Note

Genomewide Evolutionary Rates in Laboratory and Wild Yeast

James Ronald,* Hua Tang[†] and Rachel B. Brem[‡,1]

Department of Genome Sciences, University of Washington, Seattle, Washington 98195, [†]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109 and [‡]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720

ABSTRACT

As wild organisms adapt to the laboratory environment, they become less relevant as biological models. It has been suggested that a commonly used *S. cerevisiae* strain has rapidly accumulated mutations in the lab. We report a low-to-intermediate rate of protein evolution in this strain relative to wild isolates.

WHEN introduced into the lab, wild organisms often undergo selection for easier growth. This adaptation, and loss of selective pressures normally present in the wild, may have wide-ranging effects, such that the biology of a lab organism may no longer reflect that of wild populations. This concern has arisen in the recent literature for *Saccharomyces cerevisiae* (Liu *et al.* 1996; Bonhivers *et al.* 1998; Yvert *et al.* 2003; Deutschbauer and Davis 2005; Dunn *et al.* 2005; Gu *et al.* 2005; Qin and Lu 2006). The S288C yeast strain was bred in the 1950s from wild and commercial strains (Mortimer and Johnston 1986) and passed into use as a common lab strain. In a recent comparison of S288C with the clinical strain YJM789 (Gu *et al.* 2005), isolated in 1989 (Tawfik *et al.* 1989), the phylogenetic lineage to the lab strain exhibited faster protein evolution. One interpretation of this result is that S288C accumulated more mutations during its longer tenure in the lab. Here we revisit this hypothesis in the context of a third strain, the vineyard isolate RM11-1a, which was introduced into the lab in 1996 (Torok *et al.* 1996; Brem *et al.* 2002).
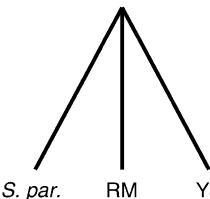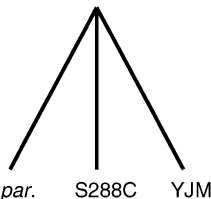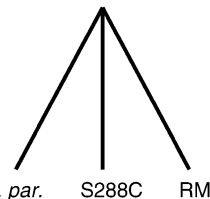
To parallel previous calculations (Gu *et al.* 2005), we first analyzed protein evolution in yeast strain pairs. We obtained ORF alignments of S288C, RM11-1a (hereafter RM), YJM789 (hereafter YJM), and *S. paradoxus* orthologs (Ronald *et al.* 2005) and eliminated frameshifts, for a total of 4162 genes. We reasoned that most assumptions of molecular evolution methods would be as valid here as when species are compared, and the approximation of constant generation time may be better in this case. As such, for each pair of *S. cerevisiae* strains

plus *S. paradoxus*, we used PAML (Yang 1997) to infer maximum-likelihood branch lengths for the star tree describing each gene, assuming an independent evolutionary rate for each lineage. We then used nonsynonymous and synonymous changes in inferred trees to estimate genomewide evolutionary rates for each lineage as described (Chimpanzee Sequencing and Analysis Consortium 2005; Gu *et al.* 2005). The results are shown in Table 1. As expected, in a comparison of S288C and YJM, the lineage to the lab strain had a faster evolutionary rate. However, in other comparisons, the rate for the RM lineage was faster still (Table 1). Thus, the vineyard strain bears the strongest signature of rapid protein evolution.

To confirm this, we sought to analyze the three *S. cerevisiae* strains and *S. paradoxus* simultaneously. As the genealogy of these genomes varies between loci (Ruderfer *et al.* 2006), we modeled each gene separately. For each gene, we inferred branch lengths, assuming independent evolutionary rates on all branches, fixing in turn each of the first three topologies in Figure 1. We then identified the maximum-likelihood tree from among the three inferred trees. If this best tree had an internal branch of length zero, we considered the gene to follow the star topology. When multiple runs of phylogenetic inference did not converge to the same topology, we discarded the gene. The remaining data set comprised 3682 genes. Their maximum-likelihood topologies, shown in Figure 1, indicate that for the majority of genes, RM diverged after the divergence of YJM and S288C. Consistent with this, the estimated time since the internal branch is longer at loci where YJM and S288C are the most closely related strain pair, relative to the rest of the genes (data not shown). We conclude that longer waiting times were required for

