Similar Mutation Rates but Highly Diverse Mutation Spectra in Ascomycete and Basidiomycete Yeasts

Hongan Long*, Megan G. Behringer, Emily Williams, Ronald Te, and Michael Lynch

Department of Biology, Indiana University, Bloomington, IN

*Corresponding author: E-mail: longhongan@gmail.com.

Accepted: November 23, 2016

Data deposition: This project has been deposited in NCBI SRA with the BioProject accession number of PRJNA321129 (study no.: SRP074684).

Abstract

Yeast species are extremely diverse and not monophyletic. Because the majority of yeast research focuses on ascomycetes, the mutational determinants of genetic diversity across yeast species are not well understood. By combining mutation-accumulation techniques with whole-genome sequencing, we resolved the genomic mutation rate and spectrum of the oleaginous (oil-producing) 'red yeast' Rhodotorula toruloides, the first such study in the fungal phylum Basidiomycota. We find that the mutation spectrum is quite different from what has been observed in all other studied unicellular eukaryotes, but similar to that in most bacteria—a predominance of transitions relative to transversions. Rhodotorula toruloides has a significantly higher A:T→G:C transition rate—possibly elevated by the abundant flanking G/C nucleotides in the GC-rich genome, as well as a much lower G:C→T:A transversion rate. In spite of these striking differences, there are substantial consistencies between R. toruloides and the ascomycete model yeasts: a spontaneous base-substitution mutation rate of 1.90\times10^{-10} per site per cell division as well as an elevated mutation rate at non-methylated 5'CpG3' sites. These results imply the evolution of variable mutation spectra in the face of similar mutation rates in yeasts.

Key words: spontaneous mutation, mutation accumulation, yeast evolution, evolutionary genomics.

Introduction

In order to systematically understand the evolution of mutational features, such as the mutation rate and spectrum, investigations of organisms broadly sampled from the Tree of Life are required. In contrast to recent progress on bacterial mutation rates (Lee et al. 2012; Long et al. 2015a; Dillon et al. 2015; Sung et al. 2015; Kucukyildirim et al. 2016), studies on unicellular eukaryotes remain rare. Currently unbiased mutation spectra are available in only a handful of species, and the only two fungi are ascomycetes (Sung et al. 2012; Zhu et al. 2014; Behringer and Hall 2015; Farlow et al. 2015; Ness et al. 2015), which comprise less than 1% of all fungal species (Kurtzman and Piskur 2006). Most recently, studies of spontaneous mutation have primarily relied on mutation-accumulation techniques, repeatedly bottlenecks large numbers of parallel lineages through single individual/colony transfers (Bateman 1959; Mukai 1964; Kibota and Lynch 1996; Halligan and Keightley 2009). Such experimental design maximizes genetic drift, overwhelming the ability of selection to eliminate mutations prior to accumulation. Therefore, unless associated with extremely large fitness effects, most mutations have an equal chance of accumulating in the genome.

Rhodotorula toruloides, a yeast in the phylum Basidiomycota (~62% GC content; ~20 Mb genome size; synonym: Rhodosporidium toruloides) (Wang et al. 2015), has been given the nickname ‘red yeast’ due to the red color conferred by cellular carotenoids. This particular yeast is also oleaginous (oil-producing), with 70% of its cellular dry weight consisting of lipids. Previous spontaneous mutation studies on yeasts have been confined to the two AT-rich model species Saccharomyces cerevisiae (budding yeast, ~38% GC content) and Schizosaccharomyces pombe (fission yeast, ~36%), both of which belong to the phylum Ascomycota and are approximately 800 million years diverged from R. toruloides (Lynch et al. 2008; Parfrey et al. 2011; Zhu et al. 2014; Behringer and Hall 2015; Farlow et al. 2015). Recent comparison of spontaneous mutations in fission and budding yeasts showed that the per site mutation rate is fairly constant, while the mutation spectrum differs slightly with respect to G:C→T:A transversions (Behringer and Hall 2015; Farlow et al. 2015). By exploring spontaneous mutation in a...
GC-rich yeast of another phylum, we will gain a greater understanding of variation in the mutation rate and spectrum of yeasts. Thus, we have conducted a mutation-accumulation experiment on *R. toruloides*, resolved its mutation rate and spectrum, and contrasted these with the patterns in *S. pombe* and *S. cerevisiae*.

