonymous sites.

Phenotypic effects of ZRT1 alleles
We found that the three distinct ZRT1 alleles conferred no detectable
phenotypic differences from one another. Although the alleles were not integrated at the endogenous ZRT1 locus, each was able to fully
rescue the ZRT1 deletion phenotype (Figure 4). This result indicates
that under the conditions tested, none of the nonsynonymous differences
among the three alleles caused a substantial loss of ZRT1 function.
The lack of phenotypic differences among the different ZRT1
alleles implies that either the alleles are functionally equivalent to one
another and so are not involved in balancing selection or that the lack
of a discernible phenotype is a consequence of the conditions tested or
an effect too small to be detected. For example, ZRT1, as a metal transporter, could also influence fitness due to transport of other
metals, such as cadmium (Gitan et al. 1998; Gomes et al. 2002; Gitan
et al. 2003).

Prevalence of balancing selection
Of the five genes that exhibited an excess of nonsynonymous polymorphism according to MK test results, only ZRT1 showed evidence of
balancing selection. The excess of nonsynonymous polymorphism in the
other four genes is most likely a consequence of loss of functional
constraint or slightly deleterious polymorphism. Interestingly, alleles of IRA2, a GTPase that negatively regulates RAS signaling, are responsible
for numerous environment-specific differences in gene expression across
the genome (Smith and Kruglyak 2008), and alleles of IRA2 also have been shown to affect high temperature growth (Parts et al. 2011). How-
ever, IRA2 does not show an excess of synonymous polymorphism as
measured by the HKA test.

The prevalence of balancing selection across the entire yeast genome is more difficult to assess. The observation that rates of nonsynonymous and synonymous polymorphism are correlated
with one another provides some evidence for the possibility of weak
balancing selection throughout the yeast genome (Cutter and Moses
2011). However, genes with high rates of synonymous polymorphism do not show a tendency toward an excess of nonsynonymous polymorphism (Kendall’s tau = −0.15) (Figure 6) as predicted by the MK test using the data of Liti et al. (2009). The challenge to interpreting genome-wide evidence for balancing selection is that many cases of balancing selection may be difficult to detect. First, the effect of balancing selection on linked variation decreases as a function of
the rate of recombination; nucleotide diversity is 1 + 1/4 Nr(1-F)
relative to a neutral locus, where N is the effective population size and r is the rate of recombination and F is the inbreeding coefficient
(Charlesworth et al. 1997). Using a rate of recombination of 3.5 × 10−8/bp, a rate of outcrossing of 2 × 10−3/generation, and an effective
population size of 1.6 × 107 cells (Ruderfer et al. 2006), we expect diversity to be increased by a factor of 10 and 2, 13 bp and
113 bp from a site underbalancing selection, respectively. Gene con-
version is expected to narrow this window even further (Andolfatto and Nordborg 1998). Second, balancing selection must act over many
generations, on the order of the effective population size (Navarro et al. 2000), to noticeably influence linked neutral variation. However,
the ability to detect balancing selection may be increased if there are
multiple selected sites at a single locus, which might be the case for
genes identified by a high rate of nonsynonymous polymorphism. Thus,
it is hard to rule out the possibility that balancing selection has inflated the rate of nonsynonymous polymorphism across many
genes without generating a strong effect on linked synonymous sites.

Why is there little evidence of adaptive evolution within
the yeast genome?
An important and persistent question in genome-wide estimates of
adaptive evolution based on the MK test is why some species show
high rates of adaptive evolution whereas others, such as yeast, do not.
A small effective population size is one explanation as adaptive
substitutions are expected to be more infrequent and deleterious
polymorphism more common. This provides a reasonable explanation
for the absence of signal in humans and many plant species (Eyre-Walker and Keightley 2009; Gossmann et al. 2010; Halligan et al. 2010; Siol et al. 2010; Slotte et al. 2010; Gossmann et al. 2012). However, it does not explain the lack of signal in yeast, which has a large effective population size, on the order of $10^7$ for $S$. paradoxus (Tsai et al. 2008) and $S$. cerevisiae (Ruderfer et al. 2006). The rate of outcrossing may also be relevant to detecting selection in yeast. Selfing helps purge recessive deleterious alleles but also limits recombination between different haplotypes. Despite the presence of selfing in yeast, $S$. cerevisiae exhibits an excess of rare nonsynonymous polymorphism indicative of deleterious alleles and a rapid decay in levels of linkage disequilibrium, an observation that can be attributed to its exceptionally high rate of recombination even if outcrossing is rare. Thus, there is no obvious aspect of $S$. cerevisiae diversity that distinguishes it from outcrossing species. Furthermore, both a selfing and an outcrossing species of Arabidopsis show no signal of adaptive evolution (Foxe et al. 2008). As it stands, neither population size nor selfing provide a particularly compelling explanation for why yeast show little adaptive evolution based on the MK test.

In the present study, we considered the possibility that balancing selection obscured patterns of positive selection in yeast. While we focused only on a small number of genes exhibiting a large excess of nonsynonymous polymorphism, we found one that exhibited evidence of balancing selection. Thus, while our results highlight the evidence of balancing selection. We conclude that balancing selection is a potentially important factor in estimating the frequency of positive selection in yeast. While not emphasized here, it is also important to consider whether adaptive evolution is rare and estimates of positive selection are inflated in other species (Fay 2011).

ACKNOWLEDGMENT

This work was supported by National Institutes of Health grant GM080669.

LITERATURE CITED


Communicating editor: D. J. Gresham

Volume 3 April 2013 | Balancing Selection of ZRT1 | 673