

## ORIGINAL ARTICLE OPEN ACCESS

# Footprints of Human Migration in the Population Structure of Wild Baker's Yeast

Jacqueline J. Peña<sup>1</sup> | Eduardo F. C. Scopel<sup>2,3</sup> | Audrey K. Ward<sup>4</sup> | Douda Bensasson<sup>1,2</sup> 

<sup>1</sup>Department of Plant Biology, University of Georgia, Athens, Georgia, USA | <sup>2</sup>Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA

<sup>3</sup>Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, Wisconsin, USA | <sup>4</sup>Department of Genetics, University of Georgia, Athens, Georgia, USA

**Correspondence:** Douda Bensasson ([dbensasson@uga.edu](mailto:dbensasson@uga.edu))

**Received:** 14 August 2024 | **Revised:** 14 December 2024 | **Accepted:** 17 January 2025

**Handling Editor:** Isheng J. Tsai

**Funding:** This work was funded by an NSF LSAMP Bridge to the Doctorate Fellowship Program award (grant number: 1702361) to JJP; by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program (grant no. GT13544) awarded to JJP and DB; and by a National Science Foundation grant (IOS, no. 1946046) awarded to DB.

**Keywords:** ecological genomics | genetic admixture | phylogeography | *Saccharomyces cerevisiae* | wine yeast | yeast ecology

## ABSTRACT

Humans have a long history of fermenting food and beverages that led to domestication of the baker's yeast, *Saccharomyces cerevisiae*. Despite their tight companionship with humans, yeast species that are domesticated or pathogenic can also live on trees. Here we used over 300 genomes of *S. cerevisiae* from oaks and other trees to determine whether tree-associated populations are genetically distinct from domesticated lineages and estimate the timing of forest lineage divergence. We found populations on trees are highly structured within Europe, Japan, and North America. Approximate estimates of when forest lineages diverged out of Asia and into North America and Europe coincide with the end of the last ice age, the spread of agriculture, and the onset of fermentation by humans. It appears that migration from human-associated environments to trees is ongoing. Indeed, patterns of ancestry in the genomes of three recent migrants from the trees of North America to Europe could be explained by the human response to the Great French Wine Blight. Our results suggest that human-assisted migration affects forest populations, albeit rarely. Such migration events may even have shaped the global distribution of *S. cerevisiae*. Given the potential for lasting impacts due to yeast migration between human and natural environments, it seems important to understand the evolution of human commensals and pathogens in wild niches.

## 1 | Introduction

Since the last ice age, humans have transitioned from a hunter-gatherer to a sedentary lifestyle and developed new technologies for preserving food, including fermentation (McGovern 2003). The baker's yeast, *Saccharomyces cerevisiae*, is a driver of such fermentations and is used to produce beer, wine, sake, cocoa, and coffee (Marsit et al. 2017). The earliest archaeological evidence of fermented rice, honey, and fruit in China dates to 7000 BCE (McGovern et al. 2004). Fermented beer was first

discovered in ancient Sumerian vessels from 6000 BCE (Michel, McGovern, and Badler 1992), and there is evidence for wine production between 6000 and 4000 BCE in Iran, the Caucasus, and Mesopotamia (Pretorius 2000; McGovern 2003). Wine production then spread throughout the Mediterranean and was prevalent across Europe and Northern Africa by 500 BCE (Pretorius 2000).

Today, the population genetics of *S. cerevisiae* shows imprints of domestication with several genetic lineages associated with distinct

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

domestication events (Almeida et al. 2015; Duan et al. 2018; Fay et al. 2019; Fay and Benavides 2005; Gallone et al. 2016; Gayevskiy, Lee, and Goddard 2016; Legras et al. 2007, 2018; Liti et al. 2009; Peter et al. 2018; Schacherer et al. 2009). Deep sampling of wild strains from Chinese and Taiwanese forests revealed high levels of lineage diversity compared to all other lineages, and the current consensus is that East Asian forests likely harboured the ancestral source populations that gave rise to all global *S. cerevisiae* lineages (Duan et al. 2018; Lee et al. 2022; Wang et al. 2012). Even outside Asia, sampling of *S. cerevisiae* from natural environments shows a wild side to this human-associated yeast species; there are wild genetic lineages that are distinct from known domesticated lineages (Almeida et al. 2015; Cromie et al. 2013; Fay and Benavides 2005; Han et al. 2021; Liti et al. 2009; Peter et al. 2018; Tilakaratna and Bensasson 2017).

How much has human activity affected the ecology and evolution of wild populations of human-associated yeast species? Human pathogenic yeast such as *Candida albicans*, *Candida glabrata*, *Candida parapsilosis*, and *Candida tropicalis* can be isolated from trees and other plant habitats (Bensasson et al. 2019; Opulente et al. 2019; Robinson, Pinharanda, and Bensasson 2016), and other forest yeast species are associated with humans (Boynton and Greig 2014; Mozzachiodi et al. 2022). Here, we make use of the extensive genome data available for *S. cerevisiae* and focus on the population structure and phylogenetic relationships of wild *S. cerevisiae* from a single ecological niche. By studying strains from oaks and other trees, we characterise populations in an ancestral niche while avoiding the complications of genetic admixture more commonly seen in *S. cerevisiae* from fruit, flowers, and insects (Hyma and Fay 2013; Tilakaratna and Bensasson 2017). Specifically, we identified tree-associated populations and estimated the timing of wild yeast migration events. We show that (i) tree-associated *S. cerevisiae* populations are highly structured, (ii) the worldwide spread of forest populations out of Asia probably occurred since the last glacial maximum, and lastly (iii) human-assisted migration is ongoing and may include migration from the USA to Europe since the Great French Wine Blight.

## 2 | Materials and Methods

### 2.1 | Yeast Strains and Genome Data

Whole-genome sequences for strains sampled from trees were compiled from publicly available data ( $N=295$ ; Table S1) (Almeida et al. 2015; Barbosa et al. 2016; Bergström et al. 2014; Duan et al. 2018; Fay et al. 2019; Gayevskiy, Lee, and Goddard 2016; Han et al. 2021; Pontes et al. 2020; Skelly et al. 2013; Song et al. 2015; Strope et al. 2015; Yue et al. 2017). We defined *S. cerevisiae* tree-sampled strains as those isolated from tree bark, exudate, and leaves from trees or litter, and we also included strains from any soil. Metadata was compiled for each genome sequence to include geographical origin, ecological substrate, and previously reported genetic clade associated with the strain (Table S1). New whole-genome sequence data was generated for strains from trees in Indiana and Kentucky ( $N=7$ ; Osburn et al. 2018), North Carolina ( $N=9$ ; Diezmann and Dietrich 2009), and Europe ( $N=3$ ; Robinson, Pinharanda, and Bensasson 2016), and for new *S. cerevisiae* strains from the bark of white oak (*Quercus alba*) and live oak (*Q. virginiana*) from Georgia, Florida, Pennsylvania, and North Carolina

( $N=15$ ; Bensasson lab). DNA was extracted from single yeast colonies using the Promega Wizard Genomic DNA purification kit following the manufacturer's protocol for yeast except that only 75 units of lyticase (Sigma) were used in an overnight incubation at 37°C. For the generation of genome data from 22 strains from the Bensasson and Osburn labs, paired-end Illumina libraries were generated by the Georgia Genomics and Bioinformatics Core using the purePlex DNA Library Preparation Kit (GGBC Project #5256) or the Nextera DNA-Seq Library Protocol (GGBC Project #5881). Paired-end sequencing was performed on the Illumina NextSeq2000 platform ( $2 \times 150$  bp). The remaining strains were sequenced at the University of Manchester as described in Almeida et al. (2015). Genome data is available on NCBI-SRA under project number PRJNA1090965 and includes a further 8 strains from Pennsylvania that are monosporic derivatives of previously studied strains (Table S1; Sniegowski, Dombrowski, and Fingerman 2002).

To examine how strains from trees are related to other yeast strains, we constructed a reference panel of strains to represent published clades (1030 strains from 42 clades; Duan et al. 2018; Peter et al. 2018). These reference-panel strains were isolated from the human body (clinical), fermentation (e.g., wine and beer), baking, bioethanol, crops (e.g., sugar cane), decaying wood, fruit, flowers, insects, mushrooms, and water (e.g., sewers and oceans).

### 2.2 | Read Mapping and Base Calling

Paired-end and single-end genomic Illumina reads were downloaded from the European Bioinformatics Institute (<https://www.ebi.ac.uk/>) or generated in this study. Reads were mapped to the *S. cerevisiae* reference genome, S288c (SacCer\_Apr2011/sacCer3 from UCSC), using Burrows-Wheeler Aligner (bwa-mem, version 0.7.17; Li and Durbin 2009). We used SAMtools to sort, index, and compress bam files and generated a consensus sequence using the mpileup function with the -I option to exclude indels (version 1.6; Li et al. 2009). Next, we used the BCFtools call function with the -c option to generate a consensus sequence (version 1.9) (Li et al. 2009) and converted from vcf format to fastq format in SAMtools using the "vcfutils.pl vcf2fq" command. Lastly, base calls with a phred-scaled quality score of less than 40 were treated as missing data (calls were converted to "N") using seqtk seq -q 40 in SAMtools. In practice, this probably leads to an overall error rate lower than 1 in 10,000 because high-depth Illumina sequences usually yield consensus base calls that are not close to the Q40 cutoff. For example, after applying a Q40 cutoff to our past Illumina sequencing of *Candida albicans*, we observed an error rate below 1 in 100,000 (<169 errors in 14 Mbp; Bensasson et al. 2019), and use of the Q40 threshold on Sanger sequence from *S. cerevisiae* yielded a similarly low error rate (<1 in 180,000 bp; Bensasson 2011). None of the analyses discussed here would be affected by such a low error rate.

### 2.3 | Quality Filtering Steps

For population structure analysis, we removed genome sequences if they were from tree-sampled strains already represented in the dataset ( $N=27$ ), from strains with no geographical information ( $N=1$ ), or if they had an average read

depth below 30 $\times$  ( $N=18$ ). Additionally, genome data were visualised to check for intraspecies cross-contamination using vcf2allelePlot.pl. (Bensasson et al. 2019; Scopel et al. 2021): a genome was removed ( $N=18$ ) if some reads produced unexpected SNP calls at a frequency over 1%. For the remaining strains sampled from trees, we estimated genome-wide levels of heterozygosity using vcf2allelePlot.pl. (Bensasson et al. 2019), which estimates the number of heterozygous point substitutions divided by the length of the high-quality genome sequence (phred-scaled quality score over 40). Most strains from trees are homozygous (Figure S1). We removed 35 tree strains that are heterozygous (heterozygosity over 0.001) because they are difficult to represent in downstream phylogenetic analyses and may be interclade hybrids. After filtering, 236 tree-sampled strains remained for population genetic analyses (Table S1 and Figure S2).