**Materials and Methods**

**Strains and Transfers**

The haploid *R. toruloides* strain was ordered from American Type Culture Collection (ATCC, catalogue number: 10788). From a single cell ancestor we created 60 MA lines. Single colonies were transferred by re-streaking onto fresh Yeast Mold agar plates (Becton Dickinson, catalogue number: 271210) at 25 °C every other day. Every month, single colonies from 10 randomly selected MA lines were cut from agar plates and serially diluted to count colony forming units (CFU), and the mean cell divisions passed from a single cell to a colony were estimated by log₂(CFU). The total number of cell divisions of each MA line is the product of the grand mean (19.3) of all cell division estimates and the total transfer number for each line. On an average, each MA line experienced 181 transfers. The MA lines transferring took 362 days. We excluded two MA lines (RT6, RT10) from the final mutation analyses due to genome library construction failure.

DNA Extraction, Library Construction, and Genome Sequencing

DNA was extracted from the resulting MA lines using a phenol–chloroform protocol (Sambrook et al. 1989) and libraries constructed using the Nextera® DNA Library Preparation Kit (Illumina). Fragments were size-selected for an insert size of 300 bp and sequenced at the Hubbard Center for Genome Studies, University of New Hampshire on a HiSeq2500 2 × 150 bp rapid run. Across the 58 samples, a median coverage of 70× coverage was achieved.

Mutation Analyses

We first trimmed library adaptors of paired-end reads with Trimmmomatic 0.32 (Bolger et al. 2014), and then mapped reads to the reference genome of *R. toruloides* NP11 (GenBank genome accession number: GCA_000320785.2) using BWA mem -ver. 0.7.10 (Li and Durbin 2009). Duplicate reads were removed using picardtools-1.141 and reads mapping around indels were realigned across all 58 MA lines using GATK-3.5 before preforming SNP and indel discovery with standard hard filtering parameters described by GATK Best Practices recommendations (except that we set the Phred-scaled quality score QUAL > 100 and RMS mapping quality MQ > 59 for both variant and non-variant sites) (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). Base-pair substitutions and small indels were called using the UnifiedGenotyper in GATK. We also required greater than 99% of reads in a line to determine the line-specific consensus nucleotide at a candidate site—1% was set to allow for aberrant reads originated from sequencing errors, not absolutely pure indexes during library construction, or barcodes degeneracy during sequence demultiplexing. In order to confirm the presence of mutations that were identified bioinformatically, 96 base-substitution and 36 indel calls were randomly sampled and PCR primers were designed using Batch Primer3 (You et al. 2008). We successfully PCR amplified and Sanger sequenced 89 base-substitution mutations and found no false positives (0 false positive rate) and validated 27 small indels out of 28 successfully sequenced candidates, i.e. 3.57% false positive rate for indels. These low error rates showed the high fidelity of our mutation calls. Mutation rate μ was calculated by

\[
μ = \frac{m}{N \times T}
\]

where \( m \) is the total number of mutations across all MA lines, \( N \) is the total number of lines, \( N \) is the analyzed sites in one line, and \( T \) is the number of cell divisions passed in the entire mutation accumulation process of the MA line. 95% Poisson confidence intervals of mutation rates were calculated by using an exact Poisson test in R (poisson.test). Equilibrium GC content under mutation pressure alone was calculated by \( p = \frac{C}{G+p} \) (Lynch 2007), where \( u \) is the mutation rate from A/T to G/C nucleotides and \( v \) is the sum of the mutation rate in the other direction; standard error of \( p \) was calculated by following Lynch and Walsh (1998):

\[
SE(p) = \sqrt{\frac{p \times (1-p) \times \sigma^2(u) + 2 \sigma(u,v) \times \sigma^2(v) + \sigma^2(v)}{N^2}}
\]

