Evidence for Domesticated and Wild Populations of *Saccharomyces cerevisiae*

Justin C. Fay[*], **Joseph A. Benavides**

Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, United States of America

Saccharomyces cerevisiae is predominantly found in association with human activities, particularly the production of alcoholic beverages. S. paradoxus, the closest known relative of S. cerevisiae, is commonly found on exudates and bark of deciduous trees and in associated soils. This has lead to the idea that S. cerevisiae is a domesticated species, specialized for the fermentation of alcoholic beverages, and isolates of S. cerevisiae from other sources simply represent migrants from these fermentations. We have surveyed DNA sequence diversity at five loci in 81 strains of S. cerevisiae that were isolated from a variety of human and natural fermentations as well as sources unrelated to alcoholic beverage production, such as tree exudates and immunocompromised patients. Diversity within vineyard strains and within saké strains is low, consistent with their status as domesticated stocks. The oldest lineages and the majority of variation are found in strains from sources unrelated to wine production. We propose a model whereby two specialized breeds of S. cerevisiae have been created, one for the production of grape wine and one for the production of saké wine. We estimate that these two breeds have remained isolated from one another for thousands of years, consistent with the earliest archeological evidence for winemaking. We conclude that although there are clearly strains of S. cerevisiae specialized for the production of alcoholic beverages, these have been derived from natural populations unassociated with alcoholic beverage production, rather than the opposite.

Introduction

Sensu strictu species of the genus *Saccharomyces*, as their scientific name implies, are yeast specialized for growth on sugar. In comparison to other yeasts, *Saccharomyces* favor aerobic fermentation over respiration in the presence of high concentrations of sugar [1]. Fermentation results in the production of ethanol and a competitive advantage, as these yeasts are tolerant to high concentrations of ethanol [2]. One of these species, *S. cerevisiae,* has served as one of the best model systems for understanding the eukaryotic cell and has served as the dominant species for the production of beer, bread, and wine [3]. However, it is worth noting that strains of *S. bayanus* are sometimes used for wine production and strains of *S. pastorianus,* hybrids between *S. cerevisiae* and *S. bayanus,* are used to brew lagers [4].

Since the discovery of yeast as the cause of fermentation [5], numerous strains of *S. cerevisiae* have been isolated, the majority of which have been found associated with the production of alcoholic beverages [6–9]. In many instances, the strains are clearly specialized for use in the lab [10] and the production of wine [11], beer [12], and bread [13]. This has lead to the common view that *S. cerevisiae* is a domesticated species that has continuously evolved in association with the production of alcoholic beverages [3,6,14]. Under this model, the occasional strains of *S. cerevisiae* found in nature are thought to be migrants from human-associated fermentations.

The first use of *S. cerevisiae* is likely to have been for the production of wine, rather then bread or beer [3,15]. *S. cerevisiae* has been associated with winemaking since 3150 BC, based on extraction of DNA from ancient wine containers [16], and the earliest evidence for winemaking is to 7000 BC from the molecular analysis of pottery jars found in China [17]. The idea that *S. cerevisiae* was first used to produce wine rather than beer or bread is further supported by the fact that the production of wine requires no inoculum of yeast [7]. In addition, strains associated with whisky, ale, and bakeries show amplified fragment length polymorphism (AFLP) profiles similar to various wine strains [18].

To examine the relationship between vineyard and non-vineyard strains of *S. cerevisiae* and to understand their evolutionary origin, we have surveyed DNA sequence variation in 81 strains isolated from geographically and ecologically diverse sources (Table 1). These include 60 strains associated with human fermentations, predominantly from vineyards, and 19 strains not associated with human fermentations, predominantly from immunocompromised patients and tree exudates.

Results/ Discussion

DNA sequence variation was examined in 81 yeast strains at five unlinked loci (see Materials and Methods). A total of 184 polymorphic sites were found. Figure 1 shows all of the variable sites along with a neighbor-joining tree constructed from these sites. There are two immediately striking features of the data. First, there are high levels of linkage disequilibrium between sites found in unlinked genes. This linkage disequilibrium cannot be explained by a lack of recombina-