For the reference panel strains, we applied the same quality filtering steps and examined levels of genome-wide heterozygosity, removing 16 out of 42 published clades because all individuals were heterozygous (Figures S1 and S2).

## 2.4 | Population Structure and Genetic Admixture

Whole-genome alignments were generated by concatenating the alignments for all 16 chromosomes into a single multiple-alignment file. Strains from trees were compared to reference strains after random selection of three strains per clade from the 26 published clades that remained after quality filtering (Table S2). One strain (BJ6) was randomly assigned to both the CHN-IV and Far East Asia clades ( $N=77$  strains). Ambiguity codes or lowercase base calls were converted to N's, and ends were filled to align to the same length. A neighbour-joining tree from genetic distances estimated by pairwise comparison of all genome sequences was constructed using MEGA-CC (version 10.0.5; Kumar et al. 2012). We used a Tamura-Nei substitution model (Tamura and Nei 1993) with a gamma distribution and 100 bootstrap replicates. Gaps or missing data were discarded from each pairwise sequence comparison. For visualisation, the neighbour-joining tree was rotated using ape (version 5.6.2; Paradis and Schliep 2018) and further visualised using ggtree (version 3.4.4; Xu et al. 2021; Yu et al. 2017, 2018).

Population structure and individual ancestry were estimated from SNP allele frequencies using ADMIXTURE (version 1.3.0; Alexander, Novembre, and Lange 2009). Genome data for all strains was merged into a single alignment in variant call format (vcf) using BCFtools, and mitochondrial DNA was removed. Non-variant sites were filtered out using the min-ac 1 function in BCFtools, which retains variants with at least one non-reference allele. Low-quality reads with a Phred-scaled quality score under 40 were removed using the minQ option in VCFtools (version 0.1.16; Danecek et al. 2011). Then we converted the single alignment vcf file to text-formatted and binary files using PLINK (version 1.9; Purcell et al. 2007) for downstream analysis with ADMIXTURE. We assigned strains to distinct populations or genetic clusters ( $K$ ) through repeated runs of ADMIXTURE. Runs assumed different numbers of genetic clusters from 4 to 40 with five replicates per  $K$ . We selected the run with the highest log likelihood value for each  $K$  and

visualised population structure across different  $K$ 's (Figure S3). We used CLUMPAK, specifically 'Distruct,' to align ancestry proportions (Q matrices) across different values of  $K$  (Kopelman et al. 2015). ADMIXTURE results were visualised as stacked bar plots using the pophelper R package (version 3.2.1; Francis 2017; Figure S4). Distinct genetic clusters were verified if they showed monophyletic clades with at least 95% bootstrap support in a neighbour-joining tree (Figure 1A and Table S2). We selected the run with the highest number of verified clusters or clades based on monophyletic groups in the neighbour-joining phylogeny and whether strains grouped by geography.

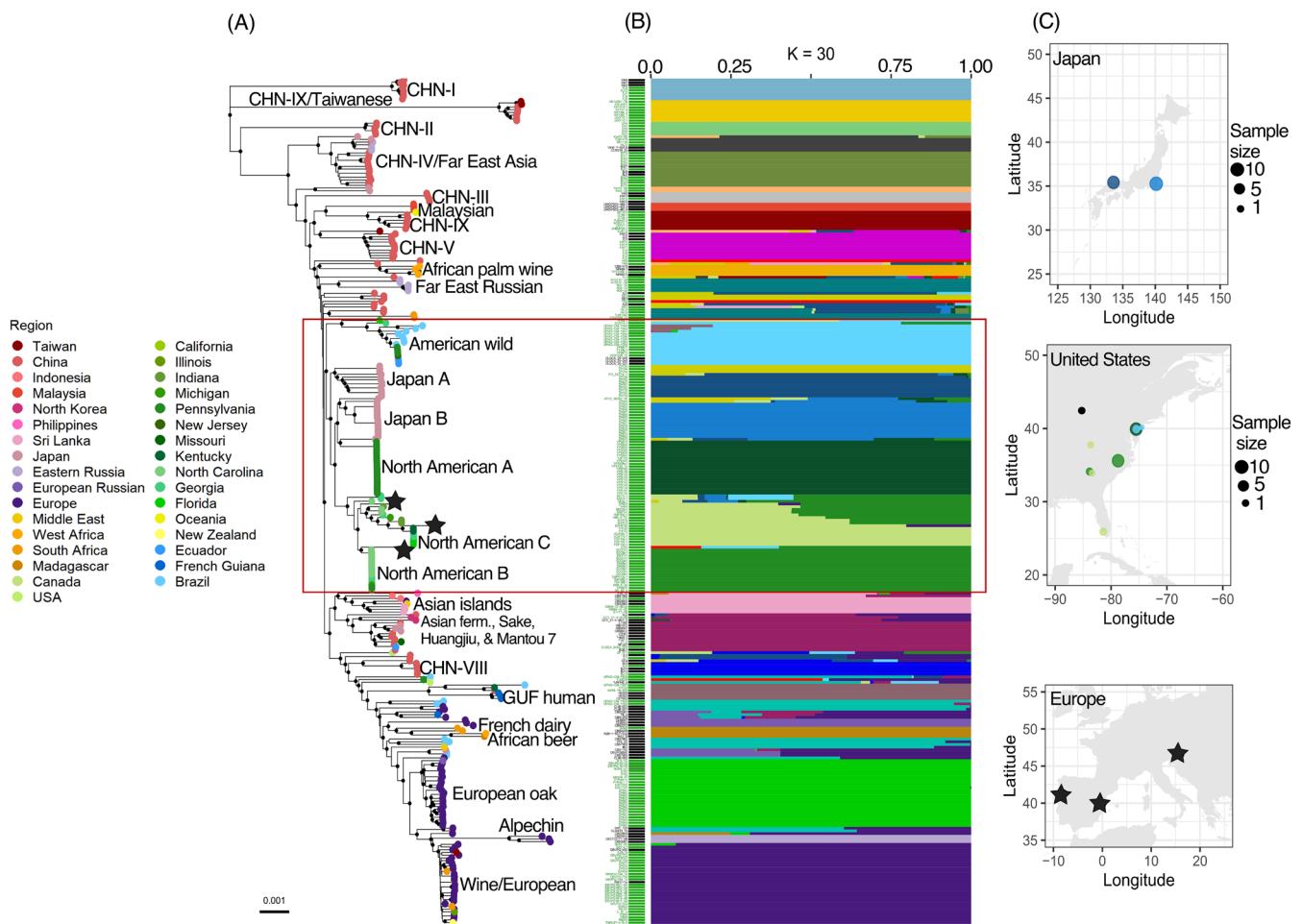
Phylogenomic relationships among strains were further examined using a maximum likelihood tree after excluding strains that showed possible recent genetic admixture when  $K=30$  (Figure S5 and Table S2). Admixed strains were defined as individuals whose percent ancestry from a single population is less than 90% in ADMIXTURE results for tree-sampled and reference panel strains (Table S2). It is possible, however, that some of the mixed ancestry invoked by the ADMIXTURE software could reflect ancestral polymorphisms or poorly sampled population subdivision. We included only three reference strains from CHN-IV and Far East Asia, which appear to be the same clade (Figure 1). We used a genome-wide alignment of SNPs mapped to the S228c reference genome to construct a phylogenomic tree with IQtree, ultrafast bootstrapping (version 1.6.12) (Minh, Nguyen, and Haeleter 2013; Nguyen et al. 2015), and a general time reversible model with a gamma distribution to estimate site heterogeneity. The maximum likelihood tree was visualised using ape and ggtree in R (version 4.2.2).

## 2.5 | Population Substructure Within Europe

Many *S. cerevisiae* strains have good quality genome data from European trees ( $N=51$  strains; Table S3) and have not been tested for population substructure. We used the same methods to analyse population substructure among European tree-sampled strains. ADMIXTURE was run by varying  $K$  from 2 to 8, and each  $K$  was repeated five times (Figures S6 and S7). Then, we constructed a phylogenomic tree using maximum likelihood estimation to infer phylogenetic relationships after removing one strain showing mixed ancestry at  $K=4$  (ZP541).

## 2.6 | *In Silico* Chromosome Painting

To identify genomic segments that could show gene flow between populations, we used a chromosome painting approach with faChrompaint.pl. and a 30kb window size (Bensasson et al. 2019). This *in silico* chromosome painting approach compares non-overlapping sliding windows of sequence to a panel of predefined clades. We identified 25 genetically distinct clades from population structure and phylogenomic analyses (Figures S4 and S5) and randomly selected three strains per clade to use as a "backbone" panel ( $N=75$  backbone strains; Figure S8 and Table S4). For a strain of interest, each 30 kb window was compared to a multiple sequence alignment of backbone strains, then "painted" a colour representing the clade of the most similar backbone sequence. Genomic regions that were diverged from all other backbone sequences (proportion of differing sites



**FIGURE 1** | Trees harbour numerous genetically distinct *S. cerevisiae* lineages with population substructure in North America and Japan. (A) Whole-genome neighbour-joining tree of 313 strains after excluding heterozygous strains (Table S1). Strains isolated from trees are shown with green text and bars, and reference panel strains with black. Black circles at nodes indicate bootstrap support > 95%. Coloured circles at the tips of the tree show geographical origin. All strains within the red box were isolated from America or Japan except for three European strains (black stars). (B) ADMIXTURE plot with  $K=30$  showing the cluster ancestry proportion for each strain. (C) Maps showing the geographic source of Japanese, North American, and European (black stars in A) tree-sampled strains from the American and Japanese lineages (red box in A). Circle sizes are based on square-root-transformed sample sizes. Japanese and North American strains are colour-coded by ancestry from ADMIXTURE plot. We removed North American and Japanese admixed strains (< 90% single-lineage ancestry) from maps for simplicity.

over 0.003) were painted white. This divergence threshold was decided based on within-clade pairwise comparisons for North American and European clades: over 90% of 30 kb windows showed a proportion of differing sites below 0.003 (< 0.3%), and most between-clade pairwise comparisons showed divergence over 0.003 (Figures S9 and S10). Exclusion of the most diverged regions reduces the chance that genetic similarity between lineages could be the result of incomplete lineage sorting. The genomic regions with low similarity to any other strain are the ones most likely to contain ancestral polymorphisms that have not yet become fixed. There were few such regions (Figure S8).

