



HAL
open science

Evidence for Two Main Domestication Trajectories in *Saccharomyces cerevisiae* Linked to Distinct Bread-Making Processes

Frédéric Bigey, Diego Segond, Anne Friedrich, Stephane Guezenec, Aurélie Bourgais, Lucie Huyghe, Nicolas Agier, Thibault Nidelet, Delphine Sicard

► **To cite this version:**

Frédéric Bigey, Diego Segond, Anne Friedrich, Stephane Guezenec, Aurélie Bourgais, et al.. Evidence for Two Main Domestication Trajectories in *Saccharomyces cerevisiae* Linked to Distinct Bread-Making Processes. *Current Biology - CB*, 2021, 31, pp.722-732. 10.1016/j.cub.2020.11.016 . hal-03110131

HAL Id: hal-03110131

<https://hal.science/hal-03110131v1>

Submitted on 2 Jun 2022

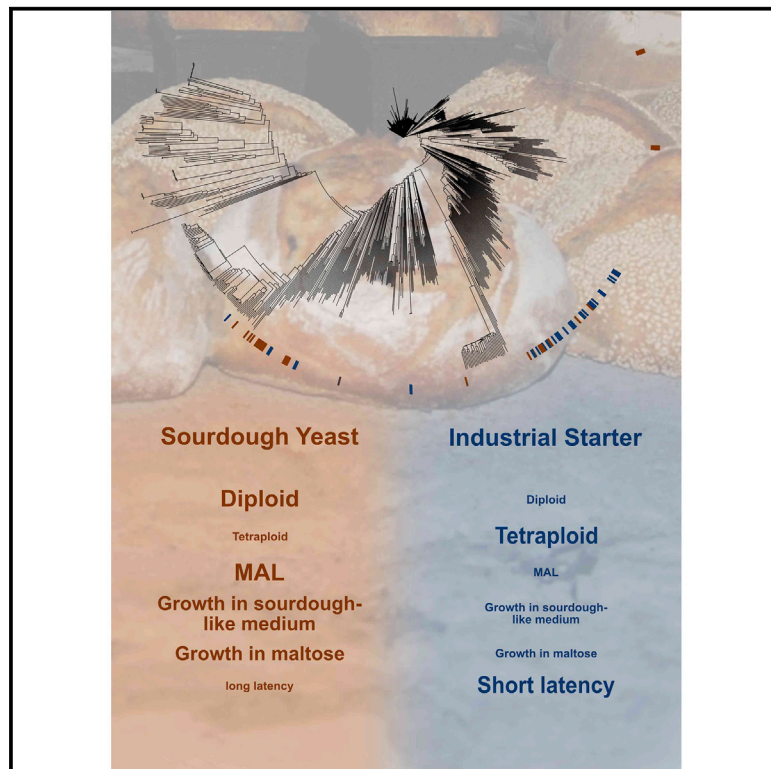
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Current Biology

Evidence for Two Main Domestication Trajectories in *Saccharomyces cerevisiae* Linked to Distinct Bread-Making Processes

Graphical Abstract



Authors

Frédéric Bigey, Diego Segond, Anne Friedrich, ..., Nicolas Agier, Thibault Nidelet, Delphine Sicard

Correspondence

delphine.sicard@inrae.fr

In Brief

The budding yeast, *Saccharomyces cerevisiae*, has long been used to make bread, but the domestication history of bread-making strains has been largely unknown. Here, Bigey et al. provide evidence that industrial and artisanal bread-making practices have led to the genomic and phenotypic divergence of two distinct bakery-strain lineages.

Highlights

- *Saccharomyces cerevisiae* bakery strains are polyphyletic
- Industrial strains have undergone at least one tetraploidization event
- Sourdough strains have a higher number of MAL genes and increased growth in maltose
- Bakery strains have specific phenotypic signatures of domestication



Article

Evidence for Two Main Domestication Trajectories in *Saccharomyces cerevisiae* Linked to Distinct Bread-Making Processes

Frédéric Bigey,¹ Diego Segond,¹ Anne Friedrich,⁴ Stephane Guezenec,¹ Aurélie Bourgeois,² Lucie Huyghe,² Nicolas Agier,³ Thibault Nidelet,¹ and Delphine Sicard^{1,5,*}

¹UMR 1083 SPO, INRAE, Université Montpellier, Montpellier SupAgro, Montpellier, France

²Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, Gif-sur-Yvette 91190, France

³Laboratory of Computational and Quantitative Biology, Institut de Biologie Paris-Seine, Sorbonne Université, CNRS, Paris 75005, France

⁴Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg 67000, France

⁵Lead Contact

*Correspondence: delphine.sicard@inrae.fr

<https://doi.org/10.1016/j.cub.2020.11.016>

SUMMARY

Production of leavened bread dates to the second millennium BCE. Since then, the art of bread making has developed, yet the evolution of bread-associated microbial species remains largely unknown. Nowadays, leavened bread is made either by using a pure commercial culture of the yeast *Saccharomyces cerevisiae* or by propagating a sourdough—a mix of flour and water spontaneously fermented by yeasts and bacteria. We studied the domestication of *S. cerevisiae* originating from industrial sources and artisanal sourdoughs and tested whether different bread-making processes led to population divergence. We found that *S. cerevisiae* bakery strains are polyphyletic with 67% of strains clustering into two main clades: most industrial strains were tetraploid and clustered with strains having diverse origins, including beer. By contrast, most sourdough strains were diploid and grouped in a second clade of strains having mosaic genomes and diverse origins, including fruits and natural environments. They harbored a higher copy number of genes involved in maltose utilization, and a high level of gene flow from multiple contributors was detected. Bakery strains displayed higher CO₂ production than do strains from other domesticated lineages (such as beer and wine), revealing a specific phenotypic signature of domestication. Interestingly, industrial strains had a shorter fermentation onset than sourdough strains, which were better adapted to a sourdough-like environment, suggesting divergent selection by industrial and artisanal processes. Our results reveal that the domestication of bakery yeast has been accompanied by dispersion, hybridization, and divergent selection through industrial and artisanal processes.

INTRODUCTION

Humans have engaged in fermented food processing since at least Neolithic times, when plant and animal domestication first occurred.^{1–7} Although plants and animals with desirable traits were consciously selected and bred, the domestication of microbes was not done intentionally until Pasteur discovered microbial fermentation. Rather, the selection of active microbial cultures was imposed by human food transformation processes and probably happened recurrently and independently since Neolithic times.⁸ Until recently, microbial domestication was poorly documented.^{2,3} Most studies focused on the filamentous fungi *Aspergillus oryzae* used in rice and soya fermentation and *Penicillium roqueforti* used for making blue cheese,^{9–11} as well as on the model yeast species *Saccharomyces cerevisiae*. This last species is found in many natural habitats (including soil, tree bark, and water) and has been domesticated for the production of a large diversity of

fermented drinks (wine, beer, sake, cachaça, coffee, and fermented milk) and foods (bread, cocoa, and olives).^{4,12–17} Selection for different fermented products has led to a genetic diversification of strains according to the fermentation type, although some gene flow between domesticated strains has been detected.¹⁶ Wine strains and the closely related group of “flor” strains likely have a single origin.^{18–21} Sake strains also appear to have evolved from a single origin,^{12,16,22} whereas cachaça strains evolved from wine strains through a secondary domestication process.²³ Coffee, cocoa, and beer strains have a more complex evolutionary history, where both migration and selection played major roles.^{24–28} Several genetic events have been detected in all these domesticated populations, including SNPs,^{20,24,28} gene duplication,^{24,29,30} horizontal gene transfer,^{18–20,31} and genome hybridization.²² Despite the cultural and historical importance of bread, the study of bakery-strain domestication has been neglected. This might be related to the fact that several industrial bakery yeast starters



have been found to be autotetraploid,³² which renders population genomic analysis complicated.^{33,34}

The earliest evidence of leavened bread was found during antiquity in the second millennium BCE in Egypt³⁵ and in the first millennium BCE in northwest China.³⁶ The art of making leavened bread developed further during ancient and medieval ages and was disseminated throughout the Mediterranean and in Middle East countries.³⁷ At that time, bread making consisted of mixing flour, water, and sourdough, a previously established mix of flour and water containing fermenting microbes. In the 19th century, the industrialization of food production and the advent of microbiology as a science resulted in the production of pure yeast cultures that were used as starter to make bread. Bread production with commercial *S. cerevisiae* yeast starter soon spread all over the world. Yet nowadays, global changes and the increase in the frequency of non-communicable diseases related to modern diets (including type 2 diabetes, obesity, and food allergies) have led to a renewed interest in traditional methods of bread making and in its local production. The appeal of traditionally prepared breads is underpinned by research showing improved flavor and nutritional benefits in sourdough bread made by artisanal bakers.^{38–40} Therefore, both ways of making leavened bread are currently found and bakers either use commercial yeasts or natural sourdoughs.

Natural sourdough is made from flour and water and is maintained by recurrent addition of flour and water, a process called backslopping. One or two prevailing species of lactic acid bacteria and one prevailing yeast species are usually found in sourdough starters,⁴¹ *S. cerevisiae* being the most widespread yeast species.^{37,42,43} It can be the dominant species in sourdough, even in bakeries where no industrial starter is used, suggesting that the species may colonize a sourdough from the bakery's environment or the baker's hands.⁴⁴ Therefore, bakery strains of *S. cerevisiae* may have undergone different domestication processes. Although industrial bread production may have led to the breeding and selection of homogeneous lineages of yeast starters, artisanal bread making may have selected strains through the continuous maintenance of sourdough microbial communities. These two types of domestication processes also apply to many different fermented foods.² The renewed interest in traditional sourdough bread production makes *S. cerevisiae* bakery populations an excellent model for studying the impact of both artisanal and industrial practices on fungal evolution and adaptation.