Abbreviations: bp, base pair

*To whom correspondence should be addressed. E-mail: jfay@genetics.wustl.edu

Synopsis

The budding yeast, *Saccharomyces cerevisiae,* has been used to make bread, beer, and wine for thousands of years. To investigate the evolutionary history of this species, the authors examined DNA sequence variation from a large collection of yeast strains isolated from a variety of sources, including saké wine, grape wine, clinical samples, tree exudates, and fruit. The DNA sequence diversity among these strains shows that both saké and grape wine strains form two distinct groups that have remained isolated for a substantial period of time. The data suggest that *S. cerevisiae* consists of both "wild" and "domesticated" populations and that at least two independent domestication events lead to extant grape wine and saké wine strains.

tion because the four gamete test [19] shows evidence of recombination both within and between loci. The high level of linkage disequilibrium is most likely caused by population subdivision and suggests that the data from these five genes provide a genomic view of population differentiation among these strains. Second, there are significant levels of population differentiation based on the source from which the samples were isolated (see Materials and Methods). A number of strains are worth noting. Y9 is very closely related to the saké strains and was obtained from Indonesian ragi, or yeast cake, which like saké is made by fermenting *koji*, a mixture of rice and the mold *Aspergillus oryzae* [20]. Y3 and Y12 were isolated from African palm wine, made from fermenting sap

Table 1. Strains Studied and Their Source

ID	Strain	Location	Source	Date
B1	Lalvin 71B	France	Vineyard (commercial)	NA
B2	Levuline ALS	NA	Vineyard (commercial)	NA
B3	Zymaflore F15	France	Vineyard (commercial)	NA
B4	Lalvin CY-3079	NA	Vineyard (commercial)	NA
B5	Lalvin BM45	NA	Vineyard (commercial)	NA
B6	Zymalfore VL3	France	Vineyard (commercial)	NA
CDB	Côte des Blancs	Germany	Vineyard (commercial)	NA
I14		Italy	Vineyard (soil)	2002
K1	Kyokai no. 1	Japan	Saké	1906
K5	Kyokai no. 5	Japan	Saké	1925
K9	Kyokai no. 9	Japan	Saké	1950s
K10	Kyokai no. 10	Japan	Saké	1952
K11	Awamori-1	Japan	Saké	1981
K12	AKU-4011	Japan	Saké (Shochu)	NA
K13	NRIC 23	Japan	Saké	NA
K14	NRIC 1413	Japan	Saké	NA
K15	NRIC 1685	Japan	Saké	NA
M1		Italy	Vineyard	1993
M2		Italy	Vineyard	1993
M3		Italy	Vineyard	1993
M4		Italy	Vineyard	1993
M5		Italy	Vineyard	1993
M6		Italy	Vineyard	1993
M7		Italy	Vineyard	1993
M8		Italy	Vineyard	1993
M9		Italy	Vineyard	1993
M11		Italy	Vineyard	1993
M12		Italy	Vineyard	1993
M13		Italy	Vineyard	1993
M15		Italy	Vineyard	1993
M17		Italy	Vineyard	NA
M19		Italy	Vineyard	NA
M20		Italy	Vineyard	NA
M21		Italy	Vineyard	NA
M22		Italy	Vineyard	NA
M24		Italy	Vineyard	NA
M29		Italy	Vineyard	1994
M30		Italy	Vineyard	1994
M31		Italy	Vineyard	1994
M32		Italy	Vineyard	NA
M33		Italy	Vineyard	NA
M34		Italy	Vineyard	NA
PR	Pasteur Red	France	Vineyard (commercial)	NA
S288C		California, United States	Nature (fig)	1937
SB	*S. boulardii*	Indonesia	Nature (lychee fruit)	NA
UC1	UCD 51	France	Vineyard	1948
UC2	UCD 175	Sicily, Italy	Vineyard	1953
UC4	UCD 529	Germany	Vineyard	Pre-1958
UC5	UCD 612	Kurashi, Japan	Saké	Pre-1974