## 2.7 | Time Divergence Analysis

Time divergence analyses were performed on a single non-admixed locus (30–60 kb) from each backbone strain per chromosome (Figure S11). *In silico* chromosome painting analyses of backbone strains confirmed whether each backbone strain

matched its primary clade assignment from allele frequency analyses using ADMIXTURE (Figure S8). For downstream time divergence analyses, we removed three backbone strains with less than 50% primary clade assignment using a chromosome painting approach, five backbone strains with over 10% secondary clade assignment, and only included a single strain from the outgroup CHN-IX/Taiwanese clade (Table S5).

To estimate the timing of divergence events in the absence of purifying selection, we considered only nucleotide sites at which synonymous changes could occur. We extracted and concatenated 435 genes, with 11–39 genes for each 30–60 kb locus, after excluding 29 genes with introns, 42 genes that overlapped with other genes, and 4 genes with low-quality sequence for at least 50% of the alignment. Using MEGA-CC, we extracted 4-fold and 2-fold degenerate sites for each locus and concatenated them into 3507 to 19,016 bp alignments with 469–2245 variable and 211–1362 informative sites each. Numbers of informative sites were estimated using SeaView (Gouy, Guindon, and Gascuel 2010).

A neighbour-joining tree was constructed for each locus on each chromosome using the same methods previously mentioned, and the phylogeny was rooted using the CHN-IX/Taiwanese strain (EN14S01) as the outgroup. Neighbour-joining trees and multiple sequence alignments were used to estimate time trees per chromosome using MEGA-CC with the RelTime-ML option (Tamura, Tao, and Kumar 2018) using a Tamura–Nei substitution model (Tamura and Nei 1993) with the default setting to consider all sites for branch length calculations (Figure S12).

Using neighbour-joining distance trees and estimates of the *S. cerevisiae* mutation rate, we estimated the approximate timing at which modern tree-associated lineages (i) migrated out of Asia; (ii) into North America; and (iii) the separation of Wine/European and European oak lineages from trees. We used genetic distance to estimate the time (T) to the most recent common ancestor (MRCA) in generations per year:  $T_{\text{MRCA}} = k/\mu$  generations per year, where  $k$  is the genetic distance to the MRCA of strains in a clade and  $\mu$  is the point mutation rate per bp. We used the mutation rate of  $1.67 \times 10^{-10}$  point substitutions per site per generation, which was estimated from hundreds of point mutations after genome sequencing of mutation accumulation lines of diploid *S. cerevisiae* (Zhu et al. 2014). Under controlled laboratory settings at 30°C, wild diploid *S. cerevisiae* and its sister species *Saccharomyces paradoxus* have an average generation (doubling) time of 65 min in the presence of glucose and 125 min on nutrient-poor growth media (Kaya et al. 2021). Although glucose, fructose, and sucrose are present in the bark of trees that harbour yeast (Sampaio and Gonçalves 2008), it is likely less available than in the lab. In regions where *S. cerevisiae* were easily sampled from trees, historic temperatures were usually below 30°C (Table S8). With these considerations in mind, we made a rough estimate of the number of generations per year for yeast on trees, assuming: (i) a 90 min generation time because while the tree niche is likely less nutrient-rich than laboratory growth media, some sugars are available; (ii) 12 h of growth per day to account for no growth at lower nighttime temperatures; and (iii) no growth for six months of the year to account for low temperatures in winter. The resulting estimate is an average of 4 generations per day or 1460 generations per year. This is much lower than the number of generations possible at 30°C in nutrient-rich media in laboratory conditions: 22 per day; 8086 generations per year. It is also lower than the estimate of 2920 generations per year used to estimate the age of the wine-associated lineage (Fay and Benavides 2005), which may be less affected by cold winters and nutrient-poor conditions.

### 3 | Results

#### 3.1 | Tree Habitats Harbour Numerous Genetically Distinct Lineages

Using genome data for 236 strains from oaks and other trees, we examined population structure among wild *S. cerevisiae* in this niche. Phylogenetic analyses of tree-sampled strains and a reference panel of 77 strains from published clades (Duan et al. 2018; Peter et al. 2018) revealed several genetically distinct lineages that only occur on trees from China, Europe, Japan, North America, Russia, and Taiwan (Figure 1, Table S2, and Figure S5). These include previously studied wild lineages such

as ‘CHN-IX,’ ‘CHN-II,’ and ‘European oak’ (Duan et al. 2018; Peter et al. 2018), and more tree-associated lineages (see below).

Before applying filters to an initial sample of 328 strains isolated from trees with good genome data ( $>30\times$  read depth), we also observed numerous strains from clades that are usually associated with humans (Table S1). For example, there were 29 strains from the ‘Wine/European’ lineage from 21 field sites on 4 continents; 24 from the ‘Asian fermentation’ lineage from 19 sites on 5 continents, 12 ‘South African Beer’ clade strains from 3 South African field sites (Han et al. 2021); 6 strains from the ‘Mixed Origin’ clade associated with baking and clinical strains (Peter et al. 2018); and occasional strains from ‘African honey wine,’ ‘French dairy,’ and ‘African palm wine’ (Table S1). Some of these were too heterozygous for further study (Table S1, Figure S2).

#### 3.2 | Population Substructure in the Forest Niches of North America and Japan

Phylogenetic analyses further revealed population substructure within North America and Japan, where each region has multiple genetically distinct populations (Figure 1A). Analysis of allele frequencies using ADMIXTURE confirmed that there are several genetically distinct populations in North America and Japan (Figure 1B). More specifically, there are at least four tree-associated (wild) American *S. cerevisiae* lineages in the eastern United States (Figure 1C) that are well-supported across phylogenetic and ADMIXTURE analyses (Figure 1, Figures S4 and S5). Most wild strains from Pennsylvania are from a previously described lineage (Liti et al. 2009), and we refer to it here as ‘North American A’ (Tables S1 and S2). There are two more North American wild lineages: ‘North American B’ and ‘North American C’, covering a broader geographical region than the North American A lineage (Figure 1C). North American B strains are from Georgia, North Carolina, a singleton strain from Michigan, and a singleton strain from Pennsylvania (Tables S1 and S2). The North American C lineage occurs in strains from the southeastern United States (Kentucky, Georgia, and Florida). Lastly, wild strains from Ecuador and Brazil cluster with North American strains from Pennsylvania and New Jersey. We are coining this lineage as ‘American wild’ to reflect its Pan-American geography (Figure 1 and Figure S5) though some of these strains from trees were previously assigned to an ‘Ecuadorian’ clade (Peter et al. 2018) or a ‘Brazil 1’ clade (Barbosa et al. 2016).

Some tree strains have mixed genetic ancestry from deeply diverged American lineages (within the red box in Figure 1A,B), which is expected given their overlapping geographic distributions (Figure 1C). It is perhaps more surprising that these lineages have remained distinct despite gene flow. The strains we and others have sampled are not from primary forests, which have almost vanished from North America (Potapov et al. 2017). A possible explanation is therefore that these forest lineages came into contact only recently because they are newly arrived in their current locations.

In Japan, there are at least two wild *S. cerevisiae* populations: one from Hiruzen Highland, ‘Japan A,’ and one from Chiba Prefecture, ‘Japan B’ (Figure 1 and Figure S5). Japan A is diverged

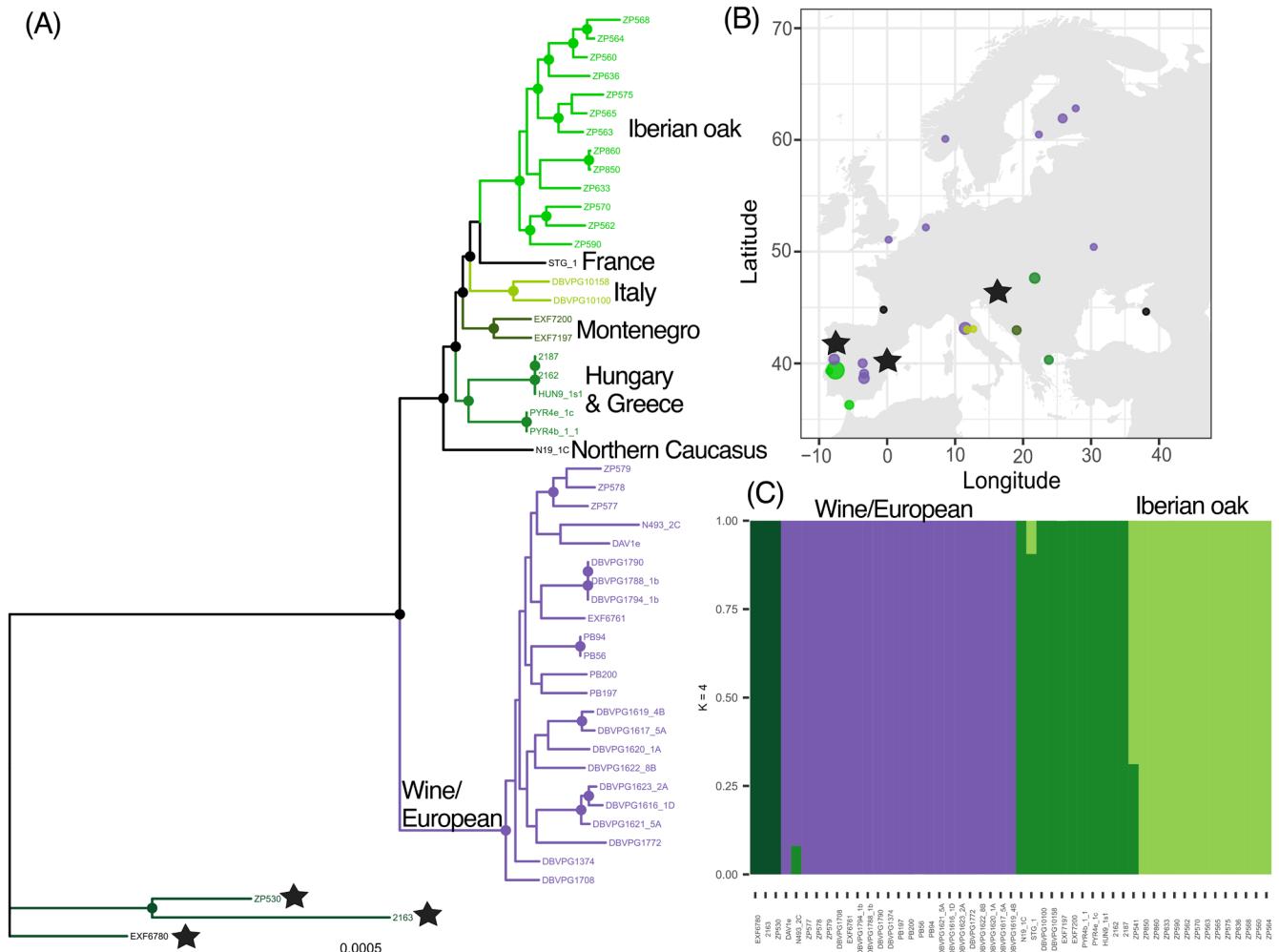
from all North American lineages and Japan B (Figure 1 and Figure S5). Japan B is most like the North American A lineage (Figure 1 and Figure S5).