The objective of this study was to investigate the evolutionary history of *S. cerevisiae* strains isolated in bakeries. We found that bakery strains are polyphyletic and that commercial and sourdough strains chiefly separated in two different clades, suggesting different domestication roads for industrial and artisanal strains. Commercial strains were most often tetraploid and displayed a shorter fermentation lag phase. On the other hand, sourdough strains most often displayed duplications of the maltose and isomaltose maltase and permease genes and appeared better adapted to synthetic sourdoughs and maltose as a carbon source than strains from any other environment. An overall high proportion of admixture was detected and some sourdough strains clustered with commercial strains, suggesting

that dispersion and gene flow have also been involved in the evolution of bakery strains.

RESULTS

Prevalent Tetraploidy and Aneuploidy in Commercial Bakery Strains

Polyploidization, which refers to the multiplication of a complete chromosome set, has been found to be associated with domestication in plants but also in *Saccharomyces cerevisiae* for brewery and baking.^{24,45} A previous analysis of 26 *S. cerevisiae* strains originating from diverse fermented products showed that most bakery strains analyzed were autotetraploid, suggesting that bakery yeast domestication was also associated with polyploidization.³² We therefore analyzed ploidy in a set of 229 bakery *S. cerevisiae* strains (Table S1) by flow cytometry DNA quantification and by analyzing the number of alleles at microsatellite markers. Thirty-one strains were commercial yeasts and 198 were isolated from European sourdoughs collected in Italy, Belgium, and France. An overall high level of tetraploidy (99 out of 229 strains; 40%) and aneuploidy (39 out of 229 strains; 17%) was observed. We found that commercial strains were significantly (two-sided Fisher's exact test; $p < 0.001$) more frequently tetraploid (21 out of 31 strains; 68%) than sourdough strains (70 out of 198 strains; 35%). On the other hand, we did not observe any significant difference in ploidy distribution between the 198 sourdough strains isolated in Belgium (10 diploids, nine tetraploids, and one aneuploid), France (66 diploids, 37 tetraploids, and 26 aneuploids), and Italy (18 diploids, 24 tetraploids, and seven aneuploids; two-sided Fisher's exact test; $p > 0.05$).

To study whether tetraploidy promoted *S. cerevisiae* adaptation to a bakery environment, competition experiments between tetraploid and diploid strains were carried out in synthetic sourdoughs. Commercial and sourdough strains of each ploidy level were included in the analysis to test not only the effect of ploidy but also of commercial/sourdough origin (Table S1). No evidence of fitness gain for tetraploids was found ($t_{43} = -1,4288$; $p = 0.16$; Figure S1).

To further study whether tetraploids benefited from any advantage in a sourdough environment, the population dynamics of six tetraploids and their diploid offsprings were compared in a synthetic sourdough. The analysis showed that most of the observed variation was related to differences between parental strains (Data S1). On average, diploid offsprings had a marginally significant lower growth rate ($F_{1,22} = 3.6$; $p = 0.071$), a lower carrying capacity ($F_{1,22} = 5.86$; $p = 0.024$), and a higher mortality rate ($F_{1,22} = 4.98$; $p = 0.036$) than their tetraploid parents. However, these diploid offsprings were generated in the laboratory, and their fitness may not reflect the one of diploids strains found in bakeries. Their reduced fitness may be related to deleterious recessive mutations accumulated in the parental tetraploids and expressed in their laboratory-derived diploid offsprings. Therefore, we also compared the population size at the end of fermentation and the mortality rate of tetraploid and diploid bakery strains (Figure 1). No significant difference in mean mortality was detected between tetraploid and diploid strains ($F_{1,28} = 0.5$; $p = 0.486$). Diploid bakery strains reached a higher population size at the end of fermentation compared to tetraploids, as expected from the cell size/population size

Tree scale: 0.1

Sequence data



Ploidy

- 2N
- 4N
- aneuploid

Isolation

- commercial
- sourdough

Population

- P1
- P2
- P3
- P4
- P6

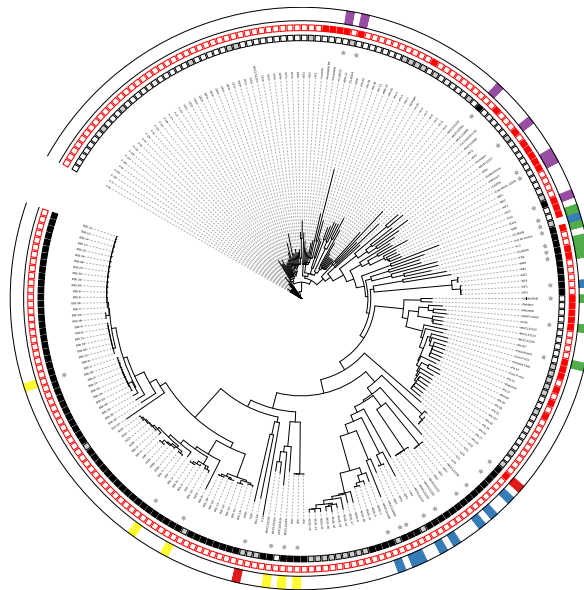


Figure 2. Neighbor Joining Tree Computed Using the Ritland Coefficient of Relatedness Obtained from Microsatellite Typing of the 229 Bakery Strains

Groups P1–P6 were defined by fastStructure analysis (Figure S3). Gray stars indicate strains for which a genome sequence was obtained. See also Figure S3 and Table S1.

analysis of principal components (DAPC), which does not rely on any life-history traits and evolutionary assumptions (Figure S3). The maximum-likelihood phylogenetic tree (Figure 3) showed that fastStructure genetic groups P3 and P4 were genetically close to each other, as were groups P2 and P6. It was concordant with the one found with microsatellite data (Figures 2 and 3) except one mosaic strain, which clustered on one side of the genomic tree (CLIV208-1). Overall, there was no clustering according to the country of origin.

We then analyzed the genetic relationship of bakery strains with the previously analyzed 1,011 strains of a worldwide *S. cerevisiae* collection.¹⁶ Adding extra bakery genomes to the 1,011 genomes tree did not change the clustering defined in Peter et al.¹⁶ (Figure 4). The bakery strains/genomes were distributed between the “wine/European”, “African beer”, “mixed origin”, “mosaic region 3”, “Asian fermentation”, and “mosaic region 1” clades. The P1 bakery group defined by fastStructure (7 bakery strains) was included in the “wine/European” clade, and group P5 (3 bakery strains) was found in the “African beer”-clade. The bakery groups P2 and P6 (23 strains, 7 from sourdough) were both located within the “mixed origin” clade that also included beer strains, clinical strains, and strains isolated from water, fruits, tree leaves, and natural environment. These two groups are indeed closely related and were not distinguished by fastStructure when the number of assumed ancestral groups *K* equaled 4 (Figure S3). Finally, the bakery groups P3 and P4 (19 strains, 16 from sourdough) were located within a group of mosaic strains that includes strains isolated from wine, sake, insect, palm wine, fruit, or clinical, fermentation, distillery, and natural environment. These two groups are also closely related and were not distinguished when *K* equaled 5.

To analyze the degree of admixture of bakery strains, we ran Admixture software on the bakery genomes as well as on 90 genomes chosen across the 1,011 genomes tree clades (Figure 5).¹⁶ The strains were chosen among all the clades except those not found to contribute to the bakery strains containing clades. A total of 48 bakery strains out of 68 presented some level of admixture

varying between strains and reaching up to 70% of the genome. For a single strain, from 2 to 11 ancestral populations were detected. There was evidence of admixture between bakery strains as well as between bakery strains and strains from the other genetic clades “Asian fermentation”, “wine/European”, “ale beer”, “Brazilian bioethanol”, “mosaic beer”, and “African beer”. One contributor, however, could not be identified, suggesting the presence of an extinct or otherwise

uncharacterized *S. cerevisiae* population. Novel alleles derived from unknown or extinct populations were also found in ale beer strains.⁵⁰ These high degrees of admixture possibly reflect dispersion and hybridization between bakery strains and strains from diverse fermented products.

The bakery strains of the “mixed origin” and “mosaic region 3” clades differed by the origin of some introgressions. Introgression from the “ale beer” clade was only detected in commercial and sourdough strains of the mixed origin clade (28% of the “mixed origin” bakery strains; Figure 5). Introgression from the “Asian fermentation” population was only found in bakery strains of the “mosaic region 3” clade (60% of the “mosaic region 3” bakery strains; Figure 5). This last result was confirmed by an analysis with Treemix⁵¹ that predicted a gene flow between “Asian fermentation” and “mosaic region 3” bakery populations ($w = 38\%$; three-population *f*₃ test; $Z = -27$; Figure S4).

An Increased Copy Number of Maltase and Isomaltase Transporter and Regulator Genes in the Sourdough Strains Clade Associated with a Better Growth on Maltose

Large copy number variations (CNVs) were previously detected in *S. cerevisiae*.¹⁶ We analyzed large CNV on our bakery strains (Table S2) and found that twenty-six strains (40%) displayed major chromosomal rearrangements or aneuploidy. Chromosome 9 was the most affected by CNV (10 strains out of 26). In addition, we analyzed the copy number of genes involved in maltose, isomaltose, and sucrose assimilation. These carbohydrates are common in cereal products, which may have led to the selection in bakery strains of an increased number of genes involved in their assimilation. In *S. cerevisiae*, the maltose gene cluster is composed of three genes, encoding the maltose transporter (permease; MAL1), maltase (MAL2), and transcription regulator (MAL3).^{47,52} These genes are clustered in five well-described MAL loci located on subtelomeric regions.^{47,52} A single MAL locus is sufficient to allow for maltose

