



Published in final edited form as:

Nat Commun. ; 5: 4044. doi:10.1038/ncomms5044.

A Gondwanan Imprint on Global Diversity and Domestication of Wine and Cider Yeast *Saccharomyces uvarum*

Pedro Almeida^{1,*}, Carla Gonçalves^{1,*}, Sara Teixeira¹, Diego Libkind², Martin Bontrager³, Isabelle Masneuf-Pomarède⁴, Warren Albertin⁵, Pascal Durrens⁶, David Sherman⁷, Philippe Marullo^{5,8}, Chris Todd Hittinger³, Paula Gonçalves^{1,#}, and José Paulo Sampaio^{1,#}

¹Centro de Recursos Microbiológicos, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

²Laboratorio de Microbiología Aplicada y Biotecnología, Institute de Investigaciones en Biodiversidad y Medio-ambiente, INIBIOMA (CONICET-UNComahue), 8400 Bariloche, Argentina

³Laboratory of Genetics, Genome Center of Wisconsin, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA

⁴Bordeaux Sciences agro, EA Œnologie 4577, ISVV, Villenave d'Ornon F-33882, France

⁵Univ. Bordeaux, EA Œnologie 4577, ISVV, Villenave d'Ornon F-33882, France

⁶CNRS UMR 5800, Univ. Bordeaux, INRIA project-team Magnome, Talence, F33400, France

⁷INRIA project-team Magnome, CNRS UMR 5800, Univ. Bordeaux, Talence, F33400, France

⁸BIOLAFFORT, Bordeaux, F-33072, France

Abstract

In addition to *Saccharomyces cerevisiae*, the cryotolerant yeast species *S. uvarum* is also used for wine and cider fermentation but nothing is known about its natural history. Here we use a population genomics approach to investigate its global phylogeography and domestication fingerprints using a collection of isolates obtained from fermented beverages and from natural environments on five continents. South American isolates contain more genetic diversity than that found in the Northern Hemisphere. Moreover, coalescence analyses suggest that a Patagonian sub-population gave rise to the Holarctic population through a recent bottleneck. Holarctic strains display multiple introgressions from other *Saccharomyces* species, those from *S. eubayanus* being prevalent in European strains associated with human-driven fermentations. These introgressions

#Corresponding authors: Paula Gonçalves, pmz@fct.unl.pt; José Paulo Sampaio, jss@fct.unl.pt.

*these authors contributed equally to this work

Author contributions: CG, CTH, DJS, DL, IMP, JPS, PA, PG, PM conceived and designed the research; PA contributed new analytic tools; CG, ST performed the phenotypic tests and crosses, CTH, DL, JPS isolated new yeast strains; CG, DL, MB, PA, PD, ST, WA performed DNA extractions, genomic library preparations and additional experiments; CG, CTH, DL, DS, IMP, JPS, PA, PD, PG, PM, WA analyzed the data; CG, CTH, PA, PG, JPS wrote the paper. All authors discussed the results and implications and commented on the manuscript at all stages.

Accession codes: The sequencing data generated in this study have been deposited in the National Center for Biotechnology Information short-read archive (NCBI-SRA) under accession codes PRJNA230139 and PRJEB5133.

Competing interests: The authors declare no competing financial interests. PM is employed by BIOLAFFORT. This does not alter the authors' adherence to all the journal policies on sharing data and materials.

are absent in the large majority of wild strains and gene ontology analyses indicate that several gene categories relevant for wine fermentation are overrepresented. Such findings constitute a first indication of domestication in *S. uvarum*.

Introduction

The budding yeast *Saccharomyces cerevisiae* has been used for millennia by humans for the fermentation of foods and beverages^{1,2}. This association is the foundation of its use as a model organism in various areas of biological research, including molecular genetics, biomedicine, biotechnology, and, more recently, evolutionary genomics³. For decades, the natural biology of *S. cerevisiae* remained elusive, but a better understanding of its natural habitats⁴ and distribution⁵, combined with the widespread use of whole genome data⁶⁻⁹, are providing a new ecological framework for the understanding of the diverse cellular, molecular, and evolutionary mechanisms that shaped the world's most biotechnologically important microbe. As a consequence, renewed attention is being paid to the importance of fieldwork for capturing unknown genetic diversity and for gathering quantitative ecological data^{5,10}.

Natural *Saccharomyces* species distinct from *S. cerevisiae* are not frequently associated with human-driven fermentations, but several are often found as contributors to interspecies hybrids used in beer and wine production. In fact, the most economically important industrial yeast after *S. cerevisiae* is the polyploid hybrid *S. pastorianus*, which is used in brewing lager beer. *S. pastorianus* is the result of domestication associated with brewing at low temperatures and combines the genome of a *S. cerevisiae* ale yeast with that of the recently described cryotolerant yeast *S. eubayanus*^{11,12}. Other hybrids harboring genomic contributions from *S. cerevisiae* and a distinct cryotolerant species, *S. kudriavzevii*, are commonly found among strains used to produce Belgian-style beers and wines fermented at low temperatures¹³⁻¹⁵. It now seems clear that, in low-temperature fermentations, *S. cerevisiae* generally tends to be replaced by hybrids that combine the genomes of *S. cerevisiae* with those of *S. eubayanus*, *S. kudriavzevii*, and a third cryotolerant species, *S. uvarum*. *S. uvarum* occupies a basal position in the *Saccharomyces* phylogeny, sister to *S. eubayanus*, making the two species the most distantly related to *S. cerevisiae* within the genus. Unlike *S. eubayanus*, *S. uvarum* does not contribute to the genome of lager-brewing strains but rather stands out as the only natural, non-hybrid species other than *S. cerevisiae* that is relevant for industrial fermentations.

S. uvarum has a fermentation profile in grape must that is different from that of *S. cerevisiae*, producing less acetic acid and ethanol but more glycerol and succinic acid, while synthesizing malic acid without posterior degradation¹⁶⁻¹⁸. Moreover, *S. uvarum* produces volatile fermentative compounds like phenylethanol and its acetate¹⁹ or volatile thiols²⁰. These phenotypic differences between *S. uvarum* and *S. cerevisiae* are associated with pronounced proteomic differences²¹. The wines typically fermented by *S. uvarum* are Tokaj (Hungary, Slovakia); Amarone (Italy); Txakoli (Spain); and in France, Sauternes and the whites of northern French vineyards in Burgundy, Champagne, Val de Loire, and

Alsace²²⁻²⁷. *S. uvarum* is also frequently responsible for cider fermentation, which is usually conducted at low temperatures^{28,29}.

Little information is available on the distribution of *S. uvarum*, and virtually nothing is known concerning possible domestication events. Apart from the appreciable number of strains that have been isolated in Europe from fermentations, only a few strains from natural environments have been obtained in Europe and North America, mostly in association with oak trees⁴. Interestingly, a radically different situation was encountered in South America. After extensive fieldwork on *Nothofagus* (southern beech), a tree genus that, along with *Quercus* (oaks), belongs to the order Fagales, we discovered a conspicuous population of *S. uvarum* in sympatry with *S. eubayanus* in Patagonia¹². Such findings encouraged us to obtain a more representative strain dataset of *S. uvarum*, especially in Australasia, the only other region of the globe where *Nothofagus* are native. Here, we describe the use of a population genomics approach to perform a comprehensive phylogeographic survey of *S. uvarum* with special focus on the detection, for the first time, of fingerprints of domestication in this species.

Results

Occurrence of *S. uvarum* in the Southern Hemisphere

We previously reported the sympatric occurrence of a wild population of *S. uvarum* and its novel sibling species *S. eubayanus* in Northwestern Patagonia, in association with *Nothofagus* and the sugar-rich fruiting structures (stromata) of its biotrophic fungal parasite, *Cyttaria hariotii*¹². In that study, 47 isolates of *S. uvarum* were recovered from 133 samples obtained from three species of *Nothofagus*, with *N. dombeyi* standing out as the species with the highest frequency of isolation with a global success rate of 35%. A much lower frequency of isolation had been previously obtained for this species in Europe when the *Quercus* habitat was explored⁴. This difference prompted us to study the ecological association of *S. uvarum* with the *Nothofagus-Cyttaria* system, not only in Patagonia but also in Australasia. Therefore, we carried out an expanded survey that included a southward extension of the Patagonian region previously covered, as well as sampling in Australia and New Zealand. In the South American survey, 218 samples that included *N. dombeyi* and five additional species of *Nothofagus* were collected in the Spring and Summer months of 2008, 2009 and 2010, yielding 59 *S. uvarum* isolates. In Australasia, 113 samples were collected in the Spring of 2009, yielding 8 isolates.