### 3.3 | Fine-Scale Population Structure of Wild *S. cerevisiae* From Europe

European yeast from trees form two genetically distinct lineages: 'Wine/European' and 'European oak' (Figure 1 and Figure S5) that were previously known (Almeida et al. 2015; Tilakaratna and Bensasson 2017). Although the Wine/European lineage is usually recovered from wine fermentations, it sometimes occurs on trees, especially in vineyards (Gayevskiy, Lee, and Goddard 2016; Hyma and Fay 2013; Robinson, Pinharanda, and Bensasson 2016). Initial analyses suggest population substructure within the European oak lineage (Figure 1 and Figure S4). To better describe this substructure, we therefore ran separate ADMIXTURE and phylogenomic analyses for all strains that were isolated from European trees ( $N=51$ ; Table S3

and Figure 2). These analyses showed population substructure within the European oak lineage that correlates with geography (Figure 2, Figures S6 and S7). Specifically, there is evidence for five sub-lineages from: (i) Portugal and Spain, which we are coining 'Iberian oak' with support from both phylogenetic and ADMIXTURE analyses (Figure 2); and phylogenetic analyses suggest other distinct populations in (ii) Italy, (iii) Montenegro, (iv) Greece and Hungary, and (v) the North Caucasus (Figure 2). Additionally, several oak trees harbour strains from the 'Wine/European' winemaking lineage (Figure 2).

Although there is fine-scale population structure within the European oak lineage, even the divergence between the European oak and Wine/European lineages is smaller than the divergences seen in Taiwan (Lee et al. 2022), China, Japan, and North America (Figure 1A). The lack of deep divergence among European tree strains ( $N=53$ ) does not seem due to a lack of sampling because the number of good-quality genomes is not substantially higher for North America ( $N=66$ ), and in other continents, deeply diverged lineages can occur in close



**FIGURE 2** | Fine-scale population structure of wild *S. cerevisiae* from Europe. (A) Whole-genome maximum likelihood phylogenetic tree of 50 strains after excluding one admixed strain, ZP541, from (C). Black circles at nodes indicate bootstrap support  $> 95\%$ . Branches are colour-coded by geography or by ecology. Black stars at tree tips denote strains that genetically cluster with North American strains in Figure 1. (B) Map of Europe showing the geographic source of strains; circles are sized by the square-root transformed sample sizes and colour-coded by branch colours in the phylogenetic tree. Singleton strains from France and Northern Caucasus are coloured in black. Black stars denote strains that genetically cluster with North American strains in Figure 1. (C) ADMIXTURE plot when  $K=4$  to examine percent ancestry per individual strain.

proximity (Lee et al. 2022; Figure 1C). Instead, the shallow divergences within Europe and within the European oak lineage suggest that *S. cerevisiae* colonised Europe and European forests more recently than in other regions.

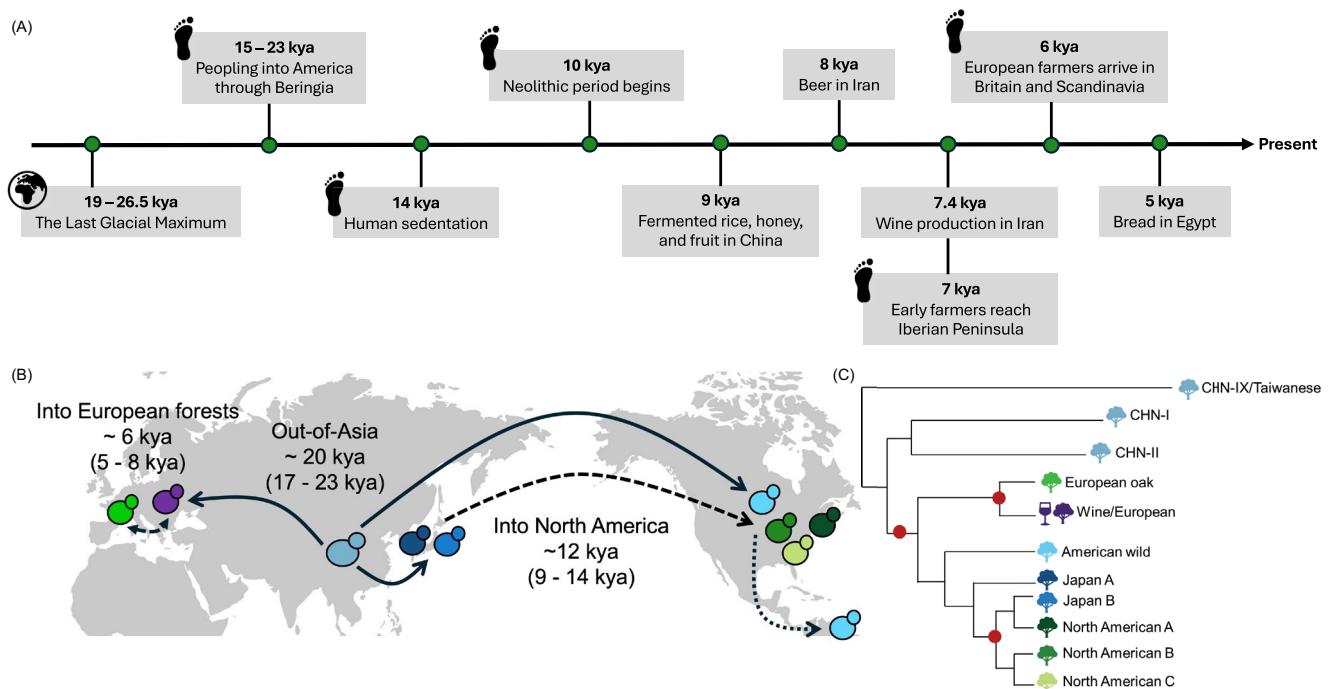
### 3.4 | Out-Of-Asia Migration of Forest Yeast Since the Last Glacial Maximum

We used a relative rate approach (Tamura, Tao, and Kumar 2018) to estimate the timing of lineage divergences that likely correspond to (i) migration events out of Asia, (ii) into North America, and (iii) the origin of a Wine lineage distinct from European forest lineages. After excluding admixture, phylogenetic analysis of individual loci (30–60 kb) from each chromosome reproduced most genetically distinct clades defined in this study (Table S6 and Figure S12). Using these phylogenetic trees and assuming a mutation rate of  $1.67 \times 10^{-10}$  point substitutions per site per generation (Zhu et al. 2014) and 4 generations per day (see Section 2), we generated rough estimates of divergence times. East Asia is the probable origin for *S. cerevisiae* (Duan et al. 2018; Han et al. 2021; Wang et al. 2012), and we estimate that non-Asian lineages diverged from those only found in Asia approximately 20 thousand years ago (mean = 19.8 kya, 95% CI 16.9–22.7 kya; Figure 3). Out of 16 individual loci, 14 loci showed a clade for European oak (Table S7 and Figure S12). Time divergence estimation for these loci suggests European oak and the domesticated Wine/European lineage diverged around 6 kya (mean = 6.1 kya, 95% CI 4.5–7.6 kya; Figure 3). Out of 16 loci, 9 loci clustered the 3 North American clades (A–C), sometimes with Japan B (5 loci) or Japan A lineages (2 loci). This mostly

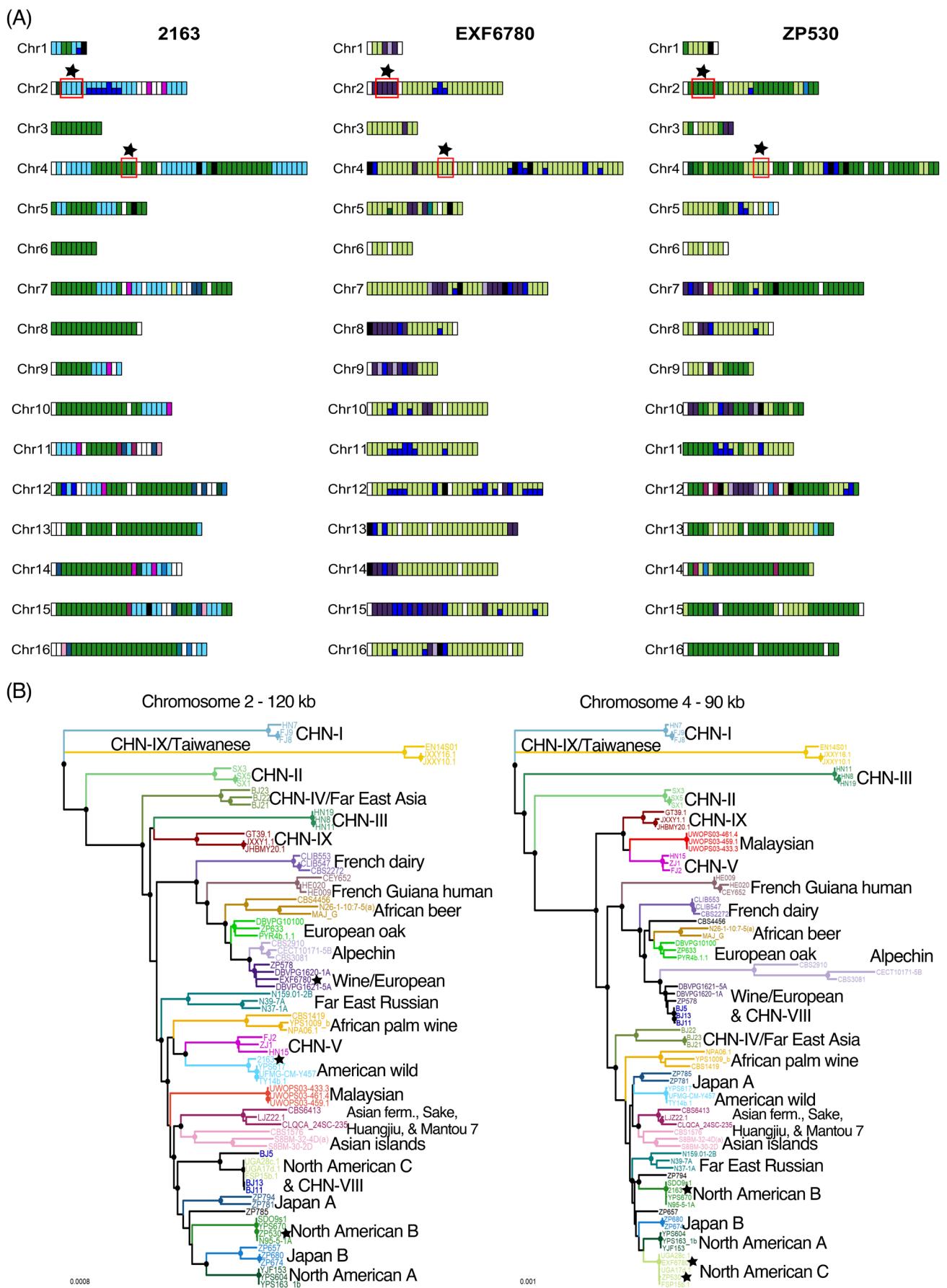
North American clade appears to diverge from others (Table S7 and Figure S12) around 12 kya (mean = 11.6 kya, 95% CI 9.4–13.7 kya; Figure 3). These estimates suggest global *S. cerevisiae* migrations occurred since the last glacial maximum (Figure 3).

