Article

# Title: **Mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species**

Authors: David Peris[1, †], Armando Arias[1, ‡], Sandi Orlić[1, §], Carmela Belloch[2], Laura Pérez-Través[2], Amparo Querol[2], Eladio Barrio[1,2,3]

[1] "Cavanilles" Institute of Biodiversity and Evolution, University of Valencia, Valencia, Spain.

[2] Department of Food Biotechnology. Institute of Agrochemistry and Food Technology (IATA), CSIC, Valencia, Spain.

[3] Department of Genetics, University of Valencia, Valencia, Spain.

Present addresses:

[†] Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA.

[‡] Biotechnology Laboratory, Botany and Zoology Department, CUCBA, University of Guadalajara, Zapopan, Jalisco, Mexico.

[§] Institute Ruđer Bošković, Zagreb, Croatia.

*Corresponding author:

David Peris Navarro

E-mail: david.perisnavarro@wisc.edu

## **Abstract** (250 words)

Horizontal Gene Transfer (HGT) in eukaryotic plastids and mitochondrial genomes is frequently observed, and plays an important role in organism evolution. In yeasts, recent mitochondrial HGT has been suggested between *S. cerevisiae* and *S. paradoxus*. However, few strains have been explored due to the lack of accurate mitochondrial genome annotations. Mitochondrial genome sequences are important to understand how frequent these introgressions occur and their role in cytonuclear incompatibilities. In fact, most of the Bateson-Dobzhansky-Muller genetic incompatibilities described in yeasts are driven by these cytonuclear incompatibilities. In this study, we have explored the mitochondrial inheritance of several worldwide distributed *Saccharomyces* species isolated from different sources and geographic origins. We demonstrated the existence of recombination hotspots in the mitochondrial region *COX2-ORF1*, likely mediated by the transfer of two different types of *ORF1,* encoding a free-standing homing endonuclease, or facilitated by AT tandem repeats and GC clusters. These introgressions were shown to occur both at intra- and interspecific levels. Based on our results we proposed a model which involve several ancestral hybridization events among *Saccharomyces* strains in wild environments.

*Keywords: Saccharomyces,* reticulate evolution, mitochondrial introgression, selfish elements, recombination, interspecies hybridization.

## Introduction

Chloroplast and mitochondrial genomes are prone to introgressions and Horizontal Gene Transfers (HGT) (Keeling 2009; Hao et al. 2010), which likely play an important role in the evolution of eukaryotes (Andersson 2009). Mitochondria are involved in multiple cellular processes (Hatefi 1985; Green and Reed 1998; Starkov 2008). In yeasts, around 750 nuclear encoded proteins must coordinate with those encoded in the mitochondrial genome (Sickmann et al. 2003). Indeed, a new interdisciplinary field, the "mitonuclear ecology", is devoted to the study of the evolutionary consequences due to mitonuclear conflicts (Hill 2015).

In yeasts, recombination in mitochondrial genomes has been mostly focused on *Saccharomyces cerevisiae* (Dujon et al. 1974; Birky et al. 1982; Taylor 1986; MacAlpine et al. 1998), the model yeast species in the genus *Saccharomyces* (Hittinger 2013). The mechanism of mitochondrial genome inheritance is well known in *Saccharomyces* (Ling et al. 2007; Basse 2010; Ling et al. 2011); but despite the recently mitochondrial genome characterization of a hundred *S. cerevisiae,* mostly clinical (Wolters et al. 2015), and few *S. paradoxus*, and the detection of mitochondrial introgression between those two species (Wu et al. 2015; Wu and Hao 2015), little is known about the genomic properties of most *Saccharomyces* species.

In this study, we inferred the mitochondrial inheritance of 517 *Saccharomyces* strains, as well as 49 natural interspecific hybrids, isolated from different sources and geographic origins, by sequencing a mitochondrial gene, *COX2*. This gene has successfully been used either for phylogenetic purposes (Belloch et al. 2000; Kurtzman and Robnett 2003) or for the identification of mtDNA inheritance (Peris et al. 2012a; Peris et al. 2012b; Badotti et al. 2013; Peris et al. 2014; Pérez-Través et al. 2014; Rodríguez et al. 2014). We extended our work to the downstream gene *ORF1*, an unstudied gene encoding a putative free-standing homing endonuclease. A homing endonuclease gene (HEG) is a selfish element described to evolve neutrally, following three steps: i) invasion of an empty site ($HEG^-$) (Colleaux et al. 1986; Burt and Koufopanou 2004), ii) accumulation of mutations generating premature stop codons or invaded by GC clusters which can disrupt their coding frame, iii) and a final loss of the gene (Goddard and Burt 1999). We also sequenced *COX3*, which encodes for cytochrome c oxidase subunit III, to demonstrate

the presence of different mitochondrial recombination hotspots in the *COX2-ORF1* region. We suggest molecular mechanisms that drive such recombinations. In addition, we propose a reticulate event model in the *Saccharomyces* mitochondrial genome as a starting hypothesis for future tests, when new complete *Saccharomyces* mitochondrial genomes will be available.

## Results

### *COX2* shows extensive reticulation among *Saccharomyces* species

To understand the mitochondrial inheritance of worldwide distributed wild and domesticated *Saccharomyces* yeasts (Figure 1), we sequenced and retrieved from public databases, the mitochondrial *COX2* gene sequences, generating a set of 566 *Saccharomyces* strain sequences (Table S1). *COX2* sequence alignment contained 80 phylogenetically informative positions (see details in Supplementary Text). The *COX2* phylogenetic tree failed to reconstruct the species tree (Figure S1), likely indicating the existence of conflicting data in our sequence alignment. A Median-Joining network showed ten haplogroups, supported by sequence inspection (Figure 2 and S2). *S. cerevisiae* were found in three haplogroups: C1a, C1b and C2. Three haplogroups differentiated *S. paradoxus* populations: P1 (Europe), P2 (America B and C) and P3 (Far East). Two haplogroups for *S. mikatae*: M1 (IFO1815) and M2 (IFO1816). *S. kudriavzevii* strains were split into K1 (European and Asia A - IFO1802) and K2 (IFO1803). Finally, one haplogroup for the rest of species: A (*S. arboricola,* CBS10644), E (*S. eubayanus*) and U (*S. uvarum*).

Extensive analysis showed the presence of several recombinant haplotypes among strains from different species (Figure S2). New recombinations were identified in Far Eastern *S. paradoxus* (haplotype 64) and *S. cerevisiae* strains (haplotype 2, 60 and 94) (Figure 2 and S2). We segmented the original *COX2* alignment based on the most common recombination point (Figure S2) and a Neighbor-Net (NN) phylogenetic network was reconstructed for each *COX2* segment (Figure S3). The 5’ end segment NN phylonetwork shows a clear differentiation of haplotypes by species with the exception of haplotypes 69, 70 and 71 belonging to American *S. paradoxus,* which share identical sequences with some *S. cerevisiae* strains (Figure S3A). Haplotypes from natural hybrids

(haplotypes 78, 79, 87, 88, 89 and 93) and *S. eubayanus* (haplotype 102) still appear in an ambiguous position in the network due to the presence of a different recombination point (Figure S2). The extensive presence of incongruent data found in the 3' end segment of *COX2* (Figure S3B) might be also due to recombination.

We were able to define the potential donors of several hybrids (*COX2* haplotypes 87, 88, 89), being *S. kudriavzevii* (European or Asian) and a European *S. paradoxus* (Figure S2). Hybrids *S. cerevisiae* x *S. kudriavzevii isolated* from wine (UvCEG) and a dietary supplement (IF6), and *S. cerevisiae* x *S. uvarum* isolated from wine (S6U) inherited an already recombinant mitochondria from *S. cerevisiae* (Figure S2, Table S1). Hybrids *S. eubayanus* x *S. uvarum* and *S. cerevisiae* x *S. eubayanus* (haplotypes 78, 79, 93) were also recombinant between *S. eubayanus* and *S. uvarum*.

*S. mikatae* IFO1815 and *S. kudriavzevii* IFO1803 haplotypes were also potentially recombinant (Figure S2). In the case of IFO1803, amino acid sequences indicates a recombination between *S. kudriavzevii* and *S. uvarum;* although, the *S. uvarum* haplotype corresponds to an unknown strain. For *S. mikatae* IFO1815, a potential donor might be a *S. cerevisiae* strain from haplogroup C1b, but its amino acid sequence is similar to IFO1816.

**All *S. cerevisiae COX2* haplogroups are worldwide distributed**

The recombination hotspot located in *COX2* makes this gene a good candidate to differentiate closely related strains. Those strains sharing a similar recombination are expected to share an ancestor. To describe the phylogeography of *S. cerevisiae,* we explored the polymorphic *COX2* gene of 418 *S. cerevisiae* from 7 continents, isolated from both human-associated (baking, beer, clinical, laboratory, fermentation, sake, wine and traditional alcoholic beverages) and wild environments (Figure 1, Table S1).

An association study among *S. cerevisiae* strain origins and their haplogroup distribution was performed (Figure S4A). Clinical and wine samples were significantly associated with haplogroup C2 ($\chi^2$ test p-values $9.9 \times 10^{-5}$ and $3 \times 10^{-4}$, respectively). We applied a similar approach to infer the distribution of haplogroups taking into account their geographic origins (Figure S4B), revealing that European *S. cerevisiae* strains were highly associated with Haplogroup C2 ($\chi^2$ test p-value $9.9 \times 10^{-5}$). No significant

distribution bias was observed in the remaining strains, possibly due to the low number of wild isolates in our study.

## Two types of the putative homing endonuclease *ORF1* define two *S. cerevisiae* Haplogroups

To define the extension of the recombination, we sequenced the *ORF1* region from 36 strains representative of the most frequent *COX2* haplotypes, exploring a total of seventy-two *ORF1* sequences (Table S1). The presence of *ORF1* gene was confirmed for the six available *Saccharomyces* species, indicating its broad dissemination across the genus. *ORF1* length was found to be highly polymorphic among strains and species, mostly due to differing content of AT-rich tandem repeats and GC clusters (see Supplementary Text, Figure S5), with lengths ranging from 1.3kbp (ZA17) to 1.5kbp (VRB).