To provide insight into the global genetic diversity and phylogeography of *S. uvarum*, a set of strains representing the entire scope of geographic origins and habitats (Figs, 1b, 1c, Supplementary Table 1) was selected for whole genome sequencing using the Illumina technology. In addition to a representative collection of isolates from the *Nothofagus* niche, our strain dataset also included wild isolates collected in North America, Eurasia, and Far East Asia, obtained mostly from the bark of oak trees and associated soil. In addition, European strains isolated from artificial environments, mostly wine and cider, were included. Finally, three confirmed hybrid strains of *S. uvarum* × *S. cerevisiae* (two of them, triple hybrids with contributions also from *S. kudriavzevii*) were also included (Supplementary Table 1). The phylogenetic relationships of 54 strains, based on 129096

high-quality polymorphic sites, are depicted in Fig. 1d (Supplementary Fig. 1 depicts a topologically similar phylogeny obtained on an expanded set of 61 strains using 3113 high-quality SNPs from chromosomes 5, 8, and 12). This broad phylogenetic analysis resolved the various representatives of *S. uvarum* into three main clades. Clade A contained all Holarctic strains and a few from South America, clade B encompassed only strains from South America, and clade C corresponded to the Australasian population and was clearly separated from the other two clades. In fact, this lineage was 4.4% divergent (divergence per site, k , with Jukes-Cantor correction) from the other two main lineages, which is similar to the divergence of the North American population of *S. paradoxus* from both the European and Far Eastern populations³⁰. In spite of their divergence, the Australasian strains are clearly more closely related to *S. uvarum* than to any other *Saccharomyces* species, as depicted in Fig. 1a. The phenotypic profiling of the Australasian isolates showed also some unique features, including delayed or absent growth on maltose and melibiose, maximum growth temperatures three to four degrees lower than those observed for other *S. uvarum* strains (31° - 32° C vs 35° C for the non-Australasian strains), and lower growth rates on melibiose at some temperatures (Supplementary Tables 2 and 3).

The observation that the Australasian strains were both genetically and phenotypically divergent from the remaining lineages, combined with their geographic isolation, led us to hypothesize that the observed divergence could be the consequence of allopatric partitioning. To help ascertain this, we performed crosses between Australasian strains and representatives of the other two main lineages represented in Fig. 1d and determined the viability of the resulting progeny. Interestingly, the four crosses involving an Australasian strain mated with a South American or European strain yielded intermediate levels of spore viability of 27-36% (Supplementary Table 4), whereas the progeny of the cross of two Australasian strains, one from Tasmania and the other from New Zealand, had 95.7% spore viability. Another observation worth noting was the relatively high spore viability (18.8%) produced by a cross between an Australasian *S. uvarum* representative and a South American strain of *S. eubayanus* (Supplementary Table 4). This value contrasts with the reported 7% of spore viability from crosses of sympatric (South American) representatives of *S. uvarum* and *S. eubayanus*¹², which could be a sign of the reinforcement of reproductive isolation that deserves further study.

Population structure and phylogeography in South America

The *S. uvarum* phylogeny (Fig. 1d) revealed that, by far, most of the diversity contained in this species is found in the Southern Hemisphere. In fact, the nucleotide diversity based on pairwise differences ($\pi * 100$) contained in the group of South American isolates was 0.689, approximately half of the nucleotide diversity observed globally for the species (1.248) and much higher than that of the Australasian isolates (0.162) and of the Holarctic group (0.141) (Supplementary Table 5). Therefore, in addition to the extremely divergent Australasian lineage, the isolates from the Southern Hemisphere encompass more genetic diversity than those from the Northern Hemisphere. This high diversity observed for the Patagonian strains can be partly attributed to the presence of two major well-separated lineages, clades A and B. The South American strains of clade A (SA-A) are phylogenetically related to the Northern Hemisphere group and were placed at the basal positions of this branch, whereas

the Holarctic isolates occupied more nested positions (Fig. 1d). Clade B was restricted to South American strains only, and is hereafter designated SA-B. The two South American lineages (SA-A and SA-B) had a genetic distance of 1.0% (Supplementary Table 6) and genetic diversities of 0.415 (SA-A) and 0.333 (SA-B) (Supplementary Table 5).

Population structure inference carried out in STRUCTURE (Fig. 1d) clearly indicated the presence of two distinct genetic clusters in our worldwide sampling, as determined by the high value of the *ad hoc* statistic, ΔK , which measures the rate of change between log probabilities for successive K values ($\Delta K_2 = 3682.6$). One of these clusters was assigned to the Australasian lineage and the other consisted of all remaining isolates (South America and Holarctic region). Because STRUCTURE inference is strongly influenced by the sampled genotypes that are used to run the software and due to the marked divergence of the Australasian population, we excluded this lineage from the dataset and repeated the analysis. Interestingly, this restricted *S. uvarum* dataset was resolved into two new genetic clusters that reflected our phylogenetic inference, one cluster that was cosmopolitan and present in both South America and the Northern Hemisphere, while the other cluster was endemic to South America (SA-B) (Fig. 1d). Increasing values of K suggested only minor possible ancestry contributions to those two clusters and did not uncover any new clean clusters (Supplementary Fig. 2). In addition to the strains clearly belonging to either one of these two populations, there were also two strains in the phylogenetic tree (CRUB 1782, CRUB 1595) that occupied positions that could be explained by a mosaic structure of their genomes, with contributions from different lineages. In fact, these strains showed mixed ancestries in STRUCTURE (Fig. 1d), which is consistent with admixture between the two populations. A sliding window analysis comparing the genome of CRUB 1782 with those of strains representing each of the two clean lineages confirmed that it was nearly identical to one or the other of the two South American populations across its genome (Supplementary Fig. 3). These analyses suggested that both strains were approximately 65% SA-A and 35% SA-B, although the exact regions contributed by each parental population were frequently non-overlapping. The existence of strains with mixed ancestries is in line with the high (97%) spore viability found in crosses between strains representing the two populations (Supplementary Table 4).

Limited genetic diversity of the Holarctic population

When compared with the genetic diversity observed for strains from the Southern Hemisphere, strains from a wide range of environments in the Northern Hemisphere show remarkably low diversity across their genomes. This seems to be consistent with the Southern Hemisphere and the *Nothofagus-Cyttaria* system being native geographical ranges and habitats of *S. uvarum*, respectively. It also suggests the hypothesis of a recent migration of the Patagonian *S. uvarum* population SA-A into the Northern Hemisphere and the consequent habitat shift to oaks and other non-*Nothofagus* trees imposed by the absence of *Nothofagus* north of the equator. Indeed, several lines of evidence support the view that the Northern Hemisphere population might be recently derived from the South American SA-A population. Firstly, the nucleotide diversity is much higher in the SA-A population (0.00415) than in the Holarctic population (0.00141), as shown in Supplementary Table 5. In addition, the Holarctic group has the highest value of Tajima's D statistic, although it is not

significant. The relative excess of intermediate frequency alleles is compatible with a recent population contraction where there has not been sufficient time for many new mutations to accumulate (Supplementary Table 5). Moreover, the analysis of shared polymorphisms involving the South American and Holarctic populations supports the hypothesis that the Northern Hemisphere population is derived from the Patagonian SA-A population. The proportion of shared polymorphisms between the SA-A and Holarctic populations is much higher than that of the comparison involving SA-B and Holarctic populations (Fig. 2). In addition, an almost negligible proportion of fixed differences were found between the SA-A and Holarctic populations, contrary to what was observed between SA-A and SA-B and, most notably, between the Holarctic population and SA-B. If the Northern Hemisphere population has derived from the Patagonian SA-A population, the time to coalescence of all Holarctic lineages in a single common ancestor is expected to be more recent than that of the coalescence of lineages of SA-A. In fact, an analysis of relative coalescence times suggested a more recent common ancestor for the Northern Hemisphere population than for either of the South American populations. Specifically, the Holarctic clade is only about 71% as old as the SA-A clade (Fig. 2), which fits in a model of colonization of the Northern Hemisphere from South America. Therefore, taken together, our analyses support the view suggested by the phylogeny in Fig. 1d that the Northern Patagonian population SA-A gave rise to the Holarctic population.