Table 1. Continued

ID	Strain	Location	Source	Date
UC6	UCD 765	Australia	Vineyard	NA
UC7	UCD 781	Switzerland	Vineyard	NA
UC8	UCD 820	South Africa	Vineyard	Pre-1988
UC9	UCD 762	Italy	Vineyard	Pre-1984
UC10	UCD 2120	California, United States	Vineyard	1998
Y1	NRRL y390	NA	Nature (mushroom)	Pre-1940
Y3	NRRL y1438	Africa	Fermentation (palm wine)	Pre-1946
Y4	NRRL y1532	Indonesia	Nature (fruit)	Pre-1947
Y5	NRRL y1546	West Africa	Fermentation (bili wine)	Pre-1947
Y6	NRRL yb1952	French Guiana	NA	Pre-1950
Y8	NRRL y2411	Turkey	Vineyard	Pre-1957
Y9	NRRL y5997	Indonesia	Fermentation (ragi)	Pre-1962
Y10	NRRL y7567	Philippines	Fermentation (coconut)	Pre-1973
Y12	NRRL y12633	Ivory Coast	Fermentation (palm wine)	Pre-1981
YJM145	seg. YJM128	Missouri, United States	Clinical	Pre-1989
YJM269		NA	Fermentation (apple juice)	1953
YJM270		Europe	Vineyard	Pre-1957
YJM280	seg. YJM273	United States	Clinical	Pre-1994
YJM308		United States	Clinical	Pre-1994
YJM320	seg. YJM309	United States	Clinical	Pre-1994
YJM326	seg. YJM310	United States	Clinical	
YJM339	seg. YJM311	United States	Clinical	Pre-1994
YJM421	seg. YJM419	United States	Clinical	Pre-1994
YJM434		Europe	Clinical	
YJM436		Europe	Clinical	Pre-1994
YJM440		United States	Clinical	Pre-1994
YJM454		United States	Clinical	Pre-1994
YJM627	seg. Y55	France	NA	
YJM1129	NRRL y-567	NA	Fermentation (distillery)	Pre-1912
YPS1000		New Jersey, United States	Nature (oak exudate)	2000
YPS1009		New Jersey, United States	Nature (oak exudate)	2000
YPS163		Pennsylvania, United States	Nature (oak exudate)	1999

NA, not available; seg., segregant.
DOI: 10.1371/journal.pgen.0010005.t001

of the oil palm, *Elaeis guineensis*. Y5 was isolated from African bili wine.

If strains of *S. cerevisiae* that are not associated with human fermentations have escaped their manmade environments, their progenitors should be closely related to strains isolated from human fermentations. Two aspects of the data indicate this is not the case. First, the oldest lineages at the root of the tree, that are most similar to *S. paradoxus,* were isolated from tree exudates in North America and Africa, or from immunocompromised patients. Although one of the clinical samples is most closely related to vineyard strains, the majority of clinical isolates are not closely related to strains obtained from human-associated fermentations. Second, strains from grape wine and saké wine production contain significantly less variation, as measured by the average number of pairwise differences between strains [21], than is found in natural and clinical isolates, which contain just as much variation as is found in the total sample (Table 2). However, diversity in strains associated with human fermentations other than grape and saké wine production is not reduced compared to the clinical and natural isolates. The four strains associated with fermentations, three of which were isolated from traditional African wines, show the greatest diversity and represent some of the oldest lineages. This raises the possibility that *S. cerevisiae* was domesticated in Africa and that most vineyard and saké strains were derived from a domesticated African strain. If so, one would expect clinical and natural isolates to be more closely related to strains isolated from vineyards, which have a cosmopolitan distribution compared to strains from traditional African wine. Clinical and natural isolates, however, show no obvious relationship to strains associated with manmade fermentations.