where \( \sigma^2(u) \) and \( \sigma^2(v) \) are the sample variance of \( u \) and \( v \) across all MA lines, respectively, \( \sigma(u,v) \) is the covariance of \( u \) and \( v \)—calculated by the mean value of \( (u - \mu_u)(v - \mu_v) \) across all lines and \( N \) is the total number of MA lines in the final mutation analysis, i.e. 56. Standard error of the transition/transversion (ts/tv) ratio was also calculated following this formula and 95% confidence intervals of the ratio using ts/tv ± 1.96 × SE.

Mutation rate outliers were determined by being greater than \( Q3 + 1.5 \times IQR \) or \( Q1 - 1.5 \times IQR \), where \( Q3 \) is the third quartile, \( Q1 \) is the first quartile, IQR—interquartile range—is the distance between the first and the third quartile (Mcgill et al. 1978); mutation enrichment analysis of genes—mutations were limited to coding regions—was adapted from Long et al. (2016): briefly, for each gene, given the expected mutation number of each gene across all MA lines, we estimated the Poisson probability of finding greater than or equal to the observed number of mutations. The expected number of mutations per gene was calculated as the product of the mutation rate per site per generation multiplied by the gene length and the total generations of all lines. The cutoff \( P \) value was determined using Bonferroni correction, i.e. \( P = 0.05/G - 1 \), where \( G \) equals 8,171—the total number of genes in the genome.
Using the BAM files for all MA lines from the above pipelines, we analyzed the presence of structural variants using the Delly software package (workflow available on GitHub, https://github.com/behrimg/R_toruloides) (Rausch et al. 2012). For structural variants present on the shorter contigs (contigs 49–94), variants were identified visually with the Integrated Genome Viewer (IGV v. 2.3.5) (Thorvaldsdottir et al. 2013). Potential large-scale insertions due to mobile genetic elements were additionally detected using RetroSeq (workflow available on GitHub, https://github.com/behrimg/R_toruloides) and a library of all reported fungal mobile elements, created using RepBase Update v. 20.12 (Jurka et al. 2013). We parsed all possible synonymous and non-synonymous changes in the coding region of the R. toruloides genome and calculated the expected non-synonymous:synonymous mutation ratio to be 2.76 in the absence of selection, which is not significantly different from the observed ratio of 2.53 ($\chi^2 = 0.22$, $df = 1$, $P = 0.64$). Thus, selection did not significantly bias the accumulation of mutations.

Among these mutations, transitions are the most abundant, with a ts/tv ratio of 1.83 (95% confidence intervals: 1.54, 2.12). This ratio is ~1.9$\times$ higher than that of S. cerevisiae (~0.95; Zhu et al. 2014) and 2.4$\times$ that of S. pombe (~0.75; Behringer and Hall 2015; Farlow et al. 2015). This ts/tv ratio is also higher than those of all other unicellular eukaryotes with resolved mutation spectra, but similar to the ts/tv ratios of bacteria (table 2).

We are also curious about if the ts/tv ratio in natural strains of R. toruloides is different from the above MA lines. We thus aligned raw genome sequences of ATCC10788 (the strain used in this study) and JCM10049 to the NP11 reference genome, analyzed SNPs (single nucleotide polymorphisms), and estimated the natural ts/tv ratio at replacement sites to be 1.72. This ratio is not significantly different from that of the MA lines, implying that natural selection may not heavily influence the segregating transitions.

The high ts/tv ratio arises because the mutation spectrum of R. toruloides is strikingly different from the ascomycete yeasts. The A:T$ightarrow$G:C transition mutation rate in R. toruloides is significantly higher than those in the ascomycete yeasts, could be from inadvertent selection on these genes, though the biological details are not clear due to the incomplete status of the genome and annotation. We thus removed the two outlier MA lines—56 MA lines were used in the final mutation analyses—and the mutations in the two enriched genes from the following analyses. The remaining 714 base-pair substitutions yield a mutation rate of 1.90$\times$10$^{-10}$ (95% Poisson confidence intervals: 1.76–2.04$\times$10$^{-10}$) (table 1 and fig. 1; supplementary tables S1–S5).