Introgressions from other *Saccharomyces* species

Since several introgression events were previously reported between *S. paradoxus* and *S. cerevisiae*³⁰⁻³², we screened the genome sequences for introgressions from other *Saccharomyces* species. Although signs of foreign genomic DNA were not found in strains isolated in the Southern Hemisphere, 20 out of 34 Holarctic strains displayed multiple introgressions that could be clearly ascribed to various *Saccharomyces* species, including from *S. kudriavzevii* (14 strains), *S. cerevisiae* (3 strains), and most notably from *S. eubayanus* (20 strains) (Supplementary Data 1 and 2). These introgressions were widespread across the *S. uvarum* genomes, but their number and size were found to differ between strains. The precise locations and distributions of the introgressions are summarized in Figs. 3 and 4. Interestingly, all but one of the 18 strains isolated from anthropic niches were found to have introgressions, either solely from *S. eubayanus* or from up to two additional *Saccharomyces* species (Fig. 1d, Supplementary Data 1).

The most striking and recurrent introgressions are from *S. kudriavzevii* and *S. eubayanus*. The *S. kudriavzevii* introgressions are all partially overlapping and relatively short (encompassing up to seven genes). Sequence comparison readily shows that these fragments of *S. kudriavzevii* DNA were acquired from the European population of the species, in line with the European origin of all but one of the strains harboring the *S. kudriavzevii* introgressions. Curiously, the sole non-European strain carrying a *S. kudriavzevii* introgression was isolated in Japan, where a genetically distinct *S. kudriavzevii* population has been identified³⁵. However, this strain is likely to be phylogenetically related with the European strains since its *S. kudriavzevii* introgression is very similar to those present in European strains, both in gene content and sequence (Supplementary Data 1).

Introgessions from *S. eubayanus* are the most prevalent and extensive. The introgessions are, on average, 99.5 % identical to the genome of the South American strain of *S. eubayanus* CRUB 1568 (FM1318), but they are 99.9% identical to the *S. eubayanus* portion of the genome of *S. pastorianus* strain Weihenstephan 34/70. This divergence between the *S. eubayanus* type strain and the *S. eubayanus* moiety found in *S. pastorianus* has already been reported¹². Since all the *S. uvarum* strains with *S. eubayanus* introgessions are European, it seems reasonable to assume that the introgessions occurred in Europe. Therefore, these results suggest that the *S. eubayanus* introgessions in *S. uvarum* either originate from *S. pastorianus* or are derived from a yet-unidentified Eurasian lineage of *S. eubayanus* that was also the contributor to the genomes of the hybrid species *S. pastorianus* and *S. bayanus*. The number of *S. eubayanus* introgessions per strain varied between seven (strain DBVPG 7787) and one (strains CBS 377, BR 6-2, NCAIM 868, ZP 646 and CID 1). The genes involved in the introgessions are listed in Supplementary Data 2, excluding those present in strain NCAIM 676, which possesses a much more extensive contribution from *S. eubayanus*. This strain has a genomic composition similar to that of *S. bayanus* strains CBS 380 and NBRC 1948 (Supplementary Table 7). NCAIM 676 has slightly more genetic material from *S. uvarum* than from *S. eubayanus*, but the extensive genomic contributions from *S. eubayanus* and limited contributions from *S. cerevisiae* suggest that it should be regarded as a representative of the artificial species *S. bayanus* sensu Libkind et al¹². The *S. uvarum* hybrids CBS 2834, CID 1, and S6U contain substantial contributions across their genomes from *S. cerevisiae* and/or *S. kudriavzevii* (Supplementary Table 7).

We searched for a signal of function-specific introgession that might suggest concerted selection by using Gene Ontology (GO) analysis, the power of which is limited by the unknown but probably considerable proportion of hitchhiking genes. Separate GO analysis of the gene set acquired by each of the strains through introgession failed to yield overrepresented GO categories that were both reasonably specific and previously associated to wine fermentations (Supplementary Data 3). However, when GO analysis was performed using the set of genes introgessed into any of the strains, the categories “GO:0006807: nitrogen compound metabolic process” and “GO:0051171: regulation of nitrogen compound metabolic process” were found to be overrepresented (corrected p values <0.05; Supplementary Data 3). Moreover, the use of a gene pool assembled so that the redundancy resulting from the presence of an introgessed gene in multiple strains was preserved, highlighted the additional categories “GO: 1900071: regulation of sulfite transport”, “GO: 0006808: regulation of nitrogen utilization”, “GO:0019740: nitrogen utilization”, and “GO: 0008238: exopeptidase activity”, among others (Supplementary Data 3). These categories include genes pertinent to wine fermentation, some of which have been previously shown in *S. cerevisiae* wine strains to be involved in HGT events^{36,37} or genomic rearrangements³⁸ associated with anthropic environments. Some other overrepresented GO categories include one or several genes located close to genes relevant for nitrogen or sulfite metabolism and could possibly appear as overrepresented as a result of hitchhiking.

All the introgessions on chromosome 2 harbor *ASPI*, a gene encoding the cytosolic L-asparaginase used to degrade asparagine to be used as nitrogen source. This locus was previously associated with low acetic acid production in *S. cerevisiae* wine strains by QTL mapping³⁹. Curiously, a similar gene induced by nitrogen starvation (*ASP3*) was acquired by

some *S. cerevisiae* strains by horizontal gene transfer³⁷. Also, nine strains contained variably sized introgressions of subtelomeric regions of chromosome 7 that contained *FZFI* (Fig. 3 and Supplementary Data 2). This gene has been shown to encode a fast-evolving transcription factor that regulates several genes, including the gene encoding Ssu1, an efflux pump involved in sulfite resistance⁴⁰. All introgressions in this set also encompass the *ZRT1* gene, which encodes a high affinity zinc transporter that was recently reported to exhibit a sequence signature of balancing selection in *S. cerevisiae*⁴¹. Whereas the 900-bp sequence of the coding region of *FZFI* is identical in all introgressed strains, it exhibits 16 and 126 SNPs when compared with reference strains of *S. eubayanus* and *S. uvarum*, respectively. Interestingly, the introgressed *ZRT1* allele is highly diverged from both reference strains (149 and 148 SNPs, respectively). For both *FZFI* and *ZRT1*, there is no evidence of introgression from other known *Saccharomyces* species. In addition to differences in the coding region, the *FZFI/ZRT1* intergenic region appears to carry a large insertion when compared with *S. eubayanus* (Supplementary Fig. 4). Since only the *S. cerevisiae* copy of the *FZFI* gene can be retrieved from the *S. pastorianus* genome, it is not possible to compare these introgressions to the *S. eubayanus* sub-genome of the hybrid species.

We also searched for the presence of non-*Saccharomyces* fungal DNA using *de-novo* assemblies constructed with the subset of reads that could not be mapped to the *S. uvarum* reference genome. Our searches did not reveal the presence of non-*Saccharomyces* ORFs, except for genes known to be present in three horizontally transferred regions previously identified in *S. cerevisiae* wine yeasts³⁶ and originating from yeasts outside this genus (Supplementary Table 8). Since all the nine strains harboring genes from these regions also had introgressions for other *Saccharomyces* species and all but one were associated with anthropic environments it is possible that non-*Saccharomyces* DNA was transmitted simultaneously with *Saccharomyces* introgressions.

Discussion

In *Saccharomyces*, several cases of genome modifications through hybridization¹¹, introgression³¹⁻³³, genome rearrangements^{38,42}, and HGT^{36,37} have been documented, suggesting that genomes modified in this manner appear to be selected for in many industrial environments. Eloquent examples are the inter-species hybrids *S. pastorianus* (*S. cerevisiae* × *S. eubayanus*) in lager beer and *S. cerevisiae* × *S. uvarum* and *S. cerevisiae* × *S. kudriavzevii* in certain types of wines and of Belgian-style beers¹³⁻¹⁵. HGT events also appear to enhance the attributes of *S. cerevisiae* wine strains^{36,37}. These changes have led to the domestication of *Saccharomyces* lineages as they have become genetically distinct from their wild relatives^{6,43,44} in ways that are beneficial to humans. Although the process shares some similarities to what has been documented for plant crops and livestock⁴⁵, HGT, interspecies hybridization, and introgression provide a broader array of genetic mechanisms that lead to domesticated phenotypes.

