NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENTO DE CIÊNCIAS DA VIDA

THE GENOMICS OF MICROBE DOMESTICATION – TESTING THE HYPOTHESIS OF SECONDARY DOMESTICATION EVENTS IN *SACCHAROMYCES CEREVISIAE*

ANA PATRÍCIA TEIXEIRA PONTES

Mestre em Microbiologia Médica

DOUTORAMENTO EM BIOLOGIA

Universidade NOVA de Lisboa

Abril, 2022

DEPARTAMENTO DE CIÊNCIAS DA VIDA

THE GENOMICS OF MICROBE DOMESTICATION – TESTING THE HYPOTHESIS OF SECONDARY DOMESTICATION EVENTS IN *SACCHAROMYCES CEREVISIAE*

ANA PATRÍCIA TEIXEIRA PONTES
Mestre em Microbiologia Médica

Orientador: Doutor José Paulo Nunes de Sousa Sampaio,
Professor Associado com Agregação, FCT - NOVA

Coorientador: Doutora Paula Maria Theriaga Mendes Bernardo Gonçalves,
Professora Associada, FCT - NOVA

Júri:

Presidente: Doutora Isabel Maria Godinho de Sá Nogueira,
Professora Catedrática, FCT-NOVA

Arguentes: Doutor João José Inácio Silva,
Professor, University of Brighton
Doutor Raúl A. Ortiz-Merino,
Investigador, Katholieke Universiteit Leuven

Orientador: Doutor José Paulo Nunes de Sousa Sampaio,
Professor Associado com Agregação, FCT-UNL

Membros: Doutor Manuel José de Carvalho Malfeito Ferreira,
Professor Associado com Agregação, ISA-UL
Doutora Isabel Maria Godinho de Sá Nogueira,
Professora Catedrática, FCT-NOVA

DOUTORAMENTO EM BIOLOGIA

Universidade NOVA de Lisboa
Abril, 2022

The genomics of microbe domestication - Testing the hypothesis of secondary domestication events in *Saccharomyces cerevisiae*

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Professor José Paulo Sampaio, who welcomed me into the laboratory and whose help was very important during this work. I also would like to highlight the importance of his guidance, trust, and advice, which were fundamental for the execution of this project. I would also like to leave a special thanks to my co-supervisor, Professor Paula Gonçalves, for the several important scientific discussions and valuable suggestions.

To the members of my thesis advisory committee, Doctor Filomena Duarte and Doctor Marco Coelho, a thank you for the follow-up, discussions, opinions and suggestions during this work and also for the suggestions and the time dispensed during the reading of the thesis.

A sincere and warm appreciation to all my colleagues in the Yeast Genomics Lab, present and former, for the support, scientific discussions, and good times. A special thanks to Carla Gonçalves, who was a great colleague and friend, and from whom I learned so much.

I would also like to express my sincere appreciation to our lab technician Nicole Soares, who always ensured that everything that was needed was ready on time and in perfect conditions. Also, a word of appreciation to the Portuguese Culture Collection (PYCC) for providing and ensuring the quality of some of the strains used in the course of this work.

A heartfelt thanks to all my friends and family. A special mention to my parents and brother for their unconditional love and support. To Pedro, my most awesome husband, a special and warm thank you for always being there for me, supporting and cheering me on.

Finally, I would like to express my appreciation for the financial support of Fundação para a Ciência e Tecnologia that funded my PhD grant (SFRH/BD/136462/2018).

ABSTRACT

For thousands of years microbes have been used, typically inadvertently, for the fermentation of a wide diversity of beverages and foods, which has led to the selection of relevant characteristics, thus promoting the domestication process. The yeast *Saccharomyces cerevisiae* is probably the most well-studied microbe with respect to domestication. However, most studies have been focused on a few industrially important lineages such as that of wine yeasts. Thus, this work aimed at deepening and extending the study of domestication trajectories in *S. cerevisiae* employing a broad perspective and using an integrated approach. For this, bioinformatic and experimental analyses were combined for a balanced perspective of genetic and phenotypic diversity of *S. cerevisiae*.

In chapter 2, the emergence of a domesticated lineage associated with cachaça, and its relationship with wine yeasts was investigated. Our findings lead us to propose different domestication trajectories. We identified initial transitions from wild to domesticate, that gave rise to primary domesticated populations with specific traits related to the fermentation in which they were selected. We also identified secondarily domesticated populations that originate from primary domesticates that have been subjected to new and distinctive selective pressures. In chapter 3 we studied a population linked to wine yeasts but that arose after an inter-species hybridization. We analyzed in detail the consequences of an ancient hybridization between *S. cerevisiae* and *S. paradoxus* for adaptation to an anthropic environment. This study further highlighted the different layers of genomic and phenotypic transitions related to domestication and lead to the proposal of the concept of quasi-domesticate. Chapter 4 provided a comprehensive view of wild and domesticated populations of *S. cerevisiae*, integrating results from several recent publications. In that study, the distribution and ecology of wild populations was evaluated, and the evolution of certain genes that code for relevant phenotypic traits was investigated.

The detailed study of domestication trajectories in *S. cerevisiae* carried out in this work revealed unanticipated levels of complexity, with cases of secondary domestication and of quasi-domestication. The results gathered in this project contribute to a better understanding of the mechanisms underpinning *S. cerevisiae* domestication. This will not only enable a better understanding the biology of this model organism from a fundamental perspective but will also be relevant for applied domains like the rational improvement of industrial fermentations.

Keywords: Microbe domestication, *Saccharomyces cerevisiae*, evolution, population genomics, microbial ecology

RESUMO

Os microrganismos são utilizados desde há milhares de anos, normalmente inadvertidamente, na fermentação de inúmeras bebidas e alimentos, o que promoveu o processo de domesticação. A levedura *Saccharomyces cerevisiae*, é provavelmente o microrganismo mais bem estudado no que diz respeito à domesticação. No entanto, a maioria dos estudos têm-se focado em linhagens industrialmente relevantes, tais como as leveduras do vinho. Neste trabalho procurou-se aprofundar e alargar o estudo das trajectórias de domesticação em *S. cerevisiae*, usando uma abordagem abrangente e integrada. Foram combinadas análises bioinformáticas e experimentais de modo a obter uma perspetiva equilibrada da diversidade genética e fenotípica em *S. cerevisiae,* dando relevo tanto a populações domesticadas como a populações selvagens.

No capítulo 2, foi investigada uma linhagem domesticada associada com a cachaça e a sua relação com as leveduras do vinho. Os nossos resultados apoiam a tese de que ocorreram múltiplas trajectórias de domesticação nesta espécie. Identificámos transições iniciais de selvagens para domesticados, que deram origem a populações domesticadas primárias com traços específicos relacionados com as fermentações em que ocorreu a selecção. Identificámos ainda populações domesticadas secundárias que têm origem em domesticados primários sujeitos a novas e distintas pressões selectivas. No capítulo 3 foi estudada uma população ligada às leveduras do vinho, mas que se originou após hibridação inter-específica. Analisámos em pormenor as consequências desta hibridação ente *S. cerevisiae* e *S. paradoxus* na adaptação ao ambiente antrópico. Este estudo destacou ainda mais a complexidade das transições genómicas e fenotípicas relacionadas com a domesticação, levando à proposta do conceito de "quasi-domesticate". O capítulo 4 focou-se numa visão abrangente tanto de populações selvagens como domesticadas. Nesse estudo, foi avaliada a distribuição e ecologia das populações selvagens, e investigada a evolução de certos genes que codificam traços fenotípicos relevantes.

O estudo detalhado das trajectórias de domesticação em *S. cerevisiae* realizado neste projecto contribuiu para suportar a tese de que este processo é complexo e multidimensional, envolvendo vários tipos de transições. Salienta-se que uma melhor compreensão dos mecanismos subjacentes à domesticação de *S. cerevisiae* permitirá não só compreender melhor a biologia deste organismo numa perspetiva fundamental, como também a sua aplicação mais racional em processos fermentativos industriais.

Palavras chave: Domesticação microbiana, *Saccharomyces cerevisiae*, evolução, genómica populacional, ecologia microbiana

INDEX

FIGURE INDEX

TABLE INDEX

ABBREVIATIONS

BC	Before Christ
bp	base pair
DNA	Deoxyribonucleic acid
CFU	Colony-forming unit
CNV	Copy number variation
GO	Gene ontology
HGT	Horizontal gene transfer
HPLC	High-performance liquid chromatography
kb	kilobase pair
NCBI	National center for biotechnology information
OD_{640nm}	Optical density at 640 nm
ORF	Open reading frame
PCR	Polymerase chain reaction
rpm	Rotations per minute
SGD	Saccharomyces genome database
SGRP	Saccharomyces genome resequencing project
SNP	Single nucleotide polymorphism
YNB	Yeast nitrogen base
w/v	Weight per volume

Abbreviations of relevant culture collections

CBS	Centraal Bureau voor Schimmelcultures, Utrecht, The Netherlands
JCM	Japan Collection of Microorganisms, Riken Bioresource Center, Saitama, Japan
NRRL	Agricultural Research Service Culture Collection (ARS), Northern Regional Research Laboratory, Peoria, Illinois, USA
PYCC	Portuguese Yeast Culture Collection, DCV-UCIBIO, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal

1.
General Introduction

1.1. Microbe domestication

Throughout history and across the world, human societies have developed fermented foods and beverages by taking advantage of locally available fermentable substrates. The widespread production of fermented drinks can be explained in part by the analgesic and disinfectant effect of ethanol. It is well-documented that alcoholic beverages have contributed to reinforce social interactions and rituals due to their pleasantness and ability to alter consciousness states. Moreover, they provided a safe source of water, devoid of disease-transmitting microbes. On the other hand, fermentation also helped to preserve and to increase the nutritional value and taste of foods and beverages. Therefore, because of the diverse beneficial effects, from pharmacological to sensory and nutritional, fermented beverages and foods have played a crucial role in the cultural and technological development of humankind (McGovern et al., 2004).

Originally, centuries or even millennia ago, beverage and food fermentation occurred spontaneously, driven by naturally contaminating microorganisms. Humans repeated and attempted to control these processes in order to maintain and in some cases to improve the final product, which resulted in well-established fermentation techniques. A critical step in many fermentations is the back-slopping, which consists in the continued passage of the dominating culture from an old fermentation batch to a new one, by adding a portion of the fermented product to a new, unfermented batch. By implementing techniques such as back-slopping and therefore propagating and improving desirable characteristics for long periods of time, humans have unwittingly promoted the adaptation and domestication of various

microorganisms to different substrates such as fruits, cereals, milk, among others (Gibbons & Rinker, 2015). Microbes co-opted in such processes have become genetically and phenotypically distinct from their ancestors and it is now clear that they are essential for successful fermentations. As such, they fit the definition of domesticate used in this work, which follows that of Diamond, 2002: an organism bred in captivity and thereby modified from its wild ancestors in ways making it more useful to humans who control its reproduction and its food supply. This broad definition of domestication was originally proposed for plants and animals. However, it also applies to microorganisms since in man-made fermentative niches, humans have control over the nutrients available and may even have some influence over reproduction, even if not directly or intentionally. Moreover, due to adaptation to these artificial environments, microorganisms acquired genetic and phenotypic changes and became distinct from their wild ancestors in ways that had a positive impact in the characteristics of the fermentative transformations favored by humans.

Even though the times when the first wild microbes were selected for food fermentations remains obscure, the domestication of bacteria, yeasts, and filamentous fungi was most probably promoted by the seasonal abundance of foods in the Neolithic era. This was a direct consequence of agriculture and of cattle breeding and led to the necessity of avoiding the deterioration of perishable resources, together with the interest in improving digestibility and flavor. For example, the production of cheese and of yogurt-like products depends on the cooperation of specific microorganisms to break down lactose into lactic acid. This process made these products more easily digested and promoted a longer shelf-life. Equivalent benefits are observed for fruits and cereals, being notable examples wine, beer and bread (Papadimitriou et al., 2015). The first archaeological evidence for cheese-making dates back to 6,000 BC in northern Europe, where milk fat was identified in sieve vessels (Salque et al., 2013). Also, records of a fermented beverage consisting of a mixture of rice, honey and fruit were identified in China dating back to 7,000 BC (McGovern et al., 2004). With respect to wine-making, archaeological evidence points to an origin in Mesopotamia and Greece around 5,000 BC, one of the testimonials being a pottery jar from Iran that contained tartaric acid from grapes and terebinth as flavoring ingredient (McGovern et al., 1996).

In line with what was observed previously for animals and plants, it was proposed that wild microbes, including bacteria, yeasts, and filamentous fungi were 'tamed' into the industrial organisms we use today. This transition from complex and competitive natural environments to more stable anthropogenic niches provided the perfect setting for artificial selection to act upon, promoting the shift from generalist life-strategies to specialist ones.

Lactic Acid Bacteria (LAB) and filamentous fungi will be briefly introduced in this section to provide a general overview of the diversity of domesticated microbes. Yeast domestication will also be described in general terms, but the particular case of *Saccharomyces cerevisiae* will be reviewed in more detail in subsequent sections.

1.1.1. Lactic acid bacteria

Lactic acid bacteria were most likely among the first microorganisms used by humans. These bacteria are characterized by their ability to produce lactic acid from simple sugars. This functional classification includes a variety of industrially important genera, like *Lactococcus*, *Enterococcus*, *Oenococcus*, *Pediococcus*, *Streptococcus*, *Leuconostoc*, and *Lactobacillus*. This apparently simple metabolism of LAB was used in fermentations to improve the taste and texture of products, as well as to protect them from spoilage by producing antimicrobial peptides and lowering the pH due to the accumulation of lactic acid. Domestication of LAB occurred, as in many other microorganisms, due to the continuous passage through human-made niches, which eventually resulted in modern-day cultures capable of carrying out stable and safe fermentations. Nowadays, several LAB species have a key role in the food industry, performing bioconversions of dairy products, vegetables (e.g., sauerkraut) and even meat (e.g., sausages). Moreover, they are also relevant in other fermentations like those of wine, coffee, sourdough, among others (Makarova et al., 2006).

Lactobacillus and *Lactococcus* are two of the LAB genera most commonly used in the production of fermented dairy products, including cheese, yoghurt, sour cream, buttermilk, and others. In both genera several modifications associated with adaptation to human-made environments have been identified, the most striking being the loss of several genes related to metabolic pathways of carbohydrates found in natural habitats but not in milk. Thus, in domesticated strains genome reduction has been observed, with loss or pseudogenization of several genes that are only relevant in natural environments. In addition, gain of function by horizontal gene transfer or other processes has been documented. Examples are the ability to utilize casein, the predominant protein in milk, and the efficient assimilation of lactose (Gibbons & Rinker, 2015; Makarova et al., 2006; Steensels et al., 2019). Similarly, to what is observed in dairy environments for *Lactobacillus* and *Lactococcus*, another LAB, *Oenococcus oeni*, that occurs in wine fermentations, also underwent a major genome reduction. This species promotes deacidification by converting L-malic acid (dicarboxylic) into L-lactic acid (monocarboxylic), the malolactic fermentation. The genes lost during wine-adaptation are involved in the mismatch repair system. This loss resulted in a higher mutation rate, which has been described as an adaptive trait to wine-related environments (Steensels et al., 2019).

1.1.2. Filamentous fungi

Apart from bacteria, filamentous fungi have also been extensively used in food and beverage production and, as such, several adaptations associated with domestication have also been identified. Some species of *Rhizopus* are used in the production of different alcoholic beverages and tempeh, *Monascus purpureus* is used to make red yeast rice, *Penicillium* species are employed during the production of various types of cheese, and *Aspergillus* species have a key role during the production of some traditional alcoholic beverages, sauces, and condiments.

A good example of a domesticated mold important to produce an alcoholic beverage is *Aspergillus oryzae*, a major participant in sake production. For producing sake, *A. oryzae* is inoculated on steamed rice to break down the starch into glucose, which is subsequently fermented by *Saccharomyces cerevisiae*. *Aspergillus oryzae*, a domesticated species, originated from *Aspergillus flavus*, a species that produces an aflatoxin and that is associated with major agricultural losses (Gibbons et al., 2012). It is though that *A. oryzae* emerged from a lineage of *A. flavus* that lost the ability to produce the toxin. This lineage was selected by sake producers because of its beneficial cooperative effect with *S. cerevisiae*, since the yeast is also sensitive to the toxin (Steensels et al., 2019). In addition to the lack of toxicity, *A. oryzae* also possesses a higher number of copies of a gene that codes for an alpha-amylase, which is responsible for breaking down starch into glucose. It has also been shown that this is the most expressed gene in *A. oryzae*, highlighting its relevance (Gibbons et al., 2012; Gibbons & Rinker, 2015; Steensels et al., 2019).

Penicillium species are ubiquitous fungi with an important role in biotechnological, biomedical, and food industries. Within the food industry and more particularly in cheese production, *P. camemberti* and *P. roqueforti* play a crucial role. *Penicillium camemberti* forms white-colored cultures and is used for the ripening of soft cheeses, like camembert and brie. This mold results from a selection process whose purpose was the improvement of cheese texture and color. This species has never been isolated outside dairy-associated environments and seems to have derived from *Penicillium commune*, a blue-greyish fungus. Recent studies indicated that *P. camemberti* was selected for its white color and fluffy smooth texture in the late 19th century (Dupont et al., 2017). Unlike *P. camemberti*, *P. roqueforti* has a wider distribution, being found in food environments, including cheese, but also on silage and natural environments. This species is used in the production of most blue cheeses but, given its ability to tolerate low temperatures, low oxygen concentrations and weak acids used as food preservatives, it is also a common spoilage agent in refrigerated foods, meats, and silage (Cheeseman et al., 2014; Dupont et al., 2017). *P. roqueforti* is a species that has been

participating in blue cheese production for a long time but originally it was not inoculated, but rather occurred spontaneously. Then, in the early 19th century, the technique of collecting spores of *P. roqueforti* from bread, purposely left to rot on shelves, and transferring them to cheese was developed. Later, bread was inoculated with spores from batches of cheeses that had the desired taste and texture, which represented a type of back-slopping. During the last 30 to 40 years, this process has been more controlled with the emergence of commercialized cultures, promoting consistency and safety of the final product. By the analysis of the genome of *P. camemberti* and *P. roqueforti*, as well as other closely related *Penicillium* species, it was possible to find multiple horizontal gene transfers, which appear to be associated with adaptation to cheese. In particular two regions with hundreds of kilobases, named *Wallaby* (Cheeseman et al., 2014) and *CheesyTer* (Ropars et al., 2015), were shown to have almost 100% nucleotide identity among all *Penicillium* species inhabiting cheese production niches and absent in species that are not isolated from these environments. These regions appear to play an important role in adaptation to cheese, since, in the case of the *Wallaby* region, some of the genes encode for proteins previously described to be involved in interactions with other organisms, thus increasing the competitiveness in these environments (Cheeseman et al., 2014). The *CheesyTer* region, on the other hand, has genes that code for a lactose permease and a beta galactosidase, which have important roles during the early stages of growth on cheese (Ropars et al., 2015). Growth and competition assays further highlighted the role of these regions in cheese production. *Penicillium roqueforti* strains possessing these regions grow faster in cheese-bound environments than strains lacking them. Together these studies revealed exemplary cases of microbe domestication involving recurrent horizontal gene transfer events that occurred during adaptation to anthropogenic environments (Dumas et al., 2020; Dupont et al., 2017).

1.1.3. Yeasts

Yeasts are the group of microorganisms that have been most studied with respect to domestication. However, studies have focused essentially on *S. cerevisiae* leaving the so-called 'non-conventional yeasts' in the background. Although non-*Saccharomyces* yeasts do not have the same recognition, it is nowadays established that many take part in human-driven fermentations. Some examples are *Brettanomyces bruxellensis, Kluyveromyces lactis* and *K. marxianus*, *Lachancea thermotolerans*, and *Torulaspora delbrueckii*.

As for *S. cerevisiae,* the complete genome sequencing of non-conventional yeasts was a landmark for the deepening of studies about the evolutionary changes and mechanisms shaped by human action. For example, *B. bruxellensis,* a species that occupies many of the anthropic niches of *S. cerevisiae,* some studies point out that it has been domesticated (Roach & Borneman, 2020). This species is mainly linked to the production of lambic beer, where its use has increased in recent years in the craft brewing industry (Basso et al., 2016), and to the spoilage of many beverages due to the production of many off-flavor metabolites (Roach & Borneman, 2020). Besides its presence and importance in some anthropic niches, some genetic markers, like increased copy number of certain genes and horizontal gene transfer, have been identified. Nonetheless, no wild isolates or populations have been described so far, thus preventing a comparison between domesticated and wild and consequently an accurate assessment of the possible domestication traits (Curtin & Pretorius, 2014; Roach & Borneman, 2020). This lack of information on wild populations and natural ecology is a general gap for yeast species that cohabit with *S. cerevesiae* in fermentative environments. A probable exception is *T. delbrueckii*, which has been isolated from both natural and anthropic environments (Albertin et al., 2014; Carvalho et al., 2018; Limtong & Koowadjanakul, 2012). *Torulaspora delbrueckii* has been for a long time associated with winemaking, although generally fails to complete wine fermentation, given that it cannot consume all the sugars. However, in mixed fermentations with *S. cerevisiae, T. delbrueckii* reduces volatile acidity in high sugar fermentations helping to increase the sensory complexity of the final product. Besides its potential in wine production, *T. delbrueckii* is also used in bread fermentation, since it has a high osmotic and freeze tolerance (Alves-Araújo et al., 2004). As mentioned before, in addition to anthropogenic niches, this species has also been isolated in nature, from soil, plants, fruits, and even insects. Recently, a study using microsatellites differentiated several populations that split according to the substrate of isolation (Albertin et al., 2014). Although this study did not focus on genetic traits linked to domestication, it did reveal the existence of wild and anthropic populations, launching a line of research for studies about possible domestication events in this species.

In summary, although recent studies have suggested that several non- *Saccharomyces* yeasts might have adapted to anthropic fermentative niches, representing therefore additional cases of yeast domestication, detailed studies on their evolutionary trajectories and the correspondent genomic mechanisms are still scarce.

1.2. *Saccharomyces cerevisiae*

The most famous budding yeast, *S. cerevisiae,* a model organism for numerous areas of Biology, is in many instances treated as a synonym of "yeast". The genus *Saccharomyces* was described by Rees in 1870, and since then has undergone several taxonomic changes. Over the years, several species were classified in this genus based on morphological and physiological characteristics, and even a division between species belonging to the so-called *Saccharomyces sensu lato* and *Saccharomyces sensu stricto* groups, the latter including *S. cerevisiae* and its closest relatives, was made. Only after 2003 did the genus *Saccharomyces* begin to resemble more the one we know today and the denominations *sensu lato* and *sensu stricto* were abandoned (Alsammar & Delneri, 2020; Boynton & Greig, 2014; Kurtzman & Robnett, 2003). The genus currently consists of eight species, *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. jurei*, *S. kudriavzevii*, *S. arboricola*, *S. eubayanus* and *S. uvarum* (Figure 1.1). Some of these species are parents of hybrids that appear to have originated mostly in human-related environments. A well-known example of one of these hybrids, due to its commercial importance, is *S. pastorianus*. This hybrid is used worldwide to produce lager beer and resulted from a cross between *S. cerevisiae* and *S. eubayanus* (Libkind et al., 2011). In addition, other hybrids have also been found, mainly in association with other types of fermentations. Some are also used industrially mostly for cider and wine production. Besides *S. cerevisiae* they combine the genomes of *S. uvarum* and *S. kudriavzevii* (Alsammar & Delneri, 2020; González et al., 2006; Sampaio & Gonçalves, 2017) (Figure 1.1).

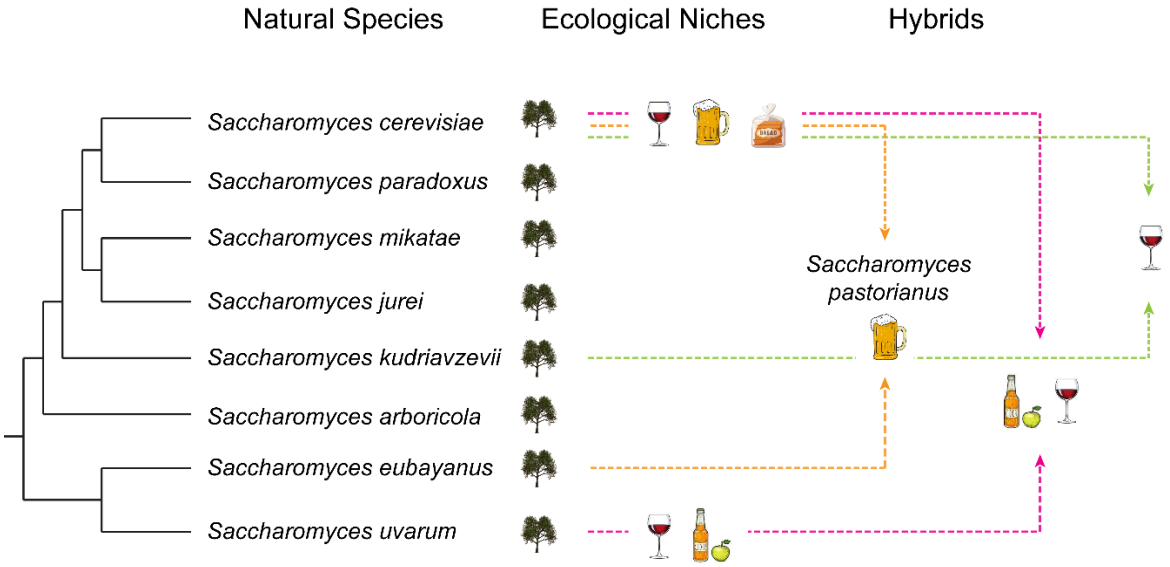


Figure 1.1 - Schematic representation of the phylogenetic relationships between the species of the genus *Saccharomyces* together with some industrially important hybrids. The ecological niches of the different species are indicated as wild (arboreal) or anthropic (human-driven fermentations).

Since the early days of yeast research up to the present, the fermentative character of domesticated strains of *Saccharomyces* has been intensively studied and consequently environments associated with these fermentations, such as vineyards and breweries, have been extensively sampled. However, most species of the genus are natural species, in the sense that they have been only isolated from environments unrelated to human activity. *Saccharomyces paradoxus*, *S. jurei*, *S. kudriavzevii*, *S. arboricola* and *S. eubayanus* are examples of strictly wild species, which means that, to date, they have only been isolated from natural niches. On the other hand, *S. cerevisiae* and *S. uvarum* are isolated from both wild and anthropic environments (Alsammar & Delneri, 2020; Sampaio & Gonçalves, 2017) (Figure 1.1).

Saccharomyces cerevisiae has long been regarded as a unique microbe, viewed by some as exclusively domesticated (Naumov, 1996; Vaughan Martini & Martini, 1995) and therefore remarkably distinct from its wild siblings. This view was supported by the traditional use of *S. cerevisiae* for centuries in baking, brewing, and winemaking, which justified the classical designations in the industry and in the scientific literature of 'baker's yeast', 'brewer's yeast', 'distiller's yeast' and 'wine yeast'. This intimate connection to human activity is at the base of Pasteur's discovery of the role of microbes in alcoholic fermentations in 1857 (reviewed in Landry et al., 2006). The research into the production of beers with different features was also one of the drivers of yeast genetics. The first crosses were conducted by Winge at the Carslberg laboratories in an attempt to produce strains with improved fermentative traits by combining the features of different strains (reviewed in Barnett, 2007; Boynton & Greig, 2014). The ease in controlling and manipulating the life cycle of *S. cerevisiae*, made the budding yeast the most coveted and employed eukaryotic model, used in countless studies aiming at answering the most diverse biological questions.

At the dawn of *S. cerevisiae* genome research, a culture designated S288c gained relevance. This laboratory strain was obtained by crossing several natural isolates and its main characteristic was a stable haploid state, which greatly facilitated the study of mutation effects. In 1996, S288c became the first representative of an eukaryotic organism to have its genome completely sequenced (Goffeau et al., 1996) which, in addition to the small genome size and availability of molecular genetics tools for *S. cerevisiae*, motivated pioneering work in functional biology. However, the reference strain provides little information on the natural history of this species since it has been bred in the laboratory and has not been exposed to environmental selective pressures. Through phenotypic studies, among others, it has been possible to determine that the reference strain is an outlier, behaving differently from most strains of this species (Warringer et al., 2011). Also, the presence of auxotrophic markers, mutations that prevent the cell from synthesizing essential compounds, although very useful for genetic engineering, has serious consequences for many traits, enforcing the idea that S288c is not a trustworthy reference of the species (Hittinger, 2013; Landry et al., 2006; Liti, 2015).

1.2.1. Life cycle

Among eukaryote microbes, the life cycle of *S. cerevisiae* is one of the most well-studied and understood in laboratory conditions. The ability to adapt its life cycle by switching between mitotic and meiotic reproduction is one of the important factors that contribute to its popularity. Paradoxically, little is known about the life cycle of *S. cerevisiae* in natural environments and what are the conditions that promote the transition from mitosis to meiosis in natural settings. Concisely, in the laboratory, diploid cells reproduce mitotically in rich media, but when they encounter nutrient scarcity (nitrogen depletion and poor carbon sources, such as acetate), these cells undergo meiosis and produce four haploid spores (ascospores) that are enclosed within a sac (ascus). Most asci contain four ascospores, two of each mating type (a and α). Ascospores are resistant structures that can endure diverse environmental stresses such as high and low temperatures and in the particular case of laboratory conditions, dissection. When nutrients are replenished, the ascospores germinate into haploid cells. The haploid cell can then divide mitotically, but usually after germination a cell fuses with another haploid cell of the opposite mating type giving rise to a diploid cell. Most crosses occur between haploid cells originating from the same meiosis, a form of self-fertilization known as intra-tetrad mating. Mating can also occur between haploids of different tetrads, which can be more or less related (inter-tetrad mating). Haploid cells can also change their mating type following mitotic division, allowing them to cross with their clone partners (autodiploidization) to form perfect homozygotic diploids. Another, albeit rarer, possibility for a haploid to return to the diploid state is to cross with an unrelated individual, which is called outcrossing (Boynton & Greig, 2014; Herskowitz, 1988).

Natural niches are not stable and have important oscillations in critical parameters such as temperature and nutrients, which could influence reproduction and proliferation. It is therefore thought that cells under these conditions spend a long time in a non-division state, called quiescence, or in spores (Liti, 2015). When conditions become favorable and there is again abundance of nutrients the cells return to their mitotic clonal growth (in the case of quiescent cells) or germinate (in the case of spores). Population genomic studies indicate that *S. cerevisiae* reproduces mostly asexually. However outcrossing can also occur, even if rarely (reviewed in Liti, 2015). Some experiments have shown that outcrossing, although apparently rare, can be promoted by another model organism, the fruit fly *Drosophila melanogaster*, which is attracted to substrates where *S. cerevisiae* is found, promoting not only diversification through mating but also dispersal (Christiaens et al., 2014; Reuter et al., 2007). Adding to the constrainments of natural environments and vectors, domestication also had a profound effect

in the normal life cycle of this species and domesticated strains tend to have a lower sporulation capacity (De Chiara et al., 2020).

1.2.2. Ecology

The close association of *S. cerevisiae* with fermented beverages contributed strongly to promote the concept that it was a 'man-made organism', restricted to anthropogenic environments (Naumov, 1996; Vaughan Martini & Martini, 1995). Studies on yeast ecology in the 1970's and subsequent decades, led by H. Phaff, did not employ enrichment techniques in an attempt to not introduce a bias in the isolation procedure. Such studies revealed that *S. cerevisiae,* although prevalent in wine fermentations, was not dominant in vineyards but could be recovered in moderate levels from damaged berries (Mortimer & Polsinelli, 1999). Such studies initiated a debate on the ecology of *S. cerevisiae* that lasts until today. A good example of the ongoing discussions is the disruptive hypothesis that *S. cerevisiae* is a nomadic species with no niche (Goddard & Greig, 2015). Adding to this debate, several studies suggest that different *S. cerevisiae* lineages are found in anthropic and in natural niches (Almeida et al., 2015; Duan et al., 2018; Peter et al., 2018).

As already mentioned, *S. cerevisiae* is found in many anthropic fermentative environments and for many years these environments were the sole focus of sampling. These environments range from wineries, breweries, bakeries to sake (or similar) production facilities in Asia. Beside these more industrial environments, *S. cerevisiae* is also commonly found in traditional fermentation settings around the world. Moreover, this species is also found among the commensal or opportunistic pathogens associated with humans. Several studies, have found *S. cerevisiae* as an agent of infection commonly associated with *Candida* infections particularly in immunocompromised patients (Muller et al., 2011; Strope et al., 2015). Clinical strains seem to have adaptations to the human body, such as being able to grow at higher temperatures and pseudohyphae formation. In addition to the association with pathogenicity, *S. cerevisiae* may also have a commensal relationship with humans. A study conducted in an Amerindian community inhabiting a remote area in French Guiana showed that in these individuals *S. cerevisiae* was more commonly isolated in stool, contrary to what was observed in more industrialized countries where *Candida albicans* predominated (Angebault et al., 2013). These gut colonizers most likely derive from the fermented food and drinks consumed, being of notice their ability to tolerate the stresses associated with the gut environment. However, such association is not novel as a famous *S. cerevisiae* strain, originally

called *S. boulardii* (Mcfarland & Bernasconi, 1993) is reported to have the same ability, being used by the pharmaceutical industry as a probiotic in the treatment of intestinal disorders (Pais et al., 2020; Zanello et al., 2009).

Despite being found associated with almost all aspects of human life, *S. cerevisiae* is also isolated from natural environments, far away from human intervention. In nature, *S. cerevisiae* is usually linked with oak trees (*Quercus* spp., family Fagaceae). This appears to be both the natural reservoir of *S. cerevisiae* and of several of the other species of the genus that have been sampled from this type of substrates across several continents (Sampaio & Gonçalves, 2008; Sniegowski et al., 2002). Additional niches include other trees of the Fagaceae family (such as beech and chestnut), other plants and also insects (Naumov et al., 2013; Stefanini et al., 2012). This association with oak trees is not easy to understand given that *S. cerevisiae* is usually associated with sugar-rich environments, and several adaptations to a lifestyle of rapid fermentation of abundant simple sugars are well-known. As such, a dominant natural association with low-sugar environments is puzzling and the exact mechanisms that *S. cerevisiae* uses to thrive in these environments are unknown. Besides this unexpected ecology, its detection in nature is not easy. For the isolation of *S. cerevisiae* from nature, usually an enrichment medium with ethanol is used, in order to favor the growth of *S. cerevisiae* in detriment of other microorganisms that may coexist in the same sample and could outcompete it (Sampaio & Gonçalves, 2008; Sniegowski et al., 2002). Thus, although most studies in the past have been based on the isolation of this species from anthropic environments, more recently a change in this trend has been observed with more and more studies exploiting the natural distribution *S. cerevisiae* (Duan et al., 2018; Peter et al., 2018; Wang et al., 2012).

1.2.3. Population genomics

Population genomics has provided a powerful means for understanding the evolutionary history of *S. cerevisiae*. In recent decades several studies based on multiple gene sequencing, microsatellite analysis, restriction-site-associated sequencing (Rad-seq) and more recently whole genome sequencing with long and short reads, have helped to further this knowledge. Early genome sequencing studies using *S. cerevisiae* isolates from different substrates led to the identification of five distinct lineages based on their technological (Sake and Wine/European) and geographical (West Africa, Malaysia and North America) origin (Liti et al., 2009; Schacherer et al., 2009). In addition to these pure lineages, i.e., populations

where the phylogenetic relationships with other populations remain the same throughout the genome regardless of the region that is compared, mosaic strains have also been identified. Mosaic strains represent situations in which gene flow through recombination is recurrent, and a signal of mixture of several strains or lineages is observed at the genome level. Mosaic strains tend to be associated with environments linked to human activity, such as bread production, food spoilage, spontaneous fermentations, and even clinical cases. Such environments appear to provide the opportunity for interbreeding (Dujon & Louis, 2017).

Some *S. cerevisiae* strains are characteristic of specific types of fermentations such as sake or wine fermentations and represent the first cases of domesticated populations that were identified at the genetic level (Fay & Benavides, 2005). For the case of wine strains, hereafter designated as the Wine population, they do not follow strict geographical boundaries despite being genetically related. Instead, these strains are dispersed across several wine-producing regions of the globe, reflecting the history of human migration and maritime trade (Goddard et al., 2010; Legras et al., 2007; Liti et al., 2009). The composition of the fermentative environments from which strains are isolated can vary greatly. This substrate variability has contributed to select different strains and genotypes that can be organized along populations. These distinct genotypes appear to be adapted for different fermentations and constitute separate populations referred to in the literature as the Wine, Beer or Sake populations (Fay & Benavides, 2005; Gonçalves et al., 2016; Liti et al., 2009). Furthermore, different genetic signatures of adaptations linked to these different anthropic environments have been reported (Marsit & Dequin, 2015; Novo et al., 2009) and will be discussed in more detail in the next section. The Wine population is one of the best understood domesticated populations. This population shows low genetic diversity in comparison with wild populations (Fay & Benavides, 2005), and it has been suggested that it results from a domestication event distinct from the one that originated the Sake population (Fay & Benavides, 2005), followed by a population expansion favored by human activity. A study that analyzed diverse natural strains, especially isolates associated with oak trees in Europe, detected a new population named the Mediterranean oaks (MO) population, which was identified as being the closest wild relative of the Wine population (Almeida et al., 2015). The estimated divergence between the two populations is consistent with the first historical evidence of wine production. Furthermore, this wild population does not present the genomic traits linked to domestication found in the Wine population (Almeida et al., 2015).

Unlike domesticated populations, wild populations tend to be restricted to a particular geographic area. Illustrative examples are oak-associated populations of North America and Japan (North America and Japan population), a population found in palm tree inflorescences in the Malaysian rainforest (Malaysian population) and the already mentioned population associated with Mediterranean oaks. In addition, a major sampling effort in China revealed several isolates that appear to possess a high genetic diversity in comparison to that already

known from populations in other regions (Wang et al., 2012). These findings spurred additional studies on phylogeography, which in turn led to the hypothesis of a single 'out-of-China' event as the most likely model to explain the observed natural diversity of *S. cerevisiae* (Bendixsen et al., 2021; Duan et al., 2018; Peter et al., 2018). According to the authors, the high genetic diversity observed in China, the highest observed in the species, helps to support this hypothesis. Duan et al. (2018), further proposed a single domestication event, where both European and Asian domesticates are derived from strains isolated in Chinese primeval forests. However, it is still unclear whether domestication had a single origin, in Asia, being those first domesticates subsequently introduced into Europe, or whether wild strains of *S. cerevisiae* dispersed from Asia to the Old World and the New World and domestication occurred independently multiple times (Dupont et al., 2017; Steensels et al., 2019).

1.3. **Adaptation to anthropic environments**

Various studies based on genomics and population genetics have revealed that *S. cerevisiae* has a highly complex population structure which, in part, appears to have been influenced by domestication. Human-made environments colonized by *S. cerevisiae* constituted a setting for the emergence of strains genetically and phenotypically distinct from their wild ancestors. One of the more important questions regarding domestication concerns the nature of the molecular modifications that have allowed the acquisition of traits valued by humans. In addition, *S. cerevisiae* is an excellent model for the study of genomic and phenotypic changes associated to anthropic environments since, besides encompassing a large panoply of domesticated populations that thrive in different fermentative environments, it also includes wild populations that allow for comparisons and the association of specific modifications to a given fermentation.

Several modifications associated to domestication have been described. These modifications can occur through different mechanisms. Small alterations at the nucleotide level (such as insertions, deletions, or substitutions), which may alter protein structure, protein interaction or gene expression; or large genomic rearrangements (duplications, translocations, aneuploidies), which could also affect expression due to change of the genomic context or even by copy number variation (CNV). In addition, cases of hybridization, introgression and even horizontal gene transfer (HGT) from other species play an important role by generating new fermentation-relevant features (Dunn et al., 2012; Marsit & Dequin, 2015) (Figure 1.2).

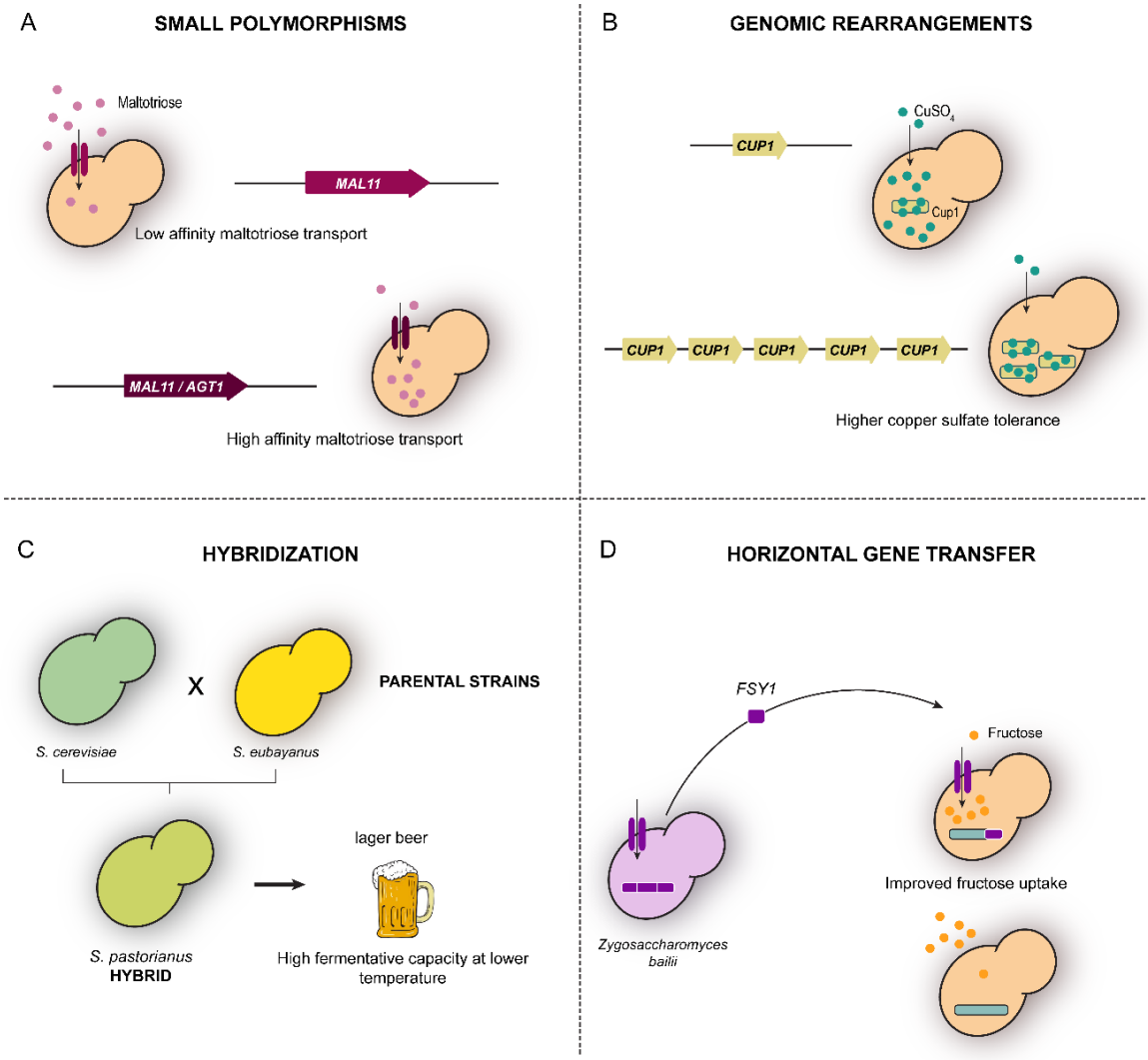


Figure 1.2 - Examples of mechanisms of adaptation to anthropic environments in *S. cerevisiae*. A - Polymorphism in the *MAL11* gene that originates a new allele (*MAL11 / AGT1*) able to transport maltotriose efficiently (Gallone et al., 2016); B - Amplification of the *CUP1* gene that allows a higher tolerance to copper sulfate by reducing the free copper in the cell (Marsit et al., 2017); C - Hybridization between *S. cerevisiae* and *S. eubayanus* that resulted in the hybrid species *S. pastorianus* that possess a higher fermentative capacity at lower temperatures (Libkind et al., 2011); D - Horizontal gene transfer of a fructose transporter (*FSY1*) from *Zygosaccharomyces bailii* enabling a better fructose assimilation (Galeote et al., 2011).

These different events, most of which have already been observed in the laboratory, have led to the emergence of numerous traits, such as resistance to toxic compounds, better acquisition of nutrients, greater ability to metabolize some carbon sources and a reduced production of unwanted compounds (Marsit et al., 2017).

1.3.1. Small polymorphisms

The generalization of whole genome sequencing made it possible to identify small changes that can lead to large phenotypic alterations, and that have been selected over time for their positive impact on traditional and industrial fermentations. Single nucleotide polymorphisms (SNPs) can have a major impact on the adaptation to fermentative environments (Marsit et al., 2017). These modifications can be substitutions that lead to loss of function or substitutions that lead to changes in function.

The *MAL11* gene, that codes for a maltose transporter, is a good example of how SNPs can have an important role during adaptation to anthropic environments. Maltose and maltotriose are the major sugars in beer wort. As such, a desirable trait in beer strains is the efficient assimilation of these two sugars. Brewer's strains have a version of *MAL11*, also known as *AGT1*, that allows active maltotriose utilization through an efficient transport (Figure 1.2A). The modifications in *MAL11* that led to the higher specificity for maltotriose have been selected for in brewing environments, and this trait has only been identified in this population, apart from some highly admixed mosaic strains (Gallone et al., 2016). Another example is the case of *HXT3,* a gene coding for a hexose transporter. Substitutions in the sequence of this gene have been associated to enhanced fructose transport (Guillaume et al., 2007). In wine must, fructose is present in high concentration, together with glucose, but *S. cerevisiae* prefers glucose over fructose. Fructose is usually left until the final stages of fermentation and when not consumed leads to stuck fermentations. The gain of an allele that contributes to the intracellular transport of this sugar has therefore a valuable positive impact in the wine industry. Therefore, it seems the modified *HXT3* transporter has been selected for in the wine population, where some of the industrial strains possess it (Guillaume et al., 2007; Marsit et al., 2017).

Domestication has also led to phenotypic variation through the acquisition of loss-of function SNPs and/or insertions and deletions causing frameshifts. One such case involves the AQY*1* and *AQY2* genes, which encode for aquaporins, i.e., water transporters. These water channels are critical in natural environments for freeze and thaw survival by rapidly transporting water to the outside of the cell preventing cellular shattering due to water crystallization. Will et al., 2010 showed that loss of function of these genes due to SNPs provides an adaptative advantage in environments with high sugar concentrations, where non-functional aquaporins increase fitness by helping to overcome the stressful effects of high osmolarity. In many domesticated strains the inactivation of these genes is indeed observed, which seems to correlate with the amount of sugar available in the fermentative environments where strains were isolated (Marsit & Dequin, 2015). However, the type of mutation causing

the inactivation differs between strains from different populations, suggesting that it happened independently several times, as a mechanism of adaptation to the different fermentations. In addition to domesticated populations with non-functional aquaporins, there is a wild population, the Malaysian population, that has a unique type of inactivation in the genes encoding for aquaporins (Will et al., 2010). The Malaysian strains were isolated from floral nectar of a palm tree that is rich in simple sugars. Therefore, inactivation of these genes appears to be associated not only with domesticated strains, but more broadly with strains isolated from high-sugar niches. A different example of gene inactivation associated with domestication is the case of *PAD1*, coding for a phenylacrylic acid decarboxylase, and *FDC1*, coding for a ferulic acid decarboxylase. These genes are involved in the decarboxylation of ferulic acid into 4-vinylguaicol, a phenolic compound with a very distinctive smoke-like aroma. The production of 4-vinylguaicol during beer fermentation is considered a phenolic off-flavor (POF). However, in some specific types of beer this aroma is a very desirable trait. Genomic studies revealed that these genes are inactivated in most beer strains, with the exception of a group of strains used in Belgian and German wheat beers where these flavors are desirable (Gallone et al., 2016; Gonçalves et al., 2016).

1.3.2. Genomic rearrangements

Several genome rearrangements associated with increased fitness in anthropic environments have been identified in *S. cerevisiae*. One of the best documented examples is related to sulfite resistance. This domesticated phenotype, common among strains of the Wine population, has been described to have evolved independently at least twice, through two distinct rearrangements. Both rearrangements result in a modification upstream of *SSU1*, a gene encoding for a sulfite pump. This modification leads to an increased expression of *SSU1* resulting in a higher resistance to sulfite, which is used as an antimicrobial agent in wine production. The first rearrangement to be described involving *SSU1*, consists of a reciprocal translocation between chromosome XVI (the chromosome where the native *SSU1* gene is located) to chromosome VIII (Goto-Yamamoto et al., 1998; Pérez-Ortín et al., 2002). The second case involves the translocation of this gene from chromosome XVI to chromosome XV (Zimmer et al., 2014). Both translocations have only been observed in the Wine population, with most strains possessing one of them, and being the XVI-VIII translocation the most frequent (Marsit & Dequin, 2015).

In addition to chromosomal translocations, copy number variations (CNVs), broadly defined as sections of the genome that differ in copy number, have also been implicated in the domestication process. These include gene deletions, gene duplications, and gene amplifications. These types of structural alterations are frequently involved in altered gene expression. One such alteration, again identified in the Wine population, is the amplification of the *CUP1* gene that codes for a copper-binding metallothionein (Figure 1.2B). Metallothioneins are a class of low molecular weight cytoplasmatic proteins rich in cysteine residues, that bind to metal ions, limiting or neutralizing their toxic effect. In *S. cerevisiae*, the *CUP1* gene encodes the most efficient copper-chelating metallothionein (Crosato et al., 2020; Fogel & Welch, 1982). The amplification of *CUP1* is associated with copper sulfate resistance and in wine strains seems to reflect a specific adaptation to the vineyard and winery environments given that copper sulfate has an important role protecting vines against powdery mildew (Marsit et al., 2017; Marsit & Dequin, 2015). Another example of gene amplification is found in the Beer population where the amplification of *MAL* genes has been noticed. Maltose metabolism is dependent on the *MAL* gene family, *MALT*, *MALS* and *MALR* comprising maltose transporters, maltases and regulators, respectively (Brown et al., 2010). Typically, beer strains contain six or more copies of the *MAL3* locus and in some German-style beer strains the number of copies can reach up to 15, while in wine strains the highest number is three (Gonçalves et al., 2016). Thus, the amplification of the genes involved in maltose uptake and breakdown, one of the main sugars in beer wort, is regarded as an adaptation to these environments, where the fast and efficient fermentation of this sugar is a desirable trait.

1.3.3. Hybridization

Sexual reproduction is associated with several advantages such as genetic recombination, which allows purging of deleterious mutations or combining beneficial mutations more efficiently. As in other organisms, mating in fungi is a tightly controlled process to ensure it only occurs in specific circumstances and between closely related individuals. This process is controlled by two main barriers, which can prevent either the formation of zygotes (pre-zygotic barrier) or zygote viability (post-zygotic barrier). However, these barriers can, in some rare occasions, be overcome, resulting in the formation of hybrids. Hybrids may reveal unique traits that are not necessarily intermediate between those displayed by the parental strains, and which may provide a selective advantage in a particular

environment. This phenomenon is known as hybrid heterosis or hybrid vigor (Gabaldón, 2020).

Saccharomyces hybrids have been commonly found in industrial fermentative environments and less frequently in natural ones. This is probably due to unique features of their phenotypes that allow them to survive and proliferate in new niches with a better fitness than that of the parental species (Giannakou et al., 2020; Morales & Dujon, 2012). The famous yeast *S. pastorianus*, is a hybrid between *S. cerevisiae* and *S. eubayanus* (Libkind et al., 2011) and is used worldwide for lager beer production (Vaughan Martini & Martini, 1987; Sampaio, 2018) (Figure 1.2C). This hybridization resulted in a yeast that combines valued brewing properties with the ability to ferment at low temperatures (Figure 1.2C), which seems to be closely linked to the onset of low-temperature fermentations in Bavaria.

Many other *Saccharomyces* hybrids have been isolated from industrial fermentations. Examples are *S. cerevisiae* X *S. kudriavzevii* and *S. cerevisiae* X *S. uvarum* hybrids found in beer, cider and wine (González et al., 2008; Hittinger, 2013; Krogerus et al., 2018; Peris et al., 2018). Hybrids tend to show more robust characteristics than their parents, such as a higher level of resistance to different stresses during the fermentation process. For example, *S. kudriavzevii* and *S. uvarum* are better adapted to growth at low temperatures, compared to *S. cerevisiae*, which is instead more tolerant to ethanol. Hybrids between these species are thus adapted to grow under ethanol stress and at low temperatures (Marsit & Dequin, 2015).

Hybrids are chimeric organisms, possessing two divergent sets of chromosomes in a single nucleus. Most evolutionary theories suggest that this could be a problematic combination since incompatibilities between the two subgenomes would be expected. The Dobzhansky-Muller model (BDM) predicts incompatibilities resulting from negative epistatic interactions between genes with independent evolutionary histories. That is, two divergent populations over time may have accumulated different mutations that when put together in the same organism by hybridization are expected to be less fit than the combination of alleles present in non-hybrid organisms. However, it is known that this is not always the case and hybridization in many cases provides rapid adaptive advantage with increased fitness compared to parental strains (Gabaldón, 2020). Following hybridization several processes can shape the hybrid genome, including duplication or loss of whole chromosomes or large chromosomal regions, gene loss, gene conversion or even genome duplication. Many of these processes result in a progressive loss of heterozygosity (LOH), which promotes genome stabilization (Smukowski Heil et al., 2017). Another possible fate of a hybrid is crossing with one of the parental strains. This corresponds to backcrossing and if successive backcrossing with one of the parents occurs this can lead to 'cleaning' of one of the subgenomes, thus helping the LOH process. In an advanced stage of this process, it is likely that this hybrid has almost the complete genome of one of the parents (the one to which it has backcrossed) and a minor number of genes from the other progenitor. Nevertheless, this minor contribution

might have a selective advantage in the environments where the hybrid was formed. These minoritarian regions that result from an initial hybridization event are called introgressions. Several cases of introgression have been reported in *S. cerevisiae* (Barbosa et al., 2016; Borneman et al., 2016; D'Angiolo et al., 2020), however the association between introgression and adaptive advantage might not always be clear.

1.3.4. Horizontal gene transfer

Horizontal gene transfer (HGT), is a process that can generate variation and the acquisition of new traits in both prokaryotes and eukaryotes (Keeling & Palmer, 2008). In yeasts, HGT has been proposed as a mechanism for genetic adaptation to specific niches (C. Gonçalves et al., 2018; Hall et al., 2005) and is defined as a non-sexual transfer of genetic material between different species. Although, it is a well described mechanism in prokaryotes, with many reported examples, in eukaryotes only in the last decade genomic studies have shown its importance.

In *S. cerevisiae* HGT has been associated with adaptation to new environments during the domestication process (Giannakou et al., 2020). One of the most intensively studied cases of HGT, was the identification of three major chromosomal segments in a commercial wine strain, EC 1118. This strain contains three regions with 120 kilobases in total named A, B and C, comprising 39 genes, of which 5 are pseudogenes (Marsit & Dequin, 2015; Novo et al., 2009). These three regions were acquired by HGT from phylogenetically distant yeast species and are mostly present in wine strains. The donor species of regions B and C, are species that also co-habit with *S. cerevisiae* in wine, *Zygosaccharomyces bailii* and *Torulaspora microellipsoides*, respectively (Galeote et al., 2011; Marsit et al., 2015). The acquired regions code for potentially important functions in wine production, including sugar and nitrogen metabolism. Some of the genes in these regions have been analyzed in detail, which has helped elucidating their effective role in wine production. One such case is *FSY1* which encodes for a high-affinity fructose transporter (Figure 1.2D). As noted earlier, fructose consumption is a highly desirable trait during wine fermentation, as failure to consume this sugar can give rise to incomplete fermentations with marked economic losses (Galeote et al., 2010). In addition to *FSY1*, two other genes, named *FOT1* and *FOT2*, are found in the same region, and have gained particular relevance. These genes code for oligopeptide transporters and have an important role during wine fermentation. These genes allow for the strains that have them to increase the range of nitrogen sources that can be used. Since nitrogen is scarce in these niches, these transporters

play a crucial role, contributing to fitness during wine fermentation (Marsit et al., 2017; Marsit & Dequin, 2015).

1.4. **Aim and structure of the thesis**

The main goal of this work is to deepen the knowledge about the domestication trajectories of the iconic model organism *S. cerevisiae*. More specifically, this research project aims at understanding in more detail the multiple layers of the domestication process. For this, we used an array of approaches that combined evolutionary and population genomics with ecological studies and phenotypic analyses. Therefore, this research integrated field work, laboratory experimentation and bioinformatics.

As discussed above, *S. cerevisiae* has several lineages, both domesticated and wild. In chapter 2, we investigated the emergence of a domesticated lineage associated with cachaça, a Brazilian distilled beverage, and its relationship with wine yeasts. Our findings led us to propose the concepts of primarily and secondarily domesticate. In chapter 3 we studied another population linked to wine yeasts but ecologically associated with processed olives that arose after an inter-species hybridization. We analyzed in detail the consequences of inter-species hybridization for adaptation at the genotypic and phenotypic level. Chapter 4 provides a comprehensive view of wild and domesticated populations of *S. cerevisiae*, integrating fragmentary results from several recent publications. It evaluates the distribution and ecology of wild populations and investigates the evolution of certain genes that code for relevant phenotypic traits. Finally, chapter 5 presents the main conclusions of this work as well as future perspectives.

2.

Multiple rounds of artificial selection promote microbe secondary domestication - the case of Cachaça yeast

The work presented in this chapter was published in:

Barbosa, R.[†]; Pontes, A.[†]; Santos, R.O.; Montandon, G.G.; Ponzzes-Gomes, C.M.; Morais, P.B.; Gonçalves, P.; Rosa, C.A.; Sampaio, J.P. Multiple Rounds of Artificial Selection Promote Microbe Secondary Domestication—The Case of Cachaça Yeasts. Genome Biology and Evolution 2018, 10(8), 1939-1955.

[†] *Raquel Barbosa and Ana Pontes contributed equally to this work*

The sequencing data generated in this study have been deposited at the European Nucleotide Archive (ENA) under the accession code PRJEB24932.

Contributions:

Ana Pontes performed all the experimental work with Raquel Barbosa.

Ana Pontes executed all whole-genome sequence data analyses with help from Raquel Barbosa.

The contributions of the various authors were as follows:

C.R., J.P.S., P.G., and R.B. conceived and designed the research.

C.M.P-G., C.R., G.G.M., J.P.S., P.B.M., R.B., and R.O.S. isolated new yeast strains.

A.P., C.R., J.P.S., P.G., and R.B. wrote the manuscript.

2.1. **Introduction**

Several features make the yeast *S. cerevisiae* an ideal model to study the mechanisms of microbe domestication. It shows clear signatures of artificial selection (Fay & Benavides, 2005; Legras et al., 2007; Liti et al., 2009; Novo et al., 2009) and the extensive knowledge available on its biological processes facilitates the interpretation of new data. The study of the genomic, genetic, and phenotypic transitions associated with microbe domestication has witnessed recent advances in the case of two major yeast products: ale beer and wine. In the first case, ale-type beer yeasts were found to be essentially distinct from other industrially relevant lineages, including wine yeasts, and different phylogenetic subgroups of beer yeasts could be associated with distinct beer types (Gallone et al., 2016; Gonçalves et al., 2016). In the second case, the closest wild relatives of wine yeasts were discovered in the Mediterranean region and in association with oak trees (Almeida et al., 2015). The detection of this unique lineage allowed the disclosure of the predomesticated version of the genome of wine strains (Almeida et al., 2015). This enabled a direct comparison of wine yeasts with their wild relatives that revealed widespread genome-wide divergence, particularly at noncoding sites and in transacting DNA binding proteins, thus implicating transcriptional regulation as a driver of divergence between these two groups (Almeida et al., 2017). These and other advances contribute to uncover a complex scenario involving the emergence of various domesticated lineages.

Given the worldwide ubiquity of fermented foods and beverages that rely on the conversion of simple sugars to ethanol implicating *S. cerevisiae* as the dominant microorganism (Legras et al., 2017; Romano et al., 2006), it can be hypothesized that this yeast was coopted multiple times by humans. If those putative domestication events involved the artificial selection of local wild *S. cerevisiae* lineages is still an open question. The case of sake yeasts, that are phylogenetically distinct from wine yeasts (Fay & Benavides, 2005; Liti et al., 2009), argues in favor of independent domestication. However, the global dispersion of domesticated wine yeast lineage (Borneman et al., 2016) and the widespread of mosaicism, that is, interpopulation recombination detected in *S. cerevisiae* (Tilakaratna & Bensasson, 2017) make it possible that a single "proto" -domesticated lineage was globally dispersed and further and independently domesticated, locally. The assessment of those alternative hypotheses and even the accommodation of an intermediate scenario that accepts both independently domesticated lineages and cases of multiple rounds of domestication of a single lineage, requires detailed knowledge of additional domestication events. Here, we use a population and phylogenomics approach to investigate the domestication of *S. cerevisiae* strains employed

in the fermentation of cachaça, a Brazilian distilled spirit based on the conversion of sugar cane juice.

Sugar cane cultivation and cachaça production were introduced by Portuguese settlers in the XVI century. Originally consumed only by slaves working on sugar cane plantations, its production gradually improved and by early XIX century it was considered a typical Brazilian beverage (Badotti et al., 2012). Nowadays cachaça is the third most popular distilled beverage in the world and 1.3 billion liters are produced annually (Badotti et al., 2012). Almost one billion liters are annually produced industrially in stainless steel distillation columns. Another 300 million liters are produced by approximately 35,000 rural producers through the distillation of the sugar cane wine (must) in copper alembics, the traditional production method (Rosa et al., 2009). Industrial cachaça producers employ baker's yeast or active dried yeast used for ethanol production as the starter for the fermentation process, whereas traditional producers use natural ferments (spontaneous fermentation) cultured by various methods (Rosa et al., 2009). The fermentation employs a dilution of the original sugar cane juice that reduces sucrose concentration from around 20% to 16-14% w/v, and a starter culture from a previous fermentation that normally represents 20% of the total volume of the fermentation vat. At the beginning of the season natural inoculum is prepared in traditional distilleries by mashing rice maize flour and salt biscuits with undiluted cane juice and lemon or orange juice to lower pH. Typically, the cachaça fermentation stage lasts 18-30 hours and proceeds at elevated temperatures that in some regions can reach 41 °C (Vianna et al., 2008). At the end of fermentation ethanol concentration is around 8% and products with organoleptic relevance are ethyl esters, aldehydes, and organic acids (Cardoso et al., 2004). More recently, multilocus sequencing was employed to analyze cachaça strains from traditional distilleries of different regions in Brazil that were resolved in two main populations, one corresponding to the wine group and the other corresponding to "native strains" (Badotti et al., 2014). According to these authors, a third group of cachaça strains corresponded to hybrids of the two groups mentioned above.

In this study, we employed whole genome data to try to understand in more detail the relationships of cachaça strains among themselves and with representatives of well-established populations of *S. cerevisiae*, in particular to the Wine group. More specifically, we wanted to know if cachaça fermentation selected for typical wine strains, for modified wine strains further adapted to this niche, or, instead for strains unrelated to the Wine group. We also wanted to assess the degree of genetic variability within the cachaça group. Also, taking advantage of the availability of complete genome data we wanted to know if domestication signatures typical of wine strains were present in cachaça yeasts or if distinct signatures could be detected. Moreover, given the recent identification of wild populations of *S. cerevisiae* in Brazil (Barbosa et al., 2016), their relationship with cachaça yeasts and the evaluation of possible genetic contact between these two groups were also investigated.