Microsatellite

★ Microsatellite

Isolation

■ Commercial
□ Sourdough

Origin

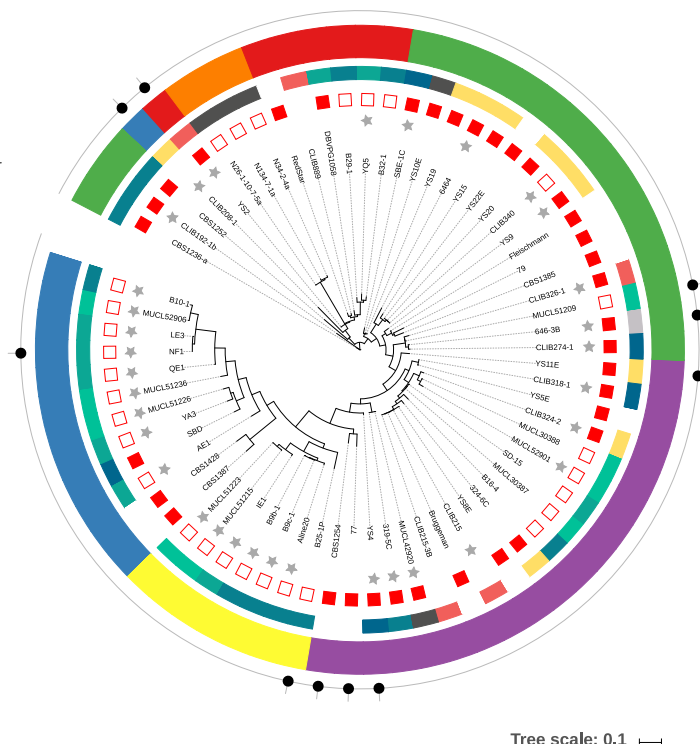
■ Europe
■ Europe/France
■ Europe/Italy
■ Europe/Belgium
■ Africa
■ Asia
■ America
■ Oceania
□ NA

Population

■ P1
■ P2
■ P3
■ P4
■ P5
■ P6

Mosaic

● Mosaic



Tree scale: 0.1

Figure 3. Maximum-Likelihood Phylogenetic Tree Obtained from 33,032 Biallelic SNPs from 68 Bakery Strains Using RAxML (Evolution Model: GTRGAMMA)

The most likely tree from 100 bootstrap replicates is presented. Groups P1–P6 were defined by fast-Structure analysis (Figure S3). See also Tables S1 and S2.

Phenotypic Signatures of Domestication

To further study whether sourdough strains have undergone selection by the baking process, we compared fermentation performances between bakery and non-bakery strains and analyzed phenotypic convergence among bakery strains. In total, eighty-four strains, originating from sourdough ($n = 27$), commercial starter ($n = 19$), wine ($n = 11$), beer ($n = 13$), Asian fermentation ($n = 9$), as well as non-fermented products (referred to as “other”; $n = 5$) were fermented in a synthetic sourdough-like medium (Table S1). Bakery strains were chosen to represent the six fastStructure genetic groups of bakery strains (Table S1). Non-bakery strains were chosen in the two main bakery strain clades “mosaic region 3” and “mixed origin” and in the clades with other domesticated lineages (“alebeer”, “mosaic beer”, “wine,” and “Asian fermentation”; Table S1).

Sourdough and commercial bakery strains had in average significantly higher maximum cumulative CO_2 production ($F_{5,73} = 14.5$; $p < 0.001$; data and Tukey test; Data S3; Figure 7A) and higher maximum CO_2 production rate ($F_{5,78} = 21.89$; $p < 0.001$; data and Tukey test; Data S3; Figure 7B) than strains coming from any other fermented or non-fermented product, suggesting that both sourdough and commercial strains have been selected for better performance in bread fermentation. Commercial bakery strains performed better than sourdough strains in terms of fermentation onset ($F_{5,74} = 3.1$; $p = 0.01$; Tukey test; Data S3; Figure 7C). However, sourdough strains had a maximum cumulative CO_2 production and a maximum CO_2 production rate as high as commercial strains. Moreover, sourdough strains reached a significantly higher population size than commercial bakery strains ($F_{5,80} = 8.1$; $p < 0.001$; Tukey test; Data S3; Figure 7E). They also reached, on average, a higher population size than strains from other fermentation origins, except strains from Asian fermentation that reached, on average, a not significantly different (albeit slightly lower) population size (Tukey test; Data S3; Figure 7E). Overall, these results showed that sourdough strains are better adapted to sourdough-mimicking media than strains from wine, beer, and non-fermented environments. Surprisingly, beer strains, which

fermentation.^{48,49} In the “mosaic region 3” clade, in which most bakery strains were isolated from sourdough, the copy number of the *MAL12* (fold change [FC] = 1.6; Wilcoxon rank sum test; $p < 10^{-3}$), *MAL32* (FC = 2.0; $p < 10^{-4}$), and *MAL31* (FC = 1.8; $p < 10^{-4}$) was on average significantly higher in bakery strains than in non-bakery strains (Figures 6 and S5; Data S2). This was also the case for the *IMA1* (FC = 2.1; $p < 10^{-6}$), *IMA3* (FC = 2.1; $p < 10^{-4}$), and *IMA4* (FC = 2.1; $p < 10^{-4}$) isomaltase genes (Figure S5). We then analyzed the growth rate and carrying capacity of strains grown on maltose as sole carbon source. We found a significant effect of the CNV on the growth rate for *MAL11*, *MAL31*, *MAL32*, *MAL33*, *IMA1*, *IMA2*, *IMA3*, and *IMA4* and a significant effect of the CNV on the carrying capacity for all these genes but *MAL32* and *MAL33* (Figure 6; data and statistical results after Bonferroni correction shown in Data S2). In the “mixed origin” clade, where many commercial bakery strains were located, we did not detect any increase of maltose and isomaltase genes copy number in bakery strains compared to others. There was no effect of the CNV on growth in maltose except when a diploid laboratory-derived beer strain (Win-8D) was included in the analysis (Figure S6; Data S2). These results suggest selection for a better assimilation of maltose and isomaltose in sourdough, where starch degradation releases maltose. In the same way, adaptation to a beer environment was found to be associated with an increased copy number of the maltase gene cluster.²⁴

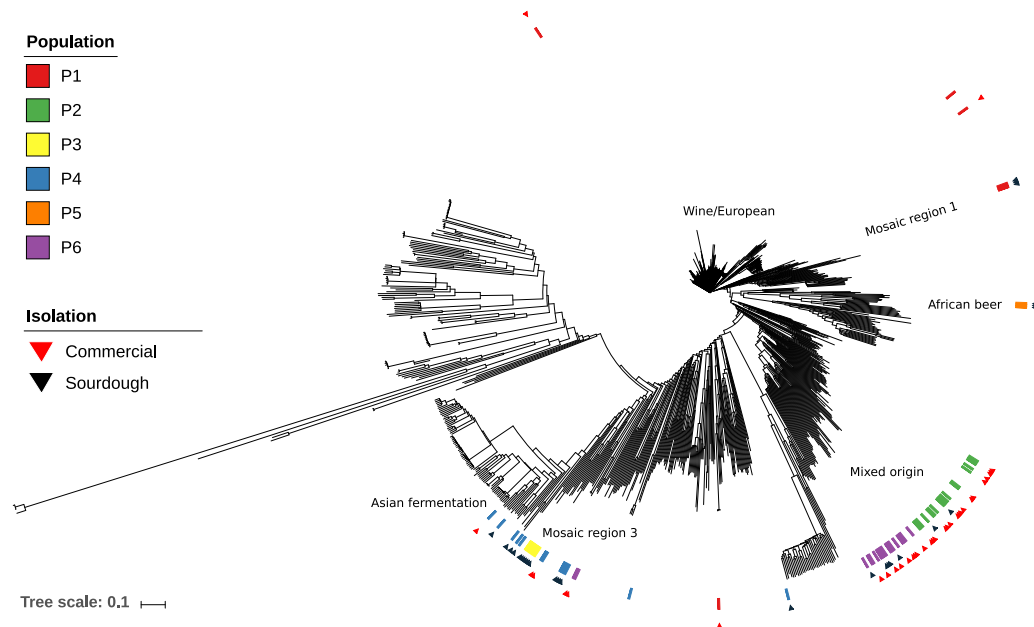


Figure 4. Phylogenetic Tree Obtained from SNPs from Strains of the *S. cerevisiae* 1,011-Genomes Project

17 newly sequenced diploid sourdough strains genomes and a representative set of 51 previously sequenced bakery strains extended the dataset to a total of 68 bakery yeast genomes (Table S1). Groups P1–P6 were defined by fastStructure analysis (Figure S3). The names of the clades are taken from the 1,011-genomes project as described in Peter et al.¹⁶ See also Figure S3 and Tables S1 and S2.

originate from a cereal-fermented product, had a significantly higher mortality rate than all other strains in a sourdough-mimicking medium ($F_{5,67} = 5.03$; $p < 0.001$; Tukey test; Data S3; Figure 7F).

DISCUSSION

We report here the first broad analysis of bakery yeast domestication using a collection of 229 *Saccharomyces cerevisiae*

K = 15

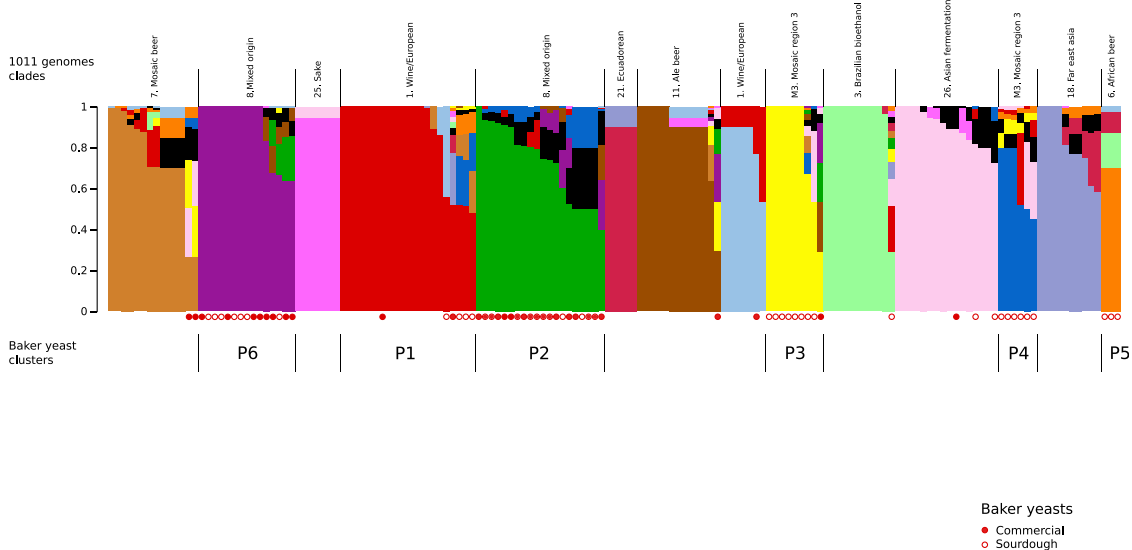


Figure 5. Population Structure Obtained from 48,482 Biallelic SNPs from 157 Genomes (Including 68 Bakery Strains; Table S1) Using Admixture Software

The vertical axis depicts the fractional representation of resolved populations (colors) within each strain (horizontal axis) for K assumed ancestral populations. The value of K = 17 exhibited the lowest cross-validation error compared to other K values and best explained the structure. Groups P1–P6 were defined by fastStructure analysis (Figure S3). The clades from the 1,011-genomes project are those described in Peter et al.¹⁶ See also Figure S4 and Table S1.