Here we present for the first time evidence of domestication in *S. uvarum*. In our study, we detected multiple introgressions, mainly of *S. eubayanus* genomic DNA, into the genomes of European wine or cider strains of *S. uvarum*. These introgressions were also found in *S. uvarum* × *S. cerevisiae* hybrids, which suggests that the *S. eubayanus* contribution is

advantageous even in “enriched” hybrid genomes. We postulate that the anthropic habitats colonized by *S. uvarum* in Europe, may have favored hybridization of *S. uvarum* with *S. eubayanus*, followed by subsequent introgression by backcrossing to *S. uvarum*. We suggest that hybrids or introgressed strains involving these two species may be less fit than the clean parental lineages in their natural South American habitat.

Many *S. cerevisiae* wine strains possess three gene clusters acquired from non-*Saccharomyces* yeasts by HGT³⁶. These regions are 120 kb long and include 34 genes, many of which are involved in key aspects of must fermentation such as the metabolism and transport of sugars and nitrogen. Hence, there seems to be some similarity between our findings in *S. uvarum* and those reported for *S. cerevisiae* because, in both cases, selective pressures in anthropic environments seem to have promoted genome adaptations involving inter-species exchange of genetic material. Nevertheless, the global pattern of domestication in *S. cerevisiae* seems to be different from that observed here in *S. uvarum*. The marked reduction of genetic diversity of *S. cerevisiae* wine domesticates relative to their wild relatives has no parallel in *S. uvarum* because the Holarctic wild and domesticated strains are indistinguishable in this respect.

It is not clear if *S. eubayanus* introgressions occurred repeatedly or if they are the result of a single hybridization event that then spread throughout the European population. The latter possibility seems to be more in line with the apparent low abundance of *S. eubayanus* in Europe. The polymorphic pattern of the introgressions when different strains are compared suggests that their fixation or elimination from the population has not yet stabilized. Judging from the widespread occurrence of introgressed strains across Europe and their almost complete absence in wild populations, we hypothesize that introgressions are rapidly lost in European natural environments.

This and other recent studies^{12,47} concur in documenting an ancestral *Nothofagus* association for the two basal *Saccharomyces* species, *S. eubayanus* and *S. uvarum*. For *S. eubayanus*, two populations have also been reported in Patagonia, with a genetic divergence of 1%⁴⁶, which is similar to the divergence measured between the two Patagonian *S. uvarum* populations. In South America, *S. eubayanus* and *S. uvarum* share the *Nothofagus* niche, but *S. eubayanus* seems to be much more elusive than *S. uvarum* in the Northern hemisphere. Only three isolates of *S. eubayanus* were recently reported in North America⁴⁶, and the evidence presented here for the presence of *S. eubayanus* in Europe remains indirect. For *S. uvarum*, we could expand the association with *Nothofagus* by identifying a highly divergent population in New Zealand and Tasmania, which constitutes a significant addition to the diversity of early diverged *Saccharomyces* lineages found in the Southern Hemisphere. The partial reproductive isolation of the Australasian population is comparable to what is observed for the most divergent populations of *S. paradoxus* (32% spore viability)³⁰ and also fit within the range of results obtained for divergent lineages of *S. cerevisiae* found in China that have 10.2 – 55% spore viabilities⁵. Therefore, until more detailed analyses are carried out, we adopt a conservative view and regard the Australasian population as a considerably divergent *S. uvarum* population, possibly the result of an ongoing process of allopatric speciation.

Although the two South American populations of *S. uvarum* are generally sympatric, their distribution ranges appear to only partially overlap. In fact, the distribution of one of these populations (SA-A) is restricted to the northern part of Patagonia, whereas the other population (SA-B) has a distribution that extends to southern Patagonia, including Tierra del Fuego (Fig. 1c). The evidence for admixture that we obtained and the minor contributions from population SA-B in all studied isolates from population SA-A revealed in the STRUCTURE analysis (Fig. 1d) suggest that the two populations are presently in contact and that there is little or no relevant reproductive isolation between them, which could be experimentally confirmed. The geographic distribution of the two Patagonian populations can be related to the phylogeography of *Nothofagus* whose species tend to have latitudinally disjunct populations as a consequence of long-lasting vicariance events related to past glaciations and to the presence of an ancient paleobasin at mid-latitudes in Patagonia⁴⁷⁻⁴⁹. Our results fit this model of two, historically isolated clades, one in the north and one in the south. Favorable climatic and geologic conditions prevailing since 25,000 years ago seem to have facilitated secondary contact. In mid-latitudes, northward migration of *Nothofagus* appears to have been more efficient than southward migration⁴⁹, which is concordant with what we have observed for *S. uvarum* (Fig. 1c). Phylogenomic and population structure analyses support the view that a restricted subset of population SA-A gave rise to the Holarctic population although vectors and mechanisms of this migration into the Northern Hemisphere remain to be elucidated. The lack of diversity found in the extant residents of the Northern Hemisphere contrasts with the vast geographical area colonized, which ranges from temperate North America to Europe and Asia and suggests that the subsequent dispersal across these northern continents occurred rapidly enough to prevent the accumulation of many local mutations.

Australasia and South America harbour the only ecosystems where *Nothofagus* trees can be found presently and were once united in the mega-continent Gondwana. We propose that the much higher genetic diversity of both *S. uvarum* and *S. eubayanus* found in the Southern Hemisphere, the high frequency of isolation and the *Nothofagus* association suggest that both species are native to the Southern Hemisphere and that their primal niche is the *Nothofagus* system. Based on the basal position of the *S. uvarum/S. eubayanus* lineage, it seems plausible to hypothesize that the last common *Saccharomyces* ancestor thrived in the Southern Hemisphere in association with the *Nothofagus* system. In this scenario, an early derived lineage may have undergone speciation while associated with *Nothofagus*, originating *S. eubayanus* and *S. uvarum*, whereas another lineage may have migrated to the Northern Hemisphere through Southeast Asia, the only region of the globe where southern beech and oak relatives of the Fagaceae overlap⁵⁰. The transition to the oak system could have been associated with the formation of other species, giving rise to endemic Asian taxa like *S. arboricola* and *S. mikatae*, as well as to species with more global Holarctic distributions, such as *S. kudriavzevii*, *S. paradoxus*, and *S. cerevisiae*.

Methods

Yeast isolation, identification, and crosses

The selective protocol used for *Saccharomyces* isolations was based on the selective enrichment in a raffinose-ethanol liquid medium as previously described^{4,12}. Putative *Saccharomyces* isolates were confirmed by the observation of *Saccharomyces*-type ascospore production. Species identifications were based on DNA-sequencing of the ITS and D1/D2 regions of the rDNA. For each pair of strains tested, ascospores were isolated and crossed to obtain intra and interspecies hybrids. Hybridization between the two parental strains was confirmed by sequencing of the *MET2* and/or *FSY1* genes and confirmation of the expected heterozygous sites. For each cross, interspecific spore viability was determined by examining at least 200 ascospores produced by two independent hybrid strains.

Quality filtering and alignment to reference genomes

Genomic Illumina sequencing libraries were prepared for most of the isolates (or their monosporic derivatives) (Supplementary Table 1) and sequenced for 100 cycles (single-end) or 2×100 cycles (paired-end) using the Illumina HiSeq2000 or GAXII systems. Some strains were multiplexed as described previously³⁵.

To make results directly comparable, all sequencing data were treated as single-end reads. Reads were filtered to include only those with a perfect index match. To optimize downstream analysis, quality control for each set of reads was carried out using FASTX toolkit v0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/). In brief, any adapter sequences were clipped from the reads, and low quality bases from the 3'-end of reads were trimmed based on the Illumina Phred scores. Reads were retained only if read length was above 32. Finally, reads were discarded if more than 5% of the positions had Phred scores below 10.

Filtered reads from each isolate were mapped to the reference genome of *Saccharomyces uvarum* CBS7001⁵¹ using SMALT v0.6.4 (Wellcome Trust Sanger Institute, www.sanger.ac.uk/resources/software/smalt/) with default parameters, except that we allowed for a step size of 2 (-k 13 -s 2) and SAM format output. In addition to reporting the best gapped alignment for each read, the default settings of SMALT v0.6.4 only generate uniquely mapped reads. Downstream analyses, such as conversion to BAM format, sorting, indexing, several mapping statistics, and consensus genotype calling were performed using the tools available in the SAM tools package v1.18⁵² with the default parameters. Consensus genotypes in VCF format were then converted to fastq format by restricting the maximum depth to 100, 250, or 500, depending on the median mapping depth, in order to avoid overrepresented regions that could be copy-number variants or artifacts. A fasta file was generated for each alignment where base calls with a consensus Phred quality score below 40 (equivalent to a 99.99% of base call accuracy) were masked to lower case. The final quality of the filtered consensus fasta files was assessed by calculating the proportion of high-quality bases (with Phred quality score above 40, Q40 hereafter) relative to the entire mapped genome. For downstream analysis, all bases with Phred quality score below Q40 were converted to an "N". The resulting fasta files were pooled together to generate multiple sequence alignments for each reference contig.