The *ORF1* phylogenetic tree (Figure S5) conflicts with the species tree (Figure S1), indicating a potential HGT from one species to another or the presence of recombinant sequences. To visualize the presence of conflicting data in the alignment, we reconstructed a NN phylogenetic network (Figure 3). Two groups of sequences were visualized in this network: the Type I group contains most *Saccharomyces ORF1* sequences, except for sequences of *S. kudriavzevii*. The Type II group is comprised of *S. cerevisiae* haplotypes from the *COX2* haplogroup C2 and the sequence of the Far Eastern *S. paradoxus* CECT11152 (M51, syn. IFO1804), isolated from Far East Russia, which also appeared as having a recombinant *COX2* (Figure S2 and S3). Haplotypes corresponding to strains of *S. cerevisiae,* some European and Far Eastern *S. paradoxus,* and *S. kudriavzevii*, as well as some hybrids, were located in an ambiguous position in the phylonetwork, between the two main *ORF1* groups (Type I and Type II), suggesting they correspond to recombinant forms.

Recombinant analyses, performed with both `RDP` and `GARD` programs, supported the presence of four partitions. A Kishino-Hasegawa test of phylogenetic congruence indicated that phylogenetic trees from each partition were incongruent with each other. The best partition model was 4 partitions (ΔAICc 143.724), and for each breakpoint the p-value was below 0.01. However, `RDP` and `GARD` disagreed with the location of the

second breakpoint due to the presence of different recombinant points, depending on the strain (Figure S6 and S7B). Some recombination breakpoints were located close to A+T-rich sequences or GC clusters (see Supplementary text, Figure 4). One of the recombination points was just on the beginning of the second LAGLIDADG domain of the encoded homing endonuclease (Figure 4). At least one of the recombinant events involved some of the haplotypes located in the ORF1 phylogeny in-between the two main types (Figure S6 and S7). Further recombinations between Type I x Type I, Type I x Type II and Type II x Type II were detected (Figure S6 and S7).

Data indicates a recent recombination event between *ORF1* from *S. kudriavzevii* and Far Eastern *S. paradoxus*, supported by the high identity among *S. kudriavzevii* and Far Eastern *S. paradoxus* (CECT11422 and CECT11424) *ORF1* amino acid sequences for two segments of the alignment (Figure S6 and S7). For example, the genetic distance for the second region was: 1.1% (*S.par* FE-*S.kud* IFO1802) and 2.4% (*S. par* FE-*S.kud* EU), much lower compared to the nuclear genetic differences, 13.66% and 13.55%, respectively, and when it was compared to the forth *ORF1* segment (Figure S6 and S7): 46.05% (*S. par* FE-*S.kud* IFO1802) and 47.91% (*S.par* FE-*S.kud* EU).

Particular attention must be given to the strains CECT11757 and L1528, representative of the C2xP3 recombinant *COX2* haplogroup. Their *COX2* sequence suggested that their *ORF1* sequence should be closely related to type II; however, their *ORF1* sequence in fact was related to type I. This result suggests a second recombination event occurred in L1528, which recombinant *ORF1* contains a type II sequence from the second to the last segment (Figure S6 and S7). However, the case of CECT11757 *ORF1* seems more complex. This *ORF1* is a type I closely related to that from European *S. paradoxus*, which suggests a second recombination event with a European *S. paradoxus ORF1* while maintaining the typical *COX2* 3' end region of a Far Eastern *S. paradoxus.* Another scenario might be that *COX2* 3' end is under a different substitution rate driving to homoplasy in that region, according to a "patchy-tachy" model (Sun et al. 2011). For this reason we extended the analysis to the *ORF1* region, where most of the recombinations were well supported by the invasion of *ORF1* or mediated by highly recombinogenic regions.

203 **_COX3_ supports a mtDNA transfer from _S. cerevisiae_ to American _S. paradoxus_**
204 **strains and HGT/introgression in the _COX2-ORF1_ region**

205 To improve the species assignment based on a mitochondrial gene, we sequenced
206 the *COX3* gene for those strains representative of the most frequent *COX2* haplotypes
207 (Table S1). The selection of *COX3* for species assignment was based on its lack of introns
208 or an overlapping homing endonuclease, minimizing a potential recombinant scenario.

209 The *COX3* NJ phylogenetic tree was congruent with the species tree (Figure S1 and
210 S8), except for American *S. paradoxus.* In addition, the position of *S. mikatae* was not
211 well resolved. The *COX3* MJ phylogenetic network assigned species by haplogroups
212 (Figure 5). We could not define the *S. eubayanus* haplogroup because its *COX3*
213 sequence was not available. We found a high number of nucleotide substitutions in the
214 *S. cerevisiae COX3* sequences retrieved from *Saccharomyces* Genome Resequencing
215 Project (SGRP) (H4, H9-11), indicative of assembling errors in these mitochondrial
216 genomes; nevertheless, the topologies of the MJ network and the NJ tree were not
217 affected.

218 The CECT11152 *COX3* sequence was located within *S. paradoxus* haplogroup. This
219 result confirms an HGT event involving the 3' end *COX2* and *ORF1* region (Figure S3B),
220 but not affecting the rest of the mitochondrial genome (Figure S8). A similar conclusion is
221 reached for *S. cerevisiae* x *S. kudriavzevii* hybrids (CECT1102, CECT11011,
222 CECT1990), where most of the mitochondrial genome is from *S. kudriavzevii*. The two
223 representative strains from the *COX2* haplogroup C2 (L1528 and CECT11757) possess
224 an *ORF1* closely related to the European *S. paradoxus* CECT10380 but have a *S.*
225 *cerevisiae COX3* sequence type, supporting another HGT for the *COX2-ORF1* region
226 (Figure 5 and S8).

227 Our two representatives from the American *S. paradoxus,* 120MX and CBS5313,
228 displaying an identical 5' end *COX2* segment sequence to *S. cerevisiae* sequences
229 (Figure S2A), also had a *COX3* sequence closely related to the *S. cerevisiae* haplogroup.
230 The *COX3* haplotype 17 of 120MX was identical to *S. cerevisiae COX3* sequences
231 (Figure 5 and S8). This result suggests that the American *S. paradoxus* likely inherited a
232 mitochondrial genome from *S. cerevisiae*.

233

# Discussion

## Recombinations in the *COX2-ORF1* region might be mediated by *ORF1*, GC clusters and/or AT tandem repeats.

Evidence of intraspecific mitochondrial recombination among *S. cerevisiae* strains has been shown (Dujon et al. 1974; Nunnari et al. 1997; Berger and Yaffe 2000). Indeed, a complete DNA recombination map in *S. cerevisiae* has been recently drawn (Fritsch et al. 2014). Mitochondrial recombination is initiated by a double-strand break (DSB) generated and resolved by mainly four nuclear encoded proteins (Lockshon et al. 1995; Ling and Shibata 2002; Ling et al. 2007; Ling et al. 2013) and it can be facilitated by the mitochondrial genome architecture, presence of GC clusters, or by the mobility of mitochondrial elements, such as introns or homing endonucleases (Dieckmann and Gandy 1987; Séraphin et al. 1987; Yang et al. 1998). We found that two *COX2 S. cerevisiae* and *S. paradoxus* groups of sequences are mainly driven by the presence of two highly divergent *ORF1* gene sequences, type I and type II. The presence of two clear *COX2* groups might be the result of the *ORF1* invasion.

A recent study, where fourteen *S. cerevisiae* and *S. paradoxus* mitochondrial genomes were sequenced, found HGTs between those two species mainly facilitated by GC clusters (Wu et al. 2015; Wu and Hao 2015). Our results showed recombination breakpoints close to GC clusters and AT tandem repeats in wild *Saccharomyces* strains, supporting that these genomic elements are facilitating mitochondrial recombination events. All these three elements, GC clusters, AT tandem repeats and homing endonucleases might be responsible of facilitating recombination in mitochondria.

The clearest example of *ORF1* transfer was between Far Eastern *S. paradoxus* CECT11152 and *S.* cerevisiae, which is in agreement to the described HGT of GC cluster GC48 between *S. cerevisiae* and IFO1804 (CECT11152) (Wu and Hao 2015), suggesting that both transfers likely occurred at the same time. For this reason, we propose that *ORF1* is a likely functional homing endonuclease similar to other *S. cerevisiae* free-standing homing endonucleases, such as *ORF3* (*ENS2*) (Séraphin et al. 1987; Nakagawa et al. 1992). Following nomenclature rules (Belfort and Roberts 1997), we suggest renaming *ORF1* to F-*Sce*IIIα or F-*Sce*IIIβ for *S. cerevisiae* and replacing the last two letters with the corresponding first two letters for each species *ORF1*, where “F” stands

for freestanding homing endonuclease, “S” for *Saccharomyces*, “ce” for *cerevisiae* and III because F-*Sce*I and F-*Sce*II are designated for Endo.*Sce*I (*ENS2*, *ORF3*) and HO endo, respectively. The type I and type II group membership is designated by using a suffix α and β, respectively.

**Wine *S. cerevisiae* domestication bottleneck fixed a mitochondrial variant**

Wild *Saccharomyces* yeasts reproduce mainly by mitotic divisions (Tsai et al. 2008; Liti et al. 2009); however, some industrial strains have been shown to be hybrids (Lopandic et al. 2007; González et al. 2008; Sipiczki 2008; Dunn and Sherlock 2008; Peris et al. 2012c; Pérez-Través et al. 2014) or experienced some degree of sexual mating with closely related or fairly distant species, as it is supported by the presence of nuclear genome introgressions (Liti et al. 2006; Novo et al. 2009; Dunn et al. 2012; Almeida et al. 2014). The result of hybridization and HGT in industrial conditions can be adaptive (Belloch et al. 2008; Gibson et al. 2013). Most of these studies have been focused in the nuclear genome, but little is known about the mitochondrial genotype of wild and industrial strains.