[1]*Corresponding author:* Department of Molecular and Cell Biology, 237 Hildebrand Hall 3206, Berkeley, CA 94720-3206.
E-mail: rbrem@berkeley.edu

TABLE 1

Lineage-specific evolutionary rates from analysis of *S. cerevisiae* strain pairs

	S. par. RM YJM	*S. par.* S288C YJM	*S. par.* S288C RM
A.			
S288C	—	0.218 (0.209–0.227)	0.209 (0.200–0.219)
RM	0.262 (0.250–0.274)	—	0.259 (0.249–0.270)
YJM	0.178 (0.169–0.186)	0.188 (0.180–0.196)	—
S. paradoxus	0.126 (0.124–0.129)	0.126 (0.124–0.129)	0.126 (0.124–0.129)
B.			
S. cerevisiae difference	$<10^{-10}$	2.02×10^{-7}	$<10^{-10}$

Each column represents one strain pair, illustrated by the trees above; branch lengths of the trees are not to scale. (A) Each row represents one branch, designated at the left by the genome at its tip. The relative rate of nonsynonymous-to-synonymous mutations along the indicated branch is given, calculated from the sum of each class of changes, and the number of each class of sites, across all inferred gene trees as described (CHIMPANZEE SEQUENCING AND ANALYSIS CONSORTIUM 2005; GU *et al.* 2005). Confidence intervals of 95% from 1000 gene resamplings with replacement are given in parentheses. (B) The χ^2 significance of the difference in the number of nonsynonymous and synonymous changes in the two *S. cerevisiae* lineages indicated in the tree is given.

migration and coalescent events between the lab and clinical strain lineages.

We next grouped all genes with the same topology in Figure 1 and, for each group, estimated evolutionary rates as above. The results are given in Table 2. In each case the lineage to RM, rather than to the lab strain, has the fastest evolutionary rate of all branches. Interestingly, accelerated protein evolution does not seem unique to RM. At loci where the lab or clinical strain shares recent ancestry with RM, both lineages below the internal branch are enriched for nonsynonymous changes, in contrast to loci where S288C and YJM are most closely related to each other (Table 2A). Certain complex demographic scenarios are consistent with this pattern. However, we favor the hypothesis that variation in the ages of branches drives apparent differences in evolutionary rates. Assuming that some existing nonsynonymous changes are mildly deleterious, these alleles are expected to be in excess on recent branches relative to ancient ones, because purifying selection has had a longer time to act on the latter (WILLIAMSON and ORIVE 2002). As the evolutionary time since the internal branch is longest for loci where S288C and YJM are most

closely related (see above), branches below the internal node represent a longer period of purifying selection and bear fewer coding changes; at loci where such branches involve the more recently diverged RM, they bear more coding changes. Recent changes in effective population size and adaptation to viticulture niches (MORTIMER 2000; TOWNSEND *et al.* 2003; AA *et al.* 2006), and to laboratory and pathogenic niches, likely account for additional variation in protein evolutionary rates.