Although the genealogical relationships among strains of *S. cerevisiae* show that the species as a whole is not domesticated, the data do support the hypothesis that some strains are domesticated. Based on the low levels of diversity within vineyard and saké strains and the clear separation of these two groups, we propose two domestication events, one for yeast used to produce grape wine and one for yeast used to produce rice wine. When might these events have occurred? Domestication would have occurred after the divergence between the vineyard and saké strains but before differentiation among the vineyard and among the saké strains. These two time points can be roughly estimated by the average number of differences per synonymous site between the saké and vineyard strains, 1.28×10^{-2}, and the average number of differences among the vineyard, 2.92×10^{-3}, and among the saké strains, 4.06×10^{-3}, respectively (see Materials and Methods). Assuming a point mutation rate of 1.84×10^{-10} per base pair (bp) per generation and 2,920 generations per year, the estimate for the divergence time between the two
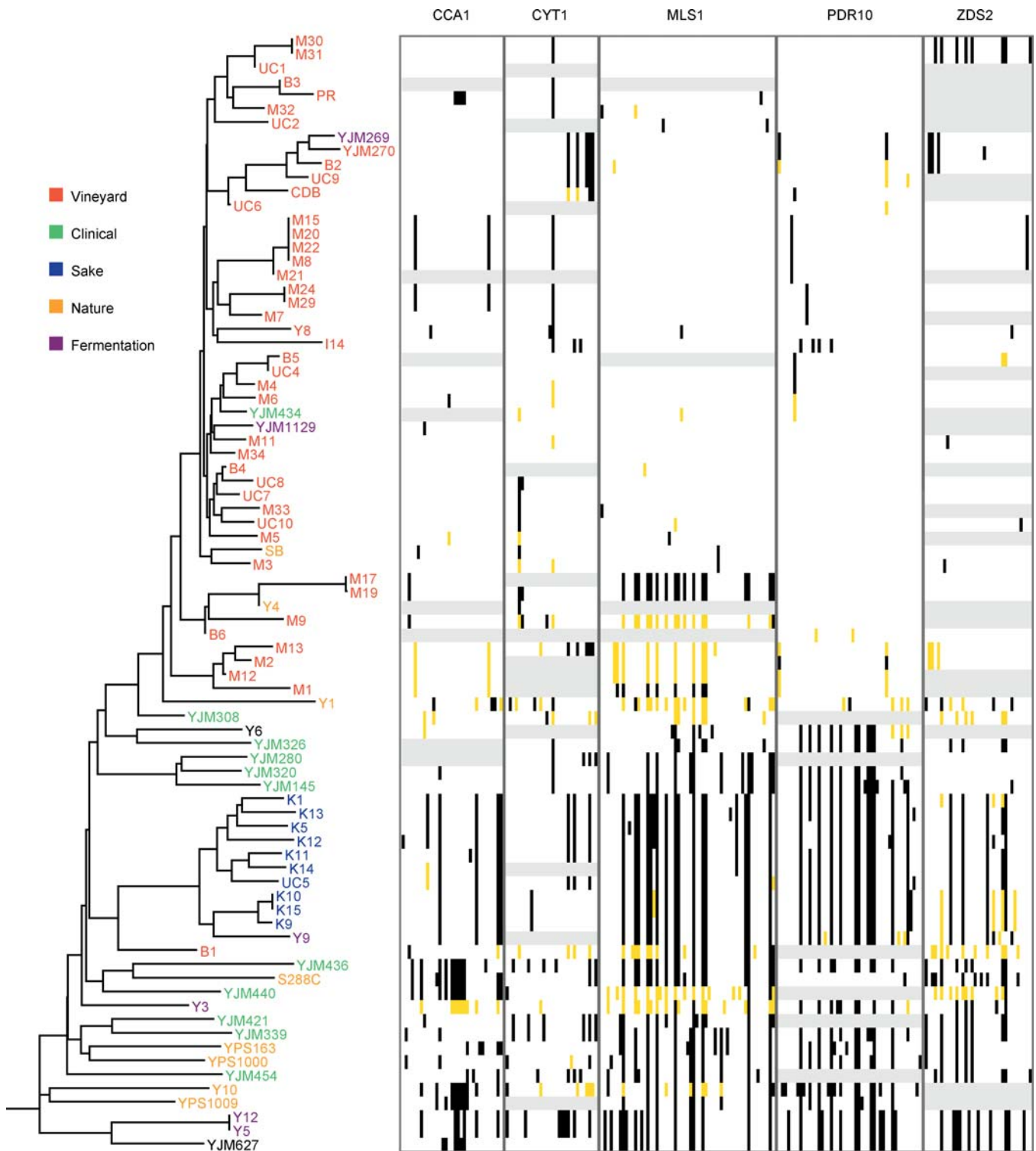
Figure 1. A Neighbor-Joining Tree Shows Differentiation among Yeast Strains Isolated from Different Sources

The tree was constructed from polymorphic sites found at five unlinked loci and was rooted using *S. paradoxus*. Strains are colored according to the substrates from which they were isolated. The right side shows color-coded polymorphism data with minor alleles shown in black, major alleles shown in white, missing data shown in light gray, and heterozygous sites shown in orange.

DOI: 10.1371/journal.pgen.0010005.g001

groups is approximately 11,900 years ago, and within the vineyard group and saké group is approximately 2,700 and approximately 3,800 years ago, respectively (see Materials and Methods). These dates could easily be an order of magnitude older if the number of generations per year is one tenth that obtained assuming an exponential growth rate. Interestingly, the time period is consistent with the earliest archeological evidence for winemaking, approximately 9,000 years ago [17].