The low mutation rate of yeast’s mitochondrial coding sequences (Clark-Walker 1991) and assembling method improvements are turning the attention to the mitochondria, which have been recently used to estimate time divergence among *Schizosaccharomyces pombe* strains (Jeffares et al. 2015), and to infer the population structure of 100 *S. cerevisiae* (Wolters et al. 2015). Although the presence of recombinations or the utilization of highly polymorphic sequences is not well suited for time divergence estimation, mitochondrial sequence is useful for tracing evolutionary relationships between closely related strains (Bartelli et al. 2013; Wolters et al. 2015).

Two independent *S. cerevisiae* domestication events from wild isolates have been inferred (Legras et al. 2007), one for sake and another for wine strains (Fay and Benavides 2005; Liti et al. 2009; Schacherer et al. 2009). Indeed, *S. cerevisiae* wine domestication is attributed to be originated in the Near East (Fay and Benavides 2005; Liti et al. 2009) probably from the wild *S. cerevisiae* stock from the Asian continent (Wang et al. 2012).The high frequency of European wine strains with haplogroup C2 suggests that the bottleneck during the domestication to winemaking fixed few *COX2* variants from

haplogroup C2. At the nuclear level, similar results were observed for the wine/European strains (Liti et al. 2009; Schacherer et al. 2009). Most clinical samples are derived from wine isolates, as the mitochondrial haplogroup indicate, in agreement with previous results of the 100 *S. cerevisiae* project (Strope et al. 2015; Wolters et al. 2015). With the expansion of domesticated wine *S. cerevisiae* strains together with vineyards throughout Europe by Phoenicians and Romans, and the migration to America after European colonization, these wine *S. cerevisiae ORF1* sequences were able to recombine with other *S. cerevisiae ORF1* sequences. Wild and wine *S. cerevisiae* crosses might have been possible, as wine isolates have been found in oak trees close to vineyards (Hyma and Fay 2013). It is also clear how in the case of the dietary supplement and wine hybrids *S. cerevisiae* x *S. kudriavzevii* (IF6 and UvCEG) and a winemaking hybrid *S. cerevisiae* x *S. uvarum* (S6U) inherited a recombinant mitochondria from *S. cerevisiae* strains mostly associated with wine from Europe, South America and Africa. Before hybridization of *S. cerevisiae* with other *Saccharomyces* species to generate the IF6/UvCEG and S6U hybrids, *S. cerevisiae* was introgressed with Far Eastern *S. paradoxus*, likely before migrating to Europe (Figure 6). Introgressions occurring before hybridization might be also a potential scenario for some other hybrids, such as *S. cerevisiae* x *S. kudriavzevii* hybrids which showed introgressions from European *S. paradoxus* (Peris et al. 2012a) or introgressions from European *S. uvarum* into mitochondria of *S. cerevisiae* x *S. eubayanus* hybrids (Peris et al. 2014).

The impact of the haplogroup C2 on the winemaking process is of interest to understand the domestication of *S. cerevisiae.* Interestingly, most of our hybrids have inherited a non-*cerevisiae* mitochondrial genome (Peris et al. 2012a; Peris et al. 2012b; Peris et al. 2014; Pérez-Través et al. 2014). Previous work (Warren *et al.* 2013) has demonstrated how the inheritance of one of the parental mitochondrial genomes impacts respiration in hybrids, highlighting the importance of taking into account the fixation of the mtDNA during the generation of artificial hybrids for specific industrial processes.

**Introgressions might influence diversification**

All potential transfers detected in our study are summarized in Figure 6 which is a starting point to model mitochondrial introgressions in *Saccharomyces* genus. Some

particular *Saccharomyces* lineages have some degree of introgression, such as *S. kudriavzevii* Asia B, *S. kudriavzevii* Asia A and European and *S. mikatae* IFO1815.

The clearest and most interesting introgression was detected in all American *S. paradoxus* from populations B and C. These strains have a *COX2* sequence with a closely related 5' end to *S. cerevisiae* from haplogroup C1, and the two selected representative strains (120MX and CBS5313) shared a *COX3* sequence with *S. cerevisiae* strains. Recent evidence of phylogenetic tree incongruence among *S. cerevisiae* and American population B *S. paradoxus* YPS138 (Leducq et al. 2014) strain was also detected (Wu et al. 2015), supporting our hypothesis. Our results indicate that the American *S. paradoxus* has inherited the mtDNA from *S. cerevisiae*. Indeed, a recent study has described a *S. cerevisiae* from an unknown source with a mitochondrial genome closely related to *S. paradoxus* CBS432 (Wolters et al. 2015), which supports the hypothesis that wild strains can survive with foreign mitochondrial genomes. The increase of new isolates resulting from the improvement of isolation methods (Sampaio and Gonçalves 2008; Sylvester et al. 2015), along with the application of new methods for assembling mitochondrial genome will shed light about other American *S. paradoxus* mitochondrial inheritance and the extension of mitochondrial introgressions among other species. The presence of a foreign mitochondrial genome brings the question is mitochondrial introgression has an influence in the diversification of species. We have previously shown how hybrid *S. cerevisiae* x *S. kudriavzevii* strains inheriting a mitochondrial genome from *S. cerevisiae* are more prone to lose *S. kudriavzevii* genes (Peris et al. 2012c) suggesting an evolutionary constriction due to the proper coordination of nuclear genes with mitochondrial genes. Nonetheless, most of Bateson-Dobzhansky-Muller incompatibilities are between nuclear and mitochondrial genes (Lee et al. 2008; Chou et al. 2010; Hou et al. 2015). In this way, the accommodation of nuclear genome to this new mitochondrial genome might drive the diversification of American *S. paradoxus* from the other *S. paradoxus* lineages.

**Mitochondrial introgressions as evidence of ancestral hybridization events in wild environments**

*Saccharomyces* double and triple hybrids have been isolated from many different industrial conditions, such as “ale” and lager beer, wine, cider, dietary supplements (Casaregola et al. 2001; Barros Lopes et al. 2002; González et al. 2008; Peris et al. 2012a; Pérez-Través et al. 2014) and clinical samples (Peris et al. 2012a). However, hybrids have not been isolated from natural samples, suggesting that hybridization is only occurring in artificial conditions where hybrids are better adapted to those stressful environments (Belloch et al. 2008) by the acquisition of beneficial traits from the parents (Peris et al. 2012c; Gibson et al. 2013). HGT between *S. cerevisiae* and *S. paradoxus* may be mediated by the formation of heterokaryons due to pseudohyphae formation of *Saccharomyces* strains found in sympatric association (Wu *et al.* 2015), such as *S. cerevisiae, S. paradoxus* and *S. kudriavzevii* in Europe (Sampaio and Gonçalves 2008). However, we cannot rule-out a complete hybridization with a rapid loss of one of the parent genomes by outcrossings with the non-hybridized sibling strains. The diversity of *S. cerevisiae* x *S. kudriavzevii* hybrids (Peris et al. 2012a), most of them with lower *S. kudriavzevii* genome content, suggests a rapid loss of *S. kudriavzevii* genes after hybridization while keeping traits important for low temperature fermentations (Peris et al. 2012c). In addition, several introgressions involving most *Saccharomyces* species suggest some gene flow among them (Liti et al. 2005; Muller and McCusker 2009; Almeida et al. 2014) that supports hybridization events in natural environments. Our study suggests that Asia is a hotspot of these hybridization events and that hybridization might be influenced by the presence of all *Saccharomyces* species, as the Asian origin model suggests (Bing et al. 2014; Liti 2015).

## Material and Methods

### Saccharomyces strains and culture media

A collection of 517 *Saccharomyces* strains and 49 natural hybrids, worldwide distributed (Figure 1, Table S1), were used in this study. Species assignment for each strain was delimited by different authors using molecular techniques, such as the sequencing of 5.8S-ITS region, Random Fragment Length Polymorphisms (RFLP), a multilocus sequence approach or whole genome sequencing (Table S1). Hybrids were mostly characterized on the basis of restriction analysis of 35 different nuclear genes

(González et al. 2008; Peris et al. 2012a; Peris et al. 2012b; Pérez-Través et al. 2014). *Naumovozyma castellii* sequences were included as outgroup references to root phylogenetic trees. Yeast strains were grown at 28 ºC in YPD (2% glucose, 2% peptone and 1% yeast extract).

**PCR amplification, sequencing and gene alignments**

Total yeast DNA was extracted following the procedure described by Querol *et al.* (1992). Partial gene sequence (585bp) of the mitochondrial gene *COX2* was amplified by PCR using the primers described in Belloch *et al.* (2000). *COX3* and *ORF1* gene sequences were amplified and sequenced, using primers described in Table S2, for seventy-two *Saccharomyces* strains, representatives of the most frequent *COX2* haplotypes (Table S1). For the sequencing of *ORF1* we followed a primer walking approach (Figure 4). *ORF1* gene amplification from IFO1815 failed to be amplified, and the reference *S. eubayanus* strain yHCT76 was not available in the period of this study. Sequences were deposited in GenBank under Accession nos. JN676363-JN676823 and JN709044-JN709115. Gene sequence accession numbers from previously sequenced strains are shown in Table S1. Sequences from *S. cerevisiae* strains from the *Saccharomyces* Genome Resequencing Project (SGRP) were retrieve using the blast server (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_cerevisiae_sgrp). A PSI-Blast search was run to retrieve *ORF1* sequences from the closest sequences of non-*Saccharomyces* species.

*COX2* and *COX3* sequences were aligned using `CLUSTALW`, as implemented in `MEGA v5` (Tamura et al. 2011), and manually trimmed. For *ORF1* we used `MUSCLE` (Edgar 2004) to align the aminoacid sequences and the nucleotide sequence alignment was further refined by visual inspection in `Jalview 4.0b2` (Waterhouse et al. 2009).