### Results

In order to resolve the mutation rate and spectrum, we sequenced 58 MA lines, which were established from a single-cell ancestor and single-colony transferred every other day for 362 days. Each MA line experienced ~181 single-cell bottlenecks and on average ~3,486 cell divisions. The single-cell bottlenecks provide strong genetic drift, which dominates selection during clonal expansion, allowing mutations to accumulate almost neutrally (Bateman 1959; Mukai 1964; Kibota and Lynch 1996).

### Base-Substitution Mutation Rate and Spectrum

We detected 789 base-substitution mutations from the 58 successfully sequenced MA lines. Two MA lines (RT17, RT30) were identified as outliers (supplementary fig. S1, Supplementary Material online), which have significantly higher mutation rate than other lines. Because specific genes enriched with mutations could bias the mutation rate and spectrum, we performed a mutation enrichment test for the coding region of each gene and found two genes with significantly higher numbers of mutations than others: gene4182 (coding one argonaute) and gene3886, a hypothetical protein without known significant homology with any other organisms in NCBI databases. Such enrichment might influence the segregation transitions.

<table>
<thead>
<tr>
<th>Substitutions</th>
<th>Count</th>
<th>$\mu$</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G:C$\rightarrow$A:T</td>
<td>288</td>
<td>1.23</td>
<td>1.10, 1.39</td>
</tr>
<tr>
<td>A:T$\rightarrow$G:C</td>
<td>174</td>
<td>1.22</td>
<td>1.04, 1.41</td>
</tr>
<tr>
<td>Transversions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A:T$\rightarrow$T:A</td>
<td>23</td>
<td>0.16</td>
<td>0.10, 0.24</td>
</tr>
<tr>
<td>G:C$\rightarrow$T:A</td>
<td>75</td>
<td>0.32</td>
<td>0.25, 0.40</td>
</tr>
<tr>
<td>A:T$\rightarrow$C:G</td>
<td>68</td>
<td>0.48</td>
<td>0.37, 0.60</td>
</tr>
<tr>
<td>G:C$\rightarrow$C:G</td>
<td>86</td>
<td>0.37</td>
<td>0.29, 0.46</td>
</tr>
<tr>
<td>Insertions</td>
<td>33</td>
<td>0.09</td>
<td>0.06, 0.12</td>
</tr>
<tr>
<td>Deletions</td>
<td>50</td>
<td>0.13</td>
<td>0.10, 0.18</td>
</tr>
</tbody>
</table>
while the G:C→T:A transversion mutation rate is significantly lower (see details below; fig. 1). Consistent with observations in fission and budding yeasts (Zhu et al. 2014; Behringer and Hall 2015; Farlow et al. 2015), mutations at 5’CpG3’ sites are significantly elevated above those at cytosine sites without a flanking 3’G (χ² = 132.84, P < 2.2 × 10⁻¹⁶). As in these other species, R. toruloides is not known to have a DNA C-5 methyltransferase. Therefore, it is unlikely that the mutational enrichment at 5’CpG3’ sites is attributable to methylated cytosines.

The mutation rate in the AT direction \( m_{G/C} = C \Rightarrow A = T \) (including G:C→A:T transitions and G:C→T:A transversions) is 1.56 × 10⁻¹⁰ (95% Poisson confidence interval: 1.40–1.72 × 10⁻¹⁰), while the mutation rate in the GC direction \( m_{A/T} = G \Rightarrow C \) (A:T→G:C transitions and A:T→C:G transversions) is 1.69 × 10⁻¹⁰ (1.49–1.92 × 10⁻¹⁰). Thus, the mutation rate in this GC-rich organism is slightly biased to G/C, with the equilibrium GC content expected from mutation pressure alone being 52.13% (SE: 2.23%), lower than the 62% genome GC content, implying that evolutionary forces such as selection are operating to increase the genome GC content in this GC-rich organism, at least at some sites.