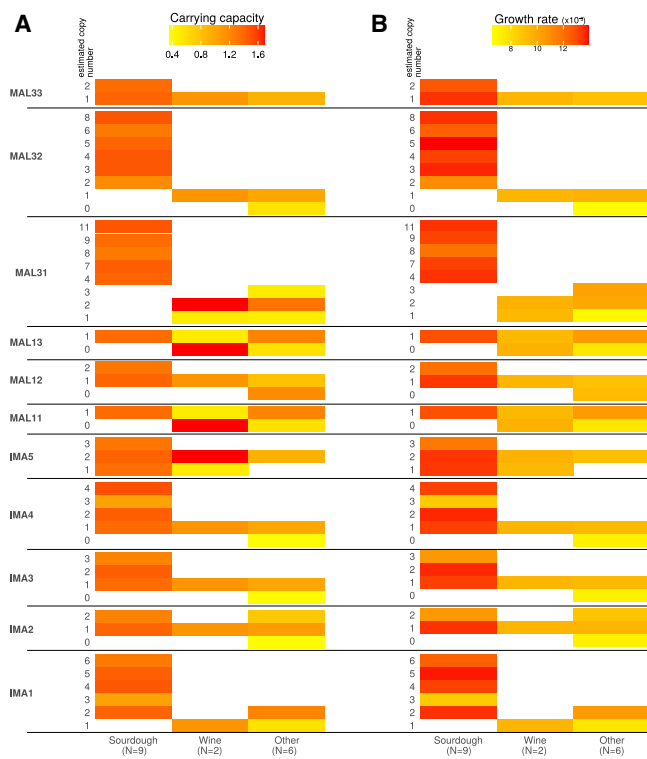


Figure 6. Population Growth of Strains Having Different Copy Numbers of Maltose and Isomaltose Genes Clusters

Copy number variations were estimated using gene reads coverage relative to a neighbor non-maltose gene on the chromosome. Strains were grown in a synthetic sourdough medium with only maltose as carbon source. All strains belong to the “mosaic region 3” genomic clade. Population growth was followed by spectrophotometry at 610 nm and compared between strains isolated from sourdough, wine, and non-fermented products (“other”). (A) Average carrying capacity (optical density 610 [OD₆₁₀]) is shown; (B) average growth rate (h⁻¹) is shown. See also [Figures S5](#) and [S6](#), [Table S1](#), and [Data S2](#).

bakery strains collected worldwide from industry or natural sourdoughs. We found that the origin of bakery strains is polyphyletic. Most clustered in two main clades, suggesting that bakery strains have followed at least two main domestication trajectories: one appeared to have led to most commercial strains, although the other led to sourdough strains.

The domestication of commercial bakery strains was associated with at least one tetraploidization event. To our knowledge, this is the first time that dispersion of widespread autotetraploids is associated with fungus domestication. Beer strains of *S. cerevisiae* were found to have a high rate of ploidy variation also associated to aneuploidy, but there was no evidence of worldwide spread of autotetraploids.^{16,24} A change in chromosomal copy number is often observed in yeasts whenever they are adapting to new, stressful environments.^{45,53,54} However, several studies showed genome instability in yeast polyploids. Chromosome loss, aneuploid mis-segregation event, chromosome translocation, and large chromosome rearrangement all occurred during the evolution of yeast polyploids.⁴⁵ Experimental evolution showed that tetraploid ancestors converged into diploids in 1,800 mitotic generations.^{55,56} Here, we found that the vast majority of commercial yeasts are autotetraploid,

as are also some strains isolated from sourdough. Bakery autotetraploids are in reproductive isolation with diploids and thus represent a new species.³² The high proportion of tetraploids among commercial strains as well as the association of tetraploids with a shorter fermentation latency phase suggest that deliberate artificial selection by industrial yeast geneticists could be at the origin of tetraploid bakery strains and that worldwide industrial distribution of these selected strains might have disseminated these tetraploids in bakery environments. The high level of aneuploidy observed in bakery strains could result from the known instability of tetraploid genomes.

Sourdough strains domestication was associated with (1) increased copy number of the maltose and isomaltose transporter (permease) and maltase and isomaltase transcription regulator encoding genes and (2) increased growth on maltose as sole carbon source ([Figure 6](#)). Although CNVs are generally deleterious, they also appear to be a key mechanism that can enable adaptation during an episode of stringent selection. CNVs are widespread in domesticated plant and animal species,⁵⁷ as well as in domesticated yeast populations.^{3,4} Because maltose and isomaltose are released through amyolytic starch breakdown, they constitute an important carbon source in dough and may directly be linked to fermentation performance (duration and CO₂ production). Therefore, natural selection in sourdough as well as unintentional selection by bakers for increased fermentation rate may both have resulted in strains having an increased growth on maltose and an increased underlying copy number of genes related to maltose and isomaltose utilization. Alternatively, another explanation might be exaptation, through which the increased ability to grow on maltose would have appeared in strains before their ancestors were introduced or propagated on sourdough.

Some sourdough strains were found in the commercial strains “mixed origin” clade and vice versa, suggesting that commercial starters may disseminate in natural sourdoughs. Some bakers may indeed add commercial starters to their sourdough or may contaminate their sourdough by using commercial yeast for making other bakery products, such as pastries. Admixture between these two clades was also detected, suggesting gene flow occurrence between commercial and sourdough strains.

Bakery strains have been hypothesized to be genetically related to beer strains.^{13,24} In the 17th century, baker’s yeast was reportedly provided by brewers.⁴ Yet, although beer strains were found in seven clades out of the 26 clades structuring the 1,011 *S. cerevisiae* worldwide strain collection,¹⁶ only two of these clades also contained bakery strains. First, the “African beer” clade contains sourdough strains isolated from maize dough coming from Ghana, suggesting that the same strains are used both to ferment maize dough and to brew beer. Second, the “mixed origin” clade contains beer strains, commercial bakery strains from all over the world, a few sourdough strains from Belgium and France, as well as strains from diverse other habitats (fruits, soil, water, humans, etc.). None of the bakery strains clustered with the “ale beer” clade. However, some introgression from this clade was detected in ten out of the 36 bakery strains of the mixed origin clade, indicating the presence of some gene flow between ale beer and bakery strains. This clade mostly contained commercial strains, suggesting that some ale beer

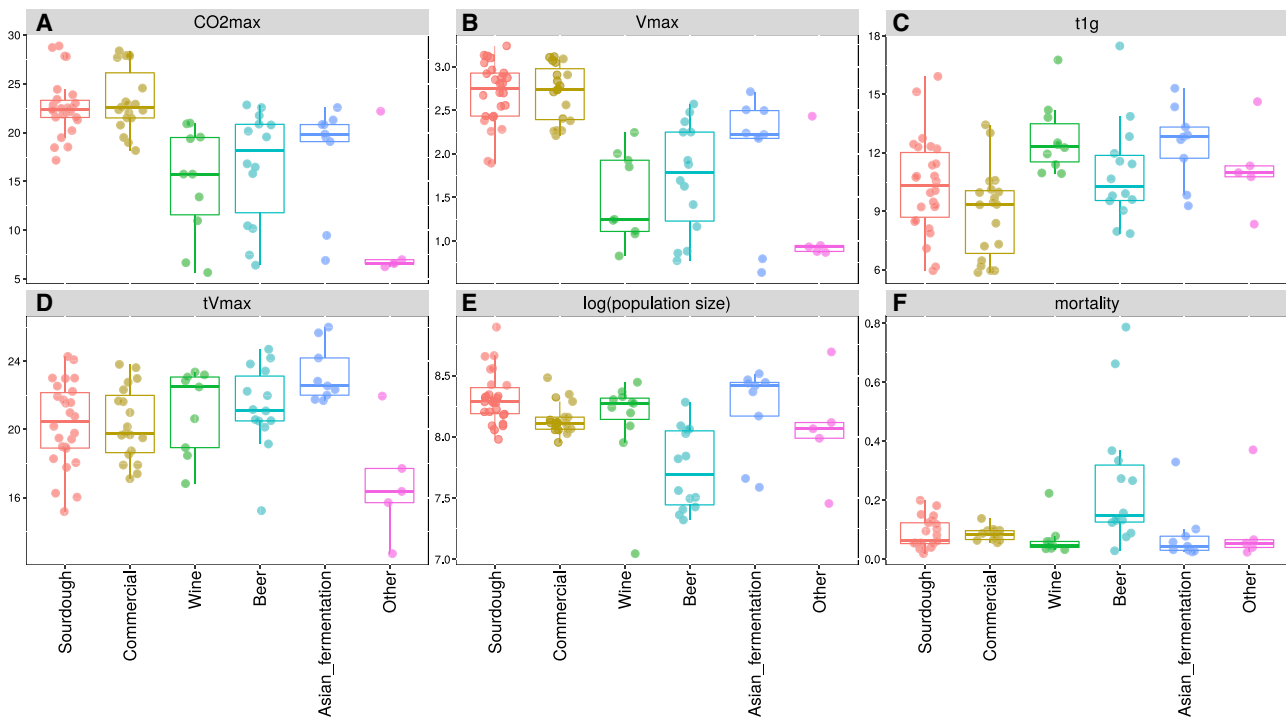


Figure 7. Phenotypic Variation of Fermentation Kinetics in a Synthetic Sourdough Medium among Strains Isolated from Sourdough, Commercial Bakery Starter, Wine, Beer, Asian Fermentations, and Non-fermented Environments (Other)

Six parameters were estimated: (A) maximum CO₂ release (g); (B) maximum CO₂ production rate Vmax (g/L/h); (C) fermentation latency-phase time (h; time between inoculation and the beginning of the fermentation calculated as 1 g of CO₂ release); (D) time of the maximum CO₂ production rate; (E) log of the population size; and (F) cell mortality as measured by flow cytometry after 27 h of fermentation. See also [Table S1](#) and [Data S3](#).

strains have been used as progenitors in bakery strains industrial breeding.