***COX2*, *ORF1* and *COX3* haplotype classification, and *COX2* genetic diversity**

`DnaSP v5` (Librado and Rozas 2009) was used to calculate the number of haplotypes of *COX2*, *ORF1* and *COX3* and genetic statistics of *COX2*, such as the number of polymorphic sites(s), average number of differences between sequences (*k*), nucleotide diversity ($\pi$) and haplotype diversity (Hd) based on the species designation.

## Phylogenetic analysis and detection of recombination

*COX2, ORF1* and *COX3* phylogenetic trees reconstructed using the Neighbor-Joining (NJ) methods was performed in `MEGA v5` (Tamura et al. 2011), performing 10000 pseudoreplicates bootstrapping for branch support. *COX2* and *COX3* median joining networks were reconstructed using `PopART v1.7.2b` (http://popart.otago.ac.nz). *COX2* and *ORF1* Neighbor-Net phylogenetic networks were also reconstructed using `SplitsTree v4` (Huson and Bryant 2006) to explore the presence of incongruence in our dataset.

An alignment with representative sequences of each *COX2* haplotype was used as an input for `RDPv3.44` (Martin et al. 2010) to detect and define recombination points. Recombination points detected by two or more methods implemented in `RDPv3.44` were considered significant, applying a Bonferroni correction for multiple comparisons. Although, different recombination points were detected, we defined two *COX2* segments using the most frequent recombinant sites. *COX2* gene was divided into two segments referred as 5'-end (positions 1-496 in the alignment [124-620 in the reference *COX2* gene sequence from the S288c strain, SGD ID: S000007281]) and 3'-end (from 497-end of alignment [621-708 in the reference *COX2* sequence]). The Maximum Likelihood phylogenetic trees for both *COX2* segments was reconstructed with the best fitted models, inferred using `jModeltest` (Posada 2008). `Tree Puzzle v5.2` (Schmidt et al. 2002) was used to test the phylogenetic congruence of the two inferred phylogenetic trees to the species *Saccharomyces* phylogenetic tree topology (Borneman and Pretorius 2015). The statistical significance of these comparisons was performed by the Shimodaira-Hasegawa (Shimodaira and Hasegawa 1999) and ELW (Expected-Likelihood Weights) (Strimmer and Rambaut 2002) tests.

A concatenated alignment of *COX2* position 621 to the end and the partial *ORF1* sequences was generated. Indels, mostly due to AT repetitive regions, and GC clusters were removed. For the detection of recombinant sites we followed the approach described above. Four segments were described for the most frequent recombinant points. First segment (224bp) takes the *COX2* region and the 246 nucleotide positions of *ORF1* (corresponding to nucleotide 292 in the S288c *ORF1* gene, SGD ID: S000007282).

The last *COX2* nineteen nucleotides are the first *ORF1* nucleotides, both CDS are overlapped. The second segment takes 247 to 644 position of *ORF1* (from 203-704 in S288c *ORF1*), the third was built from 64 to 920 (706-980 in S288c), and the forth from 921 to the end of the alignment (981-1435 in S288c). Recombinant segments were also supported by the GARD (Genetic Algorithm Recombination Detection) method implemented in `Datamonkey` (Delport et al. 2010), which also perform a Kishino-Hasegawa test (Kishino and Hasegawa 1989).

The species tree was reconstructed by using a concatenated alignment of *RIP1*, *MET2* and *FUN14*, generated with `FASCONCAT` v1.0 (Kück and Meusemann 2010). Gene sequences were retrieved from Genbank, blast searches against SGRP *S. cerevisiae* sequences and *S. arboricola* online blast database (http://www.moseslab.csb.utoronto.ca/sarb/blast/), or blast searches to local databases generated with the Saccharomyces Sensu Stricto (SSS) genomes (http://www.saccharomycessensustricto.org/cgi-bin/s3.cgi?data=Assemblies&version=current). A Maximum-Likelihood (ML) phylogenetic tree was inferred using `RAxML` v8 (Stamatakis 2014) by performing 100 heuristic searches for the best gene tree, which branches were bootstrap supported by 1000 pseudosamples.

**Statistical analysis**

$\chi^2$ test for detecting bias distributions of *S. cerevisiae* strains by country or isolation source among the *COX2* haplogroups was performed in R statistical package (Adler and Murdoch D 2009). p-values were replicated 10000 times by a Monte Carlo simulation.

## Acknowledgments

We thank to C.P. Kurtzman to provide us the *S. mikatae* NRRL Y-27342. We thank Chris Todd Hittinger and William G. Alexander for critical comments on the manuscript. This work was supported by Spanish Government grant (AGL2009-12673-C02-02) and Generalitat Valenciana grants (PROMETEUS and ACOMP/2012) to EB, and from the Spanish Government FEDER (AGL2012-39937-C02-01) and Generalitat Valenciana (PROMETEOII/2014/042) to AQ. DP acknowledges to the Spanish Government for its Ministerio de Ciencia e Innovacion (MICINN) FPI fellowship. AA received PROMEP Fellowship from SEP, Mexican government. LP acknowledges to CSIC and the Spanish Ministry of Education and Science (MEC) for an I3P fellowship. SO acknowledges to MEC for the postdoctoral research contract.

## References

Adler D, Murdoch D. 2009. rgl: 3D visualization device system (Open GL).

Albertin W, da Silva T, Rigoulet M, Salin B, Masneuf-Pomarede I, *et al.* 2013. The mitochondrial genome impacts respiration but not fermentation in interspecific *Saccharomyces* hybrids. PLoS ONE. 8:e75121

Almeida P, Gonçalves C, Teixeira S, Libkind D, Bontrager M *et al.* 2014. A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. Nat Commun. 5:

Andersson JO. 2009. Gene transfer and diversification of microbial Eukaryotes. Annu Rev Microbiol. 63:177-193.

Badotti F, Vilaça ST, Arias A, Rosa CA, Barrio E. 2013. Two interbreeding populations of *Saccharomyces cerevisiae* strains coexist in cachaça fermentations from Brazil. FEMS Yeast Res. 14:289-301.

Barros Lopes M, Bellon JR, Shirley NJ, Ganter PF. 2002. Evidence for multiple interspecific hybridization in *Saccharomyces sensu stricto* species. FEMS Yeast Res. 1:323-331.

501 Bartelli TF, Ferreira RC, Colombo AL, Briones MRS. 2013. Intraspecific comparative
502 genomics of *Candida albicans* mitochondria reveals non-coding regions under neutral
503 evolution. Infection, Genetics and Evolution. 14:302-312.
504 Basse CW. 2010. Mitochondrial inheritance in fungi. Curr Opin Microbiol. 13:712-719.
505 Belfort M, Roberts RJ. 1997. Homing endonucleases: keeping the house in order. Nucl
506 Acids Res. 25:3379-3388.
507 Belloch C, Querol A, Garcia MD, Barrio E. 2000. Phylogeny of the genus *Kluyveromyces*
508 inferred from the mitochondrial cytochrome-c oxidase II gene. Int J Syst Evol Microbiol.
509 50:405-416.
510 Belloch C, Orlic S, Barrio E, Querol A. 2008. Fermentative stress adaptation of hybrids
511 within the *Saccharomyces sensu stricto* complex. Int J Food Microbiol. 122:188-195.
512 Berger KH, Yaffe MP. 2000. Mitochondrial DNA inheritance in *Saccharomyces cerevisiae*.
513 Trends Microbiol. 8:508-513.
514 Bing J, Han PJ, Liu WQ, Wang QM, Bai FY. 2014. Evidence for a Far East Asian origin
515 of lager beer yeast. Curr Biol. 24:R380-R381.
516 Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws,
517 Mechanisms, and Models. Annu Rev Genet. 35:125-148.
518 Birky CW, Arlene RA, Rosemary D, Maureen C. 1982. Mitochondrial transmission
519 genetics: replication, recombination, and segregation of mitochondrial DNA and its
520 inheritance in crosses. Mitochondrial Genes. Cold Spring Harbor Monograph Archive.
521 Borneman AR, Pretorius IS. 2015. Genomic insights into the *Saccharomyces sensu*
522 *stricto* complex. Genetics. 199:281-291.
523 Burt A, Koufopanou V. 2004. Homing endonuclease genes: the rise and fall and rise again
524 of a selfish element. Curr Opin Genet Dev. 14:609-615.
525 Casaregola S, Nguyen HV, Lapathitis G, Kotyk A, Gaillardin C. 2001. Analysis of the
526 constitution of the beer yeast genome by PCR, sequencing and subtelomeric sequence
527 hybridization. Int J Syst Evol Microbiol. 51:1607-1618.
528 Chou JY, Hung YS, Lin KH, Lee HY, Leu JY. 2010. Multiple molecular mechanisms cause
529 reproductive isolation between three yeast species. PLoS Biol. 8:e1000432
530 Clark-Walker GD. 1991. Contrasting mutation rates in mitochondrial and nuclear genes
531 of yeasts versus mammals. Curr Genet. 20:195-198.

532 Colleaux L, d'Auriol L, Betermier M, Cottarel G, Jacquier A, Galibert F, Dujon B. 1986.
533 Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into
534 E. coli as a specific double strand endonuclease. 44:521-533.
535 Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: a
536 suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 26:2455-
537 2457.
538 Dieckmann C, Gandy B. 1987. Preferential recombination between GC clusters in yeast
539 mitochondrial DNA. EMBO J. 6:4197-4203.
540 Dujon B, Slonimski PP, Weill L. 1974. Mitochondrial genetics IX: A model for
541 recombination and segregation of mitochondrial genomes in *Saccharomyces cerevisiae*.
542 Genetics. 78:415-437.
543 Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G. 2012. Analysis of the *Saccharomyces*
544 *cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse
545 yeast strains from differing industrial environments. Genome Res. 22:908-924.
546 Dunn B, Sherlock G. 2008. Reconstruction of the genome origins and evolution of the
547 hybrid lager yeast *Saccharomyces pastorianus*. Genome Res. 18:1610-1623.
548 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
549 throughput. Nucl Acids Res. 32:1792-1797.
550 Fay JC, Benavides JA. 2005. Evidence for domesticated and wild populations of
551 *Saccharomyces cerevisiae*. PLoS Genet. 1:e5
552 Fritsch ES, Chabbert CD, Klaus B, Steinmetz LM. 2014. A genome-wide map of
553 mitochondrial DNA recombination in yeast. Genetics. 198:755-771.
554 Gibson BR, Storgårds E, Krogerus K, Vidgren V. 2013. Comparative physiology and
555 fermentation performance of Saaz and Frohberg lager yeast strains and the parental
556 species *Saccharomyces eubayanus*. Yeast. 30:255-266.
557 Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish gene. PNAS.
558 96:13880-13885.
559 González SS, Barrio E, Querol A. 2008. Molecular characterization of new natural hybrids
560 between *S. cerevisiae* and *S. kudriavzevii* from brewing. Appl Environ Microbiol. 74:2314-
561 2320.
562 Green DR, Reed JC. 1998. Mitochondria and Apoptosis. Science. 281:1309-1312.