We have shown that the lab strain S288C evolved at a slow-to-intermediate rate relative to two natural isolates, suggesting that growth in the lab has not engendered deleterious mutation on a wide scale. Other reports indicate that lab strains are not exceptionally diverged from other subpopulations (DE BARROS LOPES *et al.* 1999; WINZELER *et al.* 2003; BEN-ARI *et al.* 2005; FAY and BENAVIDES 2005; AA *et al.* 2006), although certain S288C alleles were almost certainly selected in the lab (LIU *et al.* 1996; BONHIVERS *et al.* 1998). By contrast, the vineyard isolate RM has the fastest rate of protein evolution genomewide, and nonsynonymous changes are enriched in other strains where these strains share recent ancestry with RM. Future work will be needed to test our
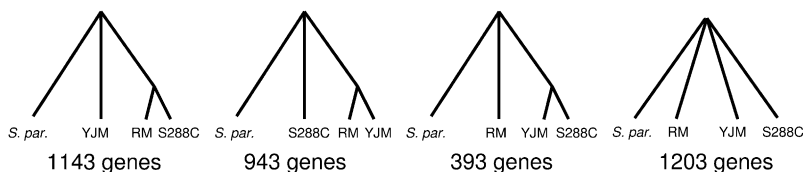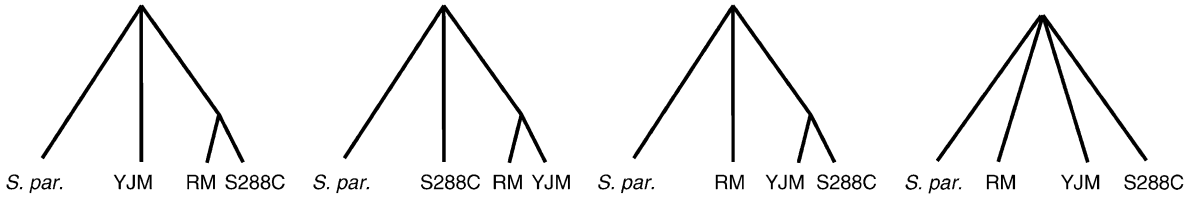

FIGURE 1.—Distribution of gene genealogies. Every tree represents a genealogy relating one *S. paradoxus* and three *S. cerevisiae* strains. Text under each tree represents the number of genes whose maximum-likelihood tree follows the indicated topology. Branch lengths are not to scale.

TABLE 2

Lineage-specific evolutionary rates from analysis of three *S. cerevisiae* strains

	S. par. / YJM / RM / S288C	S. par. / S288C / RM / YJM	S. par. / RM / YJM / S288C	S. par. / RM / YJM / S288C
A.				
S288C	*0.313 (0.284–0.346)*	0.197 (0.183–0.214)	*0.198 (0.172–0.231)*	0.314 (0.274–0.361)
RM	*0.354 (0.319–0.393)*	*0.298 (0.263–0.334)*	0.286 (0.255–0.326)	0.362 (0.316–0.417)
YJM	0.166 (0.155–0.180)	*0.258 (0.228–0.292)*	*0.199 (0.168–0.237)*	0.257 (0.217–0.301)
S. paradoxus	0.134 (0.129–0.139)	0.126 (0.121–0.132)	0.129 (0.122–0.137)	0.130 (0.119–0.145)
Internal	0.208 (0.194–0.224)	0.187 (0.173–0.201)	0.208 (0.184–0.235)	—
B.				
S288C-RM-YJM	$<10^{-10}$	3.47×10^{-9}	2.04×10^{-5}	0.004
S288C-RM	0.09	3.84×10^{-9}	1.30×10^{-4}	0.20
RM-YJM	$<10^{-10}$	0.10	1.98×10^{-4}	0.001
S288C-YJM	$<10^{-10}$	4.13×10^{-4}	0.99	0.05

Each column represents one tree topology; branch lengths are not to scale. (A) Each row represents one branch, designated at left by the genome at its tip. The evolutionary rate of the indicated branch, with 95% confidence intervals in parentheses, calculated as in Table 1 over all genes whose maximum-likelihood tree follows the indicated topology, is given. Results for the most recently diverged lineages of each branched tree are in italics. (B) The χ^2 significance of the difference in the number of nonsynonymous and synonymous changes among the indicated *S. cerevisiae* lineages is given. Results were similar when we analyzed blocks of contiguous genes with the same branched topology (data not shown), suggesting that intragenic recombination does not influence the analysis.

hypothesis that apparently accelerated protein evolution in recently diverged individuals reflects slow selection against mildly deleterious alleles destined to be lost.

LITERATURE CITED

AA, E., J. P. TOWNSEND, R. I. ADAMS, K. M. NIELSEN and J. W. TAYLOR, 2006 Population structure and gene evolution in Saccharomyces cerevisiae. FEMS Yeast Res. **6:** 702–715.