Reads from the three hybrid strains studied and from *S. bayanus* NCAIM 676 (Supplementary Table 1) were treated in the same way as described above but were mapped to a combined reference that included the genomes of *S. uvarum* CBS 7001, *S. kudriavzevii* IFO 1802⁵¹, and *S. cerevisiae* (UCSC version sacCer3). This process provided a reliable alignment of the mappable portion of reads for each hybrid strain to the *S. uvarum* reference genome that were treated in the same way as for the other isolates.

Phylogeny, population structure, and polymorphism analyses

Prior to analyses, regions with evidence for introgression of *S. eubayanus* (see below) were removed from the alignments. For phylogenetic and population structure analyses, unambiguous single nucleotide polymorphisms (SNPs) were extracted from the contig multiple sequence alignments if the evaluated site was represented only by high-confidence alleles, meaning that for each position information was available for all of the isolates.

Two phylogenetic analyses were carried out for our strain dataset. The first included the complete set of isolates and was constructed based on the concatenated SNP alignment of chromosomes 5, 8 and 12 (3113 SNPs; note that chromosome 10 was mislabeled as chromosome 12, and vice-versa, in the annotation of Scannell et al⁵¹). A more restricted phylogeny was built using the 54 isolates that exhibited more than 70% of Q40 bases. Chromosomal SNPs were then concatenated to generate a whole-genome SNP alignment (129096 SNPs). Both phylogenies were inferred by the Neighbour-Joining method with 1 000 bootstrap replicates using the p-distance to compute evolutionary distances. *S. eubayanus* CRUB 1568 was used as outgroup. Phylogenetic analyses were performed in MEGA5⁵³. For the restricted dataset a maximum likelihood phylogeny was also performed using the rapid bootstrap algorithm in RAxML⁵⁴ with GTRCAT approximation.

Population structure of *S. uvarum* was explored using the model-based Bayesian clustering method implemented in STRUCTURE v2.3.4⁵⁵. For the global sampling of *S. uvarum*, STRUCTURE was run with a subset of 10337 equally spaced SNPs. The divergent Australasian population was then removed to facilitate identification of subtle population structure among the remaining clades of *S. uvarum*. A new subset of 9391 equally spaced SNPs was generated for this dataset. For both data, the value of K was allowed to vary from K=1 to K=8, and the number of Markov chain Monte Carlo (MCMC) iterations was set to an initial burn-in period of 50000 iterations, followed by 30000 iterations of sampling. Allele frequencies were assumed to be correlated among populations, and the ancestry model allowed for admixture. Ten independent simulations were run for each value of K and stability was assessed by monitoring the standard deviation between simulations. The optimal number of K clusters were estimated from the *ad hoc* statistic K^{56} . CLUMPP v1.1.2⁵⁷ was used to compute the similarity coefficient between replicate simulations of STRUCTURE (G') using the Greedy or LargeKGreedy (K=7 and K=8) algorithms with 10000 random configurations. The highest value of H' was always found to be above 0.99 indicating high modal similarity between the replicate Q-matrices.

Nucleotide diversity within populations and divergence between populations were estimated using Variscan v2.0⁵⁸. Only positions with valid alleles in at least 90% of the individuals were used for calculations (defined with the NumNuc parameter together with

CompleteDeletion = 0 and FixNum = 1). The null distributions of Tajima's D statistic were generated in ms⁵⁹ under the standard coalescent neutral model with 10000 independent simulations conditioned on the sample size (as determined by NumNuc in Variscan) and observed estimates of θ_w . Sliding window analyses were performed on non-overlapping windows of 1000 sites, retaining only those windows where at least 500 sites (Q40) were used in calculations.

Statistics of shared polymorphisms, fixed differences, and private alleles were calculated employing sharedPoly program from the analysis package (<http://molpopgen.org/>)⁶⁰ using whole-genome SNP alignments.

Bayesian coalescent analysis

Rooted coalescent phylogenies were estimated using a subset of strains from each identified population. Potential hybrid/mosaic strains were excluded from the original alignments in order to minimize violations of the models' assumptions. High-confidence alignment sites were extracted from the contig alignments whenever a site was represented with an unambiguous high-quality allele (Q40) in all strains. These high-confidence alignments were broken into regions with no 4-gamete test violations using RminCutter.pl v1.05 (https://github.com/RILAB/rmin_cut/)⁶¹ with the following settings: -v -q -f -m -g -n. For each contig, the segment with the highest number of polymorphisms was chosen and used as a different partition in the BEAST v1.7.4 package⁶². This approach produced a total of 51159 sites with 2884 segregating positions. BEAST was run with linked partitions using the coalescent constant size tree prior, the GTR+ Γ +I model of sequence evolution, and a strict molecular clock. Monophyly was assumed for the main clades. Coalescent estimates are given in units of substitution per site. Two independent chains of 30000 MCMC generations each were performed, with sampling at every 1000 generations. Traces were combined in LogCombiner, discarding the first 10% of generations as burn-in, and convergence was assessed by examination of the effective sampling size (ESS) in Tracer v1.5.0. All estimated parameters showed ESS values above 600.

Phylogeny of the genus *Saccharomyces*

We deployed a recent multi-locus sequence approach⁶³ to examine the evolutionary relationships among the species and known populations of the genus *Saccharomyces*, including the new Australasian population of *S. uvarum* described here. Complete coding sequences for the fourteen genes were extracted from the available annotations of six *Saccharomyces* species^{51,64}, (*Saccharomyces* Genome Database). Homologous regions for a representative strain of the Far Eastern (IFO 1804) and North American (YPS 138) populations of *S. paradoxus* were retrieved from the BLAST server available at http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_paradoxus_sgrp/⁶. A local BLAST database was also set to retrieve the same regions from our *de novo* assemblies (see below) of *S. eubayanus* (CRUB 1568) and from one representative strain of the *S. uvarum* Australasian population (ZP 962), and from the reference genome of *Naumovozyima castellii* CBS 4309 (NCBI assembly ASM23734v1), which was used as an outgroup. The fourteen coding sequences were concatenated and then aligned with FSA v1.15.7⁶⁵. Sites with alignment gaps and unknown bases were removed from the alignment prior to phylogenetic

reconstruction. A maximum likelihood phylogeny was estimated in PhyML v3.0⁶⁶ using the GTR+ Γ +I model of sequence evolution, estimating the Γ distribution parameter, with 5 substitution rate categories, and the proportion of invariable sites. Tree topology moves were performed using the best of NNI and SPR searches with 5 initial random trees. Branch support was estimated with 100 non-parametric bootstrap replicates.

Screening for foreign genes in the genomes of *S. uvarum*

We searched for evidence of introgressions from other *Saccharomyces* species by mapping the reads to a combined reference including all the available annotated coding sequences of six *Saccharomyces* species^{51,64}, (*Saccharomyces* Genome Database). Reads were quality-filtered as above and mapped to this combined reference using BWA with default parameters⁶⁷. SAM tools v1.18⁵² was used for the manipulation of the resulting BAM files, following the same approach as described above. Genes showing more than 80% of Q40 bases, which provides a good initial measure for the mappability of reads to the corresponding targets, in more than 100 bases were selected for further analysis. Finally, only genes with orthologs unambiguously annotated in at least four species, including *S. uvarum*, were analysed. Exceptions are mainly due to the fact that some of the *bona fide* transfers of genes that are not annotated in *S. uvarum* genome seemed to be close neighbours to other genes that are annotated in both the *S. uvarum* and donor genomes (Supplementary Data 1). A gene was recognized as having a foreign origin if the number of reads that mapped to the putatively donor genome was higher than the sum of the reads that mapped to the orthologous genes in the other *Saccharomyces* species. Five single-gene transfer events were identified using this process but were excluded from further analysis because they corresponded to short coding sequences (less than 200 bp long) or had a low number of reads mapped.