2.2. Materials and Methods

2.2.1. Genome sequencing, read alignment and genotype calling

Whole-genome Illumina MiSeq paired-end reads (500 cycles) were obtained for monosporic or single cell derivatives. Single-cell derivatives were used when cultures did not form ascospores (Table A1.1). Genome data of strains not sequenced in this study was retrieved from public databases (Table A1.1). Where only finished genome sequences were available, the corresponding error-free Illumina reads were simulated using dwgsim (https://github.com/nh13/DWGSIM). Reads for each isolate were mapped to *S. cerevisiae* reference genome (UCSC version sacCer3) using SMALT v0.7.5 aligner (https://www.sanger.ac.uk/science/tools/smalt-0). The reference index was built with a word length of 13 and a sampling step size of 2 (-k 13 -s 2). An exhaustive search for alignments (-x) was performed during the mapping step with the random assignment of ambiguous alignments switched off (-r -1) and the base quality threshold for the look-up of the hash index set to 10 (-q 10). With these settings, SMALT v0.7.5 only reports the best unique gapped alignment for each read. For paired-end information, the insert size distribution was inferred with the "sample" command of SMALT prior to mapping. Conversion of SAM format to BAM, sorting, indexing, several mapping statistics, and consensus genotype calling were performed using the tools available in SAMtools package v1.18 (Li et al., 2009) as described previously (Almeida et al., 2014). Multiple sequence alignments for each reference chromosome were generated from the resulting fasta files. For downstream analysis, all bases with Phred quality score below Q40 (equivalent to 99.99% base call accuracy) or ambiguous base calls were converted to an "N". All strains with > 20,000 heterozygous sites with a Phred quality score above Q40 were selected for phasing. The BAM file of each strain with the paired-end read sequences mapped to the reference genome was analyzed with the phase command of SAMtools to infer both phases, thus solving the heterozygous SNPs. This algorithm uses the pair-end read information and attempts to find the best phase using the Minimum Error Correction method, solving each local haplotype with the highest probability based on the observed reads and the reference genome. The -F option was used to exclude errors from unmapped or misaligned sequences. Two haplotypes were obtained for each strain, each with reduced levels of heterozygosis. One haplotype per strain was randomly chosen and used in subsequent analyses.

2.2.2. Phylogeny and population structure

Chromosomal single nucleotide polymorphisms (SNPs) were extracted from multiple sequence alignments only if the evaluated site was represented by unambiguous high-confidence alleles in at least 85% of the isolates. SNPs were then concatenated to generate a whole-genome SNP alignment. A neighbor-joining phylogeny was estimated using the p-distance model as implemented in MEGA 6. Single gene phylogenies were prepared with MEGA 6, employed sequence alignments obtained with Muscle and were constructed using neighbor-joining method and Tamura's 3-parameter model. Population structure was explored using the model-based Bayesian clustering method implemented in Structure v2.3.4 (Falush et al., 2003) and using SNPs present in all the sequences, not allowing for gaps. STRUCTURE was run with a subset of approximately 10,000 equally spaced parsimony informative sites. The number of Markov chain Monte Carlo (MCMC) iterations was set to an initial burn-in period of 100,000 iterations, followed by 100,000 iterations of sampling. The ancestry model allowed for admixture and allele frequencies assumed to be correlated among populations. Five independent simulations were run for each value of K, varying from K = 1 to K =15, and stability was assessed by monitoring the standard deviation between simulations.

2.2.3. Survey of specific genes and GO analysis

We performed de novo genome assemblies using SPAdes v3.1.0 (Bankevich et al., 2012). Prior to assembly, reads were processed with Trimmomatic (Bolger et al., 2014) based on a quality score threshold of 20 for windowed trimming, discarding reads < 100 bp in length or harboring ambiguities. To retrieve genes of interest, a local BLAST database was set up for each genome and ORFS were searched by BLASTN (1e-4 e-value cutoff), using as queries sequences of YPS 163 for *AQY1* and *AQY2* (Will et al., 2010), sequences of EC1118 for regions A, B and C (Marsit et al., 2015; Novo et al., 2009), sequences of CEN.PK13 for *RTM1*, *BIO1* and *BIO6* (Nijkamp et al., 2012) and SGD sequences for *PAD1* and *FDC1* (Mukai et al., 2014). Blast hits were retained if sequence identity was above 90% and if the sequence was aligned to at least 10% of the query.

The standard GO term discovery was performed with GO Term Finder tool, available at SGD.

2.2.4. Polymorphism and divergence analyses

Whole-genome levels of polymorphism and divergence were estimated using Variscan v2.0 (Hutter et al., 2006). Only sites with valid high-quality alleles (Q > 40) in at least 75% of ingroup sequences were used in calculations. Sites with more than this threshold were randomly subsampled to 75% of the total number of sequences (defined with the NumNuc parameter together with CompleteDeletion = 0 and FixNum =1). For divergence estimates between populations, only positions with valid alleles in at least four ingroup individuals were used for calculations (defined with the NumNuc parameter together with CompleteDeletion = 0 and FixNum = 0).

2.2.5. Analysis of introgressions from *S. paradoxus*

We searched for evidence of introgressions from other *Saccharomyces* species by mapping the reads to a combined reference that includes all the available annotated coding sequences of *Saccharomyces* species. Reads were mapped to this combined reference using BWA v0.6.2 (Li & Durbin, 2009) with default parameters but setting the quality threshold to 10 (-q 10). SAMtools v1.1852 (Li et al., 2009) was used for the manipulation of the resulting BAM files. Only genes with orthologs unambiguously annotated in all six species were analyzed. An ORF was considered to have a foreign origin to *S. cerevisiae* if its coverage was at least higher than 1/4 of the median whole-genome coverage for the analyzed strain. The ORF coverage was defined as the product of the total number of mapped reads to the orthologous ORFs by the read size, dividing by the sum of the length of each ORF, considering only the ones with > 25% of reads mapped (relative orthologous ORF with the highest number of reads) to control for spurious alignment counts. This coverage threshold allowed for some heterogeneity in the read counts and for the eventual presence of a foreign ORF together with the native *S. cerevisiae* ORF.

Pairwise divergence between *S. paradoxus* (strain YPS 138) and *S. cerevisiae* (strain S288c) was used as a proxy to search for evidence of DNA segments of *S. paradoxus* in the genomes of *S. cerevisiae* strains. Divergence per site, k (with Jukes-Cantor correction), was calculated using a nonoverlapping sliding window of 10,000 sites, using Variscan v2.0 (Hutter et al., 2006). Using de novo genome assemblies a local BLAST database was set up for each genome in order to retrieve the introgressed genes. The introgressed ORFS were searched by BLASTN, using the correspondent *S. cerevisiae* ORF sequences available at SGD as queries.

2.2.6. PCR detection of *SSU1* translocations

Primers and PCR conditions followed those of (Pérez-Ortín et al., 2002) and (Zimmer et al., 2014). In the case of the detection of the translocation first observed by (Pérez-Ortín et al., 2002), we modified the original primer sequences and the primers used to amplify *SSU1* were NOG1_FW (5'-GAATCTGATAGACACAATGC-3') and GLR1_RV (5'-TACTCTAGTAGCGAGGTC-3'), whereas the primers used to amplify *SSU1*R were YHL044_FW (5'- CAAGTACTGGGAGGATAAG -3') and GLR1_RV. All essays employed DNA obtained from parental strains.

2.3. Results

2.3.1. Complete genome sequence of Brazilian *S. cerevisiae* strains

In this study, we obtained the complete genome sequences of 26 Brazilian *S. cerevisiae* strains – 21 strains from cachaça, three strains from jabuticaba wine, and two strains from grape wine (Table A1.1). Jabuticaba wine results from fermentation of the fruits of a tree of the Myrtaceae (Jabuticaba, *Plinia cauliflora*) that is native to Brazil, Argentina, Paraguay, Peru, and Bolivia. The cachaça strains had distinct geographical provenances and were obtained in six Brazilian states. Three additional genome sequences from bioethanol producing strains available in public databases were also used (CBS 7960, JAY 291, and BG1). Finally, a group of 26 wild Brazilian strains studied recently by Barbosa et al. (2016) was also included in our study, so that a total of 55 genomes of Brazilian *S. cerevisiae* strains were investigated. For comparison and as indicated in Table A1.1, we used representatives of all the known *S. cerevisiae* populations for which genomic data is available: wild lineages found in Brazil and in oak trees in the Mediterranean region and in North America-Japan (NA-JPN), and representatives of the Malaysian, Philippine, and West African populations, together with members of domesticated lineages (Wine, Beer, Bread, and Sake). This group of non-Brazilian representatives encompassed 135 genomes.

2.3.2. Cachaça strains are polyphyletic

Similarly to the beer and bread strains that had a high number of heterozygous sites, indicative of ploidy levels higher than 2n (Gonçalves et al., 2016) (Table A1.1), 10 out of the 21 cachaça strains studied had also > 20,000 heterozygous sites (Table A1.1) and were phased prior to further analysis. Since a preliminary analysis showed that the vast majority of phased cachaça haplotypes were phylogenetically very close to each other (Figure A1.1), one haplotype per strain was randomly chosen and used in subsequent analyses. For strain UFMG-CM-Y263 both phased haplotypes, that showed some divergence, were retained in the final phylogenetic analysis. Among the wild strains we detected a single Brazilian strain, UFMG-CM-Y456, with an elevated number of heterozygous sites. In this case, the two phased haplotypes were phylogenetically distinct and were also retained in the final phylogenetic analysis.

The phylogenetic relationships of cachaça strains were analyzed based on 1,108,048 high quality polymorphic sites as shown in Figure 2.1. Instead of forming a single group, cachaça strains were markedly polyphyletic. Two clades containing mostly cachaça strains were placed at the base of the Wine + Mediterranean oak population (MO) + Beer2 clades. One, that we designated C1 was composed of eight strains and the other (C2) contained six strains, but two strains were isolated from *Tapirira guianensis*, a candidate Brazilian natural habitat for *S. cerevisiae* (Barbosa et al., 2016). The group of three Brazilian strains from sugar cane bioethanol was also placed at the base of the Wine + MO + Beer2 clades. Whereas group C2 contained only strains from the Tocantins state (cachaça strains and wild strains), group C1 contained only cachaça strains but from four Brazilian states, none of them being Tocatins state (Table A1.1). Other cachaça strains were included in the Wine, Bread and Wild Brazilian B1 clades (one, three, and one strains, respectively). Finally, four cachaça strains (UFMG-CM-Y263, UFMG-CM-Y628, UFMG-CM-Y637, and UFMG-CM-Y638) were placed in isolated positions outside any known clade. Interestingly, the Brazilian strains from grape and jabuticaba wine clustered within the Wine group (Figure 2.1). We note that the Brazilian wine strains were placed very close to commercial wine strains, being therefore possible that they are descendants of starter cultures, instead of typically autochthonous wine strains.

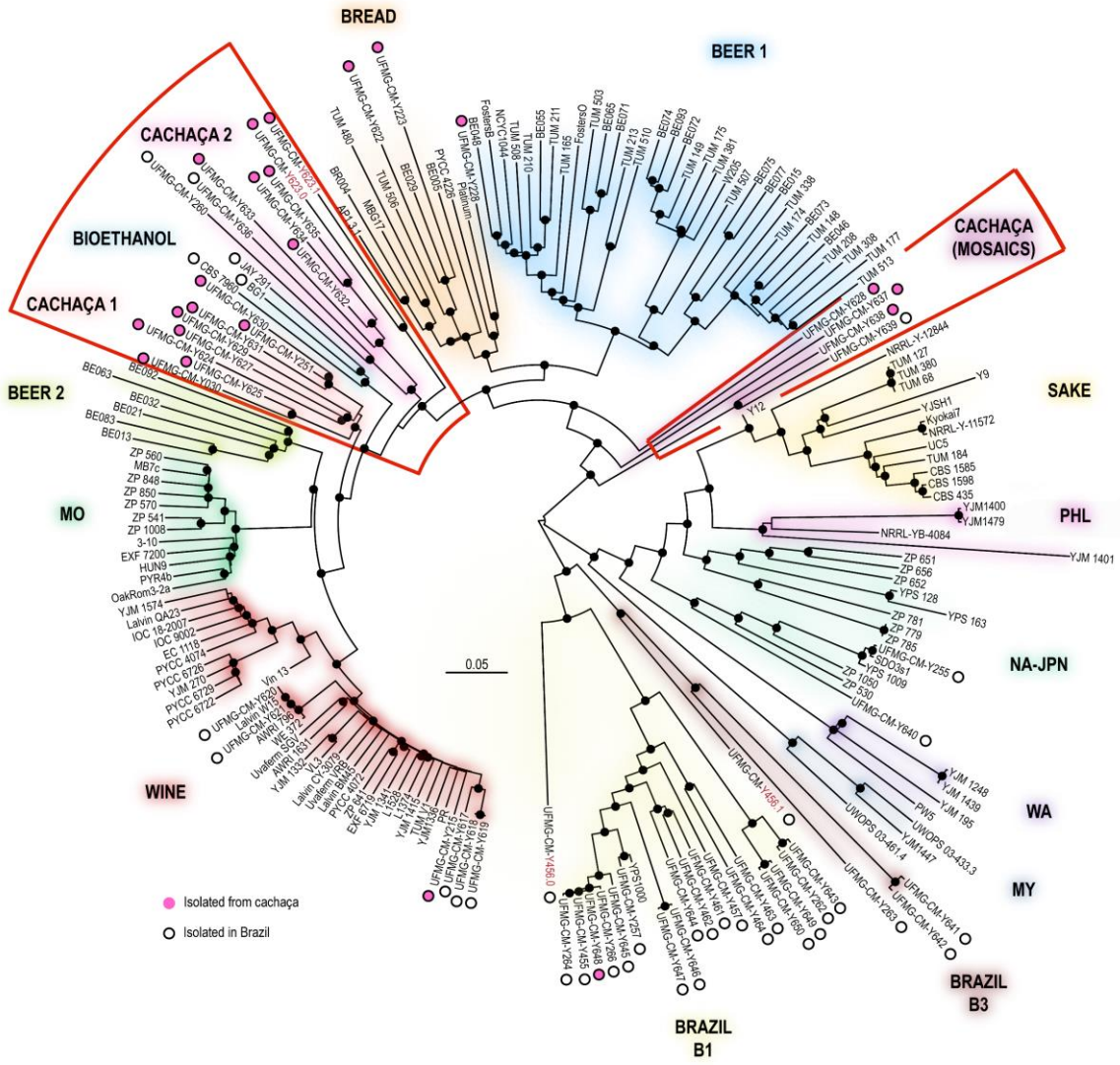


Figure 2.1 - Phylogenetic placement of cachaça strains among the known lineages of *S. cerevisiae*. Whole-genome phylogenetic tree of 188 sequences, inferred from 1,108,048 SNPs using neighbor-joining method and p-distance model of sequence evolution. The tree was rooted with *S. paradoxus*, branch lengths correspond to the expected number of substitutions per site and black dots depict bootstrap values above 90% (100 replicates). Lineages of cachaça or bioethanol strains are highlighted in red. For phased sequences (those with heterozygous sites > 20,000) a single phase is depicted except when the two phases are phylogenetically distinct (two exceptions with strain designations Indicated In red). Abbreviations of populations: MO, Mediterranean oaks; MY, Malaysia; NA-JPN, North America-Japan; PHL, Philippines; WA, West Africa.

2.3.3. Population structure

Using a selection of eight cachaça/bioethanol strains and also representatives of all populations of *S. cerevisiae* known so far (Table A1.1), we analyzed the population structure of this dataset using STRUCTURE (Falush et al., 2003) and testing from 2 to 15 ancestral (*K*) clusters. A comprehensive representation of sequence ancestry was achieved with *K* = 12

(Figure 2.2A) and analyses using higher *K* values did not reveal new meaningful clusters. Similarly to other recent studies (e.g., Almeida et al., 2015; Barbosa et al., 2016) our analysis recovered the main groups of industrial variants or geographically delimited populations such as 1) Wine – Mediterranean oak, 2) Beer, 3) Sake, 4) Philippines – North American-Japanese populations, 5) West Africa, 6) Wild Brazilian B1, and 7) Wild Brazilian B3. In this comprehensive analysis involving a wide representation of populations of *S. cerevisiae*, the cachaça strains shared most their ancestry with the Wine group.

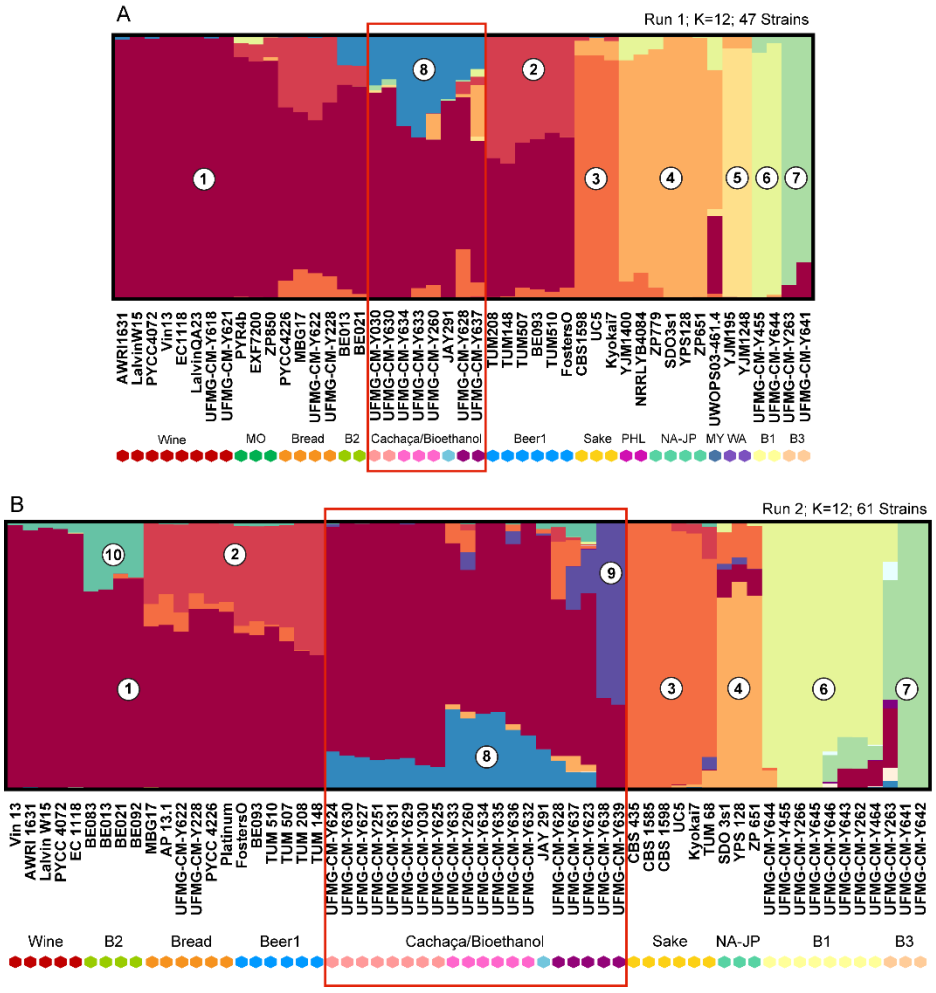


Figure 2.2 - Population structure of *S. cerevisiae*. (A) STRUCTURE plot (run1) that depicts the analysis of 47 strains from all major populations and from the cachaça lineages that emerged from the phylogeny depicted in Figure 2.1. The plot is based on a subset of 10,445 parsimony informative sites for *K = 12*. Numbers from 1 to 8 represent the different clusters that capture the maximum representation of population ancestry. The phylogenetic groups inferred in Figure 2.1 are color-coded at the bottom of the plot. (B) STRUCTURE plot (run 2) that depicts the analysis of 61 strains, excluding representatives of less relevant populations as inferred from run 1 and including all cachaça strains. The plot is based on a subset of 10,308 parsimony informative sites for *K = 12*. Numbers up to 10 represent the different clusters that capture the maximum representation of population ancestry.

However, they did not appear to be identical to the wine strains but seemed rather to present admixture between wine ancestry and a hitherto unknown genetic cluster (cluster 8). Interestingly, the two wild Brazilian populations recently described (Barbosa et al., 2016)

appear not to have relevant contributions to the genetic composition of cachaça strains (Figure 2.2). In order to confirm these results, we performed a second analysis in STRUCTURE eliminating some representatives of populations that did not show an association with cachaça strains (e.g., West Africa, Philippines, and Malaysia) and increasing the proportion of cachaça and wild Brazilian strains in the dataset. Overall, the results of this second study were consistent with those of the previous analysis and the correspondent K = 12 population representations are shown in Figure 2.2B. This analysis evidenced the important role of wine ancestry for other domesticated groups besides the Wine group, such as the main clade of ale-type beer strains (Beer 1), bread strains, the so-called Beer 2 clade, recently revealed and composed of Belgian ale beer strains and other beer strains from high ethanol beers (Gallone et al., 2016), and for the cachaça strains analyzed in this study. Except for the wine strains, all the other groups have clear signs of admixture (Figure 2.2). The resemblance between JAY 291, a strain used for the production of bioethanol, and cachaça strains is worth noting. Cluster 8, the hallmark of cachaça strains, was present in distinct proportions in different subgroups of cachaça strains. Strains of group C1 had a higher proportion of cluster 8 (85 – 90% Wine; 10 – 15% cluster 8), whereas for group C2 the proportion of cluster 8 tended to be higher (45 – 80% Wine; 20 – 30% cluster 8). For the cachaça strains not assigned to the two main cachaça groups (C1 and C2), nor to the Wine, Bread and B1 clades, more complex admixture patterns were detected and the proportion of cluster 8 was even lower than that observed for cachaça C1 group (UFMG-CM-Y628 10.1%, UFMG-CM-Y637 5.8%, and UFMG-CM-Y623 5.7%). One cachaça strain, UFMG-CM-Y638, that occupied an isolated position in the phylogeny of Figure 2.1 together with UFMG-CM-Y639, a Brazilian wild strain isolated from *Tapirira guianensis*, lost its genetic ancestry in cluster 8 with K = 12 (but not with lower K values, similarly with UFMG-CM-Y639). The K = 12 analysis of the second STRUCTURE run indicated that a new genetic cluster (cluster 9) explains most of the genetic ancestry of these two strains, being also a minor fraction of a few other cachaça strains, especially those with more admixed genomes.

2.3.4. Domestication signatures

2.3.4.1. Regions A, B, and C

One of the hallmarks of wine yeast domestication is the acquisition of three genomic regions (named A, B, and C), independently and through horizontal gene transfer from non-*Saccharomyces* yeasts. These regions encompass 39 genes potentially relevant for the winemaking process (Galeote et al., 2011; Marsit et al., 2015; Novo et al., 2009). In line with

these findings, regions A, B, and C can be found in wine yeasts, but not in beer or sake yeasts (Gonçalves et al., 2016). Given that our phylogenetic analysis pointed to a closeness of cachaça strains of clades C1 and C2 to wine strains (Figure 2.1), and that cachaça strains share with wine strains most of their genetic ancestry (Figure 2.2), we surveyed the genomes of cachaça strains for the presence of regions A, B, and C. In the eight strains of the C1 clade, six strains had at least one of these regions and the same happened with four of the six strains of C2 clade (Table A1.1). Therefore, for these two clades, and considering only strains isolated from cachaça, one of these regions was present in 67% of the strains. For comparison, among the wine strains included in this study (Brazilian wine strains excluded) at least one of the regions A, B, or C was present in 71% of the strains. Moreover, both the Brazilian grape and jabuticaba wine strains had at least one of these regions (Table A1.1). As mentioned before, two strains of the C2 cachaça clade were firstly regarded as wild because they were isolated from a natural substrate, the tree *T. guianensis* (Barbosa et al., 2016). However, these strains not only resembled cachaça strains in the phylogenetic and population structure analyses but harbored the complete region B, thus suggesting that these are feral strains, that is, domesticated cachaça strains that have escaped their original environment and have colonized natural environments. Since Brazilian jabuticaba wine and grape wine strains from Brazil were assigned to the wine group (Figure 2.1), it is not surprising that they harbor these regions. Similarly, the cachaça strains assigned to the bread group possessed these regions as it normally occurs for bread strains. Conversely, the Brazilian wild strains of clades B1 and B3 did not harbor any of these regions, as already observed for other wild populations (e.g., MO and NA-JPN) (Almeida et al., 2015).

Borneman et al. (2011) and Galeote et al. (2011) studied in more detail *S. cerevisiae* strains that harbored region B, that corresponds to a cluster of five ORFs, and reported several variants concerning copy number, synteny and chromosome location. Contrasting with the reference wine strain, EC 1118, that had three copies in chromosomes X (sinteny "e-a-b-c-d-e"), XII (sinteny "a-b-c-d-e"), and XIV (sinteny "a-b-c-d-e-a") (Figure 2.3A), the bioethanol strain JAY291 had a single copy of region B in chromosome XI with a distinct synteny (Galeote et al., 2011). The alternative synteny observed for region B is suggestive of a distinct mode of linearization of this region prior to its integration in the chromosome from a putative circular precursor (Borneman et al., 2011; Galeote et al., 2011). Interestingly, the distinctive features previously observed for JAY 291 were present in most of the cachaça strains that harbored region B, in Brazilian jabuticaba wine strains and in L1374, a wine strain from Chile (sinteny "d-e-a-b-c", Figure 2.3B, Table A1.1). Moreover, other cachaça strains or other South American wine strains had organizations of this region that match those of seen in wine strains other than EC 1118 (sinteny "e-a-b-c-d", Figure 2.3C and Figure 2.3D). Taken together, our results suggest that South American wine – bioethanol – cachaça strains are similar in terms of sinteny of region B.

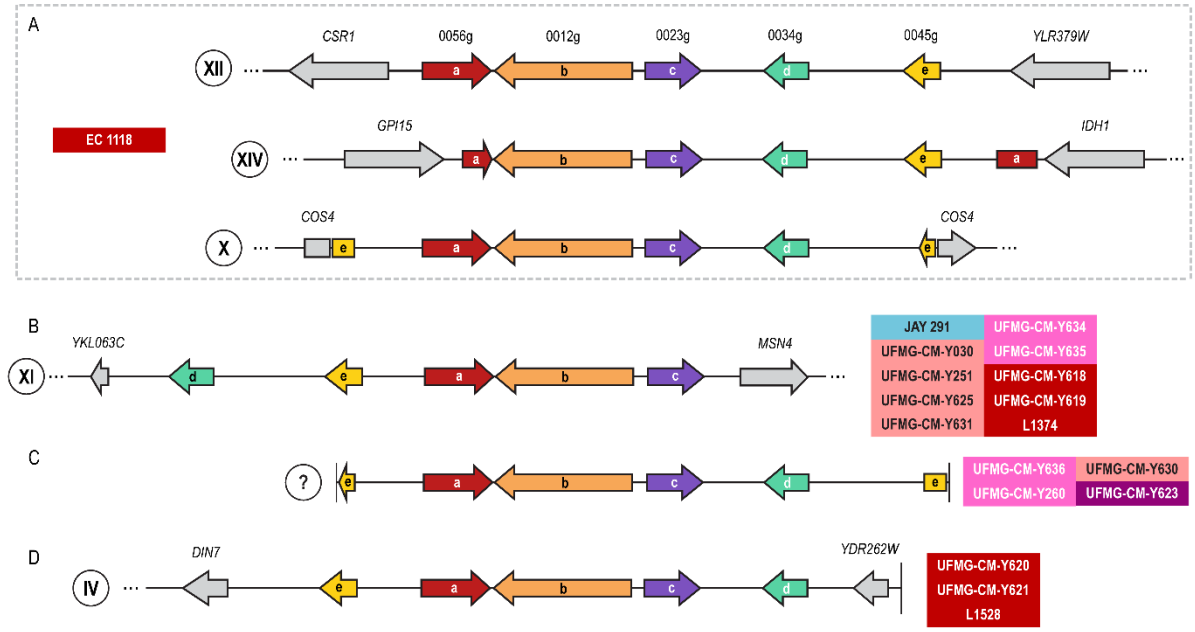


Figure 2.3 - Similar localization and organization of region B in Cachaça and other South American domesticated strains. (A) Organization of region B in EC 1118 as reported by Galeote et al. (2011). Colored arrows represent syntenic ORFS designated as in Novo et al. (2009). (B) Most cachaça strains that have region B share with bioethanol strain JAY 291 and with South American wine strains its organization and location in chromosome XI. (C) Alternative organization of region B in cachaça strains (chromosome location was not possible to determine). (D) Third organization type of region B, detected in chromosome IV in South American wine strains.

2.3.4.2. Inactivation of aquaporin genes

The domestication of wine strains and the consequent adaptation to sugar-rich and high osmolarity environments involved the adaptive loss of water channels encoded in the aquaporin genes *AQY1* and *AQY2* (Will et al., 2010). Several frame-shifting deletions or mutations giving rise to premature stop codons have been identified in the aquaporin genes of wine strains. Ale-type beer also has inactive aquaporin genes as well as other strains obtained from sugar rich environments (Gonçalves et al., 2016). All wine strains included in our study had at least one inactive aquaporin gene and 62% had the two genes inactive. For the six Brazilian strains clustering in the Wine clade (one strain from cachaça, two strains from grape wine and three strains from jabuticaba wine), the inactivation of *AQY* genes was detected in all strains (Table A1.1). All but one of the 14 strains in the cachaça clades C1 and C2 had at least one aquaporin gene inactivated and most (10 out of 14) had the two genes inactivated. The single strain that had both genes functional was UFMG-CM-Y260 (cachaça C2), a strain not isolated from cachaça and a putative feral strain. For *AQY1*, the typical inactivation of wine strains, an adenine deletion at base position 881 that renders *AQY1* inactive (A881 deletion), was also observed in cachaça strains of C1 and C2 clades (8 out of 12

cases of inactivation). The remaining four cases revealed a new type of inactivation caused by a thymine deletion at position 498 (Table A1.1). This deletion is unique of cachaça strains and has not been reported before. For *AQY2* the typical 11 bp deletion of wine strains was detected in C1 and C2 strains. For a single cachaça strain, the *AQY2* inactivation was caused by a guanine deletion at position 25, which is a typical sake strain inactivation. For the cachaça strains falling outside the C1 and C2 clades, the deeply admixed strain UFMG-CM-Y268 had both aquaporin genes inactivated but, interestingly, *AQY1* was inactivated through a new kind of mutation, detected only in that strain (A817del). The cachaça strains belonging to the Wine and Bread groups had the inactivations typical of those groups whereas Brazilian wild strains had functional aquaporin genes (Table A1.1). For the bioethanol strains, the inactivations typical of the wine group were observed, similarly to what was seen for cachaça strains. The exception was a unique 20 bp insertion at position 326 of *AQY2* detected for strain BG1.

2.3.4.3. RTM1

RTM1 provides resistance to the effects of inhibitory compounds present in molasses and is a member of a three-gene cluster that also includes *SUC* telomeric genes (Ness & Aigle, 1995). *RTM1* can be viewed as a domestication signature of beer yeasts since it is consistently present in the Beer 1 clade and absent in wine strains (Gonçalves et al., 2016). This gene is also present in other domesticated groups like the Sake clade, albeit infrequently, and in this study, we observed that is consistently present in the Bread and Beer 2 clades (Table A1.1). Interestingly, *RTM1* was present in cachaça strains of clades C1 (infrequently) and C2 (frequently) and one to three copies of this gene were detected in cachaça strains. We suggest that the *RTM1* donors of cachaça, bread and beer strains of the Beer 2 clade belong to the Beer 1 clade. In spite of being present in several domesticated groups, *RTM1* is also present consistently in wild Malaysian strains associated with sugary substrates (Table A1.1) and in West African strains. *RTM1* was also detected, infrequently, in Philippine and wild Brazilian B1 populations (Table A1.1).

2.3.4.4. *FZF1* and *SSU1*

FZF1 regulates the transcription of *SSU1*, a sulfite efflux pump conferring resistance to sulfite (Avram et al., 1999; Park & Bakalinsky, 2000) and is a target of recent adaptive evolution in *S. cerevisiae* and *S. paradoxus* (Engle & Fay, 2012). Moreover, most of the representatives of the wild population associated with Mediterranean oaks harbor an introgressed version of the allele found in the European population of *S. paradoxus* (Almeida et al., 2017). We compared the *FZF1* sequences of Brazilian strains belonging to wild populations, to cachaça, and to grape and jabuticaba wine with sequences from representatives of all other populations. With the exception of introgressed alleles from *S. paradoxus*, the diversity of native *S. cerevisiae* sequences could be resolved in two basic alleles that we designated the Cosmopolitan allele (C) and the Wine allele (W) (Figure 2.4). Whereas the W allele was detected in the Wine, Beer 2, and Bread clades (although together with C allele in some heterozygous bread strains), the C allele was found in most of the remaining populations, namely Beer 1, Sake, West Africa, NA-JPN, Philippines, Malaysia, and Brazil 1 (Table A1.1 and Figure A1.2).

Interestingly, the cachaça clades C1 and C2 presented a unique situation because they included the W and C allele in roughly equal frequencies. Two of the seven strain from clades C1 and C2 that had an increased number of heterozygozyties, suggestive of a ploidy higher than 2n, had the C and W alleles (Table A1.1 and Figure A1.2), whereas the remaining were homozygous for this gene. Similarly with the cachaça strains, the group of bioethanol strains contained also the two *FZF1* alleles. It thus appears that the presence of the C allele differentiates, at a population level, cachaça and wine strains. Another interesting observation concerned the wild Brazilian population B3. Among the studied strains of this population, the complex and probably tetraploid genome of UFMG-CM-Y456 had two copies of *FZF1*, one corresponding to allele C and the other corresponding to the *FZF1* allele present in the North American population of *S. paradoxus* (Figure 2.4A and Figure A1.2). The other two representatives of population B3 (UFMG-CM-Y641 and UFMG-CM-Y642) had *S. cerevisiae* x *S. paradoxus* recombinant sequences as illustrated for the first strains in Figure 2.4B. Therefore, in addition to the already described introgression of the European *FZF1* allele of *S. paradoxus* into the population of *S. cerevisiae* associated with Mediterranean oaks (Almeida et al., 2017), we observed here a second case of introgression implicating the North American *FZF1* allele of *S. paradoxus*.

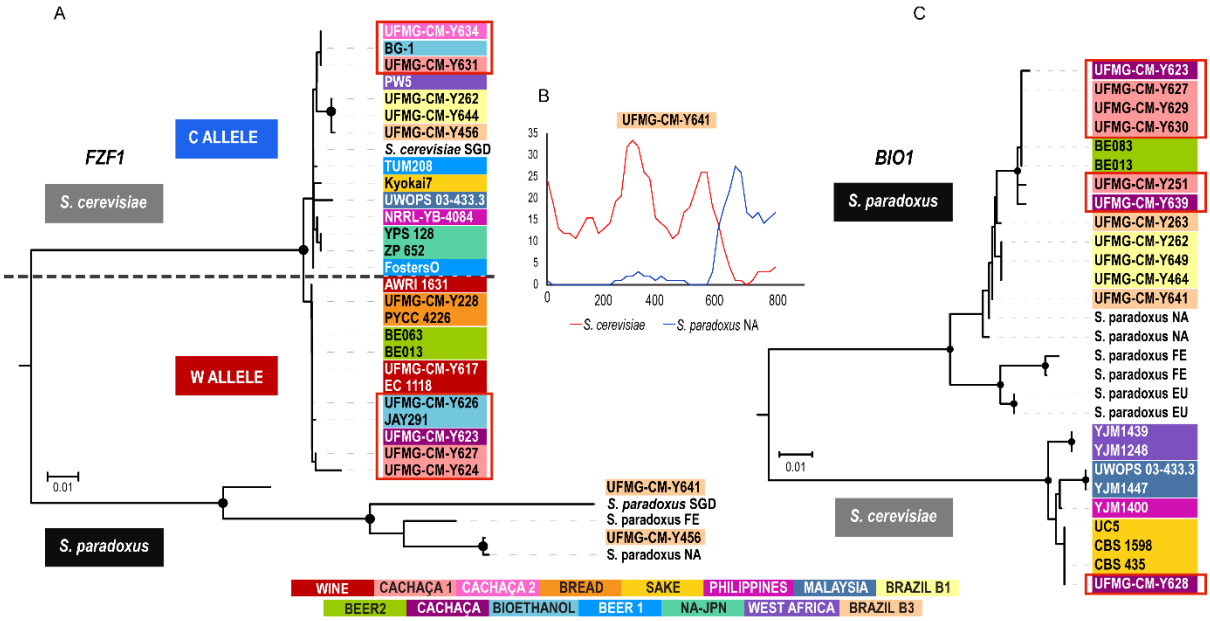


Figure 2.4 - Phylogenetic analysis of *FZF1* and *BIO1* in *Saccharomyces cerevisiae*. (A) Phylogenetic tree based on a *FZF1* alignment showing the cosmopolitan (C) and the Wine (W) allele, The C allele is present in Beer 1, Brazil 1, Malaysia, North America-Japan, Philippines, Sake and West Africa populations, and also on bioethanol and cachaça strains, whereas the W allele is present in Wine, Bread and Beer 2 populations, and also on bioethanol and cachaça strains. Populations are color-coded and sequences from cachaça/bioethanol strains are highlighted by a red border. The phylogenetic analyses employed the neighbor-joining method and Tamura's 3-parameter model with bootstraps values > 90% indicated by black circles (1000 replicates). The tree also includes *FZF1* sequences of different populations of *S. paradoxus* (EUR, Europe - SGD sequence; FE, Far East; NA, North America), and cases of introgression into the wild Brazilian B3 population and was rooted with *FZF1* (SGD) sequences of *S. mikatae* and *S. uvarum* (not included in the image). (B) Divergence plot of recombinant (North American *S. paradoxus* x *S. cerevisiae*) *FZF1* sequence of UFMG-CM-Y641. (C) Phylogenetic tree of *BIO1* from *S. cerevisiae* and *S. paradoxus* showing the origin of *BIO1* in cachaça strains. The allele present in wild Brazilian populations and in cachaça and beer strains of Beer 2 clade is the North American *S. paradoxus* allele. One cachaça strain has the *S. cerevisiae* Sake allele. The phylogeny was constructed as in (A) and was rooted with *S. arboricola*.

Two chromosomal translocations upstream of *SSU1*, a gene that codes for a sulfite efflux pump, have been implicated in sulfite resistance in wine strains because they lead to the overexpression of this gene (Pérez-Ortín et al., 2002; Zimmer et al., 2014). We compared wine strains and cachaça strains using PCR and primers designed to detect the native *SSU1* version (chromosome XVI) or a translocated and more sulfite-resistant allele (*SSU1*-R, chromosome VIII or *SSU1*-R2, chromosome XV). Among 24 wine strains analyzed, we detected *SSU1* in 20 occasions and *SSU1*-R in 10 occasions (in some cases both *SSU1* and *SSU1*-R were detected in the same genome; Table A1.1). Interestingly all of the 19 cachaça strains analyzed (clades C1, C2, and mosaic strains) were positive for *SSU1* and negative for *SSU1*-R and for *SSU1*-R2. These results suggest that sulfite resistance is not a trait selected for during cachaça fermentation, which is in line with the absence of this compound during the fermentation process.

2.3.4.5. Presence of *BIO1/BIO6*

BIO1 and *BIO6* encode enzymes involved in the synthesis of biotin, a pathway that also includes enzymes coded by four additional genes (*BIO2* - *BIO5*). Whereas strains used in sake fermentation have the complete genetic makeup necessary to synthetize biotin de novo, the remaining industrially used strains like wine, bread and beer yeasts are auxotrophic for biotin and lack functional versions of *BIO1* and *BIO6* (Borneman & Pretorius, 2015; Gonçalves et al., 2016; Hall & Dietrich, 2007). Moreover, the biotin pathway appears to have been lost in Saccharomycotina and subsequently rebuilt by a combination of horizontal gene transfer and gene duplication followed by neofunctionalization (Hall & Dietrich, 2007). Our survey for the presence or absence of *BIO1*/*BIO6* in cachaça strains and remaining populations yielded interesting results. First, we detected that these two genes are present in the Brazilian wild populations B1 and B3 and also in the Philippine population (Table A1.1). Confirming earlier reports (Borneman & Pretorius, 2015), we also recorded the presence of *BIO1*/*BIO6* in sake strains and in the Malaysian and West African populations. These findings challenge the notion that in *S. cerevisiae* only sake strains are able to synthesize biotin. Moreover, a phylogenetic analysis of *BIO1* revealed two clades, one containing native *S. cerevisiae* alleles from sake, Philippine, Malaysian, and West African populations and the other corresponding to the North American *S. paradoxus* allele but including also the Brazilian wild populations, thus suggesting that *BIO1* from *S. paradoxus* was introgressed in these *S. cerevisiae* populations (Figure 2.4C). The same pattern was observed for *BIO6* (Figure A1.3). We observed that, contrary to wine strains, some cachaça strains also harbored *BIO1* and *BIO6*. Most of the alleles found in cachaça strains could be assigned to the introgression from North American *S. paradoxus* observed in Brazilian wild strains. However, in one cachaça strain we detected the sake alleles of *BIO1* and *BIO6* (Figure 2.4C and Figure A1.3). Surprisingly, the introgressed *S. paradoxus* alleles of *BIO1* and *BIO6* were also detected in some beer strains of the Beer 2 clade (Figure 2.4C and Figure A1.3). It is not evident at this stage how this allele was acquired in beer strains.

2.3.5. Introgressions from *S. paradoxus*

We have previously reported a widespread dissemination of introgressions from North American *S. paradoxus* population in the wild populations (B1 and B3) of Brazilian *S. cerevisiae* (Barbosa et al., 2016). Although the pattern of distribution of introgressions was not uniform among strains, overall distinct patterns could be discerned for populations B1 and B3. Also,

the presence of these introgressions in Brazilian wild populations contrasted with their absence in other populations. Here we used the same approach to analyze cachaça strains. Introgressions from *S. paradoxus* (but not from other *Saccharomyces* species) were also detected but, globally, they were distinct from those previously found in Brazilian wild strains (Figure 2.5). In spite of some heterogeneity in their introgression patterns, all but one of cachaça strains of clades C1 and C2 had introgressions. Moreover, except for a single ORF in chromosome XV, none of the introgressions detected in cachaça strains were detected in the Wine group. The introgressions of cachaça strains involved single ORFs or larger regions encompassing up to four ORFs. In total 20 ORFs and five chromosomes were implicated, less than those reported by Barbosa et al., 2016 for wild Brazilian strains (62 introgressed ORFs in 13 chromosomes). Remarkably, only three ORFs were simultaneously found in cachaça and wild Brazilian strains (Figure 2.5), thus pointing to independent origins in these two groups. We could determine that the majority of introgressions originated from the most widespread North American population of *S. paradoxus*, normally designated as population B, thus pointing to the occurrence of hybridization events in the American continent, rather than in Europe. However, an exception to this pattern was recorded for cachaça strains and encompassed a considerable number of introgressions from the European population of *S. paradoxus*, all located in chromosome III (Figure 2.5).

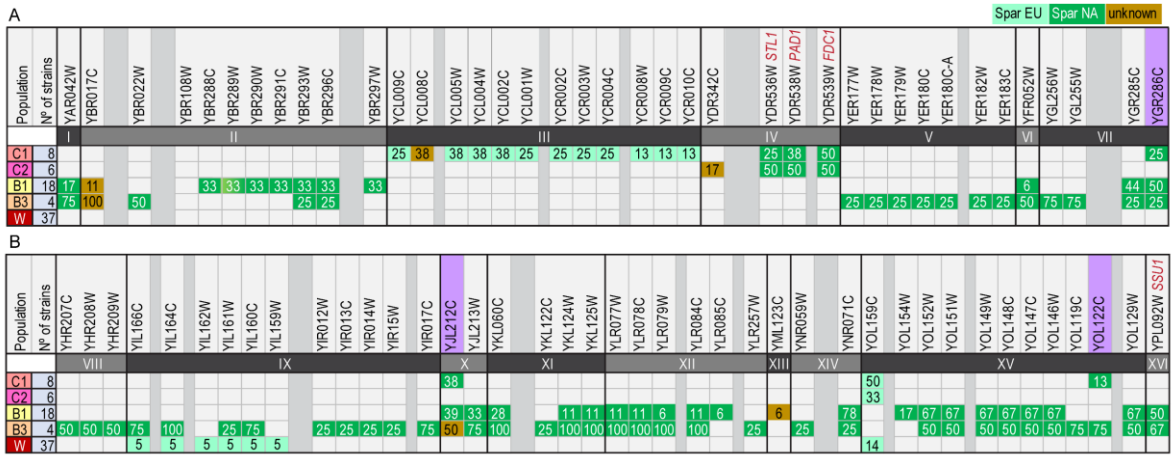


Figure 2.5 - Distinct introgressions of *Saccharomyces paradoxus* in the genomes of Brazilian *S. cerevisiae* Cachaça (C1 and C2) and wild (B1 and B3) populations and comparison with the Wine group. Introgressed ORFs are indicated in the first rows of the two panels and introgressions common to the cachaça and wild populations are highlighted in purple. (A) chromosome I - VII; (B) chromosome VIII - XVI. The presence of introgressions is color coded and refers to the donor population of *S. paradoxus*: European (light green), North American population B (dark green), unknown population or recombinant sequences (brown). The frequency (%) of each introgressed ORF in each population is indicated by numerals. The common names of the genes discussed in the text ae indicated.

As indicated above, some cachaça strains belong to the Wine and Bread groups, but these strains did not harbor the *S. paradoxus* introgressions detected for the C1 and C2 clades

(Figure A1.4). However, other cachaça strains with complex admixture patterns discussed above like UFMG-CM-Y623 and UFMG-CM-Y628 also had typical cachaça introgressions (Figure A1.4). Interestingly, the bioethanol strains shared with cachaça strains five introgressions and had also two introgressions found in Brazilian wild strains but not in cachaça strains (Figure A1.4). A Gene Ontology (GO) analysis of the complete set of introgressed genes detected in cachaça strains revealed that it was significantly enriched in genes encoding secondary active transmembrane transporters. Three genes were implicated: *STL1*, a glycerol proton symporter; *SMF1*, a broad specificity divalent and trivalent metal ion transporter; and *OPT1*, a proton-coupled oligopeptide transporter. Notably, these last two genes belong to the restricted group of three introgressed genes found simultaneously in cachaça and Brazilian wild strains (Figure 2.5). They belong also to a group of introgressed and significantly enriched genes coding for secondary active transmembrane transporters in wild Brazilian strains (Barbosa et al., 2016). For cachaça strains, the introgressed region in chromosome IV containing *STL1*, includes also *PAD1* and *FDC1* (Figure 2.5 and Table A1.1), two genes responsible for decarboxylation of aromatic acids like ferulic and cinnamic acids, thus rendering them less toxic. In an attempt to understand if the introgressed genes improved resistance to ferulic acid in acidic conditions (pH 4.5), we compared the phenotypes of cachaça strains harboring the native and the introgressed alleles (eight strains in each group). Cachaça strains harboring the introgressed genes were slightly more resistant to 0.2% ferulic acid but without statistical significance since 50% of cachaça strains harboring the introgressed genes had intermediate resistance, whereas 25% of cachaça strains with native *PAD1*/*FDC1* had the same resistance and all other strains had low resistance (Table A1.2).

2.3.6. Diversity, divergence, and fixed differences

Nucleotide diversity (pairwise differences, $\pi * 100$) of the Cachaça population for which a more representative number of isolates is available (C1) was 0.0018% (Table 2.1). This value is slightly higher than that of Wine population (0.0011%). We also measured the mean pairwise divergence between two alleles drawn from two populations (πB, estimated per site from pairwise comparisons across the total length of the genome).

Table 2.1 - Whole-genome diversity of cachaça strains and comparison with other populations of *Saccharomyces cerevisiae*.

	No. of strains	Analysed sites	Segregating sites	π
Cachaça (C1)	8	10,930,605	46,633	0.0018302
Cachaça (C2)	6	11,275,315	60,642	0.0029448
Brazil (B1)[a]	17	11,075,830	70,292	0.002169
North America-Japan[b]	42	11,348,218	119,184	0.002560
Mediterranean oaks[b]	31	11,286,153	56,053	0.0009901
Wine[b]	19	11,216,288	56,363	0.0011166

Note - Diversity values are per site estimates calculated for the length of the genome.

[a]Taken from Barbosa et al., 2016.

[b]Taken from Almeida et al., 2015.

As expected, nucleotide divergence ($\pi B * 100$) between cachaça C1 and the Wine group was lower (0.176) than the divergence between C1 and other populations (0.223 - 0.595, divergence to MO and West Africa populations, respectively). We also observed that C1 and Wine had the highest number of shared polymorphisms and the lowest number of fixed differences in pairwise comparisons of the various populations with the cachaça C1 group (Figure 2.6).

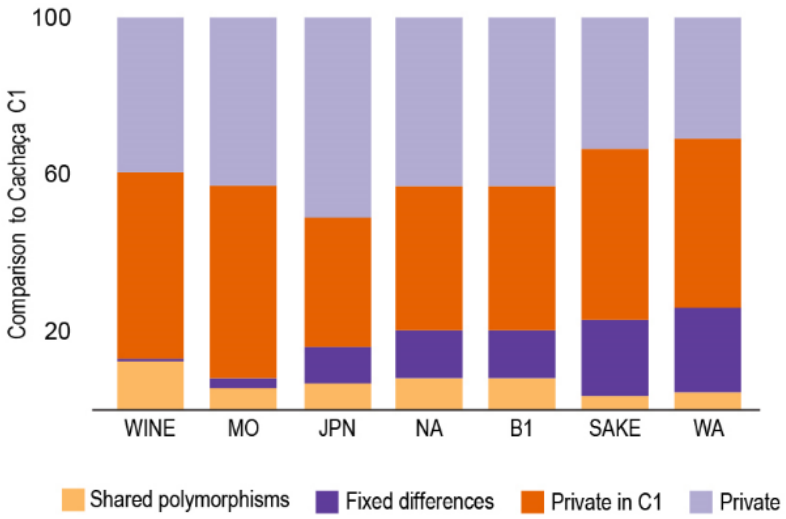


Figure 2.6 - Proportion of shared polymorphisms and fixed differences in pairwise comparisons of the Cachaça C1 group with other populations of *S. cerevisiae* (B1, Wild Brazil; JPN, Japan; MO, Mediterranean oak; NA, North America; WA, West Africa).

2.3.7. Cachaça feral strains and wild Brazilian strains in cachaça fermentations

In three cases, strains with cachaça genotypes (UFMG-CM-Y260, UFMG-CM-Y636, and UFMG-CM-Y639, Figure 2.1 and Figure 2.2) were found in a wild environment, the tree *Tapirira guianensis* (Table A1.1). Therefore, we regard these strains as feral and as evidence that domesticated strains can disperse into wild environments. Evidence of gene flow from domesticated lineages into Brazilian wild populations had already been reported (Barbosa et al., 2016) and here we also observed partial wine ancestry in the population structure analysis of the Brazilian wild population B3 (Figure 2.2). The reverse situation, that is, the occurrence of wild Brazilian strains in cachaça fermentations was also observed, but in a single occasion (UFMG-CM-Y648, Brazilian wild population B1). Overall, the limited number of cases of transitions from the original niche observed suggests moderate genetic contact. An apparent consequence of this is the absence of genetic homogeneity at regional scale, which might also be due to distinct selective pressures and a strong selective effect in anthropic and wild environments.

2.4. Discussion

Here, we analyzed a comprehensive group of *S. cerevisiae* strains used in the fermentation of sugar cane juice, the first step of cachaça production. Most strains were isolated from spontaneous fermentations in Brazil, but we also surveyed commercial cachaça strains and strains used for bioethanol production from sugar cane juice and molasses. Our analysis included, for comparison, representatives of the main domesticated and wild populations of *S. cerevisiae* known so far. The hypothesis that the yeasts that ferment cachaça are, or derive from, wine yeasts make sense from an historical and technological perspective. Alternative hypothesis worth considering are that the genetic stock of cachaça yeasts derives mostly from local (Brazilian) wild populations or that cachaça yeasts form a distinct and unique domesticated group not related to any of the known populations.

Our genome-based phylogenetic analysis represents the first population genomics study of cachaça strains and provides two important findings. First, cachaça strains are distinct from the main groups of domesticated strains (Wine, Beer 1, Beer 2, Bread, Sake). Secondly, and contrary to those domesticated groups (except for Beer 1), cachaça strains do not cluster in a single clade and are therefore not monophyletic. Indeed, most cachaça strains

belong to two clades (C1 and C2) and a third clade comprehends the bioethanol strains. Moreover, a population structure analysis indicated that cachaça strains share most of their genetic ancestry with wine yeasts. It is therefore conceivable, both from the historical and from the population genomics perspective, that the early stock of cachaça yeasts corresponded to wine yeasts transported from Europe by Portuguese settlers. This means that cachaça yeasts are modified, that is, further domesticated, wine yeasts. In addition, because Brazilian bioethanol yeasts closely resemble cachaça yeasts, our conclusion can also be extended to this technologically important group. We note that the divergence of cachaça yeasts from wine yeasts cannot be explained simply by geographic effect and by genetic isolation of South American strains. On the contrary, the genomes of Brazilian grape and jabuticaba wine strains suggest that typical wine yeasts belonging to the global population of wine strains occur in Brazil and not exhibit the differences observed in cachaça yeasts. An earlier study analyzed cachaça strains based on the sequences of four genes (Badotti et al., 2014) and recognized three groups: wine, "native" and hybrids between those two allelic variants. The higher resolution of our analysis confirmed closeness between cachaça and wine strains but failed to reveal substantial presence of native alleles in cachaça strains.

Combining the phylogenetic and the technological perspectives, three domesticated main groups emerge: the Wine-super group, including particular cases of beer yeasts (Beer 2 clade), cachaça and bioethanol strains; the Bread - Beer group (Beer 1, main group of ale-type beer strains); and the Sake group. With respect to cachaça yeasts, we find it relevant that their polyphyletic nature contrasts with the monophyletic nature of wine and sake yeasts. We also find it relevant that wine yeasts are typically diploid whereas cachaça yeasts are a mixture of diploid and probably tetraploid strains. These and other observations of discontinuities among domesticated lineages of *S. cerevisiae* (e.g., Gallone et al., 2016; Gonçalves et al., 2016) suggest that multiple domestication events have occurred in this species and also that fundamentally distinct patterns of adaptation to anthropic environments have unfolded. Based on such observations we propose here a multilayered domestication model, encompassing primary and secondary domestications with wine and cachaça as two epitomes of primary and secondary domestication, respectively (Figure 2.7).

Cachaça yeasts are likely to represent a case of a second round of domestication of primarily domesticated wine yeasts because their genomes exhibit not only signatures of wine yeast domestication but also other markers that we postulate were acquired more recently (Figure 2.7). Signatures of primary (wine yeast) domestication are the presence of regions B and C and some of the types of inactivating mutations of aquaporin genes that were detected simultaneously in wine and cachaça yeasts. The signatures of secondary domestication are additional kinds of inactivation of aquaporin genes not found in wine yeasts or in any other group and that are therefore specific of cachaça yeasts. Another evidence of secondary domestication concerns the presence of *RTM1* gene in cachaça strains. Since this cluster

provides resistance to inhibitory compounds that occur in molasses, its acquisition by cachaça yeasts might have enhanced adaptation to sugar cane juice. It is likely that this cluster was acquired from beer yeasts (Beer 1 clade) given its prevalence in this clade. Reacquisition of biotin prototrophy through cooptation of *BIO1/BIO6* from *S. paradoxus* (or from Brazilian wild populations that originally acquired these genes from *S. paradoxus*) and, albeit less frequently, from Sake strains, also configures another set of secondary modifications. The acquisition of the *FZF1* C allele and the apparent loss of *SSU1*-R, that enhances resistance to sulfite, are also two additional secondary modifications from a hypothetical ancestral wine genome. Besides a primary level of discontinuities detected between wine and cachaça yeasts at a phylogenetic and population level and a secondary category of differences that concerns gene-level domestication signatures, a third category of changes corresponds to the introgressions from *S. paradoxus*. We detected the introgression in cachaça strains, but not in wine strains, of eight ORFs from the North American population of *S. paradoxus*. As discussed earlier it is likely that this population is present also in South America and therefore these introgressions were acquired after the introduction of wine yeasts in South America.

We propose a three-step model for the domestication of *S. cerevisiae*. From wild populations, some of which known only from wild representatives (e.g., North-America Japan or Malaysian populations) whereas others include strains from both wild and anthropic environments (e.g., Philippines population), primarily domesticated populations have evolved. Wine and sake yeasts are the best examples of such events and beer yeasts of Beer 1 clade might also represent a primary domestication event (Figure 2.7). We tentatively place the Beer 1 clade in an intermediary position between primarily and secondarily domesticated lineages mostly because the genomes of beer yeasts are tetraploid, contrary to the genomes of wine and sake yeasts that are diploid. It was recently shown that polyploid *S. cerevisiae* cells explore more genotypic and phenotypic space than lower ploidy cells (Scott et al., 2017). According to our model, > 2n genomes are common in secondarily domesticated populations, a situation that can relate to more derived states in yeast domestication trajectories (Figure 2.7). The Bread, Beer 2, and Cachaça populations would represent cases of secondary domestications where the genome of a wine yeast acquired some additional domestication-related features (e.g., *RTM1* and *BIO1/BIO6*) while other features were lost (e.g., region A and *SSU1*-R in cachaça yeasts) (Figure 2.7).

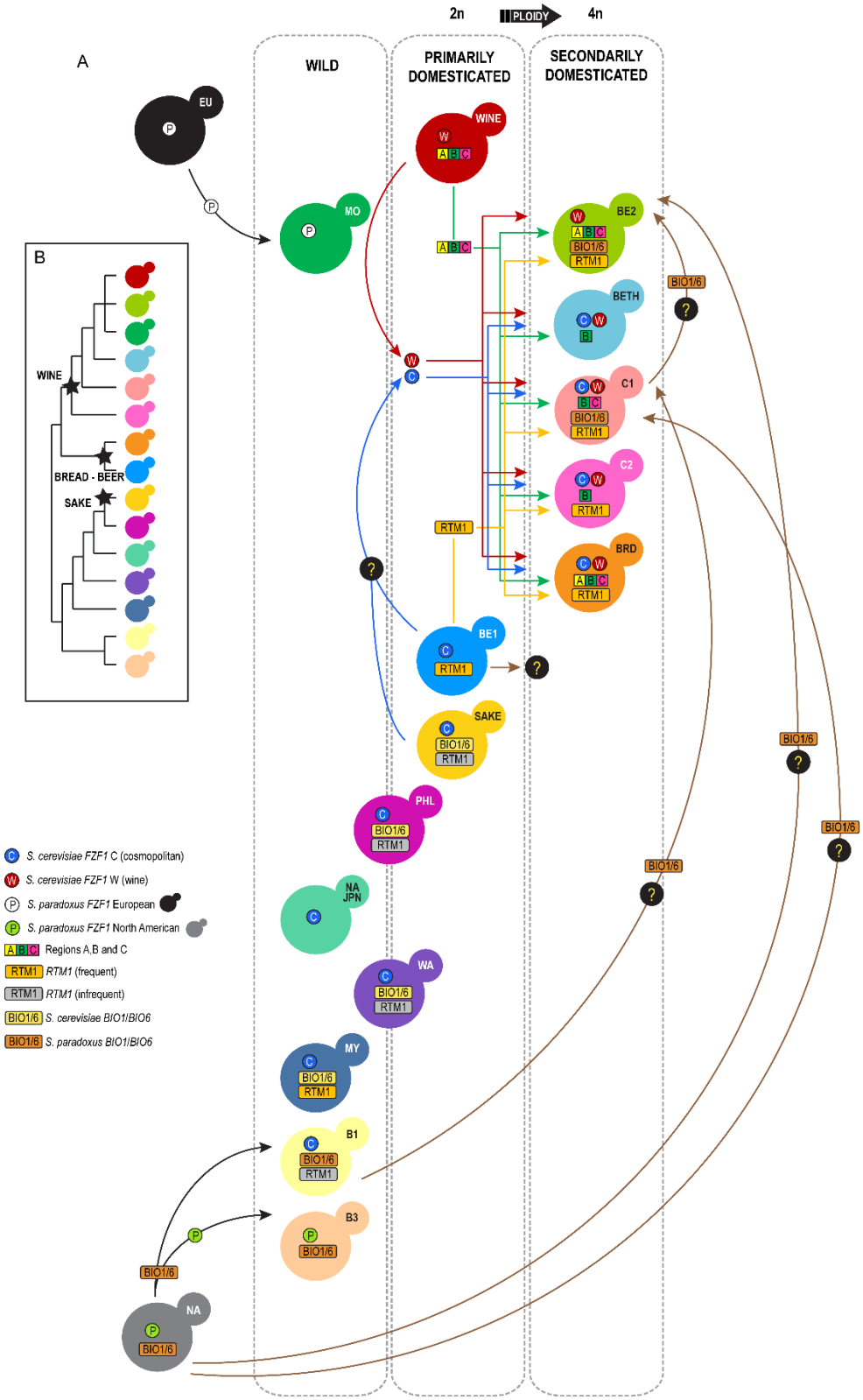


Figure 2.7 - A model of secondary domestication in *Saccharomyces cerevisiae*. (A) Primarily domesticated lineages, Wine, Beer 1 (BE1) and Sake, contribute to the secondarily domesticated genomes of Cachaça (C1, C2), Bioethanol (BETH), Bread (BRD) and Beer 2 (BE2) yeasts. Populations including simultaneously wild and domesticated strains are placed at the borderline between wild and primarily domesticated populations. Hypothetical genetic transfers are marked with a question mark (abbreviations of other populations: B1 and B3, Wild Brazil B1 and B3; MO, Mediterranean Oak; MY, Malaysia; NA-JPN, north America-Japan; PHL, Philippines). (B) Schematic representation of the phylogenetic relationships of wild and domesticated populations of *S. cerevisiae*. Stars depict the main domesticated groups.

Recent studies on the domestication of *S. cerevisiae* not only clarified the acquisition of several domestication traits (Coi et al., 2017; Marsit et al., 2015; Novo et al., 2009), and structural rearrangements (Borneman et al., 2012), but also unveiled unanticipated levels of complexity. For example, they showed that beer yeasts are genetically very diverse and contrary to wine yeasts do not have an obvious wild ancestor (Sampaio & Gonçalves, 2017). Their ploidy is higher than 2n, a trait shared with bread yeasts, but the exact causes that elicited this change in some anthropic environments but not in others is unknown. Furthermore, it is possible that the evolution of their genomes included multiple interpopulation hybridizations, thus making it challenging to trace back their deep roots in one (or more) wild lineage. Also, Asian wild yeasts are not related to domesticated sake yeasts, this rendering the origins of this domesticated group also obscure. These and other questions call for a better understanding of the domestication trajectories of *S. cerevisiae* which in turn will help ascertain the weight of independent domestications versus the role of the wine group in the global pattern of *S. cerevisiae* domestication. The results presented here reinforce this last aspect because cachaça yeasts clearly emerge as secondarily domesticated wine yeasts. The relevance of meta-domestication of wine yeasts is strengthened by our observation that bread and beer yeasts of the Beer 2 clade are also additional cases of secondary domestication of wine yeasts (Figure 2.7).

In spite of being used worldwide, wine yeasts are better circumscribed phylogenetically than cachaça yeasts. The latter are not monophyletic as clade C1 and C2 demonstrate, and even include strongly admixed strains as well as strains from other industrial groups as wine and bread strains. It might be speculated that because of its recent origins, cachaça fermentations have not yet shaped into a single and typical population, contrary to what is observed for the historically much older wine fermentations. Here, we have shown that wine yeasts constituted the main genetic source of cachaça strains. However multiple additional contributions originating in other domesticated populations and possibly also in native wild lineages or even in sister species *S. paradoxus* also contributed to shape the unique genomes of cachaça yeasts. Therefore, from the three hypothesis stated earlier, the first one stating that cachaça yeasts are, or derive from, wine yeasts gets stronger support.

3.

A quasi-domesticate relic hybrid population of *Saccharomyces cerevisiae* x *S. paradoxus* adapted to olive brine

The work presented in this chapter was published in:

Pontes, A.; Čadež, N.; Gonçalves, P.; Sampaio, J.P. A Quasi-Domesticate Relic Hybrid Population of *Saccharomyces cerevisiae* x *S. paradoxus* Adapted to Olive Brine. Frontiers in Genetics 2019, 10, 449.

The sequencing data generated in this study have been deposited at the European Nucleotide Archive (ENA) under the accession code PRJEB30431.

Contributions:

Ana Pontes performed all the experimental work.

Ana Pontes executed all whole-genome sequence data analyses.

The contributions of the various authors were as follows:

A.P., J.P.S., and P.G. conceived the study, analyzed the data, and wrote the manuscript.

A.P., J.P.S, and N. Č. collected and identified the strains.

3.1. Introduction

The domestication of plants and animals was a major revolution in human history because it drove the emergence civilizations with the associated demographic and technological consequences that last until today (Diamond, 1997). In many instances domestication represents a dramatic case of adaptive divergence in response to human selection (Doebley et al., 2006; Ross-Ibarra et al., 2007). Domestication consists in the selective and controlled propagation of an organism that genetically acquires modifications that not only distinguish it from its wild ancestors, but also make it more useful to humans (Diamond, 2002). Since archeological and biomolecular evidence indicates that fermented beverages reminiscent of rice wine were produced as far back as 9,000 years ago in China (McGovern, 2009) and the forebear of modern beer was consumed 8,000 years ago in Sumeria (Hornsey, 2003), the case of microbe domestication has a context and a time scale comparable to the much better understood aspects of plant and animal domestication. In fact, the mechanisms and consequences of artificial selection of microbes such as yeasts, carried out in most cases unconscientiously by innumerable generations of brewers, winemakers and other artisans, are starting to be understood (Almeida et al., 2015, 2017; Barbosa et al., 2018; Duan et al., 2018; Gallone et al., 2016; Gonçalves et al., 2016; Legras et al., 2018; Peter et al., 2018). Although the mechanisms that gave rise to phenotypes of domesticated strains are currently the focus of intense scientific inquiry, the detailed comprehension of the multiple transformations that gave rise to domesticated yeast lineages is far from been achieved.