BEN-ARI, G., D. ZENVIRTH, A. SHERMAN, G. SIMCHEN, U. LAVI et al., 2005 Application of SNPs for assessing biodiversity and phylogeny among yeast strains. Heredity **95:** 493–501.

BONHIVERS, M., J. M. CARBREY, S. J. GOULD and P. AGRE, 1998 Aquaporins in Saccharomyces. Genetic and functional distinctions between laboratory and wild-type strains. J. Biol. Chem. **273:** 27565–27572.

BREM, R. B., G. YVERT, R. CLINTON and L. KRUGLYAK, 2002 Genetic dissection of transcriptional regulation in budding yeast. Science **296:** 752–755.

CHIMPANZEE SEQUENCING AND ANALYSIS CONSORTIUM, 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437:** 69–87.

DE BARROS LOPES, M., S. RAINIERI, P. A. HENSCHKE and P. LANDGRIDGE, 1999 AFLP fingerprinting for analysis of yeast genetic variation. Int. J. Syst. Bacteriol. **49:** 915–924.

DEUTSCHBAUER, A. M., and R. W. DAVIS, 2005 Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat. Genet. **37:** 1333–1340.

DUNN, B., R. P. LEVINE and G. SHERLOCK, 2005 Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC Genomics **6:** 53.

FAY, J. C., and J. A. BENAVIDES, 2005 Evidence for domesticated and wild populations of Saccharomyces cerevisiae. PLoS Genet. **1:** 66–71.

GU, Z., L. DAVID, D. PETROV, T. JONES, R. W. DAVIS et al., 2005 Elevated evolutionary rates in the laboratory strain of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA **102:** 1092–1097.

LIU, H., C. A. STYLES and G. R. FINK, 1996 *Saccharomyces cerevisiae* S288C has a mutation in FLO8, a gene required for filamentous growth. Genetics **144:** 967–978.

MORTIMER, R. K., 2000 Evolution and variation of the yeast (Saccharomyces) genome. Genome Res. **10:** 403–409.

MORTIMER, R. K., and J. R. JOHNSTON, 1986 Genealogy of principal strains of the yeast genetic stock center. Genetics **113:** 35–43.

QIN, H., and M. LU, 2006 Natural variation in replicative and chronological life spans of Saccharomyces cerevisiae. Exp. Gerontol. **41:** 448–456.

RONALD, J., R. B. BREM, J. WHITTLE and L. KRUGLYAK, 2005 Local regulatory variation in Saccharomyces cerevisiae. PLoS Genet **1:** e25.

RUDERFER, D. M., S. C. PRATT, H. S. SEIDEL and L. KRUGLYAK, 2006 Population genomic analysis of outcrossing and recombination in yeast. Nat. Genet. (in press).

TAWFIK, O. W., C. J. PAPASIAN, A. Y. DIXON and L. M. POTTER, 1989 Saccharomyces cerevisiae pneumonia in a patient with acquired immune deficiency syndrome. J. Clin. Microbiol. **27:** 1689–1691.

TOROK, T., R. K. MORTIMER, P. ROMANO, G. SUZZI and M. POLSINELLI, 1996 Quest for wine yeasts—an old story revisited. J. Ind. Microbiol. **17:** 303–313.

TOWNSEND, J. P., D. CAVALIERI and D. L. HARTL, 2003 Population genetic variation in genome-wide gene expression. Mol. Biol. Evol. **20:** 955–963.

Williamson, S., and M. E. Orive, 2002   The genealogy of a sequence subject to purifying selection at multiple sites. Mol. Biol. Evol. **19:** 1376–1384.

Winzeler, E. A., C. I. Castillo-Davis, G. Oshiro, D. Liang, D. R. Richards *et al.*, 2003   Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics **163:** 79–89.

Yang, Z., 1997   PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Yvert, G., R. B. Brem, J. Whittle, J. M. Akey, E. Foss *et al.*, 2003   Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat. Genet. **35:** 57–64.

Communicating editor: G. Gibson