Introgressions from other fungal non-*Saccharomyces* species were searched by assembling the reads that failed to map to the reference genome of *S. uvarum* into contigs with Velvet v. 1.2.08⁶⁸. Assembled contigs with more than 500 bp length were used as query and “blasted” against the ORF sequences of the three horizontally transferred regions identified by Novo *et al.*³⁶ and to the NCBI nr database using BLASTN (1e-10 e-value cutoff). Only the best blast hit for each query was retained and hits to the rDNA region were excluded from the analysis.

Screening for introgressions from *S. eubayanus*

Saccharomyces eubayanus does not yet have a reference assembly against which reads can be mapped. We generated additional reads from the type strain of *S. eubayanus* (CRUB 1568) in this study, and the corresponding reads were treated with the methods described above for the alignment to the reference genome of *S. uvarum*. Pairwise divergence between the *S. uvarum* and *S. eubayanus* type strains was used as a proxy to search for evidence of DNA segments of *S. eubayanus* in the genomes of *S. uvarum*. Divergence per site, k , (with Jukes-Cantor correction) was calculated in Variscan v2.0⁵⁸ using a non-overlapping sliding window of 1000 sites. Only the windows with at least 500 high-quality sites used in the pairwise comparisons (Q40) were retained for further analyses. For easier visualization and interpretation of these results, each datapoint in the plots (Fig. 2) represents the average divergence of itself plus three windows on each side.

Gene Ontology (GO) analysis

Standard GO term discovery was performed with the Generic GO Term Finder⁶⁹. This procedure, however, does not consider the observation that the same gene can be found within introgressed regions of multiple strains. To account for this redundancy, we first determined the number of times each GO attribute was annotated for the set of introgressed genes in each strain and then summed over all the strains considered. Fisher's exact test was used to test the hypothesis that a specific GO attribute was overrepresented in the pooled introgressed gene set relative to the expected number of genes in the reference gene annotation file (retrieved from YeastMine on December 2013). The Bonferroni procedure was applied to correct for multiple comparisons using the total number of tests performed for each ontology term (process, component or function). Significance was assessed for all data with a corrected p-value cutoff of 0.05.

De novo assemblies and analysis of *FZF1* and *ZRT1* sequences

For a more detailed analysis of the introgression on chromosome 7, which harbors *FZF1* and *ZRT1*, we performed *de novo* genome assemblies of the Illumina single-end or paired-end reads for all strains included in this study, using Velvetv.1.2.08⁶⁸. The expected coverage for each region was calculated, and the contigs with a coverage value below 10× were discarded, while the regions with a coverage value below 15× were masked. To retrieve the gene sequences, we set up a local BLAST database for each genome and searched for the aforementioned genes by a TBLASTX search using the *S. uvarum* *FZF1* and *ZRT1* sequences from strain CBS 7001 (*Saccharomyces* Genome Database) as queries [PORF 7762; PORF 7764]. The gene sequences were then aligned using MEGA v.5.05 (ClustalW)⁵³. For the promoter analysis of each gene, a 1000 bp upstream sequence was retrieved from SGD and the aforementioned procedure was followed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Andrey Yurkov, Benedetta Turchetti, Gabor Peter, Gennadi Naumov, Kyria Boundy-Mills, and Michael Weiss for providing strains used in this study; Katie Hyma for samples; and Douda Bensasson for advice on some of the Perl scripts used. The Argentinean National Park Administration is thanked for issuing sampling permits. The following individuals and organizations are thankfully acknowledged for their help during preparation and execution of field work in Australia and New Zealand: Genevieve Gates and David Ratkowsky, University of Tasmania, AUS; Nigel Fechner, Queensland Herbarium, AUS; Peter Johnston, Landcare Research, NZ; Tony Young, The University of Queensland, AUS; Department of the Environment, Water, Heritage and the Arts, Australian Government; Department of Environment and Resource Management, Toowoomba Office, Queensland, AUS; and Department of Conservation (West Coast Conservancy and Otago Conservancy), NZ. This work was supported by grants PTDC/BIA-EVF/118618/2010 (JPS, PG, ST), PTDC/AGR-ALI/118590/2010 (JPS, PG), PTDC/AGR-ALI/112802/2009 (PG), SFRH/BD/77390/2011 (PA), SFRH/BD/89489/2012 (CG), Pest-OE/BIA/UI/457/2011 (JPS, PG) from Fundação para a Ciência e a Tecnologia, Portugal; UNComahue (Argentina) grant CRUB B171 and ANPCyT PICT2011-1814 (DL); National Science Foundation Grant No. DEB-1253634 and in part DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494) (CTH). Aquitaine Region grants 20030304002FA, 20040305003FA and SAGESSE supported the Genomic and Transcriptomic Facility of Bordeaux and PM and DJS received a grant from Conseil Regional d'Aquitaine (SAGESSE).

References

1. Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsineli M. Evidence for *S. cerevisiae* fermentation in ancient wine. *J Mol Evol.* 2003; 57(Suppl 1):S226–S232. [PubMed: 15008419]
2. McGovern PE, et al. Fermented beverages of pre- and proto-historic China. *Proc Natl Acad Sci USA.* 2004; 101:17593–17598. [PubMed: 15590771]
3. Hittinger CT. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* 2013; 29:309–317. [PubMed: 23395329]
4. Sampaio JP, Goncalves P. Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl Environ Microbiol.* 2008; 74:2144–2152. [PubMed: 18281431]
5. Wang QM, Liu WQ, Liti G, Wang SA, Bai FY. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol Ecol.* 2012; 21:5404–5417. [PubMed: 22913817]
6. Liti G, et al. Population genomics of domestic and wild yeasts. *Nature.* 2009; 458:337–341. [PubMed: 19212322]
7. Cromie GA, et al. Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* Assessed by RAD-seq. G3 (Bethesda). 2013; 3:2163–2171. [PubMed: 24122055]
8. Schacherer J, Shapiro J, Ruderfer DM, Kruglyak L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature.* 2009; 458:342–345. [PubMed: 19212320]
9. Bergström A, et al. A high-definition view of functional genetic variation from natural yeast genomes. *MBE.* 2014
10. Hyma KE, Fay JC. Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. *Mol Ecol.* 2013; 22:2917–2930. [PubMed: 23286354]
11. Dunn B, Sherlock G. Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res.* 2008; 18:1610–1623. [PubMed: 18787083]
12. Libkind D, et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA.* 2011; 108:14539–14544. [PubMed: 21873232]
13. González SS, Barrio E, Gafner J, Querol A. Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res.* 2006; 6:1221–1234. [PubMed: 17156019]
14. Lopandic K, et al. Genetically different wine yeasts isolated from Austrian vine-growing regions influence wine aroma differently and contain putative hybrids between *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. *FEMS Yeast Res.* 2007; 7:953–965. [PubMed: 17484739]
15. González SS, Barrio E, Querol A. Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. *Appl Environ Microbiol.* 2008; 74:2314–2320. [PubMed: 18296532]
16. Kishimoto M, Shinohara T, Soma E, Goto S. Identification and enological characteristics of cryophilic wine yeasts. *J Brew Soc Jpn.* 1993; 88:708–713.
17. Giudici P, Zambonelli C, Passarelli P, Castellari L. Improvement of wine composition with cryotolerant *Saccharomyces* strains. *Am J Enol Vitic.* 1995; 46:143–147.
18. Bertolini L, Zambonelli C, Giudici P, Castellari L. Higher alcohol production by cryotolerant *Saccharomyces* strains. *Am J Enol Vitic.* 1996; 47:343–345.
19. Masneuf-Pomarède I, Bely M, Marullo P, Lonvaud-Funel A, Dubourdieu D. Reassessment of phenotypic traits for *Saccharomyces bayanus* var. *uvarum* wine yeast strains. *Int J Food Microbiol.* 2010; 139:79–86. [PubMed: 20188428]
20. Masneuf I, Murat ML, Naumov GI, Tominaga T, Dubourdieu D. Hybrids *Saccharomyces cerevisiae* X *Saccharomyces bayanus* var. *uvarum* having a high liberating ability of some sulfur varietal aromas of *Vitis vinifera* Sauvignon blanc wines. *J Int Sci Vigne Vin.* 2002; 36:205–212.
21. Blein-Nicolas M, et al. Yeast proteome variations reveal different adaptive responses to grape must fermentation. *Mol Biol Evol.* 2013; 30:1368–1383. [PubMed: 23493259]