563 Hao W, Richardson AO, Zheng Y, Palmer JD. 2010. Gorgeous mosaic of mitochondrial
564 genes created by horizontal transfer and gene conversion. Proc Natl Acad Sci U S A.
565 107:21576-21581.
566 Hatefi Y. 1985. The mitochondrial electron transport and oxidative phosphorylation
567 system. Annu Rev Biochem. 54:1015-1069.
568 Hill GE. 2015. Mitonuclear ecology. Mol Biol Evol.
569 Hittinger CT. 2013. *Saccharomyces* diversity and evolution: a budding model genus.
570 29:309-317.
571 Hou J, Friedrich A, Gounot JS, Schacherer J. 2015. Comprehensive survey of condition-
572 specific reproductive isolation reveals genetic incompatibility in yeast. Nat Commun. 6:
573 Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies.
574 Mol Biol Evol. 23:254-267.
575 Hyma KE, Fay JC. 2013. Mixing of vineyard and oak-tree ecotypes of *Saccharomyces*
576 *cerevisiae* in North American vineyards. Mol Ecol. 22:2917-2930.
577 Jeffares DC, Rallis C, Rieux A *et al.* 2015. The genomic and phenotypic diversity of
578 *Schizosaccharomyces pombe*. Nat Genet. 47:235-241.
579 Keeling P. 2009. Role of Horizontal Gene Transfer in the evolution of photosynthetic
580 Eukaryotes and their plastids. In: Gogarten M, Gogarten J, Olendzenski L, editors.
581 Horizontal Gene Transfer. Humana Press. p. 501-515.
582 Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the
583 evolutionary tree topologies from DNA sequence data, and the branching order in
584 hominoidea. J Mol Evol. 29:170-179.
585 Kück P, Meusemann K. 2010. FASconCAT, Version 1.0. Zool. Forschungsmuseum A.
586 Koenig, Germany.
587 Kurtzman CP, Robnett CJ. 2003. Phylogenetic relationships among yeasts of the
588 '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast
589 Res. 3:417-432.
590 Leducq JB, Charron G, Samani P *et al*. 2014. Local climatic adaptation in a widespread
591 microorganism. Proc R Soc Lond B Biol Sci. 281:

592 Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. 2008. Incompatibility of
593 nuclear and mitochondrial genomes causes hybrid sterility between two yeast species.
594 135:1065-1073.
595 Legras JL, Merdinoglu D, Cornuet JM, Karst F. 2007. Bread, beer and wine:
596 *Saccharomyces cerevisiae* diversity reflects human history. Mol Ecol. 16:2091-2102.
597 Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA
598 polymorphism data. Bioinformatics. 25:1451-1452.
599 Ling F, Hori A, Shibata T. 2007. DNA recombination-initiation plays a role in the extremely
600 biased inheritance of yeast [rho-] mitochondrial DNA that contains the replication origin
601 ori5. Mol Cell Biol. 27:1133-1145.
602 Ling F, Hori A, Yoshitani A, Niu R, Yoshida M, Shibata T. 2013. Din7 and Mhr1 expression
603 levels regulate double-strand-break□induced replication and recombination of mtDNA at
604 ori5 in yeast. Nucl Acids Res. 41:5799-5816.
605 Ling F, Mikawa T, Shibata T. 2011. Enlightenment of yeast mitochondrial homoplasmy:
606 diversified roles of gene conversion. Genes. 2:169-190.
607 Ling F, Shibata T. 2002. Recombination-dependent mtDNA partitioning: in vivo role of
608 Mhr1p to promote pairing of homologous DNA. EMBO J. 21:4730-4740.
609 Liti G. 2015. The fascinating and secret wild life of the budding yeast *S. cerevisiae*. eLife
610 Sciences. 4:
611 Liti G, Barton DB, Louis EJ. 2006. Sequence diversity, reproductive isolation and species
612 concepts in *Saccharomyces*. Genetics. 174:839-850.
613 Liti G, Carter DM, Moses AM *et al*. 2009. Population genomics of domestic and wild
614 yeasts. Nature. 458:337-341.
615 Liti G, Peruffo A, James SA, Roberts IN, Louis EJ. 2005. Inferences of evolutionary
616 relationships from a population survey of LTR-retrotransposons and telomeric-associated
617 sequences in the *Saccharomyces sensu stricto* complex. Yeast. 22:177-192.
618 Lockshon D, Zweifel SG, Freeman-Cook LL, Lorimer HE, Brewer BJ, Fangman WL. 1995.
619 A role for recombination junctions in the segregation of mitochondrial DNA in yeast. Cell.
620 81:947-955.
621 Lopandic K, Wallner E, Tscheik G, Leitner G, Querol A, Borth N, Tiefenbrunner W. 2007.
622 Genetically different wine yeasts isolated from Austrian vine-growing regions influence

623 wine aroma differently and contain putative hybrids between *Saccharomyces cerevisiae*
624 and *Saccharomyces kudriavzevii*. FEMS Yeast Res. 7:953-965.
625 MacAlpine DM, Perlman PS, Butow RA. 1998. The high mobility group protein Abf2p
626 influences the level of yeast mitochondrial DNA recombination intermediates in vivo. Proc
627 Natl Acad Sci U S A. 95:6739-6743.
628 Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. 2010. RDP3: a flexible
629 and fast computer program for analyzing recombination. Bioinformatics. 26:2462-2463.
630 Muller LAH, McCusker JH. 2009. A multispecies-based taxonomic microarray reveals
631 interspecies hybridization and introgression in *Saccharomyces cerevisiae*. FEMS Yeast
632 Res. 9:143-152.
633 Nakagawa K, Morishima N, Shibata T. 1992. An endonuclease with multiple cutting sites,
634 Endo.SceI, initiates genetic recombination at its cutting site in yeast mitochondria. EMBO
635 J. 11:2707-2715.
636 Novo M, Bigey F, Beyne E *et al*. 2009. Eukaryote-to-eukaryote gene transfer events
637 revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118.
638 Proc Natl Acad Sci U S A. 106:16333-16338.
639 Nunnari J, Marshall WF, Straight A, Murray A, Sedat JW, Walter P. 1997. Mitochondrial
640 transmission during mating in *Saccharomyces cerevisiae* is determined by mitochondrial
641 fusion and fission and the intramitochondrial segregation of mitochondrial DNA. Mol Biol
642 Cell. 8:1233-1242.
643 Pérez-Través L, Lopes CA, Querol A, Barrio E. 2014. On the complexity of the
644 *Saccharomyces bayanus* taxon: hybridization and potential hybrid speciation. PLoS ONE.
645 9:e93729
646 Peris D, Belloch C, Lopandic K, Álvarez-Pérez JM, Querol A, Barrio E. 2012a. The
647 molecular characterization of new types of *S. cerevisiae* x *S. kudriavzevii* hybrid yeasts
648 unveils a high genetic diversity. Yeast. 29:81-91.
649 Peris D, Lopes CA, Arias A, Barrio E. 2012b. Reconstruction of the evolutionary history
650 of *Saccharomyces cerevisiae* x *S. kudriavzevii* hybrids based on multilocus sequence
651 analysis. PLoS ONE. 7:e45527

652 Peris D, Lopes CA, Belloch C, Querol A, Barrio E. 2012c. Comparative genomics among
653 *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* natural hybrid strains isolated
654 from wine and beer reveals different origins. BMC Genomics. 13:407
655 Peris D, Sylvester K, Libkind D, Gonçalves P, Sampaio JP, Alexander WG, Hittinger CT.
656 2014. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its
657 lager-brewing hybrids. Mol Ecol. 23:2031-2045.
658 Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol. 25:1253-
659 1256.
660 Querol A, Barrio E, Huerta T, Ramon D. 1992. Molecular monitoring of wine fermentations
661 conducted by active dry yeast strains. Appl Environ Microbiol. 58:2948-2953.
662 Rodríguez ME, Pérez-Través L, Sangorrín MP, Barrio E, Lopes CA. 2014.
663 *Saccharomyces eubayanus* and *Saccharomyces uvarum* associated with the
664 fermentation of *Araucaria araucana* seeds in Patagonia. FEMS Yeast Res. n/a-n/a.
665 Sampaio JP, Gonçalves P. 2008. Natural populations of *Saccharomyces kudriavzevii* in
666 Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S.*
667 *paradoxus*. Appl Environ Microbiol. 74:2144-2152.
668 Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. 2009. Comprehensive
669 polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*.
670 Nature. 458:342-345.
671 Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum
672 likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics.
673 18:502-504.
674 Séraphin B, Simon M, Faye G. 1987. The mitochondrial reading frame RF3 is a functional
675 gene in *Saccharomyces uvarum*. J Biol Chem. 262:10146-10153.
676 Shimodaira H, Hasegawa M. 1999. Multiple comparisons of Log-Likelihoods with
677 applications to Phylogenetic inference. Mol Biol Evol. 16:1114-1116.
678 Sickmann A, Reinders J, Wagner Y *et al*. 2003. The proteome of *Saccharomyces*
679 *cerevisiae* mitochondria. Proc Natl Acad Sci U S A. 100:13207-13212.
680 Sipiczki M. 2008. Interspecies hybridization and recombination in *Saccharomyces* wine
681 yeasts. FEMS Yeast Res. 8:996-1007.