A previous study also proposed that bakery strains could originate from a tetraploidization event between ale beer and wine strains.¹³ Moreover, a recent study found that ale beer strains were derived from admixture between populations closely related to European grape wine strains and Asian rice wine strains.⁵⁰ Here, all bakery strains but three clustered separately from the well-defined wine and sake lineages, clearly suggesting a distinct evolutionary history. The admixture analysis provided no evidence of gene flow between bakery and sake strains. However, there was evidence of low introgression from wine populations and from Asian fermentation populations. Interestingly, signs of introgression from wine populations were found both in sourdough and commercial strains although introgressions from Asian fermentation populations were detected only in sourdough strains (Figure 5). Artisanal bakers often experiment with other fermentation processes, including Asian fermentation, which may explain this finding.

Sourdough is a human-made habitat. One may consider that any yeast population originating from sourdough is domesticated because this environment would not exist without human intervention. Alternatively, one may consider that sourdough yeast populations can only be considered domesticated if some selection imposed by the baking process has indeed occurred. Here, we suggest that sourdough strains are not only present in their environment by chance or by recurrent introduction but could have been selected for better fermentation

performance through bakery processes. They compared to commercial strains in terms of CO₂ production but reached a higher population size at the end of fermentation in synthetic sourdoughs. They also displayed better growth on maltose than commercial strains. These results suggest that sourdough strains are indeed better adapted to a sourdough environment and provide interesting genetic resources for improving sourdough-bread-making processes. Sourdough yeasts co-occur with lactic acid bacteria. Future work on yeasts-bacteria interaction will likely shed additional light on their possible coevolution in a sourdough environment.

The relative contribution of selection or neutral evolutionary processes in the evolution of bakery strains is difficult to estimate; yet introgressions from beer, wine, or Asian fermentation strains alone did not appear to be responsible for improved fermentation performance in bakery strains because commercial and sourdough strains produced more CO₂ and at a higher rate than all three other domesticated lineages (beer, wine, and Asian fermentation; Figure 7). Indeed, bakery strains had a better fermentation performance in a synthetic sourdough than strains from any other origin, in line with the hypothesis of a specific bread-making process domestication.

In summary, we revealed that bakery strains have followed at least two main evolutionary trajectories that mobilized different genetic events (tetraploidization and CNV) and were associated with phenotypic divergence. At least one tetraploidization event appeared to have occurred in the past of most commercial strains that also have a shortened latency phase of fermentation.

On the other side, several duplications of some maltose genes appeared to have occurred in sourdough strains that are also better adapted to grow on maltose as sole carbon source than strains from any other origin. Finally, we provide some evidence that different sources of introgression have occurred in sourdough and commercial strains (respectively, Asian and ale beer sources), suggesting that dispersion and gene flow are also important evolutionary processes in bakery strains. Overall, this is the first time that the analysis of fungal domestication reveals that artisanal and industrial domestication led to divergent populations. This demonstrates the need to conserve different fermentation practices to maintain microbial genetic diversity.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Yeast collection and cultivation procedure
- **METHOD DETAILS**
 - Ploidy analysis by flow cytometry
 - Fitness and fermentation analysis
 - Microsatellite typing and analysis
 - Genome sequencing and read processing
 - Variant calling
 - Population structure
 - Phylogenetic tree imputation
 - Analysis across the 1,011 genomes data
 - Copy number variations
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2020.11.016>.

ACKNOWLEDGMENTS

We acknowledge Pierre Gomes for curation of microsatellite data, a panel of bakers and farmer-bakers for providing sourdoughs, F. Minervini and the University of Perugia for providing Italian sourdough strains, as well as all the international yeast collections for providing strains. We are also grateful to Agnes Masquin and Isabelle Delmas for their administrative support and Jean-Luc Legras for discussion and critical reading of the manuscript. This work benefited from the support of the project BAKERY ANR-13-ALID-0005 of the French National Research Agency.

AUTHOR CONTRIBUTIONS

D. Sicard supervised the project and designed the experiment; F.B., T.N., and D. Sicard analyzed the data; F.B. and D. Sicard wrote the manuscript; D. Sicard performed fitness experiments and conducted fermentation in synthetic medium; T.N. and D. Sicard performed statistical analysis on fitness and fermentation data; S.G. was in charge of the yeast collection and genomic DNA extraction for sequencing; L.H., N.A., and A.B. performed microsatellite typing and cytometry analysis; and F.B. and A.F. performed genomes analysis.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 27, 2020

Revised: September 7, 2020

Accepted: November 5, 2020

Published: December 9, 2020

REFERENCES

1. McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nuñez, A., Butrym, E.D., Richards, M.P., Wang, C.S., et al. (2004). Fermented beverages of pre- and proto-historic China. *Proc. Natl. Acad. Sci. USA* *101*, 17593–17598.
2. Gibbons, J.G., and Rinker, D.C. (2015). The genomics of microbial domestication in the fermented food environment. *Curr. Opin. Genet. Dev.* *35*, 1–8.
3. Steensels, J., Gallone, B., Voordeckers, K., and Verstrepen, K.J. (2019). Domestication of industrial microbes. *Curr. Biol.* *29*, R381–R393.
4. Sicard, D., and Legras, J.-L. (2011). Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C. R. Biol.* *334*, 229–236.
5. Brüßow, H. (2020). Bioarchaeology: a profitable dialogue between microbiology and archaeology. *Microb. Biotechnol.* *13*, 406–409.
6. Aouizerat, T., Gutman, I., Paz, Y., Maeir, A.M., Gadot, Y., Gelman, D., Szitenberg, A., Drori, E., Pinkus, A., Schoemann, M., et al. (2019). Isolation and characterization of live yeast cells from ancient vessels as a tool in bio-archaeology. *MBio* *10*, e00388–19.
7. Salque, M., Bogucki, P.I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M., and Evershed, R.P. (2013). Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature* *493*, 522–525.
8. Bodinaku, I., Shaffer, J., Connors, A.B., Steenwyk, J.L., Biango-Daniels, M.N., Kastman, E.K., Rokas, A., Robbat, A., and Wolfe, B.E. (2019). Rapid phenotypic and metabolomic domestication of wild *Penicillium* molds on cheese. *MBio* *10*, e02445–19.
9. Gibbons, J.G., Salichos, L., Slot, J.C., Rinker, D.C., McGary, K.L., King, J.G., Klich, M.A., Tabb, D.L., McDonald, W.H., and Rokas, A. (2012). The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Curr. Biol.* *22*, 1403–1409.
10. Ropars, J., Rodríguez de la Vega, R.C., López-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, É., Lacoste, S., Debuchy, R., Dupont, J., Branca, A., and Giraud, T. (2015). Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Curr. Biol.* *25*, 2562–2569.
11. Dumas, E., Feurtey, A., de la Vega, R.C.R., Prieur, S.L., Snirc, A., Coton, M., Thierry, A., Coton, E., Piver, M.L., Roueyre, D., et al. (2020). Independent domestication events in the blue-cheese fungus *Penicillium roqueforti*. *Mol. Ecol.* *29*, 2639–2660.
12. Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* *458*, 337–341.
13. Legras, J.-L., Merdinoglu, D., Cornuet, J.-M., and Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol. Ecol.* *16*, 2091–2102.
14. Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., and Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat. Commun.* *9*, 2690.
15. Goddard, M.R., and Greig, D. (2015). *Saccharomyces cerevisiae*: a nomadic yeast with no niche? *FEMS Yeast Res.* *15*, fov009.
16. Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freil, K., Llored, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* *556*, 339–344.

17. Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A., and Bai, F.-Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol. Ecol.* **21**, 5404–5417.
18. Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J.-L., Serra, M., Dequin, S., Couloux, A., et al. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol. Ecol.* **24**, 5412–5427.
19. Marsit, S., Mena, A., Bigey, F., Sauvage, F.-X., Couloux, A., Guy, J., Legras, J.-L., Barrio, E., Dequin, S., and Galeote, V. (2015). Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Mol. Biol. Evol.* **32**, 1695–1707.
20. Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., et al. (2018). Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* **35**, 1712–1727.
21. Coi, A.L., Bigey, F., Mallet, S., Marsit, S., Zara, G., Gladieux, P., Galeote, V., Budroni, M., Dequin, S., and Legras, J.L. (2017). Genomic signatures of adaptation to wine biological ageing conditions in biofilm-forming flor yeasts. *Mol. Ecol.* **26**, 2150–2166.
22. Fay, J.C., and Benavides, J.A. (2005). Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, 66–71.
23. Barbosa, R., Pontes, A., Santos, R.O., Montandon, G.G., de Ponzzes-Gomes, C.M., Morais, P.B., Gonçalves, P., Rosa, C.A., and Sampaio, J.P. (2018). Multiple rounds of artificial selection promote microbe secondary domestication—the case of cachaça yeasts. *Genome Biol. Evol.* **10**, 1939–1955.
24. Gallone, B., Steensels, J., Prah, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166**, 1397–1410.e16.
25. Gallone, B., Mertens, S., Gordon, J.L., Maere, S., Verstrepen, K.J., and Steensels, J. (2018). Origins, evolution, domestication and diversity of *Saccharomyces* beer yeasts. *Curr. Opin. Biotechnol.* **49**, 148–155.
26. Libkind, D., Hittinger, C.T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P., and Sampaio, J.P. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. USA* **108**, 14539–14544.
27. Ludlow, C.L., Cromie, G.A., Garmendia-Torres, C., Sirt, A., Hays, M., Field, C., Jeffery, E.W., Fay, J.C., and Dudley, A.M. (2016). Independent origins of yeast associated with coffee and cacao fermentation. *Curr. Biol.* **26**, 965–971.
28. Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., and Sampaio, J.P. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* **26**, 2750–2761.
29. Dunn, B., Richter, C., Kvitek, D.J., Pugh, T., and Sherlock, G. (2012). Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res.* **22**, 908–924.
30. Steenwyk, J., and Rokas, A. (2017). Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. *G3* **7**, 1475–1485.
31. Marsit, S., Sanchez, I., Galeote, V., and Dequin, S. (2016). Horizontally acquired oligopeptide transporters favour adaptation of *Saccharomyces cerevisiae* wine yeast to oenological environment. *Environ. Microbiol.* **18**, 1148–1161.
32. Albertin, W., Marullo, P., Aigle, M., Bourgais, A., Bely, M., Dillmann, C., DE Vienne, D., and Sicard, D. (2009). Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. *J. Evol. Biol.* **22**, 2157–2170.
33. Meirmans, P.G., Liu, S., and van Tienderen, P.H. (2018). The analysis of polyploid genetic data. *J. Hered.* **109**, 283–296.
34. Dufresne, F., Stift, M., Vergilino, R., and Mable, B.K. (2014). Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Mol. Ecol.* **23**, 40–69.
35. Samuel, D. (1996). Investigation of ancient Egyptian baking and brewing methods by correlative microscopy. *Science* **273**, 488–490.
36. Shevchenko, A., Yang, Y., Knaust, A., Thomas, H., Jiang, H., Lu, E., Wang, C., and Shevchenko, A. (2014). Proteomics identifies the composition and manufacturing recipe of the 2500-year old sourdough bread from Subeixi cemetery in China. *J. Proteomics* **105**, 363–371.
37. Carbonetto, B., Ramsayer, J., Nidelet, T., Legrand, J., and Sicard, D. (2018). Bakery yeasts, a new model for studies in ecology and evolution. *Yeast* **35**, 591–603.
38. Pétel, C., Onno, B., and Prost, C. (2017). Sourdough volatile compounds and their contribution to bread: A review. *Trends Food Sci. Technol.* **59**, 105–123.
39. Gänzle, M.G., and Zheng, J. (2019). Lifestyles of sourdough lactobacilli - Do they matter for microbial ecology and bread quality? *Int. J. Food Microbiol.* **302**, 15–23.
40. Gobbetti, M., De Angelis, M., Di Cagno, R., Calasso, M., Archetti, G., and Rizzello, C.G. (2019). Novel insights on the functional/nutritional features of the sourdough fermentation. *Int. J. Food Microbiol.* **302**, 103–113.
41. Corsetti, A., and Settanni, L. (2007). Lactobacilli in sourdough fermentation. *Food Res. Int.* **40**, 539–558.
42. Michel, E., Masson, E., Bubbendorf, S., Lapique, L., Legrand, J., Guézennec, S., Nidelet, T., Marlin, T., Rué, O., Onno, B., et al. (2019). Artisanal and farmers bread making practices differently shape fungal species diversity in French sourdoughs. *bioRxiv*. <https://doi.org/10.1101/679472>.
43. Urien, C., Legrand, J., Montalent, P., Casaregola, S., and Sicard, D. (2019). Fungal species diversity in French bread sourdoughs made of organic wheat flour. *Front. Microbiol.* **10**, 201.
44. Reese, A.T., Madden, A.A., Joossens, M., Lacaze, G., and Dunn, R.R. (2020). Influences of ingredients and bakers on the bacteria and fungi in sourdough starters and bread. *MSphere* **5**, e00950-19.
45. Albertin, W., and Marullo, P. (2012). Polyploidy in fungi: evolution after whole-genome duplication. *Proc. Biol. Sci.* **279**, 2497–2509.
46. Spor, A., Nidelet, T., Simon, J., Bourgais, A., de Vienne, D., and Sicard, D. (2009). Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*. *BMC Evol. Biol.* **9**, 296.
47. Naumov, G.I., Naumova, E.S., and Michels, C.A. (1994). Genetic variation of the repeated MAL loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genetics* **136**, 803–812.
48. Needleman, R.B., Kaback, D.B., Dubin, R.A., Perkins, E.L., Rosenberg, N.G., Sutherland, K.A., Forrest, D.B., and Michels, C.A. (1984). MAL6 of *Saccharomyces*: a complex genetic locus containing three genes required for maltose fermentation. *Proc. Natl. Acad. Sci. USA* **81**, 2811–2815.
49. Needleman, R. (1991). Control of maltase synthesis in yeast. *Mol. Microbiol.* **5**, 2079–2084.
50. Fay, J.C., Liu, P., Ong, G.T., Dunham, M.J., Cromie, G.A., Jeffery, E.W., Ludlow, C.L., and Dudley, A.M. (2019). A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. *PLoS Biol.* **17**, e3000147.
51. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967.
52. Charron, M.J., Read, E., Haut, S.R., and Michels, C.A. (1989). Molecular evolution of the telomere-associated MAL loci of *Saccharomyces*. *Genetics* **122**, 307–316.
53. Yona, A.H., Manor, Y.S., Herbst, R.H., Romano, G.H., Mitchell, A., Kupiec, M., Pilpel, Y., and Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. USA* **109**, 21010–21015.

54. Dhar, R., Sägesser, R., Weikert, C., Yuan, J., and Wagner, A. (2011). Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. *J. Evol. Biol.* *24*, 1135–1153.
55. Gerstein, A.C., Chun, H.-J.E., Grant, A., and Otto, S.P. (2006). Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genet.* *2*, e145.
56. Gerstein, A.C., McBride, R.M., and Otto, S.P. (2008). Ploidy reduction in *Saccharomyces cerevisiae*. *Biol. Lett.* *4*, 91–94.
57. Lye, Z.N., and Purugganan, M.D. (2019). Copy number variation in domestication. *Trends Plant Sci.* *24*, 352–365.
58. Lhomme, E., Lattanzi, A., Dousset, X., Minervini, F., De Angelis, M., Lacaze, G., Onno, B., and Gobbetti, M. (2015). Lactic acid bacterium and yeast microbiotas of sixteen French traditional sourdoughs. *Int. J. Food Microbiol.* *215*, 161–170.
59. Vrancken, G., De Vuyst, L., Van der Meulen, R., Huys, G., Vandamme, P., and Daniel, H.-M. (2010). Yeast species composition differs between artisan bakery and spontaneous laboratory sourdoughs. *FEMS Yeast Res.* *10*, 471–481.
60. Minervini, F., Lattanzi, A., De Angelis, M., Di Cagno, R., and Gobbetti, M. (2012). Influence of artisan bakery- or laboratory-propagated sourdoughs on the diversity of lactic acid bacterium and yeast microbiotas. *Appl. Environ. Microbiol.* *78*, 5328–5340.
61. Breslow, D.K., Cameron, D.M., Collins, S.R., Schuldiner, M., Stewart-Ornstein, J., Newman, H.W., Braun, S., Madhani, H.D., Krogan, N.J., and Weissman, J.S. (2008). A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* *5*, 711–718.
62. Vrancken, G., Rimaux, T., De Vuyst, L., and Leroy, F. (2008). Kinetic analysis of growth and sugar consumption by *Lactobacillus fermentum* IMDO 130101 reveals adaptation to the acidic sourdough ecosystem. *Int. J. Food Microbiol.* *128*, 58–66.
63. Bloem, A., Rollero, S., Seguinot, P., Crépin, L., Perez, M., Picou, C., and Camarasa, C. (2018). Workflow based on the combination of isotopic tracer experiments to investigate microbial metabolism of multiple nutrient sources. *J. Vis. Exp.*, 56393.
64. Delobel, P., Pradal, M., Blondin, B., and Tesniere, C. (2012). A 'fragile cell' sub-population revealed during cytometric assessment of *Saccharomyces cerevisiae* viability in lipid-limited alcoholic fermentation. *Letts. Appl. Microbiol.* *55*, 338–344.
65. Legras, J.-L., Ruh, O., Merdinoglu, D., and Karst, F. (2005). Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. *Int. J. Food Microbiol.* *102*, 73–83.
66. Kamvar, Z.N., Tabima, J.F., and Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* *2*, e281.
67. Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet. Res.* *67*, 175–185.
68. Excoffier, L., Smouse, P.E., and Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* *131*, 479–491.
69. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114–2120.
70. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* *20*, 1297–1303.
71. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* *25*, 1754–1760.
72. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* *81*, 559–575.
73. Raj, A., Stephens, M., and Pritchard, J.K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* *197*, 573–589.
74. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* *15*, 1179–1191.
75. Ramasamy, R.K., Ramasamy, S., Bindroo, B.B., and Naik, V.G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus* *3*, 431.
76. Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* *11*, 94.
77. Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* *27*, 3070–3071.
78. Lee, T.-H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* *15*, 162.
79. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* *30*, 1312–1313.
80. Huang, K., Ritland, K., Guo, S., Dunn, D.W., Chen, D., Ren, Y., Qi, X., Zhang, P., He, G., and Li, B. (2015). Estimating pairwise relatedness between individuals with different levels of ploidy. *Mol. Ecol. Resour.* *15*, 772–784.
81. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* *19*, 1655–1664.
82. Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* *23*, 1801–1806.
83. Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. *Nature* *461*, 489–494.
84. Wiser, M.J., and Lenski, R.E. (2015). A comparison of methods to measure fitness in *Escherichia coli*. *PLoS ONE* *10*, e0126210.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Newly sequenced <i>S. cerevisiae</i> genomes	This paper	European Nucleotide Archive accessions PRJEB36058
Experimental Models: Organisms/Strains		
<i>S. cerevisiae</i> strains	N/A	See Table S1
Chemicals		
Wheat peptone	Sigma-Aldrich	93492-500G-F
Di potassium hydrogenophosphate	Sigma-Aldrich	P3786
Potassium dihydrogen phosphate	Sigma-Aldrich	P5379
Magnesium sulfate heptahydrate	Sigma-Aldrich	M9397
Manganese sulfate (II)	Sigma-Aldrich	31425
TWEEN®80	Sigma-Aldrich	P1754
Citric acid monohydrate	Sigma-Aldrich	C0759
Glucose	Sigma-Aldrich	D8375
Maltose	Sigma-Aldrich	M0875
Thiamine	Sigma-Aldrich	T4625
Folic acid	Sigma-Aldrich	F7876
Nicotinamide	Sigma-Aldrich	N4126
Pantothenic acid	Sigma-Aldrich	P2250
pyridoxal-phosphate	Sigma-Aldrich	P9255-1
Cyanocobalamin	Sigma-Aldrich	47869
Yeast extract	BIOKAR Diagnostics	A1202 HA
Oligonucleotides		
GFPNATMXtoHO-for 5'- GCTATTGAGTAAGTTCGATCC GTTTGGCGTCTTT TGGGGTGT AACGCCAAGATCTGTTAGCTT GCCTTGTC	This paper	N/A
GFPNATMXtoHO-rev 5'- GAGCCCGCGGACAGCATC AAACTGTAAGATTC CGCCAC ATTTTATACACTCATGAATTCG AGCTCGTTGTC	This paper	N/A
Software and Algorithms		
Poppr version 2.8.2 (R package)	N/A	CRAN
Trimmomatic version 0.322	N/A	http://www.usadellab.org/cms/?page=trimmomatic
GATK version 3.6	N/A	https://gatk.broadinstitute.org/hc/en-us
BWA version 0.7.12	N/A	http://bio-bwa.sourceforge.net/
fastStructure version 1.0	N/A	https://rajanil.github.io/fastStructure/
Structure Plot	N/A	http://omicsspeaks.com/strplot2/
CLUMPAK	N/A	http://clumpak.tau.ac.il/
adegenet version 2.1.1 (R package)	N/A	CRAN
PLINK version 1.9-beta3j	N/A	https://www.cog-genomics.org/plink2
CLUMPP	N/A	https://web.stanford.edu/group/rosenberglab/clumpp.html
TreeMix	N/A	https://bitbucket.org/nygcresearch/treemix/downloads/