### 3.5 | Occasional Strains in Europe Resemble Present-Day North American Lineages

There are three strains from European trees (ZP530, 2163, and EXF6780) that resemble North American strains (black stars in Figure 1) and differ from all other European lineages (Figure 2). These genomes are from two different investigations where (i) ZP530 was isolated from chestnut (*Castanea sativa*) from Marão, Campeã, Portugal (Almeida et al. 2015), (ii) EXF6780 was isolated from sessile oak (*Quercus petraea*) from Velike Lašče, Kobilna hill, Slovenia (Almeida et al. 2015), and (iii) 2163 was isolated from Portuguese oak (*Quercus faginea*) from Castellon, Spain (Peter et al. 2018). In the phylogenetic analysis, these strains differ from their most closely related clades (North American B and C in Figure 1A) and appear to show some genetic admixture (Figure 1B). Did these strains arrive on European trees as a result of ancient migration, or could their genetic distance from other North American strains be explained by recent admixture? To find out, we “painted” their chromosomes according to the clade of the most closely related strain (Figure 4A). EXF6780, ZP530, and 2163 were compared to our backbone phylogeny (Table S5), which revealed the strains are a mix of three lineages found in North America and that two strains have admixture from the lineage used to ferment grape wine (Figure 4).



**FIGURE 3 |** Out-of-Asia migration of forest yeast since the Last Ice Age. (A) Timeline showing early archaeological evidence of fermentation and human migration (dates from Clark et al. 2009; Marsit et al. 2017; Nielsen et al. 2017). (B) Map showing forest yeast migration events depicted with a red dot at nodes in (C) (i) out of Asia, (ii) into North America, and (iii) into European forests. Lineages are colour-coded by the lineages in the cladogram in (C). Dashed arrows indicate secondary migration events. (C) Cladogram showing phylogenetic relationships of lineages of interest for date estimation.



**FIGURE 4** | Legend on next page.

**FIGURE 4** | Occasional strains in Europe resemble present-day North American lineages. (A) Painted chromosomes of 2163, EXF6780, and ZP530 show admixture between multiple lineages. Genomic regions were “painted” based on the clade assignment of the most similar strain in 30 kb non-overlapping windows. Diverged regions were not coloured (white) and were defined as regions that differed by 0.003 from all other strains in the backbone phylogeny. Black-coloured regions indicate low coverage. Colours are as in (B): North American B is forest green, American wild is light blue, North American C is light green, Wine/European is dark purple, and CHN-VIII is blue. Genomic regions (90–120 kb) were selected for phylogenetic analysis (red boxes with black stars). (B) Neighbour-joining phylogenetic trees for two loci. Solid circles at nodes indicate bootstrap support >95%. Branches are colour-coded by clade. Phylogenetic analyses show that in the absence of admixture, 2163, EXF6780, and ZP530 (black stars) are very similar to strains from American wild, Wine/European, North American B, and C.

To assess genetic distances from modern North American lineages while accounting for admixture, we selected genomic regions from chromosomes 2 and 4 for phylogenetic analysis that did not show admixture from multiple clades (Figure 4A). Phylogenies with the effects of admixture removed in this way showed that strains 2163, EXF6780, and ZP530 seem almost identical to modern North American lineages across (Figure 4B). It therefore seems most likely that these ‘American European’ strains arrived very recently in Europe and that admixture among North American and Wine/European lineages (Figure 4) explains their genetic distance from modern strains in the whole-genome phylogeny (Figure 1A).

Intriguingly, all three of these American European strains show some regions with close sequence similarity to CHN-VIII (blue windows in Figure 4A). One potential explanation is that the CHN-VIII lineage also has some recent domesticated Wine/European ancestry because all reference Wine/European strains resemble CHN-VIII in many genomic regions (Figure S8). Chromosome painting shows that one of the 3 strains representing CHN-VIII is very similar to reference strains from North American C, American Wild, and Wine/European lineages (>10%, Figure S8). Locus-by-locus phylogenetic analysis (Figure S11) suggests that the remaining two CHN-VIII strains are diverged from the European oak lineage, yet indistinguishable in large genomic regions from the Wine/European lineage (chromosomes 4–7 and 11) and North American C (chromosomes 2 and 12). CHN-VIII genomes are also similar to the American wild lineage (chromosomes 3 and 15) and show regions diverged from any other lineages (chromosomes 1, 8, 9, and 16). The degree of similarity between modern Wine/European and North American C strains suggests very recent gene flow from Europe. Consistent with this proposal, CHN-VIII occurs in apple orchards and secondary forests near Beijing, and Wine/European strains have also been isolated from orchards in China (Duan et al. 2018; Wang et al. 2012).

## 4 | Discussion

### 4.1 | Genetic Isolation in Forest Niches

Before molecular methods showed that *S. cerevisiae* lives on trees (Naumov et al. 1992; Sniegowski, Dombrowski, and Fingerman 2002), people thought it lived only with humans and not in natural environments (Vaughan-Martini and Martini 1995). This resembles thinking for *Candida* pathogenic species before their recent discovery on trees and other plants (Bensasson et al. 2019; Opulente et al. 2019). Yet our results suggest that *S. cerevisiae* recently colonised woodlands many times and lived

there in relative isolation from humans. Phylogeographic analysis of forest *S. cerevisiae* populations shows that strains are recognisably from Iberian, French, or Italian trees, with further lineages occurring in Eastern Europe (Figures 1 and 2). There are at least four forest lineages in North America with regional differences; for example, North American C occurs in the southern USA, and North American A in Pennsylvania. The higher divergence seen among North American lineages (A–C) suggests that their most recent common ancestor must considerably predate the common ancestor of European *S. cerevisiae* and therefore an earlier arrival (see below). The genetic similarity of North American and Japanese lineages (A and B) suggests that the migration into North America, after the earlier arrival of American Wild, came from a different Asian source, perhaps close to Japan. There is also genetic (mtDNA) and archaeological evidence suggesting shared ancestry among the people of China, Japan, and the Americas, suggesting a human migration route from northern coastal China into the Americas (Li et al. 2023).

Past analyses also show high population structure in *S. cerevisiae* from the primaeval forests of East Asia (Wang et al. 2012; Lee et al. 2022) and in its sister species, *S. paradoxus* (Hénault et al. 2019; Leducq et al. 2014). *Saccharomyces* yeast are probably not usually air-dispersed (Mortimer 2000); therefore, it is not surprising that they show more population structure than other fungal microbes. Our observations for the tree niche contrast with the broader distribution of domesticated and fruit-associated lineages (Almeida et al. 2015; Duan et al. 2018; Gallone et al. 2016; Gonçalves et al. 2016; Lee et al. 2022; Legras et al. 2007, 2018; Peter et al. 2018) and are consistent with the proposal that animal-assisted long-distance migration is relatively rare in forests (Magwene et al. 2011; Tilakaratna and Bensasson 2017).

The evidence for isolated forest populations includes phylogenetic analyses using whole genomes, single chromosomes, and single locus data (Figures 1A, 2A, 4B, Figures S5 and S12) in addition to analyses of allele frequencies (Figures 1B and 2C, Figure S7). Monophyletic tree-associated clades were reproducible across most chromosomes at the tips of phylogenetic trees, suggesting the fixation of many alleles for each lineage (Figures 1A, 2A, 4B, Figures S5 and S12). Why might *S. cerevisiae* show many phylogenetically distinct lineages within continents? According to past estimates, *S. cerevisiae* reproduces sexually only once in hundreds or thousands of generations (Magwene et al. 2011; Lee et al. 2022), and the same is true for *S. paradoxus* (Tsai et al. 2008). Even when meiosis does occur, *S. paradoxus* are almost always selfing (99% of sexual cycles, Tsai et al. 2008). Asexual reproduction of *S. cerevisiae* might lead to

population bottlenecks and the local fixation of alleles by genetic drift.

## 4.2 | Global Spread of *S. cerevisiae* Forest Populations Since the Last Glacial Maximum

Mutation rate estimates for *S. cerevisiae* applied to phylogenetic analyses suggest that the expansion of forest *S. cerevisiae* out of Asia likely occurred in the last 20,000 years (Figure 3 and Figure S12). Divergence among the forest lineages of Europe and America is less deep than among the lineages of Asia (Figure 1A), so our analyses support an East Asian species origin (Wang et al. 2012; Lee et al. 2022) with the caveat that better sampling in other regions such as Central Asia or Central Africa could also reveal high genetic diversity. Averages estimated from phylogenies of 16 loci suggest that *S. cerevisiae* lineages migrated out of Asia around 17–23 kya (Figure 3 and Figure S12), which was when climate started warming after the last glacial maximum (Clark et al. 2009). The forest lineages that first diverged from East Asian populations include those occurring in South America (American wild, French Guiana human; Figure S12). The fine-scale population structure within North American forests arose more recently; since the divergence of North American lineages (A–C) from most Asian lineages around 9–14 kya. The population structure occurring on European trees arose since these lineages diverged from the Wine lineage approximately 5–8 kya.