### Table 2

Transition to transversion ratios (ts/tv) of bacteria and unicellular-eukaryote MA lines with whole genomes sequenced

<table>
<thead>
<tr>
<th>Organisms</th>
<th>ts/tv</th>
<th>GC content</th>
<th>AT→G:C transition rate</th>
<th>GC→AT transition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unicellular eukaryotes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>0.76</td>
<td>0.64</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Rhodotorula toruloides</td>
<td>1.83</td>
<td>0.62</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>0.95</td>
<td>0.38</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>0.75*</td>
<td>0.36</td>
<td>0.48*</td>
<td>Behringer and Hall (2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus subtilis</td>
<td>2.46</td>
<td>0.43</td>
<td>0.76</td>
<td>Sung et al. (2015)</td>
</tr>
<tr>
<td>Burkholderia cenocepacia</td>
<td>1.21</td>
<td>0.67</td>
<td>0.86</td>
<td>Dillon et al. (2015)</td>
</tr>
<tr>
<td>Deinococcus radiodurans</td>
<td>1.71</td>
<td>0.67</td>
<td>1.14</td>
<td>Long et al. (2015a)</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>1.28</td>
<td>0.51</td>
<td>0.62</td>
<td>Lee et al. (2012)</td>
</tr>
<tr>
<td>Mesoplasma florum</td>
<td>1.26</td>
<td>0.27</td>
<td>0.08</td>
<td>Sung et al. (2012)</td>
</tr>
<tr>
<td>Mycobacterium smegmatis</td>
<td>1.48</td>
<td>0.66</td>
<td>1.43</td>
<td>Kucukyildirim et al. (2016)</td>
</tr>
</tbody>
</table>

*Mutations from strains CC2344 and CC2931 are not included due to presence of MA lines with affected DNA repair.

**Average value of Behringer and Hall (2015) and Farlow et al. (2015).**
We asked if the relatively high A:T → G:C transition mutation rate could be related to flanking nucleotide contexts. About 98.28% of the A/T sites with A:T → G:C transitions are flanked by at least one G or C nucleotide, and this proportion is significantly different from the proportion of A/T sites with at least one G or C nucleotide in the R. toruloides NP11 genome (92.54%) (χ² test, P = 0.02). Flanking G:C pair(s) may elevate the mutation rate by orders of magnitude due to their stable anchoring and base-stacking power (Yakovchuk et al. 2006; Lee et al. 2012; Long et al. 2015b; Ness et al. 2015; Sung et al. 2015). Indeed, the transition mutation rate for A/T sites flanked by G or C is ~3-fold higher than that without such flanking (F-test for equal variance—F = 1.06, P = 0.83; two-sample Student t test P = 0.04; fig. 2). Therefore, the nucleotide context is a significant factor accounting for the unusually high A:T → G:C transition mutation rate in this GC-rich yeast.

In contrast to the elevated A:T → G:C transition mutation rate found in R. toruloides, the G:C → T:A transversion mutation rate of 3.22 × 10⁻¹¹ per site per cell division is much lower than that in the ascomycetous yeasts (~1/5 of the S. pombe and ~1/2 of the S. cerevisiae rates; fig. 2). This may be explained by the extremely high concentration of cellular triglycerides (70% of cellular dry weight) and carotenoids in R. toruloides. A large fraction of G:C → T:A transversions are thought to result from DNA oxidative damage (Grollman and Moriya 1993; Foster et al. 2015), and triglycerides and carotenoids are known to reduce oxidative mutation (Kobayashi et al. 2008; Rock 2009), for example, by scavenging reactive oxygen species.

Insertions and Deletions (Indels)

We detected 50 small deletions (total size of 267 bp; < 29 bp) and 33 small insertions (317 bp; < 31 bp) (supplementary tables S1 and S3, Supplementary Material online). These indels yield an indel mutation rate of 2.21 × 10⁻¹² per site per cell division (95% Poisson confidence intervals: 1.76–2.74 × 10⁻¹¹), or 11.51% of the base-substitution mutation rate. This small-indel rate is ~4.4 times higher than that of S. cerevisiae (Zhu et al. 2014), but the proportion of indels occurring next to simple sequence repeats is much lower than that of the budding yeast (25.30% vs. 61.54%). Such difference may reflect the high divergence in the genome content or the DNA replication and repair machineries of these two yeasts from different phyla.