Table 2. Diversity among Strains

Source[a]	Strains	$\pi \times 100$[b]
Saké wine	9	0.10 (0.01)
Grape wine	23	0.14 (0.03)
Clinical	4	0.42 (0.10)
Nature	5	0.50 (0.08)
Fermentation	4	0.54 (0.15)
Total	45	0.42 (0.03)

[a]Only strains without missing data are used.
[b]π is the average number of pairwise differences between strains, per basepair. The standard deviation is shown in parentheses.
DOI: 10.1371/journal.pgen.0010005.t002

It should be noted that proof that these strains are domesticated requires evidence that they have acquired characteristics advantageous to humans through human activity, whether intentional or not. The alternative hypothesis to domestication is that initial fermentations selected those natural isolates most amenable to alcoholic beverage production and that these initial isolates have been used by humans ever since.

The source population for both the saké and grape wine strains is not clear, but is likely similar to the source population for the clinical strains. Insects, particularly fruit flies, present one possibility [22,23]. Numerous strains of *S. cerevisiae* and *S. paradoxus* have been isolated from oak tree exudates in North America [24], and tree exudates are often visited by insects [22]. Three of these oak tree isolates were included in our study and are among the most diverse of the strains (Figure 1). Given that *S. paradoxus* is most often found in association with tree exudates from both Europe [25,26] and North America [24], strains of *S. cerevisiae* isolated from tree exudates may be truly "wild" yeast. Whether the yeast isolated from African palm wine is domesticated remains an open question, although it is worth noting that African palm wine is made by collecting sap tapped from oil palm trees and fermentation occurs naturally without the addition of yeast.

Materials and Methods

Strains were obtained from a number of individuals and stock centers. B1–B6 were obtained from B. Dunn; I14 from J. Fay; CDB and PR from Red Star, Berkeley, California, United States; K1–K15 from N. Goto-Yamamoto and the NODAI culture collection; M1–M34 from R. Mortimer; SB from Whole Foods, Berkeley, California, United States; UC1–UC10 from the University of California, Davis stock center; Y1–Y12 from C. Kurtzman and the ARS culture collection; YJM145–YJM1129 from J. McCusker; and YPS163–YPS1009 were from the collection of P. Sniegowski.

Five genes, *CCA1, CYT1, MLS1, PDR10,* and *ZDS2,* and their promoters were sequenced in 81 strains (see Table 1). These genes were randomly chosen from all divergently transcribed intergenic sequences upstream of functionally annotated genes with clear orthologs in *S. paradoxus*. The sequenced regions include 3,671 bp of coding sequence and 3,561 bp of noncoding sequence. For each gene, both strands of purified PCR products were sequenced using Big Dye (Perkin Elmer, Boston, Massachusetts, United States) termination reactions. Sequence variation was identified using phred, phrap, and consed [27]. For construction of the neighbor-joining tree, a single allele was used from strains with heterozygous sites. The allele was randomly chosen from the two haplotypes inferred by PHASE [28].

Sequence data were analyzed using DNASP [29]. Population subdivision was tested by a permutations test according to the source categories from which each strain was obtained (Table 1). The average time since divergence of two strains was obtained by $k = 2\mu t$, where k is the substitution rate, μ is the mutation rate per bp and t is the time in generations. The mutation rate has been estimated at *CAN1* and *SUP3* at 2.25×10^{-10} per base pair per generation [30]. Given that 82% of spontaneous mutations are single base substitutions [31], we estimate the point mutation rate is 1.84×10^{-10} per bp per generation. *S. cerevisiae* can reproduce in 90 min, or 16 generations per day. However, even under optimal laboratory conditions the number of generations over a 24-h period is typically much less. To obtain divergence time in years rather than generations, we assumed *S. cerevisiae* can go through a maximum of eight generations per day or 2,920 generations per year.