The timing of these yeast migrations seems to roughly coincide with human migration into America (15–23 kya), sedentarization (~14 kya), widespread settlements in the Americas (12.6–13 kya), the origins of agriculture (~10 kya), fermentation practices in Asia (~9 kya), European agriculture (6–7 kya), and European winemaking (~4 kya; Figure 3A; Marsit et al. 2017; McGovern 2003; Nielsen et al. 2017). It is therefore possible that humans or their commensals carried yeast with their food as they moved around the world. The alternative, that yeast migrated naturally across the globe as the climate warmed, seems less likely for multiple reasons: (i) *S. cerevisiae* are only rarely isolated from trees in northern Europe, and upon further genetic investigation, northern populations appear feral or domesticated (the Wine/European lineage or admixed; Robinson, Pinharanda, and Bensasson 2016). Tree-associated genetic lineages of *S. cerevisiae* have not been reported from northern cool temperate regions such as Canada or eastern Russia, despite documentation of other *Saccharomyces* species (Charron, Leducq, and Landry 2014; Naumov et al. 2000). Indeed, tree-associated genetic lineages may have a mostly subtropical or tropical distribution (Robinson, Pinharanda, and Bensasson 2016). A natural expansion through the cold climate of the Bering land bridge therefore seems unlikely. (ii) Such an expansion seems more likely for *S. paradoxus*, which inhabits northern forests in Canada, northern Europe, and Siberia, yet that species shows much greater genetic isolation between America, Asia, and Europe (Liti et al. 2009), perhaps because it does not live in association with humans. (iii) *S. cerevisiae* differs from *S. paradoxus* in that it is common on fruit. Humans might have inadvertently carried *S. cerevisiae* with their food, vessels, or the insects that travelled with them. (iv) Human-associated migration can explain a relatively recent

origin of European lineages that is roughly coincident with the expansion of agriculture into Europe (5–8 kya), the much later spread to New Zealand in the last 1000 years, and their concentration near New Zealand vineyards (Gayevskiy, Lee, and Goddard 2016). A less likely possibility is that *S. cerevisiae* was naturally dispersed by insects capable of migrating from warm regions in Asia to North America. Monarch butterflies have dispersed from America to Australia via Pacific islands since the last glacial maximum (Zhan et al. 2014). Such long-distance dispersal is rare, however, even for known long-distance dispersers like locusts, which do not have cosmopolitan distributions (Lovejoy et al. 2005) like those of humans or other human commensal organisms.

Our time estimates are based on the well-studied mutation rate of *S. cerevisiae*. The estimate we use ( $1.67 \times 10^{-10}$  per base per generation) is the most accurate, from hundreds of point mutations (867) observed in 145 genome sequences from diploid mutation accumulation lines (Zhu et al. 2014). Earlier mutation rate estimates did not differ greatly:  $2.9 \times 10^{-10}$  from mutations accumulated in a different diploid background (Nishant et al. 2010),  $3.3 \times 10^{-10}$  in haploids (Lynch et al. 2008), or  $1.84 \times 10^{-10}$  from reporter assays in haploid strains at individual loci (Drake 1991; Fay and Benavides 2005). The number of generations occurring in natural forest environments is more difficult to measure (Mozzachiodi et al. 2022). As in past analysis of wine yeast by Fay and Benavides (ca. 2920 generations per year; 2005), we assume a lower growth rate than in the laboratory and only 12 h of growth per day. The rate we use for trees (ca. 1460 generations per year) assumes slower growth because of fewer nutrients on trees and no growth for 6 months of the year because of low temperatures (see Methods). For some parts of the species range, such as Florida and Georgia, generation times could be underestimated because some of the maximum temperatures in the coldest 6 months (18°C–29°C) and minimum nighttime temperatures in the hottest months (16°C–23°C) are also high enough for yeast growth (Table S8; Sweeney, Kuehne, and Sniegowski 2004). While the fossil record is excellent for estimating divergences among yeast families or genera (Douzery et al. 2004; Marcet-Houben and Gabaldón 2015; Shen et al. 2018), the most recent fossil for ascomycotans dates to 417 million years ago (Douzery et al. 2004) and therefore may not be accurate for intraspecies divergences. Our timings suggest an older divergence (17–23 kya) than that of Fay and Benavides (2005) for the split of wine and East Asian sake strains (11.9 kya), and are consistent with past estimates for divergence among wine strains (3.7 kya; Fay and Benavides 2005), the split between European oak and wine 10.3–1.3 kya (Almeida et al. 2015), and the arrival of *S. cerevisiae* in New Zealand less than 1 kya (Gayevskiy, Lee, and Goddard 2016).

## 4.3 | Ongoing Migration Between Human and Tree Environments

Not all *S. cerevisiae* strains on trees are from tree-associated lineages. Strains from the European grape wine lineage and other human-associated lineages also live on trees (Gayevskiy, Lee, and Goddard 2016; Hyma and Fay 2013; Robinson, Pinharanda, and Bensasson 2016). Indeed, there were at least 10 migration events from Europe to New Zealand trees that happened

in the last 1000 years since humans arrived in New Zealand (Gayevskiy, Lee, and Goddard 2016). Here we observe strains on trees from clades connected with grape and African wines, Asian fermentations, brewing, baking, and clinical strains (Table S1, Figure S2), which suggests transmission between humans and trees is ongoing.

#### 4.4 | Footprints of Human Activity in the Genomes of European Tree Strains

Three strains from trees in Portugal, Spain, and Slovenia had genome sequences that were predominantly from North American forest clades. Chromosome painting of their genomes shows these strains must be descended from three different transatlantic migrants (Figure 4A). These migrations from North America must have been recent because analysis of large (90 and 120 kb) loci shows genomic tracts that look typical of current North American B, C, and American wild lineages (Figure 4B). None of these American European strains resemble strains from the North American A lineage, which only occurred in Pennsylvania (Table S1). Interestingly, strains from Portugal and Slovenia (ZP530 and EXF6780) resemble the North American C lineage (Figure 4), which we only observe in the southern USA (Figure 1C).

A potential explanation for the presence of North American tree-associated lineages in Europe is the human response to the Great French Wine Blight. In the 1850s, humans accidentally introduced an insect pest, *Phylloxera*, from North America to Europe that destroyed most European vineyards. Native American vines are naturally resistant to *Phylloxera*. The European wine industry was rescued by the mass import of vines from the southern USA to Europe, and from the late 1800s to the present day, most European vines are grafted onto resistant North American grapevine rootstock (Campbell 2004). The imported American grapevines could have harboured North American yeast. In support of an explanation, Portuguese and Slovenian American strains show admixture from the Wine/European lineage, suggesting recent association with vineyards, whereas European oak and North American forest lineages rarely show admixture from the wine lineage (Figure 4). The puzzling occurrence of long stretches of Wine/European and North American DNA in yeast strains from orchards and secondary forests near Beijing in China could then be explained by very recent human-assisted migration from Europe.

#### 5 | Conclusion

In summary, our analyses show forests harbour many isolated *S. cerevisiae* populations that are distinct from human-associated lineages. The phylogeographic structure of tree-associated lineages implies that migrants from humans rarely establish in forest niches. Yet even rare events can shape the distribution of a species. The postglacial spread of forest *S. cerevisiae* out of Asia and into North America and Europe suggests that this substantial impact was driven by people. Consistent with this, we also observe footprints of ongoing human-assisted movement of forest yeast. Fungal microbes introduced into forests can transform landscapes when they are mutualists or parasites

(Hoeksema et al. 2020; Averill et al. 2022). For intimate human commensals and occasional pathogens, such as *S. cerevisiae* and *Candida* sp., it seems important to consider their evolution in non-human environments—especially since environmental fungal microbes may adapt to fungicide use or rising temperatures (Garcia-Solache and Casadevall 2010; Leducq et al. 2014; Kang et al. 2022; Lockhart, Chowdhary, and Gold 2023).

#### Author Contributions

The research was conceptualised and designed by J.J.P. and D.B. A.K.W., J.J.P., and E.F.C.S. performed DNA extractions for genome sequencing. J.J.P. and E.F.C.S. developed bioinformatic pipelines and obtained and curated the data. J.J.P. performed analyses and data visualisation. J.J.P. and D.B. wrote the paper with input from E.F.C.S. and A.K.W.

#### Acknowledgements

This work was funded by an NSF LSAMP Bridge to the Doctorate Fellowship Program award (grant number: 1702361) to JJP; by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program (grant no. GT13544) awarded to JJP and DB; and by a National Science Foundation grant (IOS, no. 1946046) awarded to DB. For some of the new genomes used, we would like to thank the Florida State Parks Service, the Duke Forest Teaching and Research Laboratory, Tyler Arboretum, Thompson Mills Forest, Skidaway Institute of Oceanography, and the University of Georgia for their support of fieldwork, and Diana Ambrocio, Domenic Won, and Linda Habersham for strain isolation. Past and current members of the Bensasson lab provided helpful feedback. We would like to acknowledge that most of the research performed for this study was on Cherokee, Yuchi, and Muscogee indigenous lands.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

Short-read genome data are available in the NCBI-SRA under project number PRJNA1090965. Consensus genome sequences generated by mapping to the sacCer3 reference, alignments, and PLINK files used in ADMIXTURE analyses are available on Dryad under <https://doi.org/10.5061/dryad.pnvx0k6zq>. Benefits from this research include this sharing of our data and results on public databases and the sharing of the yeast strains isolated for this work, which are available from our lab or public yeast collections.

#### References

- Alexander, D. H., J. Novembre, and K. Lange. 2009. “Fast Model-Based Estimation of Ancestry in Unrelated Individuals.” *Genome Research* 19, no. 9: 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Almeida, P., R. Barbosa, P. Zalar, et al. 2015. “A Population Genomics Insight Into the Mediterranean Origins of Wine Yeast Domestication.” *Molecular Ecology* 24, no. 21: 5412–5427. <https://doi.org/10.1111/mec.13341>.
- Averill, C., M. A. Anthony, P. Baldrian, et al. 2022. “Defending Earth’s Terrestrial Microbiome.” *Nature Microbiology* 7, no. 11: 1717–1725. <https://doi.org/10.1038/s41564-022-01228-3>.
- Barbosa, R., P. Almeida, S. V. B. Safar, et al. 2016. “Evidence of Natural Hybridization in Brazilian Wild Lineages of *Saccharomyces cerevisiae*.” *Genome Biology and Evolution* 8, no. 2: 317–329. <https://doi.org/10.1093/gbe/evv263>.

Bensasson, D. 2011. "Evidence for a High Mutation Rate at Rapidly Evolving Yeast Centromeres." *BMC Evolutionary Biology* 11: 211. <https://doi.org/10.1186/1471-2148-11-211>.

Bensasson, D., J. Dicks, J. M. Ludwig, et al. 2019. "Diverse Lineages of *Candida albicans* Live on Old Oaks." *Genetics* 211, no. 1: 277–288. <https://doi.org/10.1534/genetics.118.301482>.

Bergström, A., J. T. Simpson, F. Salinas, et al. 2014. "A High-Definition View of Functional Genetic Variation From Natural Yeast Genomes." *Molecular Biology and Evolution* 31, no. 4: 872–888. <https://doi.org/10.1093/molbev/msu037>.