682 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis
683 of large phylogenies. Bioinformatics. 30:1312-1313.
684 Starkov AA. 2008. The role of mitochondria in reactive oxygen species metabolism and
685 signaling. Annals of the New York Academy of Sciences. 1147:37-52.
686 Strimmer K, Rambaut A. 2002. Inferring confidence sets of possibly misspecified gene
687 trees. Proceedings Biological sciences / The Royal Society. 269:137-142.
688 Strope PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, Dietrich FS,
689 McCusker JH. 2015. The 100-genomes strains, an *S. cerevisiae* resource that illuminates
690 its natural phenotypic and genotypic variation and emergence as an opportunistic
691 pathogen. Genome Res.
692 Sun S, Evans BJ, Golding GB. 2011. "Patchy-Tachy" leads to false positives for
693 recombination. Mol Biol Evol. 28:2549-2559.
694 Sylvester K, Wang QM, James B, Mendez R, Hulfachor AB, Hittinger CT. 2015.
695 Temperature and host preferences drive the diversification of *Saccharomyces* and other
696 yeasts: a survey and the discovery of eight new yeast species. FEMS Yeast Res.
697 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular
698 Evolutionary Genetics Analysis using Maximum Likelihood, evolutionary distance, and
699 Maximum Parsimony methods. Mol Biol Evol. 28:2731-2739.
700 Taylor JW. 1986. Fungal evolutionary biology and mitochondrial DNA. Exp Mycol. 10:259-
701 269.
702 Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast
703 *Saccharomyces paradoxus*: quantifying the life cycle. Proc Natl Acad Sci U S A.
704 105:4957-4962.
705 Wang QM, Liu WQ, Liti G, Wang SA, Bai FY. 2012. Surprisingly diverged populations of
706 *Saccharomyces cerevisiae* in natural environments remote from human activity. Mol Ecol.
707 21:5404-5417.
708 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-
709 -a multiple sequence alignment editor and analysis workbench. Bioinformatics. 25:1189-
710 1191.
711 Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in
712 *Saccharomyces cerevisiae*. BMC Genomics. 16:451

713 Wu B, Buljic A, Hao W. 2015. Extensive horizontal transfer and homologous
714 recombination generate highly chimeric mitochondrial genomes in yeast. Mol Biol Evol.
715 Wu B, Hao W. 2015. A dynamic mobile DNA family in the yeast mitochondrial genome.
716 G3: Genes|Genomes|Genetics.
717 Yang J, Mohr G, Perlman PS, Lambowitz AM. 1998. Group II intron mobility in yeast
718 mitochondria: target DNA-primed reverse transcription activity of ai1 and reverse splicing
719 into DNA transposition sites in vitro1. J Mol Biol. 282:505-523.
720

## 721 Figure Legends

722 **Figure 1. *Saccharomyces* worldwide distribution map.** *Saccharomyces* strains were
723 colored according to the species designation or the hybrid character (Table S1). Highly
724 sampled regions were zoom in, for better resolution.
725 **Figure 2. *COX2* Median-Joining phylogenetic network.** A hundred and ten *COX2*
726 haplotypes (Table S1) were represented by circles. Circles sizes are scaled according to
727 the haplotype frequency. Pie charts show the frequency of each species in that particular
728 haplotype. Number of mutations from one haplotype to another are indicated by lines in
729 the edges connecting haplotypes.
730 **Figure 3. *ORF1* Neighbor-Net phylonetwork.** The phylonetwork differentiated two
731 groups of *ORF1* sequences, type I and type II. Haplotypes for *ORF1* are represented with
732 M[number] according to Table S1. Haplotype numbers are colored according to the
733 species having them. Scale is given in nucleotide substitution per site.
734 **Figure 4. *COX2-ORF1* mitochondrial region scheme.** A representation of the region
735 *COX2-ORF1* was shown. Homing endonuclease domains were indicated by empty
736 arrows. Primer names and their targets are shown by arrows. Compatible primers are
737 indicated by showing identical arrow types. Location of GC clusters and AT tandem
738 repeats are drawn. Strains and haplotypes containing GC clusters and AT tandem
739 repeats are indicated in Table S3, Figure S5 and S9. *COX2* and *ORF1* coding sequences
740 are overlapped and represented by the grey region.
741 **Figure 5. *COX3* Median-Joining phylogenetic network.** Thirty-four *COX3* haplotypes
742 (Table S1) were represented by circles. Circles sizes are scaled according to the number
743 of sequences in a haplotype. Pie charts show the frequency of each species in that

744 particular haplotype. Asterisks indicate haplotype sequences with nucleotide errors.
745 Number of mutations from one haplotype to another are indicated by lines in the edges
746 connecting haplotypes.

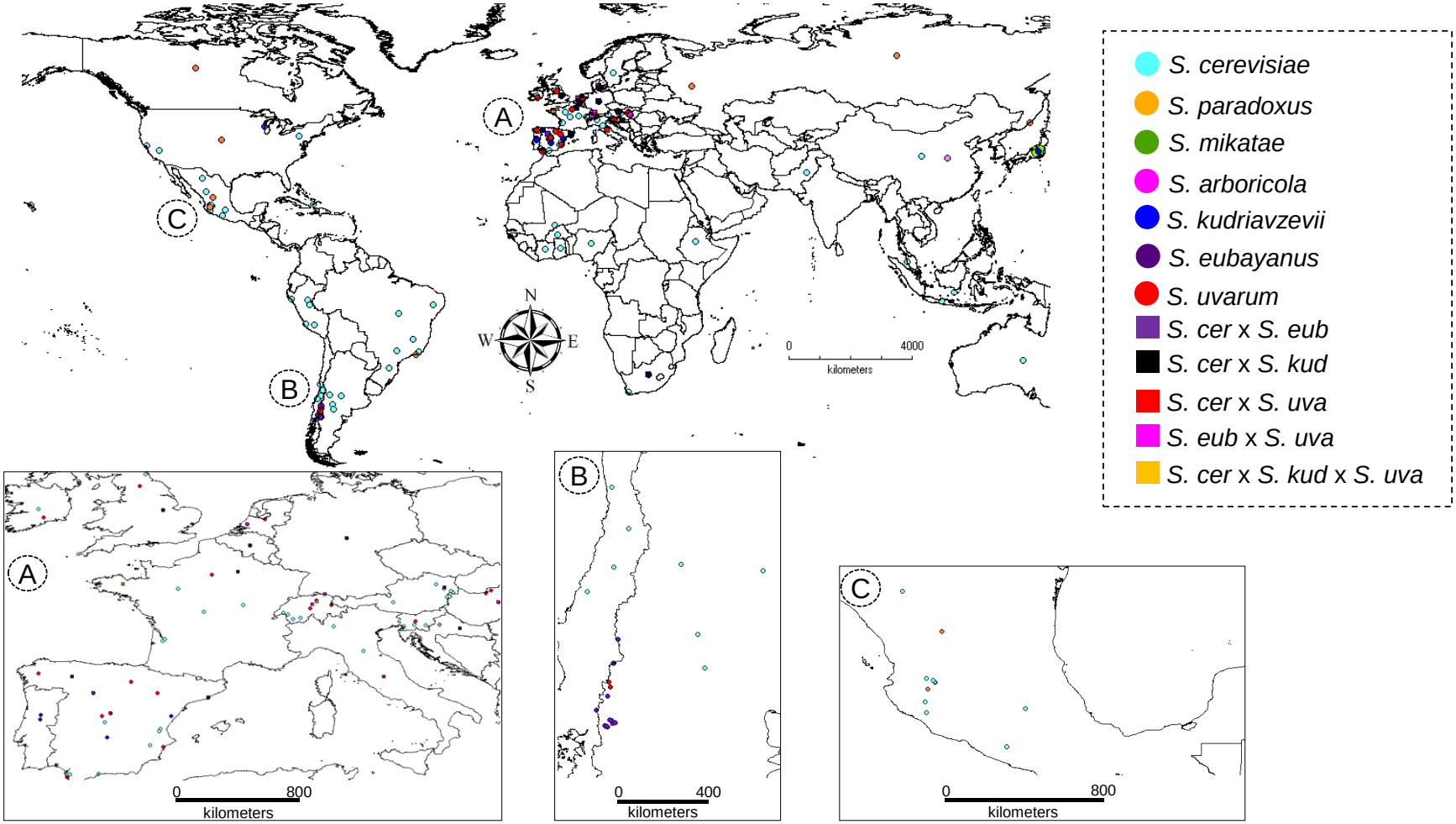
747 **Figure 6. A summary model of potential *COX2-ORF1* introgression in wild**
748 **environments and during domestication.** A starting model of potential introgressions
749 in *Saccharomyces* is described. Black, light blue, orange, green, dark blue, pink, red and
750 purple yeasts represent the yeast ancestors, *S. cerevisiae*, *S. paradoxus*, *S. mikatae,*
751 *S. kudriavzevii, S. arboricola, S. uvarum* and *S. eubayanus* yeasts, respectively. Small
752 circles inside yeasts represent the mtDNA, which is colored according to the species
753 designation. A *S. paradoxus* yeast with a blue small circle indicates an inheritance of
754 mtDNA from *S. cerevisiae*. Red, green or red/green boxes represent type I, type II or
755 recombinant *ORF1* sequences, respectively. A question tag indicates dubious scenario,
756 unknown *ORF1* or unknown mtDNA sequence due to the absence of *COX3* sequence to
757 support the mtDNA inheritance.