Supporting Information

Acknowledgments

References

1. Otterstedt K, Larsson C, Bill RM, Stahlberg A, Boles E, et al. (2004) Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. EMBO Rep 5: 532–537.
2. Sipiczk M, Romano P, Lipani G, Miklos I, Antunovics Z (2001) Analysis of yeasts derived from natural fermentation in a Tokaj winery. Antonie Van Leeuwenhoek 79: 97–105.
3. Mortimer RK (2000) Evolution and variation of the yeast *(Saccharomyces)* genome. Genome Res 10: 403–409.
4. Nguyen HV, Gaillardin C (2005) Evolutionary relationships between the former species *Saccharomyces uvarum* and the hybrids *Saccharomyces bayanus* and *Saccharomyces pastorianus;* reinstatement of *Saccharomyces uvarum* (Beijerinck) as a distinct species. FEMS Yeast Res 5: 471–483.
5. Pasteur L (1866) Études sur le vin. Paris (France): Imprimeurs Imperials. 266 p.
6. Martini A (1993) Origin and domestication of the wine yeast *Saccharomyces cerevisiae*. J Wine Res 4: 165–176.
7. Mortimer R, Polsinelli M (1999) On the origins of wine yeast. Res Microbiol 150: 199–204.
8. Naumova ES, Bulat SA, Mironenko NV, Naumov GI (2003) Differentiation of six sibling species in the *Saccharomyces* sensu stricto complex by multilocus enzyme electrophoresis and UP-PCR analysis. Antonie Van Leeuwenhoek 83: 155–166.
9. Teresa F-EM, Barrio E, Querol A (2003) Analysis of the genetic variability in the species of the *Saccharomyces* sensu stricto complex. Yeast 20: 1213–1226.
10. Mortimer RK, Johnston JR (1986) Genealogy of principal strains of the yeast genetic stock center. Genetics 113: 35–43.
11. Kunkee RE, Bisson LF (1993) Wine-making yeasts. In: Rose AH, Harrison JS, editors. The yeasts. Volume 5, Yeast technology. New York: Academic Press. pp. 69–126
12. Hammond JRM (1993) Brewers yeast. In: Rose AH, Harrison JS, editors. The yeasts. Volume 5, Yeast technology. New York: Academic Press. pp. 7–67
13. Rose AH, Vijayalakshmi G (1993) Baker's yeasts. In: Rose AH, Harrison JS, editors. The yeasts. Volume 5, Yeast technology. New York: Academic Press. pp. 357–397.
14. Naumov GI (1996) Genetic identification of biological species in the *Saccharomyces* sensu stricto complex. J Ind Appl Microbiol 17: 295–302.
15. McGovern PE (2003) Ancient wine: The search for the origins of viniculture. Princeton (New Jersey): Princeton University Press. 335 p.
16. Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsinelli M (2003)

Evidence for *S. cerevisiae* fermentation in ancient wine. J Mol Evol 57 (Suppl 1): S226–S232.

17. McGovern PE, Zhang J, Tang J, Zhang Z, Hall GR, et al. (2004) Fermented beverages of pre- and proto-historic China. Proc Natl Acad Sci U S A 101: 17593–17598.

18. Azumi M, Goto-Yamamoto N (2001) AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces* sensu stricto and its application to phenetic clustering. Yeast 18: 1145–1154.

19. Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111: 147–164.

20. Kodama K (1993) Saké-brewing yeasts. In: Rose AH, Harrison JS, editors. The yeasts. Volume 5, Yeast technology. New York: Academic Press. pp. 129–168

21. Nei M (1987) Molecular evolutionary genetics. New York: Columbia University Press. 512 p.

22. Phaff HJ, Starmer WT (1987) Yeasts associated with plants, insects and soil. In: Rose AH, Harrison JS, editors. The yeasts. Volume 1, Biology of yeasts. New York: Academic Press. pp. 123–180

23. Naumov GI, Naumova ES, Sniegowski PD (1998) *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. Can J Microbiol 44: 1045–1050.

24. Sniegowski PD, Dombrowski PG, Fingerman E (2002) *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEM Yeast Res 1: 299–306.

25. Johnson LJ, Koufopanou V, Goddard MR, Hetherington R, Schafer SM, et al. (2004) Population genetics of the wild yeast *Saccharomyces paradoxus*. Genetics 166: 43–52.

26. Naumov GI, Naumova ES, Sniegowski PD (1997) Differentiation of European and Far East Asian populations of *Saccharomyces paradoxus* by allozyme analysis. Int J Syst Bacteriol 47: 341–344.

27. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8: 186–194.

28. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978–989.

29. Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15: 174–175.

30. Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci U S A 88: 7160–7164.

31. Kang XL, Yadao F, Gietz RD, Kunz BA (1992) Elimination of the yeast RAD6 ubiquitin conjugase enhances base-pair transitions and G.C——T.A transversions as well as transposition of the Ty element: Implications for the control of spontaneous mutation. Genetics 130: 285–294.