Besides the obvious cases of the emergence of wine, beer or sake variants, several other presumably domesticated lineages of *Saccharomyces cerevisiae* have also been recently revealed (Duan et al., 2018; Preiss et al., 2018). Moreover, the proximity to humans appears to have elicited the emergence of a new niche to which *S. cerevisiae* is adapting to, as the isolation of commensal (Angebault et al., 2013) and opportunistic (Enache-Angoulvant & Hennequin, 2005; Muñoz et al., 2005) strains suggests. Recently we and others have explored the domestication space of *S. cerevisiae* (Almeida et al., 2015; Barbosa et al., 2018; Gallone et al., 2016; Gonçalves et al., 2016; Legras et al., 2018) and one of the main findings is that the domestication routes of this yeast are multiple and independent, and most remain poorly known. The picture that is gradually emerging depicts a complex population structure rich of different wild populations, most of them showing geographical partitioning, and numerous domesticated populations, each associated with a given fermented product and showing specific adaptations related to the type of fermentation in which they participate (Duan et al., 2018; Legras et al., 2018; Peter et al., 2018). This complex scenario is further complicated by the occurrence of admixture between certain populations that gives rise to mosaic genotypes

(Tilakaratna & Bensasson, 2017) and to transitions from primary to secondary domestications (Barbosa et al., 2018). Here we address a yet unexplored aspect of *S. cerevisiae* domestication - that of the emergence of lineages harboring some domestication signatures but that do not fit completely in the definition of a domesticated yeast because artificial selection, even if unintentional, is not easy to accommodate with the emergence of a phenotype that provides identifiable benefits to humans. Specifically, we survey a set of *S. cerevisiae* strains associated with the maturation of table olives, where they occur spontaneously without propagation from one batch to other. Olive brine strains were first studied by Cromie et al. (2013) using restriction site-associated sequencing (RAD-seq), a reduced genome sequencing strategy. These authors detected a small group of *S. cerevisiae* strains isolated from European olives that clustered next to wine strains and were defined by a unique set of sequence variants not present in other populations. Subsequently, Strope et al. (2015), in a genomics survey of 100 *S. cerevisiae* strains, found that a restricted group of three strains that included YJM 1252 (PYCC 6732 = CBS 3081), isolated from alpechin (olive mill wastes) in Spain had a considerable number of ORFs (> 200) from *S. paradoxus*. The other two strains were YJM 1078 (NRRL YB-4348 = PYCC 8028) and YJM 248 (NRRL Y-12659 = CBS 2910 = PYCC 8034) isolated in Portugal in the 1950's from human feces. In a more recent study involving the genome analyses of more than 1,000 *S. cerevisiae* strains, Peter et al. (2018) identified 17 strains in a so-called alpechin clade sharing the already reported *S. paradoxus* contribution. Because the strains originating from the olive niche have never been studied separately but rather on comprehensive surveys that included hundreds or thousands of strains from other provenances, thus precluding their detailed analysis, we carried out an investigation on the ecological, physiological, and genomic particularities of olive strains aiming at understanding their origins and specific adaptations within the global framework of yeast domestication. We found that olive strains are more fit to grow and survive in olive brine than control *S. cerevisiae* wine strains and that they appear to be adapted to cope with the presence of NaCl in olive brine. Moreover, the ecological range of these strains includes the processed olives niche but not olive trees or olives in natural conditions. We postulate that an ancient hybridization between a *S. cerevisiae* wine strain and *S. paradoxus*, provided the genetic diversity that allowed the adaptation to the new niche and that this process was accompanied by the adaptive loss of most the non-*cerevisiae* sub-genome.

3.2. Materials and Methods

3.2.1. Yeast isolation, identification, and crosses

The isolation of *Saccharomyces* strains was conducted using a selective enrichment protocol previously described (Sampaio & Gonçalves, 2008). For strains isolated in Slovenia, samples were directly used for yeast isolation without enrichment. Preliminary species-level identifications were performed by sequencing the D1/D2 region of the 26S rDNA. Crosses involved ascospore micro-manipulation. Positive crosses between two parental strains were confirmed by sequencing of *GAL1* and confirmation of the expected heterozygous sites. For each cross, interspecific spore viability was determined by examining 200 ascospores.

3.2.2. Olive brine medium and absolute fitness

Olives of the Oblica variety (approximately 200 g) collected in Évora, Portugal were used to prepare olive brine (200 ml H_2O, 8% NaCl w/v) during 3 months at 17 °C. After this period, the brine was sterilized by filtration and kept at 4 °C. This brine was analyzed by HPLC to identify and quantify the sugars and sugar-like compounds present. Sugar concentrations were determined using a carbohydrate analysis column (250 mm x 4 mm + Aminotrap, Dionex Carbopac PA10; DIONEX ICS3000). The column was kept at 25 °C and 0.1 M NaOH was used as the mobile phase at 1 ml min^{-1}. For the phenolic compounds the concentrations (%) were determined using a Waters Novapack C18 15 mm column (DIONEX ICS3000). The column was kept at 30 °C and 2% methanol was used as the mobile phase at 0.5 ml min^{-1}.

Two experiments were done independently for the group of strains selected. Pre-cultures (20 ml in 50 ml flasks) grown for 24 hours in YNB (Yeast Nitrogen Base, Difco) supplemented with 1% (w/v) glucose (incubation at 25 °C) were used to inoculate approximately 1 x 10^5 cells/ml in a volume of 2 ml of olive brine in 2 ml micro-centrifuge tubes. Cells were grown in batch cultures for 70 days at 17 °C without shaking, and cell viability was estimated by performing regular plate counts after preliminary counts in a hemocytometer. Statistical significance was tested using an unpaired *t*-test with Welch's correction, implemented in GraphPad Prism v5 (*p*-value cut-off < 0.01). At the end of the experiment, the supernatants of two cultures from the Olives population and two cultures

from the Wine population were randomly chosen for HPLC analysis to determine the residual concentrations of sugars.

3.2.3. Growth rates in NaCl and sugar consumption in phosphate buffer

To determine the growth rate in NaCl, strains were pre-grown overnight in liquid YPD medium [yeast extract 1% (w/v), peptone 2% (w/v), D-glucose 2% (w/v)], at 25 °C and were subsequently transferred to fresh medium (20 ml YPD or YPD supplemented with NaCl 6% or 8% (w/v)) and incubated with orbital shaking (180 rpm) at 30 °C in 50 ml flasks. The initial OD_{640nm} was 0.1 - 0.2. Growth rates were calculated in the exponential phase using OD_{640nm} measurements.

For the sugar consumption measurements, strains were pre-grown overnight at 25 °C in YNB medium supplemented with 1% (w/v) glucose. Cells were then transferred to phosphate buffer at pH 5.8 (30 ml in 50 ml flasks) supplemented with 0.6% (w/v) glucose, 0.1% (w/v) fructose, and 8% (w/v) NaCl, to mimic the conditions in olive brine and incubated at 20 °C. A similar experiment was also conducted in the absence of NaCl. Sugar consumption was monitored for 10 days by HPLC. Extracellular concentrations of fructose and glucose were determined using a carbohydrate analysis column (300 mm x 7 mm, Thermo HyperREZ XP Carbohydrate Ca++; KNAUER Smartline) and a differential refractometer. The column was kept at 85 °C and H_2O was used as a mobile phase at 0.6 ml min^{-1}.

3.2.4. Genome sequencing, read alignment, and genotype calling

DNA was extracted from overnight grown cultures of monosporic or single-cell derivatives and paired-end Illumina MiSeq 250 bp genomic reads were obtained after sequencing for 500 cycles. Genomic data for other strains was obtained from the NCBI-SRA archive and from the *Saccharomyces* genome resequencing project v2 (SGRP2) (Bergström et al., 2014). When only finished genome sequences were available in public databases (NCBI), the corresponding error-free Illumina reads were simulated using dwgsim (https://sourceforge.net/projects/dnaa/files/dwgsim/). Reads for each isolate were mapped to the *S. cerevisiae* reference genome (UCSC version SacCer3) using SMALT v0.7.5 aligner (http://www.sanger.ac.uk/resources/software/smalt/). The reference index was

built with a word length of 13 and a sampling step size of 2 (-k 13 -s 2). An intensive search for alignments (-x) was performed during the mapping step with the random assignment of ambiguous alignments switched off (-r 1) and the base quality threshold for the look-up of the hash index set to 10 (-q 10). With these settings, SMALT v0.7.5 only reports the best unique gapped alignment for each read. For the paired-end data the insert size distribution was inferred with the "sample" command of SMALT prior to mapping. Conversion of SAM format to BAM, sorting, indexing, several mapping statistics, and consensus genotype calling were performed using the tools available in the SAMtools package v.18 (Li et al., 2009) and as described previously (Almeida et al., 2014). Multiple sequence alignments for each reference chromosome were generated from the resulting fasta files. For downstream analysis, all bases with Phred quality score below Q40 (equivalent to a 99.99% base call accuracy) or ambiguous base calls were converted to "N". For obtaining the *S. cerevisiae* and *S. paradoxus* sub-genomes of the hybrid strains, reads for each strain were mapped to an extended *Saccharomyces* spp. reference with assembled sequences from the genomes of *S. cerevisiae* (UCSC version sacCer3), *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. uvarum* (Scannell et al., 2011), and *S. arboricola* (Liti et al., 2013).

3.2.5. Phylogenetic inference and divergence across the genome

Chromosomal single nucleotide polymorphisms (SNPs) were extracted from multiple sequence alignments only if the evaluated site was represented by unambiguous high confidence alleles in at least 85% of the strains. SNPs were then concatenated to generate the whole-genome SNP alignment. The phylogeny was inferred using maximum likelihood as implemented in IQ-TREE v1.6.7 (Nguyen et al., 2015) using an empirically determined substitution model, SH-like approximate likelihood ration test (1,000 replications) (Guindon et al., 2010), and rooted with *S. paradoxus*. The phylogeny was visualized using iTOL v3.0 (Letunic & Bork, 2016). Whole-genome levels of divergence were estimated using Variscan v2.0 (Hutter et al., 2006). Divergence was calculated for each mapped strain in comparison with the reference genome of *S. cerevisiae* using RunMode 21. The results were processed using a 10 kb sliding window with 10 kb step increments.

3.2.6. Screening for non-*S. cerevisiae* genes

Evidence of the presence of genes from other *Saccharomyces* species was investigated by mapping the reads to a combined reference that includes the annotated coding sequences of *S. arboricola*, *S. cerevisiae*, *S. kudriavzevi*, *S. mikatae*, *S. paradoxus*, and *S. uvarum* (Liti et al., 2013; Scannell et al., 2011). Reads were mapped to this combined reference using BWA v0.6.2 (Li & Durbin, 2009) with default parameters, but setting the quality threshold to 10 (-q 10). SAMtools v18 (Li et al., 2009) was then used for manipulation of the resulting BAM files. Only ORFs with orthologs unambiguously annotated in all species were analyzed. An ORF was considered to have a foreign origin to *S. cerevisiae* if its coverage was higher at least one-fourth of the median whole genome coverage for a given strain. The ORF coverage was defined as the product of the total number of mapped reads to a given ORF by the read length, dividing by the sum of all the ORFS length (considering ORFs that have at least 25% of reads mapped to, when comparing to the orthologous ORF with the highest number of reads). This measure was taken to control spurious alignment counts. The coverage threshold allowed for some heterogeneity in the read count and for the eventual presence of a foreign ORF together with native *S. cerevisiae* ORF. For some of the *S. paradoxus* genes detected in the hybrid genomes, their assignment to this species was confirmed with phylogenetic analyses involving homologous sequences from other *Saccharomyces* species.

3.2.7. GO analyses and survey of specific genes and of gene copy variation

Standard gene ontology (GO) term find was performed with the GO TERM FINDER tool v0.83, available at SGD, using a *p*-value cut-off of < 0.01. We performed de novo genome assemblies using SPAdes v3.11.1. Prior to assembly, reads were processed with Trimmomatic v0.36 based on quality score threshold of 20 for windowed trimming, discarding reads less than 100 bp in length or harboring ambiguities. To retrieve genes of interest, a local BLAST database was set up for each genome. Copy number variation of the two *CUP1* genes (*CUP1-1* and *CUP1-2*) and the three *ENA* genes (*ENA1*, *ENA2*, and *ENA5*) was investigated using CNVNator (Abyzov et al., 2011) on mapped genomes and using *ACT1* as control. The query sequences were defined by the coordinates in the reference sequence of *S. cerevisiae* for the coding regions of the genes of interest. The results obtained were manually validated by

checking the chromosomal context of the hits using UGENE (Okonechnikov et al., 2012) and by analyzing the copy number if the genes flanking the genes of interest.

3.3. Results

3.3.1. Ecology and phylogeny

Given earlier reports on the occurrence of *S. cerevisiae* in association with table olives (Arroyo-López et al., 2008; Bonatsou et al., 2018; Cromie et al., 2013) and with alpechin (Santa María, 1958), we asked if the original source of these yeasts was the olive tree itself. For this reason, we conducted an isolation program employing samples of olive tree bark, leaves, fruits and soil underneath the trees, and a selective enrichment protocol for yeasts of the genus *Saccharomyces*. In parallel, processed products such as olive oil and olive brine from table olives were also investigated. In total 163 samples from olive trees were investigated, together with 53 samples from olive oil and 7 samples from olive brine. Although the number of samples collected from olive trees was much higher, the frequency of isolation of *Saccharomyces* spp. was very low (3.7%), and only six strains were collected. The frequency of isolation in olive oil was higher (7.6%) and yielded four strains but was still markedly lower than that of olive brine (85.7%, six strains). Therefore, in total 16 new strains of *Saccharomyces* spp. were isolated from the olive niche.

Next, we obtained draft genome sequences of the new isolates in order to ascertain if they belonged to *S. cerevisiae* and, if so, to determine to which population they belonged to. As shown in Figure 3.1A, the new isolates were all identified as members of *S. cerevisiae*, thus showing that *S. paradoxus* was not isolated during our survey. Interestingly, the *S. cerevisiae* strains were found to belong to different populations. The six strains isolated directly from olive tree bark or ripe olives did not belong to the same population (Figure 3.1A and Table A2.1). A single strain had substantial genomic contributions from *S. paradoxus* and, accordingly, was assigned to the Olives clade. Two strains belonged to the Wine clade, one strain to the Sake clade and two additional strains occupied an isolated position in the phylogeny and subsequent analyses showed that they had "mosaic" genomes with major contributions from the Wine and North American - Japan clades. With respect to the four strains isolated from olive oil, three of them had *S. paradoxus* contributions and belonged to the Olives clade. The remaining strain belonged to the Wine clade. For the strains isolated

from olive brine, more homogenous results were obtained and all of them were found to belong to the Olives clade (Figure 3.1A).

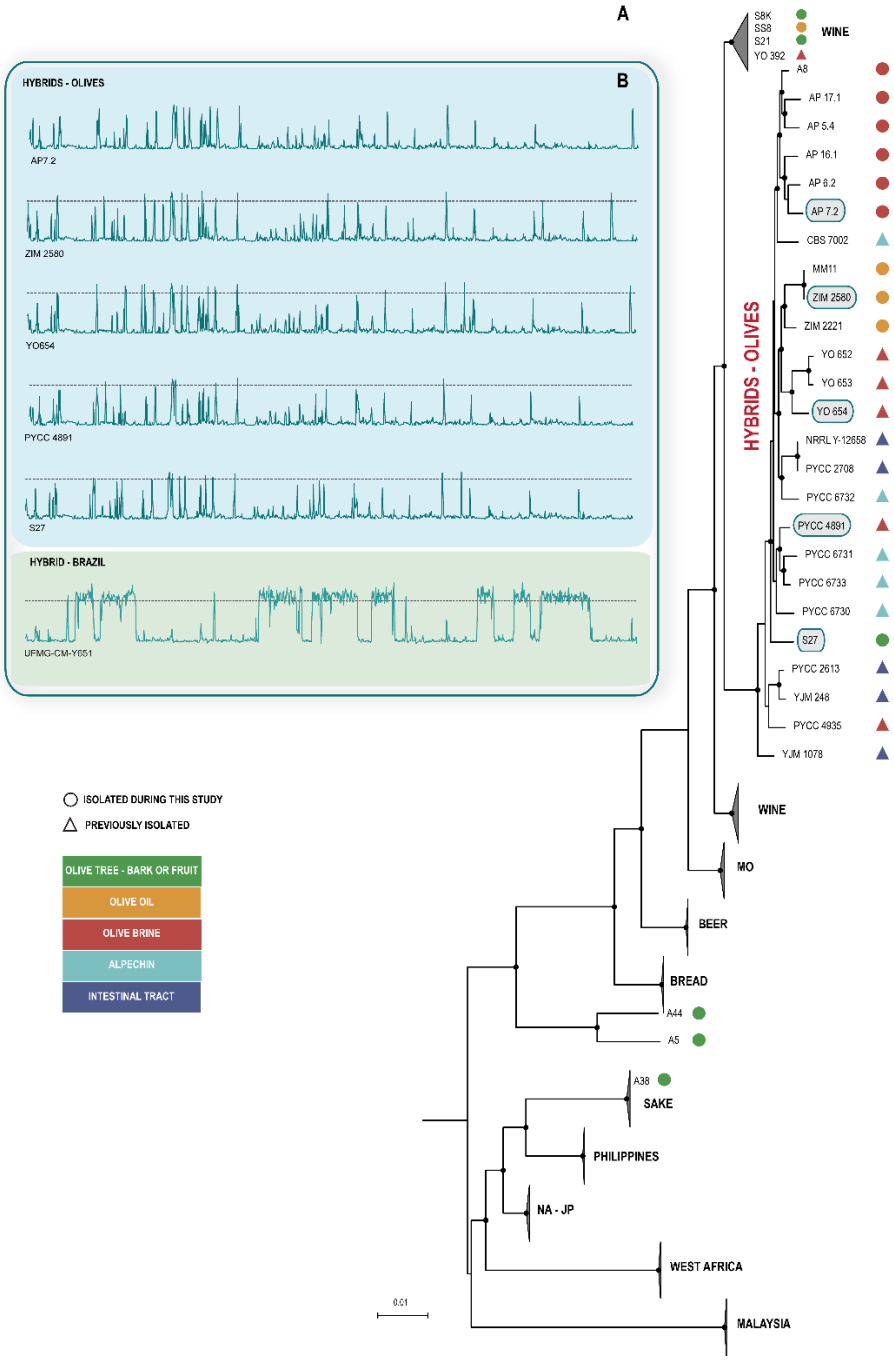


Figure 3.1 - Phylogenetic and divergence analysis of the olive strains hybrids. (A) Phylogenetic placement of hybrid olive strains among the known lineages of *S. cerevisiae* (MO, Mediterranean oaks; NA-JP, North America - Japan). Whole-genome phylogenetic tree constructed after discarding *S. paradoxus* regions in all genomes. The phylogeny was inferred from 93 sequences and 913,590 SNPs using the maximum likelihood method as implemented in IQ-TREE with the TVM+F+G4 model of sequence evolution and was rooted with *S. paradoxus*. Branch lengths correspond to the expected number of substitutions per site and black dots in tree nodes depict bootstrap support values above 85% (1000 replicates). Strains isolated from the olive niche are distinguished based on the specific isolation source (see color codes). (B) Divergence plots of the genomes of selected hybrid strains (highlighted in the phylogeny) to the reference genome of *S. cerevisiae*. The dotted lines depict the 10% divergence threshold that represents the average divergence between *S. cerevisiae* and *S. paradoxus*. The substantially distinct divergence plot of a Brazilian *S. cerevisiae* x *S. paradoxus* hybrid strain (UFMG-CM-Y651) previously reported by us (Barbosa et al., 2016) is included for comparison.

Our phylogenetic analysis included also other strains that belong to the Olives clade and that had been isolated as far back as 1957 from olives, olive brine, alpechin, and from the gut or feces of humans and pigs (Santa María, 1958, 1962; Van Uden & Assis-Lopes, 1957). It is noteworthy that the 25 *S. cerevisiae* x *S. paradoxus* hybrid strains isolated during this study or in previous studies formed a monophylum even if the phylogeny of Figure 3.1A was prepared only with *S. cerevisiae* ORFs, thus avoiding the strong bias that would be introduced if *S. paradoxus* ORFs were considered. This suggests that all hybrid strains share the same *S. cerevisiae* ancestor irrespective of the geographical origin and particular substrate from which they were collected, a possibility also supported by the divergence plots depicting the *S. cerevisiae* and *S. paradoxus* blocks along the genome, that were similar for all strains of the Olives clade (Figure 3.1B). Moreover, it appears the *S. cerevisiae* ancestor of the hybrids was a member of the Wine population.

In conclusion, the results from our ecological survey do not support the hypothesis that the members of the Olives clade reside in the olive tree environment. Although the possibility that such strains are associated with olive trees cannot be entirely ruled out, it appears more likely that the ecological niche of this clade is processed olives and their products like olive oil, alpechin, which is the corresponding waste product, and table olives / olive brine. Also, the occurrence of hybrid strains in the intestinal tract is of notice. Besides the two strains (YJM 248 and YJM 1078) already reported in Strope et al. (2015) we found three additional strains from this source (PYCC 2613, PYCC 2708, and PYCC 8033).

3.3.2. Fertility

Most strains (70%) of the Olives clade were sexually competent (Table A2.1). Spore viability for two strains in this clade (PYCC 4935 and YO652) ranged 95.5 - 96% and a cross between them was also fertile (90% spore viability), thus suggesting that sexual recombination within the clade can occur. Also, a cross between YO652 and the Wine strain EXF 6719 (97% spore viability) had an ascospore fertility of 87%, thus indicating that sexual contact between the Olives and Wine populations appears not to be significantly hampered.

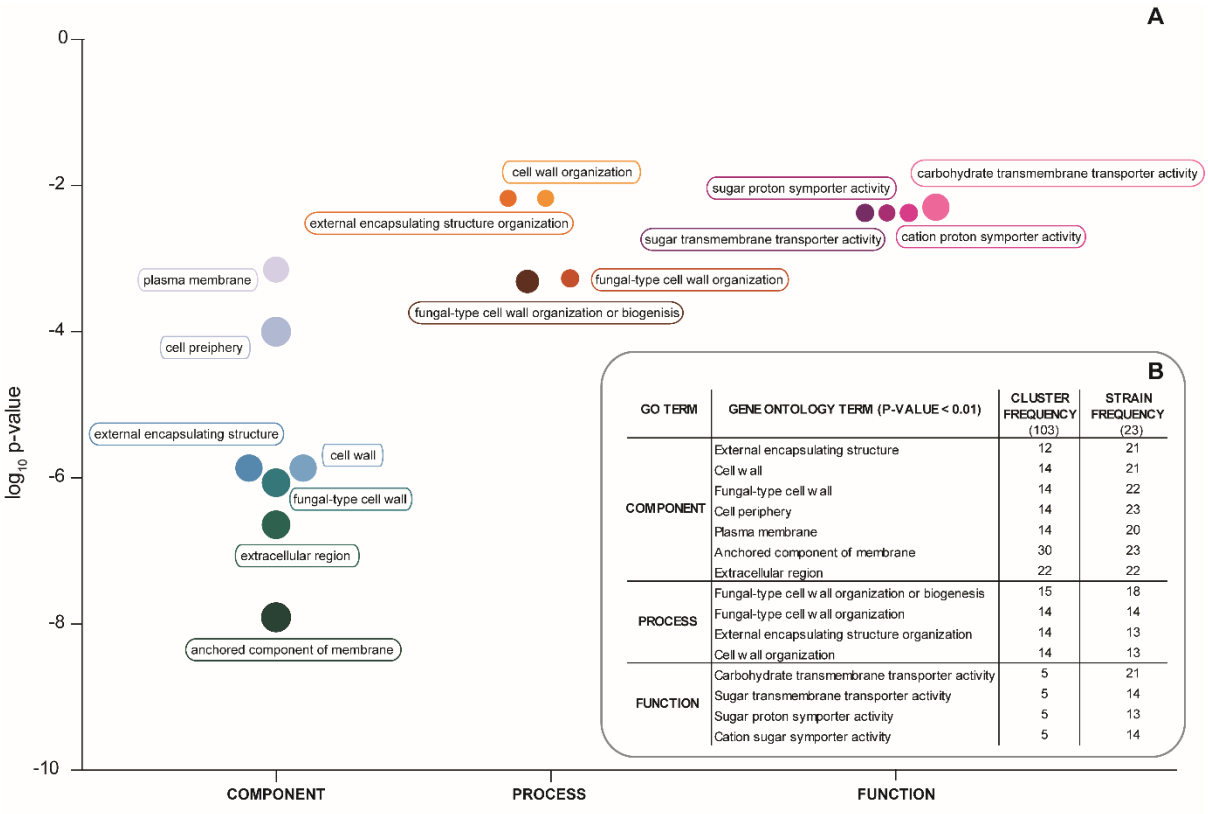
3.3.3. Genomic analysis

In order to characterize the genomic nature of the hybrids, we analyzed in detail 23 strains (Table A2.4). We detected a total of 540 *S. paradoxus* ORFs and between 193 and 314 *S. paradoxus* ORFs per strain, with 103 ORFs being shared among all strains. The *S. paradoxus* ORFs originated in the European population of this species (Figure A2.2), thus suggesting that the hybridization event occurred in this continent. The co-existence of *S. cerevisiae* and *S. paradoxus* alleles for a given ORF was not frequent. In total 148 ORFs (27.4% of the total number of *S. paradoxus* ORFs) were found to occur in that configuration in at least one strain. One strain was devoid of ORFs represented in the genome by alleles belonging to the two species and 16 strains had only two to four ORFs with *S. cerevisiae* and *S. paradoxus* alleles. Together, these strains represent 74% of the total number of strains analyzed. Three strains had between 12 and 25 ORFs with *S. cerevisiae* and *S. paradoxus* alleles and another tree strains had between 34 and 54 ORFs with *S. cerevisiae* and *S. paradoxus* alleles. The distribution of strains having more heterozygous ORFs did not show any association with the isolation substrate or with the phylogeny.

Gene ontology analysis of the 103 ORFs shared among all the strains revealed a significant enrichment in genes encoding for proteins of the fungal cell wall and plasma membrane, like *TIP1*, *HLR1*, *DAN1*, *FCY2*, and *STL1*. However, for several strains, when an individual analysis was performed, a significant result for an enrichment in hexose transporters was also observed (Figure 3.2 and Table A2.2). Taken together, these results support the view that the hybrid strains have a similar core genomic composition, thus suggesting that they share a common (hybrid) ancestor and also that after hybridization have evolved adaptations to the processed olives niche.

Given that the hybrid strains descend from a *S. cerevisiae* wine strain (Figure 3.1A), we surveyed the hybrid genomes for the presence of typical domestication signatures of wine strains (Almeida et al., 2015; Barbosa et al., 2018). For regions A, B, and C, that encompass 39 genes potentially relevant for the winemaking process acquired by horizontal gene transfer from non-*Saccharomyces* species, at least one of these regions was present in 68.8% of the control group of 32 Wine strains listed in Table A2.1, whereas only 13% (3 out of 23) hybrid genomes shared the same characteristic. It thus appears that these regions are less prevalent in hybrid strains, either because their ancestor *S. cerevisiae* wine strain already lacked most of them and/or because they are not relevant in the olives niche and were therefore lost. With respect to the inactivation of aquaporin genes *AQY1* and *AQY2*, associated with the domestication of wine strains and with the adaptive loss of those water channels, a trait that increases fitness in sugar-rich environments (Will et al., 2010), no differences were found between the two groups and all strains had at least one aquaporin gene coding for a non-

functional protein. We also investigated the variation of the number of copies of *CUP1*, a gene involved in resistance to copper toxicity in *S. cerevisiae*, especially in wine strains, due probably to their expose to copper sulfate used in vineyards (Fay et al., 2004; Strope et al., 2015). Copy number variation (CNV) of the two paralogs of *CUP1* among reference wine and wild (oak-associated) strains is shown in Table A2.1. Whereas among wine strains CNV of *CUP1* could exceed 30 (in two cases), the Mediterranean Oak (MO) strains did not show an enrichment in the number of *CUP1* copies. Some of the hybrid strains showed also elevated numbers of copies of *CUP1*, with the two most enriched genomes having 33 and 35 copies. Statistically, wine and olive strains could not be distinguished in terms of presence and expansion of *CUP1*.



B

GO TERM	GENE ONTOLOGY TERM (P-VALUE < 0.01)	CLUSTER FREQUENCY (103)	STRAIN FREQUENCY (23)
COMPONENT	External encapsulating structure	12	21
	Cell wall	14	21
	Fungal-type cell wall	14	22
	Cell periphery	14	23
	Plasma membrane	14	20
	Anchored component of membrane	30	23
	Extracellular region	22	22
PROCESS	Fungal-type cell wall organization or biogenesis	15	18
	Fungal-type cell wall organization	14	14
	External encapsulating structure organization	14	13
	Cell wall organization	14	13
FUNCTION	Carbohydrate transmembrane transporter activity	5	21
	Sugar transmembrane transporter activity	5	14
	Sugar proton symporter activity	5	13
	Cation sugar symporter activity	5	14

Figure 3.2 - Gene ontology of *S. paradoxus* genes found in hybrid strains. (A) Gene ontology terms with *p*-value < 0.01 and organized under "Component", "Process", and "Function" for the 103 *S. paradoxus* genes shared between the 23 hybrid genomes analyzed. The size of the circles is proportional to the number of genomes that contribute for that term. (B) Number of *S. paradoxus* genes and number of strains by GO (gene ontology) term.

3.3.4. Absolute fitness in olive brine

In order to investigate whether strains of the Olive clade were adapted to thrive in the processed olives niche, we estimated absolute fitness of a set of six strains isolated from olive

brine, olive oil, and alpechin, and compared it to six *S. cerevisiae* strains from the Wine population. We measured absolute fitness as the number of viable cells maintained in a long-term batch culture of table olive brine, here used as proxy for a habitat to which the *S. cerevisiae* hybrids are adapted to (Figure 3.3, Figure A2.1 and Table A2.3).

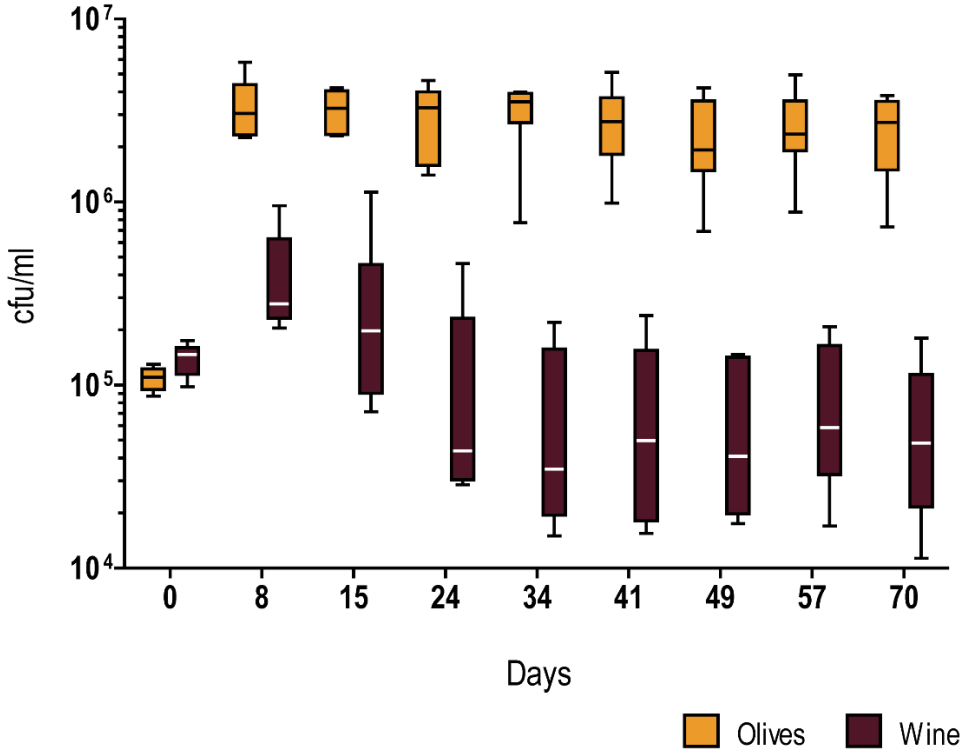


Figure 3.3 - Whiskers plots of the relative fitness (growth and survival) in olive brine of representatives of the Olives (AP 5.4, AP 7.2, YO 654, ZIM 2580, PYCC 4891, and PYCC 6732) and Wine (AWRI 1631, Lalvin W15, PR, PYCC 4072, TUM V1, and Uvaferm VRB) populations of *S. cerevisiae*. The results are based on counts of colony forming units/ml of 6 strains from each group inoculated individually in two duplicate and independent experiments.

Employing freshly collected ripe olives, we prepared a brine containing 8% (w/v) NaCl (see section "Materials and methods" for details) which was used to test each strain separately. The absolute fitness of the olive and wine strain cohorts was inferred by measuring viable cell numbers for 70 days in two independent experiments (Figure 3.3, Figure A2.1 and Table A2.3). Although strain fitness varies within both groups, variation is much more pronounced in the Wine group. In spite of the within-group differences among strains, a clear difference is observable between fitness of the Olive group and the Wine group ($p < 0.0001$, unpaired *t*-test with Welch's correction), the former being able both to attain higher cell numbers and to sustain viability throughout the duration of the experiment (70 days). On the contrary, wine strains tended to start losing viability already during the first month of incubation.

We reasoned that one possible cause for the difference in fitness between the two groups might be related to their ability to use nutrients available in olive brine. Contrary to what is

typical of the initial stages of wine fermentation, olive brine has low concentrations of sugars. To analyze this in more detail, we identified and quantified the sugars and sugar-related compounds present in olive brine and measured their consumption by two representatives of the wine and two representatives of the olive cohorts (Table 3.1). While strains belonging to the Olives clade virtually exhausted the glucose and fructose present initially in the brine, the representatives of the Wine clade consume only about half of the available sugars. Mannitol was left untouched in both cases.

Table 3.1 - Sugar consumption in olive brine by strains of the Olives and Wine population of *S. cerevisiae*.

Time (days)	**Population**	**Strain**	**pH**	**Glucose % (w/v)**	**Fructose % (w/v)**	**Mannitol % (w/v)**
0			5.1	0.57	0.1	0.14
70	Olives	YO 654	5.0	0.0055	0.0023	0.13
		AP 7.2	4.9	0.0058	0.0039	0.14
	Wine	Lalvin W15	5.2	0.26	0.077	0.13
		TUM V1	5.2	0.38	0.083	0.14

This finding was intriguing because some wine strains were previously found to have an impairment in high affinity hexose transport (the type of transporters expressed under the low sugar concentrations measured in olive brine), a trait that was subsequently associated to certain variants of the *HXT* hexose transporter genes (Luyten et al., 2002). Numerous *HXT* genes are present in the genomes of the species of the genus *Saccharomyces*, encoding transporters with different affinities for their substrates. Since hexose transporter genes encoding the main high affinity transporters (*HXT6/HXT7*) were among those "replaced" in the hybrids by their *S. paradoxus* counterparts (Table A2.4), we asked if these substitutions might have contributed to improve high affinity hexose transport in the hybrids.

To assess this, we compared the ability of the same strains of the Wine and Olives clades used in the previous experiment (see Table 3.1) to consume the sugars present in olive brine in the course of the first 8 days after inoculation of brine (Figure 3.4A). Surprisingly, and although in this period olive strains grew on average in the brine fitness experiments shown in Figure 3.3 an order of magnitude more than wine strains, sugar consumption was very similar between wine and olive strains. Fructose consumption in particular was indistinguishable, while olive strains seemed to assimilate glucose slightly better, an observation that nevertheless does not explain the differences in growth in brine between the two sets of strains (Figure 3.3).

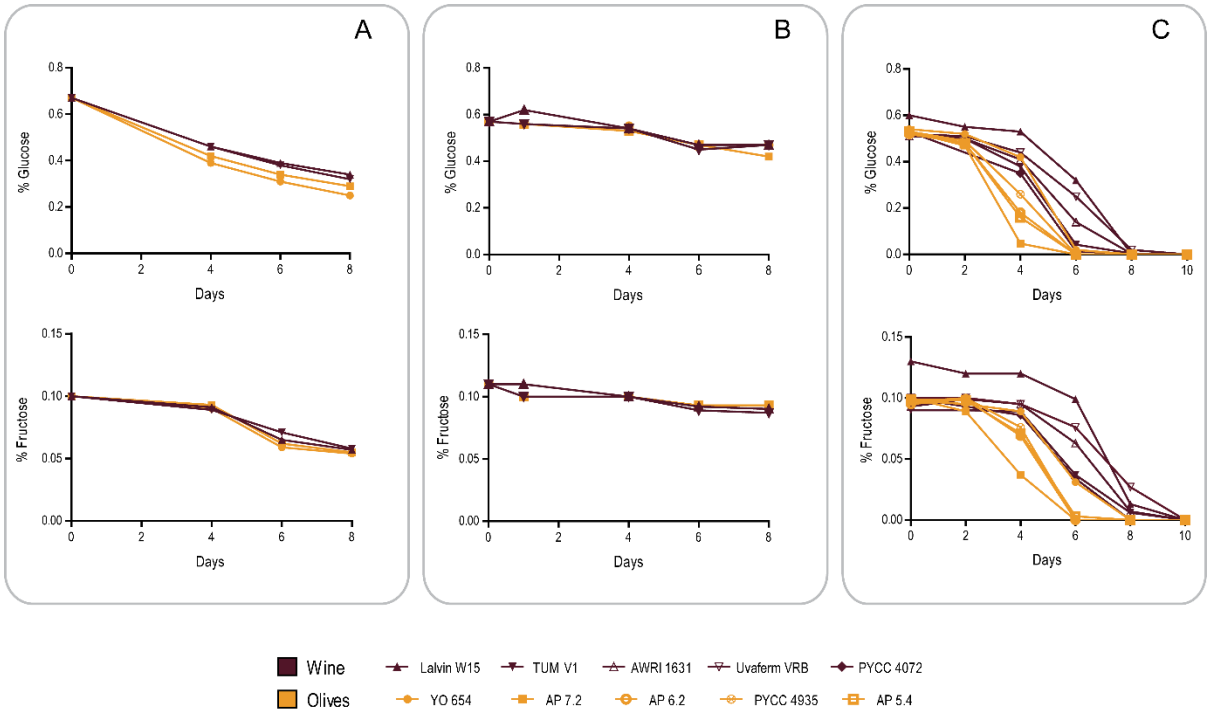


Figure 3.4 - Comparison of glucose (initial concentration 0.6% w/v) and fructose (initial concentration 0.1% w/v) consumption by representatives of the Olives and Wine population of *S. cerevisiae* in different conditions. (A) Olive brine. (B) Phosphate buffer supplemented with 8% w/v NaCl. (C) Phosphate buffer supplemented with 8% w/v NaCl and 0.1% w/v yeast extract.

A distinct experiment was subsequently performed in which brine was replaced by phosphate buffer supplemented with NaCl, glucose and fructose in concentrations identical to those found in brine. This time Wine strains seemed to be slightly more proficient in fructose assimilation while in glucose no clear differences were observed (Figure 3.4B). Nevertheless, when this experiment was performed without NaCl, glucose and fructose were totally consumed after 2 days by wine and olive strains. Taken together, these results, suggest that no considerable differences in sugar uptake capacities exist between the two groups of strains that justify the better growth of olive strains. It seems therefore that the observed differences in growth is due to a better capacity to adapt to the harsh conditions of olive brine, of which high salt concentration stand out, resulting in better growth for olive strains during the first 8 days while consuming the same amount of sugar in the same period as Wine strains. The same experiment as shown in Figure 3.4B was subsequently performed but this time adding 0.1% yeast extract to the phosphate buffer and adding more strains to both cohorts to increase representativeness (Figure 3.4C). The sugar consumption profiles of both strain cohorts were different in these conditions, with olive strains exhausting the available sugars significantly faster ($p < 0.05$, unpaired *t*-test), which means that in the presence of the required nutrients olive strains are better equipped to adapt to the high salt medium. Interestingly, under these conditions wine strains were capable of consuming all the sugar available after 8 - 10 days while they only consumed about 50% of the available sugars during the fitness

experiment in brine. This could mean that brine contains other inhibitors that affect the metabolism of wine strains more than that of olive strains, in addition to NaCl.

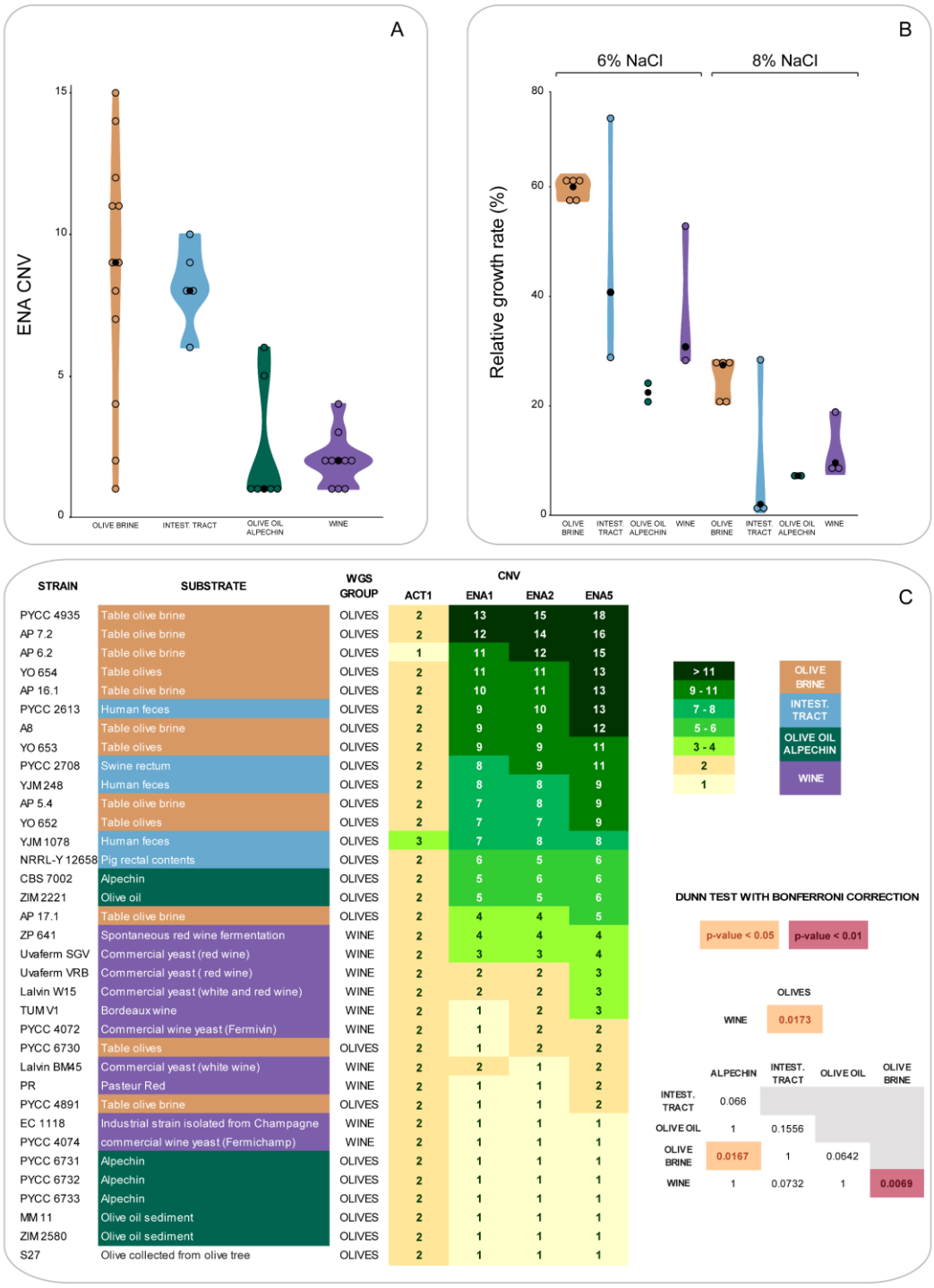
In summary, while *S. paradoxus HXT* genes may confer a slight advantage for glucose consumption in brine, this advantage does not explain the considerable difference in the ability of wine and olive strains to grow in brine. Instead, this difference seems to be derived from a better adaptation of olive strains to the particular conditions of brine of which the high NaCl concentration appears as a relevant factor.

3.3.5. Adaptation to NaCl

The results of the experiments shown in Figures 3.3 and 3.4 suggested that fitness and sugar consumption aptitude in brine might be related, at least partly, to salt resistance. To investigate this hypothesis, we started by determining the copy number of the three *ENA* genes found in *S. cerevisiae* reference genome (*ENA1*, *ENA2*, and *ENA5*) in the hybrid strains of the Olives clade and compared their abundance using 10 representative strains of the Wine clade. The ENA proteins are sodium pumps that help the cells to cope with an excess of sodium ions in their environment (Ruiz & Ariño, 2007). *ENA* copy number variation is shown in Figures 3.5A and 3.5C. Interestingly, the highest *ENA* copy number (14 - 18 copies) was detected among strains isolated from olive brine (Figure 3.5C). Overall, strains isolated from olive brine and the intestinal tract were more likely to have a higher number of *ENA* copies than strains isolated from olive oil, alpechin or wine (Figure 3.5A). It is possible that the strains found in the intestinal tract originate from olive brine environment, having been subsequently ingested. This would explain their increased number of *ENA* copies. The differences between the number of *ENA* copies were found to be statistically significant between the Wine and Olive Brine populations (Figure 3.5C, $p < 0.01$, Dunn *post hoc* test with Bonferroni correction). The comparison of *ENA* copy numbers was also significantly distinct when all hybrid strains from the olives niche were compared with wine strains and for the comparison between the olive brine and alpechin groups ($p < 0.05$).

To investigate to which extent *ENA* gene copy numbers determined fitness of the strains under study in the presence of salt, the ability of the various strains to grow in the presence of 6% and 8% NaCl was also tested (Figure 3.5B). There was, as expected, a correlation between *ENA* gene copy numbers and the ability to grow in the presence of NaCl, but this correlation was not complete. For example, while all strains isolated from olive brine performed well in the growth tests even if they had only moderately high *ENA* copy numbers (e.g., AP17.1; 4 - 5 copies), strains associated with the intestinal tract behaved heterogeneously,

varying between an excellent performance (PYCC 2613; 9 - 13 copies) and a very poor performance (e.g., YJM 1078; 7 - 8 copies).



STRAIN	SUBSTRATE	WGS GROUP	CNV ACT1	ENA1	ENA2	ENA5
PYCC 4935	Table olive brine	OLIVES	2	13	15	18
AP 7.2	Table olive brine	OLIVES	2	12	14	16
AP 6.2	Table olive brine	OLIVES	1	11	12	15
YO 654	Table olives	OLIVES	2	11	11	13
AP 16.1	Table olive brine	OLIVES	2	10	11	13
PYCC 2613	Human feces	OLIVES	2	9	10	13
A8	Table olive brine	OLIVES	2	9	9	12
YO 653	Table olives	OLIVES	2	9	9	11
PYCC 2708	Swine rectum	OLIVES	2	8	9	11
YJM 248	Human feces	OLIVES	2	8	8	9
AP 5.4	Table olive brine	OLIVES	2	7	8	9
YO 652	Table olives	OLIVES	2	7	7	9
YJM 1078	Human feces	OLIVES	3	7	8	8
NRRL-Y 12658	Pig rectal contents	OLIVES	2	6	5	6
CBS 7002	Alpechin	OLIVES	2	5	6	6
ZIM 2221	Olive oil	OLIVES	2	5	5	6
AP 17.1	Table olive brine	OLIVES	2	4	4	5
ZP 641	Spontaneous red wine fermentation	WINE	2	4	4	4
Uvaferm SGV	Commercial yeast (red wine)	WINE	2	3	3	4
Uvaferm VRB	Commercial yeast (red wine)	WINE	2	2	2	3
Lalvin W15	Commercial yeast (white and red wine)	WINE	2	2	2	3
TUM V1	Bordeaux wine	WINE	2	1	2	3
PYCC 4072	Commercial wine yeast (Fermivin)	WINE	2	1	2	2
PYCC 6730	Table olives	OLIVES	2	1	2	2
Lalvin BM45	Commercial yeast (white wine)	WINE	2	2	1	2
PR	Pasteur Red	WINE	2	1	1	2
PYCC 4891	Table olive brine	OLIVES	2	1	1	2
EC 1118	Industrial strain isolated from Champagne	WINE	2	1	1	1
PYCC 4074	commercial wine yeast (Fermichamp)	WINE	2	1	1	1
PYCC 6731	Alpechin	OLIVES	2	1	1	1
PYCC 6732	Alpechin	OLIVES	2	1	1	1
PYCC 6733	Alpechin	OLIVES	2	1	1	1
MM 11	Olive oil sediment	OLIVES	2	1	1	1
ZIM 2580	Olive oil sediment	OLIVES	2	1	1	1
S27	Olive collected from olive tree	OLIVES	2	1	1	1

Figure 3.5 - Copy number variation (CNV) of *ENA* genes and growth rates in the presence of NaCl of strains from the Olives and Wine populations. (A) Violin plots describing the number of *ENA* genes (*ENA1*, *ENA2* and *ENA5* average value for each strain) among olive brine, intestinal tract, olive oil-alpechin, and wine strains (black dots indicate the median within each group). (B) Violin plots of relative growth rates in the presence of 6 and 8% (w/v) NaCl (reference: medium without NaCl) among olive brine, intestinal tract, olive oil-alpechin, and wine strains. (C) Numbers of *ENA* copies shown in tabular format for each strain. Darker green color shades correspond to increased numbers of gene copies. CNV of actin (*ACT1*) is indicated as reference. Statistically significant differences of CNV between groups of strains are highlighted.

3.4. **Discussion**

Here we analyzed in detail a *S. cerevisiae* x *S. paradoxus* hybrid lineage associated with a distinctive artificial environment, that of processed olives. Even when only *S. cerevisiae* ORFs are considered, these hybrid strains form a distinct and exclusive monophyletic lineage among those already known for *S. cerevisiae*. This suggests that the olive hybrids are genetically isolated from the other *S. cerevisiae* populations and that it is likely that they descend from a single ancestral hybridization event. A set of other additional features also suggests that olive hybrids constitute a "natural" population in an evolutionary and ecological sense. First, sexual recombination appears to be possible within members of this population and secondly the olive hybrids have a considerable dissemination both in space and in time, since in this study we analyzed representatives from the Iberian Peninsula (Portugal and Spain), Southeast Europe (Slovenia and Croatia), and also strains from the United States. Moreover, these strains were collected during a period of time that spans six decades (1957 - 2018). The Olives population exhibits a particular ecological preference to environments having in common the presence of processed olives or their products, but not the olive tree itself. Therefore, it appears that the origin of this population is linked to human activities and to artificial substrates they create. Although some strains were found associated to the intestinal tract, these strains exhibit the characteristic expansion of the *ENA* gene copy number typical of olive brine strains, suggesting that they may have been ingested together with cured olives. The occurrence of *S. cerevisiae* hybrids in the intestinal tract parallels other reports of association of *S. cerevisiae* with humans (Angebault et al., 2013; Strope et al., 2015) and warrants the need for investigating if these strains are better adapted to survive in the intestinal tract.

A striking feature of the genomes of the strains of the Olives population is the markedly unbalanced contribution of the two parental sub-genomes, *S. cerevisiae* being the clearly prevalent sub-genome given that *S. paradoxus* contributes only around 3.7%. A likely scenario for the origin of the hybrid that originated the Olives lineage is homoploid hybridization. This would have corresponded to the fusion of a *S. paradoxus* meiospore with a *S. cerevisiae* meiospore resulting in a "normal" diploid hybrid genome that subsequently underwent the adaptive loss of most *S. paradoxus* sub-genome (Figure 3.6). Therefore, the observed genomic organization of the strains of the Olives lineage can be seen as reminiscent of a relic hybridization. This hybridization appears to have occurred in Europe since the *S. paradoxus* progenitor belongs to the recognized population of this species. We could determine that *S. cerevisiae* progenitor belonged to the Wine population and that the hybrids still exhibit some of the domestication signatures of this population such as the loss of functional aquaporins, the expansion of *CUP1* genes and the presence of region B, reminiscent of the presence of the

regions A, B, and C, typical in wine strains. It is noteworthy that these relic hybrids are capable of sexual reproduction, which would facilitate the emergence of a population adapted to a newly colonized niche. The ecological barrier between the processed olives niche and the vineyards/winery environment, even if incipient would also have promoted, together with selection, the ecological specialization of the new genotypes. Therefore, the model we propose to explain emergence of the Olives population is based on an original homoploid interspecies hybridization followed by a massive adaptive loss of heterozygosity (LOH) by replacement of most *S. paradoxus* alleles by their *S. cerevisiae* orthologs, combined with intra-population gene flow through sexual recombination and evolution of new ecological adaptations, with backcrossing with *S. cerevisiae* parent probably playing a very limited role. Similar cases of apparent reduction of the non-*cerevisiae* sub-genome have also been reported for artificially generated hybrids involving *S. kudriavzevii* (Lopandic et al., 2016) and *S. uvarum* (Antunovics et al., 2005).

Most importantly, LOH following hybridization, i.e., after a dramatic gain of genetic variation through interspecies hybridization, has been revealed as a major adaptation mechanism of populations when they invade new ecological niches (Smukowski Heil et al., 2017). Contrary to previous examples known exclusively from experimental evolution studies in the laboratory (e.g., Dunn et al., 2013; Smukowski Heil et al., 2017), the Olives population illustrates the fate of a relic hybridization in real conditions. It is also relevant to mention that although the fraction of *S. paradoxus* genome is relatively small, it is still much larger than instances of introgression of *S. paradoxus* in *S. cerevisiae* reported so far (Barbosa et al., 2016; Doniger et al., 2008), excluding the cases reported by Muller & McCusker (2009), in which information on strain origin was not given but that in fact correspond to the intestinal strains studied here. The relatively high number of homozygous *S. paradoxus* ORFs (342 out of 540) found in the Olives clade suggests that the genomic contribution of this species to adaptation to the processed olives environment is likely to involve a multiplicity of cellular processes. Gene ontology analysis of the set of *S. paradoxus* genes present in homozygosity in all hybrid strains examined here, suggested that cell wall function and hexose transport were likely among the cellular process benefiting from the *S. paradoxus* genomic contribution. Because inefficient high affinity hexose transport was found to be associated with specific *HXT* alleles carried by some wine strains (Luyten et al., 2002), we investigated whether the *S. paradoxus HXT* alleles retained by the olives strains were likely to contribute to the observed superior ability of these strains to consume sugars in brine. However, our experiments indicate that sugar consumption profiles are different between members of the Olives and Wine populations solely in the presence of NaCl and only when the cells are cultivated in nutritional conditions that support adaptation of their proteome to the high salt environment. Assuming that the *HXT* transporters operating when cells are cultivated with or without a nitrogen source are in both instances the high affinity *HXT6/7* transporters, these observations suggest

that the differences perceived are due to a better overall fitness of Olives strains in the presence of high NaCl concentrations, rather than to a better intrinsic ability of the *S. paradoxus HXT6/7* versions to operate in the presence of salt. We observed that the strains isolated from olive brine had a tendency for having an increased number of copies of *ENA* genes, a feature known to increase tolerance to NaCl. However, this tendency was not universal among olive brine strains and even strains with a lower number of *ENA* copies grew relatively well in the presence of NaCl, thus suggesting that other mechanisms might also be involved in the adaptation to NaCl of olive brine strains, as has been already documented (Dhar et al., 2011; Posas et al., 2000; Saito & Posas, 2012).

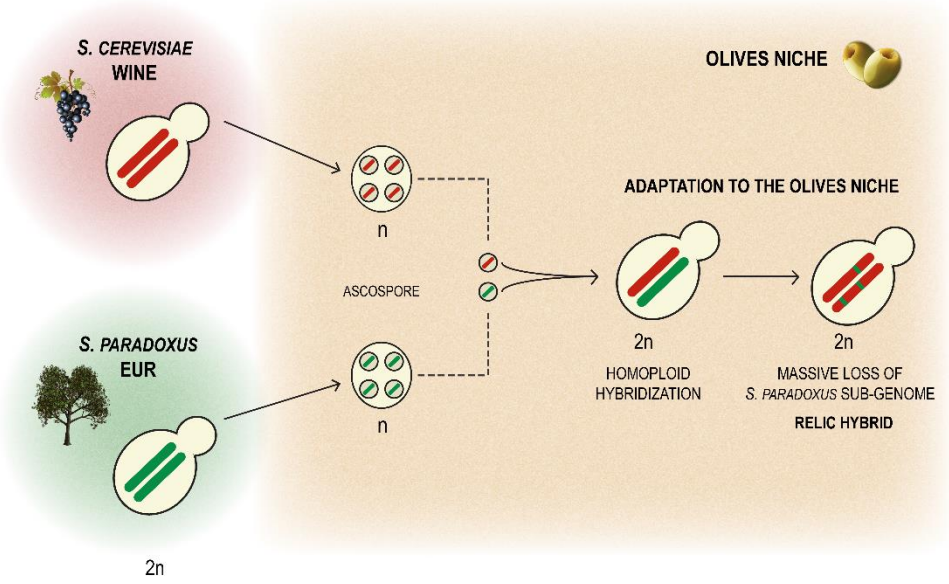


Figure 3.6 - Model for the origin of relic *S. cerevisiae* hybrids as a consequence of homoploid hybridization between *S. cerevisiae* wine strain and a *S. paradoxus* member of the European population, followed by adaptive LOH corresponding to a massive loss of the *S. paradoxus* sub-genome.

Interestingly, wine strains were able to exhaust glucose and fructose in the medium containing high NaCl concentrations while they failed to do so in brine even after 70 days, suggesting that inhibiting components other than NaCl are affecting the performance of wine strains in brine. Also, according to our results, these inhibitory components were not the phenolic compounds likely present in brine, since we failed to detect differences in the sensitivity of wine and olives strains (4 strains from each group) to oleuropein (3% w/v in YPD medium, pH 5), and ferulic acid (2% w/v in YPD medium, pH 4.5).

The emergence of relic olive hybrids from an already domesticated lineage (wine yeasts) in the artificial environment of processed olives can be seen as another instance of yeast domestication, or even a case of secondary domestication *sensu* Barbosa et al. (2018). However, contrary to wine and beer domestication, where genomic and phenotypic changes can be linked to characteristics of these beverages valued and improved over time by humans, in the

present case the beneficial role of the relic hybrids has not been clearly demonstrated in olive brine fermentations and therefore their origin and prevalence in the processed olives niche can be viewed as inconsequential to humans. One illustrative example of a tangible consequence of domestication is the inactivation of *PAD1* and *FDC1* genes in beer yeasts which overcomes the phenolic off flavor defect (Gallone et al., 2016; Gonçalves et al., 2016). The phenolic aroma, due to the formation of 4-vinyl guaicol, is negatively valued in most beers but not in wine where it can be even considered as desirable. The consequence of artificial selection is that beer yeasts differ from wine and wild strains in having acquired inactivating mutations in *PAD1* and *FDC1*. Therefore, if domestication is viewed as the controlled bred of an organism that becomes genetically distinct from its wild relatives in ways making it more useful to humans (Diamond, 2002), relic olive hybrids can be seen as a case of adaptation to the human environment but without the emergence of traits that we can readily recognize as useful. In order to reflect this distinct stage of "incomplete" domestication we define these changes as a quasi-domestication event.

4.

Revisiting the taxonomic synonyms and populations of *Saccharomyces cerevisiae* - phylogeny, phenotypes, ecology and domestication

The work presented in this chapter was published in:

Pontes, A.; Hutzler, M.; Brito, P.H.; Sampaio, J.P. Revisiting the Taxonomic Synonyms and Populations of *Saccharomyces cerevisiae*—Phylogeny, Phenotypes, Ecology and Domestication. Microorganisms 2020, 8, 903.

The sequencing data generated in this study have been deposited at the European Nucleotide Archive (ENA) under the accession code PRJEB36095.

Contributions:

Ana Pontes performed all the experimental work.

Ana Pontes executed all whole-genome sequence data analyses.

The contributions of the various authors were as follows:

A.P., and J.P.S. conceived the study, analyzed the data, and wrote the manuscript.

A.P., M.H., and P.H.B. executed experimental and bioinformatics analysis.

4.1. **Introduction**

Saccharomyces cerevisiae - the most emblematic and industrially relevant yeast species - was defined in 1838 by Meyen (Meyen, 1839) and typified, i.e., linked to a living type strain, by Hansen in 1883. In the following decades, physiological and morphologic characters were the sole criteria available to yeast taxonomists and even minor phenotypic differences were considered adequate for species delimitations, a practice that promoted a continuous rise in the number of accepted species. As consequence, Stelling-Dekker in 1931 (Stelling Dekker, 1931), Lodder and Kreger-van Rij in 1952 (Lodder & Kreger-van Rij, 1952) and van der Walt in 1970 (Van der Walt, 1970) recognized 23, 30 and 41 *Saccharomyces* species, respectively. As in other microbial groups, yeast classification based on phenotypic markers was challenged when DNA based methods started to be implemented as taxonomic tools. As a result, numerous *Saccharomyces* species were recognized as synonyms of *S. cerevisiae* (Yarrow, 1984) and a list of approximately 100 species or varietal names was merged into that species. Such taxonomical changes were continued in subsequent decades and generated an understandable confusion among scientists and industry professionals not versed in yeast systematics.

More recently, complete genome sequencing has provided the necessary resolution to dissect *S. cerevisiae* at the population level and various studies have revealed a complex scenario with wild, domesticated, and admixed groups (Almeida et al., 2015; Liti et al., 2009; Peter et al., 2018; Strope et al., 2015). Here, we use genomics to investigate a group of 45 reference strains (type strains) of former *Saccharomyces* species - currently regarded as synonyms of *S. cerevisiae* - many of which represent strains isolated from human-made fermentations worldwide. We show that these variants are distributed across the phylogenetic spectrum of domesticated lineages of *S. cerevisiae*, with a majority belonging to the most relevant technological groups, but totally absent in wild lineages. We also examine salient features of domesticated lineages and reassess the biogeography and ecology of wild lineages.

4.2. Materials and Methods

4.2.1. Genome Sequencing, Read Alignment and Genotype Calling

Paired-end Illumina Miseq (500 cycles) or NextSeq (300 cycles) reads were obtained for 37 strains. Genomic data for other strains were retrieved from public datasets as indicated in Table A3.1. When only finished genome sequences were available, the corresponding error-free Illumina reads were simulated using dwgsim.

Reads for each isolate were mapped to an extended *Saccharomyces* spp. reference containing sequences of *S. cerevisiae* (UCSC version sacCer3), *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. uvarum* (Scannell et al., 2011) and *S. arboricola* (Liti et al., 2013), as previously described (Borneman et al., 2016) and using SMALT v. 0.7.5 aligner. The reference Index was built with a word length of 13 and a sampling step size of 2 (-k1 3 -s 2). An intensive search for alignments (-x) was performed during the mapping step with the random assignment of ambiguous alignments switched off (-r 1) and the base quality threshold for the look-up of the hash index set to 10 (-q 10). With these settings, SMALT v. 0.7.5 only reports the best unique gapped alignment for each read. The insert size distribution was inferred with the "sample" command of SMALT prior to mapping. Conversion of SAM format to BAM, sorting, indexing, mapping statistics and consensus genotype calling were performed using the tools available in the SAMtools package v 1.18 (Li et al., 2009) and as described previously (Almeida et al., 2014). Multiple sequence alignments for each reference chromosome were generated from the resulting fasta files. For downstream analysis, all bases with a Phred quality score below Q40 (equivalent to a 99.99% base call accuracy) or ambiguous base calls were converted to 'N'.

4.2.2. Phylogenetic analysis and survey of specific genes

For the construction of the main phylogeny, if contributions from non-*S. cerevisiae* species were detected, only the *S. cerevisiae* sub-genome was considered. Chromosomal single nucleotide polymorphisms (SNPs) were extracted from a multiple sequence alignment only if the SNP was present unambiguously in at least 85% of the strains in the alignment. SNPs were then concatenated to generate a whole genome SNP alignment. Strains with more than 20,000 heterozygous sites with a Phred quality score above Q40 were selected for phasing. The BAM file of each strain with paired-end read sequences mapped to the reference genome was analyzed with the phase command of SAMtools to infer both phases, thus solving the

heterozygous SNPs. The -F option was used to exclude errors from unmapped or misaligned sequences. One haplotype per strain was randomly chosen and used in subsequent analysis. The main phylogeny was inferred using the maximum likelihood method as implemented in IQ-TREE v. 1.6.11 (Nguyen et al., 2015), using the TVM+F+ASC+G4 model of sequence evolution and the ultrafast bootstrap approximation with 1000 replicates (Hoang et al., 2018). The software iTOL v. 3.0 (Letunic & Bork, 2016) was used for visualization. Single gene phylogenies were constructed with MEGA 7 (Kumar et al., 2016), using Tamura 3-parameter model and the maximum likelihood method.