758

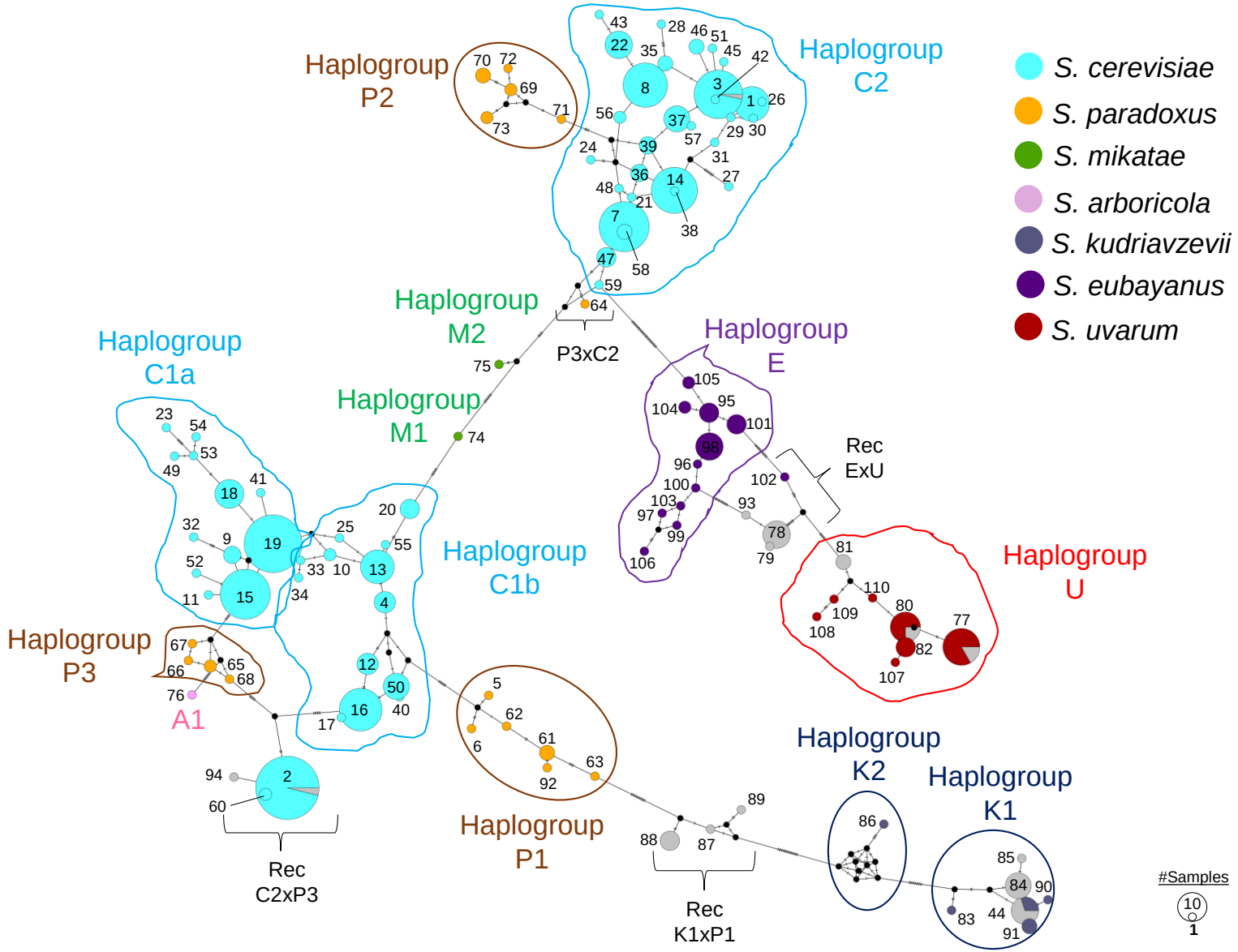
759 ## Tables and Figures

760

**Figure 1**


A
B
C
N
W
E
S
0
4000
kilometers
S. cerevisiae
S. paradoxus
S. mikatae
S. arboricola
S. kudriavzevii
S. eubayanus
S. uvarum
S. cer x S. eub
S. cer x S. kud
S. cer x S. uva
S. eub x S. uva
S. cer x S. kud x S. uva
A
0
800
kilometers
B
0
400
kilometers
C
0
800
kilometers

764 **Figure 2**




765

**Figure 3**

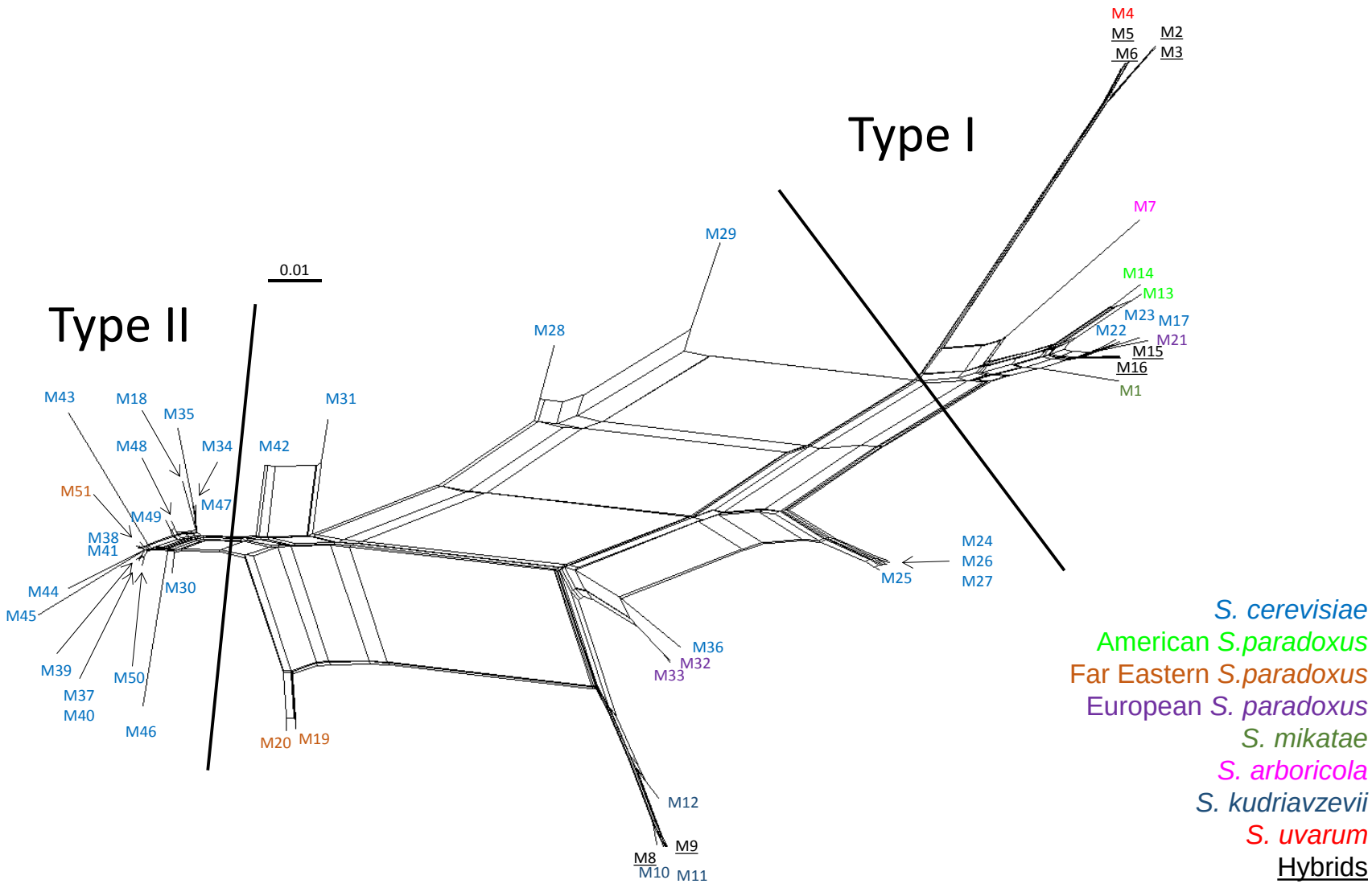
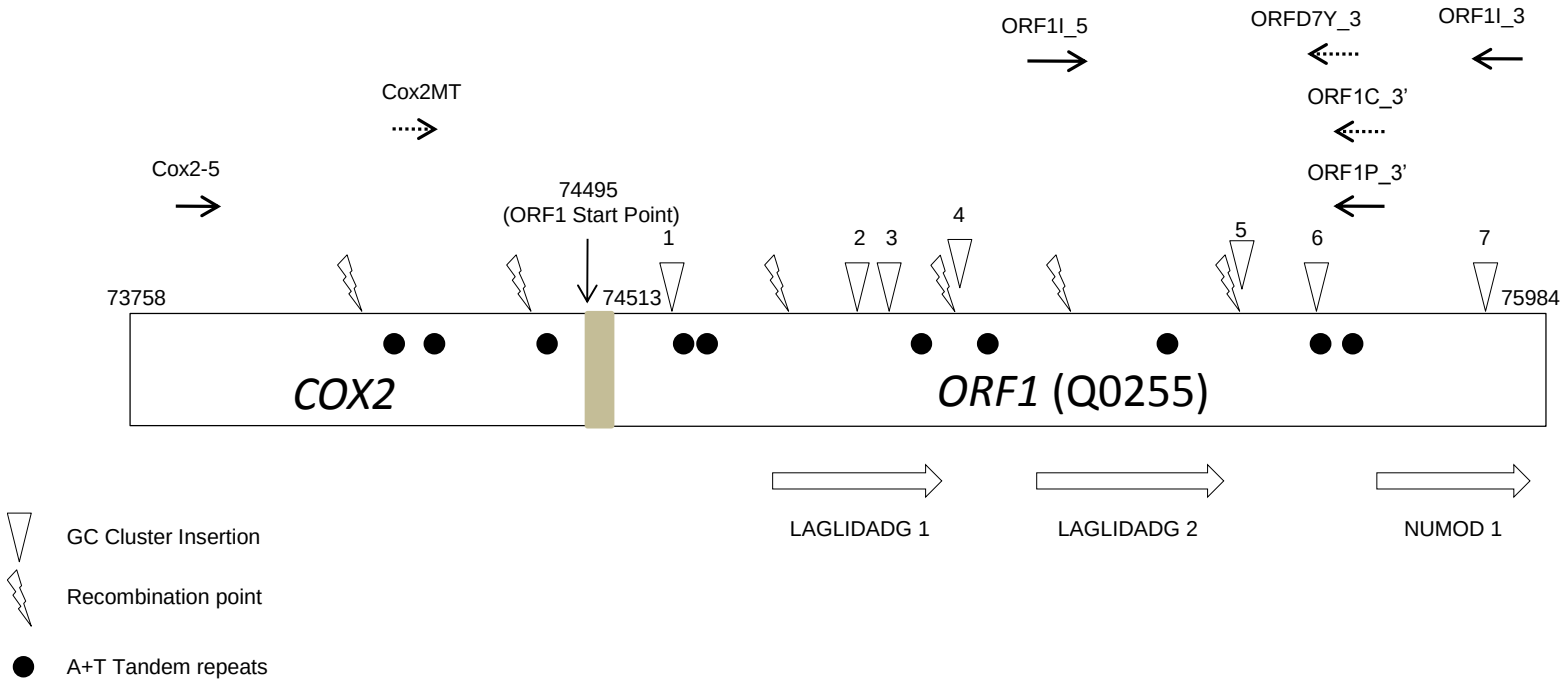
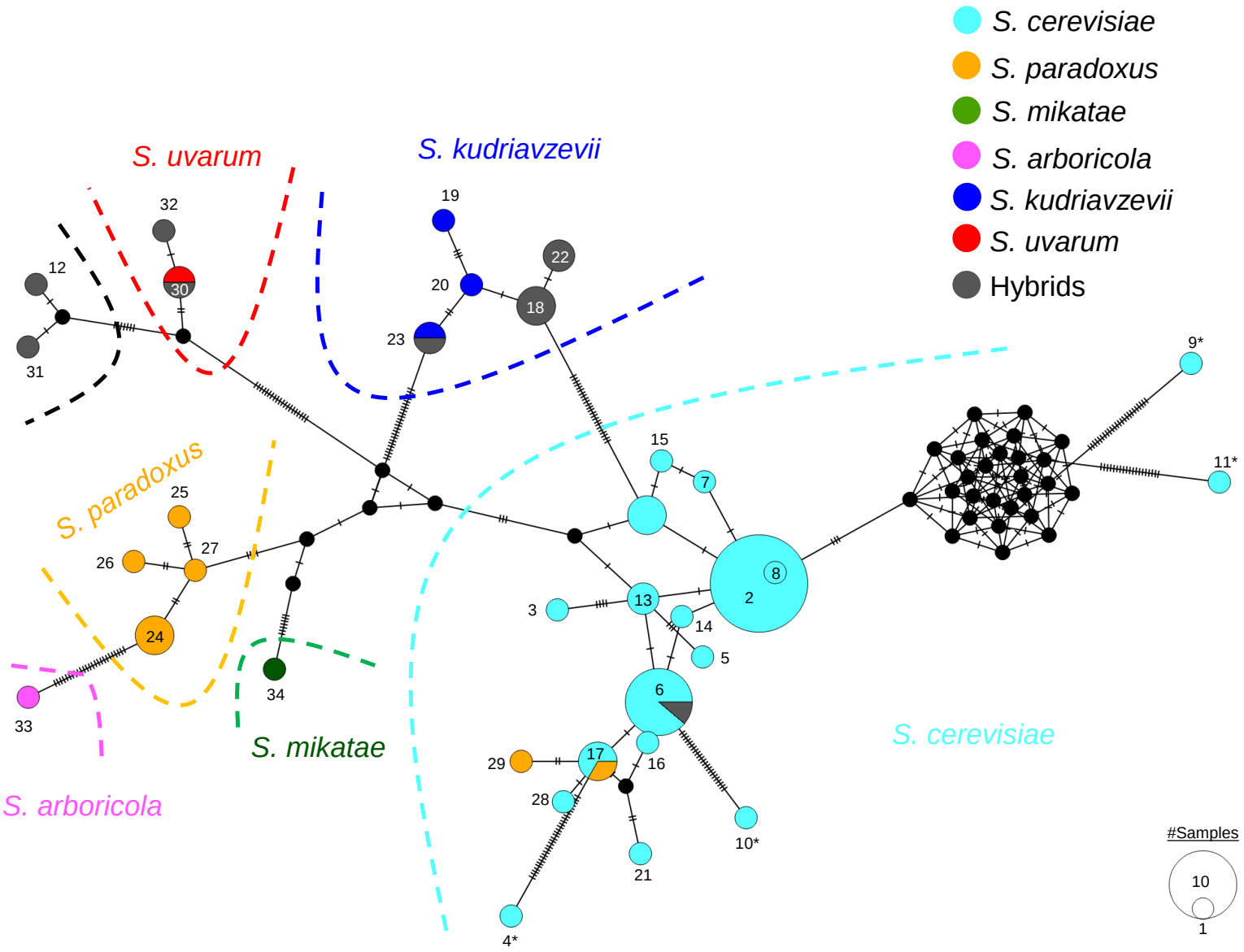