RESOURCE AVAILABILITY

Lead Contact

Requests for further information and for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Delphine Sicard (delphine.sicard@inrae.fr).

Materials Availability

Bakery strains used in this study are available upon request. See also [Table S1](#).

Data and Code Availability

The datasets and codes generated during this study are available at Zenodo under DOI: <https://zenodo.org/record/3759218>. The accession number for the newly sequenced genomes reported in this paper is European Nucleotide Archive accessions :PRJEB36058

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experiments were conducted using *S. cerevisiae*. All the yeast strains used in this study are listed in [Table S1](#).

Yeast collection and cultivation procedure

One hundred and twenty-nine strains were collected from sourdoughs in France as previously described^{43,58}. Twenty sourdough strains from Belgium isolated by⁵⁹ were kindly provided by the Belgium MUCL collection. Finally, 47 sourdough strains isolated by⁶⁰ in Italy were kindly provided by Fabio Minervini through the University of Perugia yeast collection and two sourdough strains from Sicily were kindly provided by Jean-Luc Legras²⁰. Thirty-one commercial strains were ordered from different international yeast collection or bought as starters.

Cultures were performed in 10 mL of liquid YE medium and were inoculated with yeast cells either from frozen stocks (stored at -80°C) or by picking a single colony from a previous culture plate.

METHOD DETAILS

Ploidy analysis by flow cytometry

Strain ploidy was analyzed by flow cytometry. Namely, 10^7 yeast cells, recovered at the beginning of stationary phase, were fixed in 70% ethanol for 16 h at 4°C . They were washed with 200 μL sodium citrate buffer (50 mM, pH 7) and then dispersed in 1 mL of the same buffer. 100 μL of cell suspension were transferred to a microtube and treated with 1 μL RNase A (100 mg/mL) for 2 h at 37°C . Labeling was performed by addition of 400 μL of a staining solution (50 $\mu\text{g}/\text{mL}$ propidium iodide in citrate buffer), and incubating for 40 min at 20°C in the dark. Cells were recovered by centrifugation and resuspended in 500 μL citrate buffer.

Flow cytometry analysis of 30,000 cells was carried out using the MACSQuant® Analyzer from Miltenyi Biotec GmbH (Germany) and the data analysis by the MACSQuant® software. A haploid strain (S288C) and a tetraploid strain (Levante³²) were used as calibration controls.

Fitness and fermentation analysis

GFP labeling

To study the impact of ploidy and strain origin (sourdough/commercial) on fitness, three diploid and three tetraploid strains were tagged with GFP. The pFA6a-TEF2Pr-eGFP-ADH1-NATMX4 plasmid⁶¹ was used as a template to amplify a cassette containing the TEF2 promoter, eGFP, ADH1 terminator and NATMX4 conferring resistance to clonNat. The PCR fragment obtained with primers GFPNATMXtoHO-for and GFPNATMXtoHO-rev was inserted into the HO locus of the selected strains. All *S. cerevisiae* strains used here are listed in [Table S1](#). The fitness and fermentation cost of carrying the GFP construction was assessed by competing the GFP-labeled and its unlabelled ancestral strain, as well as by comparing fermentation kinetics of these two strains. Relative fitness of the GFP-labeled strain relative to the unlabelled one was not significantly different from 1 (One-sample t test, two-sided, data not shown) and neither did GFP labeling change fermentation kinetics parameters (ANOVA, data not shown).

Competition and Fermentation conditions

Competition between GFP-labeled and unlabeled strains and single strain fermentations kinetics were performed in a synthetic sourdough medium (SSM) that was adapted from⁶² to better mimic sourdoughs' average composition. SSM contained, per liter: wheat peptone, 24 g; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.2 g; $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0.05 g; KH_2PO_4 , 4 g; K_2HPO_4 , 4 g; Tween 80, 1 mL; glucose, 15 g; maltose, 35 g; cobalamine, 0.2 mg; folic acid, 0.2 mg; nicotinamide, 0.2 mg; pantothenic acid, 0.2 mg; pyridoxal-phosphate, 0.2 mg and thiamine, 0.2 mg. The pH was adjusted to 4.5 with citric acid and the solution was sterilized by steam pasteurization for 15 min. Sterile-filtered vitamin and sugar solutions were added after pasteurization. Overnight pre-cultures in YPD were titrated with a C6 flow cytometer (Accuri, BD Biosciences). For competition experiments, $5 \cdot 10^5$ cells/mL of each strain were then inoculated in 15 mL SSM. For single strain analysis, 10^6 cells/mL of pre-culture were inoculated in SSM. Population dynamics were followed by measuring population size

over time using the flow cytometer. Fermentations were carried out at 24°C with constant magnetic stirring (300 rpm) in 20 mL glass tubes closed with a filter tip to allow CO₂ release. Fermentations were monitored during 24 h for CO₂ release, by measuring weight loss every 40 min using an automated robotic system⁶³. At the end of fermentation, cultures were centrifuged and pellets were re-suspended in PBS for flow cytometry analysis (C6 cytometer, Accuri, BD Biosciences). Population size and cell viability were determined as described in⁶⁴. Relative fitness of tetraploid versus diploid strains was estimated based on the proportion of GFP-labeled versus unlabelled strains in mixed cultures. GFP fluorescence (excitation 488 nm, emission 530 nm) was collected in the FL1 channel.

For growth on maltose assays, four independent overnight cultures of each yeast strain listed in [Table S1](#) were inoculated at 10⁶ cells/mL into 150 μL of SSM with maltose (50 g/L) as sole carbon source. Yeast cells were incubated at 24°C under continuous shaking in 96-wells microtiter plates, and growth was measured every 20 min for 24 h (OD₆₁₀) using a TECAN Infinite M200 Microplate Reader (TECAN, Männedorf, Switzerland).

Genomic DNA extraction

Cells from an overnight culture were recovered by centrifugation of 5 mL of culture medium at 4500 rpm for 3 min. The yeast pellet was suspended in a sorbitol solution containing 20 μL Zymolyase 20T (1 mg/mL). Cell wall lysis was performed at 37°C for 1 h. DNA extraction was then carried out using the DNAeasy Blood & Tissue Kit (QIAGEN).

Microsatellite typing and analysis

In order to study the genetic diversity of diploids and tetraploids, 15 microsatellites loci previously described in⁶⁵ were used: C3, C4, C5, C6, C8, C9, C11, ScAAT1, ScAAT3, SCYOR267C, YKL172w, YKR072c, YLL049, YLR, YPL009c. Multiplexing was used in order to amplify 4 or 5 microsatellites per PCR run, with two distinct markings, one at 700 nm and the second at 800 nm. Once amplified, amplicons were diluted to 1/20 in formamide, denatured for 5 minutes at 85°C, then separated by electrophoresis on a 13% polyacrylamide gel containing 39% urea in 1.2X TBE buffer at 2000 V for 15 h (50°C) on an automatic sequencer.

Analyses of microsatellite data were performed using Poppr version 2.8.2⁶⁶. A microsatellite genetic relatedness tree was constructed based on the Ritland relatedness coefficient⁶⁷. Analysis of Molecular Variance (AMOVA) was used as a method of estimating population differentiation⁶⁸.