Boynton, P. J., and D. Greig. 2014. "The Ecology and Evolution of Non-Domesticated *Saccharomyces* Species." *Yeast* 31, no. 12: 449–462. <https://doi.org/10.1002/yea.3040>.

Campbell, C. 2004. *Phylloxera: How Wine Was Saved for the World*. London: HarperCollins Publishers. <http://archive.org/details/phylloxerawin0000camp>.

Charron, G., J. Leducq, and C. R. Landry. 2014. "Chromosomal Variation Segregates Within Incipient Species and Correlates With Reproductive Isolation." *Molecular Ecology* 23, no. 17: 4362–4372. <https://doi.org/10.1111/mec.12864>.

Clark, P. U., A. S. Dyke, J. D. Shakun, et al. 2009. "The Last Glacial Maximum." *Science* 325, no. 5941: 710–714. <https://doi.org/10.1126/science.1172873>.

Cromie, G. A., K. E. Hyma, C. L. Ludlow, et al. 2013. "Genomic Sequence Diversity and Population Structure of *Saccharomyces cerevisiae* Assessed by RAD-Seq." *G3: Genes, Genomes, Genetics* 3, no. 12: 2163–2171. <https://doi.org/10.1534/g3.113.007492>.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27, no. 15: 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.

Diezmann, S., and F. S. Dietrich. 2009. "Saccharomyces cerevisiae: Population Divergence and Resistance to Oxidative Stress in Clinical, Domesticated and Wild Isolates." *PLoS One* 4, no. 4: e5317. <https://doi.org/10.1371/journal.pone.0005317>.

Douzery, E. J. P., E. A. Snell, E. Baptiste, F. Delsuc, and H. Philippe. 2004. "The Timing of Eukaryotic Evolution: Does a Relaxed Molecular Clock Reconcile Proteins and Fossils?" *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 43: 15386–15391. <https://doi.org/10.1073/pnas.0403984101>.

Drake, J. W. 1991. "A Constant Rate of Spontaneous Mutation in DNA-Based Microbes." *Proceedings of the National Academy of Sciences of the United States of America* 88, no. 16: 7160–7164. <https://doi.org/10.1073/pnas.88.16.7160>.

Duan, S.-F., P.-J. Han, Q.-M. Wang, et al. 2018. "The Origin and Adaptive Evolution of Domesticated Populations of Yeast From Far East Asia." *Nature Communications* 9, no. 1: 2690. <https://doi.org/10.1038/s41467-018-05106-7>.

Fay, J. C., and J. A. Benavides. 2005. "Evidence for Domesticated and Wild Populations of *Saccharomyces cerevisiae*." *PLoS Genetics* 1, no. 1: e5. <https://doi.org/10.1371/journal.pgen.0010005>.

Fay, J. C., P. Liu, G. T. Ong, et al. 2019. "A Polyploid Admixed Origin of Beer Yeasts Derived From European and Asian Wine Populations." *PLoS Biology* 17, no. 3: e3000147. <https://doi.org/10.1371/journal.pbio.3000147>.

Francis, R. M. 2017. "pophelper: An R Package and Web App to Analyse and Visualize Population Structure." *Molecular Ecology Resources* 17, no. 1: 27–32. <https://doi.org/10.1111/1755-0998.12509>.

Gallone, B., J. Steensels, T. Prahl, et al. 2016. "Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts." *Cell* 166, no. 6: 1397–1410.e16. <https://doi.org/10.1016/j.cell.2016.08.020>.

Garcia-Solache, M. A., and A. Casadevall. 2010. "Global Warming Will Bring New Fungal Diseases for Mammals." *mBio* 1, no. 1: e00061-10. <https://doi.org/10.1128/mbio.00061-10>.

Gayevskiy, V., S. Lee, and M. R. Goddard. 2016. "European Derived *Saccharomyces cerevisiae* Colonisation of New Zealand Vineyards Aided by Humans." *FEMS Yeast Research* 16, no. 7: fow091. <https://doi.org/10.1093/femsyr/fow091>.

Gonçalves, M., A. Pontes, P. Almeida, et al. 2016. "Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts." *Current Biology* 26, no. 20: 2750–2761. <https://doi.org/10.1016/j.cub.2016.08.040>.

Gouy, M., S. Guindon, and O. Gascuel. 2010. "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." *Molecular Biology and Evolution* 27, no. 2: 221–224.

Han, D.-Y., P.-J. Han, K. Rumbold, et al. 2021. "Adaptive Gene Content and Allele Distribution Variations in the Wild and Domesticated Populations of *Saccharomyces cerevisiae*." *Frontiers in Microbiology* 12: 631250. <https://doi.org/10.3389/fmicb.2021.631250>.

Hénault, M., C. Eberlein, G. Charron, et al. 2019. "Yeast Population Genomics Goes Wild: The Case of *Saccharomyces paradoxus*." In *Population Genomics: Microorganisms*, edited by M. F. Polz and O. P. Rajora, 207–230. Cham: Springer International Publishing. [https://doi.org/10.1007/13836\\_2017\\_4](https://doi.org/10.1007/13836_2017_4).

Hoeksema, J. D., C. Averill, J. M. Bhatnagar, et al. 2020. "Ectomycorrhizal Plant-Fungal Co-Invasions as Natural Experiments for Connecting Plant and Fungal Traits to Their Ecosystem Consequences." *Frontiers in Forests and Global Change* 3: 84. <https://doi.org/10.3389/ffgc.2020.00084>.

Hyma, K. E., and J. C. Fay. 2013. "Mixing of Vineyard and Oak-Tree Ecotypes of *Saccharomyces cerevisiae* in North American Vineyards." *Molecular Ecology* 22, no. 11: 2917–2930. <https://doi.org/10.1111/mec.12155>.

Kang, S. E., L. G. Sumabat, T. Melie, B. Mangum, M. Momany, and M. T. Brewer. 2022. "Evidence for the Agricultural Origin of Resistance to Multiple Antimicrobials in *Aspergillus Fumigatus*, a Fungal Pathogen of Humans." *G3: Genes, Genomes, Genetics* 12, no. 2: jkab427. <https://doi.org/10.1093/g3journal/jkab427>.

Kaya, A., C. Z. J. Phua, M. Lee, et al. 2021. "Evolution of Natural Lifespan Variation and Molecular Strategies of Extended Lifespan in Yeast." *eLife* 10: e64860. <https://doi.org/10.7554/eLife.64860>.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. "Clumpak: A Program for Identifying Clustering Modes and Packaging Population Structure Inferences Across K." *Molecular Ecology Resources* 15, no. 5: 1179–1191. <https://doi.org/10.1111/1755-0998.12387>.

Kumar, S., G. Stecher, D. Peterson, and K. Tamura. 2012. "MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis." *Bioinformatics* 28, no. 20: 2685–2686. <https://doi.org/10.1093/bioinformatics/bts507>.

Leducq, J.-B., G. Charron, P. Samani, et al. 2014. "Local Climatic Adaptation in a Widespread Microorganism." *Proceedings of the Royal Society B: Biological Sciences* 281, no. 1777: 20132472. <https://doi.org/10.1098/rspb.2013.2472>.

Lee, T. J., Y.-C. Liu, W.-A. Liu, et al. 2022. "Extensive Sampling of *Saccharomyces Cerevisiae* in Taiwan Reveals Ecology and Evolution of Predomesticated Lineages." *Genome Research* 32, no. 5: 864–877. <https://doi.org/10.1101/gr.276286.121>.

Legras, J., D. Merdinoglu, J. Cornuet, and F. Karst. 2007. "Bread, Beer and Wine: *Saccharomyces Cerevisiae* Diversity Reflects Human History." *Molecular Ecology* 16, no. 10: 2091–2102. <https://doi.org/10.1111/j.1365-294x.2007.03266.x>.

Legras, J.-L., V. Galeote, F. Bigey, et al. 2018. "Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication." *Molecular Biology and Evolution* 35, no. 7: 1712–1727. <https://doi.org/10.1093/molbev/msy066>.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.

Li, Y.-C., Z.-L. Gao, K.-J. Liu, et al. 2023. "Mitogenome Evidence Shows Two Radiation Events and Dispersals of Matrilineal Ancestry From Northern Coastal China to the Americas and Japan." *Cell Reports* 42, no. 5: 112413. <https://doi.org/10.1016/j.celrep.2023.112413>.

Liti, G., D. M. Carter, A. M. Moses, et al. 2009. "Population Genomics of Domestic and Wild Yeasts." *Nature* 458, no. 7236: 337–341. <https://doi.org/10.1038/nature07743>.

Lockhart, S. R., A. Chowdhary, and J. A. W. Gold. 2023. "The Rapid Emergence of Antifungal-Resistant Human-Pathogenic Fungi." *Nature Reviews Microbiology* 21, no. 12: 818–832. <https://doi.org/10.1038/s41579-023-00960-9>.

Lovejoy, N. R., S. P. Mullen, G. A. Sword, et al. 2005. "Ancient Trans-Atlantic Flight Explains Locust Biogeography: Molecular Phylogenetics of *Schistocerca*." *Proceedings of the Royal Society B: Biological Sciences* 273, no. 1588: 767–774. <https://doi.org/10.1098/rspb.2005.3381>.

Lynch, M., W. Sung, K. Morris, et al. 2008. "A Genome-Wide View of the Spectrum of Spontaneous Mutations in Yeast." *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 27: 9272–9277. <https://doi.org/10.1073/pnas.0803466105>.

Magwene, P. M., Ö. Kayıkçı, J. A. Granek, J. M. Reininga, Z. Scholl, and D. Murray. 2011. "Outcrossing, Mitotic Recombination, and Life-History Trade-Offs Shape Genome Evolution in *Saccharomyces cerevisiae*." *Proceedings of the National Academy of Sciences of the United States of America* 108, no. 5: 1987–1992. <https://doi.org/10.1073/pnas.1012544108>.

Marçet-Houben, M., and T. Gabaldón. 2015. "Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage." *PLoS Biology* 13, no. 8: e1002220. <https://doi.org/10.1371/journal.pbio.1002220>.

Marsit, S., J.-B. Leducq, É. Durand, A. Marchant, M. Filteau, and C. R. Landry. 2017. "Evolutionary Biology Through the Lens of Budding Yeast Comparative Genomics." *Nature Reviews Genetics* 18, no. 10: 581–598. <https://doi.org/10.1038/nrg.2017.49>.

McGovern, P. E. 2003. *Ancient Wine: The Search for the Origins of Viniculture*. Princeton, NJ: Princeton University Press. <https://doi.org/10.2307/j.ctvfd0bk>.

McGovern, P. E., J. Zhang, J. Tang, et al. 2004. "Fermented Beverages of Pre- and Proto-Historic China." *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 51: 17593–17598. <https://doi.org/10.1073/pnas.0407921102>.