For the study of specific genes, whole genome assemblies were prepared with SPAdes v. 3.13.1. Prior to assembly, reads were processed with trimmomatic v. 0.36 to remove adapter sequences. In order to retrieve genes of interest, a local BLAST database was set for each genome and ORFS were searched with BLASTN using as queries: *AQY1* and *AQY2* of YPS163 (Will et al., 2010), regions A, B and C of EC1118 (Novo et al., 2009), *RTM1*, *BIO1* and *BIO6* of CEN-PK13 (Nijkamp et al., 2012), *MEL1* of UWOPS 03-461.4 and UWOPS 91-917.1 (Warringer et al., 2011) and *STA1* (Krogerus et al., 2019). Gene copy number was estimated by mapping the reads from each strain against the genome of the strain UWOPS 03-461.4 (Yue et al., 2017), as described above. The median genome coverage was estimated as the coverage of each nuclear chromosome. Following a previous study (Peter et al., 2018), the ratio between the median coverage of each individual gene and the values of the genome coverage was considered a good estimate of the gene copy number.

4.2.3. Phenotypic analyses

Strains investigated for the diastase phenotype were precultured in 20 mL of YPD (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose) overnight at 25 °C. Washed cells were then inoculated in 50 mL of beer wort (11° Brix) and incubated for 15 days at 25 °C. The decrease of the °Brix was monitored and strains that in the end of the incubation period had a °Brix lower than 4 were considered positive for the diastase phenotype. Strains investigated for the ability to grow on melibiose were precultured in 200 µL of 1% (w/v) melibiose and YNB. Growth was followed by measuring absorbance at 640 nm for three days in a Tecan Spark (Tecan Trading, Männedorf, Switzerland) microplate reader incubated at 25 °C.

4.3. **Results**

4.3.1. Global phylogenetic analysis reveals an unequal distribution of type strains of former species

We analyzed the genomes of 45 strains considered in the past as distinct *Saccharomyces* species (or varieties), but now regarded as *S. cerevisiae* synonyms, together with the valid type strain of *S. cerevisiae* (Table 4.1) and an additional 202 genomes representing the population diversity and technological variants currently known (Table A3.1). This is the most comprehensive dataset representing the diversity in this species. The phylogenetic analysis of these 247 genomes is shown in Figure 4.1 and a simplified unrooted phylogeny is shown in Figure A3.1. We identified 27 clades that represent the various technological variants of *S. cerevisiae* or geographical or ecological populations, many of which have been recognized previously by us or by other authors. In order to contribute to standardizing the designations of populations of *S. cerevisiae*, we provide a list of designations used here and a comparison with designations employed in other studies in Table 4.2.

We also classified the 27 clades according to their association with domesticate or wild populations (Figure 4.1 and Table 4.2). In the first case the strains are associated with processes like wine or beer fermentations and are expected to have acquired specific domestication signatures. In the second case the populations have an ecological or geographic circumscription, like for example the Mediterranean oaks population. Except for PYCC 4654, the type strain of *S. fructuum*, that occupies an isolated position in the phylogeny outside any of the recognized clades, all other type strains formed clades with technological, and therefore domesticated groups, namely Wine (two clades), Olives, Beer (two clades), Bread, Dairy, Sake and West African groups (Figure 4.1). Therefore, the distribution of type strains former species is markedly unbalanced across the spectrum of genetic diversity of *S. cerevisiae*. The Wine and Olives groups congregated most (29 out of 45) of the type strains analyzed, followed by two Beer groups that together gathered seven type strains, including the recognized type strain of *S. cerevisiae* (PYCC 4455), a beer strain assigned to the Beer 1 clade. Overall, nine of the 27 clades that represent distinct populations of *S. cerevisiae*, included at least one type strain. Remarkably, all the wild populations like those found in arboreal niches in China, Europe or the Americas did not include any of the type strains of former species. A more detailed analysis of the distribution of former *Saccharomyces* type strains in the context of the groups to which they were assigned is presented below.

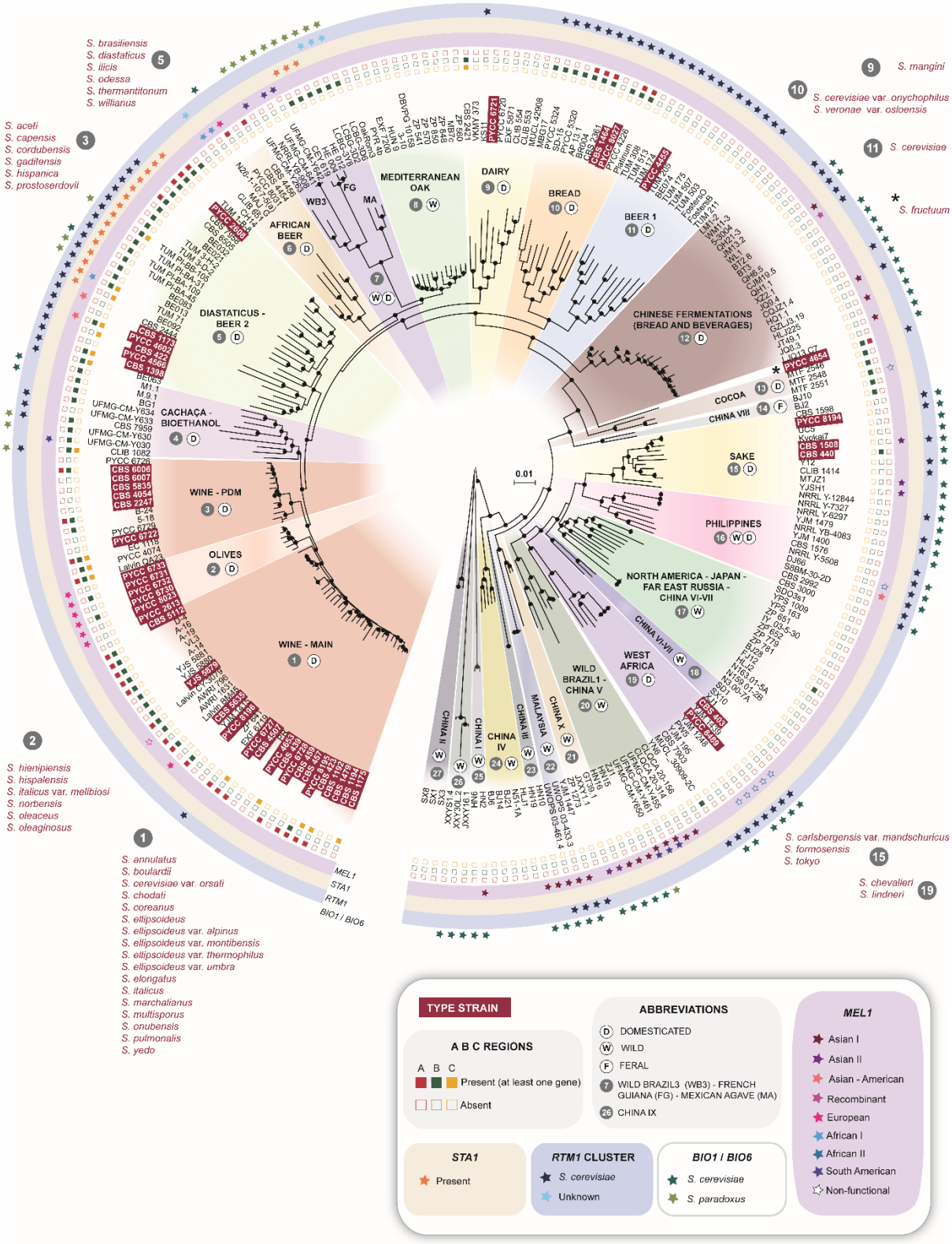


Figure 4.1 - Current synonyms of *S. cerevisiae* are unequally distributed among the known lineages of the species. The phylogeny was inferred from 248 sequences and 1,520,302 single nucleotide polymorphisms using TVM+F+ASC+G4 model of sequence evolution and the maximum likelihood method as implemented in IQ-TREE and was rooted with *S. paradoxus*. Branch lengths correspond to the expected number of substitutions per site and black dots in tree nodes depict bootstrap support values above 95% (1000 replicates). The 27 clades detected are numbered in gray circles, the former type strains of current synonyms of *S. cerevisiae* are highlighted in red rectangles.

Table 4.1 - Synonyms of *Saccharomyces cerevisiae* analyzed in this study and relevant information pertaining to them (the type strain of *S. cerevisiae* is also included).

Original Name	Strain Number#	Strain Number (Other Collections)	Strain Status	Population	Substrate of Isolation	Locality of Isolation
Saccharomyces aceti Santa María 1958	CBS 4054	-	type strain	WINE-PDM	red wine	Spain
Saccharomyces annulatus Negroni 1929	PYCC 6727	CBS 1227	type strain	WINE-MAIN	abscess on epididymis	New Zealand
Saccharomyces boulardii Seguela, Bastide & Massot 1984 (invalid name)	YJS 5879	-	type strain	WINE-MAIN	lychee	Vietnam
Saccharomyces brasiliensis Lindner 1909; *Saccharomyces logos* van Laer ex Jörgenssen 1909	PYCC 4602	CBS 382	type strain	BEER 2 - DIASTATICUS	beer (Logos brewery)	Rio de Janeiro, Brazil
Saccharomyces capensis van der Walt & Tscheuschner 1956	CBS 2247	-	type strain	WINE-PDM	grape must	South Africa
Saccharomyces cerevisiae Meyen ex Hansen 1883	PYCC 4455	CBS 1171	neotype strain	BEER 1	brewer's top yeast	Oranjeboom brewery, Rotterdam, Netherlands
Saccharomyces cerevisiae Hansen var. *ony-chophilus* Zach 1934	CBS 1464	-	type strain	BREAD	nail of 4-year-old girl	Austria
Saccharomyces cerevisiae Hansen var. *orsati* Steiner 1924	CBS 1175	-	type strain	WINE-MAIN	wine	unknown
Saccharomyces chevalieri Guilliermond 1914	PYCC 8489	NRRL Y-12633; CBS 400	type strain	WEST AFRICA	palm wine from Elaeis guineensis	Ivory Coast
Saccharomyces chodati Steiner 1924	CBS 423	-	type strain	WINE-MAIN	wine	Switzerland
Saccharomyces cordubensis Santa María 1970	CBS 6007	-	type strain	WINE-PDM	wine	Spain
Saccharomyces coreanus Saito 1910	CBS 5635	-	neotype strain	WINE-MAIN	grape must	South Africa
Saccharomyces diastaticus Andrews & Gilli-land ex van der Walt 1965	PYCC 2608	CBS 1782	type strain	BEER 2 - DIASTATICUS	super-attenuated beer	unknown
Saccharomyces carlsbergensis Hansen var. *mandschuricus* (Saito) Stelling-Dekker 1931	CBS 1508	-	type strain	SAKE	starter for sorghum brandy	unknown
Saccharomyces ellipsoideus Reess 1870	PYCC 4653	CBS 1395; NRRL Y-1529	neotype strain	WINE-MAIN	unknown	unknown

Table 4.1 - (continued).

Original Name	Strain Number#	Strain Number (Other Collections)	Strain Status	Population	Substrate of Isolation	Locality of Isolation
Saccharomyces ellipsoideus Hansen var. *umbra* Castelli 1933	CBS 457	-	type strain	WINE-PDM	grape must	Italy
Saccharomyces ellipsoideus Hansen var. *alpinus* Steiner 1924	CBS 1192	-	type strain	WINE-MAIN	wine	unknown
Saccharomyces ellipsoideus Hansen var. *montibensis* Steiner 1924	CBS 1479	-	type strain	WINE-MAIN	wine	unknown
Saccharomyces ellipsoideus Hansen ssp. *thermophilus* Steiner 1924	CBS 1194	-	type strain	WINE-MAIN	wine	unknown
Saccharomyces fructuum Lodder & Kreger-van Rij 1952	PYCC 4654	CBS 1544	type strain	OUTLIER	fermenting fruit juice	Netherlands
Saccharomyces elongatus Krumbholz 1932	CBS 439	-	type strain	WINE-MAIN	silvaner grapes	Germany
Saccharomyces formosensis Nakazawa 1933	CBS 440	-	type strain	SAKE	molasses	Taiwan
Saccharomyces gaditensis Santa María 1970	CBS 6006	-	type strain	WINE-PDM	wine	Spain
Saccharomyces hienipiensis Santa María 1962	PYCC 6733	VKM Y-1235	type strain	OLIVES	alpechin	Spain
Saccharomyces hispalensis Santa María 1978	PYCC 8023	CBS 7002	type strain	OLIVES	alpechin	Seville, Spain
Saccharomyces hispanica Santa María 1968	CBS 5835	-	type strain	WINE-PDM	wine	Spain
Saccharomyces ilicis Grönlund 1893	CBS 1173	-	type strain	OUTLIER	fruit of *Ilex aquifolium*	unknown
Saccharomyces italicus Castelli 1939	CBS 459	-	type strain	WINE-MAIN	grape must	Italy
Saccharomyces italicus var. *melibiosi* van Uden & Assis-Lopes 1957	PYCC 2613	CBS 2909	type strain	OLIVES	feces of man	Portugal
Saccharomyces lindneri Guilliermond 1914	CBS 403	PYCC 4571	type strain	WEST AFRICA	ginger beer from *Zinziber officinale*	West Africa
Saccharomyces mangini var. *casei* Saccchetti 1933	PYCC 6721	CBS 420; VKM Y-482	type strain	DAIRY	stracchino cheese	Italy
Saccharomyces marchalianus Ku_erath 1920	PYCC 8196	CBS 1460	type strain	WINE-MAIN	fermenting fruit	Indonesia
Saccharomyces multisporus Jörgensen 1909	CBS 4507	-	type strain	WINE-MAIN	English top brewing yeast	unknown

Table 4.2 - Populations of *S. cerevisiae* and their designations.

Population	Designation in Other Studies	Life Style[a]	Number of Type Strains of Former Saccharomyces Species	Comments
1- Wine-main	Wine (Legras et al., 2018); Wine/European (Liti et al., 2009; Peter et al., 2018)	D	17	Global distribution probably associated with the widespread dissemination of winemaking
2-Olives	Alpechin (Peter et al., 2018); Olives (Cromie et al., 2013; Pontes et al., 2019)	D	6	Ancestral hybrids with *S. paradoxus* associated with processed olives
3-Wine-PDM	Flor (Legras et al., 2007, 2018); Georgian (Peter et al., 2018); PDM-Prise de Mousse [12]	D	6	Lalvin EC1118, a well-known commercial strain in this group is also known as "Prise de mousse" or Champagne strain (Legras et al., 2007; Novo et al., 2009). This population also contains the Spanish flor yeasts of Xerez wine (Sherry) and similar wines (Legras et al., 2007)
4-Cachaça-Bioethanol	Brazilian bioethanol (Peter et al., 2018); Cachaça C1 and C2 (Barbosa et al., 2018); Rum and bioethanol (Legras et al., 2018)	D	-	This group is a secondary domesticate derived from wine strains and includes two types of cachaça strains (C1 and C2) and bioethanol strains (Barbosa et al., 2018)
5-Beer 2-Diastaticus	Beer 2 (Gallone et al., 2016); Mosaic beer (Peter et al., 2018)	D	6	Contains Saison-type low-gravity beer strains and beer spoilage strains with diastase activity
6-African beer	African beer (Peter et al., 2018)	D	-	Includes strains that ferment malted millet to produce bantu beer or malted sorghum to produce bili-bili or kaffir beer
7-Wild Brazil - French Guiana - Mexican agave	Mexican agave, French Guiana human (Peter et al., 2018); Wild Brazil B3 (Barbosa et al., 2016)	W/D	-	Complex clade composed of three subclades, one containing wild strains found in Brazil, another containing strains used in artisanal mezcal fermentation in Mexico and a third one containing hut associated strains from French Guiana aborigines; this last group contains *STA1* and was also found in cachiri, a traditional beer made from chewed and fermented starch-rich manioc (Angebault et al., 2013)

[a] Life Style: W - Wild; D - Domesticated; F - Feral.

Table 4.2 - (continued).

Population	Designation in Other Studies	Life Style [a]	Number of Type Strains of Former Saccharomyces Species	Comments
8-Mediterranean oak	-	W	-	Unique wild population with no evident links to the core group of Asian wild populations; no changes in designation since the original description (Almeida et al., 2015)
9-Dairy	Cheese (Legras et al., 2018); French dairy (Peter et al., 2018); Milk (Duan et al., 2018)	D	1	Recently revealed population associated with dairy products and adapted to galactose utilization
10-Bread	Active dry yeast (Duan et al., 2018); Mixed (Gallone et al., 2016)	D	2	
11-Beer 1	Ale beer (Peter et al., 2018); Beer (Legras et al., 2018); Beer 1 (Gallone et al., 2016)	D	1	Various sub-populations associated with different ale-beer types
12-Chinese fermentations	Mantou (bread)/Baijiu (distilled)/Huangjiu (rice wine)/Qingkejiu (barely wine)/fermented milk (Duan et al., 2018)	D	-	Predominantly Chinese domesticated strains that are distinct from strains of the Sake clade
13-Cocoa	West African cocoa (Peter et al., 2018)	D	-	
14-China VIII	-	F	-	Found in the arboreal niche in China, but with several domestication signatures (*MAL* gene expansion, *AQY* inactivation, presence of region B)
15-Sake	Asia (Cromie et al., 2013; Gallone et al., 2016); Asian fermentation, Sake (Liti et al., 2009; Peter et al., 2018)	D	3	
16-Philippines	Asian islands (Peter et al., 2018); Philippines (Cromie et al., 2013)	W/D	-	

[a] Life Style: W - Wild; D - Domesticated; F - Feral.

Table 4.2 - (continued).

Population	Designation in Other Studies	Life Style[a]	Number of Type Strains of Former Saccharomyces Species	Comments
17-North America-Japan - Far East Russia - China VI-VII	North America (Liti et al., 2009)	W	-	A wild Chinese population also found in North America, Japan, and Russia
18-China VI-VII	-	W	-	No changes in designation since the original description (Wang et al., 2012)
19-West Africa	African palm wine (Peter et al., 2018); West Africa (Liti et al., 2009)	D	2	
20-Wild Brazil 1 - China V	Ecuadorean (Peter et al., 2018);Wild Brazil B1 (Barbosa et al., 2016)	W	-	A wild Chinese population also found in Brazil and Ecuador
21-China I	-	W	-	No changes in designation since the original description (Wang et al., 2012)
22-Malaysia	-	W	-	No changes in designation since the original description (Liti et al., 2009)
23-ChinaIII	-	W	-	No changes in designation since the original description (Wang et al., 2012)
24-China IV	-	W	-	No changes in designation since the original description (Wang et al., 2012)
25-China I	-	W	-	No changes in designation since the original description (Wang et al., 2012)
26-China IX	-	W	-	No changes in designation since the original description (Wang et al., 2012)
27-China II	-	W	-	No changes in designation since the original description (Wang et al., 2012)

[a] Life Style: W - Wild; D - Domesticated; F - Feral.

4.3.2. The Wine and Olives clades harbour most type strains of former species

The largest number of type strains of former *Saccharomyces* species (23) was assigned to the two clades that congregate wine yeasts (clades 1 and 3, the main Wine clade and the Prise de Mousse clade, respectively). Most of these former type strains (16) were isolated from wine or grapes and two are clinical isolates. The remaining strains were isolated from other fruits (2) or other fermented beverages (2) and one strain has unknown origin. Previous studies have revealed that an important feature of the Wine clade is the inclusion of strains exhibiting typical genomic markers of a wine strain, but that were isolated in a clinical setting, mostly as human opportunists (Peter et al., 2018; Strope et al., 2015). The two former type strains mentioned above were isolated from similar settings, *S. annulatus* from a human skin infection in 1929 and *S. pulmonalis* from the sputum of a tuberculosis patient in 1925, thus confirming that the association between clinical and wine strains is, at least, one century old. Another clinically relevant strain is the reference strain of *S. boulardii*, widely employed as a probiotic (Zanello et al., 2009). This synonym of *S. cerevisiae* was isolated from a lychee in Vietnam but has the main genomic features of a wine strain (Figure 4.1, Table A3.1). Therefore, this clade includes both clinical and probiotic strains, which do not have clear discernible features at the genome level.

As already mentioned, all the former type strains isolated from wine investigated here joined clades 1 or 3. Clade 3 gathers a particular type of wine yeasts relevant for Jerez (Sherry) wine, a type of wine that undergoes an aging process in which an adapted variant of wine yeasts, the flor yeasts, forms a surface biofilm and contributes to the production of specific flavors and aromas (Legras et al., 2014). Other genomically similar strains isolated in other countries like Hungary, France or Italy, are particularly resistant to the stresses associated with the advanced stages of fermentation. Probably due to this, commercial strains from this group like the well-known EC 1118 (Novo et al., 2009), proved to be adequate for secondary fermentations in champagne, a process also designated as "prise de mousse" (Torresi et al., 2011) or to re-start stuck fermentations. Four of the six former type strains associated with the Prise de mousse clade were isolated in Spain and the remaining two in Armenia and South Africa. The Prise de Mousse clade includes also Georgian wine strains (Figure 4.1, Table A3.1) isolated from amphorae (Peter et al., 2018). It was recently hypothesized that since the Caucasus is thought to be the birthplace of winemaking, these Georgian isolates could represent the closest relatives of the first domesticated wine strains (Peter et al., 2018). However, not only other Georgian strains cluster in the main wine clade and not the in the

Prise the Mousse clade (Figure1, Table S1), but, as indicated above, several other strains of different origins and with highly specialized phenotypes are part of the Prise de Mousse clade, that therefore may be derived instead of ancestral.

As previously discussed (Barbosa et al., 2018; Legras et al., 2018; Marsit et al., 2017; Novo et al., 2009), three genomic regions usually designated as A, B and C, were horizontally transferred from non-*Saccharomyces* yeasts. They encompass 39 genes potentially relevant for the winemaking process. These regions are well-represented in clades 1 and 3 (Figure 4.1). In line with this, these regions were also more prevalent in the type strains that joined the two wine clades than in type strains that joined other clades.

Six former type strains, mostly isolated from alpechin, the wastewater of olive oil mills, were assigned to a clade distinct, but related to those of wine yeasts (Figure 4.1, Table 4.1, Table A3.1). These strains are *S. cerevisiae* x *S. paradoxus* hybrids (Pontes et al., 2019) and their ecological niche is related to processed olives, including table olives, olive oil and wastewater from olive oil mills. The *S. cerevisiae* sub-genomes of these strains (the *S. paradoxus* sub-genome was excluded from this analysis) cluster close to the main wine clade (Figure 4.1), which indicates that the *S. cerevisiae* ancestor of these hybrids was a wine strain. The species to which these strains were associated with were described between 1957 and 1978 (Table 4.1) and all had the peculiarity of being unable to assimilate sucrose, an uncommon feature for *S. cerevisiae*. We confirmed that in all strains of the Olives clade, *SUC2* has a premature stop codon, thus explaining the observed phenotype. Another peculiar trait was the capacity to assimilate melibiose, again an uncommon trait for *S. cerevisiae*. Traditionally, melibiose utilization was regarded as a distinctive marker for the discrimination of Mel- *S. cerevisiae* wine and top-brewing yeasts from Mel+ *S. pastorianus*, the bottom-brewing yeasts.

4.3.3. The beer and bread clades contain multiple type strains of former species

S. cerevisiae ale beer yeasts are assigned to two main clades, Clade 11 (Beer 1) that includes most beer types (British, German, wheat beer and Belgian beers) and Clade 5 (Beer 2) that includes low-gravity Saison-type beers (Figure 4.1). Whereas the main beer group (Beer 1) included only one type strain, in fact the valid type strain of *S. cerevisiae*, which was isolated from a brewery in the Netherlands, the Beer 2 clade included six type strains of former species. Not surprisingly, most type strains in this clade were isolated from beer and the most notable one being the type strain of *S. diastaticus*. This species was described more than 60 years ago as an unusual *S. cerevisiae*-like yeast occurring preferentially as a contaminant in beer fermentations (Andrews et al., 1952; van der Walt, 1965). The diastase (starch-degrading)

ability of *S. diastaticus* is encoded by the chimeric gene *STA1* that gives rise to an extra-cellular glucoamylase that allows the conversion of soluble starch and dextrin into fermentable sugars. This leads to an abnormal and undesirable attenuation of beer, corresponding to a specific gravity much lower than what is typical for most beers (Gilliland, 1952). Our assignment of the type strain of *S. diastaticus* to the Beer 2 clade is in line with a recent report indicating the association of diastase-positive strains with this clade (Krogerus et al., 2019). Adding to the elucidation of the contribution of *STA1* and its polymorphisms to desirable and undesirable brewing properties and its association mostly with the Beer 2 clade (Krogerus et al., 2019), we provide here the phylogenetic placement of seven additional beer-spoiling strains. The phenotypes of these strains were recently characterized (Meier-Dörnberg et al., 2018), and here they were all assigned to the Beer 2 clade (Figure 4.1). Therefore, this clade, that we rename as "Beer 2-Diastaticus" (Table 4.2), combines not only strains adequate for beer production, but also an important group of beer-spoilage yeasts. Most notably, these two technologically distinct yeast types are not easy to differentiate at the genome level since brewing strains and spoilage strains appear intermingled in a detailed view of the Beer 2-Diastaticus clade (Figure 4.2). Interestingly, several of the members of this clade had an excess of heterozygous sites which are likely an indication of ploidy levels higher than 2n (Figure 4.2), as already shown for Beer 1 strains that are tetraploid (Gallone et al., 2016; Gonçalves et al., 2016). However, contrary to Beer 1 and bread yeasts, which invariably have a ploidy higher than 2n, the Beer 2-Diastaticus clade appears to contain both 2n and > 2n strains, which is an uncommon situation among the different populations of *S. cerevisiae* known so far. Our results are in line with previous findings on the strongly variable nature of the diastase activity (Krogerus et al., 2019). Moreover, they support the division of the Beer 2-Diastaticus clade into two subclades, one where the presence of *STA1* could not be detected (subclade *STA1*-negative) and another one positive for *STA1*. The *STA1*-negative subclade congregates most of the former type strains found in this group, some of which are associated with brewing or with beer deterioration, but not with diastase activity (Figure 4.2). The *STA1*-positive subclade harbors the type strain of *S. diastaticus* and all diastase-positive spoilage strains, together with several brewing strains. A 1162-bp deletion in the promoter of *STA1* has been shown to explain a weak diastase activity in *STA1*-positive strains (Krogerus et al., 2019). This could explain the occurrence in the same lineage of spoiling and beer-production strains. Our survey of this deletion among the genomes of the *STA1*-positive subclade revealed that all the five strains that exhibited a diastase activity in vitro (Figure A3.2) had the promoter region intact, whereas the diastase-negative strains for which *STA1* was detected (three strains) had the deletion (Figure 4.2).

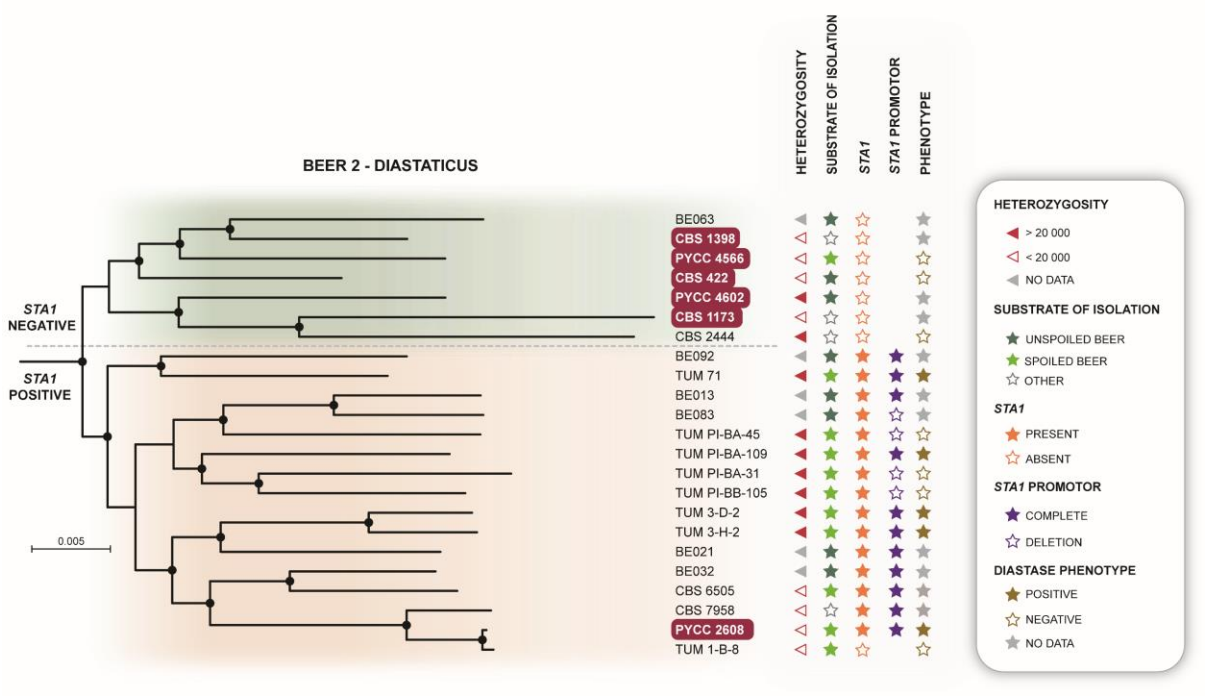


Figure 4.2 - The Beer 2 - Diastaticus clade of *S. cerevisiae* contains *SAT1*-positive and *STA1*-negative strains. Detail of the main phylogeny showing the phylogenetic relationships within the Beer 2 - Diastaticus clade together with relevant genetic and phylogenetic attributes.

The Bread clade contains strains used for bread leavening and is closely related to the beer yeasts of the Beer1 clade. It includes two former type strains that are clinical isolates, *S. cerevisiae* var. *onychophilus* and *S. veronae* var. *osloensis*. This suggests that besides de Wine clade, the Bread clade also hosts opportunistic strains.

4.3.4. Additional clades harboring type strains of former species

The remaining six former type strains are placed in three additional clusters, all associated with human-made fermentations (Figure 4.1). The type strain of *S. mangini*, found in cheese, was associated with Dairy clade (Clade 9). This clade was recently described (Legras et al., 2018) and gathers strains exclusively isolated from dairy products. Three other former type strains belonged to the Sake clade (Clade 15) that gathers strains used to ferment sake in Japan and other cereal-based beverages in other parts of Asia such as Laos, Philippines, and Tibet. Not surprisingly, the former type strains of the Sake clade were isolated in China, Japan and Taiwan from sake and other fermentations. Finally, two reference strains clustered in the West Africa clade (Clade 19), and both were isolated from African artisanal fermented beverages, one from ginger beer (*S. lindneri*) and the other from palm wine (*S. chevalieri*). For

this last species we sequenced strain PYCC 8489, obtained from the NRRL collection (NRRL Y-12633) in 2009. We found its genome sequence to be markedly distinct from that of supposed copy of NRRL Y-12633, strain Y12 (Bergström et al., 2014; Fay & Benavides, 2005). Whereas PYCC 8489 clustered in the West African clade, Y12 belonged to the Sake clade (Figure 4.1). Given the isolation of *S. chevalieri* from an African fermented beverage, we believe that the sequence of PYCC 8489 represents the correct placement of *S. chevalieri*, whereas Y12 probably does not represent the original African isolate.

4.3.5. The biogeography and ecology of wild populations

Although the recognition and study of wild populations of *S. cerevisiae* is of major importance for the detailed understanding of the emergence of domesticates, our knowledge of wild lineages and of their natural biology is much less advanced than that of domesticated populations and their features. In our analysis, we identified 13 wild clades that therefore represent almost half of the total of 27 clades recognized (Figure 4.1). Our main criteria for identifying a wild clade are the ecological association with a natural niche, sometimes coupled with a well-defined geographical distribution, and the absence of domestication signatures (see below). In two cases, we characterized a clade simultaneously as wild and domesticated. Clade 7 combines wild strains from Brazil, aboriginal human-associated strains from French Guiana and Mexican agave fermentation strains. Clade 16 combines strains directly isolated from fruits or tree sap and palm wine strains. In one situation (Clade 14) we characterized a clade as feral, since while its members originate in arboreal wild environments in China, they possess several typical domestication signatures (Table 4.1). The wild clades did not include any former type strains, thus emphasizing their recent discovery.

Some wild lineages considered here result from the combination of previously recognized lineages. The most striking example is Clade 17, which combines the North America-Japan group (Almeida et al., 2015), the Far East Russia group (Peter et al., 2018) and the China VI-VII groups (Duan et al., 2018). In this case, a wild population predominantly associated with arboreal niche spans from China to Far East Russia, Japan, and North America. Other cases involve the already mentioned Clade 7 in Central and South America and Clade 20 in China, Ecuador, and Brazil.

Figure 4.3 summarizes our current understanding of the biogeography and ecology of wild populations of *S. cerevisiae*. As already discussed (Wang et al., 2012), most wild lineages are Chinese or have been originally detected in China. Seven of the 13 wild lineages are exclusively composed of Chinese strains and three additional lineages contain strains from China and from other regions (Figure 4.3). Moreover, most wild lineages (10 lineages) are associated with arboreal niches, like tree bark, decaying wood and soil underneath trees. In

such habitats simple sugars are scarce and therefore the "make-accumulate-consume" (ethanol) ecological strategy of *S. cerevisiae* (Piškur et al., 2006), may not take place. On the contrary, the Malaysian population (Clade 22), for which a very limited number of clonally related isolates is known, appears to fulfill the anticipated ecological profile of a wild *S. cerevisiae* population, since it thrives in a sugar-rich environment where ethanol accumulates (Wiens et al., 2008). Additional wild populations from sugar-rich environments are scarce. The Philippines clade (Clade 16) is a candidate wild population since it contains strains isolated from sap or fruits of palm trees (Table A3.1). However, several strains in this clade came from different kinds of palm wine, which could suggest a domesticated origin.

The populations found in low-sugar niches appear to occur predominantly in temperate climates and the populations found in high-sugar niches seem to be associated with tropical regions (Figure 4.3). In China, three wild populations from low-sugar environments were found in the limit of the tropical zone and one of them (Clade 20) is also found in tropical South America (Brazil and Ecuador) (Figure 4.3). The other wild neotropical population (Clade 7) has a complex ecology and is associated with low and high sugar environments, and also with human gut (Figure 4.3).

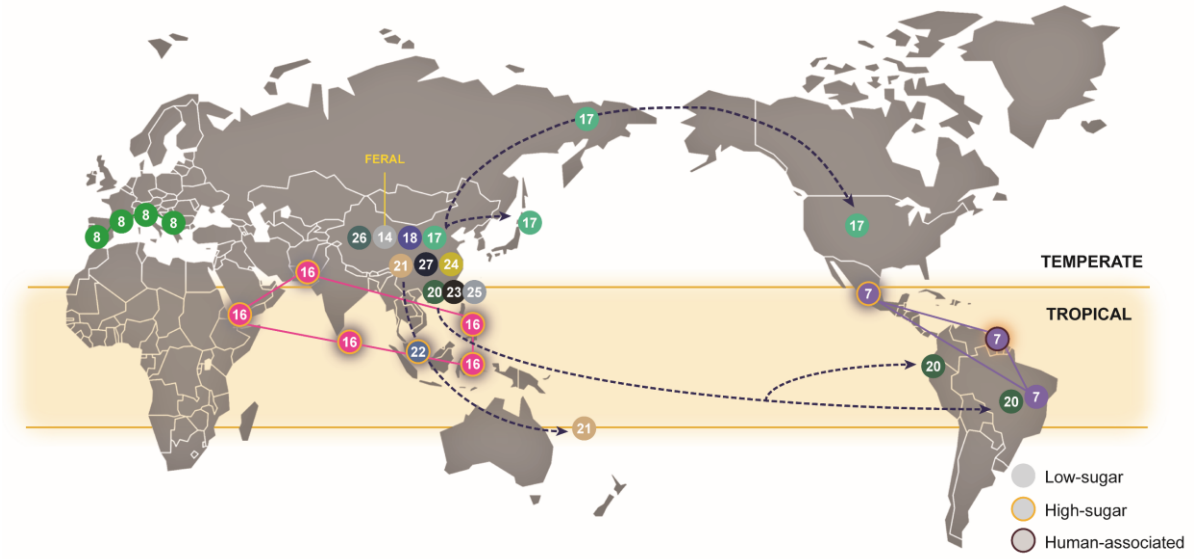


Figure 4.3 - The global biogeography and ecology of wild populations of *S. cerevisiae*. Populations are numbered and color-coded according to the clades depicted in Figure 4.1. Circles without outline color correspond to populations associated with low-sugar environments and circles outlined in orange correspond to populations associated with high-sugar environments (in the case of Clade 7, the violet outline color indicates association to human gut). Possible migration routes are indicated with arrowed dashed lines and population ranges are depicted with solid lines.

4.3.6. Lifestyle drives fate of aquaporins

The loss of function of aquaporins, which are membrane water channels that facilitate the transport of water in and out of the cell, is viewed as a consequence of domestication. Functional aquaporins, encoded by the genes *AQY1* and *AQY2*, decrease fitness in environments with a high osmolarity caused by an elevated concentration of sugars, like wine must or beer wort, although they contribute to the fitness of wild strains (Will et al., 2010). We observed that all the wild populations associated with low-sugar arboreal niche (Figure 4.3) have functional aquaporins (Table A3.1), while in the Wine, Beer 1, Sake and other domesticated clades the inactivation of aquaporins is the norm (Barbosa et al., 2018; Gonçalves et al., 2016) (Table A3.1). As discussed elsewhere, the inactivating mutations of *AQY1* and *AQY2* are distinct across the various populations (Gonçalves et al., 2016), which suggests that the loss of function occurred independently and multiple times. The Wine and related secondarily domesticated populations like the Cachaça-Bioethanol population, and also the Dairy population, share the same type of aquaporin gene inactivation, while Sake and Chinese Fermentations populations exhibit other types of inactivating mutations (Table A3.1). Moreover, we also observed the inactivation of aquaporins in wild populations associated with high-sugar environments, namely the Malaysian and Philippines populations (Table A3.1). Interestingly in Clade 7, found in the neotropics, the subclade associated with the arboreal niche in Brazil has functional aquaporins whereas the subclade associated with the human gut (French Guiana) and agave fermentation (Mexico) have non-functional aquaporins, whose genes exhibit unique mutations within each subclade (Table A3.1). Whereas the Malaysian and neotropical populations have distinctive inactivating mutations, not found in any other populations, the inactivating mutations in the Philippines population (Clade 16) are the same as those found in Sake (Clade 15) and Chinese Fermentations (Clade 12), which could indicate a wild-domesticate relationship in the Asian region. Taking into consideration this broader population-level analysis, the argument that aquaporin genes are maintained in two states by balancing selection (Will et al., 2010) gets additional support. Moreover, it may be hypothesized that in *S. cerevisiae* aquaporin gene inactivation occurs whenever a transition to a high-sugar niche occurs, thus including the occurrence of inactivation in wild populations. Therefore, this loss of function may predate human domestication and may be a general and natural response of *S. cerevisiae* to nutritionally rich environments that also occurs, independently and with distinct mutations, in different domesticated lineages.

4.3.7. The natural reservoirs of *RTM1* and *BIO1/BIO6*

RTM1 is a subtelomeric gene associated with the locus of sucrose utilization and provides resistance to inhibitory compounds present in molasses (Ness & Aigle, 1995). This gene tends to be present in domesticated populations that are grown in molasses or equivalent substrates like beer wort, sugar cane juice or dough, but not in wine yeasts (Borneman & Pretorius, 2015). *RTM1* can be viewed as a domestication signature of several clades (Figure 4.1), and twelve type strains contained this gene, including the valid type strain of *S. cerevisiae* (Beer 1 clade), five former type strains in the Beer 2-Diastaticus clade, and another six type strains distributed in the Bread, Sake and West Africa clades. The population analysis of *RTM1* occurrence shows that it is also present in wild strains, like those of the Philippines (Clade 16), China X (Clade 21) and Malaysia (Clade 22), thus suggesting that domesticated strains may have acquired this gene from natural reservoirs in wild populations. A distinct *RTM1* allele was found in strains of the African beer population (Clade 6) and in strains used to ferment Mexican agave (Clade 7). Given the absence of *RTM1* homologs in *S. paradoxus*, it is difficult to access at this stage if these divergent sequences are intraspecific *S. cerevisiae* alleles or if they have a foreign origin.

BIO1 and *BIO6* encode enzymes involved in the synthesis of biotin and were considered to have a restricted distribution in *S. cerevisiae*, being present in strains used in sake fermentation, but absent in wine strains (Borneman & Pretorius, 2015; Gonçalves et al., 2016; Hall & Dietrich, 2007). More recently, *BIO1/BIO6* were found to be present in cachaça strains and in wild Brazilian strains (Barbosa et al., 2018). Here, we detected *BIO1/BIO6* in 15 of the 27 clades depicted in Figure 4.1. North American *S. paradoxus* alleles of *BIO1/BIO6* appear to have introgressed into American populations. Earlier we reported such alleles in wild Brazilian populations (Clade 7 and Clade 20) and in the Cachaça-Bioethanol group (Clade 4) and also supposedly European brewing strains of the Beer 2- Diastaticus clade (Legras et al., 2007). Here we confirmed and expanded those observations and detected the same type of *S. paradoxus* alleles in additional representatives of the Beer2-Diastaticus clade and also in all representatives of Clade 7. As discussed above for *RTM1*, wild populations in Asia (China, Malaysia, Philippines) may have been the original reservoirs of BIO1/BIO6 that later disseminated into domesticated lineages propagated in substrates where biotin is scarce, like sake must.

4.3.8. The complex distribution of *MEL* alleles

The utilization of melibiose requires the expression of an α-galactosidase that hydrolyzes this disaccharide into glucose and galactose. This property is encoded by a set of polymorphic *MEL* genes (Naumov et al., 1990, 1991; H. Turakainen et al., 1993) that are present in *S. eubayanus*, *S. uvarum* and *S. mikatae*, but that are rare in *S. cerevisiae* and *S. paradoxus* (Naumov et al., 2002; Naumova et al., 2011). In a series of studies, Naumov and coworkers characterized by genetic mapping 15 genes (*MEL1* to *MEL15*) in *S. cerevisiae* (Naumov et al., 2002; Naumova et al., 2003), that were found to be distributed in 11 chromosomes in telomeric regions. Together, these studies revealed that wine strains rarely contained this gene, but that strains related to the niche of processed olives could have multiple, highly similar, *MEL* genes. Another study reported that the Malaysian population (Clade 22) had a good ability to utilize melibiose (Warringer et al., 2011). This study also revealed a case of melibiose utilization associated with the West African population (Clade 19), but with a distinct evolutionary history since the West African *MEL* gene appeared to represent an introgression from *S. paradoxus*.

Our analysis indicated that, although infrequent within the phylogenetic spectrum of *S. cerevisiae*, *MEL* occurrence is wider, and its divergence is more complex, than previously anticipated. Of the 27 clades of *S. cerevisiae* depicted in Figure 4.1, 16 contained *MEL* alleles. However, the frequency of MEL+ genomes in those 16 clades was in most cases lower than 50% and only in two cases (considering only clades with five or more sequences), the complete fixation of the *MEL* gene was observed (Olives and Wild Brazil 1-China V). Moreover, five of the 16 clades with *MEL* genes included former type strains. From the phylogenetic comparison of all *MEL* sequences detected, we inferred that the native *S. cerevisiae* sequences could be resolved into three alleles, with four additional alleles of uncertain origin (Figure 4.4A). One allele, "Asian I" (*MEL1*), was present in genomes assigned to six clades that have in common an Asian origin. This allele was present in several strains of Chinese Fermentations clade (Clade 12) and of the Sake clade (Clade 15), including the type strain of *S. carlsbergensis* var. *mandschuricus* that belongs to this last clade (Figure 4.4A).

This allele was also present in wild populations: Wild-Brazil1-ChinaV (Clade 20), China III (Clade 23) and China I (Clade 25). Another allele, "Asian II", was present in China X (Clade 21). A third allele, "Asian-American", was found in an Asian strain isolated in Sri Lanka and belonging to the Philippines clade (Clade 16) and in two genomes of the Beer 2-Diastaticus clade (Clade 5). These two Asian-American alleles were found in the type strain of *S. brasiliensis*, a Brazilian brewing strain and in a North American distillery strain. Recombination between Asian II and Asian-American alleles was detected in strain BT3, a Chinese domesticated strain from Clade 12. A similar recombinant allele corresponds to *MEL2*

found in a strain isolated from dewberries in Russia (Naumova et al., 2003). A fourth *MEL* allele, "European", was detected in two wine strains of Clade 1 (Wine) and in all strains of Clade 2 (Olives). In line with the information available in the literature, we confirmed that in wine strains (clades 1 and 3) the occurrence if this gene is rare, as we did not detect additional *MEL* genes in wine strains. The two genomes that had the *MEL* genes were that of CBS 5635, the type strain of *S. coreanus* and the genome of the commercial wine strain AWRI 796, but in this case the gene had inactivating mutations. Curiously, both strains were isolated in South Africa and another South African strain, from millet beer (Clade 6), also had this allele. A survey in our database of 170 genomes of *S. cerevisiae* wine strains revealed that additional genomes contained *MEL* genes, but only in two cases the genes were functional. In contrast, all the strains of the Olives clades, i.e., six former type strains, were MEL+. Even in a larger strain dataset of more than 20 genomes of this clade related to a previous study (Pontes et al., 2019), we could detect the presence of *MEL* in all strains, which makes the contrast with *MEL* scarcity in the Wine clades even more striking. This allele corresponded to several *MEL* genes described previously, from *MEL3* to *MEL11* (Naumova et al., 2003), which appear to represent a recent event of gene expansion in *S. cerevisiae* associated with domestication. All strains of the Olives clade had four or more *MEL* copies, with the highest copy numbers being recorded for *S. hienipiensis* (10) and *S. norbensis* (11). Interestingly, the European allele was also found in three strains of *S. paradoxus* (Figure 4.4A). These strains were isolated in Greece and belong to the European population of that species. The *S. cerevisiae* and *S. paradoxus* sequences of the European allele are almost identical, which suggests a case of introgression between these two species, but whose donor and recipient cannot be unequivocally identified at present.

A fifth allele (African I), is found mostly in *S. cerevisiae* strains isolated in Africa (Clade 6, African beer) and corresponds to *MEL12* to *MEL15* (Naumov et al., 1990). A similar allele (African II) is also present in African strains, mostly of Clade 19 (West Africa), but is a pseudogene, as previously reported (Warringer et al., 2011). The type strains of *S. chevalieri* and *S. lindneri* contains this allele. The last allele (South American) is present in wild Brazilian strains (clades 7 and 20) and in Brazilian domesticated strains used for sugar cane (cachaça) fermentations (Clade 4). Interestingly, Clade 20 that combines wild Chinese and wild Brazilian and Ecuadorean strains, contains the Asian I allele, present in all strains and the South American allele that co-occurs with the Asian I allele in Brazilian strains.

The African and South American alleles are not only considerably divergent from the Asian and European ones, but also share a phylogenetic proximity with *S. paradoxus MEL* genes found in representatives of the North American and Hawaiian population of this species (Figure 4.4A). Again, an introgression involving *S. cerevisiae* and *S. paradoxus* can be hypothesized, but the donor and the recipient species cannot be identified at present. The frequency of *MEL* genes in *S. paradoxus* appears to be low. Among the European strains, we

surveyed 40 genomes and only in three cases did we found these genes. For the other populations we detected two positive cases among 30 surveyed genomes (Table A3.2).

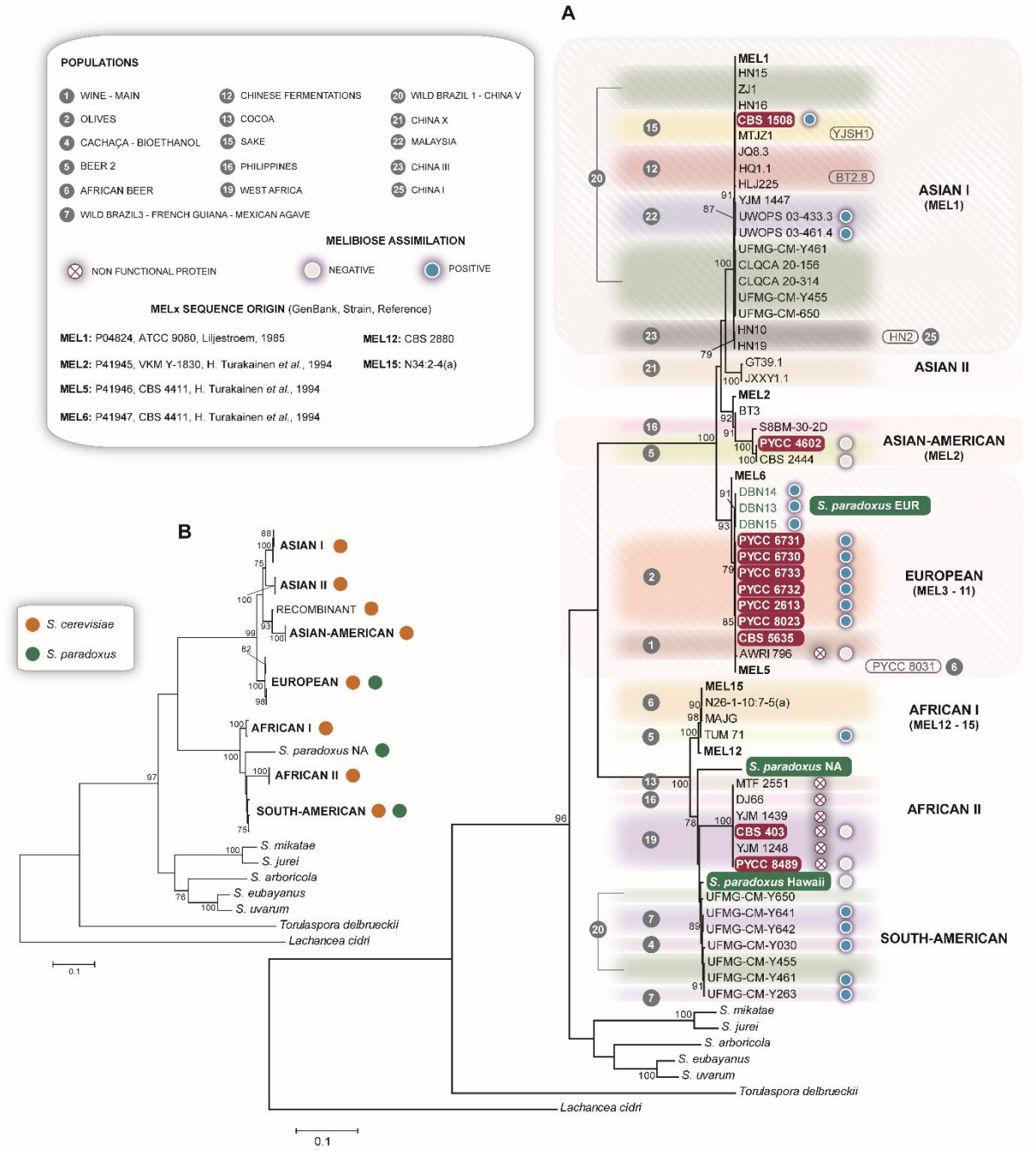


Figure 4.4 - Phylogenetic analysis of *MEL* alleles of *S. cerevisiae*. (A) Phylogenetic overview of *MEL* alleles from 16 populations of *S. cerevisiae*. For comparison, *MEL* sequences from other *Saccharomyces* species were included together with *S. cerevisiae MEL* alleles (*MEL1*, *MEL2*, *MEL5*, *MEL6*, *MEL12* and *MEL15*) retrieved from the literature and from the NCBI database (*MEL12* and *MEL15* were retrieved directly from genomic assemblies). For *S. paradoxus*, sequences from representatives of the European (EUR), North American (NA) and Hawaiian populations were added. Incomplete sequences that could be assigned to an allelic group, but that were not used in the phylogeny, are indicated next to the tree in a rounded rectangle. The former type strains of *S. cerevisiae* are indicated in red rectangles; (B) simplified phylogeny of *MEL* alleles in the genus *Saccharomyces*. The two phylogenetic trees were constructed using maximum likelihood method and the Tamura 3-parameter model. Bootstrap values > 75% are indicated (1000 replicates).

We tested strains harboring distinct alleles for the ability to grow on melibiose and observed some cases of absence of growth on this compound (Figure 4.4 and Figure A3.3). As expected, the inactivating mutations in the European allele of wine strains and in African II allele resulted in MEL- phenotype. Moreover, the strains with the Asian-American alleles and the North American and Hawaiian strains of *S. paradoxus* were also incapable of growing on melibiose in spite of having apparently functional genes.

A simplified phylogeny of the *MEL* gene for the genus *Saccharomyces* is presented in Figure 4.4B. Given the suspected introgressions to or from *S. paradoxus*, the boundaries between *S. cerevisiae* and *S. paradoxus* are not easily discernible. Since the average genome divergence between *S. cerevisiae* and *S. paradoxus* is 10%, the approximate 10% pairwise sequence divergence observed between the Asian I, II and Asian-American and the European allele could represent the above-mentioned species-level divergence (Figure 4.4b). In this scenario, Asian I, II and Asian-American would represent the *S. cerevisiae* native allele and European would represent the *S. paradoxus* allele. Consequently, the European alleles found in *S. cerevisiae* wine and olives strains would represent an introgression from *S. paradoxus*. However, the origin of the remaining African and South African alleles, which are 20% divergent from the Asian I allele, is puzzling and can only be explained by an introgression from an unknown *Saccharomyces* species that is phylogenetically more related to *S. cerevisiae* and *S. paradoxus* than any currently known *Saccharomyces* species. An alternative explanation is to view the Asian I, II, Asian-American and the European alleles as native to *S. cerevisiae*, in spite of their remarkable divergence, and the African-South American ones as native to *S. paradoxus*. In this scenario, *MEL* divergence between *S. cerevisiae* and *S. paradoxus* is approximately double the average genomic divergence. This hypothesis also implies that the European allele was introgressed from *S. cerevisiae* into *S. paradoxus* and the African-South American alleles were introgressed from *S. paradoxus* into *S. cerevisiae*.

4.4. **Discussion**

In this study, we attempted to compare the diversity of a group of 45 type strains of species now viewed as *S. cerevisiae* synonyms, while simultaneously reassessing the population landscape of the species at a global scale. Under the lens of genomics, this group of former type strains represents a valuable historical record of diversity, ecology and biogeography that spans from 1870 (*S. ellipsoideus*) to 1984 (*S. boulardii*). When this diversity is superimposed on the global population diversity of *S. cerevisiae*, former type strains map reasonably well on the known domesticated populations but are absent in wild populations. This is an eloquent illustration of the roots of our biased understanding of *S. cerevisiae*

primarily as a utility in wine, beer and bread production (among much other fermented products), rather than as a microbe shaped by its natural history. Another interesting observation is the strong redundancy of old species delimitations, based mostly on strains that belong to the Wine and Olives populations. A minor redundancy is also present in the Beer2-Diastaticus, Bread, West Africa and Sake clades, but not in the Dairy and Beer 1 population. Our analysis also clarifies the status of historically and technologically relevant *Saccharomyces* species names for users in the fermentation industry.

Various population genomics studies of *S. cerevisiae* have been published in recent years and consequently distinct designations have been used to name the same population. Such lack of consistency, together with incomplete sampling, strong redundancy of certain domesticated lineages or the improper use of the "wild" descriptor, call for the need of an informed debate. Population VIII from China (Clade 14) provides a striking example of difficulties in recognizing a wild lineage. Undisputedly considered as "wild", it was even used as evidence supporting the East Asian origin of all domesticated *S. cerevisiae* since it carries duplicated genes involved in maltose metabolism, thus appearing as a direct link between wild and domesticated lineages (Duan et al., 2018; Fay et al., 2019). However, *MAL* expansion can be viewed as a domestication footprint and in fact this population has additional signatures of domestication like a type of inactivation mutation of *AQY1* and the presence of region B (Table A3.1) that are typical of wine strains. Consequently, China VIII must be regarded as a feral population, not a wild one.

A recent study included a comprehensive genomic survey of *S. cerevisiae* that involved 362 strains from the Wine clade (Peter et al., 2018). Following earlier publications (Liti et al., 2009), this clade was referred to as "Wine/European" in spite of the well-recognized global distribution of wine strains, not only from Europe, but also from North and South America, South Africa and Oceania, the enrichment in European strains in this clade being simply the likely consequence of an historical sampling bias in Europe. Moreover, the truly European population corresponds to a wild South European population, closely related, but phylogenetically distinct from the wine yeasts (Sampaio & Gonçalves, 2017). This is so far the only population to which the epithet "European" can be appropriately applied since all members of this population were found exclusively in Europe. In addition, the claim that wild strains represent 16% of the isolates of Wine clade (Peter et al., 2018), needs to be verified since the finding of a strain exhibiting key genomic features of Wine clade, but isolated from soil or another non-wine substrate, does not warrant the classification as wild, as feral may be more adequate descriptor, this avoiding confusion with truly wild strains.

Recently, the genomes of 266 wild and domesticated Asian isolates were sequenced and it was proposed that China and Far East Asia are the center of domestication of *S. cerevisiae* (Duan et al., 2018). In another recent study, 1011 genomes were analyzed and although the "out-of-China" model for wild populations of *S. cerevisiae* was supported, it was suggested

that several independent domestication events explain the emergence of domesticated populations (Peter et al., 2018). Here, we combined key representative sequences of both studies and although most of the phylogenetic structure of *S. cerevisiae* from previous studies could be recovered, a supposedly distinctive position of Asian domesticated lineages, compatible with an initial domestication in China from which all other domestications derive, could not be confirmed. Instead, most Chinese domesticated strains appear closely related and were gathered in a new clade (Clade 12, Chinese fermentations) that is nevertheless distinct from the Sake clade (Clade 15), a key group in most population studies, but that was not adequately represented in the study of Asian strains.

By combining genome sequences obtained in recent studies by other authors and those sequenced by us, we could observe that some Chinese wild populations, which globally represent the most diverse assemblage of populations in the species, have members in other regions, including other continents. This illustrates not only the broad range of the distribution of some wild populations, but also their ability to migrate over thousands of kilometers, for example, across the Indian or Pacific oceans. The mechanism and consequences of such long range dispersal still awaits a formal study. Another relevant aspect is the coexistence in sympatry of distinct wild lineages without apparent relevant genetic contact. Again, the drivers for such long-term sympatric coexistence of different genetic lineages remain unknown.

In our current understanding of *S. cerevisiae* populations is taken into consideration, a wide diversity of lifestyles appears as the hallmark of this species. Such diversity is still remarkable when only wild populations are considered since arboreal (low-sugar), and sap-fruit-nectar (high-sugar) lifestyles are known. This has not been observed on *S. cerevisiae* closest relative, *S. paradoxus*, in spite their close genetic resemblance and even sympatry of these two species in certain temperate regions (Robinson et al., 2016; Sampaio & Gonçalves, 2008). We speculate that the better adaptation of *S. cerevisiae* to higher temperatures, reflected in its biogeography (Robinson et al., 2016), allowed the colonization of tropical regions and the consequent transition to high-sugar niches. In several ways, the transitions promoted by humans during the fermentation of beverages or foodstuffs, mimic the transitions in nature from low-sugar to high-sugar niches. Independent inactivation of aquaporin genes in the wild and in the winery are an indication of such plasticity.

Recently, we proposed a domestication model with transitions from a primarily domesticated state, exemplified by wine yeasts, to a secondarily domesticated state, exemplified by cachaça-bioethanol yeasts (Barbosa et al., 2018). The regions A, B and C, likely acquired first by wine strains (Novo et al., 2009), are also present (especially region B) in secondarily domesticated lineages that derive from wine yeasts like Bread, Beer2-Diastaticus and Cachaça-Bioethanol (Figure 4.1). In sharp contrast, these regions are absent in wild lineages.

Contrary to regions A, B and C that were acquired solely during domestication, other traits like aquaporin inactivation, the presence of *RTM1* cluster o the presence of the *BIO1*/*BIO6* genes appear to have predated domestication. We observed that such traits already occur in wild populations, especially in Asian ones. Therefore, the presence of the *RTM1* cluster or the *BIO1*/*BIO6* genes in certain domesticated populations may originate in the wild genetic stock of the species. Likewise, the propensity for aquaporin inactivation observed in domesticated populations may recapitulate a natural phenomenon that occurs in wild populations. For *MEL* alleles and in spite of their unclear evolutionary history, which also involves *S. paradoxus*, their origin in *S. cerevisiae* is again centered in Asia and in wild populations.

5.
Concluding remarks and future perspectives

Our understanding of microbial domestication in general, and of *S. cerevisiae* domestication in particular, has increased rapidly in the last decades. This is due firstly to the dissemination of whole-genome studies that allowed for a variety of analyses not available before. Nevertheless, uncovering past events in the evolutionary history of microbes during their association with humans is hampered by several factors. First, the lack of awareness, let alone knowledge, on microorganisms and microbiology prevented an accurate historical record. Secondly, given that microbial ecology is still understudied when compared to other of biology, the natural biology of many microorganisms, including *S. cerevisiae*, remains obscure. Finally, the history of domestication of *S. cerevisiae* is a global and ancient event that includes, but is not exclusive to, wine, beer, and bread. It is relevant to note that the association *S. cerevisiae* - humans is as ancient as human civilization since the period of early human exploitation of microbes is likely to be contemporary with that of plant and animal domestication and consequently on the availability of seasonal crop surpluses or animal products such as milk. However, the widespread sequencing effort of representatives of wild and domesticated lineages, allied with the depth and speed at which hypotheses can be tested in the laboratory, have contributed to illuminate several important aspects. Furthermore, the gradual recognition of the importance of wild environments and of natural ecology has been a decisive step forward. As such, a more balanced view has emerged in which the routes of domestication are viewed within an evolutionary and ecological context.

During this work, we have attempted to deepen the knowledge on the domestication trajectories of *S. cerevisiae*. One of the main conclusions is that domestication is not a linear and straightforward process as initially anticipated. In chapter 2, the results of the study of a new population associated with cachaça, allowed us to develop a new model of domestication

in *S. cerevisiae*. This model accommodates cases of transition from wild to primary domesticate and cases of transition of primary domesticate to secondary domesticate. This differs substantially from the traditional view that focused on a more linear transition from wild to domesticate (Figure 5.1). The model presented here views microbe domestication as a multilayered phenomenon, in which transitions might have different levels of complexity. More specifically, it accommodates multiple rounds of domestication and therefore secondarily domesticated populations. The concept of secondary domestication refers to certain lineages that, contrary to the primary domesticated, do not derive from wild populations, but instead from other domesticated populations. That is, a primarily domesticated population subjected to a new fermentative environment and novel selective pressures undergoes a second round of domestication, which leads to the conservation of some beneficial characteristics of the primary population, and the gain of new desirable traits. In addition to the Cachaça population described during this work, it was also possible to extend this classification of secondary domestication to the populations associated with bread leavening and to beer (Beer 2 population).