Genome sequencing and read processing

Genome sequencing data were obtained for 68 yeast strains from numerous sources according to [Table S1](#). For the present study, 17 yeast genomes were newly sequenced in our laboratory: DNA samples were processed to generate libraries of short 400-bp inserts. After passing quality control, the libraries were sequenced using an Illumina HiSeq 2000 platform. Sequencing from both ends generated paired-end reads of 2 × 100 bp, resulting in a 100X average sequencing depth. This dataset was deposited in the European Nucleotide Archive (ENA) under study accessions number PRJEB36058. Trimming low quality regions and adapters in Illumina data was performed using Trimmomatic version 0.322⁶⁹ with sequencing parameters:

```
ILLUMINACLIP:adapterFile:2:30:7,LEADING:20, TRAILING:20, SLIDINGWINDOW:20:25, MINLEN:75.
```

Variant calling

We used the Genome Analysis Toolkit (GATK)⁷⁰ version 3.6 for SNP and indel calling. Briefly, the workflow is divided into four sequential steps: initial mapping, refinement of the initial reads, multi-sample indel and SNP calling, and finally variant quality score recalibration.

First, reads were aligned to the S288c reference genome (release number R64-1-1, downloaded from SGD) using BWA version 0.7.12⁷¹ resulting in aligned reads in a BAM file format.

Second, optical and PCR duplicates were removed using MarkDuplicate from the Picard Tools version 2.6.0 (<http://picard.sourceforge.net>). Base quality scores were recalibrated using BaseRecalibrator/PrintReads (GATK). These recalibrated scores in the output BAM file are closer to their actual probability of mismatching to the reference genome, and are subsequently more accurate. Moreover, the recalibration tool attempts to correct for variation in quality with machine cycle and sequence context. At the end of this step we obtained analysis-ready reads.

Third, we performed SNP and indel discovery using HaplotypeCaller (GATK) on each sample separately in BP_RESOLUTION mode, to produce an intermediate file format termed GVCF (for Genomic VCF). These per-sample GVCFs were then run through a joint genotyping step using GenotypeGVCFs (GATK) to produce a raw multi-sample VCF callset.

Fourth, we used Variant Quality Score Recalibration (VQSR) to build an adaptive error model (VariantRecalibrator tool) using an unpublished dataset of known SNPs and Indels obtained from 86 genomes²⁰. Then this model was applied (ApplyRecalibration tool) to estimate the probability that each variant in the callset is a true genetic variant or a machine / alignment artifact. This step assigns to each variant a VQSLOD score that is much more reliable than the raw QUAL score calculated by the caller. We used this variant quality score to filter the raw call set, thus producing a subset with our desired level of quality, fine-tuned to balance specificity and sensitivity. This genotyping pipeline resulted in a VCF file containing 302,290 biallelic SNPs and 21,045 indels discovered across 68 samples to which were associated a genotyping quality for each strain.

Population structure

The set of 302,290 biallelic SNPs sites identified above was further filtered by removing SNPs with missing genotypes above 0.10, minimum alternate allele frequency (MAF) below 0.03 and SNPs in linkage-disequilibrium using PLINK⁷² version 1.9-beta3j with a

window size of 50 SNPs, a step of 5 SNPs at a time and a r^2 threshold of 0.5. The resulting filtered dataset contained 32,379 SNPs positions.

Twenty independent runs of fastStructure version 1.0⁷³ were performed varying ancestral populations from 1 to 10 using the simple prior. The number of iterations varied from 10 ($K = 1$) to 70 ($K = 10$). The highest likelihood was obtained for 6 ancestral populations (likelihood: -0.75599 ; total iterations: 30). CLUMPAK⁷⁴ was used for analyzing the result of multiple independent runs of fastStructure. CLUMPAK identifies an optimal alignment of inferred clusters across different values of K , simplifying the comparison of clustering results across different K values. Structure plots were obtained using the interactive web application Structure Plot⁷⁵.

Stratification was also studied using discriminant analysis of principal components (DAPC)⁷⁶ a multivariate method designed to identify clusters of genetically related individuals. This analysis was performed using the R package adegenet version 2.1.1⁷⁷. DAPC was performed (function dapc) using clusters defined by K -means where we specified a number of clusters from 4 to 7.

Phylogenetic tree imputation

The set of 302,290 biallelic SNPs sites was filtered by removing SNPs at positions with missing genotypes above 0.10 and minimum alternate allele frequency (MAF) below 0.10. The resulting filtered dataset contained 99,128 SNPs positions. The VCF file was converted into Fasta sequences using generate_snp_sequence.R from the R-package SNPhylo⁷⁸. A phylogenetic tree was computed with RAxML version 7.2.8⁷⁹ performing a complete analysis (ML search + 100 bootstrapping) using the GTRGAMMA evolution model.

For the microsatellite typing obtained from 229 baker yeasts, Ritland relatedness coefficient⁶⁷ was estimated using PolyRelatedness, a software able to estimating pairwise relatedness between individuals with different levels of ploidy⁸⁰. A neighbor joining tree was obtained using the nj function from the ape R package.

Analysis across the 1,011 genomes data

A genotyping matrix was constructed with the GenotypeGVCFs function of GATK that was run with 1,011 gvcf. files constructed in¹⁶ as well as the 26 bakery strains gvcf. files that were not included in this previous study (17 newly sequenced diploid sourdough strains genomes and a 9 previously sequenced bakery strains). This extended the dataset to a total of 68 bakery yeast genomes.

The neighbor-joining tree was constructed with the R packages ape and SNPrelate. The gvcf. matrix was first converted into a gds file and individual dissimilarities were estimated for each pair of individuals with the snpgdsDiss function. The bionj algorithm was then run on the distance matrix that was obtained.

A set of 552,093 biallelic SNPs was obtained from the gvcf. matrix selecting 157 genomes (including 68 bakery strains; Table S1). SNPs were further filtered by removing SNPs with missing genotypes above 0.10, minimum alternate allele frequency (MAF) below 0.03 and SNPs in linkage-disequilibrium using PLINK version 1.9-beta3j with a window size of 50 SNPs, a step of 5 SNPs at a time and r^2 threshold of 0.5. The resulting filtered dataset contains 49,482 SNPs positions. Twenty independent runs of Admixture version 1.3.0⁸¹ were performed varying ancestral population from 2 to 20. The value of $K = 17$ exhibited the lowest cross-validation error compared to other K values. CLUMPAK⁸² was used for analysis the results of multiple independent runs of Admixture. Structure plots were obtained using the interactive web application Structure Plot⁷⁵.

To evaluate the bakery strains admixture, we also used the TreeMix algorithm⁵¹, which builds population trees and tests for the presence of gene flow between populations. We estimated a maximum-likelihood tree rooted using the China population (CHNV), the likely ancestral population of *S. cerevisiae*¹⁷. To test for gene flow, we used the three-population f_3 test⁸³ as suggested by⁵¹.

Copy number variations

Reads were mapped using BWA (default parameters). *MAL1-3* and isomaltase genes copy numbers were expressed as the read mapping depth of the gene relative to that of a neighboring gene used as reference (expressed as a log₂ ratio).

QUANTIFICATION AND STATISTICAL ANALYSIS

All competition and single strain fermentations were carried out with 3 or 4 replicates.

The effect of ploidy was analyzed on relative and absolute fitness.

Relative fitness was calculated as:

$$w = \frac{\ln\left(\frac{A_f}{A_i}\right)}{\ln\left(\frac{B_f}{B_i}\right)}$$

where w is fitness, A and B are the population sizes of the two competitors, while subscripts i and f indicate the initial and final time point (24h) of the competition kinetics⁸⁴. To statistically compare the fitness of tetraploids and diploids, the mean relative fitness of tetraploids relative to diploids was compared to 1 using a one-sample t test, two-sided.

Absolute fitness was estimated using different proxy depending on the collection of strain. First we followed the population dynamic of a collection of six parental tetraploids and their six derived diploids spore and estimated the growth rate, carrying capacity, area under the population dynamic curve and mortality rate at the end of fermentation. We then tested the effect of ploidy using the following model: $Y_{ijkl} = \mu + ploidy_i + geneticbackground_j + ploidy \times geneticbackground_{ij} + Bloc_k + \epsilon_{ijkl}$. Second, we compared a collection of 19 tetraploids and 25 diploids bakery strains from sourdough or industrial origin and measured their population

size and mortality rate at the end of fermentation in SSM. We then used the following mixed linear model $Y_{ijklm} = \mu + ploidy_i + habitat_j + ploidy \times habitat_{ij} + Strain_k + Block_l + \epsilon_{ijklm}$, where Y_{ijk} is either the population size or the mortality rate, and habitat indicated strain origin s (sourdough/commercial).

The growth rate and carrying capacity in synthetic sourdough with maltose as sole carbon source were estimated by fitting an optical density curve thanks to the summarizegrowth R function from the growthcurver package. The effect of the number of maltose gene on growth was tested using the following model: $Y_{ijk} = \mu + habitat_i + CNV_j + \epsilon_{ijk}$

The effect of domestication was analyzed by comparing the fitness and fermentation kinetics of commercial and sourdough bakery strains to strains originating from wine, beer, asian fermentation and non-fermented product (called “other”). Population size and mortality rate after 27h of fermentation in synthetic sourdough media were used as proxies of absolute fitness. To analyze fermentation kinetics, the cumulative CO₂ production curve was calculated and the kinetics of CO₂ production rate over time was estimated. Four parameters were then estimated: the maximum CO₂ release (g), the fermentation latency-phase time (h) (time between inoculation and the beginning of the fermentation calculated as 1g of CO₂ release), the maximum CO₂ production rate Vmax (g/L/h) and the time of the maximum CO₂ production rate.

To test the “habitat of origin” effect (sourdough/commercial/wine/beer/asian fermentation/other) the following mixed linear model was used $Y_{ijk} = \mu + habitat_i + Strain_j + Block_k + \epsilon_{ijk}$, where Y_{ijk} is the fitness component or fermentation kinetics variable and habitat, the “habitat of origin” effect (sourdough/commercial/wine/beer/asian fermentation/other).

Ploidy or habitat were considered as fixed effects while Block and Strain were considered random effects. Tukey tests were used to compare pairs of means.