22. Torriani S, Zapparoli G, Suzzi G. Genetic and phenotypic diversity of *Saccharomyces* sensu stricto strains isolated from Amarone wine. *Antonie Van Leeuwenhoek*. 1999; 75:207–215. [PubMed: 10427409]
23. Naumov GI, Masneuf I, Naumova ES, Aigle M, Dubourdiou D. Association of *Saccharomyces bayanus* var. *uvarum* with some French wines: genetic analysis of yeast population. *Res Microbiol*. 2000; 151:683–691. [PubMed: 11081582]
24. Sipiczki M, Romano P, Lipani G, Miklos I, Antunovics Z. Analysis of yeasts derived from natural fermentation in a Tokaj winery. *Antonie Van Leeuwenhoek*. 2001; 79:97–105. [PubMed: 11392490]
25. Rementería A, et al. Yeasts associated with spontaneous fermentations of white wines from the “Txakoli de Bizkaia” region (Basque Country, North Spain). *Int J Food Microbiol*. 2003; 86:201–207. [PubMed: 12892935]
26. Demuyter C, Lollier M, Legras JL, Le Jeune C. Predominance of *Saccharomyces uvarum* during spontaneous alcoholic fermentation for three consecutive years in an Alsatian winery. *J Appl Microbiol*. 2004; 97:1140–1148. [PubMed: 15546404]
27. Naumov GI, Naumova ES, Antunovics A, Sipiczki M. *Saccharomyces bayanus* var. *uvarum* in Tokaj wine-making of Slovakia and Hungary. *Appl Microbiol Biotechnol*. 2002; 59:727–730. [PubMed: 12226732]
28. Naumov GI, et al. Genetic identification of *Saccharomyces bayanus* var. *uvarum*, a cider-fermenting yeast. *Int J Food Microbiol*. 2001; 65:163–171. [PubMed: 11393685]
29. Valles BS, Bedriñana RP, Tascón NF, Querol A, Madrera R. Yeast species associated with spontaneous fermentation of cider. *Food Microbiol*. 2007; 24:25–31. [PubMed: 16943091]
30. Liti G, Barton DB, Louis EJ. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics*. 2006; 174:839–850. [PubMed: 16951060]
31. Doniger SW, et al. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet*. 2008; 4:e1000183. [PubMed: 18769710]
32. Muller LA, McCusker JH. A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. 2009; 9:143–52. [PubMed: 19054123]
33. Warringer J, et al. Trait variation in yeast is defined by population history. *PLoS Genet*. 2011; 7(6):e1002111. [PubMed: 21698134]
34. Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res*. 2012; 22:908–924. [PubMed: 22369888]
35. Hittinger CT, et al. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature*. 2010; 464:54–58. [PubMed: 20164837]
36. Novo M, et al. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc Natl Acad Sci USA*. 2009; 106:16333–16338. [PubMed: 19805302]
37. League GP, Slot JC, Rokas A. The *ASP3* locus in *Saccharomyces cerevisiae* originated by horizontal gene transfer from *Wickerhamomyces*. *FEMS Yeast Res*. 2012; 12:859–863. [PubMed: 22776361]
38. Pérez-Ortín JE, Querol A, Puig S, Barrio E. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res*. 2002; 10:1533–1539.
39. Marullo P, et al. Single QTL mapping and nucleotide-level resolution of a physiologic trait in wine *Saccharomyces cerevisiae* strains. *FEMS Yeast Res*. 2007; 7:941–952. [PubMed: 17537182]
40. Engle EK, Fay JC. Divergence of the yeast transcription factor FZF1 affects sulfite resistance. *PLoS Genet*. 2012; 8:e1002763. [PubMed: 22719269]
41. Engle EK, Fay JC. *ZRT1* harbors an excess of nonsynonymous polymorphism and shows evidence of balancing selection in *Saccharomyces cerevisiae*. *G3 (Bethesda)*. 2013; 3:665–673.
42. Zimmer A, et al. QTL Dissection of lag phase in wine fermentation reveals a new translocation responsible for *Saccharomyces cerevisiae* adaptation to sulphite. *Plos one*. 2014; 9(1):e86298. [PubMed: 24489712]

43. Fay JC, Benavides JA. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* 2005; 1:e5.
44. Legras JL, Merdinoglu D, Cornuet J, Karst F. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol Ecol.* 2007; 16:2091–2102. [PubMed: 17498234]
45. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. *Cell.* 2006; 127:1309–1321. [PubMed: 17190597]
46. Peris D, et al. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol Ecol.* 2014; 23:2031–2045. [PubMed: 24612382]
47. Markgraf V, McGlone M, Hope G. Neogene paleoenvironmental and paleoclimatic change in southern temperate ecosystems - a southern perspective. *Trends Ecol Evol.* 1995; 10:143–147. [PubMed: 21236983]
48. Markgraf, V.; Romero, E.; Villagrán, C.; In Veblen, TT.; Hill, RS.; Read, J., editors. *The Ecology and Biogeography of Nothofagus Forests*. Yale University Press; New Haven, USA: 1996.
49. Mathiasen P, Premoli AC. Out in the cold: genetic variation of *Nothofagus pumilio* (Nothofagaceae) provides evidence for latitudinally distinct evolutionary histories in austral South America. *Mol Ecol.* 2010; 19:371–385. [PubMed: 20002584]
50. Hill RS. *Nothofagus*: Evolution from a southern perspective. *Trends Ecol Evol.* 1992; 7:190–194. [PubMed: 21236005]
51. Scannell DR, et al. The awesome power of yeast evolutionary genetics: New genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)*. 2011; 1:11–25. [PubMed: 22384314]
52. Li H, et al. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079. [PubMed: 19505943]
53. Tamura K, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28:2731–2739. [PubMed: 21546353]
54. Stamatakis A. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006; 22:2688–2690. [PubMed: 16928733]
55. Falush D, Stephens M, Pritchard JK. Inference of population structure: Extensions to linked loci and correlated allele frequencies. *Genetics.* 2003; 164:1567–1587. [PubMed: 12930761]
56. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005; 14:2611–2620. [PubMed: 15969739]
57. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007; 23:1801–1806. [PubMed: 17485429]
58. Hutter S, Vilella AJ, Rozas J. Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics.* 2006; 7:409. [PubMed: 16968531]
59. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics.* 2002; 18:337–338. [PubMed: 11847089]
60. Thornton K. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics.* 2003; 19:2325–2327. [PubMed: 14630667]
61. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics.* 1985; 111:147–164. [PubMed: 4029609]
62. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 2012; 29:1969–1973. [PubMed: 22367748]
63. Koufopanou V, Swire J, Lomas S, Burt A. Primers for fourteen protein-coding genes and the deep phylogeny of the true yeasts. *FEMS Yeast Res.* 2013; 13:574–584. [PubMed: 23786589]
64. Liti G, et al. High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC Genomics.* 2013; 31:69. 14.
65. Bradley RK, et al. Fast statistical alignment. *PLoS Comput Biol.* 2009; 5:e1000392. [PubMed: 19478997]
66. Guindon S, Gascuel O. PhyML: A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003; 52:696–704. [PubMed: 14530136]

67. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
68. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18:821–829. [PubMed: 18349386]
69. Boyle EI, et al. GO::TermFinder-open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*. 2004; 20:3710–3715. [PubMed: 15297299]