Across all 56 MA lines, we detected 39 large deletions (> 50 bp) for a total loss of 40,183 bp (supplementary table S4, Supplementary Material online). The ends of chromosomes are particularly susceptible to deletion, most likely due to recombination between subtelomeric regions (McEachern 2008). This is particularly true for the chromosome represented by the scaffold KB722717.1, which experienced a deletion in 7 out of the 56 MA lines, with the nucleotide sequence affected being flanked by CCCTAA repeats in all the seven cases, characteristic of telomeres (Meyne...
et al. 1989; Guzman and Sanchez 1994; Underwood et al. 1996). The second most common hotspot for deletions is located on the chromosome represented by the scaffold KB722690.1, where 4 out of the 56 MA lines are affected. Again, the sequencing reads mapping around these deletions contain CCCCTAA repeats.

We additionally attempted to identify novel mobile-element insertions, but none were detected. This may be caused by the incomplete status of the genome assembly and/or annotation as mobile elements are difficult to assemble, but another possibility is that fungi, and yeast species in particular, are extremely adept at inactivating mobile-genetic elements through various pathways (Hood et al. 2005; Hu et al. 2013). The latter possibility is further supported when comparing the genome size of R. toruloides (~20 Mb) to that of other members of subphylum Pucciniomycotina such as Puccinia graminis (~89 Mb) and Melampsora larici-populina (~101 Mb), which have high numbers of mobile elements, while Sporobolomyces roseus (~10Mb), also a ‘red yeast’, has extremely low mobile-element content (Muszewska et al. 2011, 2013).

**Discussion**

Despite being separated by 800 million years of evolution, the total base-substitution mutation rates of the three yeasts are ~ 2 × 10^{-10} per nucleotide site per cell division. This is particularly striking in light of their dramatic biological differences: such as gene structure (e.g., only 5% of S. cerevisiae genes have introns, whereas most genes have introns in S. pombe and R. toruloides), genome size, chromosome number, GC content, etc. (Spingola et al. 1999; Cherry et al. 2012; Wood et al. 2002; Zhu et al. 2012). This shared mutation rate implies limited variation of effective population sizes in yeasts, under the drift-barrier hypothesis (Sung et al. 2012; Farlow et al. 2015), where selection rates of DNA polymerases at these sites or different pre-replication repair biases may be involved. Given that oxidative damage is a very strong endogenous mutation determinant (Foster et al. 2015), and antioxidant triglycerides and carotenoids are extremely abundant in R. toruloides, the much lower G:C→T:A transversion mutation rate suggests that this oleaginous yeast may provide benefits as an antioxidant nutritional oil candidate. Based on this observation, we predict that organisms with high cellular antioxidants will generally have a lower G:C→T:A transversion mutation rate relative to organisms with low antioxidants or without. However, these differences may also be due to undetermined fundamental differences between ascomycete and basidiomycete yeasts. By investigating more fungal species, especially filamentous ones, it may be possible to determine the mechanisms underlying the variation in the fungal mutation spectrum.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Acknowledgments**

This research is funded by a Multidisciplinary University Research Initiative award (W911NF-09-1-0444) from the US Army Research Office and a NIH R01 grant (R01GM036827). We thank Thomas G. Doak and Samuel F. Miller for technical assistance and comments on the manuscript.

**Literature Cited**


Behringer MG, Hall DW. 2015. Genome wide estimates of mutation rates and spectrum in Schizosaccharomyces pombe indicate CpG sites are highly mutagenic despite the absence of DNA methylation. G3 (Bethesda) 6:149–160.


Similar Mutation Rates but Highly Diverse Mutation Spectra


Assocate editor: Charles Baer