Type I
Type II
0.01
M4
M5
M6
M2
M3
M7
M29
M14
M13
M23
M17
M22
M21
M15
M16
M1
M28
M31
M43
M18
M35
M48
M34
M42
M51
M47
M49
M38
M41
M24
M26
M27
M25
M44
M45
M30
M36
M32
M33
M39
M50
M37
M40
M46
M20
M19
M12
M9
M8
M10
M11
S. cerevisiae
American S.paradoxus
Far Eastern S.paradoxus
European S. paradoxus
S. mikatae
S. arboricola
S. kudriavzevii
S. uvarum
Hybrids

**Figure 4**

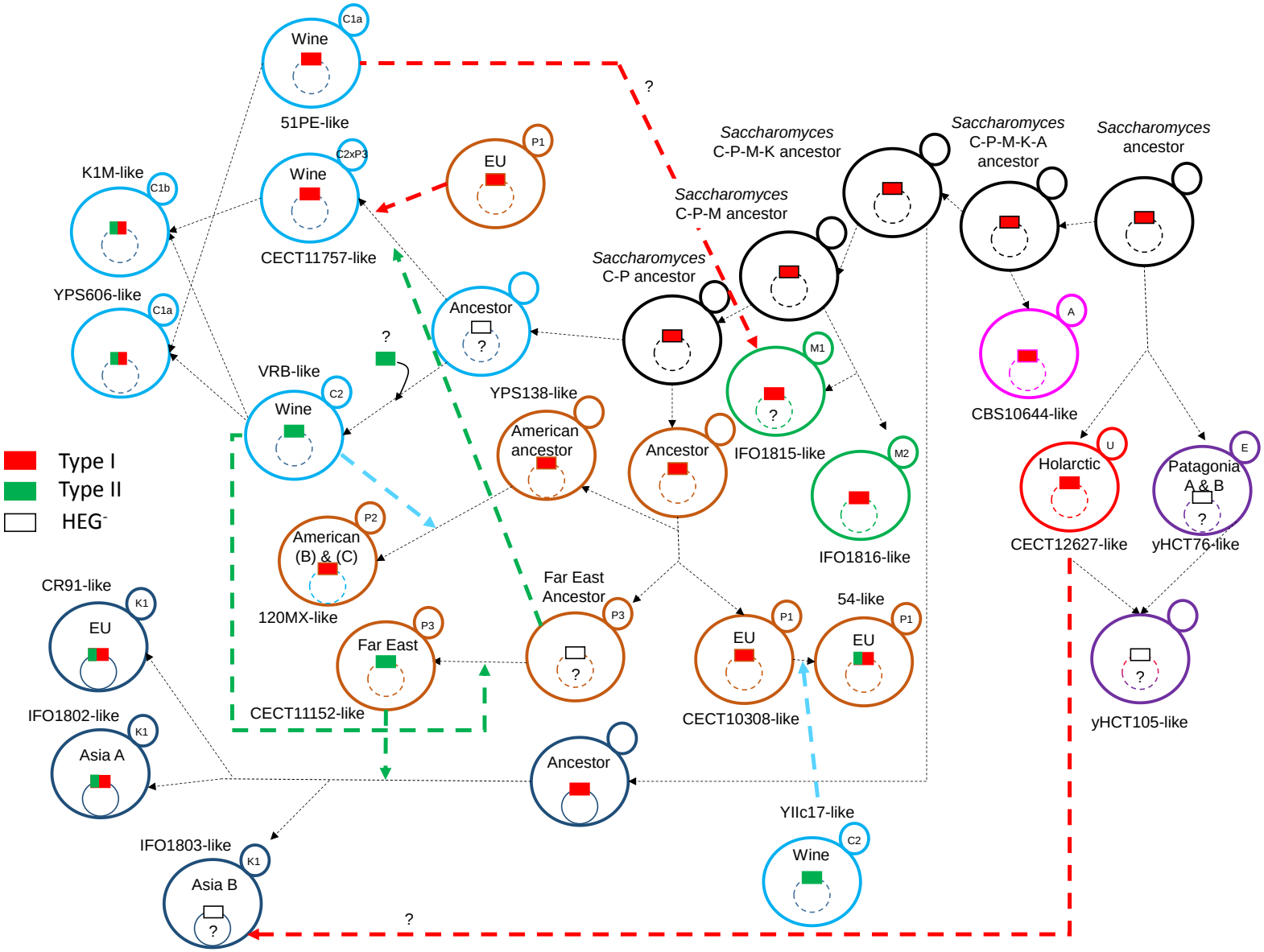

ORF1I_5
ORFD7Y_3
ORF1I_3
Cox2MT
ORF1C_3'
Cox2-5
ORF1P_3'
74495
(ORF1 Start Point)
1
2
3
4
5
6
7
73758
74513
75984
COX2
ORF1 (Q0255)
LAGLIDADG 1
LAGLIDADG 2
NUMOD 1
GC Cluster Insertion
Recombination point
A+T Tandem repeats

772 **Figure 5**




773

774 **Figure 6**




775