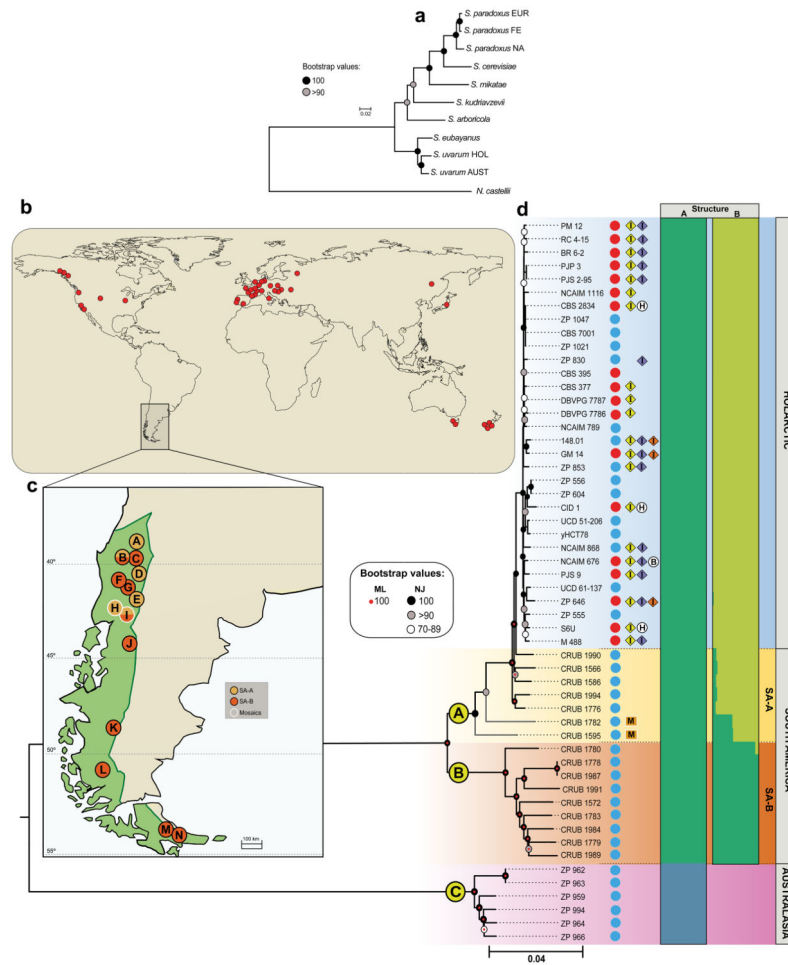


Fig. 1. Geographic distribution, phylogeny, and population structure of *Saccharomyces uvarum*
 A, Maximum Likelihood (ML) phylogeny of the genus *Saccharomyces* based on a concatenated alignment of 14 gene sequences and rooted with *Naumovozyma castellii*. Representatives of populations of *S. paradoxus* (EUR, European; FE, Far Eastern; and NA, North American) and of the Australasian (AUST) and Holarctic (HOL) populations of *S. uvarum* are included. Support values correspond to 100 bootstrap replicates and branch lengths correspond to the mean number of nucleotide substitutions per site. B, World map depicting the geographic origin of Holarctic and Australasian *S. uvarum* strains. C, Map of South America depicting the collecting sites. Locations are indicated by capital letters surrounded by light- or dark-orange colored circles corresponding to populations SA-A and SA-B, respectively. The co-occurrence of strains of the two populations is depicted by light- and dark-orange semi-circles, and the occurrence of mosaic strains is depicted by white contour lines. The distribution of *Nothofagus* is shown in green. D, Whole-genome Neighbor-Joining (NJ) phylogeny of 54 strains based on 129096 SNPs and excluding introgressed regions from *S. eubayanus*, inferred with p-distance and rooted with *S. eubayanus*. The three main clades are marked by letters A, B, and C. Support values from 1000 bootstrap NJ and ML (RAxML rapid bootstrap) trees are included. Incongruent topologies between NJ and ML are denoted by grey lines. Branch lengths correspond to the mean number of base differences per site. Strains isolated from natural environments are

marked by blue circles whereas strains isolated from anthropic environments are marked by red circles. Strains having introgressions are marked with an “I” in a yellow (introgressions from *S. eubayanus*), blue (introgressions from *S. kudriavzevii*), or orange (introgressions from *S. cerevisiae*) diamond. For the hybrid strains (marked with an “H” in a white circle) or *S. bayanus* (“B” in a white circle), only introgressions in the *S. uvarum* sub-genome are depicted. South American mosaic strains are marked with an “M” in an orange square. Clusters inferred with STRUCTURE considering all sequences (A) and all but the Australasian sequences (B) are depicted.

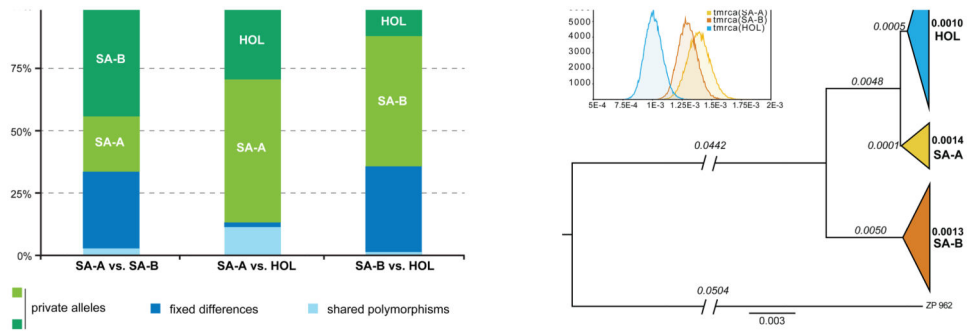


Fig. 2. Proportion of shared and privately segregating polymorphisms and coalescence analysis in the South American and Holarctic populations

A, Proportion of private alleles, fixed differences, and shared polymorphisms among SNPs found in all possible pairwise comparisons involving the Holarctic (HOL) and the two South American populations, SA-A and SA-B (South American mosaic strains were excluded from the analysis). B, Genome-wide estimation of relative time to coalescence for the Holarctic and the two South American populations. The tree was built from an alignment of 51159 high quality sites (Phred quality score >Q40) partitioned over the 16 chromosomes. Each partition represents regions without evidence for intra-locus recombination. The scale bar depicts estimated substitutions per site. Node ages are printed in bold near the nodes of the population whose coalescence they estimate. Branch lengths are printed in italics above the branches. One representative strain of the divergent Australasian population was used to root the tree. The insert shows the marginal posterior densities of the time to the most recent common ancestor (tmrca) in each population.

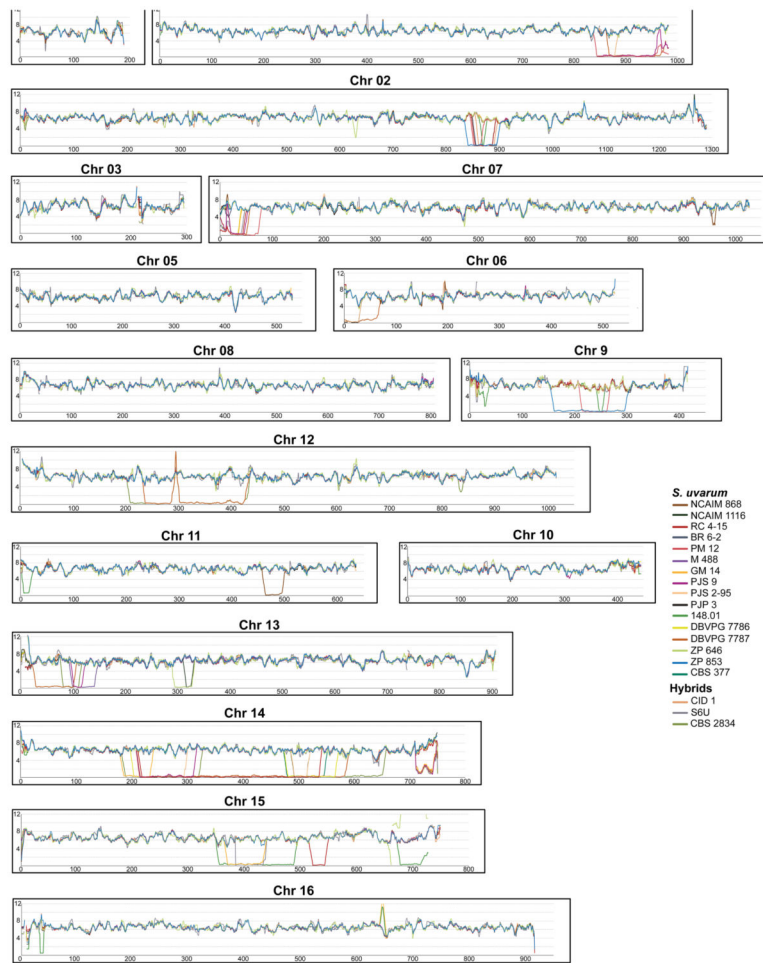


Fig. 3. Sliding window analysis of *Saccharomyces eubayanus* introgressions in *S. uvarum* genomes Each plot represents the divergence k (with Jukes-Cantor correction) relative to *S. eubayanus* reference strain CRUB 1568. Introgressed strains are color-coded according to the key.



Fig. 4. Chromosome maps of *S. uvarum* strains showing the location and extent of introgressions from *S. eubayanus*

The maps are shown for the 16 strains that have *S. eubayanus* introgressions (hybrid strains excluded).