Michel, R. H., P. E. McGovern, and V. R. Badler. 1992. "Chemical Evidence for Ancient Beer." *Nature* 360, no. 6399: 24. <https://doi.org/10.1038/360024b0>.

Minh, B. Q., M. A. T. Nguyen, and A. von Haeseler. 2013. "Ultrafast Approximation for Phylogenetic Bootstrap." *Molecular Biology and Evolution* 30, no. 5: 1188–1195. <https://doi.org/10.1093/molbev/mst024>.

Mortimer, R. K. 2000. "Evolution and Variation of the Yeast (*Saccharomyces*) Genome." *Genome Research* 10, no. 4: 403–409. <https://doi.org/10.1101/gr.10.4.403>.

Mozzachiodi, S., F. Bai, P. Baldrian, et al. 2022. "Yeasts From Temperate Forests." *Yeast* 39, no. 1–2: 4–24. <https://doi.org/10.1002/yea.3699>.

Naumov, G., S. A. James, E. S. Naumova, E. J. Louis, and I. N. Roberts. 2000. "Three New Species in the *Saccharomyces* Sensu Stricto Complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*." *International Journal of Systematic and Evolutionary Microbiology* 50, no. 5: 1931–1942.

Naumov, G. I., E. S. Naumova, R. A. Lantto, E. J. Louis, and M. Korhola. 1992. "Genetic Homology Between *Saccharomyces Cerevisiae* and Its Sibling Species *S. Paradoxus* and *S. Bayanus*: Electrophoretic Karyotypes." *Yeast* 8, no. 8: 599–612. <https://doi.org/10.1002/yea.320080804>.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32, no. 1: 268–274. <https://doi.org/10.1093/molbev/msu300>.

Nielsen, R., J. M. Akey, M. Jakobsson, J. K. Pritchard, S. Tishkoff, and E. Willerslev. 2017. "Tracing the Peopling of the World Through Genomics." *Nature* 541, no. 7637: 302–310. <https://doi.org/10.1038/nature21347>.

Nishant, K. T., W. Wei, E. Mancera, et al. 2010. "The Baker's Yeast Diploid Genome Is Remarkably Stable in Vegetative Growth and Meiosis." *PLoS Genetics* 6, no. 9: e1001109. <https://doi.org/10.1371/journal.pgen.1001109>.

Opulente, D. A., Q. K. Langdon, K. V. Buh, et al. 2019. "Pathogenic Budding Yeasts Isolated Outside of Clinical Settings." *FEMS Yeast Research* 19, no. 3: foz032. <https://doi.org/10.1093/femsyr/foz032>.

Osburn, K., J. Amaral, S. R. Metcalf, et al. 2018. "Primary Souring: A Novel Bacteria-Free Method for Sour Beer Production." *Food Microbiology* 70: 76–84. <https://doi.org/10.1016/j.fm.2017.09.007>.

Paradis, E., and K. Schliep. 2018. "ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35, no. 3: 526–528. <https://doi.org/10.1093/bioinformatics/bty633>.

Peter, J., M. D. Chiara, A. Friedrich, et al. 2018. "Genome Evolution Across 1,011 *Saccharomyces Cerevisiae* Isolates." *Nature* 556, no. 7701: 339. <https://doi.org/10.1038/s41586-018-0030-5>.

Pontes, A., M. Hutzler, P. H. Brito, and J. P. Sampaio. 2020. "Revisiting the Taxonomic Synonyms and Populations of *Saccharomyces cerevisiae*—Phylogeny, Phenotypes, Ecology and Domestication." *Microorganisms* 8, no. 6: 903. <https://doi.org/10.3390/microorganisms8060903>.

Potapov, P., M. C. Hansen, L. Laestadius, et al. 2017. "The Last Frontiers of Wilderness: Tracking Loss of Intact Forest Landscapes From 2000 to 2013." *Science Advances* 3, no. 1: e1600821. <https://doi.org/10.1126/sciadv.1600821>.

Pretorius, I. S. 2000. "Tailoring Wine Yeast for the New Millennium: Novel Approaches to the Ancient Art of Winemaking." *Yeast* 16, no. 8: 675–729. [https://doi.org/10.1002/1097-0061\(20000615\)16:8<675::aid-yea585>3.0.co;2-b](https://doi.org/10.1002/1097-0061(20000615)16:8<675::aid-yea585>3.0.co;2-b).

Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81, no. 3: 559–575. <https://doi.org/10.1086/519795>.

Robinson, H. A., A. Pinharanda, and D. Bensasson. 2016. "Summer Temperature Can Predict the Distribution of Wild Yeast Populations." *Ecology and Evolution* 6, no. 4: 1236–1250. <https://doi.org/10.1002/ece3.1919>.

Schacherer, J., J. A. Shapiro, D. M. Ruderfer, and L. Kruglyak. 2009. "Comprehensive Polymorphism Survey Elucidates Population Structure of *Saccharomyces cerevisiae*." *Nature* 458, no. 7236: 342–345. <https://doi.org/10.1038/nature07670>.

Scopel, E. F. C., J. Hose, D. Bensasson, and A. P. Gasch. 2021. "Genetic Variation in Aneuploidy Prevalence and Tolerance Across *Saccharomyces Cerevisiae* Lineages." *Genetics* 217, no. 4: iyab015. <https://doi.org/10.1093/genetics/fyab015>.

Shen, X.-X., D. A. Opulente, J. Kominek, et al. 2018. "Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum." *Cell* 175, no. 6: 1533–1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>.

Skelly, D. A., G. E. Merrihew, M. Riffle, et al. 2013. "Integrative Phenomics Reveals Insight Into the Structure of Phenotypic Diversity

in Budding Yeast." *Genome Research* 23, no. 9: 1496–1504. <https://doi.org/10.1101/gr.155762.113>.

Sniegowski, P. D., P. G. Dombrowski, and E. Fingerman. 2002. "Saccharomyces Cerevisiae and Saccharomyces Paradoxus Coexist in a Natural Woodland Site in North America and Display Different Levels of Reproductive Isolation From European Conspecifics." *FEMS Yeast Research* 1, no. 4: 299–306. <https://doi.org/10.1111/j.1567-1364.2002.tb00048.x>.

Song, G., B. J. A. Dickins, J. Demeter, et al. 2015. "AGAPE (Automated Genome Analysis PipelinE) for Pan-Genome Analysis of *Saccharomyces cerevisiae*." *PLoS One* 10, no. 3: e0120671. <https://doi.org/10.1371/journal.pone.0120671>.

Strope, P. K., D. A. Skelly, S. G. Kozmin, et al. 2015. "The 100-Genomes Strains, an *S. cerevisiae* Resource That Illuminates Its Natural Phenotypic and Genotypic Variation and Emergence as an Opportunistic Pathogen." *Genome Research* 25, no. 5: 762–774. <https://doi.org/10.1101/gr.185538.114>.

Sweeney, J. Y., H. A. Kuehne, and P. D. Sniegowski. 2004. "Sympatric Natural *Saccharomyces cerevisiae* and *S. paradoxus* Populations Have Different Thermal Growth Profiles." *FEMS Yeast Research* 4, no. 4–5: 521–525. [https://doi.org/10.1016/S1567-1356\(03\)00171-5](https://doi.org/10.1016/S1567-1356(03)00171-5).

Tamura, K., and M. Nei. 1993. "Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees." *Molecular Biology and Evolution* 10, no. 3: 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>.

Tamura, K., Q. Tao, and S. Kumar. 2018. "Theoretical Foundation of the RelTime Method for Estimating Divergence Times From Variable Evolutionary Rates." *Molecular Biology and Evolution* 35, no. 7: 1770–1782. <https://doi.org/10.1093/molbev/msy044>.

Tilakaratna, V., and D. Bensasson. 2017. "Habitat Predicts Levels of Genetic Admixture in *Saccharomyces cerevisiae*." *G3: Genes, Genomes, Genetics* 7, no. 9: 2919–2929. <https://doi.org/10.1534/g3.117.041806>.

Tsai, I. J., D. Bensasson, A. Burt, and V. Koufopanou. 2008. "Population Genomics of the Wild Yeast *Saccharomyces paradoxus*: Quantifying the Life Cycle." *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 12: 4957–4962. <https://doi.org/10.1073/pnas.0707314105>.

Wang, Q.-M., W.-Q. Liu, G. Liti, S.-A. Wang, and F.-Y. Bai. 2012. "Surprisingly Diverged Populations of *Saccharomyces cerevisiae* in Natural Environments Remote From Human Activity." *Molecular Ecology* 21, no. 22: 5404–5417. <https://doi.org/10.1111/j.1365-294X.2012.05732.x>.

Xu, S., Z. Dai, P. Guo, et al. 2021. "ggtreeExtra: Compact Visualization of Richly Annotated Phylogenetic Data." *Molecular Biology and Evolution* 38, no. 9: 4039–4042. <https://doi.org/10.1093/molbev/msab166>.

Yu, G., T. T.-Y. Lam, H. Zhu, and Y. Guan. 2018. "Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree." *Molecular Biology and Evolution* 35, no. 12: 3041–3043. <https://doi.org/10.1093/molbev/msy194>.

Sampaio, J. P., and P. Gonçalves. 2008. "Natural Populations of *Saccharomyces kudriavzevii* in Portugal Are Associated With Oak Bark and Are Sympatric With *S. cerevisiae* and *S. paradoxus*." *Applied and Environmental Microbiology* 74, no. 7: 2144–2152. <https://doi.org/10.1128/AEM.02396-07>.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T. Lam. 2017. "Ggtree: An R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data." *Methods in Ecology and Evolution* 8, no. 1: 28–36. <https://doi.org/10.1111/2041-210x.12628>.

Vaughan-Martini, A., and A. Martini. 1995. "Facts, Myths and Legends on the Prime Industrial Microorganism." *Journal of Industrial Microbiology and Biotechnology (Historical Archive)* 14, no. 6: 514–522.

Yue, J.-X., J. Li, L. Aigrain, et al. 2017. "Contrasting Evolutionary Genome Dynamics Between Domesticated and Wild Yeasts." *Nature Genetics* 49, no. 6: 913–924. <https://doi.org/10.1038/ng.3847>.

Zhan, S., W. Zhang, K. Niitepõld, et al. 2014. "The Genetics of Monarch Butterfly Migration and Warning Colouration." *Nature* 514, no. 7522: 317–321. <https://doi.org/10.1038/nature13812>.

Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov. 2014. "Precise Estimates of Mutation Rate and Spectrum in Yeast." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 22: E2310–E2318. <https://doi.org/10.1073/pnas.1323011111>.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.