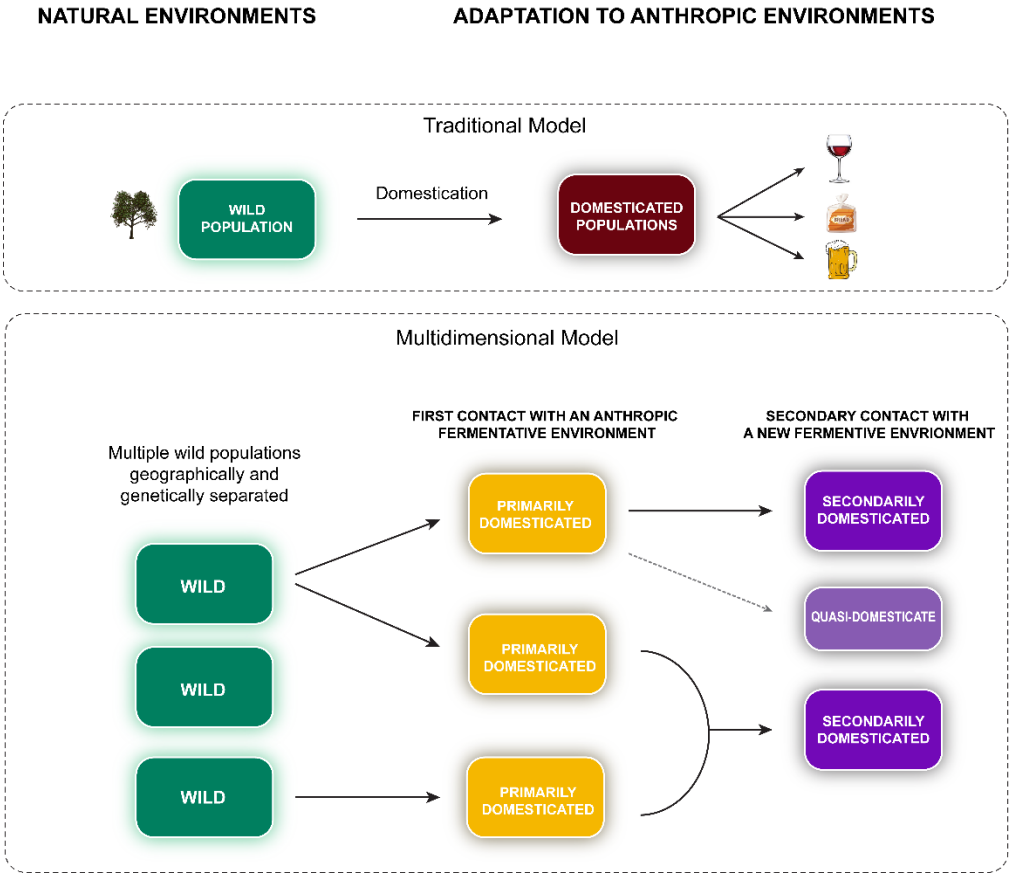


Figure 5.1 - Schematic representation of the two domestication models. The traditional model shows a simplistic and straightforward view, with domesticated populations originating from wild populations. The modern model represents a layered model maintaining the idea of domesticated populations originating from wild populations, but where new domesticated populations can arise from already established domesticated populations.

Another aspect of domestication emerged from the case presented in chapter 3 that refers to a very peculiar population found in products related to processed olives. This unique lineage resulted from a hybridization event between *S. cerevisiae* and *S. paradoxus*. We defined these "Olives" hybrids as relic hybrids, in order to distinguish them from typical hybrids and from a typically introgressed genome. Whereas a hybrid genome normally has an approximately similar contribution of both parental genomes with respect to gene content, introgression normally involves a minor contribution of a "foreign" genome (Harrison & Larson, 2014). In the case of relic hybrid genomes, we found a minor contribution of *S. paradoxus*, but dispersed across the entire *S. cerevisiae* backbone genome. This particular case of hybridization can be viewed from the perspective of our model of domestication as a secondary domesticate. In this case one of the parents is a *S. cerevisiae* wine strain, i.e., a primary domesticate, and the relic hybrid, the secondary domesticate, that shows an improved adaptation to the processed olives environment by comparison to the wine yeast progenitor. This case illustrates an event of secondary domestication in which hybridization drives the transition from primary to secondary domesticate.

Another important aspect concerns the role of the *S. cerevisiae* relic hybrid in the environments associated with processed olives, like table olives, in the context of domestication. Do the observed genomic and phenotypic changes fit in the definition of domestication used in this work? This definition was summarized in the General Introduction as follows: an organism bred in captivity and thereby modified from its wild ancestors in ways making it more useful to humans who control its reproduction and its food supply (Diamond, 2002). At this stage, it is not clear if the changes detected in these hybrids contributed to the aroma or flavor properties of table olives that are valued by humans. Given this situation we label this and similar lineages as quasi-domesticate, i.e., lineages in which domestication-like traits are detected but for which a clear link between the genetic and phenotypic changes and valued properties of the final product are not evident.

We argue that the study of *S. cerevisiae* domestication must rely on a solid knowledge of its natural biology, especially its natural ecology. Unfortunately, this knowledge is for the most part fragmentary (Liti, 2015; Sampaio & Gonçalves, 2017). Moreover, gaps in a detailed understanding of *S. cerevisiae* lifestyle in nature are frequently filled with assumptions not supported by direct evidence, being instead based in laboratory observations or on fermentations of artificial substrates like wine and beer, among others (Goddard et al., 2010; Stefanini et al., 2012). A strong bias is now evident in the type of isolation substrates from which most cultures of *S. cerevisiae* available for studying originate. This bias favors artificial environments and overlooks the natural ones. Historically, this bias has strongly influenced the way *S. cerevisiae* was perceived and has fostered several anthropocentric views that persist until today. Chapter 4, which was focused on a comprehensive view of the species, highlighted this isolation bias. Most of the currently accepted synonyms of *S. cerevisiae*, i.e.,

cultures that represent former type strains, belong to the Wine and Olives populations, both domesticated. Moreover, none of the former type strains clustered among wild populations. This study also highlighted problems on the nomenclature of industrially relevant strains. This is especially evident in the case of industrially important variants like *S. boulardii*, widely used as a probiotic (Pais et al., 2020), and *S. diastaticus*, an important brewing variant that can also be a serious beer contaminant (Meier-Dörnberg et al., 2018). These names still generate some confusion, especially among non-specialists, and we have attempted to clarify their status at the biological and nomenclature level.

In chapter 4, a comprehensive and balanced integration of all known populations of *S. cerevisiae*, including recently described Chinese populations (Duan et al., 2018; Wang et al., 2012), confirmed that the Chinese wild isolates represent the more divergent lineages found in *S. cerevisiae*, as already demonstrated in other studies (Duan et al., 2018; Peter et al., 2018). It also showed that wild isolates from other regions like North and South America and Japan, fit into some of the Chinese populations, thus deepening the global phylogeographic assessment of *S. cerevisiae*. Such analysis is also essential for a proper integration and interpretation of domestication. Recently, one hypothesis was put forward favoring a single domestication even in China (Duan et al., 2018). However, our results, together with those of other authors (e.g. Peter et al., 2018), do not support the hypothesis of a single domestication, in which a primal domesticated Chinese lineage then expanded geographically and acquired novel modifications that explain the present-day occurrence of *S. cerevisiae* in a wide diversity of fermentations. On the contrary, the data shown in chapter 4 suggests multiple independent domestication events from different wild populations, as is the case of the Wine population and its closest wild relative the Mediterranean Oaks population (Almeida et al., 2015), in the Mediterranean region. An independent domestication is also exemplified by the Sake population and a not yet fully characterized Asian wild population. Consequently, the dispersal model of wild populations from continental Asia to the rest of the world and the multiple domestication model need to be combined for a proper understanding of the biology and biogeography of *S. cerevisiae*. This is evidently a complex task that would benefit from additional isolation efforts in poorly sampled regions such as Africa, South America, and Southeast Asia.

The extensive analysis performed in chapter 4, also helped to highlight a few genomic particularities of some of the populations. One striking example is the case of the distribution of the *MEL* gene that codes for an alpha-galactosidase. It was already known from earlier studies that this gene was rarely found among *S. cerevisiae* and *S. paradoxus* strains (Naumov et al., 1995). By studying all known populations of *S. cerevisiae* to date, it was possible to observe that its presence is indeed rare, yet this gene is fixed in some populations. One of these populations is the Olives population, where all strains (i.e., those discussed in chapters 3 and 4) possess this gene. In this work we provided, for the first time, a population genomics

perspective of the distribution of *MEL* genes. As such, we could integrate previous dispersed records and confirm cases of *MEL* gene expansion (Naumov et al., 1991, 1995; Naumova et al., 2003; Hilkka Turakainen et al., 1994). Hybridization between *S. cerevisiae* and *S. paradoxus* seems to have been crucial for the adaptation to the niche of processed olives since, among *Saccharomyces* spp., only relic hybrids are isolated from these environments. Moreover, the presence of *MEL* genes in increased copy number appears to have contributed to adaptation. However, their precise role remains unclear, since melibiose or other substrates of alpha-galactosidase like raffinose or stachyose, are not detected in brine or in other olive-related products. Additional studies are therefore necessary to fully understand the role of *MEL* genes that might be more related to detoxification processes than to nutrient assimilation. Thus, the Olives population emerges as an relevant case-study that offers interesting opportunities for detailed studies on the mechanisms of genomic stabilization and reorganization in anthropic environments, and for the study of the evolution of specific genes in the context of their adaptive role.

In conclusion, the detailed study of domestication trajectories in *S. cerevisiae* carried out in this thesis revealed unanticipated levels of complexity, with cases of secondary domestication and of quasi-domestication. A better understanding of the mechanisms underpinning *S. cerevisiae* domestication will not only enable scientists to better understand the biology of this model organism from a fundamental perspective, but also from an applied one. The knowledge gained on the mechanisms that gave rise to domesticates and their resulting phenotypes can be employed for the rational improvement of industrial strains, a much-needed tool in many industries that rely on the products of yeast fermentation.

REFERENCES

[1] Abyzov, A., Urban, A. E., Snyder, M., & Gerstein, M. (2011). CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Research, 21(6), 974–984. https://doi.org/10.1101/gr.114876.110

[2] Akao, T., Yashiro, I., Hosoyama, A., Kitagaki, H., Horikawa, H., Watanabe, D., Akada, R., Ando, Y., Harashima, S., Inoue, T., Inoue, Y., Kajiwara, S., Kitamoto, K., Kitamoto, N., Kobayashi, O., Kuhara, S., Masubuchi, T., Mizoguchi, H., Nakao, Y., … Shimoi, H. (2011). Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. DNA Research, 18(6), 423–434. https://doi.org/10.1093/dnares/dsr029

[3] Albertin, W., Chasseriaud, L., Comte, G., Panfili, A., Delcamp, A., Salin, F., Marullo, P., & Bely, M. (2014). Winemaking and bioprocesses strongly shaped the genetic diversity of the ubiquitous yeast *Torulaspora delbrueckii*. PLoS ONE, *9*(4). https://doi.org/10.1371/journal.pone.0094246

[4] Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2017). Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites. Molecular Ecology, *26*(7), 2167–2182. https://doi.org/10.1111/mec.14071

[5] Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J. L., Serra, M., Dequin, S., Couloux, A., Guy, J., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. Molecular Ecology, *24*(21), 5412–5427. https://doi.org/10.1111/mec.13341

[6] Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., Albertin, W., Durrens, P., Sherman, D. J., Marullo, P., Todd Hittinger, C., Gonçalves, P., & Sampaio, J. P. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. Nature Communications, *5*. https://doi.org/10.1038/ncomms5044

[7] Alsammar, H., & Delneri, D. (2020). An update on the diversity, ecology and biogeography of the *Saccharomyces* genus. FEMS Yeast Research, 20(3), 1–12. https://doi.org/10.1093/femsyr/foaa013

[8] Alves-Araújo, C., Almeida, M. J., Sousa, M. J., & Leão, C. (2004). Freeze tolerance of the yeast *Torulaspora delbrueckii*: Cellular and biochemical basis. FEMS Microbiology Letters, 240(1), 7–14. https://doi.org/10.1016/j.femsle.2004.09.008

[9] Andrews, B. J., Gilliland, R. B., & Limit, A. (1952). Super-attenuation is in all cases due. Journal of the Institute of Brewing, 58.

[10] Angebault, C., Djossou, F., Abélanet, S., Permal, E., Soltana, M. Ben, Diancourt, L., Bouchier, C., Woerther, P. L., Catzeflis, F., Andremont, A., D'enfert, C., & Bougnoux, M. E. (2013). *Candida albicans* is not always the preferential yeast colonizing humans: A study inwayampi amerindians. Journal of Infectious Diseases, 208(10), 1705–1716. https://doi.org/10.1093/infdis/jit389

[11] Antunovics, Z., Nguyen, H. V., Gaillardin, C., & Sipiczki, M. (2005). Gradual genome stabilisation by progressive reduction of the *Saccharomyces uvarum* genome in an interspecific hybrid with *Saccharomyces cerevisiae*. FEMS Yeast Research, 5(12), 1141–1150. https://doi.org/10.1016/j.femsyr.2005.04.008

[12] Argueso, J. L., Carazzolle, M. F., Mieczkowski, P. A., Duarte, F. M., Netto, O. V. C., Missawa, S. K., Galzerani, F., Costa, G. G. L., Vidal, R. O., Noronha, M. F., Dominska, M., Andrietta, M. G. S., Andrietta, S. R., Cunha, A. F., Gomes, L. H., Tavares, F. C. A., Alcarde, A. R., Dietrich, F. S., McCusker, J. H., Petes, D. T, & Pereira, G. A. G. (2009). Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. Genome Research, 19(12), 2258–2270. https://doi.org/10.1101/gr.091777.109

[13] Arroyo-López, F. N., Querol, A., Bautista-Gallego, J., & Garrido-Fernández, A. (2008). Role of yeasts in table olive production. International Journal of Food Microbiology, 128(2), 189–196. https://doi.org/10.1016/j.ijfoodmicro.2008.08.018

[14] Avram, D., Leid, M., & Bakalinsky, A. T. (1999). Fzf1p of Saccharomyces cerevisiae is a positive regulator of *SSU1* transcription and its first zinc finger region is required for DNA binding. Yeast, 15(6), 473–480. https://doi.org/10.1002/(SICI)1097-0061(199904)15:6<473::AID-YEA388>3.0.CO;2-Y

[15] Badotti, F., Gomes, F. C. O., & Rosa, C. A. (2012). Brazilian cachaça: Fermentation and production. Handbook of Plant-Based Fermented Food and Beverage Technology, Second Edition, 639–648. https://doi.org/10.1201/b12055

[16] Badotti, F., Vilaça, S. T., Arias, A., Rosa, C. A., & Barrio, E. (2014). Two interbreeding populations of *Saccharomyces cerevisiae* strains coexist in cachaça fermentations from Brazil. FEMS Yeast Research, 14(2), 289–301. https://doi.org/10.1111/1567-1364.12108

[17] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

[18] Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J. B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of natural hybridization in Brazilian wild lineages of *Saccharomyces cerevisiae*. Genome Biology and Evolution, 8(2), 317–329. https://doi.org/10.1093/gbe/evv263

[19] Barbosa, R., Pontes, A., Santos, R. O., Montandon, G. G., De Ponzzes-Gomes, C. M., Morais, P. B., Gonçalves, P., Rosa, C. A., & Sampaio, J. D. S. P. (2018). Multiple Rounds of Artificial Selection Promote Microbe Secondary Domestication- The Case of Cachaça Yeasts. Genome Biology and Evolution, 10(8), 1939–1955. https://doi.org/10.1093/gbe/evy132

[20] Barnett, J. A. (2007). A history of research on yeasts 10: foundations of yeast genetics. Yeast, 24(10), 799–845. https://doi.org/10.1002/yea.1513

[21] Basso, R. F., Alcarde, A. R., & Portugal, C. B. (2016). Could non-*Saccharomyces* yeasts contribute on innovative brewing fermentations? Food Research International, 86, 112–120. https://doi.org/10.1016/j.foodres.2016.06.002

[22] Bendixsen, D. P., Gettle, N., Gilchrist, C., Zhang, Z., & Stelkens, R. (2021). Genomic Evidence of an Ancient East Asian Divergence Event in Wild *Saccharomyces cerevisiae*. Genome Biology and Evolution, 13(2), 1–12. https://doi.org/10.1093/gbe/evab001

[23] Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. Molecular Biology and Evolution, 31(4), 872–888. https://doi.org/10.1093/molbev/msu037

[24] Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

[25] Bonatsou, S., Paramithiotis, S., & Panagou, E. Z. (2018). Evolution of yeast consortia during the fermentation of Kalamata natural black olives upon two initial acidification treatments. Frontiers in Microbiology, *8*, 1–13. https://doi.org/10.3389/fmicb.2017.02673

[26] Borneman, A. R., Desany, B. A., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., Egholm, M., & Chambers, P. J. (2011). Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genetics, 7(2). https://doi.org/10.1371/journal.pgen.1001287

[27] Borneman, A. R., Desany, B. A., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., Egholm, M., & Chambers, P. J. (2012). The genome sequence of the wine yeast VIN7 reveals an allotriploid hybrid genome with *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii* origins. FEMS Yeast Research, 12(1), 88–96. https://doi.org/10.1111/j.1567-1364.2011.00773.x

[28] Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., & Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. G3: Genes, Genomes, Genetics, 6(4), 957–971. https://doi.org/10.1534/g3.115.025692

[29] Borneman, A. R., Forgan, A. H., Pretorius, I. S., & Chambers, P. J. (2008). Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. FEMS Yeast Research, 8(7), 1185–1195. https://doi.org/10.1111/j.1567-1364.2008.00434.x

[30] Borneman, A. R., & Pretorius, I. S. (2015). Genomic insights into the *Saccharomyces* sensu stricto complex. Genetics, 199(2), 281–291. https://doi.org/10.1534/genetics.114.173633

[31] Boynton, P. J., & Greig, D. (2014). The ecology and evolution of non-domesticated *Saccharomyces* species. Yeast, 31(October), 449–462. https://doi.org/10.1002/yea.3040

[32] Brown, C. A., Murray, A. W., & Verstrepen, K. J. (2010). Rapid Expansion and Functional Divergence of Subtelomeric Gene Families in Yeasts. Current Biology, 20(10), 895–903. https://doi.org/10.1016/j.cub.2010.04.027

[33] Cardoso, D. R., Andrade-Sobrinho, L. G., Leite-Neto, A. F., Reche, R. V., Isique, W. D., Ferreira, M. M. C., Lima-Neto, B. S., & Franco, D. W. (2004). Comparison between cachaça and rum using pattern recognition methods. Journal of Agricultural and Food Chemistry, 52(11), 3429–3433. https://doi.org/10.1021/jf035262+

[34] Carvalho, C., Tomás, A., Libkind, D., Imanishi, Y., & Sampaio, J. P. (2018). *Zygotorulaspora chibaensis* sp. nov. and *Zygotorulaspora danielsina* sp. nov., novel ascomycetous yeast species from tree bark and soil. International Journal of Systematic and Evolutionary Microbiology, 68(8), 2633–2637. https://doi.org/10.1099/ijsem.0.002889

[35] Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A. L., Ceppi, M., Conseiller, E., Debuchy, R., Malagnac, F., Goarin, A., Silar, P., Lacoste, S., Sallet, E., Bensimon, A., Giraud, T., & Brygoo, Y. (2014). Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nature Communications, 5. https://doi.org/10.1038/ncomms3876

[36] Christiaens, J. F., Franco, L. M., Cools, T. L., de Meester, L., Michiels, J., Wenseleers, T., Hassan, B. A., Yaksi, E., & Verstrepen, K. J. (2014). The fungal aroma gene *ATF1* promotes dispersal of yeast cells through insect vectors. Cell Reports, 9(2), 425–432. https://doi.org/10.1016/j.celrep.2014.09.009

[37] Coi, A. L., Bigey, F., Mallet, S., Marsit, S., Zara, G., Gladieux, P., Galeote, V., Budroni, M., Dequin, S., & Legras, J. L. (2017). Genomic signatures of adaptation to wine biological ageing conditions in biofilm-forming flor yeasts. Molecular Ecology, 26(7), 2150–2166. https://doi.org/10.1111/mec.14053

[38] Coutouné, N., Mulato, A. T. N., Riaño-Pachón, D. M., & Oliveira, J. V. de C. (2017). Draft Genome Sequence of *Saccharomyces cerevisiae* Barra Grande (BG-1), a Brazilian Industrial Bioethanol-Producing Strain. Genome Announcements, 5(13), 1–2. https://doi.org/10.1128/genomeA.00111-17

[39] Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., Huang, A. A., Dudley, A. M., & Fay, J. C. (2013). Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. G3: Genes, Genomes, Genetics, 3(12), 2163–2171. https://doi.org/10.1534/g3.113.007492

[40] Crosato, G., Nadai, C., Carlot, M., Garavaglia, J., Ziegler, D. R., Rossi, R. C., De Castilhos, J., Campanaro, S., Treu, L., Giacomini, A., & Corich, V. (2020). The impact of *CUP1* gene copy-number and XVI-VIII/XV-XVI translocations on copper and sulfite tolerance in vineyard *Saccharomyces cerevisiae* strain populations. FEMS Yeast Research, 20(4), 1–11. https://doi.org/10.1093/femsyr/foaa028

[41] Curtin, C. D., & Pretorius, I. S. (2014). Genomic insights into the evolution of industrial yeast species *Brettanomyces bruxellensis*. FEMS Yeast Research, 14(7), 997–1005. https://doi.org/10.1111/1567-1364.12198

[42] D'Angiolo, M., De Chiara, M., Yue, J. X., Irizar, A., Stenberg, S., Persson, K., Llored, A., Barré, B., Schacherer, J., Marangoni, R., Gilson, E., Warringer, J., & Liti, G. (2020). A yeast living ancestor reveals the origin of genomic introgressions. Nature, 587(7834), 420–425. https://doi.org/10.1038/s41586-020-2889-1

[43] De Chiara, M., Barré, B., Persson, K., Chioma, A. O., Irizar, A., Warringer, J., & Liti, G. (2020). Domestication reprogrammed the budding yeast life cycle. *BioRxiv*, 1–28. https://doi.org/10.1101/2020.02.08.939314

[44] Dhar, R., Sägesser, R., Weikert, C., Yuan, J., & Wagner, A. (2011). Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. Journal of Evolutionary Biology, 24(5), 1135–1153. https://doi.org/10.1111/j.1420-9101.2011.02249.x

[45] Diamond, J. (1997). Guns, germs, and steel: the fates of human societies. New York, NY: W.W. Norton & Company.

[46] Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. Nature, 418(6898), 700–707. https://doi.org/10.1038/nature01019

[47] Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The Molecular Genetics of Crop Domestication. Cell, 127(7), 1309–1321. https://doi.org/10.1016/j.cell.2006.12.006

[48] Doniger, S. W., Kim, H. S., Swain, D., Corcuera, D., Williams, M., Yang, S. P., & Fay, J. C. (2008). A catalog of neutral and deleterious polymorphism in yeast. PLoS Genetics, 4(8). https://doi.org/10.1371/journal.pgen.1000183

[49] Duan, S. F., Han, P. J., Wang, Q. M., Liu, W. Q., Shi, J. Y., Li, K., Zhang, X. L., & Bai, F. Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nature Communications, 9(1). https://doi.org/10.1038/s41467-018-05106-7

[50] Dujon, B. A., & Louis, E. J. (2017). Genome diversity and evolution in the budding yeasts (Saccharomycotina). Genetics, 206(2), 717–750. https://doi.org/10.1534/genetics.116.199216

[51] Dumas, E., Feurtey, A., Rodríguez de la Vega, R. C., Le Prieur, S., Snirc, A., Coton, M., Thierry, A., Coton, E., Le Piver, M., Roueyre, D., Ropars, J., Branca, A., & Giraud, T. (2020). Independent domestication events in the blue-cheese fungus *Penicillium roqueforti*. Molecular Ecology, 29(14), 2639–2660. https://doi.org/10.1111/mec.15359

[52] Dunn, B., Paulish, T., Stanbery, A., Piotrowski, J., Koniges, G., Kroll, E., Louis, E. J., Liti, G., Sherlock, G., & Rosenzweig, F. (2013). Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression. PLoS Genetics, 9(3). https://doi.org/10.1371/journal.pgen.1003366

[53] Dunn, B., Richter, C., Kvitek, D. J., Pugh, T., & 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 Research, 22(5), 908–924. https://doi.org/10.1101/gr.130310.111

[54] Dupont, J., Dequin, S., Giraud, T., Le Tacon, F., Marsit, S., Ropars, J., Richard, F., & Selosse, M. A. (2017). Fungi as a source of food. The Fungal Kingdom, 8, 1063–1085. https://doi.org/10.1128/9781555819583.ch53

[55] Enache-Angoulvant, A., & Hennequin, C. (2005). Invasive *Saccharomyces* infection: A comprehensive review. Clinical Infectious Diseases, 41(11), 1559–1568. https://doi.org/10.1086/497832

[56] Engle, E. K., & Fay, J. C. (2012). Divergence of the yeast transcription factor *FZF1* affects sulfite resistance. PLoS Genetics, 8(6). https://doi.org/10.1371/journal.pgen.1002763

[57] Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics, 164(4), 1567–1587. https://www.genetics.org/content/164/4/1567

[58] Fay, J. C., & Benavides, J. A. (2005). Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLoS Genetics, 1(1), 0066–0071. https://doi.org/10.1371/journal.pgen.0010005

[59] Fay, J. C., Liu, P., Ong, G. T., Dunham, M. J., Cromie, G. A., Jeffery, E. W., Ludlow, C. L., & Dudley, A. M. (2019). A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. PLOS Biology, 17(3), 466276. https://doi.org/10.1101/466276

[60] Fay, J. C., McCullough, H. L., Sniegowski, P. D., & Eisen, M. B. (2004). Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. Genome Biology, 5(4), 1–14. https://doi.org/10.1186/gb-2004-5-4-r26

[61] Fogel, S., & Welch, J. W. (1982). Tandem gene amplification mediates copper resistance in yeast. Proceedings of the National Academy of Sciences of the United States of America, 79(17 I), 5342–5346. https://doi.org/10.1073/pnas.79.17.5342

[62] Gabaldón, T. (2020). Hybridization and the origin of new yeast lineages. FEMS Yeast Research, 20(5), 1–8. https://doi.org/10.1093/femsyr/foaa040

[63] Galeote, V., Bigey, F., Beyne, E., Novo, M., Legras, J. L., Casaregola, S., & Dequin, S. (2011). Amplification of a *Zygosaccharomyces bailii* DNA segment in wine yeast genomes by extrachromosomal circular DNA formation. PLoS ONE, 6(3), 1–10. https://doi.org/10.1371/journal.pone.0017872

[64] Galeote, V., Novo, M., Salema-Oom, M., Brion, C., Valério, E., Gonçalves, P., & Dequin, S. (2010). *FSY1*, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter. Microbiology, 156(12), 3754–3761. https://doi.org/10.1099/mic.0.041673-0

[65] Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., & Verstrepen, K. J. (2016). Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. Cell, 166(6), 1397-1410.e16. https://doi.org/10.1016/j.cell.2016.08.020

[66] Giannakou, K., Cotterrell, M., & Delneri, D. (2020). Genomic Adaptation of *Saccharomyces* Species to Industrial Environments. Frontiers in Genetics, 11(8), 1–10. https://doi.org/10.3389/fgene.2020.00916

[67] Gibbons, J. G., & Rinker, D. C. (2015). The genomics of microbial domestication in the fermented food environment. Current Opinion in Genetics and Development, 35, 1–8. https://doi.org/10.1016/j.gde.2015.07.003

[68] 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., & Rokas, A. (2012). The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. Current Biology, 22(15), 1403–1409. https://doi.org/10.1016/j.cub.2012.05.033

[69] Gilliland, R. B. (1952). *Saccharomyces diastaticus* – A strach-fermenting yeast.

[70] Goddard, M. R., Anfang, N., Tang, R., Gardner, R. C., & Jun, C. (2010). A distinct population of *Saccharomyces cerevisiae* in New Zealand: Evidence for local dispersal by insects and human-aided global dispersal in oak barrels. Environmental Microbiology, 12(1), 63–73. https://doi.org/10.1111/j.1462-2920.2009.02035.x

[71] Goddard, M. R., & Greig, D. (2015). *Saccharomyces cerevisiae*: A nomadic yeast with no niche? FEMS Yeast Research, 15(3), 1–6. https://doi.org/10.1093/femsyr/fov009

[72] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 Genes. Science, 274(10), 546–567.

[73] Gonçalves, C., Wisecaver, J. H., Kominek, J., Salema Oom, M., Leandro, M. J., Shen, X. X., Opulente, D. A., Zhou, X., Peris, D., Kurtzman, C. P., Hittinger, C. T., Rokas, A., & Gonçalves, P. (2018). Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. ELife, *7*, 1–28. https://doi.org/10.7554/eLife.33034

[74] Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., & Sampaio, J. P. (2016). Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts. Current Biology, 26(20), 2750–2761. https://doi.org/10.1016/j.cub.2016.08.040

[75] González, S. S., Barrio, E., Gafner, J., & Querol, A. (2006). Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. FEMS Yeast Research, 6(8), 1221–1234. https://doi.org/10.1111/j.1567-1364.2006.00126.x

[76] González, S. S., Barrio, E., & Querol, A. (2008). Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. Applied and Environmental Microbiology, 74(8), 2314–2320. https://doi.org/10.1128/AEM.01867-07

[77] Goto-Yamamoto, N., Kitano, K., Shiki, K., Yoshida, Y., Suzuki, T., Iwata, T., Yamane, Y., & Hara, S. (1998). *SSU1*-R, a sulfite resistance gene of wine yeast, is an allele of *SSU1* with a different upstream sequence. Journal of Fermentation and Bioengineering, 86(5), 427–433. https://doi.org/10.1016/S0922-338X(98)80146-3

[78] Guillaume, C., Delobel, P., Sablayrolles, J. M., & Blondin, B. (2007). Molecular basis of fructose utilization by the wine yeast *Saccharomyces cerevisiae*: A mutated *HXT3* allele enhances fructose fermentation. Applied and Environmental Microbiology, 73(8), 2432–2439. https://doi.org/10.1128/AEM.02269-06

[79] Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307–321. https://doi.org/10.1093/sysbio/syq010

[80] Hall, C., Brachat, S., & Dietrich, F. S. (2005). Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. Eukaryotic Cell, 4(6), 1102–1115. https://doi.org/10.1128/EC.4.6.1102-1115.2005

[81] Hall, C., & Dietrich, F. S. (2007). The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. Genetics, 177(4), 2293–2307. https://doi.org/10.1534/genetics.107.074963

[82] Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. Journal of Heredity, 105(S1), 795–809. https://doi.org/10.1093/jhered/esu033

[83] Herskowitz, I. (1988). Life cycle of the budding yeast *Saccharomyces cerevisiae*. Microbiological Reviews, 52(4), 536–553. https://doi.org/10.1128/mmbr.52.4.536-553.1988

[84] Hittinger, C. T. (2013). *Saccharomyces* diversity and evolution: A budding model genus. Trends in Genetics, 29(5), 309–317. https://doi.org/10.1016/j.tig.2013.01.002

[85] Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution, 35(2), 518–522. https://doi.org/10.1093/molbev/msx281

[86] Hornsey, I. S. (2003). A history of beer and brewing (Vol. 34). Royal Society of Chemistry.

[87] Hutter, S., Vilella, A. J., & Rozas, J. (2006). Genome-wide DNA polymorphism analyses using VariScan. BMC Bioinformatics, 7, 1–10. https://doi.org/10.1186/1471-2105-7-409

[88] Keeling, P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics, 9(8), 605–618. https://doi.org/10.1038/nrg2386

[89] Krogerus, K., Magalhães, F., Kuivanen, J., & Gibson, B. (2019). A deletion in the *STA1* promoter determines maltotriose and starch utilization in *STA1*+ *Saccharomyces cerevisiae* strains. Applied Microbiology and Biotechnology, 103(18), 7597–7615. https://doi.org/10.1007/s00253-019-10021-y

[90] Krogerus, K., Preiss, R., & Gibson, B. (2018). A unique *Saccharomyces cerevisiae Saccharomyces uvarum* hybrid isolated from norwegian farmhouse beer: Characterization and reconstruction. Frontiers in Microbiology, 9(9), 1–15. https://doi.org/10.3389/fmicb.2018.02253

[91] Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution, 33(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

[92] Kurtzman, C. P., & Robnett, C. J. (2003). Phylogenetic relationships among yeasts of the "*Saccharomyces* complex" determined from multigene sequence analyses. FEMS Yeast Research, 3(4), 417–432. https://doi.org/10.1016/S1567-1356(03)00012-6

[93] Landry, C. R., Townsend, J. P., Hartl, D. L., & Cavalieri, D. (2006). Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. Molecular Ecology, 15(3), 575–591. https://doi.org/10.1111/j.1365-294X.2006.02778.x

[94] Legras, J. L., Erny, C., & Charpentier, C. (2014). Population structure and comparative genome hybridization of European flor yeast reveal a unique group of *Saccharomyces cerevisiae* strains with few gene duplications in their genome. PLoS ONE, 9(10). https://doi.org/10.1371/journal.pone.0108089

[95] Legras, J. L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldon, T., Schuller, D., Sampaio, J. P., & Dequin, S. (2018). Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. Molecular Biology and Evolution, 35(7), 1712–1727. https://doi.org/10.1093/molbev/msy066

[96] Legras, J. L., Galeote, V., Camarasa, C., Blondin, B., & Dequin, S. (2017). Ecology, Diversity and Applications of *Saccharomyces* Yeasts in Food and Beverages. In T. Satyanarayana & G. Kunze (Eds.), Yeast Diversity in Human Welfare (pp. 283–321). Springer Singapore. https://doi.org/10.1007/978-981-10-2621-8_12

[97] Legras, J. L., Merdinoglu, D., Cornuet, J. M., & Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Molecular Ecology, 16(10), 2091–2102. https://doi.org/10.1111/j.1365-294X.2007.03266.x

[98] Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research, 44(W1), W242–W245. https://doi.org/10.1093/nar/gkw290

[99] Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

[100] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

[101] Libkind, D., Hittinger, C. T., Valeŕio, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P., & Sampaio, J. P. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proceedings of the National Academy of Sciences of the United States of America, 108(35), 14539–14544. https://doi.org/10.1073/pnas.1105430108

[102] Limtong, S., & Koowadjanakul, N. (2012). Yeasts from phylloplane and their capability to produce indole-3-acetic acid. World Journal of Microbiology and Biotechnology, 28(12), 3323–3335. https://doi.org/10.1007/s11274-012-1144-9

[103] Liti, G. (2015). The fascinating and secret wild life of the budding yeast *S. cerevisiae*. ELife, 4, 1–9. https://doi.org/10.7554/eLife.05835

[104] Liti, G., Ba, A. N. N., Blythe, M., Müller, C. A., Bergström, A., Cubillos, F. A., Dafhnis-Calas, F., Khoshraftar, S., Malla, S., Mehta, N., Siow, C. C., Warringer, J., Moses, A. M., Louis, E. J., & Nieduszynski, C. A. (2013). High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome. BMC Genomics, 14(1), 69. https://doi.org/10.1186/1471-2164-14-69

[105] 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., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J. T., Van Oudenaarden, A., Barton, D. B. H., Bailes, E., Nguyen, A. N., Jones, M., Quail, A. M., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., & Louis, E. J. (2009). Population genomics of domestic and wild yeasts. Nature, 458(7236), 337–341. https://doi.org/10.1038/nature07743

[106] Lodder, J., & Kreger-van Rij, N. J. W. (1952). The yeasts-a taxonomic study. The Yeasts-a Taxonomic Study.

[107] Lopandic, K., Pfliegler, W. P., Tiefenbrunner, W., Gangl, H., Sipiczki, M., & Sterflinger, K. (2016). Genotypic and phenotypic evolution of yeast interspecies hybrids during high-sugar fermentation. Applied Microbiology and Biotechnology, 100(14), 6331–6343. https://doi.org/10.1007/s00253-016-7481-0

[108] Luyten, K., Riou, C., & Blondin, B. (2002). The hexose transporters of *Saccharomyces cerevisiae* play different roles during enological fermentation. Yeast, 19(8), 713–726. https://doi.org/10.1002/yea.869

[109] Makarova, K., Slesarev, A., Wolf, Y., & A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. P. and V. S. (2006). Comparative genomics of the lactic acid bacteria. Proceedings of the National Academy of Sciences, *103*(42), 15611-15616.

[110] Marsit, S., & Dequin, S. (2015). Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. FEMS Yeast Research, 15(7), 1–12. https://doi.org/10.1093/femsyr/fov067

[111] Marsit, S., Leducq, J. B., Durand, É., Marchant, A., Filteau, M., & Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. In Nature Reviews Genetics. https://doi.org/10.1038/nrg.2017.49

[112] Marsit, S., Mena, A., Bigey, F., Sauvage, F. X., Couloux, A., Guy, J., Legras, J. L., Barrio, E., Dequin, S., & Galeote, V. (2015). Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. Molecular Biology and Evolution, 32(7), 1695–1707. https://doi.org/10.1093/molbev/msv057

[113] Mcfarland, L. V., & Bernasconi, P. (1993). *Saccharomyces boulardii*: A review of an innovative biotherapeutic agent. Microbial Ecology in Health and Disease, 6(4), 157–171. https://doi.org/10.3109/08910609309141323

[114] McGovern, P. E. (2009). Uncorking the Past: The Quest for Wine, Beer, and Other Alcoholic Beverages. University of California Press. https://doi.org/doi:10.1525/9780520944688

[115] McGovern, P. E., Glusker, D. L., & Exner, L. J. (1996). Neolithic resinated wine (pp. 480–481). Scientific Correspondence. https://www.nature.com/articles/381480a0

[116] 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., Cheng, G., Zhao, Z., & Wang, C. (2004). Fermented beverages of pre- and proto-historic China. Proceedings of the National Academy of Sciences of the United States of America, 101(51), 17593–17598. https://doi.org/10.1073/pnas.0407921102

[117] Meier-Dörnberg, T., Kory, O. I., Jacob, F., Michel, M., & Hutzler, M. (2018). *Saccharomyces cerevisiae* variety *diastaticus* friend or foe? - spoilage potential and brewing ability of different *Saccharomyces cerevisiae* variety *diastaticus* yeast isolates by genetic, phenotypic and physiological characterization. FEMS Yeast Research, 18(4). https://doi.org/10.1093/femsyr/foy023

[118] Meyen, F. J. F. (1839). Jahresbericht über die Resultate der Arbeiten im Felde der physiologischen Botanik: von dem Jahre 1837 (1839). Arch Naturgesch. 4, 1-186.

[119] Morales, L., & Dujon, B. (2012). Evolutionary Role of Interspecies Hybridization and Genetic Exchanges in Yeasts. Microbiology and Molecular Biology Reviews, 76(4), 721–739. https://doi.org/https://doi.org/10.1128/MMBR.00022-12

[120] Mortimer, R., & Polsinelli, M. (1999). On the origins of wine yeast. Research in Microbiology, 150(3), 199–204. https://doi.org/10.1016/S0923-2508(99)80036-9

[121] Mukai, N., Masaki, K., Fujii, T., & Iefuji, H. (2014). Single nucleotide polymorphisms of *PAD1* and *FDC1* show a positive relationship with ferulic acid decarboxylation ability among industrial yeasts used in alcoholic beverage production. Journal of Bioscience and Bioengineering, 118(1), 50–55. https://doi.org/10.1016/j.jbiosc.2013.12.017

[122] Muller, L. A. H., Lucas, J. E., Georgianna, D. R., & McCusker, J. H. (2011). Genome-wide association analysis of clinical vs. nonclinical origin provides insights into *Saccharomyces cerevisiae* pathogenesis. Molecular Ecology, 20(19), 4085–4097. https://doi.org/10.1111/j.1365-294X.2011.05225.x

[123] Muller, L. A. H., & McCusker, J. H. (2009). A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *Saccharomyces cerevisiae*. FEMS Yeast Research, 9(1), 143–152. https://doi.org/10.1111/j.1567-1364.2008.00464.x

[124] Muñoz, P., Bouza, E., Cuenca-Estrella, M., Eiros, J. M., Pérez, M. J., Sánchez-Somolinos, M., Rincón, C., Hortal, J., & Peláez, T. (2005). *Saccharomyces cerevisiae* fungemia: An emerging infectious disease. Clinical Infectious Diseases, 40(11), 1625–1634. https://doi.org/10.1086/429916

[125] Naumov, G. (1996). Genetic identification of biological species in the *Saccharomyces* sensu stricto complex. *Journal of Industrial Microbiology and Biotechnology, 17*(3–4), 295–302. https://doi.org/10.1007/bf01574704

[126] Naumov, G., Lee, C. F., & Naumova, E. S. (2013). Molecular genetic diversity of the *Saccharomyces* yeasts in Taiwan: *Saccharomyces arboricola*, *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 103(1), 217–228. https://doi.org/10.1007/s10482-012-9803-2

[127] Naumov, G., Naumova, E. S., & Korhola, M. P. (1995). Chromosomal polymorphism of *MEL* genes in some populations of *Saccharomyces cerevisiae*. FEMS Microbiology Letters, 127(1–2), 41–45. https://doi.org/10.1111/j.1574-6968.1995.tb07447.x

[128] Naumov, G., Naumova, E. S., Korshunova, I. V., & Jakobsen, M. (2002). Yeast Comparative Genetics: A New *MEL15* α-Galactosidase Gene of *Saccharomyces cerevisiae*. Russian Journal of Genetics, 38(10), 1127–1132. https://doi.org/10.1023/A:1020640418358

[129] Naumov, G., Naumova, E., Turakainen, H., Suominen, P., & Korhola, M. (1991). Polymeric genes *MEL8*, *MEL9* and *MEL10* - new members of α-galactosidase gene family in *Saccharomyces cerevisiae*. Current Genetics, 20(4), 269–276. https://doi.org/10.1007/BF00318514

[130] Naumov, G., Turakainen, H., Naumova, E., Aho, S., & Korhola, M. (1990). A new family of polymorphic genes in *Saccharomyces cerevisiae*: α-galactosidase genes *MEL1-MEL7*. MGG Molecular & General Genetics, 224(1), 119–128. https://doi.org/10.1007/BF00259458

[131] Naumova, E. S., Korshunova, I. V., & Naumov, G. I. (2003). Molecular Analysis of the α-Galactosidase *MEL* Genes in Yeast *Saccharomyces* sensu stricto. In Molecular Biology (Vol. 37, Issue 5, pp. 699–706). https://doi.org/10.1023/A:1026032926357

[132] Naumova, E. S., Serpova, E. V., Korshunova, I. V., & Naumov, G. I. (2011). Molecular polymorphism of α-galactosidase *MEL* genes of *Saccharomyces* yeasts. Microbiology, 80(4), 502–513. https://doi.org/10.1134/S0026261711040151

[133] Ness, F., & Aigle, M. (1995). RTM1: A member of a new family of telomeric repeated genes in yeast. Genetics, 140(3), 945–956. https://doi.org/10.1093/genetics/140.3.945

[134] Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300

[135] Nijkamp, J. F., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M. A., Daran-Lapujade, P., Vongsangnak, W., Nielsen, J., Heijne, W. H. M., Klaassen, P., Paddon, C. J., Platt, D., Kötter, P., van Ham, R. C., Reinders, M. J. T., Pronk, J. T., de Ridder, D., & Daran, J. M. (2012). De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. Microbial Cell Factories, 11(3). https://doi.org/10.1186/1475-2859-11-36

[136] Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J. L., Wincker, P., Casaregola, S., & Dequin, S. (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proceedings of the National Academy of Sciences of the United States of America, 106(38), 16333–16338. https://doi.org/10.1073/pnas.0904673106

[137] Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O. G., Kandrov, D., Rasputin, K., Syabro, M., & Tleukenov, T. (2012). Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics, 28(8), 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

[138] Pais, P., Almeida, V., Yılmaz, M., & Teixeira, M. C. (2020). *Saccharomyces boulardii*: What makes it tick as successful probiotic? Journal of Fungi, 6(2), 1–15. https://doi.org/10.3390/jof6020078

[139] Papadimitriou, K., Pot, B., & Tsakalidou, E. (2015). How microbes adapt to a diversity of food niches. Current Opinion in Food Science, 2, 29–35. https://doi.org/10.1016/j.cofs.2015.01.001

[140] Park, H., & Bakalinsky, A. T. (2000). *SSU1* mediates sulphite efflux in *Saccharomyces cerevisiae*. Yeast, 16(10), 881–888. https://doi.org/10.1002/1097-0061(200007)16:10<881::aid-yea576>3.3.co;2-v

[141] Pérez-Ortín, J. E., Querol, A., Puig, S., & Barrio, E. (2002). Molecular characterization of a chromosomal rearrangement involved in the adaptie evolution of yeast strains. Genome Research, 12(10), 1533–1539. https://doi.org/10.1101/gr.436602

[142] Peris, D., Pérez-Torrado, R., Hittinger, C. T., Barrio, E., & Querol, A. (2018). On the origins and industrial applications of *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* hybrids. Yeast, 35(1), 51–69. https://doi.org/10.1002/yea.3283

[143] Peter, J., Chiara, M. De, Friedrich, A., Yue, J., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., & Liti, G. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature, 556, 339–344.

[144] Piškur, J., Rozpedowska, E., Polakova, S., Merico, A., & Compagno, C. (2006). How did *Saccharomyces* evolve to become a good brewer? Trends in Genetics, 22(4), 183–186. https://doi.org/10.1016/j.tig.2006.02.002

[145] Pontes, A., Čadež, N., Gonçalves, P., & Sampaio, J. P. (2019). A Quasi-Domesticate Relic Hybrid Population of *Saccharomyces cerevisiae* × *S. paradoxu*s Adapted to Olive Brine. Frontiers in Genetics, 10(5), 1–14. https://doi.org/10.3389/fgene.2019.00449

[146] Posas, F., Chamber, J. R., Heyman, J. A., Hoeffler, J. P., De Nadal, E., & Ariño, J. (2000). The transcriptional response of yeast to saline stress. Journal of Biological Chemistry, 275(23), 17249–17255. https://doi.org/10.1074/jbc.M910016199

[147] Preiss, R., Tyrawa, C., Krogerus, K., Garshol, L. M., & Van Der Merwe, G. (2018). Traditional Norwegian Kveik are a genetically distinct group of domesticated *Saccharomyces cerevisiae* brewing yeasts. Frontiers in Microbiology, 9(9). https://doi.org/10.3389/fmicb.2018.02137

[148] Reuter, M., Bell, G., & Greig, D. (2007). Increased outbreeding in yeast in response to dispersal by an insect vector. Current Biology, 17(3), 81–83. https://doi.org/10.1016/j.cub.2006.11.059

[149] Roach, M. J., & Borneman, A. R. (2020). New genome assemblies reveal patterns of domestication and adaptation across *Brettanomyces* (Dekkera) species. BMC Genomics, 21(1), 1–14. https://doi.org/10.1186/s12864-020-6595-z

[150] Robinson, H. A., Pinharanda, A., & Bensasson, D. (2016). Summer temperature can predict the distribution of wild yeast populations. Ecology and Evolution, 6(4), 1236–1250. https://doi.org/10.1002/ece3.1919

[151] Romano, P., Capece, A., & Jespersen, L. (2006). Taxonomic and Ecological Diversity of Food and Beverage Yeasts. In Yeasts in Food and Beverages. https://doi.org/10.1007/978-3-540-28398-0_2

[152] 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., & Giraud, T. (2015). Adaptive horizontal gene transfers between multiple cheese-associated fungi. Current Biology, 25(19), 2562–2569. https://doi.org/10.1016/j.cub.2015.08.025

[153] Rosa C, AM S, Faria J. (2009). Cachaça production. In Ingledew WM, Kersal DR, Austin G., Kluhsipies C. The alcohol textbook. 5th ed. Nottingham University Press. 481-490.

[154] Ross-Ibarra, J., Morrell, P. L., & Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. In the Light of Evolution, 1, 205–223. https://doi.org/10.17226/11790

[155] Ruiz, A., & Ariño, J. (2007). Function and regulation of the *Saccharomyces cerevisiae* ENA sodium ATPase system. Eukaryotic Cell, *6*(12), 2175–2183. https://doi.org/10.1128/EC.00337-07

[156] Saito, H., & Posas, F. (2012). Response to hyperosmotic stress. Genetics, 192(2), 289–318. https://doi.org/10.1534/genetics.112.140863

[157] Salque, M., Bogucki, P. I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M., & Evershed, R. P. (2013). Earliest evidence for cheese making in the sixth millennium bc in northern Europe. Nature, 493(7433), 522–525. https://doi.org/10.1038/nature11698

[158] Sampaio, J. P. (2018). Microbe Profile: *Saccharomyces eubayanus*, the missing link to lager beer yeasts. Microbiology, 164(9), 1069–1071. https://doi.org/10.1099/mic.0.000677

[159] Sampaio, J. P., & Gonçalves, P. (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(7), 2144–2152. https://doi.org/10.1128/AEM.02396-07

[160] Sampaio, J. P., & Gonçalves, P. (2017). Biogeography and Ecology of the Genus *Saccharomyces*. In P. Buzzini, M.-A. Lachance, & A. Yurkov (Eds.), Yeasts in Natural Ecosystems: Ecology (pp. 131–153). Springer International Publishing. https://doi.org/10.1007/978-3-319-61575-2_5

[161] Santa María, J. (1958). Ecología de las levaduras. II. Levaduras del azúcar, de la leche condensada y de la remolacha. An. Del Inst. Nac. Investig. Agronómicas, 8, 779–796.

[162] Santa María, J. (1962). *Saccharomyces hienipiensis*, a New Melibiose-fermenting Yeast, Unable to Assimilate Raffinose. Journal of General Microbiology, 28(2), 375–384. https://doi.org/10.1099/00221287-28-2-375

[163] Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., Rine, J., Johnston, M., & Hittinger, C. T. (2011). The awesome power of yeast evolutionary genetics: New genome sequences and strain resources for the *Saccharomyces* sensu stricto genus. G3: Genes, Genomes, Genetics, 1(1), 11–25. https://doi.org/10.1534/g3.111.000273

[164] Schacherer, J., Shapiro, J. A., Ruderfer, D. M., & Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. Nature, 458(7236), 342–345. https://doi.org/10.1038/nature07670

[165] Scott, A. L., Richmond, P. A., Dowell, R. D., & Selmecki, A. M. (2017). The Influence of Polyploidy on the Evolution of Yeast Grown in a Sub-Optimal Carbon Source. Molecular Biology and Evolution, 34(10), 2690–2703. https://doi.org/10.1093/molbev/msx205

[166] Smukowski Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., & Dunham, M. J. (2017). Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. Molecular Biology and Evolution, 34(7), 1596–1612. https://doi.org/10.1093/molbev/msx098

[167] Sniegowski, P. D., Dombrowski, P. G., & Fingerman, E. (2002). *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEMS Yeast Research, *1*(4), 299–306. https://doi.org/10.1016/S1567-1356(01)00043-5

[168] Steensels, J., Gallone, B., Voordeckers, K., & Verstrepen, K. J. (2019). Domestication of Industrial Microbes. Current Biology, 29(10), R381–R393. https://doi.org/10.1016/j.cub.2019.04.025

[169] Stefanini, I., Dapporto, L., Legras, J. L., Calabretta, A., Di Paola, M., De Filippo, C., Viola, R., Capretti, P., Polsinelli, M., Turillazzi, S., & Cavalieri, D. (2012). Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. Proceedings of the National Academy of Sciences of the United States of America, 109(33), 13398–13403. https://doi.org/10.1073/pnas.1208362109

[170] Stelling Dekker, N. M. (1931). Die sporogenen Hefen. Verh. K. Ned. Akad. Wetensch. Afd. Natuurk.

[171] Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., & McCusker, J. H. (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, 125(5), 762–774. https://doi.org/10.1101/gr.185538.114

[172] Tilakaratna, V., & Bensasson, D. (2017). Habitat predicts levels of genetic admixture in *Saccharomyces cerevisiae*. G3: Genes, Genomes, Genetics, 7(9), 2919–2929. https://doi.org/10.1534/g3.117.041806

[173] Torresi, S., Frangipane, M. T., & Anelli, G. (2011). Biotechnologies in sparkling wine production. Interesting approaches for quality improvement: A review. Food Chemistry, 129(3), 1232–1241. https://doi.org/10.1016/j.foodchem.2011.05.006

[174] Turakainen, H., Aho, S., & Korhola, M. (1993). *MEL* gene polymorphism in the genus *Saccharomyces*. Applied and Environmental Microbiology, 59(8), 2622–2630. https://doi.org/10.1128/aem.59.8.2622-2630.1993

[175] Turakainen, Hilkka, Kristo, P., & Korhola, M. (1994). Consideration of the evolution of the *Saccharomyces cerevisiae MEL* gene family on the basis of the nucleotide sequences of the genes and their flanking regions. *Yeast, 10(1*2), 1559–1568. https://doi.org/10.1002/yea.320101205

[176] van der Walt, J. P. (1965). *Saccharomyces* vafer and *S. inconspicuus* spp.n. Antonie van Leeuwenhoek, 31(1), 187–192. https://doi.org/10.1007/BF02045890

[177] Van der Walt, J. P. (1970). *Saccharomyces* Meyen emend. Reess. The Yeasts, a Taxonomic Study. Amsterdam: North-Holland Publishing Company, 555–718.

[178] Van Uden, N., & Assis-Lopes, L. (1957). Fermentation of raffinose in the absence of invertase by *Saccharomyces italiens castelli* var. *melibiosi* nov. var. Portugaliae Acta Biologica, 4(4).

[179] Vaughan Martini, A., & Martini, A. (1987). Three newly delimited species of *Saccharomyces* sensu stricto. Antonie van Leeuwenhoek, 53(2), 77–84. https://doi.org/10.1007/bf00419503

[180] Vaughan Martini, A., & Martini, A. (1995). Facts, myths and legends on the prime industrial microorganism. Journal of Industrial Microbiology, 14(6), 514–522. https://doi.org/10.1007/BF01573967

[181] Vianna, C. R., Silva, C. L. C., Neves, M. J., & Rosa, C. A. (2008). *Saccharomyces cerevisiae* strains from traditional fermentations of Brazilian cachaça: Trehalose metabolism, heat and ethanol resistance. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 93(1–2), 205–217. https://doi.org/10.1007/s10482-007-9194-y

[182] Wang, Q. M., Liu, W. Q., Liti, G., Wang, S. A., & Bai, F. Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. Molecular Ecology, 21(22), 5404–5417. https://doi.org/10.1111/j.1365-294X.2012.05732.x

[183] Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., Forsmark, A., Durbin, R., Omholt, S. W., Louis, E. J., Liti, G., Moses, A., & Blomberg, A. (2011). Trait variation in yeast is defined by population history. PLoS Genetics, 7(6). https://doi.org/10.1371/journal.pgen.1002111

[184] Wiens, F., Zitzmann, A., Lachance, M. A., Yegles, M., Pragst, F., Wurst, F. M., Von Holst, D., Saw, L. G., & Spanagel, R. (2008). Chronic intake of fermented floral nectar by wild treeshrews. Proceedings of the National Academy of Sciences of the United States of America, 105(30), 10426–10431. https://doi.org/10.1073/pnas.0801628105

[185] Will, J. L., Kim, H. S., Clarke, J., Painter, J. C., Fay, J. C., & Gasch, A. P. (2010). Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* populations. PLoS Genetics, 6(4). https://doi.org/10.1371/journal.pgen.1000893

[186] Yarrow, D. (1984). *Saccharomyces* Meyen ex Reess. The Yeasts: A Taxonomic Study, 379–395.

[187] Yue, J., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. Nature Publishing Group, 49(6), 913–924. https://doi.org/10.1038/ng.3847

[188] Zanello, G., Meurens, F., Berri, M., & Salmon, H. (2009). *Saccharomyces boulardii* effects on gastrointestinal diseases. Current Issues in Molecular Biology, 11(33), 47–58. https://doi.org/10.21775/cimb.011.047

[189] Zheng, D. Q., Wang, P. M., Chen, J., Zhang, K., Liu, T. Z., Wu, X. C., Li, Y. D., & Zhao, Y. H. (2012). Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJS329. BMC Genomics, 13(1). https://doi.org/10.1186/1471-2164-13-479

[190] Zimmer, A., Durand, C., Loira, N., Durrens, P., Sherman, D. J., & Marullo, P. (2014). QTL dissection of lag phase in wine fermentation reveals a new translocation responsible for *Saccharomyces cerevisiae* adaptation to sulfite. PLoS ONE, 9(1), 37–39. https://doi.org/10.1371/journal.pone.0086298

A1
Appendix 1

Supplementary information pertaining to Chapter 2

Barbosa, R.†; Pontes, A.†; Santos, R.O.; Montandon, G.G.; Ponzzes-Gomes, C.M.; Morais, P.B.; Gonçalves, P.; Rosa, C.A.; Sampaio, J.P. Multiple Rounds of Artificial Selection Promote Microbe Secondary Domestication—The Case of Cachaça Yeasts. Genome Biology and Evolution 2018, 10(8), 1939-1955.

https://doi.org/10.1093/gbe/evy132

Table A1.1 - Strains and genomes used in the work presented in chapter 2 and relevant information pertaining to them.

Strain	Substrate of isolation	Geographic location	Phylogeny	Heterozigous sites (Q >40)	MD / SCD	A	B	C	AQY1		AQY2		RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
AWRI 1631	Wine	Australia	Wine	n.d.				8/19	NF	A881del	NF	11bp del			F	F	W					PRJNA30553	Borneman et al., 2008
AWRI 796	Commercial wine yeast	South Africa	Wine	1041 *			DUP		F		NF	11bp del			F	F						PRJNA48559	Borneman et al., 2011
EXF 6719	Wine must	Cuber, Slovenia	Wine	2053				17/19	NF	A881del	NF	11bp del			F	F	W					PRJEB7601	Almeida et al., 2015
ZP 641	Spontaneous red wine fermentation	Castelo de Vide, Portugal	Wine	1484					NF	A881del	NF	11bp del			F	F						PRJEB7601	Almeida et al., 2015
L1374	Wine (must from País variety)	Cauquenes, Chile	Wine	1550			5/5		NF	A881del	NF	11bp del			F	F						SGRP2	Bergström et al., 2014
L1528	Wine (must from Cabernet variety)	Cauquenes, Chile	Wine	1334			5/5		NF	A881del	NF	11bp del			F	F						SGRP2	Bergström et al., 2014
Lalvin BM45	Commercial yeast (white wine)		Wine	1280			5/5	2/19	NF	A881del	NF	11bp del			F	F						PRJEB19382	Almeida et al., 2017
Lalvin CY-3079	Commercial yeast (red wine)		Wine	1974			5/5	4/19	NF	A881del	NF	11bp del			F	F						PRJEB19382	Almeida et al., 2017
Lalvin W15	Commercial yeast (white nad red wine)		Wine	2209			5/5		F		NF	11bp del			F	F						PRJEB19382	Almeida et al., 2017
PR	Pasteur Red		Wine	1715					NF	A881del	NF	11bp del			F	F	W					PRJEB19382	Almeida et al., 2017
PYCC 4072	commercial wine yeast (Fermivin)	Portugal	Wine	1889					NF	A881del	NF	11bp del			F	F						PRJEB19382	Almeida et al., 2017
TUM V1	Bordeaux wine		Wine	1827					NF	A881del	NF	11bp del			F	F						PRJEB19382	Almeida et al., 2017
Uvaferm SGv	Commercial yeast (red wine)		Wine	1734				4/19	NF	A881del	NF	11bp del			F	F						PRJEB19382	Almeida et al., 2017
Uvaferm VRB	Commercial yeast (red wine)		Wine	2750				19/19	NF	A881del	NF	11bp del			F	F						PRJEB19382	Almeida et al., 2017
Vin 13	Commercial wine yeast		Wine	15216 *			DUP	2/19	F		NF	11bp del			F	F						PRJNA48563	Borneman et al., 2011
VL3	Commercial wine yeast		Wine	9904 *			5/5	8/19	NF	A881del	NF	11bp del			F	F						PRJNA48565	Borneman et al., 2011
WE 372	Commercial wine yeast	Cape Town, South Africa	Wine	4354			5/5		F		NF	11bp del			F	F						PRJNA60199	Justin Fay, Washington University
YJM 1332	Wine	Italy	Wine	1660				2/19	NF	A881del	NF	11bp del			F	F						PRJNA189896	Strope et al., 2015
YJM 1336	Wine	Italy	Wine	1547					NF	A881del	NF	11bp del			F	F						PRJNA189897	Strope et al., 2015
YJM 1341	Grape must	South Africa	Wine	2350					NF	A881del	NF	11bp del			F	F						PRJNA189899	Strope et al., 2015
YJM 1415	Wine	France	Wine	1256			5/5	19/19	NF	A881del	NF	11bp del			F	F						PRJNA189914	Strope et al., 2015
EC 1118	Industrial strain isolated from Champagne	France	Wine	n.d.		15/15	5/5	19/19	NF	A881del	NF	11bp del			F	F	W					PRJEA37863	Novo et al., 2009
IOC 18-2007	Commercial yeast		Wine	2181			5/5		NF	A881del	F				F	F						PRJEB19382	Almeida et al., 2017
IOC 9002	Commercial wine yeast		Wine	1224			5/5	4/19	NF	A881del	NF	11bp del			F	F	W					PRJNA264372	Almeida et al., 2015

Strain	Substrate of isolation	Geographic location	Phylogeny	Het-erozi-gous sites (Q >40)	MD / SCD	A	B	C	AQY1			AQY2			RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
Lalvin QA23	Commercial wine yeast	Portugal	Wine	18861 *		15/15	TRIP	19/19	NF	A881del		F					F	F						PRJNA48561	Borneman et al., 2011
YJM 1574	Wine		Wine	1796			5/5		NF	A881del		NF	11bp del				F	F						PRJNA189934	Strope et al., 2015
YJM 270	Wine	Slovenia	Wine	1662				19/19	NF	A881del		F					F	F						PRJNA189852	Strope et al., 2015
PYCC 4074	commercial wine yeast (Fermichamp)	Portugal	Wine	2263		15/15	5/5		NF	A881del		F					F	F						PRJEB19382	Almeida et al., 2017
PYCC 6722	wine	South Armenia	Wine	2121					NF	A881del		F					F	F						PRJEB19382	Almeida et al., 2017
PYCC 6726	Jerez-wine	Spain	Wine	4939					NF	A881del		F					F	F						PRJEB19382	Almeida et al., 2017
PYCC 6729	Jerez-wine	Armenia	Wine	2901					NF	A881del		F					F	F						PRJEB19382	Almeida et al., 2017
UFMG-CM-Y215	Cachaça	São Paulo, Brazil	Wine	3242	MD				NF	A881del		NF	11bp del				F	F						PRJEB24932	This study
UFMG-CM-Y617	Jabuticaba wine	Santa Bárbara, Minas Gerais, Brazil	Wine	2805	SCD				NF	A881del		NF	11bp del				F	F	W					PRJEB24932	This study
UFMG-CM-Y618	Jabuticaba wine	Santa Bárbara, Minas Gerais, Brazil	Wine	3394	MD		5/5	17/19	NF	A881del		NF	11bp del				F	F						PRJEB24932	This study
UFMG-CM-Y619	Jabuticaba wine	Santa Bárbara, Minas Gerais, Brazil	Wine	3317	MD		5/5	18/19	NF	A881del		NF	11bp del				F	F						PRJEB24932	This study
UFMG-CM-Y620	Wine	Casa Nova, Bahia, Brazil	Wine	3036	MD		5/5		F			NF	11bp del				F	F						PRJEB24932	This study
UFMG-CM-Y621	Wine	Casa Nova, Bahia, Brazil	Wine	3277	MD		5/5		F			NF	11bp del				F	F						PRJEB24932	This study
PYR 4b	*Quercus pubescens*	Halkidiki, Greece	MO	2738					F			F					F	F						PRJEB7601	Almeida et al., 2015
OakRom 3-2a	Oak	Near Bucarest, Romania	MO	2366					F			F					F	F						PRJEB7675	Almeida et al., 2015
EXF 7200	*Quercus robur*	Jasenovo polje, Montenegro	MO	3001					F			F					F	F						PRJNA264372	Almeida et al., 2015
HUN 9.1s1	Oak	Hungary	MO	2745					F			F					F	F						PRJEB7601	Almeida et al., 2015
3.10	*Quercus cerris*	Parco del Monte Subasio, Italy	MO	3800					F			F					F	F						PRJEB7601	Almeida et al., 2015
ZP 541	*Fagus sylvatica*	Adagoi, Portugal	MO	2684					F			F					F	F						PRJEB7601	Almeida et al., 2015
ZP 1008	*Quercus faginea*	Eja, Melres, Douro, Portugal	MO	2409					n.d.			n.d.			n.d.	n.d.	n.d.	n.d.						PRJNA264372	Almeida et al., 2015
ZP 570	*Fraxinus* sp.	Paul Boquilobo, Portugal	MO	2813					F			F					F	F						PRJEB7601	Almeida et al., 2015
ZP 848	*Quercus ilex*	Alter do Chão, Portugal	MO	2656					F			F					F	F						PRJEB7675	Almeida et al., 2015
ZP 850	*Quercus ilex*	Alconorales Natural Park, Andaluzia, Spain	MO	2869					F			F					F	F						PRJEB7601	Almeida et al., 2015
MB 7c	*Quercus pubescens*	Montbarri, Southern France	MO	3752					n.d.			n.d.			n.d.	n.d.	n.d.	n.d.						PRJNA264372	Almeida et al., 2015
ZP 560	*Quercus pyrenaica*	Castelo de Vide, Portugal	MO	2553					F			F					F	F						PRJEB7601	Almeida et al., 2015

Strain	Substrate of isolation	Geographic location	Phylogeny	Heterozigous sites (Q >40)	MD / SCD	A	B	C	AQY1		AQY2		RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
BE013	Beer (Ale refermentation)	Belgium	Beer 2	n.d.			5/5		NF	A881del	F				F	F	W					MBZP00000000	Gallone et al., 2016
BE021	Beer (Pale ale)	Canada	Beer 2	n.d.			5/5		NF	A881del	F				F	F						MBZH00000000	Gallone et al., 2016
BE032	Beer	England	Beer 2	n.d.			5/5	4/19	NF	A881del	NF	11bp del			F	F						MBYW00000000	Gallone et al., 2016
BE063	Beer	England	Beer 2	n.d.		15/15	5/5		NF	V121M	F				F	F	W					MBXR00000000	Gallone et al., 2016
BE083	Beer (Saison)	Belgium	Beer 2	n.d.			5/5		NF	A881del	NF	11bp del			F	F						MBWX00000000	Gallone et al., 2016
BE092	Beer (Strong ale)	Belgium	Beer 2	n.d.		15/15			NF	A881del	F				F	F						MBWO00000000	Gallone et al., 2016
UFMG-CM-Y030	Cachaça (commercial strain)	Minas Gerais, Brazil	C1	31074	SCD		5/5		NF	A881del	NF	G25del			F	F	C W					PRJEB24932	This study
UFMG-CM-Y625	Cachaça	Novo Cruzeiro, Minas Gerais, Brazil	C1	4737	SCD		5/5		NF	T498del	F				F	F	W					PRJEB24932	This study
UFMG-CM-Y251	Cachaça	Santa Catarina, Brazil	C1	12655	SCD		5/5	19/19	NF	A881del	NF	11bp del			F	F	C					PRJEB24932	This study
UFMG-CM-Y630	Cachaça	Santa Catarina, Brazil	C1	3570	MD		5/5		F		NF	11bp del			F	F	C					PRJEB24932	This study
UFMG-CM-Y631	Cachaça	Santa Catarina, Brazil	C1	4036	MD		5/5		NF	A881del	NF	11bp del			F	F	C					PRJEB24932	This study
UFMG-CM-Y627	Cachaça	Rio de Janiero, Brazil	C1	28953	SCD		5/5		NF	A881del	NF	11bp del			F	F	W					PRJEB24932	This study
UFMG-CM-Y629	Cachaça	Pernambuco, Brazil	C1	4332	SCD				NF	T498del	NF	11bp del			F	F	C					PRJEB24932	This study
UFMG-CM-Y624	Cachaça	Salinas, Minas Gerais, Brazil	C1	27658	MD				NF	T498del	F				F	F	W					PRJEB24932	This study
UFMG-CM-Y632	Cachaça	Tocantins, Brazil	C2	12225	SCD				NF	A881del	NF	11bp del			F	F	C					PRJEB24932	This study
UFMG-CM-Y633	Cachaça	Tocantins, Brazil	C2	6833	MD				NF	A881del	F				F	F	W					PRJEB24932	This study
UFMG-CM-Y634	Cachaça	Tocantins, Brazil	C2	21251	MD		5/5		NF	A881del	NF	11bp del			F	F	C					PRJEB24932	This study
UFMG-CM-Y635	Cachaça	Tocantins, Brazil	C2	22600	MD		5/5		NF	A881del	NF	11bp del			F	F	C					PRJEB24932	This study
UFMG-CM-Y260	*Tapirira guianensis*	Tocantins, Brazil	C2	3236	MD		5/5		F		F				F	F	W					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y636	*Tapirira guianensis*	Tocantins, Brazil	C2	3747	MD		5/5		NF	T498del	NF	11bp del			F	F	W					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y623	Cachaça	Betim, Minas Gerais, Brazil	?	34631	SCD		5/5		F		NF	11bp del			F	F	W					PRJEB24932	This study
UFMG-CM-Y628	Cachaça	Rio de Janeiro, Brazil	?	81659	SCD		5/5		NF	A817del	NF	11bp del			F	F	C W					PRJEB24932	This study
UFMG-CM-Y637	Cachaça	Pernambuco, Brazil	?	4809	MD		4/5	4/19	F		F				F	F	W					PRJEB24932	This study
UFMG-CM-Y638	Cachaça	Tocantins, Brazil	?	5932	MD		5/5		F		F				F	F	W					PRJEB24932	This study
UFMG-CM-Y639	*Tapirira guianensis*	Tocantins, Brazil	?	5249	MD		5/5		F		F				F	F	W					PRJEB11698	Barbosa et al., 2016

Strain	Substrate of isolation	Geographic location	Phylogeny	Het-erozi-gous sites (Q >40)	MD / SCD	A	B	C	AQY1			AQY2			RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
UFMG-CM-Y626	Bioethanol-produc-ing strain from sugar cane	Brazil	Bioethanol	n.d.			4/5		NF	A881del		NF	11bp del				F	F	W					PRJNA60391	Justin Fay, Washington University
JAY 291	Haploid derivative of bioethanol-producing strain JAY 270	Brazil	Bioethanol	n.d.			5/5		NF	A881del		NF	11bp del				F	F	W					PRJNA32809	Argueso et al., 2009
BG-1	Bioethanol-produc-ing strain from sugar cane	Brazil	Bioethanol	n.d.			4/5		NF	A881del		NF	20bp ins				F	F	C					PRJNA352845	Coutouné et al., 2017
PYCC 4226	Commercial baker's yeast		Bread	30081			4/5	2/19	NF	A881del		NF		792stop			F	F	W					ERS1108635	Gonçalves et al., 2016
Platinum	Commercial baker's yeast		Bread	42278			4/5	15/19	NF	V121M	A822ins	NF		792stop			F	F	W					ERS1108633	Gonçalves et al., 2016
AP13.1	Baker's yeast	Portugal	Bread	53091			5/5	17/19	NF	V121M		NF	11bp del				NF	NF	C W					PRJEB24932	This study
Morocco Bread G17	Baker's yeast		Bread	61910			5/5		NF	V121M	A822ins	F					F	NF	C W					SRR403236	Dunn et al., 2012
TUM 506	British ale	Great Britain	Bread	50225			4/5	16/19	NF	V121M	A822ins	NF	11bp del				F	F	W					ERS1108634	Gonçalves et al., 2016
TUM 480	Opaque beer	South Africa	Bread	41556			5/5		NF		A822ins	NF	11bp del				F	NF	C W					ERS1108636	Gonçalves et al., 2016
BE005	Ale beer	Belgium	Bread	n.d.			4/5	17/19	NF	V121M		F					F	NF	W					MBZX00000000	Gallone et al., 2016
BE029	Ale beer (refermentation)	Belgium	Bread	n.d.			4/5	17/19	NF	V121M		NF	11bp del				F	NF	W					MBYZ00000000	Gallone et al., 2016
BR004	Bread	Belgium	Bread	n.d.			5/5	17/19	NF	V121M		NF	11bp del				F	NF	W					MBVV00000000	Gallone et al., 2016
UFMG-CM-Y223	Cachaça	Rio de Janeiro, Brazil	Bread	46892	MD		3/5		NF	V121M	A881del	F					F	F	C					PRJEB24932	This study
UFMG-CM-Y622	Cachaça	Pernambuco, Brazil	Bread	37077	SCD		4/5	17/19	NF	V121M	A822ins	NF		786stop			F	F	C W					PRJEB24932	This study
UFMG-CM-Y228	Cachaça	São Paulo, Brazil	Bread	53617	MD		4/5	17/19	NF		A822ins	NF	11bp del				F	F	W					PRJEB24932	This study
TUM 208	Alt beer	Rhineland-Palatinate, Germany	Beer 1 - German	33689					NF	V121M		NF	11bp del	792stop			NF	NF	C					ERS1108617	Gonçalves et al., 2016
TUM 513	California ale	Unknown	Beer 1 - German	33061					NF	V121M		NF	11bp del	792stop			NF	NF						ERS1108618	Gonçalves et al., 2016
TUM 308	Alt beer	Rhineland-Palatinate, Germany	Beer 1 - German	33689					NF	V121M		NF	11bp del	792stop			NF	NF						ERS1108619	Gonçalves et al., 2016
TUM 177	Alt / Kölsch beer	Krefeld, Germany	Beer 1 - German	20371					NF	V121M		NF	11bp del	792stop			NF	NF						ERS1108620	Gonçalves et al., 2016
TUM 174	Regional beer similar to Kölsch beer	Mülheim, Germany	Beer 1 - German	21227					NF	V121M		NF	11bp del	792stop			NF	NF						ERS1108621	Gonçalves et al., 2016
TUM 148	Alt beer	Düsseldorf, Germany	Beer 1 - German	32602					NF	V121M		NF	11bp del	792stop			NF	NF						ERS1108622	Gonçalves et al., 2016
TUM 338	Alt beer	Düsseldorf, Germany	Beer 1 - German	35423					NF	V121M		NF	11bp del	792stop			NF	NF						ERS1108623	Gonçalves et al., 2016
BE015	Beer	Germany	Beer 1 - German	n.d.					NF	V121M		NF	11bp del	792stop			NF	NF						MBZN00000000	Gallone et al., 2016
BE046	Beer	Bergium	Beer 1 - German	n.d.					NF	V121M		NF	11bp del	792stop			NF	NF						MBYI00000000	Gallone et al., 2016
BE073	Beer (Hefeweizen)	Germany	Beer 1 - German	n.d.					NF	V121M		NF	11bp del	792stop			NF	NF	C					MBXH00000000	Gallone et al., 2016

Strain	Substrate of isolation	Geographic location	Phylogeny	Heterozigous sites (Q >40)	MD / SCD	A	B	C	AQY1			AQY2			RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
BE075	Wheat beer	Belgium	Beer 1 - German	n.d.					NF	V121M	A822ins	NF	11bp del	792stop			F	F						MBXF00000000	Gallone et al., 2016
BE077	Beer (Trappist)	Belgium	Beer 1 - German	n.d.					NF	V121M	A822ins	NF	11bp del	792stop			F	F						MBXD00000000	Gallone et al., 2016
TUM 213	British ale / stout	Great Britain	Beer 1 - British	38169					NF	V121M	A822ins	F					NF	NF						ERS1108624	Gonçalves et al., 2016
TUM 510	British ale	Great Britain	Beer 1 - British	37397					NF	V121M	A822ins	F					NF	NF						ERS1108625	Gonçalves et al., 2016
TUM 503	California ale	Unknown	Beer 1 - British	30324					NF	V121M	A822ins	NF	11bp del				NF	NF						ERS1108626	Gonçalves et al., 2016
FostersO	Commercial brewing ale strain		Beer 1 - British	34874					NF	V121M	A822ins	NF	11bp del				NF	NF	C					ERS1108627	Gonçalves et al., 2016
TUM 165	British ale	Burton-upon-Trend, Great Britain	Beer 1 - British	28301					NF	V121M	A822ins	NF	11bp del				NF	NF						ERS1108628	Gonçalves et al., 2016
NCYC 1044	British ale		Beer 1 - British	43373					NF	V121M	A822ins	NF	11bp del				NF	NF						SRR403240	Dunn et al., 2012
FostersB	Commercial brewing ale strain		Beer 1 - British	43000					NF	V121M	A822ins	NF	11bp del				NF	NF	C					ERS1108629	Gonçalves et al., 2016
TUM 508	Irish ale / stout	Ireland	Beer 1 - British	44826					NF	V121M	A822ins	NF	11bp del				NF	NF						ERS1108630	Gonçalves et al., 2016
TUM 210	British ale / stout	Great Britain	Beer 1 - British	29225					NF	V121M	A822ins	NF	11bp del				NF	NF						ERS1108631	Gonçalves et al., 2016
TUM 211	British ale / stout	Great Britain	Beer 1 - British	26842					NF	V121M	A822ins	NF		792stop			NF	NF						ERS1108632	Gonçalves et al., 2016
BE065	Beer	US (California)	Beer 1 - British	n.d.					NF	V121M	A822ins	NF	11bp del	792stop			NF	NF						MBXP00000000	Gallone et al., 2016
BE071	Beer	US (San Diego)	Beer 1 - British	n.d.					NF	V121M	A822ins	NF	11bp del				NF	NF						MBXJ00000000	Gallone et al., 2016
BE055	Beer	England	Beer 1 - British	n.d.					NF	V121M	A822ins	F					NF	NF						MBXZ00000000	Gallone et al., 2016
BE048	Beer (Pale ale)	England	Beer 1 - British	n.d.					NF	V121M	A822ins	F					NF	NF						MBYG00000000	Gallone et al., 2016
TUM 381	Belgian beer	Belgium	Beer 1 - Wheat	30358					NF	V121M		F					F	F	C					ERS1108612	Gonçalves et al., 2016
TUM 149	Wheat beer	Munich, Germany	Beer 1 - Wheat	21254					NF	V121M		F					F	F						ERS1108613	Gonçalves et al., 2016
W 205	Hefeweizen ale yeast		Beer 1 - Wheat	31479					NF	V121M		F					F	F						SRR403241	Dunn et al., 2012
TUM 175	Wheat beer	Freising-Weihenstephan, Germany	Beer 1 - Wheat	29902					NF	V121M		F					F	NF						ERS1108615	Gonçalves et al., 2016
TUM 507	Ale Beer from wheatmalt	Unknown	Beer 1 - Wheat	40936					NF	V121M		F					NF	NF	C					ERS1108616	Gonçalves et al., 2016
BE074	Beer (Hefeweizen)	Germany	Beer 1 - Wheat	n.d.					NF	V121M		F					F	F						MBXG00000000	Gallone et al., 2016
BE093	Beer (Hefeweizen)	Germany	Beer 1 - Wheat	n.d.					NF	V121M		F					F	F						MBWN00000000	Gallone et al., 2016
BE072	Beer (Hefeweizen)	Germany	Beer 1 - Wheat	n.d.					NF	V121M		F					F	F						MBXI00000000	Gallone et al., 2016
CBS 435	Sake moto	Nakazawa, Japan	Sake	4906					NF	V121M		NF	G25del				F	NF						ERS1108637	Gonçalves et al., 2016

Strain	Substrate of isolation	Geographic location	Phylogeny	Heterozigous sites (Q >40)	MD / SCD	A	B	C	AQY1			AQY2			RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
CBS 1585	Sake moto	Nakazawa, Japan	Sake	2960					NF	V121M		NF	G25del				F	NF						PRJNA189638	Gonçalves et al., 2016
CBS 1598	Sake moto	Nakazawa, Japan	Sake	3517					NF	V121M		NF	G25del				F	NF						PRJNA189639	Gonçalves et al., 2016
UC 5	Sene sake	Kurashi, Japan	Sake	n.d.					NF	V121M		NF	G25del				F	NF	C					PRJNA60197	Justin Fay, Washington University
TUM 184	Alt beer	Düsseldorf, Germany	Sake	4558					NF	V121M		NF	G25del				F	NF						PRJNA189640	Gonçalves et al., 2016
Kyokai-no.7	Japanese sake brewerie	Japan	Sake	n.d.					NF	V121M	A822ins	NF	G25del			TRIP	F	NF	C					PRJNA45827	Akao et al., 2011
NRRL-Y-11572	Sake moto	Japan	Sake	3270					NF	V121M		NF	G25del				F	NF						PRJNA189641	Gonçalves et al., 2016
Y9	Fermentation (ragi)	Indonesia	Sake	2932					NF	V121M		NF	G25del				F	NF						PRJNA60205	Justin Fay, Washington University
YJSH1	Bioethanol-producing strain	China	Sake	3084					NF	V121M		NF	G25del				F	NF						PRJNA72403	Zheng et al., 2012
TUM 380	Starter culture for Lambic beer	Belgium	Sake	6522					NF	V121M		NF	G25del				F	F						PRJNA189642	Gonçalves et al., 2016
TUM 68	Wheat beer	Freising-Weihenstephan, Germany	Sake	7094					NF	V121M		NF	G25del				F	F						PRJNA189643	Gonçalves et al., 2016
TUM 127	Wheat beer	Freising-Weihenstephan, Germany (progeny of TUM 68)	Sake	5979					NF	V121M		NF	G25del				F	F						PRJNA189644	Gonçalves et al., 2016
NRRL-Y-12844	Budod	Phillipines	Sake	10363					NF	V121M		NF	G25del				F	NF						PRJNA189645	Gonçalves et al., 2016
Y12	Fermentation (palm wine)	Ivory Coast	Sake	15534					NF	V121M		NF	G25del											SGRP2	Bergström et al., 2014
YJM 1479	Coconut tuba (palm wine)	Philippines	Philippines	3264					n.d.			n.d.			n.d.	n.d.	n.d.	n.d.						PRJNA189929	Strope et al., 2015
YJM 1400	Guava (fruit)	Philippines	Philippines	2960					NF	V121M		NF	G25del				F	F	C					PRJNA189911	Strope et al., 2015
YJM 1401	Papaya (fruit)	Philippines	Philippines	2762					n.d.			n.d.			n.d.	n.d.	n.d.	n.d.						PRJNA189912	Strope et al., 2015
NRRL-YB-4084	Coconut sap	Phillipines	Philippines	10363					NF	V121M		F					F	F	C					ERS1108646	Gonçalves et al., 2016
ZP 779	*Quercus acutissima*	Hirusen highland, Okayama Prefecture, Japan	NA / Japan	3509					F			F					F	F						PRJEB7601	Almeida et al., 2015
ZP 781	*Quercus serrata*	Hirusen highland, Okayama Prefecture, Japan	NA / Japan	3333					F			F					F	F						PRJEB7601	Almeida et al., 2015
ZP 785	*Quercus dentata*	Hirusen highland, Okayama Prefecture, Japan	NA / Japan	3510					F			F					F	F						PRJEB7601	Almeida et al., 2015
SDO 3s1	Oak	North Carolina, USA	NA / Japan	3693					F			F					F	F						PRJEB7601	Almeida et al., 2015
YPS 1009	Oak exudate	New Jersey, USA	NA / Japan	2650					n.d.			n.d.			n.d.	n.d.	n.d.	n.d.						PRJNA60223	Justin Fay, Washington University
ZP 1050	*Quercus ilex*	Vendinha, Reguengos de Monsaraz, Beja, Portugal	NA / Japan	3119					F			F					F	F						PRJEB7675	Almeida et al., 2015

Strain	Substrate of isolation	Geographic location	Phylogeny	Het-erozi-gous sites (Q >40)	MD / SCD	A	B	C	AQY1		AQY2		RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
YPS 128	*Quercus alba*	Pennsylvania, USA	NA / Japan	3118					F		F				F	F	C					SGRP2	Bergström et al., 2014
ZP 530	*Castanea sativa*	Marão, Campeã, Portugal	NA / Japan	3661					F		F				F	F						PRJEB7601	Almeida et al., 2015
YPS 163	*Quercus rubra*	Pennsylvania, USA	NA / Japan	2673					F		F				F	F						PRJNA28813	Doniger et al., 2008
ZP 656	*Quercus acuta*	Chiba Prefecture, Japan	NA / Japan	3723					F		F				F	F	C					PRJEB7601	Almeida et al., 2015
ZP 651	*Quercus acutissima*	Chiba Prefecture, Japan	NA / Japan	3623					F		F				F	F						PRJEB7601	Almeida et al., 2015
ZP 652	*Quercus acutissima*	Chiba Prefecture, Japan	NA / Japan	3699					F		F				F	F	C					PRJEB7601	Almeida et al., 2015
UFMG-CM-Y255	Non-identified tree	Minas Gerais, Brazil	NA / Japan	5103					F		F				F	F						PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y640	Soil	Amazonia, Brazil	NA / Japan	5426					F		F				F	F						PRJEB11698	Barbosa et al., 2016
YJM 1447	Bertram palm	Malaysia	Malaysia	n.d.					NF	106bpdel	NF	G528del			F	F						PRJNA189923	Strope et al., 2015
UWOPS 03-433.3	Nectar of Bertram palm	Malaysia	Malaysia	1908					NF	106bpdel	NF	G528del			n.d.	n.d.	C					ERS1108647	Gonçalves et al., 2016
UWOPS 03-461.4	Nectar of Bertram palm	Malaysia	Malaysia	18636					NF	106bpdel	NF	G528del			F	F	C					SGRP2	Bergström et al., 2014
YJM 195	Palm wine	Nigeria	West Africa	2977					NF	A881del	F				F	F						PRJNA189849	Strope et al., 2015
PW5	Raphia palm wine	Aba, Abia state, Nigeria	West Africa	1175					NF	A881del	F				NF	F	C					PRJNA60181	Justin Fay, Washington University
YJM 1439	Ginger beer from *Z. officinale*	West Africa	West Africa	2129					F		F				NF	F	C					PRJNA189920	Strope et al., 2015
YJM 1248	Bili wine from *Osbeckia grandiflora*	West Africa	West Africa	2600					NF	A881del	F				NF	F	C					PRJNA189888	Strope et al., 2015
UFMG-CM-Y643	*Quercus rubra*	Minas Gerais, Brazil	B1	6562					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y262	*Tapirira guianensis*	Tocantins, Brazil	B1	7501					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y649	*Tapirira guianensis*	Tocantins, Brazil	B1	6059					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y650	*Tapirira guianensis*	Tocantins, Brazil	B1	4749					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y463	Non-identified tree	Roraima, Brazil	B1	6665					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y464	Non-identified tree	Roraima, Brazil	B1	7034					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y457	Non-identified tree	Roraima, Brazil	B1	6856					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y461	Mushroom	Roraima, Brazil	B1	7540					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y462	Mushroom	Roraima, Brazil	B1	6917					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y644	Non-identified tree	Roraima, Brazil	B1	6988					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y646	*Tapirira guianensis*	Tocantins, Brazil	B1	8246					F		F				F	F	C					PRJEB11698	Barbosa et al., 2016

Strain	Substrate of isolation	Geographic location	Phylogeny	Het-erozi-gous sites (Q >40)	MD / SCD	A	B	C	AQY1	AQY2	RTM1	BIO1 BIO6	PAD1	FDC1	FZF1	FZF1 Fig 4	FZF1 Fig S2	SSU1	SSU1R	Genome data	Reference
UFMG-CM-Y647	*Tapirira guianensis*	Tocantins, Brazil	B1	8165					F	F			F	F	C					PRJEB11698	Barbosa et al., 2016
YPS 1000	Oak exudate	New Jersey, USA	B1	5103					F	F			F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y257	*Quercus rubra*	Minas Gerais, Brazil	B1	WINE					F	F			F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y645	*Tapirira guianensis*	Tocantins, Brazil	B1	11245					F	F			F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y266	*Tapirira guianensis*	Tocantins, Brazil	B1	8175					F	F			F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y455	*Tapirira guianensis*	Tocantins, Brazil	B1	8263					F	F			F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y264	*Tapirira guianensis*	Tocantins, Brazil	B1	8271					F	F			F	F	C					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y648	Cachaça	Tocantins, Brazil	B1	7267	MD				F	F			F	F	C					PRJEB24932	Barbosa et al., 2016
UFMG-CM-Y263	*Tapirira guianensis*	Tocantins, Brazil	B3	7580					F	F			F	F	W					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y456	Jequitibá tree	Tocantins, Brazil	B3	46044					F	F			F	F	C NA					PRJEB11698	This study
UFMG-CM-Y641	*Tapirira guianensis*	Tocantins, Brazil	B3	5385					F	F			F	F	R					PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y642	*Tapirira guianensis*	Tocantins, Brazil	B3	4749					F	F			F	F	R					PRJEB11698	Barbosa et al., 2016

Cachaça
Wine - Brazil
Bioethanol
Wild - Brazil
Brazilian strain

> 20000

* values taken from the literature (see citation)

F - functional
NF - non functional
n.d. - not determined

Sake
Spar NA
Spar allele only
Scer+Spar alleles
Cosmopolitan
Wine
Recombinant

detected
not detected

Table A1.2 - Tolerance to ferulic acid (0.2% w/v), pH 4.5) of Cachaça strains with and without *PAD1* / *FDC1* introgression from *S. paradoxus*. Growth was assessed in Yeast-Peptone-Glucose agar medium supplemented with ferulic acid. Strong (+++), intermediate (++) or low (+) growth was recorded after inoculation of serially diluted (10^{-2}, 10-3 and 10^{-4}) 10 µl drops of a culture grown in liquid medium without ferulic acid up to OD_{640nm} of 0.3 - 0.4 and 48 hours incubation at 25 °C. For comparison the results of two wild Brazilian strains with high resistance to ferulic acid are included.

Strain	***PAD1* / *FDC1* *S. paradoxus* introgressions**	**Phylogeny**	**Tolerance to ferulic acid**
CAY 1007	present	cachaça	++
CAY 1834	present	cachaça	+
CAY 2170	present	cachaça	+
RJ1	present	cachaça	++
TO1	present	cachaça	++
TOC1301	present	cachaça	+
TOC1346	present	cachaça	++
CAY44	present	cachaça	+
RJW03	absent	cachaça	+
SC1	absent	cachaça	++
SC3	absent	cachaça	+
SC7	absent	cachaça	+
T09	absent	cachaça	+
YEF036	absent	cachaça	+
YEF034	absent	cachaça	++
RJ15	absent	cachaça	+
49-30	absent	wild Brazil B1	+++
TOC 1345	absent	wild Brazil B3	+++

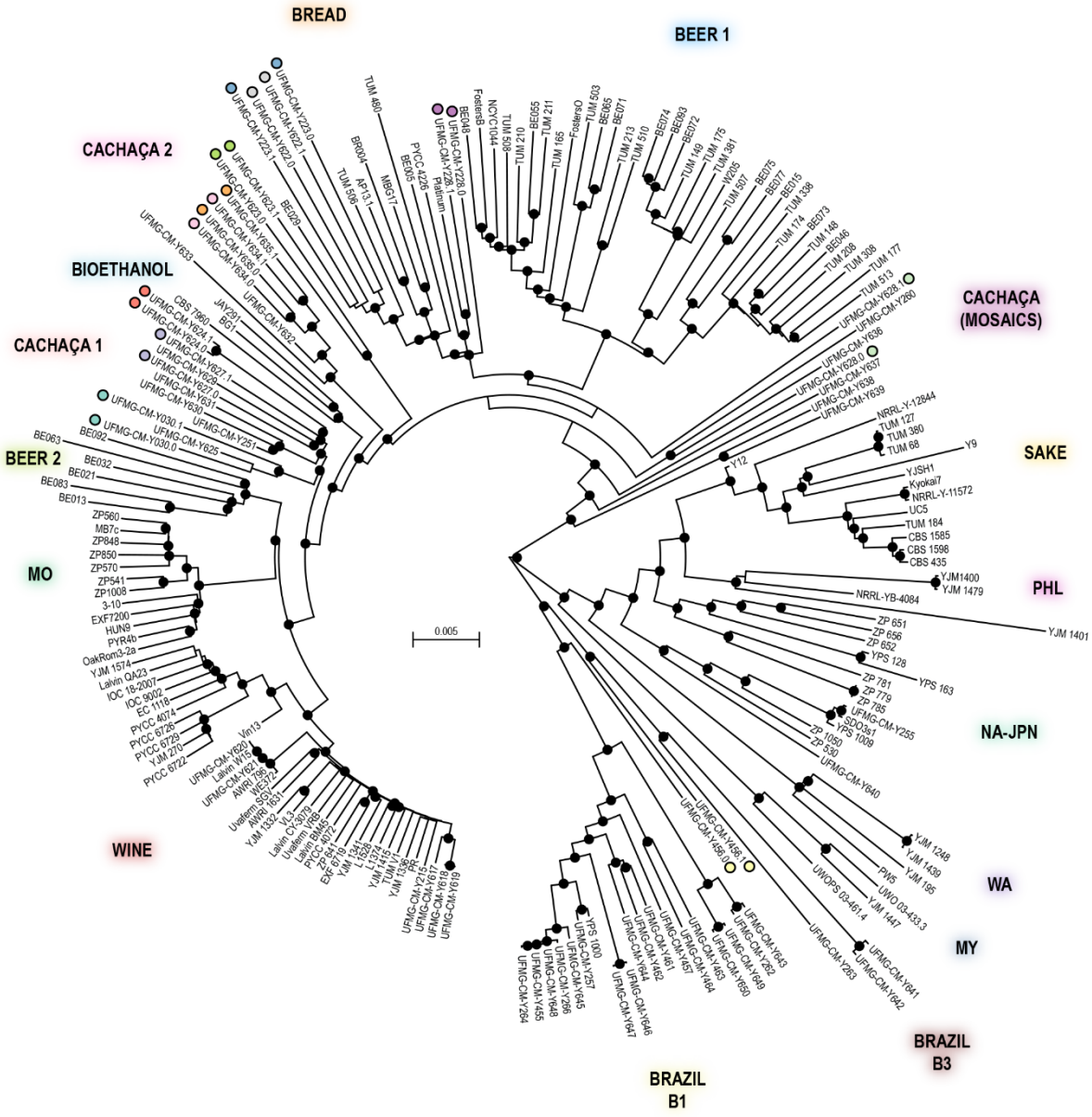


Figure A1.1 - Phylogenetic placement of phased haplotypes of Cachaça strains using the same dataset and the same methodology as in Figure 2.1. The tree was rooted with *S. paradoxus*, branch lengths correspond to the expected number of substitutions per site and black dots depict bootstrap support values above 90% (100 replicates). Phased haplotypes are highlighted with a circle colored with the same color.

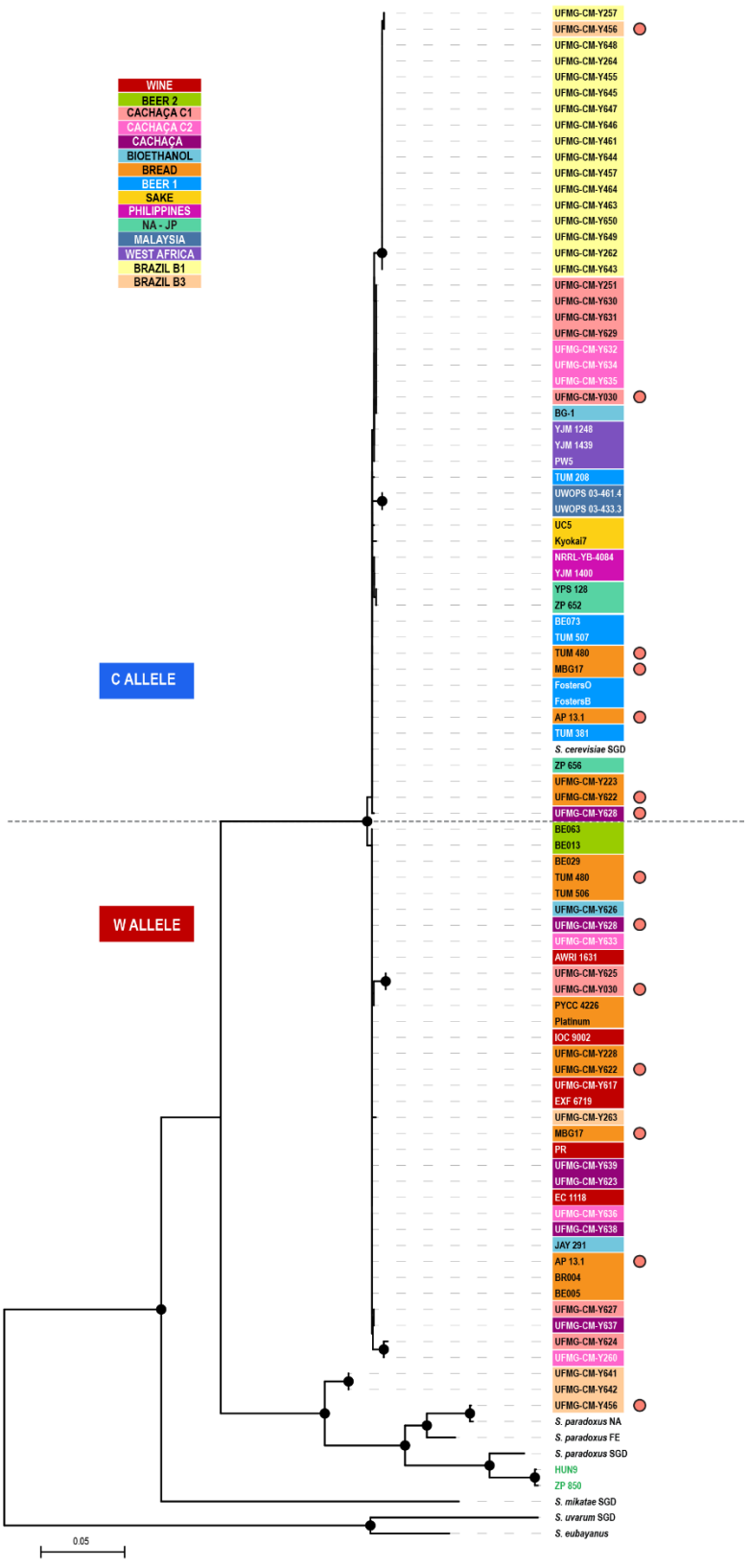


Figure A1.2 - Comprehensive *FZF1* phylogeny in *S. cerevisiae*. Phylogenetic tree based on the alignment of 90 strains and constructed with the neighbor-joining method using the Tamura's 3-parameter model and with bootstrap values > 90% indicated by black circles. The two alleles C (cosmopolitan), and W (wine) are highlighted, and strain designations are color-coded according to their phylogenetic placement. Strains harboring the two alleles are highlighted with a colored dot.

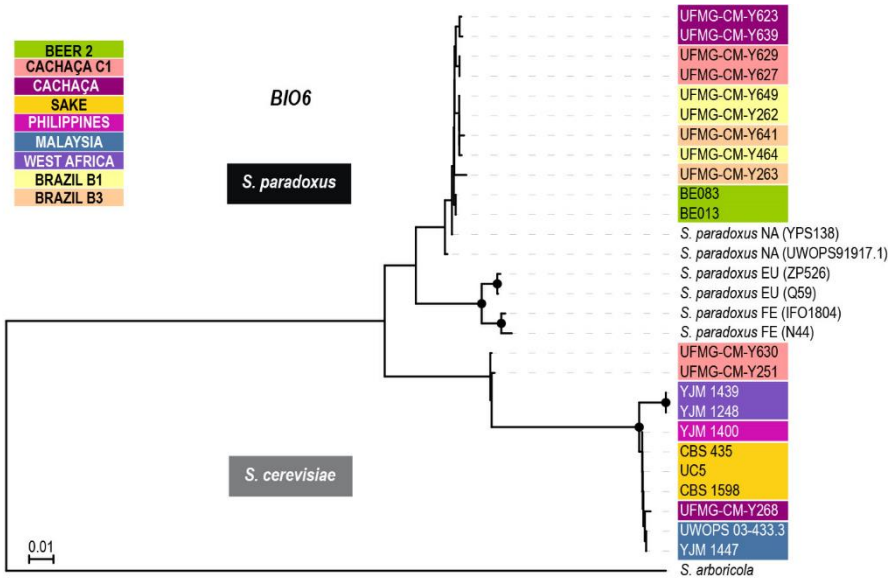


Figure A1.3 - Phylogeny of *BIO6* sequences of *Saccharomyces* spp.. Phylogenetic tree based on an alignment of 29 sequences and constructed with the neighbor-joining method using the Tamura's 3-parameter model and with bootstrap values > 90% indicated by black circles. Strain designations are color-coded according to their phylogenetic placement.

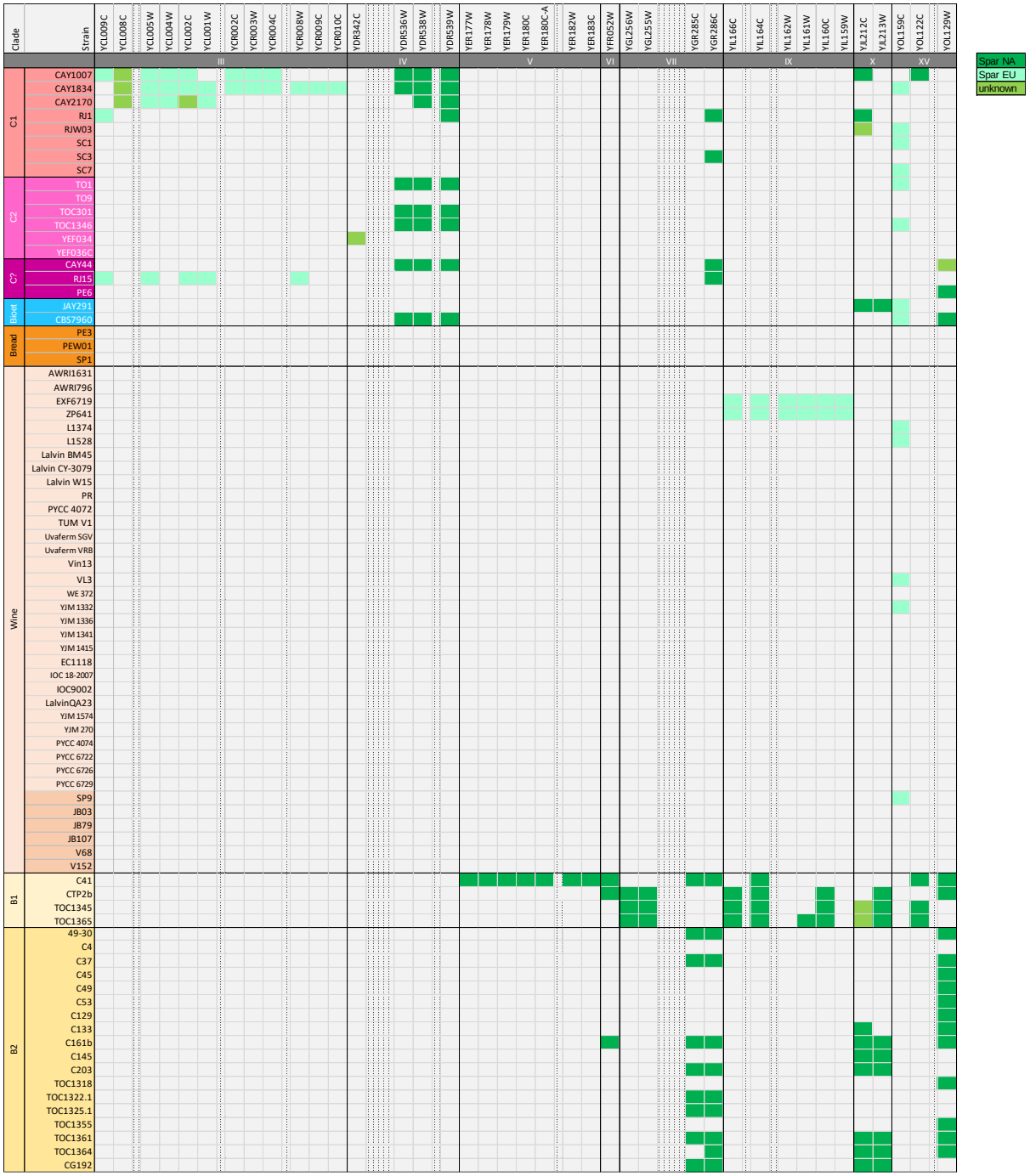


Figure A1.4 - Introgressions of *Saccharomyces paradoxus* in the genomes of Brazilian *S. cerevisiae* Cachaça and wild populations and in Bioethanol and Wine strains (depicted strain by strain). The presence of introgressions is color-coded and refers to the donor population of *S. paradoxus*: European (light green), North American population B (dark green), unknown population or recombinant sequences (brown).

A2
Appendix 2

Supplementary information pertaining to Chapter 3

Pontes, A.; Čadež, N.; Gonçalves, P.; Sampaio, J.P. A Quasi-Domesticate Relic Hybrid Population of *Saccharomyces cerevisiae* x *S. paradoxus* Adapted to Olive Brine. Frontiers in Genetics 2019, 10, 449.

https://doi.org/10.3389/fgene.2019.00449

Table A2.1 - Strains and genomes used in the work presented in chapter 3 and relevant information pertaining to them.

Strain	Other designations	Substrate	Origin	WGS group	Heterozygous sites (Q > 40)	Sporulation	Regions			*AQY1*	*AQY2*	CNV			Genome data	Reference
							A	B	C			*ACT1*	*CUP1-1*	*CUP1-2*		
AWRI 1631	N 96	Wine	Australia	WINE	n.d.	n.d.			8/19	NF	NF	n.d.	n.d.	n.d.	PRJNA30553	Borneman et al., 2008
AWRI 796	Active Dry Wine Yeast	Commercial wine yeast	South Africa	WINE	1041 *	n.d.		DUP		F	NF	n.d.	n.d.	n.d.	PRJNA48559	Borneman et al., 2011
EXF 6719		Wine must	Cuber, Slovenia	WINE	2053	+			17/19	NF	NF	2	1	1	PRJEB7601	Almeida et al., 2015
ZP 641		Spontaneous red wine fermentation	Castelo de Vide, Portugal	WINE	1484	+				NF	NF	2	4	3	PRJEB7601	Almeida et al., 2015
L1374		Wine (must from País variety)	Cauquenes, Chile	WINE	1550	n.d.		5/5		NF	NF	2	2	1	SGRP2	Bergström et al., 2014
L1528		Wine (must from Cabernet variety)	Cauquenes, Chile	WINE	1334	n.d.		5/5		NF	NF	2	16	17	SGRP2	Bergström et al., 2014
Lalvin BM45	AWRI 1486	Commercial yeast (white wine)		WINE	1280	+		5/5	2/19	NF	NF	2	2	2	PRJEB19382	Almeida et al., 2017
Lalvin CY-3079	AWRI 2078	Commercial yeast (red wine)		WINE	1974	+		5/5	4/19	NF	NF	2	3	2	PRJEB19382	Almeida et al., 2017
Lalvin W15		Commercial yeast (white and red wine)		WINE	2209	+		5/5		F	NF	2	10	14	PRJEB19382	Almeida et al., 2017
PR		Pasteur Red		WINE	1715	+				NF	NF	2	2	2	PRJEB19382	Almeida et al., 2017
PYCC 4072		Commercial wine yeast (Fermivin)	Portugal	WINE	1889	+				NF	NF	2	2	3	PRJEB19382	Almeida et al., 2017
TUM V1		Bordeaux wine		WINE	1827	+				NF	NF	2	1	1	PRJEB19382	Almeida et al., 2017
Uvaferm SGV		Commercial yeast (red wine)		WINE	1734	+			4/19	NF	NF	2	1	2	PRJEB19382	Almeida et al., 2017
Uvaferm VRB		Commercial yeast (red wine)		WINE	2750	+			19/19	NF	NF	2	5	5	PRJEB19382	Almeida et al., 2017
Vin 13	AWRI 1537	Commercial wine yeast		WINE	15216 *	n.d.		DUP	2/19	F	NF	n.d.	n.d.	n.d.	PRJNA48563	Borneman et al., 2011
VL3	B 6, AWRI 1688	Commercial wine yeast		WINE	9904 *	n.d.		5/5	8/19	NF	NF	n.d.	n.d.	n.d.	PRJNA48565	Borneman et al., 2011
WE 372		Commercial wine yeast	Cape Town, South Africa	WINE	4354	n.d.		5/5		F	NF	2	8	10	PRJNA60199	Justin Fay, Washington University
YJM 1332		Wine	Italy	WINE	1660	n.d.			2/19	NF	NF	3	2	2	PRJNA189896	Strope et al., 2015
YJM 1336		Wine	Italy	WINE	1547	n.d.				NF	NF	n.d.	n.d.	n.d.	PRJNA189897	Strope et al., 2015
YJM 1341	NRRL Y-12637	Grape must	South Africa	WINE	2350	n.d.				NF	NF	2	4	4	PRJNA189899	Strope et al., 2015
YJM 1415	NRRL Y-268	Wine	France	WINE	1256	n.d.		5/5	19/19	NF	NF	2	1	1	PRJNA189914	Strope et al., 2015
S8K		Bark of olive tree (variety Leccino) close to the soil	Kolomban, Slovenia	WINE	919	+		5/5	5/19	NF	NF	2	17	15	PRJEB30431	This study
YO 392		Table olive	Purchased in Tecate, Mexico	WINE	1859	+				NF	NF	2	3	3	PRJEB30431	This study
SS8		Extra virgin olive oil with macerated perforate St John's-wort (*Hypericum perforatum*)	Bozava, Dugi otok, Croatia	WINE	2565	-		5/5	5/19	NF	NF	2	19	11	PRJEB30431	This study
S21		Bark of olive tree variety, sampled close to the soil	Kolomban, Slovenia	WINE	3101	-				NF	NF	2	7	5	PRJEB30431	This study

Strain	Other designations	Substrate	Origin	WGS group	Heterozygous sites (Q > 40)	Sporulation	Regions			*AQY1*	*AQY2*	CNV			Genome data	Reference
							A	B	C			*ACT1*	*CUP1-1*	*CUP1-2*		
EC 1118	Prise de Mousse	Industrial strain isolated from Champagne	France	WINE	n.d.	n.d.	15/15	5/5	19/19	NF	NF	2	2	2	PRJEA37863	Novo et al., 2009
IOC 18-2007	AWRI 2340	Commercial yeast		WINE	2181	+		5/5		NF	F	2	6	5	PRJEB19382	Almeida et al., 2017
IOC 9002		Commercial wine yeast		WINE	1224	-		5/5	4/19	NF	NF	2	11	6	PRJNA264372	Almeida et al., 2015
Lalvin QA23		Commercial wine yeast	Portugal	WINE	18861 *	+	15/15	TRIP	19/19	NF	F	2	2	2	PRJNA48561	Borneman et al., 2011
YJM 1574	AWRI 1775	Wine		WINE	1796	n.d.		5/5		NF	NF	3	7	7	PRJNA189934	Strope et al., 2015
YJM 270	CBS 2807	Wine	Slovenia	WINE	1662	n.d.			19/19	NF	F	2	1	1	PRJNA189852	Strope et al., 2015
PYCC 4074		Commercial wine yeast (Fermichamp)	Portugal	WINE	2263	+	15/15	5/5		NF	F	2	6	5	PRJEB19382	Almeida et al., 2017
PYCC 6722	CBS 5155	Wine	South Armenia	WINE	2121	+				NF	F	2	1	1	PRJEB19382	Almeida et al., 2017
PYCC 6726		Jerez-wine	Spain	WINE	4939	+				NF	F	2	1	1	PRJEB19382	Almeida et al., 2017
PYCC 6729		Jerez-wine	Armenia	WINE	2901	+				NF	F	2	1	1	PRJEB19382	Almeida et al., 2017
YO 652	Kalamata 1	Table olives (variety Kalamata)	Purchased in Seattle, Washington, USA	OLIVES	2851	+				NF	NF	2	6	6	PRJEB30431	This study
YO 653	Kalamata 2	Table olives (variety Kalamata)	Purchased in Seattle, Washington, USA	OLIVES	3342	+				NF	NF	2	4	4	PRJEB30431	This study
YO 654	Kalamata 3	Table olive brine (variety Kalamata)	Purchased in Seattle, Washington, USA	OLIVES	3466	+				NF	NF	2	8	7	PRJEB30431	This study
PYCC 4935		Table olive brine	Portalegre, Portugal	OLIVES	1713	+				NF	NF	2	1	1	PRJEB30431	This study
PYCC 4891		Table olive brine	Elvas, Portugal	OLIVES	10653	-				NF	NF	2	2	2	PRJEB30431	This study
PYCC 6730		Table olives	Spain	OLIVES	2965	+		5/5		NF	NF	2	2	1	PRJEB30431	This study
PYCC 6731		Alpechin	Spain	OLIVES	1393	+		4/5		NF	NF	2	1	1	PRJEB30431	This study
PYCC 6732	NRRL Y-6679; CBS 3081; YJM 1252	Alpechin	Spain	OLIVES	1886	-				NF	NF	2	1	1	PRJEB30431	This study
PYCC 6733		Alpechin	Spain	OLIVES	5258	-				NF	NF	2	1	1	PRJEB30431	This study
CBS 7002	PYCC 8023	Alpechin	Sevilla, Spain	OLIVES	2471	+				NF	NF	2	4	3	PRJEB30431	This study
A8		Table olive brine	Purchased in Oeiras, Portugal	OLIVES	16637	-				NF	NF	2	3	2	PRJEB30431	This study
AP 5.4		Table olive brine	Purchased in Lisbon, Portugal	OLIVES	2914	+				NF	NF	2	7	8	PRJEB30431	This study
AP 6.2		Table olive brine	Purchased in Lisbon, Portugal	OLIVES	2031	+				NF	NF	1	12	10	PRJEB30431	This study
AP 7.2		Table olive brine	Purchased in Lisbon, Portugal	OLIVES	3116	+				NF	NF	2	4	3	PRJEB30431	This study
AP 16.1		Table olive brine	Douro, Portugal	OLIVES	2618	+				NF	NF	2	20	15	PRJEB30431	This study
AP 17.1		Table olive brine	Douro, Portugal	OLIVES	2961	+				NF	NF	2	3	3	PRJEB30431	This study
MM 11		Sediment of olive oil	Kolomban, Slovenia	OLIVES	3153	-				NF	NF	2	1	1	PRJEB30431	This study

Strain	Other designations	Substrate	Origin	WGS group	Heterozygous sites (Q > 40)	Sporulation	Regions			*AQY1*	*AQY2*	CNV			Genome data	Reference
							A	B	C			*ACT1*	*CUP1-1*	*CUP1-2*		
ZIM 2221		Olive oil (variety Oblica)	Split, Croatia	OLIVES	14951	-		5/5		NF	NF	2	8	10	PRJEB30431	This study
ZIM 2580		Sediment of olive oil	Kolomban, Slovenia	OLIVES	3117	+				NF	NF	2	1	1	PRJEB30431	This study
S27		Ripe olive (variety Belica)	Kolomban, Slovenia	OLIVES	3217	-				NF	NF	2	3	2	PRJEB30431	This study
PYCC 2708		Swine rectum	Portugal	OLIVES	3559	+				NF	NF	2	7	8	PRJEB30431	This study
PYCC 2613	CBS 2909; NRRL YB-6041	Human feces	Portugal	OLIVES	6783	+				NF	NF	2	2	4	PRJEB30431	This study
NRRL-Y 12658	CBS 4411; PYCC 8033	Pig rectal contents	unknown	OLIVES	2850	+				NF	NF	2	17	16	PRJEB30431	This study
YJM 248	NRRL Y-12659; CBS 2910; PYCC 8034	Human feces	Portugal	OLIVES	6185	n.d.				NF	NF	2	5	5	SRR800854	Strope et al., 2015
YJM 1078	NRRL YB-4348; PYCC 2625; PYCC 8028	Human feces	Portugal	OLIVES	2752	n.d.				NF	NF	3	6	5	SRR800768	Strope et al., 2015
PYR 4b	DBQ 26	*Quercus pubescens*	Halkidiki, Greece	MO	2738	n.d.				F	F	2	1	1	PRJEB7601	Almeida et al., 2015
EXF 7200		*Quercus robur*	Jasenovo Polje, Montenegro	MO	3001	n.d.				F	F	2	1	1	PRJNA264372	Almeida et al., 2015
HUN 9.1s1	DBS 14	Oak	Hungary	MO	2745	n.d.				F	F	2	1	1	PRJEB7601	Almeida et al., 2015
ZP 541		*Fagus sylvatica*	Adagoi, Portugal	MO	2684	+				F	F	2	1	1	PRJEB7601	Almeida et al., 2015
ZP 570		*Fraxinus* sp.	Paul Boquilobo, Portugal	MO	2813	+				F	F	2	1	1	PRJEB7601	Almeida et al., 2015
ZP 848		*Quercus ilex*	Alter do Chão, Portugal	MO	2656	n.d.				F	F	3	1	1	PRJEB7675	Almeida et al., 2015
PYCC 4226		Commercial baker's yeast		BREAD	30081	-		4/5	2/19	NF	NF	2	1	2	ERS1108635	Gonçalves et al., 2016
Platinum		Commercial baker's yeast		BREAD	42278	-		4/5	15/19	NF	NF	2	1	1	ERS1108633	Gonçalves et al., 2016
AP 13.1		Baker's yeast	Portugal	BREAD	53091	-		5/5	17/19	NF	NF	2	2	2	PRJEB24932	Barbosa et al., 2018
TUM 381		Belgian beer	Belgium	BEER 1 - WHEAT	30358	-				NF	F				ERS1108612	Gonçalves et al., 2016
TUM 508		Irish ale / stout	Ireland	BEER 1 - BRITISH	44826	-				NF	NF				ERS1108630	Gonçalves et al., 2016
TUM 208		Alt beer	Rhineland-Palatinate, Germany	BEER 1 - GERMAN	33689	-				NF	NF				ERS1108617	Gonçalves et al., 2016
CBS 1598		Sake moto	Nakazawa, Japan	SAKE	3517	-				NF	NF	2	3	1	ERS1108639	Gonçalves et al., 2016
UC5	UCD612	Sene sake	Kurashi, Japan	SAKE	n.d.	n.d.				NF	NF				PRJNA60197	Justin Fay, Washington University
Kyokai7		Sake yeast	Japan	SAKE	n.d.	n.d.				NF	NF				PRJNA45827	Akao et al., 2011
A38		Olive collected from olive tree	Alqueva, Portugal	SAKE	2747	-				NF	NF	1	4	4	PRJEB30431	This study
TUM 127	progeny of TUM 68	Wheat beer	Freising-Weihenstephan, Germany	SAKE	5979	-				NF	NF				ERS1108644	Gonçalves et al., 2016
NRRL-YB-4084		Coconut sap	Philippines	PHILIPPINES	10363	-				NF	F				ERS1108646	Gonçalves et al., 2016

Strain	Other designations	Substrate	Origin	WGS group	Heterozygous sites (Q > 40)	Sporulation	Regions			*AQY1*	*AQY2*	CNV			Genome data	Reference
							A	B	C			*ACT1*	*CUP1-1*	*CUP1-2*		
YJM 1400	NRRL YB-4081	Guava (fruit)	Philippines	PHILIPPINES	2960	n.d.				NF	NF				PRJNA189911	Strope et al., 2015
ZP 779		*Quercus acutissima*	Okayama prefecture, Japan	NA & JP	3509	+				F	F				PRJEB7601	Almeida et al., 2015
ZP 781		*Quercus serrata*	Okayama prefecture, Japan	NA & JP	3333	+				F	F				PRJEB7601	Almeida et al., 2015
YPS 128		*Quercus alba*	Pennsylvania, USA	NA & JP	3118	n.d.				F	F				SGRP2	Bergström et al., 2014
YPS 163		*Quercus rubra*	Pennsylvania, USA	NA & JP	2673	n.d.				F	F				PRJNA28813	Doniger et al., 2008
ZP 651		*Quercus acutissima*	Chiba prefecture, Japan	NA & JP	3623	+				F	F				PRJEB7601	Almeida et al., 2015
ZP 656		*Quercus acuta*	Chiba prefecture, Japan	NA & JP	3723	+				F	F				PRJEB7601	Almeida et al., 2015
PW5		Raphia palm wine	Aba, Abia state, Nigeria	WEST AFRICA	1175	n.d.				NF	F				PRJNA60181	Justin Fay, Washington University
YJM 1248	NRRL Y-1546	Bili wine from *Osbeckia grandiflora*	West Africa	WEST AFRICA	2600	n.d.				NF	F				PRJNA189888	Strope et al., 2015
YJM 1439	NCYC 110	Ginger beer from *Z. officinale*	West Africa	WEST AFRICA	2129	n.d.				NF	F				PRJNA189920	Strope et al., 2015
UWOPS 03-461-4		Nectar of Bertram palm	Malaysia	MALAYSIA	18636	n.d.				NF	NF				SGRP2	Bergström et al., 2014
UWOPS 03-433-3		Nectar of Bertram palm	Malaysia	MALAYSIA	1908	+				--	--				ERS1108647	Gonçalves et al., 2016
YJM 1447	UWOPS 05-227.2	Bertram palm	Malaysia	MALAYSIA	n.d.	n.d.				NF	NF	n.d.	n.d.	n.d.	PRJNA189923	Strope et al., 2015
A44		Olive collected from olive tree	Alqueva, Portugal	MOSAIC	3768	+				NF	NF	2	9	6	PRJEB30431	This study
A5		Olive collected from olive tree (variety Gaeta)	Mação, Portugal	MOSAIC	4551	+				F	NF	2	2	1	PRJEB30431	This study

* values taken from the literature (see reference)

MO, Mediterranean oaks

NA & JP, North America and Japan

F Functional

NF Non Functional

Table A2.2 - Gene ontology (GO) analysis of the *S. paradoxus* sub-genome of relic-hybrid strains.

		OLIVE BRINE									ALPECHIN / OLIVE OIL			OLIVE TREE				INTESTINAL TRACT							
GO TERM	**GENE ONTOLOGY TERM (P-VALUE < 0.01)**	PYCC 4935	PYCC 4891	A8	AP 5.4	AP 6.2	AP 7.2	AP 16.1	AP 17.1	YO 652	YO 653	YO 654	PYCC 6730	PYCC 6731	PYCC 6732	PYCC 6733	MM 11	ZIM 2221	ZIM 2580	S27	PYCC 2613	PYCC 2708	YJM 248	YJM 1078	SHARED
PROCESS	Fungal-type cell wall organization or biogenesis																								
	Fungal-type cell wall organization																								
	External encapsulating structure organization																								
	Cell wall organization																								
	Cell wall organization or biogenesis																								
	Carbohydrate transmembrane transport																								
	Carbohydrate transport																								
	Glucose transport																								
	Fructose transport																								
	Mannose transport																								
	Monosaccharide transport																								
	Monosaccharide transmembrane transport																								
	Glucose transmembrane transport																								
	Hexose transmembrane transport																								
	Hexose transport																								
FUNCTION	Glucose transmembrane transporter activity																								
	Monosaccharide transmembrane transporter activity																								
	Hexose transmembrane transporter activity																								
	Carbohydrate transmembrane transporter activity																								
	Mannose transmembrane transporter activity																								
	Fructose transmembrane transporter activity																								
	Glucose transmembrane transporter activity																								
	Sugar transmembrane transporter activity																								
	Symporter activity																								
	Solute proton symporter activity																								
	Solute cation symporter activity																								
	Sugar proton symporter activity																								
	Cation sugar symporter activity																								
	Structural constituent of cell wall																								
	Transporter activity																								
	Transmembrane transporter activity																								
COMPONENT	External encapsulating structure																								
	Cell wall																								
	Fungal-type cell wall																								
	Cell periphery																								
	Plasma membrane																								
	Plasma membrane part																								
	Membrane part																								
	Cell part																								
	Cell																								
	Integral component of plasma membrane																								
	Intrinsic component of plasma membrane																								
	Intrinsic component of membrane																								
	Anchored component of membrane																								
	Extracellular region																								
	Membrane																								
	Number of Genes Analysed	251	254	314	288	246	237	249	275	193	216	282	227	241	271	259	269	269	270	236	268	250	259	244	103

Table A2.3 - Cfu/ml counts per strain and per replicate for the growth and survival in olive brine experiment shown in Figure 3.2.

Strain	Replicate	Days								
		0	**8**	**15**	**24**	**34**	**41**	**49**	**57**	**70**
YO 654	1	9.8×10^4	3.1×10^6	4.1×10^6	3.2×10^6	3.3×10^6	3.0×10^6	3.5×10^6	1.9×10^6	3.2×10^6
	2	9.6×10^4	3.1×10^6	3.9×10^6	3.9×10^6	3.4×10^6	2.5×10^6	3.2×10^6	2.8×10^6	3.7×10^6
ZIM 2580	1	9.8×10^4	4.2×10^6	2.1×10^6	1.3×10^6	8.1×10^5	6.7×10^5	5.6×10^5	9.6×10^5	7.5×10^5
	2	7.5×10^4	1.8×10^6	1.96×10^6	1.5×10^6	7.3×10^5	1.3×10^6	8.2×10^5	8.0×10^5	7.1×10^5
AP 7.2	1	1.1×10^5	8.1×10^6	5.0×10^6	3.9×10^6	2.8×10^6	6.4×10^6	4.2×10^6	5.0×10^6	3.9×10^6
	2	1.1×10^5	3.5×10^6	3.4×10^6	5.3×10^6	5.2×10^6	3.8×10^6	4.2×10^6	4.9×10^6	3.7×10^6
PYCC 4891	1	1.2×10^5	2.0×10^6	2.0×10^6	1.8×10^6	3.9×10^6	2.2×10^6	2.6×10^6	3.7×10^6	3.1×10^6
	2	1.4×10^5	2.5×10^6	3.0×10^6	1.5×10^6	3.0×10^6	3.3×10^6	1.4×10^6	2.5×10^6	2.6×10^6
AP 5.4	1	1.2×10^5	3.4×10^6	3.5×10^6	3.5×10^6	3.4×10^6	2.1×10^6	1.8×10^6	1.7×10^6	1.5×10^6
	2	1.2×10^5	4.4×10^6	4.5×10^6	4.1×10^6	4.4×10^6	2.1×10^6	1.9×10^6	2.8×10^6	2.0×10^6
PYCC 6732	1	1.0×10^5	2.5×10^6	2.3×10^6	3.6×10^6	4.9×10^6	2.9×10^6	1.6×10^6	2.1×10^6	2.3×10^6
	2	1.2×10^5	2.2×10^6	2.4×10^6	2.4×10^6	2.3×10^6	3.6×10^6	1.9×10^6	2.6×10^6	2.9×10^6
Lalvin W15	1	1.2×10^5	1.0×10^6	9.6×10^5	4.1×10^5	9.1×10^4	9.1×10^4	6.4×10^4	7.6×10^4	5.5×10^4
	2	1.9×10^5	9.1×10^5	1.3×10^6	4.5×10^5	1.8×10^5	1.6×10^5	2.3×10^5	3.4×10^5	1.3×10^5
AWRI 1631	1	9.7×10^4	3.87×10^5	2.2×10^5	2.9×10^4	1.2×10^4	1.6×10^4	2.2×10^4	1.8×10^4	1.3×10^4
	2	1.9×10^5	6.5×10^5	2.0×10^5	2.8×10^4	1.8×10^4	2.2×10^4	1.3×10^4	1.6×10^4	9.6×10^3
PR	1	1.0×10^5	4.8×10^5	2.4×10^5	1.2×10^5	2.6×10^5	2.4×10^5	1.4×10^5	1.5×10^5	1.8×10^5
	2	1.4×10^5	1.5×10^5	2.2×10^5	1.9×10^5	1.8×10^5	n.d.	n.d.	n.d.	n.d.
Uvaferm VRB	1	9.6×10^4	2.5×10^5	1.5×10^5	6.0×10^4	2.6×10^4	3.9×10^4	2.9×10^4	6.7×10^4	3.7×10^4
	2	1.0×10^5	2.3×10^5	2.2×10^5	2.5×10^4	2.9×10^4	2.5×10^4	1.2×10^4	3.3×10^4	1.3×10^4
PYCC 4072	1	1.2×10^5	1.5×10^5	9.3×10^4	3.5×10^4	2.9×10^4	6.4×10^4	2.4×10^4	4.4×10^4	9.4×10^4
	2	1.8×10^5	2.6×10^5	1.0×10^5	5.5×10^4	5.5×10^4	7.1×10^4	4.7×10^4	9.0×10^4	3.9×10^4
TUM V1	1	2.1×10^5	2.8×10^5	4.6×10^4	5.2×10^4	2.6×10^4	1.7×10^4	2.3×10^4	2.1×10^4	4.5×10^4
	2	1.4×10^5	2.0×10^5	9.7×10^4	1.0×10^4	1.6×10^4	1.4×10^4	6.9×10^4	5.4×10^4	1.5×10^4

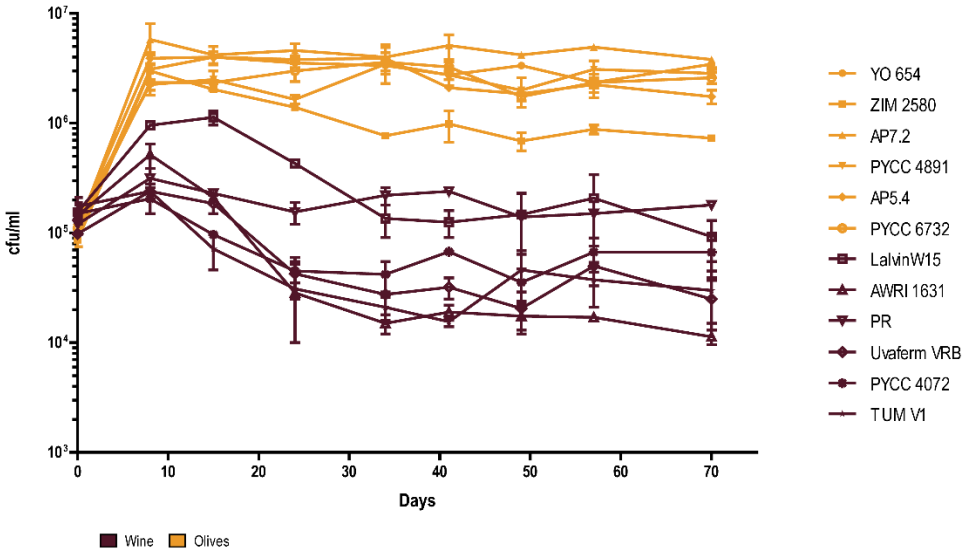


Figure A2.1 - Growth and survival in olive brine of six strains of the Olives population and six strains of the Win population. Two independent experiments were performed for each strain.

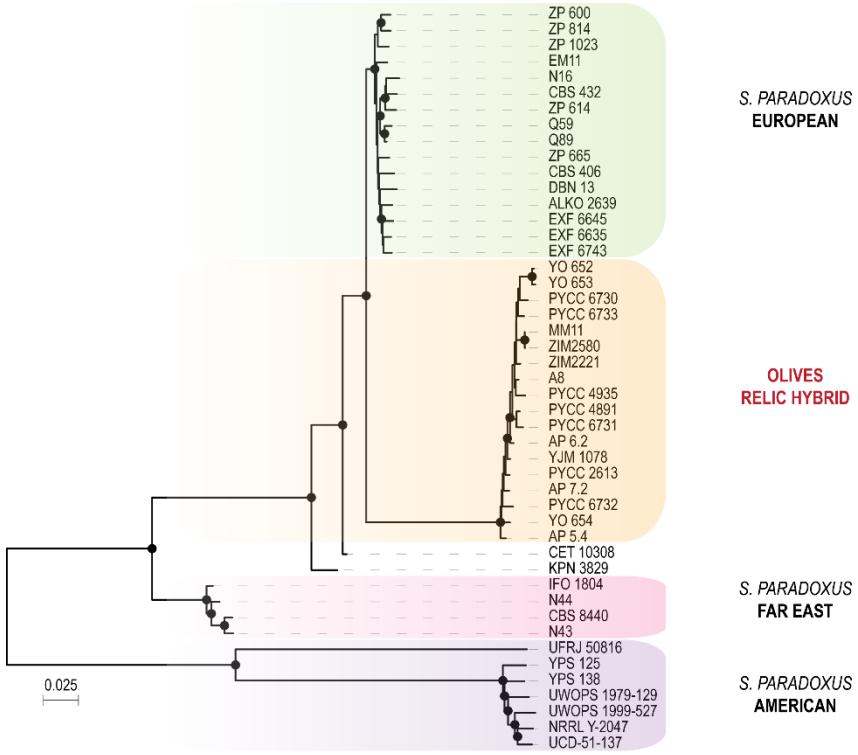
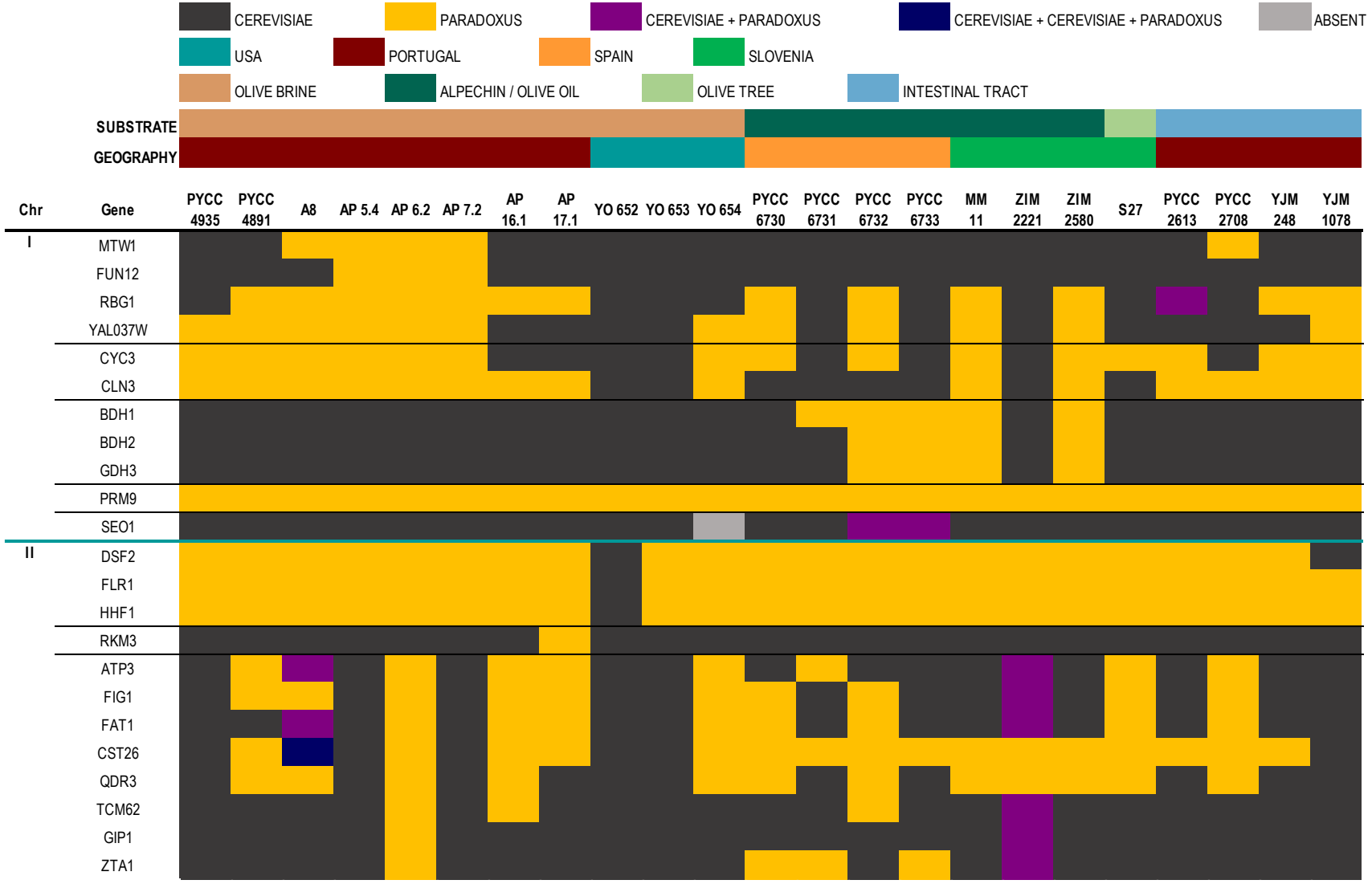
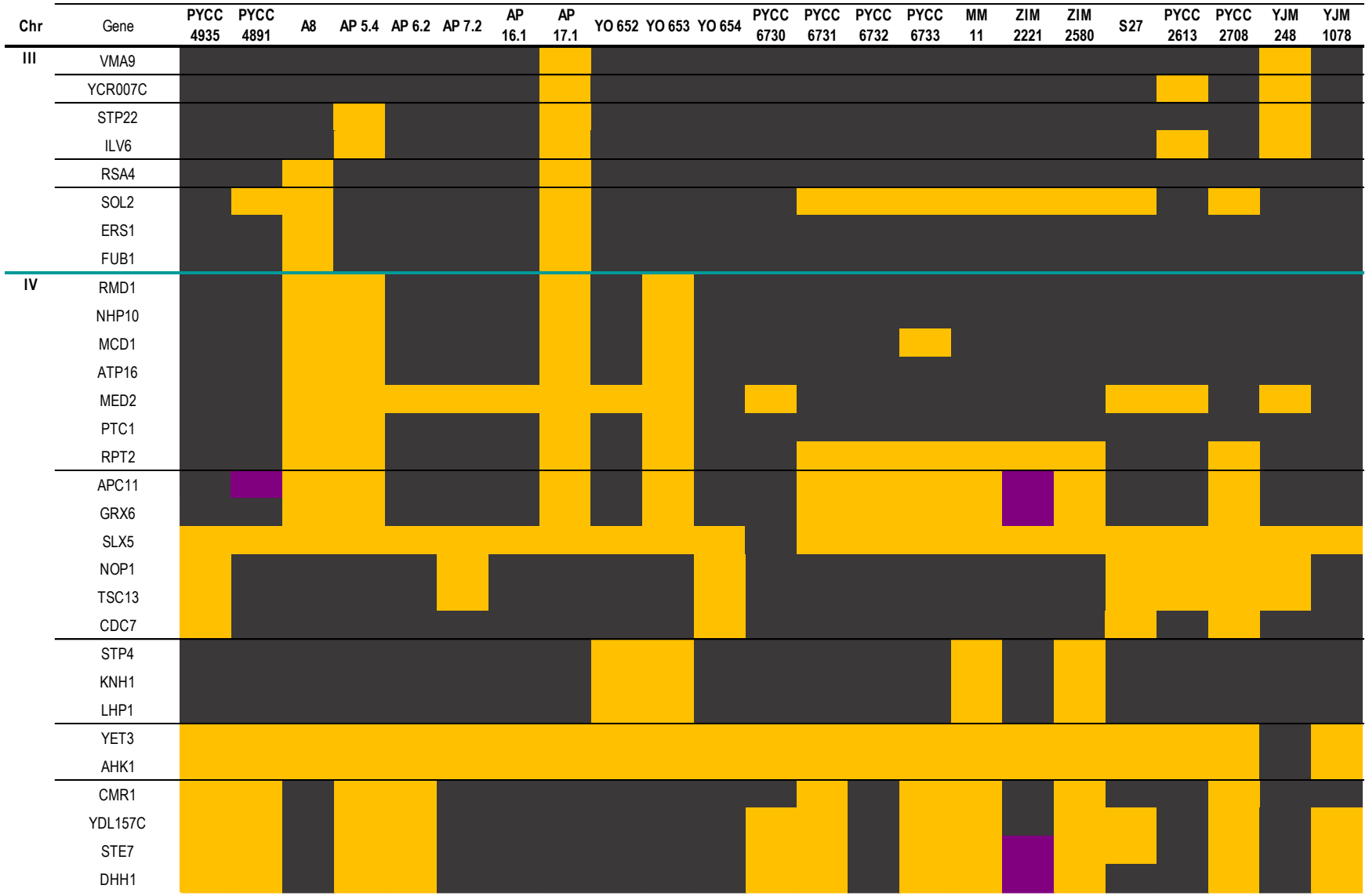


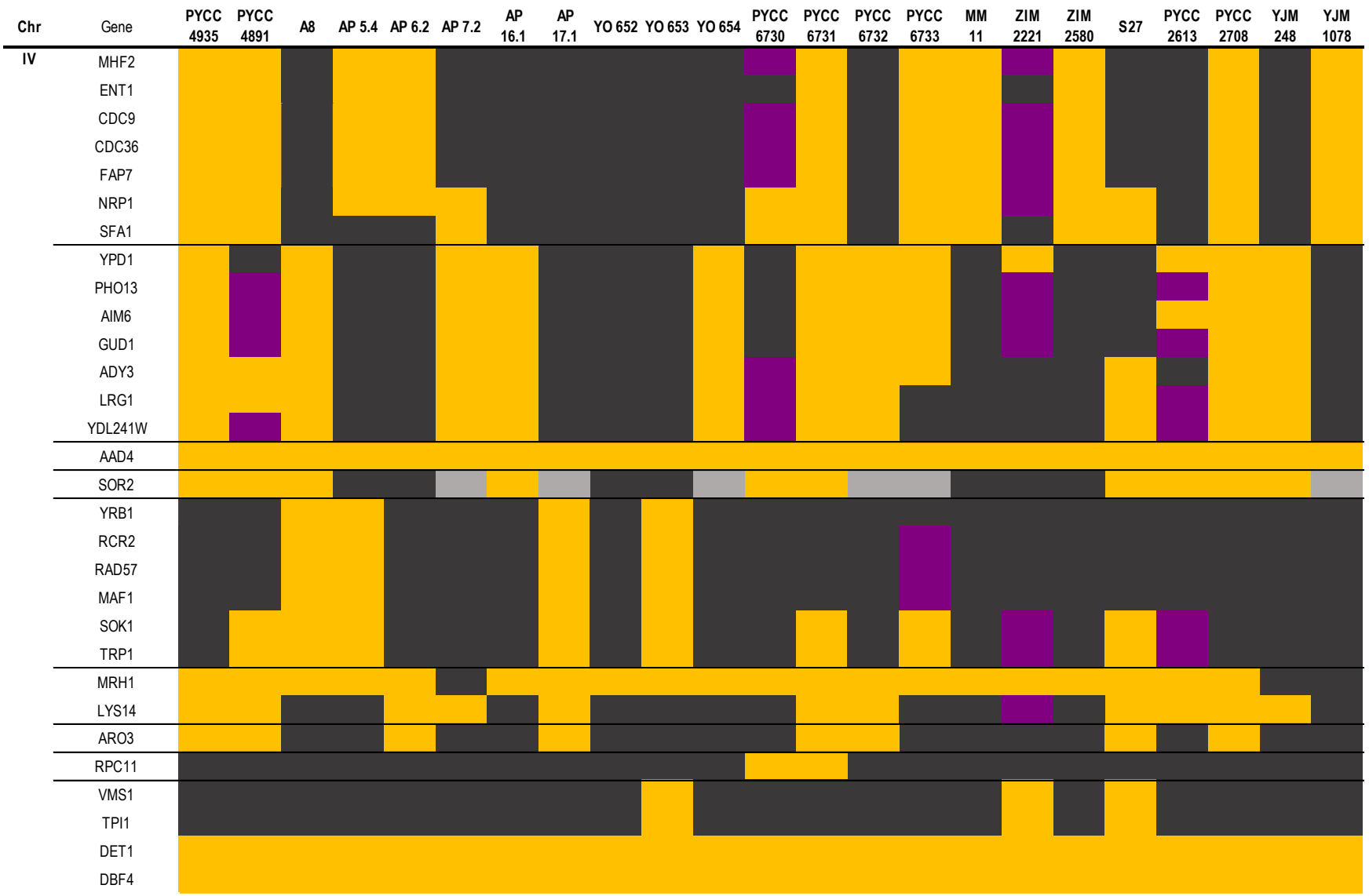
Figure A2.2 - The *S. paradoxus* sub-genome of hybrid strains originate in the European population of *S. paradoxus*. Phylogenetic tree constructed with the *S. paradoxus* fraction of the hybrid genomes and correspondent regions from representatives of the European, American, and Far Eastern population of *S. paradoxus*. The phylogeny was inferred from 47 sequences and 22,052 SNPs using the neighbor-joining method and the p-distance model of sequence evolution. Branch lengths correspond to the expected number of substitutions per site and black dots in tree nodes depict bootstrap values above 90% (1,000) replicates.

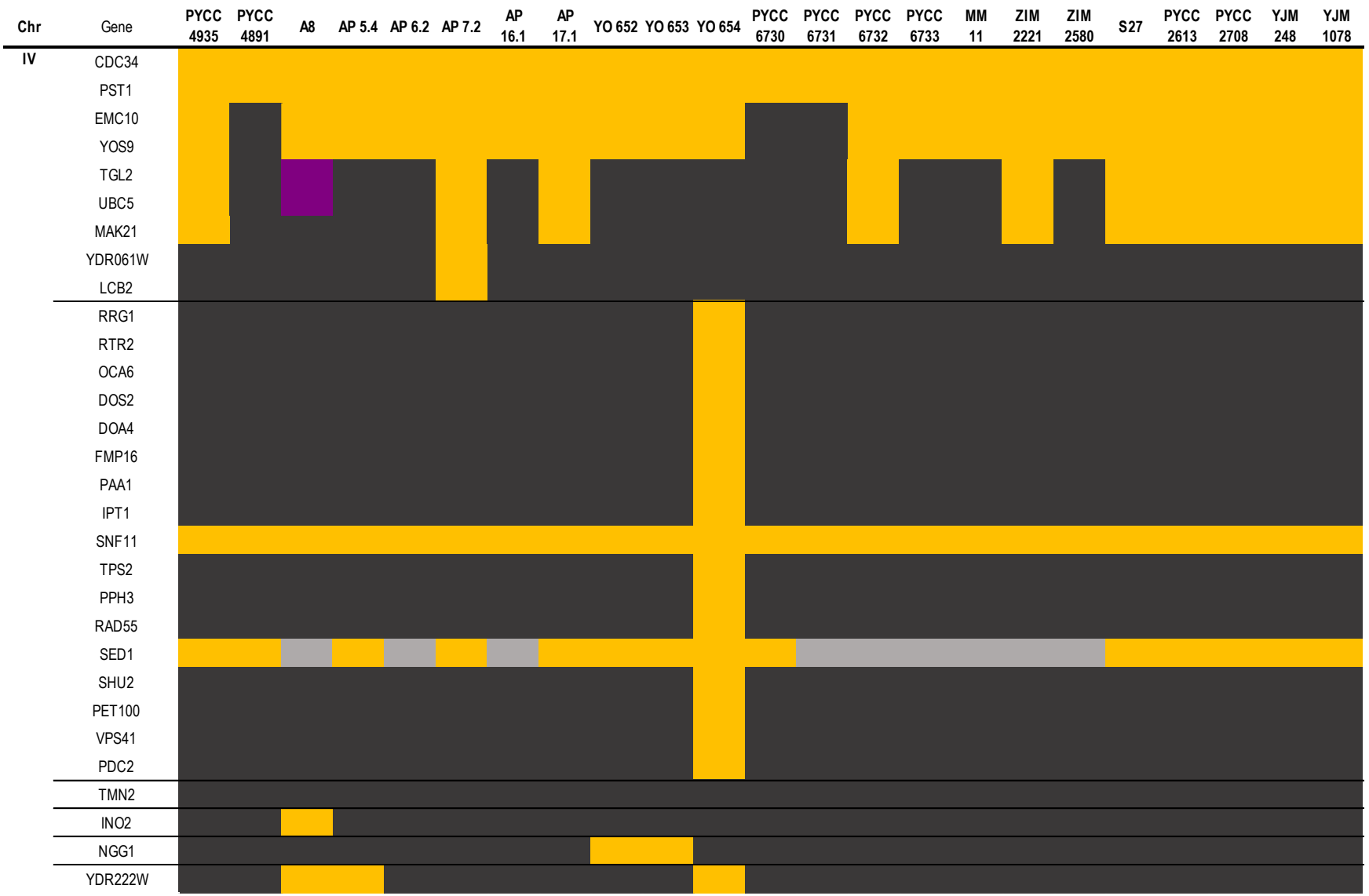
Table A2.4 - Gene content of the *S. paradoxus* sub-genome of relic hybrid strains.

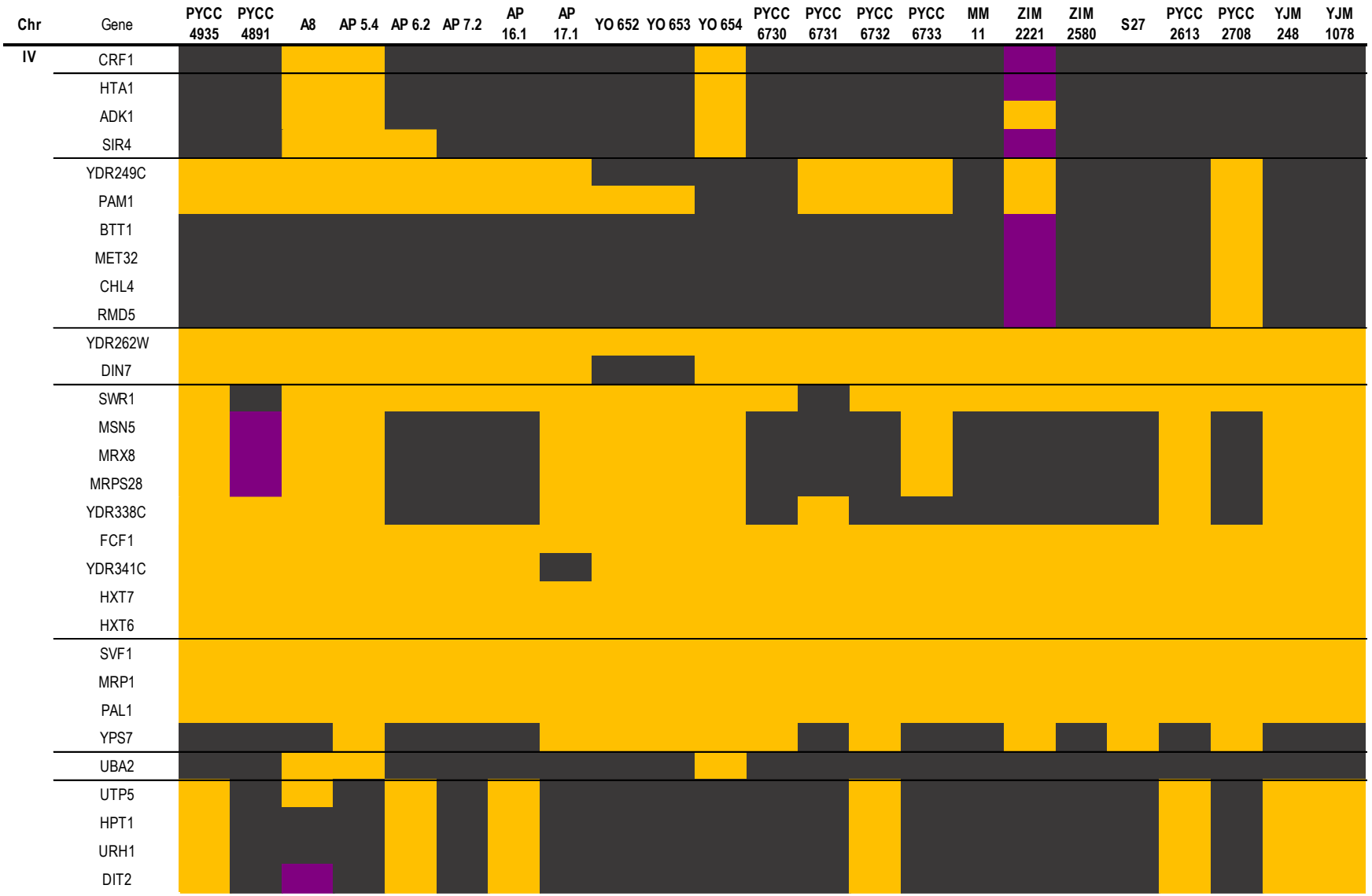


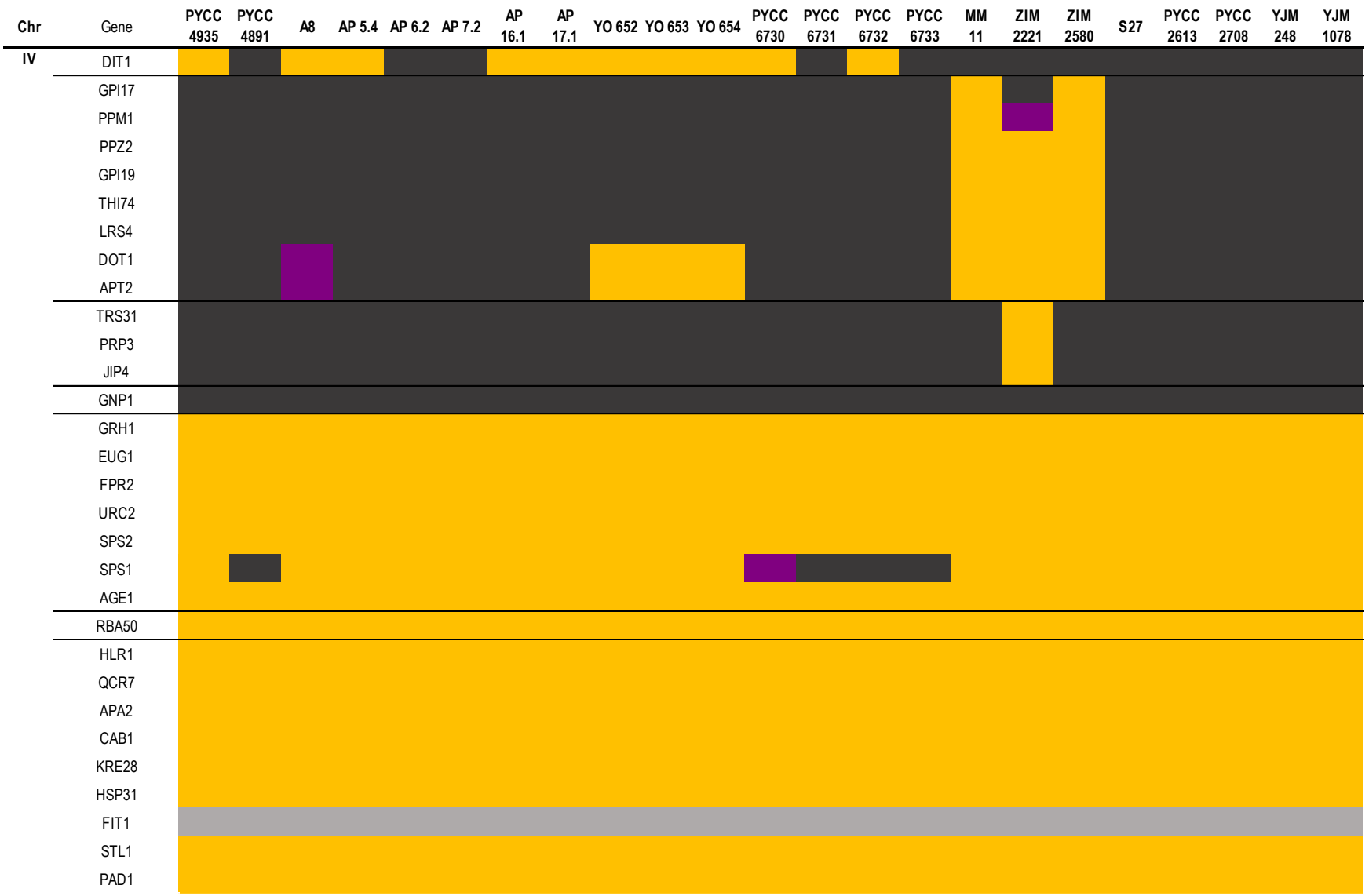

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
II
FMP23
REB1
ORC2
TRM7
YBR062C
YBR063C
ECM2
NRG2
TIP1
BAP2
TAT1
ALG14
YBR071W
HSP26
RDH54
PFF1
ECM8
SLM4
ECM33
III
GIT1
CDC10
MRPL32
YCP4
SAT4
RVS161
ADY2
RER1
YCL002C
PGS1
LDB16

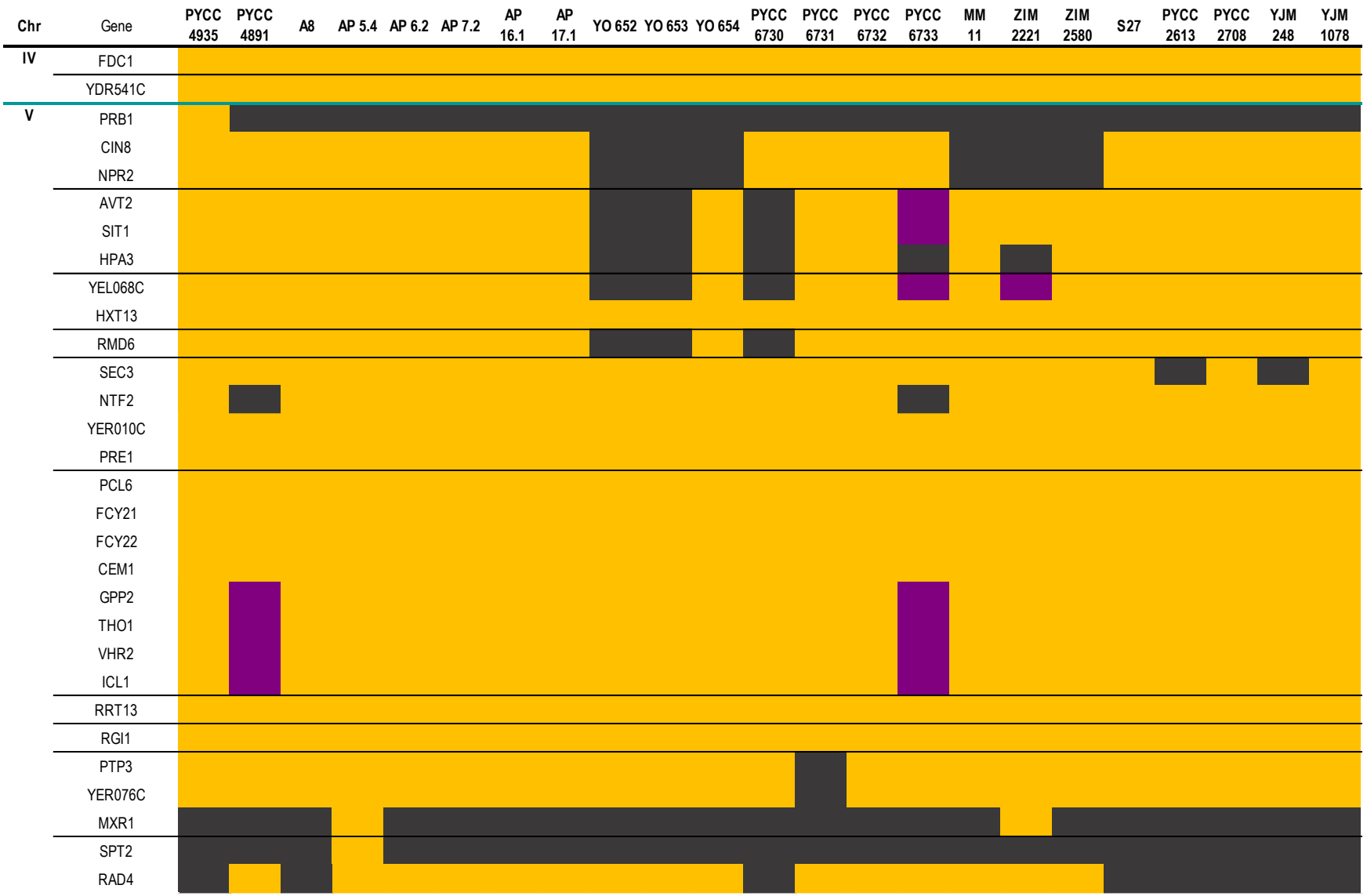

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
III
VMA9
YCR007C
STP22
ILV6
RSA4
SOL2
ERS1
FUB1
IV
RMD1
NHP10
MCD1
ATP16
MED2
PTC1
RPT2
APC11
GRX6
SLX5
NOP1
TSC13
CDC7
STP4
KNH1
LHP1
YET3
AHK1
CMR1
YDL157C
STE7
DHH1

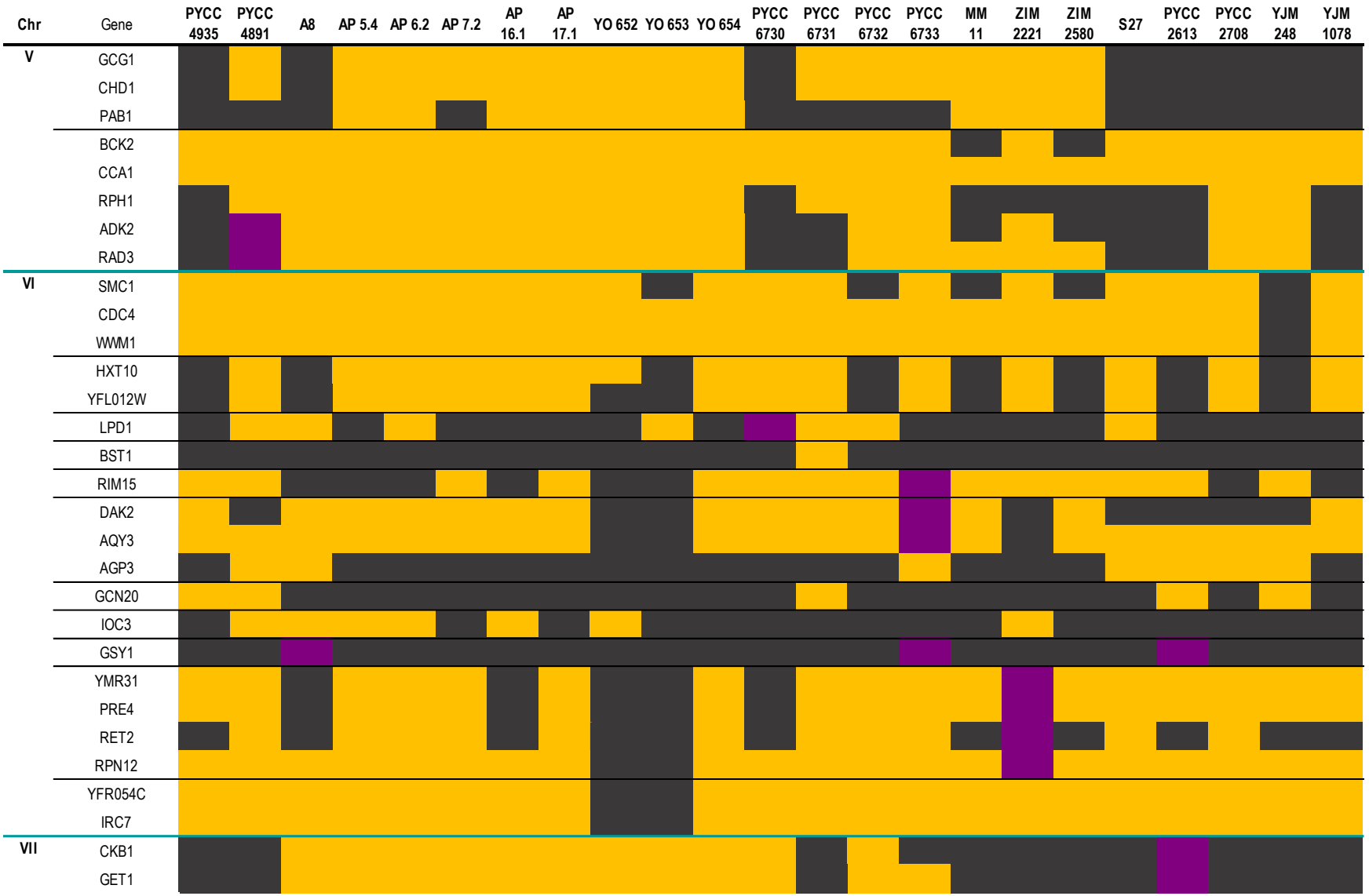

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
IV
MHF2
ENT1
CDC9
CDC36
FAP7
NRP1
SFA1
YPD1
PHO13
AIM6
GUD1
ADY3
LRG1
YDL241W
AAD4
SOR2
YRB1
RCR2
RAD57
MAF1
SOK1
TRP1
MRH1
LYS14
ARO3
RPC11
VMS1
TPI1
DET1
DBF4

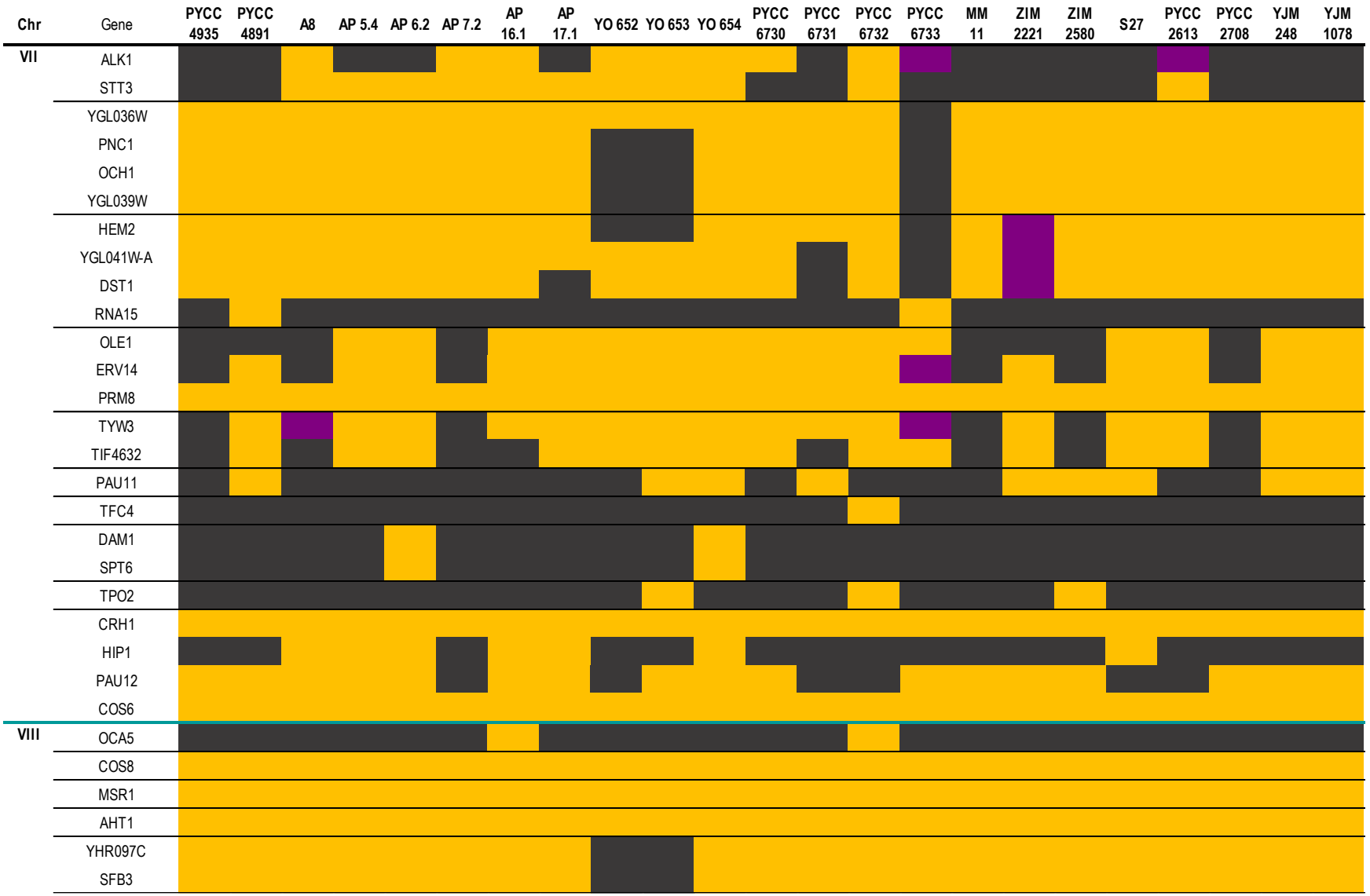

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
IV
CDC34
PST1
EMC10
YOS9
TGL2
UBC5
MAK21
YDR061W
LCB2
RRG1
RTR2
OCA6
DOS2
DOA4
FMP16
PAA1
IPT1
SNF11
TPS2
PPH3
RAD55
SED1
SHU2
PET100
VPS41
PDC2
TMN2
INO2
NGG1
YDR222W

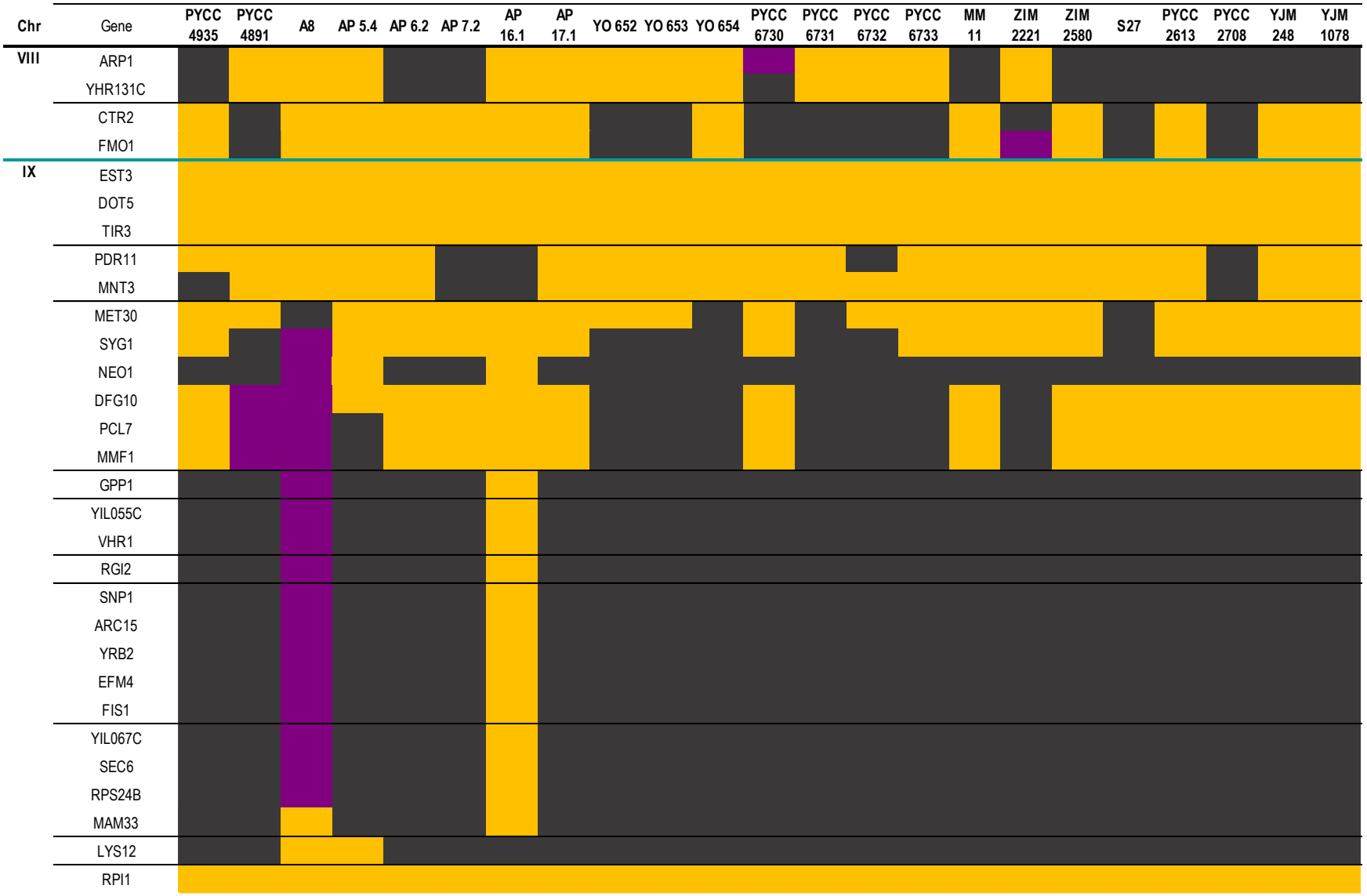

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
IV
CRF1
HTA1
ADK1
SIR4
YDR249C
PAM1
BTT1
MET32
CHL4
RMD5
YDR262W
DIN7
SWR1
MSN5
MRX8
MRPS28
YDR338C
FCF1
YDR341C
HXT7
HXT6
SVF1
MRP1
PAL1
YPS7
UBA2
UTP5
HPT1
URH1
DIT2

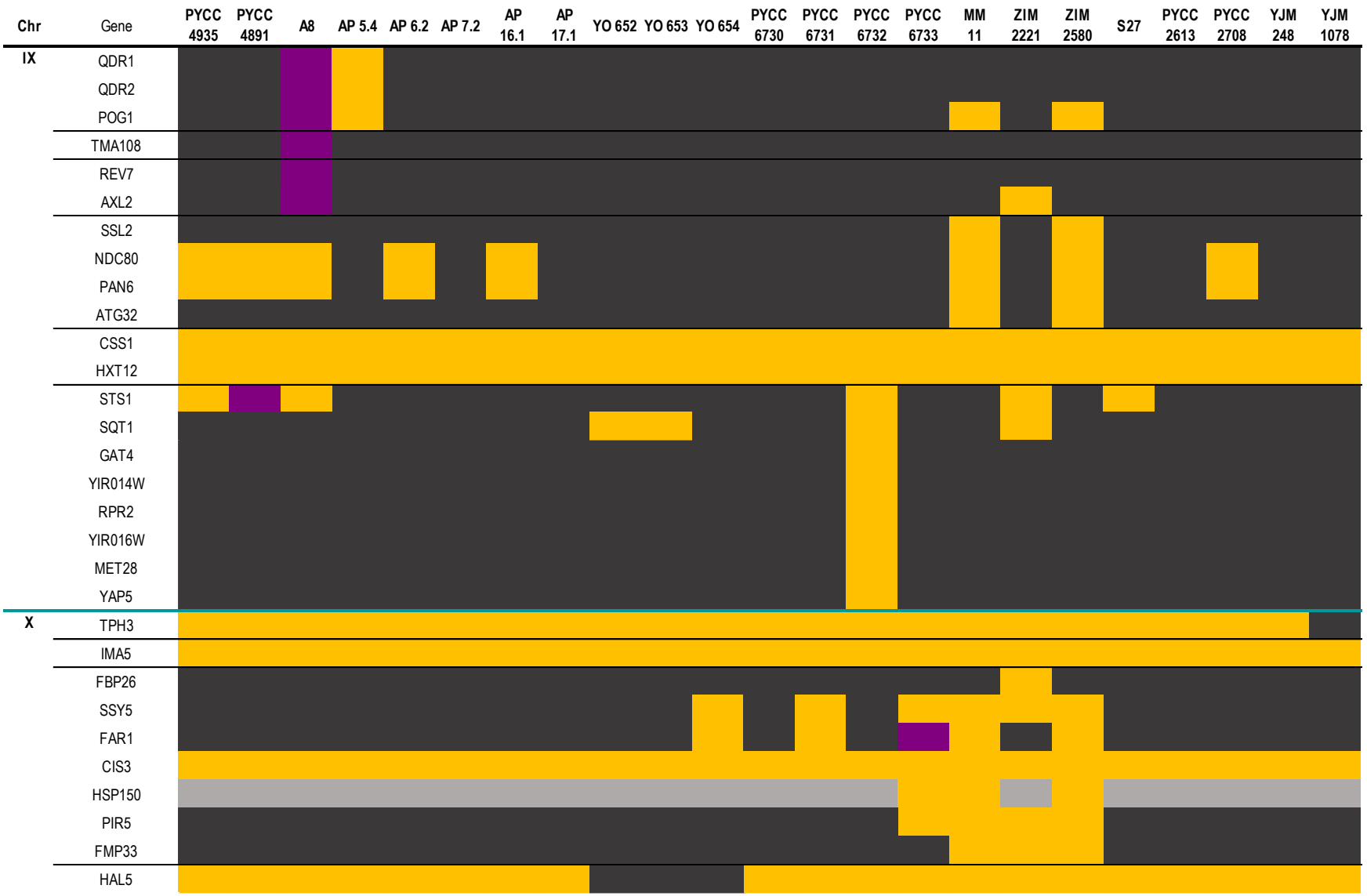

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
IV
DIT1
GPI17
PPM1
PPZ2
GPI19
THI74
LRS4
DOT1
APT2
TRS31
PRP3
JIP4
GNP1
GRH1
EUG1
FPR2
URC2
SPS2
SPS1
AGE1
RBA50
HLR1
QCR7
APA2
CAB1
KRE28
HSP31
FIT1
STL1
PAD1

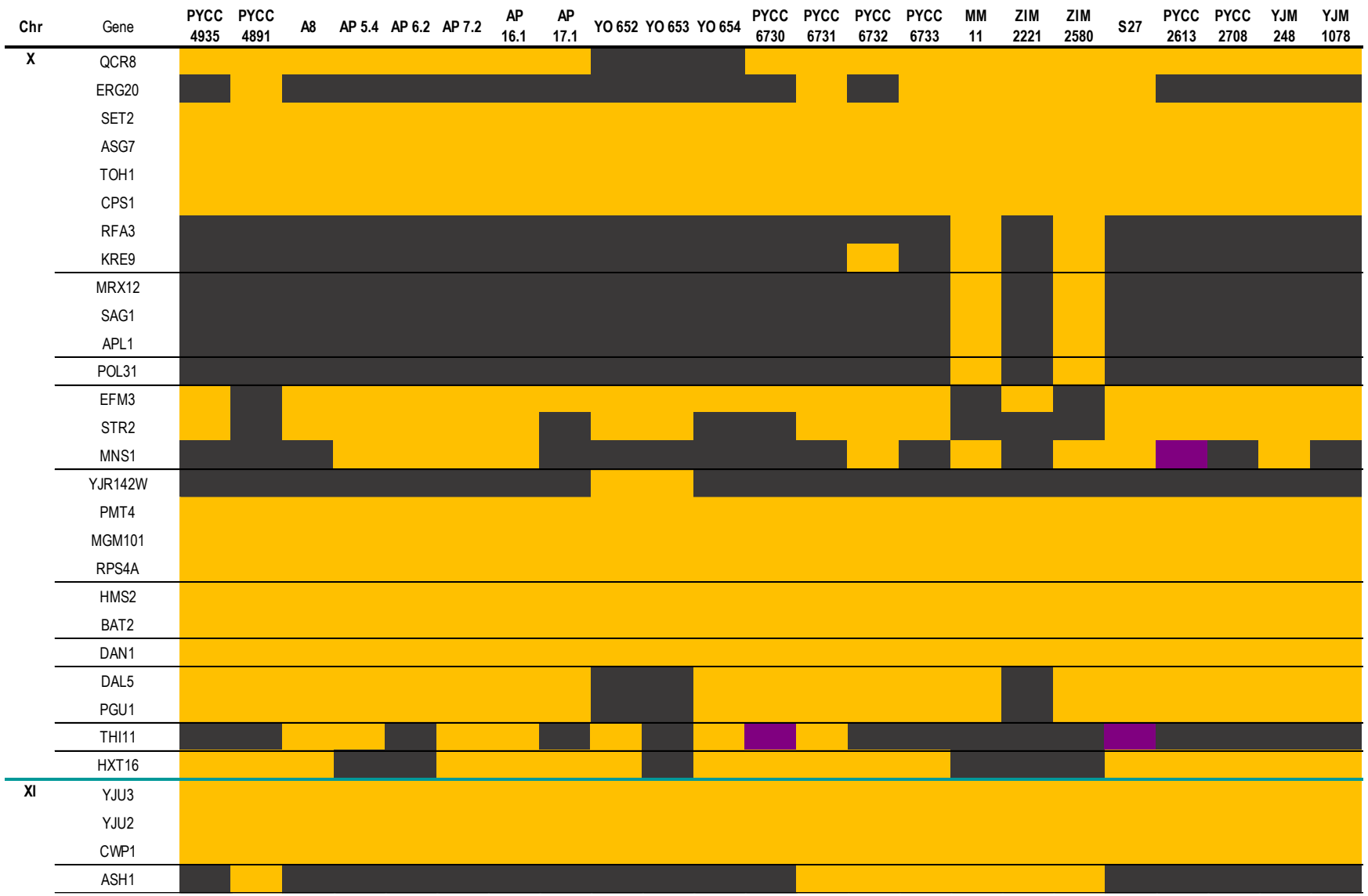

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
IV
FDC1
YDR541C
V
PRB1
CIN8
NPR2
AVT2
SIT1
HPA3
YEL068C
HXT13
RMD6
SEC3
NTF2
YER010C
PRE1
PCL6
FCY21
FCY22
CEM1
GPP2
THO1
VHR2
ICL1
RRT13
RGI1
PTP3
YER076C
MXR1
SPT2
RAD4

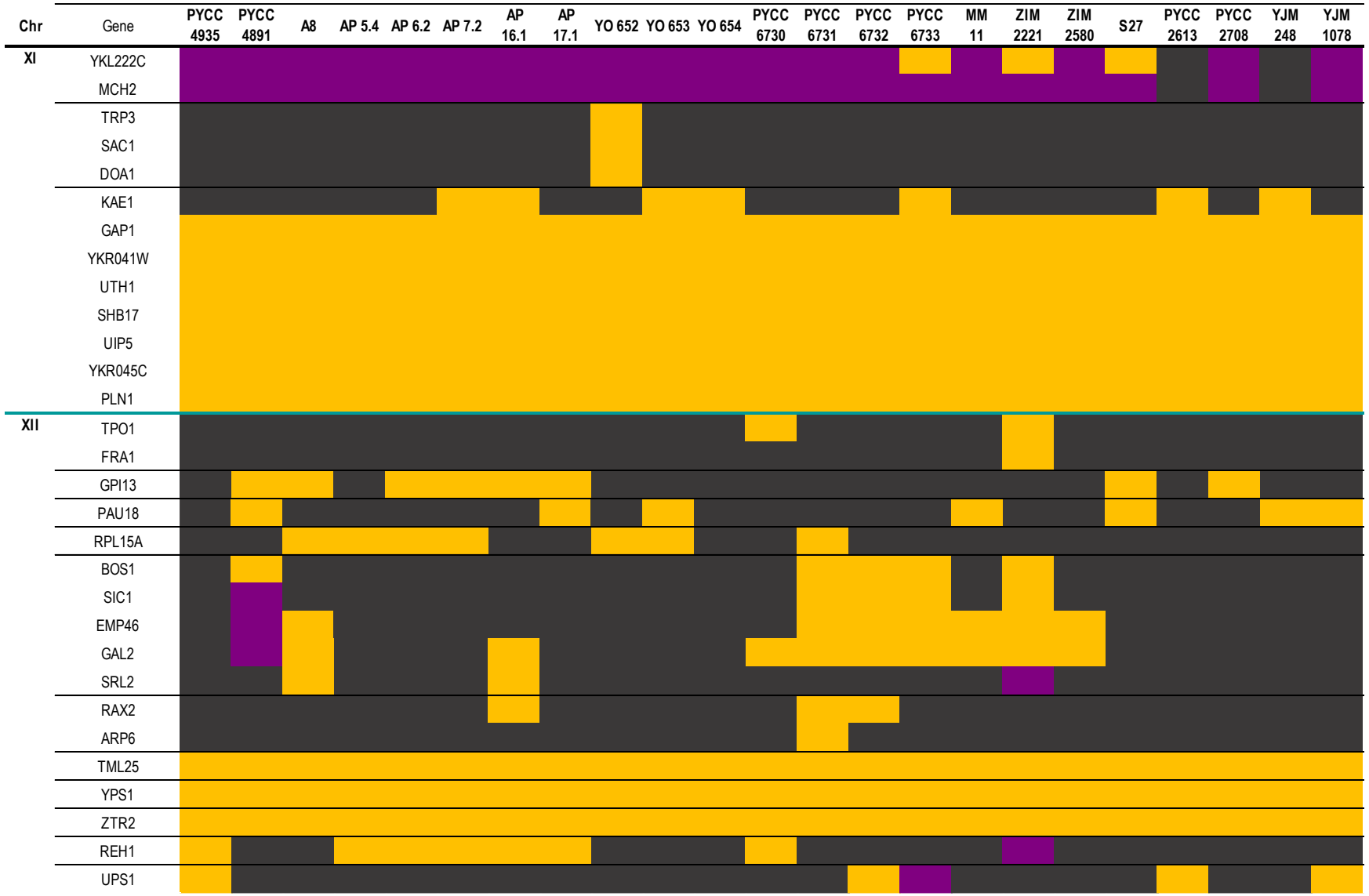

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
V
GCG1
CHD1
PAB1
BCK2
CCA1
RPH1
ADK2
RAD3
VI
SMC1
CDC4
WWM1
HXT10
YFL012W
LPD1
BST1
RIM15
DAK2
AQY3
AGP3
GCN20
IOC3
GSY1
YMR31
PRE4
RET2
RPN12
YFR054C
IRC7
VII
CKB1
GET1

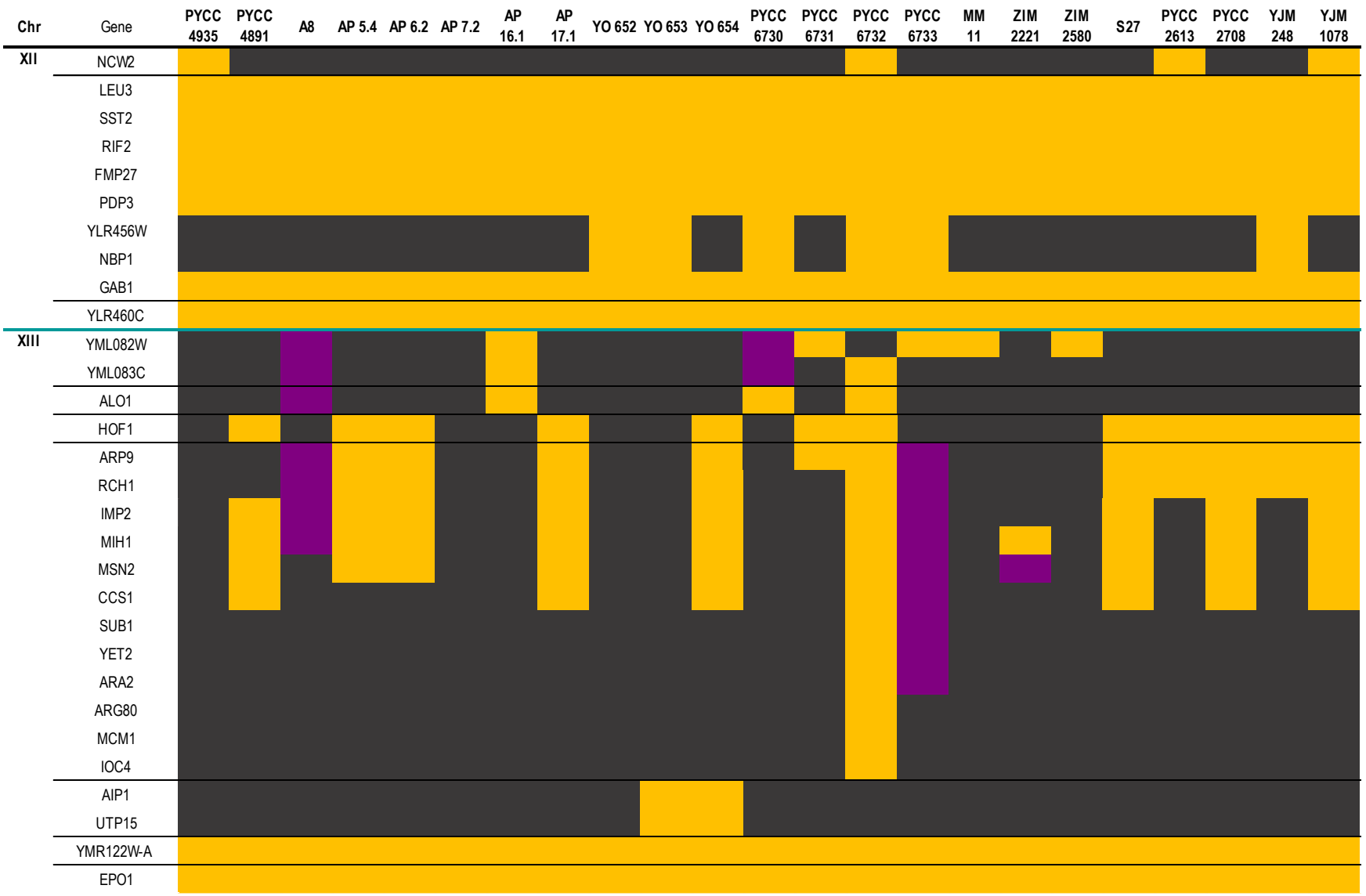

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
VII
ALK1
STT3
YGL036W
PNC1
OCH1
YGL039W
HEM2
YGL041W-A
DST1
RNA15
OLE1
ERV14
PRM8
TYW3
TIF4632
PAU11
TFC4
DAM1
SPT6
TPO2
CRH1
HIP1
PAU12
COS6
VIII
OCA5
COS8
MSR1
AHT1
YHR097C
SFB3

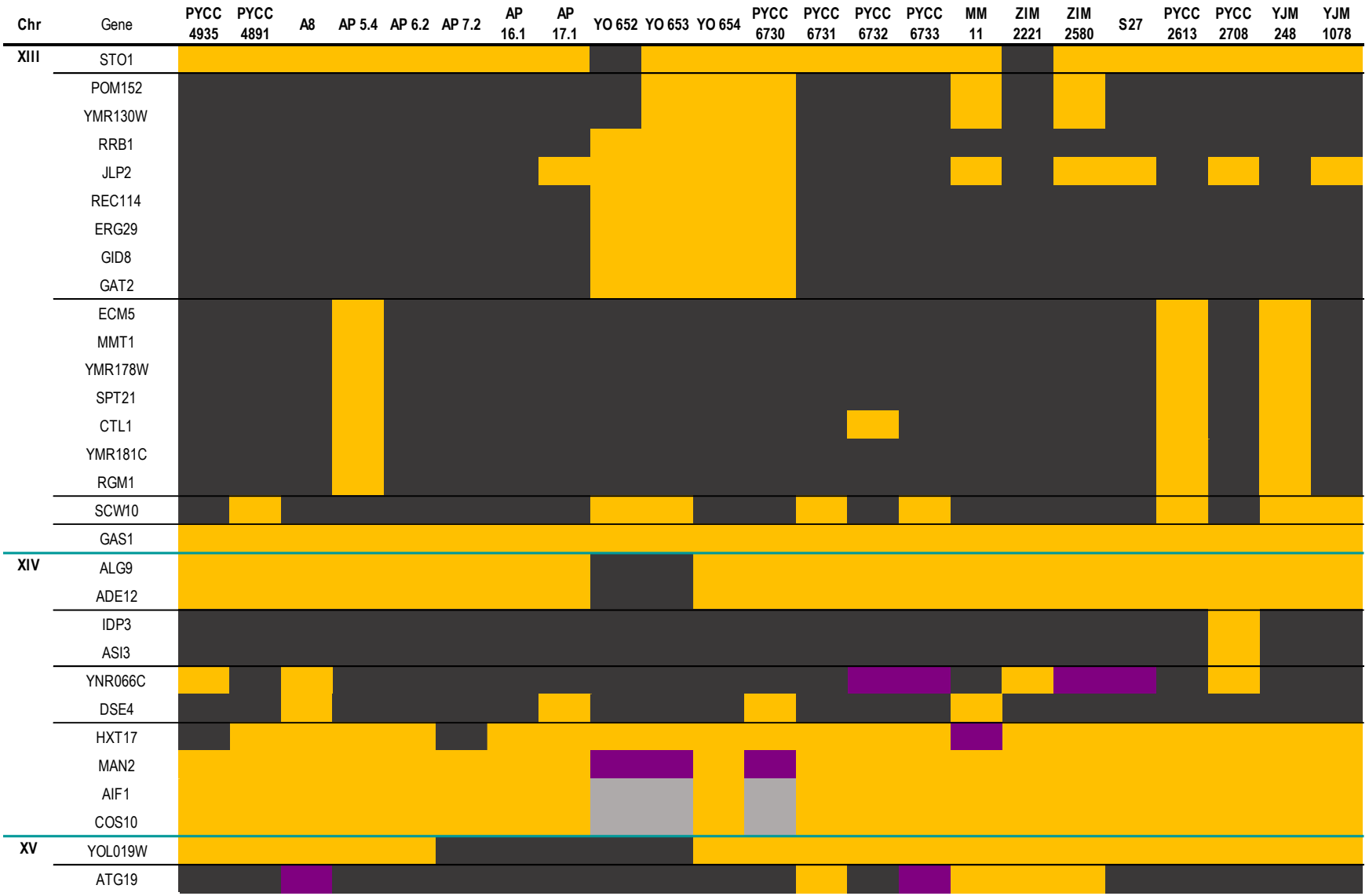

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
VIII
ARP1
YHR131C
CTR2
FMO1
IX
EST3
DOT5
TIR3
PDR11
MNT3
MET30
SYG1
NEO1
DFG10
PCL7
MMF1
GPP1
YIL055C
VHR1
RGI2
SNP1
ARC15
YRB2
EFM4
FIS1
YIL067C
SEC6
RPS24B
MAM33
LYS12
RPI1

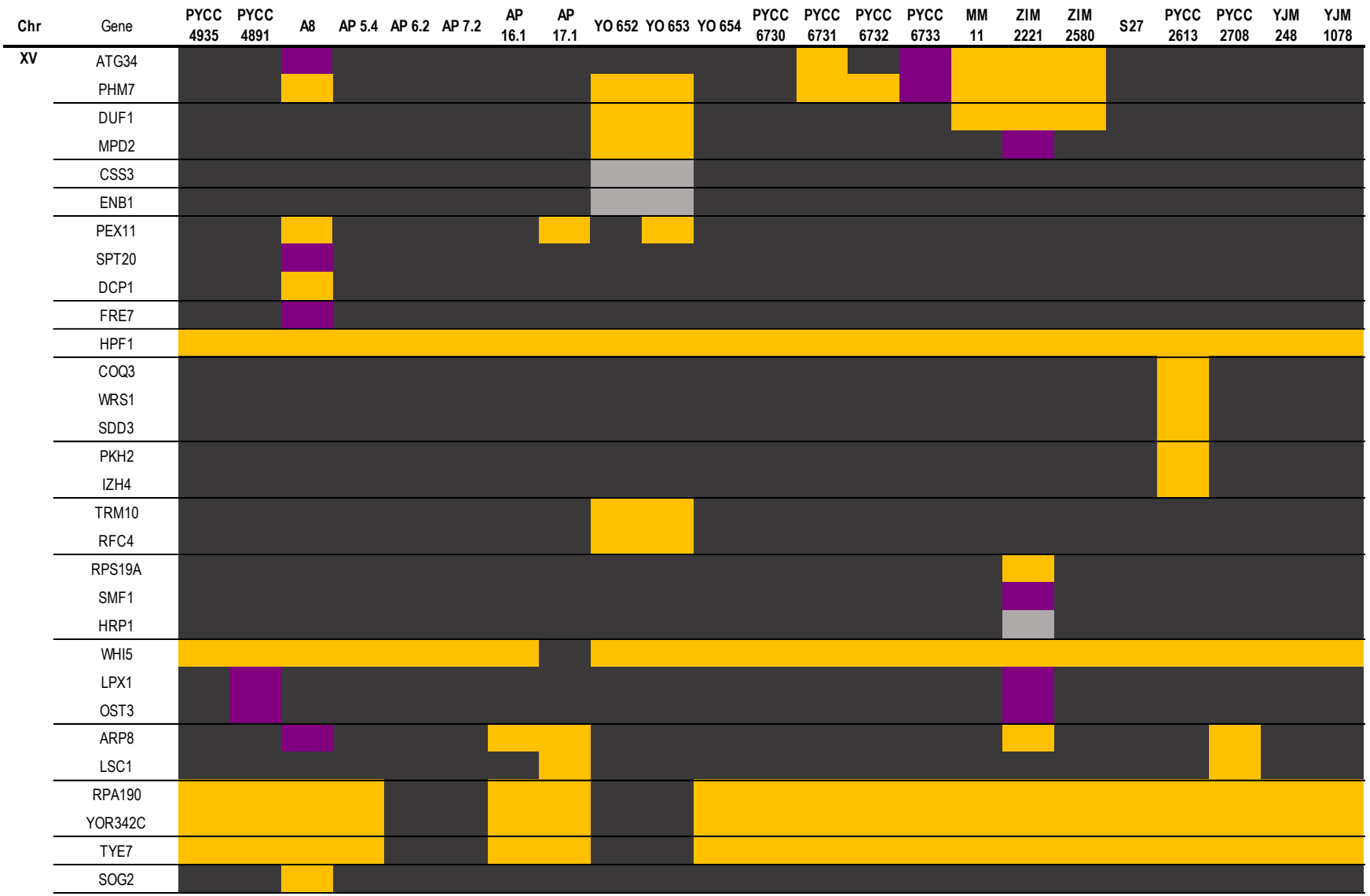

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
IX
QDR1
QDR2
POG1
TMA108
REV7
AXL2
SSL2
NDC80
PAN6
ATG32
CSS1
HXT12
STS1
SQT1
GAT4
YIR014W
RPR2
YIR016W
MET28
YAP5
X
TPH3
IMA5
FBP26
SSY5
FAR1
CIS3
HSP150
PIR5
FMP33
HAL5

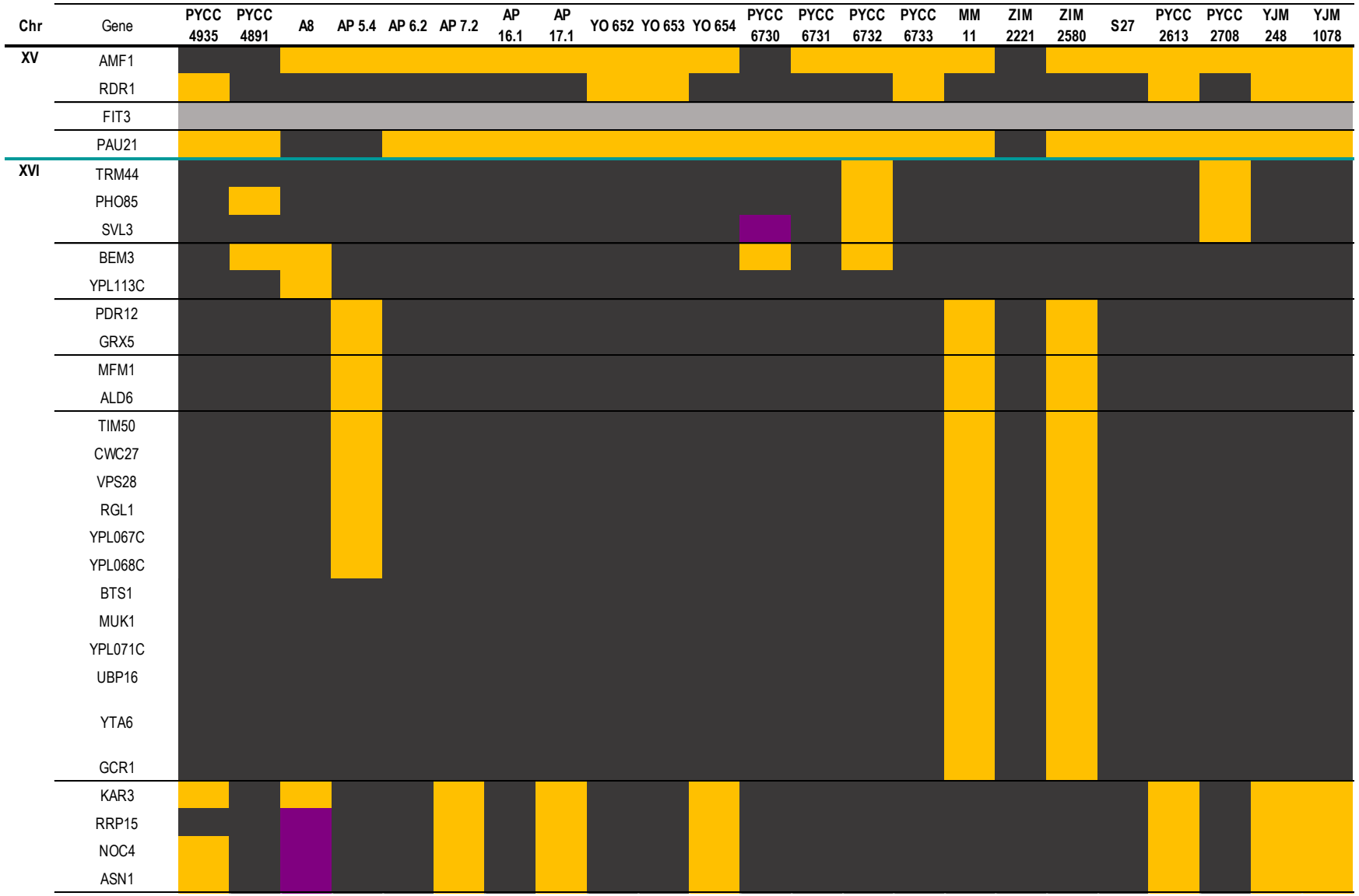

Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
X
QCR8
ERG20
SET2
ASG7
TOH1
CPS1
RFA3
KRE9
MRX12
SAG1
APL1
POL31
EFM3
STR2
MNS1
YJR142W
PMT4
MGM101
RPS4A
HMS2
BAT2
DAN1
DAL5
PGU1
THI11
HXT16
XI
YJU3
YJU2
CWP1
ASH1


Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
XI
YKL222C
MCH2
TRP3
SAC1
DOA1
KAE1
GAP1
YKR041W
UTH1
SHB17
UIP5
YKR045C
PLN1
XII
TPO1
FRA1
GPI13
PAU18
RPL15A
BOS1
SIC1
EMP46
GAL2
SRL2
RAX2
ARP6
TML25
YPS1
ZTR2
REH1
UPS1


Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
XII
NCW2
LEU3
SST2
RIF2
FMP27
PDP3
YLR456W
NBP1
GAB1
YLR460C
XIII
YML082W
YML083C
ALO1
HOF1
ARP9
RCH1
IMP2
MIH1
MSN2
CCS1
SUB1
YET2
ARA2
ARG80
MCM1
IOC4
AIP1
UTP15
YMR122W-A
EPO1


Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
XIII
STO1
POM152
YMR130W
RRB1
JLP2
REC114
ERG29
GID8
GAT2
ECM5
MMT1
YMR178W
SPT21
CTL1
YMR181C
RGM1
SCW10
GAS1
XIV
ALG9
ADE12
IDP3
ASI3
YNR066C
DSE4
HXT17
MAN2
AIF1
COS10
XV
YOL019W
ATG19


Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
XV
ATG34
PHM7
DUF1
MPD2
CSS3
ENB1
PEX11
SPT20
DCP1
FRE7
HPF1
COQ3
WRS1
SDD3
PKH2
IZH4
TRM10
RFC4
RPS19A
SMF1
HRP1
WHI5
LPX1
OST3
ARP8
LSC1
RPA190
YOR342C
TYE7
SOG2


Chr
Gene
PYCC 4935
PYCC 4891
A8
AP 5.4
AP 6.2
AP 7.2
AP 16.1
AP 17.1
YO 652
YO 653
YO 654
PYCC 6730
PYCC 6731
PYCC 6732
PYCC 6733
MM 11
ZIM 2221
ZIM 2580
S27
PYCC 2613
PYCC 2708
YJM 248
YJM 1078
XV
AMF1
RDR1
FIT3
PAU21
XVI
TRM44
PHO85
SVL3
BEM3
YPL113C
PDR12
GRX5
MFM1
ALD6
TIM50
CWC27
VPS28
RGL1
YPL067C
YPL068C
BTS1
MUK1
YPL071C
UBP16
YTA6
GCR1
KAR3
RRP15
NOC4
ASN1

Chr	Gene	PYCC 4935	PYCC 4891	A8	AP 5.4	AP 6.2	AP 7.2	AP 16.1	AP 17.1	YO 652	YO 653	YO 654	PYCC 6730	PYCC 6731	PYCC 6732	PYCC 6733	MM 11	ZIM 2221	ZIM 2580	S27	PYCC 2613	PYCC 2708	YJM 248	YJM 1078
XVI	YPR147C																							
	YPR148C																							
	NCE102																							
	SUE1																							
	URN1																							
	MAY24																							
	PIN3																							
	NCA2																							
	TPO3																							
	TDA6																							
	***S. paradoxus* genes**	**250**	**254**	**314**	**288**	**246**	**237**	**249**	**275**	**193**	**216**	**282**	**227**	**241**	**271**	**259**	**269**	**269**	**270**	**236**	**268**	**250**	**259**	**244**

Shared *S. paradoxus* genes **103**

Blue horizantal line separate chromosomes

Black horizontal line separate gene blocks

A3
Appendix 3

Supplementary information pertaining to Chapter 4

Pontes, A.; Hutzler, M.; Brito, P.H.; Sampaio, J.P. Revisiting the Taxonomic Synonyms and Populations of *Saccharomyces cerevisiae*—Phylogeny, Phenotypes, Ecology and Domestication. Microorganisms 2020, 8, 903.

https://doi.org/10.3390/microorganisms8060903

Table A3.1 - Strains and genomes used in the work presented in chapter 4 and relevant information pertaining to them.

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]		AQY2[3]		RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]	GENOME DATA	REFERENCE
									A	B	C										
CBS 1175		*Saccharomyces cerevisiae* var. *orsati*	Wine	Switzerland		WINE - Main	1	3522	15/15			NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
CBS 1194		*Saccharomyces ellipsoideus* var. *thermophilus*	Wine	Unknown		WINE - Main	1	1597				NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
CBS 1479		*Saccharomyces ellipsoideus* var. *montibensis*	Wine	Unknown		WINE - Main	1	1083				NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
CBS 1192		*Saccharomyces ellipsoideus* var. *alpinus*	Wine	Unknown		WINE - Main	1	1183	15/15		18/19	NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
CBS 423	PYCC 2607	*Saccharomyces chodati*	Wine	Switzerland		WINE - Main	1	4813	15/15		18/19	NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
CBS 436	PYCC 8195	*Saccharomyces yedo*	Sake moto	Japan		WINE - Main	1	n.d.		5/5	18/19	NF	A881del	NF	11bpdel	x	x	x	x	PRJEB36095	This study
CBS 457		*Saccharomyces ellipsoideus* var. *umbra*	Grape must	Italy		WINE - Main	1	5501			18/19	NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
CBS 459		*Saccharomyces italicus*	Grape must	Italy		WINE - Main	1	1095				NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
PYCC 6728	CBS 1489	*Saccharomyces pulmonalis*	Sputum of tuberculosis patient	Pavia, Italy		WINE - Main	1	1873			18/19	NF	A881del	NF	11bpdel	x	x	x	x	PRJEB36095	This study
CBS 439		*Saccharomyces elongatus*	Silvaner grapes	Germany		WINE - Main	1	1758	15/15			NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
PYCC 4653	CBS 1395; NRRL Y-1529	*Saccharomyces ellipsoideus*	Unknown	Unknown		WINE - Main	1	1988				NF	A881del	NF	11bpdel	x	x	x	x	PRJEB36095	This study
PR			Pasteur Red	Unknown		WINE - Main	1	1715				NF	A881del	NF	11bpdel	x	x	x	x	PRJEB19382	Almeida et al., 2017
CBS 4507		*Saccharomyces multisporus*	Brewer's yeast, English top yeast	Unknown		WINE - Main	1	2995			17/19	NF	A881del	NF	11bpdel	x	x	x	x	ERP014555	Peter et al., 2018
PYCC 6727	CBS 1227	*Saccharomyces annulatus*	Human skin infection	France		WINE - Main	1	1415	15/15		18/19	NF	A881del	NF	11bpdel	x	x	x	x	PRJEB36095	This study
EXF 6719			Wine must	Cuber, Slovenia		WINE - Main	1	2053			17/19	NF	A881del	NF	11bpdel	✓	x	x	x	PRJEB7601	Almeida et al., 2015
ZP 641			Spontaneous red wine fermentation	Castelo de Vide, Portugal		WINE - Main	1	1484				NF	A881del	NF	11bpdel	x	x	x	x	PRJEB7601	Almeida et al., 2015
Lalvin BM45	AWRI 1486		Commercial yeast (white wine)	Italy		WINE - Main	1	1280		5/5	2/19	NF	A881del	NF	11bpdel	x	x	x	x	PRJEB19382	Almeida et al., 2017
Lalvin CY-3079	AWRI 2078		Commercial yeast (red wine)	Unknown		WINE - Main	1	1974		5/5	4/19	NF	A881del	NF	11bpdel	x	x	x	x	PRJEB19382	Almeida et al., 2017
PYCC 8196	CBS 1460	*Saccharomyces marchalianus*	Fermenting fruit	Indonesia		WINE - Main	1	1855				NF	A881del	NF	11bpdel	x	x	x	x	PRJEB36095	This study
YJM 1415	NRRL Y-268		Wine	France		WINE - Main	1	1256		5/5	19/19	NF	A881del	NF	11bpdel	x	x	x	x	PRJNA189914	Strope et al., 2015
AWRI 1631	N 96		Wine	Australia		WINE - Main	1	n.d.			8/19	NF	A881del	NF	11bpdel	x	x	x	x	PRJNA30553	Borneman et al., 2008

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]		AQY2[3]		RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C												
AWRI 796	Active Dry Wine Yeast		Commercial wine yeast	South Africa		WINE - Main	1	n.d.		DUP		F		NF	11bpdel	x	x	x	✓	NF	EUR	PRJNA48559	Borneman et al., 2011
CBS 5635		*Saccharomyces coreanus*	Grape must	South Africa		WINE - Main	1	1552				NF	A881del	NF	11bpdel	x	x	x	✓	F	EUR	ERP014555	Peter et al., 2018
YJS 5879	ULG84F88I90	*Saccharomyces boulardii*	Lychee (reference strain Biocodex)	Vietnam		WINE - Main	1	3918		5/5		NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
YJS 5881	perentol		Perenterol capsule	Brussels, Belgium		WINE - Main	1	3948		5/5		NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
YJS 5880	ultra lev. 1127		UltraLevure packet	Paris, France		WINE - Main	1	3956		5/5		NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
VL3	B6, AWRI 1688		Commercial wine yeast	Laffort		WINE - Main	1	n.d.		5/5	8/19	NF	A881del	NF	11bpdel	x	x	x	x			PRJNA48565	Borneman et al., 2011
A-14			Wine conserved in amphora	Georgia		WINE - Main	1	9361		5/5	2/19	NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
A-16			Wine conserved in amphora	Georgia		WINE - Main	1	12369		5/5	2/19	NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
A-19			Wine conserved in amphora	Georgia		WINE - Main	1	13819		5/5	2/19	NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
D-4			Wine conserved in amphora	Georgia		WINE - Main	1	14899		5/5	2/19	NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
CBS 5112		*Saccharomyces onubensis*	Grape must	Spain		WINE - Main	1	1284				NF	A881del	NF	11bpdel	x	x	x	x			ERP014555	Peter et al., 2018
PYCC 2613	CBS 2909; NRRL YB-6041	*Saccharomyces italicus var. melibiosi*	Faeces of man	Portugal		OLIVES	2	6783				NF	A881del	NF	11bpdel	x	x	x	✓	F	EUR	PRJEB30431	Pontes et al., 2019
CBS 7002	PYCC 8023	*Saccharomyces hispalensis*	Alpechin	Sevilla, Spain		OLIVES	2	2471				NF	A881del	NF	11bpdel	x	x	x	✓	F	EUR	PRJEB30431	Pontes et al., 2019
PYCC 6730	VKM Y-1231	*Saccharomyces oleaceus*	Olives	Spain		OLIVES	2	2965		5/5		NF	A881del	NF	11bpdel	x	x	x	✓	F	EUR	PRJEB30431	Pontes et al., 2019
PYCC 6732	VKM Y-1234	*Saccharomyces oleaginosus*	Alpechin	Spain		OLIVES	2	1886				NF	A881del	NF	11bpdel	x	x	x	✓	F	EUR	PRJEB30431	Pontes et al., 2019
PYCC 6731	VKM Y-1232	*Saccharomyces norbensis*	Alpechin	Spain		OLIVES	2	1393		4/5		NF	A881del	NF	11bpdel	x	x	x	✓	F	EUR	PRJEB30431	Pontes et al., 2019
PYCC 6733	VKM Y-1235	*Saccharomyces hienipiensis*	Alpechin	Spain		OLIVES	2	5258				NF	A881del	NF	11bpdel	x	x	x	✓	F	EUR	PRJEB30431	Pontes et al., 2019
Lalvin QA23			Commercial wine yeast	Portugal		WINE - PDM	3	n.d.	15/15	TRIP	19/19	NF	A881del	F		x	x	x	x			PRJNA48561	Borneman et al., 2011
PYCC 4074			Commercial wine yeast (Fermichamp)	Portugal		WINE - PDM	3	2263	15/15	5/5		NF	A881del	F		x	x	x	x			PRJEB19382	Almeida et al., 2017
EC 1118	Prise de Mousse		Industrial strain isolated from Champagne	France		WINE - PDM	3	n.d.	15/15	5/5	19/19	NF	A881del	NF	11bpdel	x	x	x	x			PRJEA37863	Novo et al., 2009
PYCC 6722	CBS 5155; VKM Y-515	*Saccharomyces prostoserdovii*	Wine	South Armenia		WINE - PDM	3	2121				NF	A881del	F		x	x	x	x			PRJEB19382	Almeida et al., 2017
PYCC 6729			Jerez-wine	Armenia		WINE - PDM	3	2901				NF	A881del	F		x	x	x	x			PRJEB19382	Almeida et al., 2017
5-18			Wine conserved in amphora	Georgia		WINE - PDM	3	4034		5/5		NF	A881del	F		x	x	x	x			ERP014555	Peter et al., 2018
B-24			Wine conserved in amphora	Georgia		WINE - PDM	3	5115		5/5	19/19	NF	A881del	F		x	x	x	x			ERP014555	Peter et al., 2018
CBS 2247		*Saccharomyces capensis*	Wine	South Africa		WINE - PDM	3	1547				NF	A881del	F		x	x	x	x			ERP014555	Peter et al., 2018

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]			AQY2[3]			RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C														
CBS 4054		*Saccharomyces aceti*	Wine	Spain		WINE - PDM	3	1712				NF	A881del		F			x	x	x	x			ERP014555	Peter et al., 2018
CBS 5835		*Saccharomyces hispanica*	Wine	Spain		WINE - PDM	3	3825				NF	A881del		F			x	x	x	x			ERP014555	Peter et al., 2018
CBS 6006		*Saccharomyces gaditensis*	Wine	Spain		WINE - PDM	3	3867		5/5	19/19	NF	A881del		F			x	x	x	x			ERP014555	Peter et al., 2018
CBS 6007		*Saccharomyces cordubensis*	Wine	Spain		WINE - PDM	3	1520				NF	A881del		F			x	x	x	x			ERP014555	Peter et al., 2018
PYCC 6726			Jerez-wine	Spain		WINE - PDM	3	4939				NF	A881del		F			x	x	x	x			PRJEB19382	Almeida et al., 2017
CLIB 1082			Cider brewery, dry cider	Upper Normandy, France		CACHAÇA-BIOETHANOL	4	22616	15/15	5/5		NF	A881del	V121M	F			✓	✓	x	x			ERP014555	Peter et al., 2018
UFMG-CM-Y030	CAY1007		Cachaça (commercial strain)	Minas Gerais, Brazil		CACHAÇA-BIOETHANOL	4	31074		5/5		NF	A881del		NF		G25del	x	x	x	✓	F	SA	PRJEB24932	Barbosa et al., 2018
UFMG-CM-Y630	SC3		Cachaça	Santa Catarina, Brazil		CACHAÇA-BIOETHANOL	4	3570		5/5		F			NF	11bpdel		x	✓	x	x			PRJEB24932	Barbosa et al., 2018
CBS 7959			Factory producing ethanol (sugar cane syrup)	Sao Paulo, Brazil		CACHAÇA-BIOETHANOL	4	2365		5/5		NF	A881del		NF	11bpdel		✓	✓	x	x			ERP014555	Peter et al., 2018
UFMG-CM-Y633	T09		Cachaça	Tocantins, Brazil		CACHAÇA-BIOETHANOL	4	6833				NF	A881del		F			✓	x	x	x			PRJEB24932	Barbosa et al., 2018
UFMG-CM-Y634	TOC 1301		Cachaça	Tocantins, Brazil		CACHAÇA-BIOETHANOL	4	21251		5/5		NF	A881del		NF	11bpdel		✓	x	x	x			PRJEB24932	Barbosa et al., 2018
BG-1			Bioethanol-producing strain from sugar cane	Brazil		CACHAÇA-BIOETHANOL	4	n.d.		4/5		NF	A881del		NF		20bp ins	x	x	x	x			PRJNA352845	Coutouné et al., 2017
M.9.1			Cachaça	Brazil		CACHAÇA-BIOETHANOL	4	26964		5/5		NF		V121M	NF	11bpdel		x	✓	x	x			ERP014555	Peter et al., 2018
M1.1			Cachaça	Brazil		CACHAÇA-BIOETHANOL	4	23740		5/5		NF	A881del		NF	11bpdel		x	✓	x	x			ERP014555	Peter et al., 2018
BE063			Beer	England		BEER 2	5	n.d.	15/15	5/5		NF		V121M	F			✓	x	x	x			MBXR00000000	Gallone et al., 2016
CBS 1398		*Saccharomyces thermantitonum*	Leaf of *Eucalyptus* sp.	Unknown		BEER 2	5	16595	15/15	5/5		NF		V121M	NF	11bpdel		✓	x	x	x			ERP014555	Peter et al., 2018
CBS 381	PYCC 4566	*Saccharomyces willianus*	Spoiled beer	Japan		BEER 2	5	12832				NF		V121M	NF		G25del	✓	x	x	x			PRJEB36095	This study
CBS 422	PYCC 4572	*Saccharomyces odessa*	Beer	Ukraine		BEER 2	5	1771	15/15			NF	A881del	V121M	NF	11bpdel		✓	x	x	x			ERP014555	Peter et al., 2018
PYCC 4602	CBS 382	*Saccharomyces brasiliensis*	Logos brewer's	Rio de Janeiro, Brazil		BEER 2	5	20999	15/15	5/5		NF	A881del		NF	11bpdel	G25del	✓	x	x	✓	F	AA	PRJEB36095	This study
CBS 1173		*Saccharomyces ilicis*	Fruit of *Ilex aquifolium*	Unknown		BEER 2	5	12355				NF	A822ins	V121M	NF	11bpdel		✓	x	x	x			ERP014555	Peter et al., 2018
CBS 2444	PYCC 4574		Seagrams distillery	USA		BEER 2	5	41509				NF	A881del	V121M	NF	11bpdel		✓	x	x	✓	F	AA	PRJEB36095	This study
BE092			Beer (Strong ale)	Belgium		BEER 2	5	n.d.	15/15			NF	A881del		F			✓	x	✓	x			MBWO00000000	Gallone et al., 2016
TUM 71			Spoilage yeast Isolated from brewery	Germany, State North-Rhine Westphalia (Nordrhein-Westfalen)		BEER 2	5	20555	15/15	5/5		NF		V121M	NF	11bpdel		✓	x	✓	✓	F	AF I	PRJEB36095	This study
BE013			Beer (Ale refermentation)	Belgium		BEER 2	5	n.d.		5/5		NF	A881del		F			✓	✓	✓	x			MBZP00000000	Gallone et al., 2016

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]			AQY2[3]			RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C														
BE083			Beer (Saison)	Belgium		BEER 2	5	n.d.		5/5		NF	A881del		NF	11bp del		✓	✓	✓	x			MBWX00000000	Gallone et al., 2016
TUM PI-BA-45			Spoilage yeast Isolated from brewery	Germany, State Lower Saxony (Niedersachsen)		BEER 2	5	25787		5/5		NF		V121M	NF		786 G>A stop	✓	✓	✓	x			PRJEB36095	This study
TUM PI-BA-109			Spoilage yeast Isolated from brewery	Germany, State North-Rhine Westphalia (Nordrhein-Westfalen)		BEER 2	5	21268		5/5		NF	A881del		NF	11bp del		✓	x	✓	x			PRJEB36095	This study
TUM PI-BA-31			Spoilage yeast Isolated from brewery	Germany, State Rhineland-Palatinate (Rheinland-Pfalz)		BEER 2	5	27650		DUP		NF		V121M	NF	11bp del		✓	x	✓	x			PRJEB36095	This study
TUM PI-BB-105			Spoilage yeast Isolated from beer-mixed beverage	Unknown		BEER 2	5	23206	15/15	5/5		NF	A881del		NF	11bp del		✓	✓	✓	x			PRJEB36095	This study
TUM 3-D-2			Spoilage yeast Isolated from brewery	Germany, State Lower Saxony (Niedersachsen)		BEER 2	5	28706		5/5		NF	A881del		F			✓	x	✓	x			PRJEB36095	This study
TUM 3-H-2			Spoilage yeast Isolated from brewery	Germany, State Lower Saxony (Niedersachsen)		BEER 2	5	28485		5/5		NF	A881del		F			✓	x	✓	x			PRJEB36095	This study
BE021			Beer (Pale ale)	Canada		BEER 2	5	n.d.		5/5		NF	A881del		F			✓	x	✓	x			MBZH00000000	Gallone et al., 2016
BE032			Beer	England		BEER 2	5	n.d.		5/5	4/19	NF	A881del		NF	11bp del		✓	x	✓	x			MBYW00000000	Gallone et al., 2016
CBS 6505			Brewery (killer strain responsible for death of brewing yeast)	UK		BEER 2	5	2212		5/5		NF	A881del		NF	11bp del		✓	x	✓	x			ERP014555	Peter et al., 2018
CBS 7958			Factory producing cassava flour	Sao Paulo, Brazil		BEER 2	5	2183		5/5	19/19	NF	A881del		F			x	x	✓	x			ERP014555	Peter et al., 2018
PYCC 2608	CBS 1782	*Saccharomyces diastaticus*	Super-attenuated beer	Unknown		BEER 2	5	2770		5/5	19/19	NF	A881del		F			x	x	✓	x			PRJEB36095	This study
TUM 1-B-8			Spoilage yeast isolated from brewery	Germany, State Bavaria (Bayern)		BEER 2	5	4543		5/5	4/19	NF	A881del		F			x	x	x	x			PRJEB36095	This study
CH14			Pearl millet beer	Abengourou, Ivory Coast		AFRICAN BEER	6	27976		5/5		F			NF		424 C>T Stop	x	x	x	x			ERP014555	Peter et al., 2018
CLIB 651			Beer leaven for Bili Bili beer, brewery	Chad		AFRICAN BEER	6	16287				F			NF		424 C>T Stop	x	x	x	x			ERP014555	Peter et al., 2018
MAJ_G			Plant material	Majunga, Madagascar		AFRICAN BEER	6	7094				F			NF		GT158ins	✓	x	x	✓	F	AF I	ERP014555	Peter et al., 2018
N26-1-10:7-5(a)			Maize dough	Ghana		AFRICAN BEER	6	3781				F			NF		424 C>T Stop	x	x	x	✓	F	AF I	ERP014555	Peter et al., 2018
PYCC 8031	NRRL Y-7184		Bantu beer	van der Walt, Pretoria, South Africa		AFRICAN BEER	6	50093			5/19	NF		321 G>A Stop	NF	G25del		x	✓	x	✓	F	EUR	PRJEB36095	This study
CBS 4454			Kaffir beer	South Africa		AFRICAN BEER	6	17328				NF		T498del	NF	11bp del		x	x	x	x			ERP014555	Peter et al., 2018
CBS 4456			Kaffir beer	South Africa		AFRICAN BEER	6	40939			5/19	NF		T498del	NF	11bpdel		x	x	x	x			PRJEB36095	This study

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]			AQY2[3]			RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C														
UFMG-CM-Y263	C41		*Tapirira guianensis*	Tocantins, Brazil	TROP	WB3 - FG -MA	7	1862				F			F			x	✓	x	✓	F	SA	PRJEB11698	Barbosa et al., 2016
NRRL YB-908			Wild cherry tree gum	Wickerham, NRRL GM	TROP	WB3 - FG -MA	7	2715				F			F			x	✓	x	x			PRJEB36095	This study
UFMG-CM-Y641	TOC 1345		Tapirira guianensis	Tocantins, Brazil	TROP	WB3 - FG -MA	7	2543				F			F			x	✓	x	✓	F	SA	PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y642	TOC 1365		Tapirira guianensis	Tocantins, Brazil	TROP	WB3 - FG -MA	7	1862				F			F			x	✓	x	✓	F	SA	PRJEB11698	Barbosa et al., 2016
CEY 619			Human feces	French Guiana	TROP	WB3 - FG -MA	7	10245		5/5		NF		T239del	F			x	✓	✓	x			ERP014555	Peter et al., 2018
HE 004			Human feces	French Guiana	TROP	WB3 - FG -MA	7	11945		5/5		NF	A822ins	T239del	F			x	✓	✓	x			ERP014555	Peter et al., 2018
HE 012			Human feces	French Guiana	TROP	WB3 - FG -MA	7	9103		5/5		NF	A822ins	T239del	F			x	✓	✓	x			ERP014555	Peter et al., 2018
LCBG-3D6			*Agave* spp. fermentation during rustic mezcal production	Tamaulipas, Mexico	TROP	WB3 - FG -MA	7	17358		5/5		NF		T230del	F			✓	✓	x	x			ERP014555	Peter et al., 2018
LCBG-3Y8			*Agave* spp. fermentation during rustic mezcal production	Tamaulipas, Mexico	TROP	WB3 - FG -MA	7	22105		5/5		NF		T230del	F	PAR NA		✓	✓	x	x			ERP014555	Peter et al., 2018
LCBG-3D2			*Agave* spp. fermentation during rustic mezcal production	Tamaulipas, Mexico	TROP	WB3 - FG -MA	7	17358		5/5		NF		T230del	NF		G107 del	✓	✓	x	x			ERP014555	Peter et al., 2018
MB 7c			*Quercus pubescens*	Montbarri, Southern France	TEMP	MO	8	3752				F			F			x	x	x	x			PRJNA264372	Almeida et al., 2015
ZP 560			*Quercus pyrenaica*	Castelo de Vide, Portugal	TEMP	MO	8	2553				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
ZP 848			*Quercus ilex*	Alter do Chão, Portugal	TEMP	MO	8	2656				F			F			x	x	x	x			PRJEB7675	Almeida et al., 2015
ZP 570			*Fraxinus* sp.	Paul Boquilobo, Portugal	TEMP	MO	8	2813				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
ZP 850			*Quercus ilex*	Alconorales Natural Park, Andaluzia, Spain	TEMP	MO	8	2869				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
ZP 541			*Fagus sylvatica*	Adagoi, Portugal	TEMP	MO	8	2684				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
3.10	DBVPG 10100		*Quercus cerris*	Parco del Monte Subasio, Italy	TEMP	MO	8	3800				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
DBVPG 10158			soil underneath *Quercus cerris*	Riserva Naturale Lucciolabella/Chianciano (SI) 2012	TEMP	MO	8	n.d.				F			F			x	x	x	x			PRJEB36095	This study
HUN 9.1s1	DBS 14		Oak	Hungary	TEMP	MO	8	2745				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
EXF 7200			*Quercus robur*	Jasenovo polje, Montenegro	TEMP	MO	8	3001				F			F			x	x	x	x			PRJNA264372	Almeida et al., 2015
OakRom 3-2a			Oak	Near Bucarest, Romania	TEMP	MO	8	2366				F			F			x	x	x	x			PRJEB7675	Almeida et al., 2015
PYR 4b	DBQ 26		*Quercus pubescens*	Halkidiki, Greece	TEMP	MO	8	2738				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]			AQY2[3]			RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]		GENOME DATA	REFERENCE
									A	B	C													
CBS 2421	PYCC 2503		Japanese kefyr grains	Unknown		DAIRY	9	27903	15/15	5/5		NF	A881del		NF	11bpdel		x	x	x	x		ERP014555	Peter et al., 2018
VKMY 373			Camembert cheese	France		DAIRY	9	n.d.				NF	A881del		NF	11bpdel		x	x	x	x		PRJEB7675	J. L. Legras et al., 2018
KS11			Kumiss	Kazakhistan		DAIRY	9	n.d.				NF	A881del		NF	11bpdel		✓	x	x	x		PRJEB7675	J. L. Legras et al., 2018
PYCC 6721	CBS 420; VKM Y-482	*Saccharomyces mangini*	Cheese	Italy		DAIRY	9	37556				NF	A881del	88 C>T Stop	NF	11bpdel		x	x	x	x		PRJEB36095	This study
PYCC 6720	CBS 437, NRRL Y-1545		Stracchino Cheese	Italy		DAIRY	9	23692				NF	V121M		NF		185 T>A Stop	x	x	x	x		PRJEB36095	This study
EXF 5871			Homemade cheese	Slovenia		DAIRY	9	27590				NF	A881del		NF		185 T>A Stop	x	x	x	x		ERP014555	Peter et al., 2018
CLIB 554			Dairy, brine	France		DAIRY	9	24196				NF	A881del		NF	11bpdel		x	x	x	x		ERP014555	Peter et al., 2018
CLIB 553			Dairy, raw milk	Normandy, France		DAIRY	9	10209				NF	A881del		NF		185 T>A Stop	x	x	x	x		ERP014555	Peter et al., 2018
MUCL 42908			Kefir	Morrocco		DAIRY	9	n.d.				NF	A881del		NF		185 T>A Stop	x	x	x	x		PRJEB7675	J. L. Legras et al., 2018
Morocco Bread G17	MBG17		Baker's yeast			BREAD	10	61910		5/5		NF	V121M	A822ins	F			✓	x	x	x		SRR403240	Dunn et al., 2012
PYCC 5324			Homemade corn and rye bread dough	Braga, Portugal		BREAD	10	53843		5/5		NF	V121M	A822ins	NF	11bpdel		✓	x	x	x		PRJEB36095	This study
SD-15			Sourdough levan	Matera, Italy		BREAD	10	50578		5/5		NF	V121M	A822ins	F			✓	x	x	x		ERP014555	Peter et al., 2018
PYCC 5320			Homemade corn and rye bread dough	Braga, Portugal		BREAD	10	54495		4/5		NF	V121M		NF	11bpdel		✓	x	x	x		PRJEB36095	This study
AP13.1			Fermented sour-dough	Cinfães, Portugal		BREAD	10	53091		5/5	17/19	NF	V121M		NF	11bpdel		✓	x	x	x		PRJEB24932	Barbosa et al., 2018
BR004			Bread	Belgium		BREAD	10	n.d.		5/5	17/19	NF	V121M		NF	11bpdel		✓	x	x	x		MBVV00000000	Gallone et al., 2016
CBS 2361			Trinidadian cane sugar	UK		BREAD	10	37044		5/5	17/19	NF	A881del		NF		786 G>A stop	✓	x	x	x		PRJEB36095	This study
CBS 1464		*Saccharomyces cerevisiae var. onychophilus*	Infected nail of 4 year old girl	Austria		BREAD	10	2407				NF	V121M	A822ins	NF	11bpdel		✓	x	x	x		ERP014555	Peter et al., 2018
NRRL YB-3916	CBS 2919	*Saccharomyces. veronae var. osloensis*	Sputum	Oslo, Norway		BREAD	10	44584		5/5		NF	A881del		NF	11bpdel		✓	x	x	x		PRJEB36095	This study
PYCC 4226			Commercial baker's yeast	Unknown		BREAD	10	30081		4/5	2/19	NF	A881del		NF		792stop	✓	x	x	x		ERS1108635	Gonçalves et al., 2016
Platinum			Commercial baker's yeast	Unknown		BREAD	10	42278		4/5	15/19	NF	V121M	A822ins	NF		792stop	✓	x	x	x		ERS1108633	Gonçalves et al., 2016
TUM 308			Alt beer	Rhineland-Palatinate, Germany		BEER 1	11	33689				NF	V121M		NF	11bpdel	792stop	✓	x	x	x		ERS1108619	Gonçalves et al., 2016
TUM 513			California ale	Unknown		BEER 1	11	33061				NF	V121M		NF	11bpdel	792stop	✓	x	x	x		ERS1108618	Gonçalves et al., 2016
TUM 174			Regional beer similar to Kölsch beer	Mülheim, Germany		BEER 1	11	21227				NF	V121M		NF	11bpdel	792stop	✓	x	x	x		ERS1108621	Gonçalves et al., 2016
PYCC 4455	CBS 1171; NRRL Y-12632	*Saccharomyces cerevisiae*	Brewer's top yeast	Oranjeboom brewery, Rotterdam, Netherlands		BEER 1	11	22852				NF	V121M		NF	11bpdel		✓	x	x	x		PRJEB36095	This study

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]			AQY2[3]			RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C														
TUM 205			Wheat Beer			BEER 1	11	41573				NF	V121M		NF	11bpdel		✓	x	x	x			ERS1108614	Gonçalves et al., 2016
BE074			Beer (Hefeweizen)	Germany		BEER 1	11	n.d.				NF	V121M		F			✓	x	x	x			MBXG00000000	Gallone et al., 2016
TUM 175			Wheat beer	Freising-Weihenstephan, Germany		BEER 1	11	29902				NF	V121M		F			✓	x	x	x			ERS1108615	Gonçalves et al., 2016
TUM 507			Ale Beer from wheatmalt	Unknown		BEER 1	11	40936				NF	V121M		F			✓	x	x	x			ERS1108616	Gonçalves et al., 2016
TUM 503			California ale	Unknown		BEER 1	11	30324				NF	V121M	A822ins	NF	11bpdel		✓	x	x	x			ERS1108626	Gonçalves et al., 2016
FostersO			Commercial brewing ale strain			BEER 1	11	34874				NF	V121M	A822ins	NF	11bpdel		✓	x	x	x			ERS1108627	Gonçalves et al., 2016
FostersB			Commercial brewing ale strain			BEER 1	11	43000				NF	V121M	A822ins	NF	11bpdel		✓	x	x	x			ERS1108629	Gonçalves et al., 2016
TUM 211			British ale / stout	Great Britain		BEER 1	11	26842				NF	V121M	A822ins	NF		792stop	✓	x	x	x			ERS1108632	Gonçalves et al., 2016
LM1-2			Fermented yak milk	Naqu, Tibet		CHN FERMENTATION	12	7132				NF	A881del		NF		185 T>A Stop	x	x	x	x			PRJNA396809	Duan et al., 2018
WM11-3			Fermented mixed goat and yak milk	Mongolia		CHN FERMENTATION	12	7118				NF	A881del	88 C>T Stop	NF	11bpdel		x	x	x	x			PRJNA396809	Duan et al., 2018
5-3004			Fermented mare's milk (kumis)	Ulaanbaatar, Mongolia		CHN FERMENTATION	12	4978				NF	A881del		NF	11bpdel		x	x	x	x			PRJNA396809	Duan et al., 2018
QH21-3			Fermented yak milk	Delingha, Qinghai		CHN FERMENTATION	12	7574				NF	V121M		NF	11bpdel		✓	x	x	x			PRJNA396809	Duan et al., 2018
JM13.2			Fermenting dough	Yuncheng, Shanxi		CHN FERMENTATION	12	10349				NF	V121M		NF	G25del		✓	✓	x	x			PRJNA396809	Duan et al., 2018
WL1			Fermenting dough	Wulanchabu, Inner Mongolia		CHN FERMENTATION	12	10255				NF	V121M		NF	G25del		✓	✓	x	x			PRJNA396809	Duan et al., 2018
BT2.8			Fermenting dough	Baotou, Inner Mongolia		CHN FERMENTATION	12	9069				NF	V121M		NF	G25del		✓	x	x	✓	F	AS I	PRJNA396809	Duan et al., 2018
BT3			Fermenting dough	Baotou, Inner Mongolia		CHN FERMENTATION	12	7761				NF	V121M		F			✓	x	x	✓	F	R	PRJNA396809	Duan et al., 2018
QH6.5			Fermenting dough	Minhe, Qinghai		CHN FERMENTATION	12	9259				NF	V121M		n.d.			x	✓	x	x			PRJNA396809	Duan et al., 2018
CJM19.5			Fermenting dough	Yuncheng, Shandong		CHN FERMENTATION	12	10592				NF	V121M		NF	G25del		x	✓	x	x			PRJNA396809	Duan et al., 2018
QH1.1			Fermenting dough	Qinghai		CHN FERMENTATION	12	3753				NF	V121M		F			x	✓	x	x			PRJNA396809	Duan et al., 2018
XZ2.1			Daqu (starter) of Qingkejiu (highland barley wine)	Lasa, Tibet		CHN FERMENTATION	12	3232				NF	V121M		NF	G25del		x	✓	x	x			PRJNA396809	Duan et al., 2018
JQ9.4			Fermenting dough	Nanyang, Henan		CHN FERMENTATION	12	7371				NF	V121M		NF	G25del		✓	✓	x	x			PRJNA396809	Duan et al., 2018
CQJZ1.4			Fermented rice	Shaoxing, Zhejiang		CHN FERMENTATION	12	540				NF	V121M	A822ins	NF	G25del		x	✓	x	x			PRJNA396809	Duan et al., 2018
HQ1.1			Huangjiu (rice wine) starter with fermented monascus	Gutian, Fujian		CHN FERMENTATION	12	3753				NF	V121M	A863del	NF	G25del		✓	✓	x	✓	F	AS I	PRJNA396809	Duan et al., 2018

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]			AQY2[3]			RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C														
GZLJ3.19			Fermented sorghum grain	Changshun, Guizhou		CHN FERMENTATION	12	3807				NF	V121M	A822ins	NF	G25del		x	✓	x	x			PRJNA396809	Duan et al., 2018
HLJ225			Fermented sorghum grain	Qiqihaer, Heilongjiang		CHN FERMENTATION	12	5146				NF	V121M	A822ins	F			x	✓	x	✓	F	AS I	PRJNA396809	Duan et al., 2018
JQ8.3			Fermenting dough	Chengcheng, Shanxi		CHN FERMENTATION	12	3302				NF	V121M		NF	G25del		x	✓	x	✓	F	AS I	PRJNA396809	Duan et al., 2018
JT49.1			Dry fermented dough	Zhouzhi, Shanxi		CHN FERMENTATION	12	7371				NF	V121M	A822ins	NF	G25del		x	✓	x	x			PRJNA396809	Duan et al., 2018
LJQ13.C7			Fermenting dough	Yanan, Shanxi		CHN FERMENTATION	12	3044				NF	V121M		NF	G25del		x	✓	x	x			PRJNA396809	Duan et al., 2018
CBS 1544	PYCC 4654	*Saccharomyces fructuum*	Fermenting fruit juice	Netherlands				5165		5/5	19/19	NF	A881del		NF	11bpdel		✓	x	x	x			PRJEB36095	This study
MTF 2546			Cocoa bean fermentation	West Africa		COCOA	13	14517		5/5		NF	A881del		F			x	x	x	x			ERP014555	Peter et al., 2018
MTF 2548			Cocoa bean fermentation	West Africa		COCOA	13	23290		5/5		NF	A881del		F			x	x	x	x			ERP014555	Peter et al., 2018
MTF 2551			Cocoa bean fermentation	West Africa		COCOA	13	21583		5/5		NF	A881del		F			x	x	x	✓	NF	AF II	ERP014555	Peter et al., 2018
BJ10			Bark of Ulmus sp., secondary forest	Wuling Mountain, Beijing	TEMP	CHN VIII	14	2086		5/5		NF	A881del		F			x	x	x	x			PRJNA396809	Duan et al., 2018
BJ2			Apple, orchard	Shunyi, Beijing	TEMP	CHN VIII	14	2647		5/5		F			F			x	x	x	x			PRJNA396809	Duan et al., 2018
CBS 1598			Sake moto	Nakazawa, Japan		SAKE	15	3517				NF	V121M		NF	G25del		✓	✓	x	x			ERS1108639	Gonçalves et al., 2016
CBS 435	PYCC 8194	*Saccharomyces tokyo*	Sake moto	Nakazawa, Japan		SAKE	15	4906				NF	V121M		NF	G25del		✓	✓	x	x			ERS1108637	Gonçalves et al., 2016
UC 5	UCD612		Sene sake	Kurashi, Japan		SAKE	15	n.d.				NF	V121M		NF	G25del		x	✓	x	x			PRJNA60197	Justin Fay, Washington University
Kyokai-no.7	K7		Japanese sake brewerie	Japan		SAKE	15	n.d.				NF	V121M	A822ins	NF	G25del		x	TRIP	x	x			PRJNA45827	Akao et al., 2011
CBS 1508	PYCC 5182	*Saccharomyces carlsbergensis var. mandschuricus*	Starter for sorghum brandy	Unknown		SAKE	15	1898				NF	V121M	A822ins	NF	G25del		✓	✓	x	✓	F	AS I	ERP014555	Peter et al., 2018
CBS 440		*Saccharomyces formosensis*	Molasses	Taiwan		SAKE	15	2410		5/5		NF	A881del		NF	G25del		x	✓	x	x			ERP014555	Peter et al., 2018
Y12	NRRL Y-12663, CBS 400		Fermentation (palm wine)	Ivory Coast		SAKE	15	15534				NF	V121M		NF	G25del		✓	✓	x	x			SGRP2	Bergström et al., 2014
CLIB 1414			Rice wine	Laos		SAKE	15	11848				NF	V121M	A822ins	NF	G25del		✓	x	x	x			ERP014555	Peter et al., 2018
MTJZ1			Fermented sorghum grain	Renhuai, Guizhou, China		SAKE	15	2515				NF	V121M		NF	G25del		x	✓	x	✓	F	AS I	PRJNA396809	Duan et al., 2018
YJSH1			Bioethanol-producing strain	China		SAKE	15	3084				NF	V121M		NF	G25del		x	✓	x	✓	F	AS I	PRJNA72403	Zheng et al., 2012
NRRL Y-12844			Budod (rice-based starter)	Phillipines		SAKE	15	10363				NF	V121M		NF	G25del		x	✓	x	x			ERS1108645	Gonçalves et al., 2016
NRRL Y-7327	PYCC 8032		Tibetan beer starter	H. Douglas, University of Washington, Seattle, Washington		SAKE	15	31305				NF	V121M	C625ins	NF	G25del		✓	✓	x	x			PRJEB36095	This study

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]			AQY2[3]			RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C														
NRRL Y-6297	PYCC 8030		Coconut tuba (palm wine)	Hipolito, NIST, Manila, Philippines	TROP	PHILIPPINES	16	3044				NF	V121M		NF	G25del		✓	✓	x	x			PRJEB36095	This study
YJM 1479	NRRL Y-6297		Coconut tuba (palm wine)	Philippines	TROP	PHILIPPINES	16	3264				NF	V121M		NF	G25del		✓	✓	x	x			PRJNA189929	Strope et al., 2015
NRRL YB-4083			Coconut sap	Philippines	TROP	PHILIPPINES	16	n.d.				NF	V121M		NF	G25del		✓	✓	x	x			PRJEB36095	This study
YJM 1400	NRRL YB-4081		Guava (fruit)	Philippines	TROP	PHILIPPINES	16	2960				NF	V121M		NF	G25del		x	✓	x	x			PRJNA189911	Strope et al., 2015
CBS 1576			Exudate of *Arenga* sp. (palm tree)	Celebes, Sulavesi, Indonesia	TROP	PHILIPPINES	16	2576				NF	V121M		F			✓	x	x	x			ERP014555	Peter et al., 2018
NRRL Y-5508	PYCC 8029		Coconut pod	Arguelles, Manila, Philippines	TROP	PHILIPPINES	16	19036				NF	V121M		F			✓	✓	x	x			PRJEB36095	This study
DJ66			Palm wine	Yoboki, Djibouti	TROP	PHILIPPINES	16	19012				NF	T178 del	A822ins	F			DUP	✓	x	✓	NF	AF II	ERP014555	Peter et al., 2018
S8BM-30-2D			Phloem sap of *Caryota urens* (Toddy palm)	Sri Lanka	TROP	PHILIPPINES	16	3172				NF	V121M		F			✓	✓	x	✓	F	AA	ERP014555	Peter et al., 2018
CBS 2992			Palm wine	Pakistan	TROP	PHILIPPINES	16	2405				NF	V121M	A822ins	NF	G25del		✓	x	x	x			PRJEB36095	This study
CBS 3000			Palm wine	Pakistan	TROP	PHILIPPINES	16	2325				NF	V121M		NF	G25del		✓	✓	x	x			PRJEB36095	This study
SDO 3s1			Oak	North Carolina, USA	TEMP	NA & JP - CHN VI/VII - FER	17	3693				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
YPS 1009			Oak exudate	New Jersey, USA	TEMP	NA & JP - CHN VI/VII - FER	17	2650				F			F			x	x	x	x			PRJNA60223	Justin Fay, Washington University
YPS 163			*Quercus rubra*	Pennsylvania, USA	TEMP	NA & JP - CHN VI/VII - FER	17	2673				F			F			x	x	x	x			PRJNA28813	Doniger et al., 2008
ZP 651			*Quercus acutissima*	Chiba Prefecture, Japan	TEMP	NA & JP - CHN VI/VII - FER	17	3623				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
IY_03-5-30-1-1-1_(1)			*Chauliodes pectinicornis* (male) - Corydalidae	Louisiana, USA	TEMP	NA & JP - CHN VI/VII - FER	17	2782				F			F			x	x	x	x			ERP014555	Peter et al., 2018
ZP 652			*Quercus acutissima*	Chiba Prefecture, Japan	TEMP	NA & JP - CHN VI/VII - FER	17	3699				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
ZP 779			*Quercus acutissima*	Hirusen highland, Okayama Prefecture, Japan	TEMP	NA & JP - CHN VI/VII - FER	17	3509				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
ZP 781			*Quercus serrata*	Hirusen highland, Okayama Prefecture, Japan	TEMP	NA & JP - CHN VI/VII - FER	17	3333				F			F			x	x	x	x			PRJEB7601	Almeida et al., 2015
BJ28			Bark of *Castanea* sp., secondary forest	Wuling Mountain, Beijing	TEMP	NA & JP - CHN VI/VII - FER	17	8648		3/5		F			F			x	x	x	x			PRJNA396809	Duan et al., 2018
FJ12			Peach from a market	Wuyi Mountain, Fujian	TEMP	NA & JP - CHN VI/VII - FER	17	3074				F			F			x	x	x	x			PRJNA396809	Duan et al., 2018
HLJ2			Bark of *Quercus mongolica*, secondary forest	Jingbo lake, Heilongjiang	TEMP	NA & JP - CHN VI/VII - FER	17	2803				F			F			x	x	x	x			PRJNA396809	Duan et al., 2018
N163.01-5A			Exudate of *Quercus mongolica*	Botanical garden, Blagoveshchensk, Russia	TEMP	NA & JP - CHN VI/VII - FER	17	3049				F			F			x	x	x	x			ERP014555	Peter et al., 2018

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]		AQY2[3]		RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C												
N3.00-7A			Grape berries of wild *Vitus amurensis*	Botanical garden, Blagoveshchensk, Russia	TEMP	NA & JP - CHN VI/VII - FER	17	2432				F		F		x	x	x	x			ERP014555	Peter et al., 2018
N159.01-2B			Exudate of *Quercus mongolica*	Botanical garden, Blagoveshchensk, Russia	TEMP	NA & JP - CHN VI/VII - FER	17	2863				F		F		x	x	x	x			ERP014555	Peter et al., 2018
SD1			Persimmon, planted tree	Anqiu, Shandong	TEMP	CHN VI / CHN VII	18	2095				F		F		x	x	x	x			PRJNA396809	Duan et al., 2018
SX10			Persimmon, planted tree	Qinling Mountain, Shaanxi	TEMP	CHN VI / CHN VII	18	2509				F		F		x	x	x	x			PRJNA396809	Duan et al., 2018
XJ1			Soil, orchard	Xinjiang	TEMP	CHN VI / CHN VII	18	2880				F		F		x	x	x	x			PRJNA396809	Duan et al., 2018
CBS 403	PYCC 4571	*Saccharomyces lindneri*	Ginger beer from *Zingiber officinale*	West Africa		WEST AFRICA	19	2031				NF	A881del	F		✓	✓	x	✓	NF	AF II	ERP014555	Peter et al., 2018
YJM 1439	NCYC 110		Ginger beer from *Zingiber officinale*	West Africa		WEST AFRICA	19	2129				F		F		✓	✓	x	✓	NF	AF II	PRJNA189920	Strope et al., 2015
PYCC8489	NRRL Y-12633; CBS 400	*Saccharomyces chevalieri*	Palm wine from *Elaeis guineensis*	Ivory Coast		WEST AFRICA	19	3572				NF	A881del	F		✓	✓	x	✓	NF	AF II	PRJEB36095	This study
YJM 1248	NRRL Y-1546		Bili wine from *Osbeckia grandiflora*	West Africa		WEST AFRICA	19	2600				NF	A881del	F		✓	✓	x	✓	NF	AF II	PRJNA189888	Strope et al., 2015
PW5			Raphia palm wine	Aba, Abia state, Nigeria		WEST AFRICA	19	1175				NF	A881del	F		✓	x	x	✓	NF	AF II	PRJNA60181	Justin Fay, Washington University
YJM 195	NCYC 762		Palm wine	Nigeria		WEST AFRICA	19	2977				NF	A881del	F		✓	✓	x	x			PRJNA189849	Strope et al., 2015
CBS 7903	NCYC 761		Palm wine	Nigeria		WEST AFRICA	19	2737				NF	A881del	F		✓	✓	x	x			PRJEB36095	This study
MUCL_30909-2C			Cassava	Burundi		WEST AFRICA	19	3847		5/5		F		NF	11bpdel	✓	x	x	x			ERP014555	Peter et al., 2018
YN6			Bark of *Castanopsis orthacantha*, suburban tree	Zixi Mountain, Yunnan				2463				F		F		x	x	x	x			PRJNA396809	Duan et al., 2018
CLQCA 20-156			Flower (*Gesneriaceae* sp.)	Yasuni, Orellana, Ecuador	TROP	WB1 - CHN V	20	2991				F		F		x	x	x	✓	F	AS I	ERP014555	Peter et al., 2018
CLQCA 20-314			Termite mound	Yasuni, Orellana, Ecuador	TROP	WB1 - CHN V	20	3468				F		F		x	x	x	✓	F	AS I	ERP014555	Peter et al., 2018
UFMG-CM-Y455	C45		*Tapirira guianensis*	Tocantins, Brazil	TROP	WB1 - CHN V	20	4167				F		F		x	x	x	✓	F	AS I	SA	PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y461	C145		Mushroom	Roraima, Brazil	TROP	WB1 - CHN V	20	3675				F		F		x	x	x	✓	F	AS I	SA	PRJEB11698	Barbosa et al., 2016
UFMG-CM-Y650	TOC1364		*Tapirira guianensis*	Tocantins, Brazil	TROP	WB1 - CHN V	20	2373				F		F		x	✓	x	✓	F	AS I	SA	PRJEB11698	Barbosa et al., 2016
ZJ1			Rotten wood, primeval forest	Gutian Mountain, Zhejiang province, China	TROP	WB1 - CHN V	20	929				F		F		x	✓	x	✓	F	AS I		PRJNA396809	Duan et al., 2018
HN15			Rotten wood	Wuzhi Mountain, Hainan province, China	TROP	WB1 - CHN V	20	3077				F		F		x	✓	x	✓	F	AS I		ERP014555	Peter et al., 2018
HN16			Soil	Wuzhi Mountain, Hainan province, China	TROP	WB1 - CHN V	20	2933				F		F		x	✓	x	✓	F	AS I		ERP014555	Peter et al., 2018

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]		AQY2[3]		RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]			GENOME DATA	REFERENCE
									A	B	C												
GT 39.1			Bark, primeval forest	Gutian Mountain, Zhejiang province, China	TEMP	CHN X	21	969				F		F		x	✓	x	✓	F	AS II	PRJNA396809	Duan et al., 2018
JXXY1.1			Bark, primeval forest	Xiangxiyuan, Shennongjia, Hubei province, China	TEMP	CHN X	21	955				F		F		x	✓	x	✓	F	AS II	PRJNA396809	Duan et al., 2018
ZP 1273			Soil	New Caledonia	TROP	CHN X	21	3776				F		F		✓	x	x	x			PRJEB36095	This study
YJM 1447	UWOPS 05-227.2		Bertram palm	Malaysia	TROP	MALAYSIA	22	n.d.				NF	106bp del	NF	G528del	✓	✓	x	✓	F	AS I	PRJNA189923	Strope et al., 2015
UWOPS 03-433.3			Nectar of Bertram palm	Malaysia	TROP	MALAYSIA	22	1908				NF	106bp del	NF	G528del	✓	✓	x	✓	F	AS I	ERS1108647	Gonçalves et al., 2016
UWOPS 03-461.4			Nectar of Bertram palm	Malaysia	TROP	MALAYSIA	22	18636				NF	106bp del	NF	G528del	✓	✓	x	✓	F	AS I	SGRP2	Bergström et al., 2014
HN10			Rotten wood	Wuzhi Mountain, Hainan province, China	TROP	CHN III	23	4751				F		F		x	✓	x	✓	F	AS I	ERP014555	Peter et al., 2018
HN19			Bark from tree in Fagaceae family	Wuzhi Mountain, Hainan province, China	TROP	CHN III	23	2896				F		F		x	✓	x	✓	F	AS I	ERP014555	Peter et al., 2018
HLJ1			Bark of *Quercus mongoLica*, secondary forest	Jingbo lake, Heilongjiang	TEMP	CHN IV	24	2865				F		F		x	x	x	x			PRJNA396809	Duan et al., 2018
N51-1A			Exudate of *Quercus mongolica*	Vladivostok, Russian Far East	TEMP	CHN IV	24	3074				F		F		x	x	x	x			ERP014555	Peter et al., 2018
BJ21			Bark of *Carya cathayensis*, secondary forest	Dongling Mountain, Beijing	TEMP	CHN IV	24	3286				F		F		x	x	x	x			PRJNA396809	Duan et al., 2018
BJ14			Forest soil, secondary forest	Wuling Mountain, Beijing	TEMP	CHN IV	24	2922				F		F		x	x	x	x			PRJNA396809	Duan et al., 2018
BJ6			Persimmon	Changping, Beijing, China	TEMP	CHN IV	24	3068				F		F		x	x	x	x			ERP014555	Peter et al., 2018
HN2			Rotten wood, primeval forest	Diaoluo Mountain, Hainan province, China	TROP	CHN I	25	3283				F		F		x	✓	x	✓	F	AS I	PRJNA396809	Duan et al., 2018
HN6			Rotten wood, primeval forest	Wuzhi Mountain, Hainan province, China	TROP	CHN I	25	2780				F		F		x	✓	x	x			ERP014555	Peter et al., 2018
JXXY16.1			Bark, primeval forest	Xiangxiyuan, Shennongjia, Hubei province, China	TEMP	CHN IX	26	1234				F		F		x	✓	x	x			PRJNA396809	Duan et al., 2018
XXY30L.2			Bark, primeval forest	Xiangxiyuan, Shennongjia, Hubei province, China	TEMP	CHN IX	26	3191				F		F		x	✓	x	x			PRJNA396809	Duan et al., 2018
XXYS1.4			Bark, primeval forest	Xiangxiyuan, Shennongjia, Hubei province, China	TEMP	CHN IX	26	992				F		F		x	✓	x	x			PRJNA396809	Duan et al., 2018
SX3			Bark from *Carya* sp.	Qinling Mountain, Shaanxi province, China	TEMP	CHN II	27	2833				F		F		x	x	x	x			ERP014555	Peter et al., 2018
SX1			Persimmon, planted tree	Qinling Mountain, Shaanxi province, China	TEMP	CHN II	27	2509				F		F		x	x	x	x			ERP014555	Peter et al., 2018

Strain	Other designations	Original species designation (correspondent type strain)	Substrate	Geographic location	Clime Zone (wild /feral pop.)	Phylogeny	Clade	Heterozygosity[1]	Regions[2]			AQY1[3]		AQY2[3]		RTM1[4]	BIO1 BIO6[5]	STA1[6]	MEL[7]		GENOME DATA	REFERENCE
									A	B	C											
SX8			Bark of *Quercus fabri*, primeval forest	Qinling Mountain, Shaanxi province, China	TEMP	CHN II	27	2602				F		F		x	x	x	x		PRJNA396809	Duan et al., 2018

[1] Count of heterozygous sites across the genome. Values higher than 20,000 are highlighted.

[2] Three genomic regions horizontally transferred from non-Saccharomyces yeasts (Novo et al., 2009) that encompass genes potentially relevant for winemaking process.

[3] Aquaporins (encoded by *AQY1* and *AQY2*) are membrane water channels that facilitate the transport of water in and out of the cell. Their inactivation is viewed as a consequence of domestication.

[4] Subtelomeric gene associated with the locus of sucrose utilization that provides resistance to inhibitory compounds present in molasses.

[5] These genes encode enzymes involved in the synthesis of biotin and were originally found in strains used in sake fermentation.

TROP -Tropical
TEMP -Temperate

CHN - China
FER - Far East Russia
FG - French Guiana
MA - Mexican agavae fermentation
MO - Mediterranean oak
NA & JP - North America and Japan
PDM - Prise de mousse
WB - Wild Brazil

F Functional
NF Not Functional

x Absent
✓ Present

✓ ✓ *S. cerevisiae*
✓ *S. paradoxus*
✓ Unknown

AS I - ASIAN I
AS II - ASIAN II
AA - ASIAN - AMERICAN
R - RECOMBINANT ASII-AA
EUR - EUROPEAN
AF I - AFRICAN I
AF II - AFRICAN II
SA - SOUTH AMERICAN

AQY1 INACTIVATIONS

A881del	adenine deletion at position 881
V121M	valine to methionine at aminoacid 121
A822ins	adenine insertion at position 822
321 G>A Stop	guanine to adenine at position 321 leads to a premature stop codon
T498del	thymine deletion at position 498
T239 del	thymine deletion at position 239
T230 del	thymine deletion at position 230
88 C>T Stop	cytosine to thymine at position 88 leads to a premature stop codon
A863del	adenine deletion at position 863
C625ins	cytosine insertion at position 625
T178 del	thymine deletion at postion 178
106bp del	deletion of the first 106 bp

AQY2 INACTIVATIONS

11bp del	11 bp deletion
G25del	guanine deletion at position 25
20bp ins	20 bp insertion
786 G>A stop	guanine to adenine at position 786 leads to a premature stop codon
424 C>T Stop	cytosine to thymine at position 424 leads to a premature stop codon
GT158ins	guanine and thymine insertion at position 158
G107 del	guanine deletion at position 107
185 T >A Stop	thymine to adenine at position 185 leads to a premature stop codon
792stop	premature stop codon at position 792
G528del	guanine deletion at position 528

[6] Chimeric gene that gives rise to an extra-cellular glucoamylase that allows the conversion of soluble starch and dextrins into fermentable sugars.

[7] Polymorphic genes that encode an α-galactosidase that hydrolyses melibiose.

Table A3.2 - Presence (+) or absence (-) of the *MEL* genes in the genomes of *S. paradoxus* from different populations.

STRAIN	ORIGIN	POPULATION	MEL
DBN13	Greece	EUROPEAN	+
DBN14	Greece	EUROPEAN	+
DBN15	Greece	EUROPEAN	+
TAX15c	Greece	EUROPEAN	-
DBN10	Greece	EUROPEAN	-
TAX3d	Greece	EUROPEAN	-
TAX2	Greece	EUROPEAN	-
PAR10c	Greece	EUROPEAN	-
TAX1a	Greece	EUROPEAN	-
ZP 1157	Portugal	EUROPEAN	-
ZP 1023	Portugal	EUROPEAN	-
ZP600	Portugal	EUROPEAN	-
CET10308	Spain	EUROPEAN	-
ZP 1167	Spain	EUROPEAN	-
ZP 1233	Spain	EUROPEAN	-
MB13a	France	EUROPEAN	-
MB11d	France	EUROPEAN	-
MB15	France	EUROPEAN	-
MB4	France	EUROPEAN	-
Chr116	France	EUROPEAN	-
DBVPG 4650	Italy	EUROPEAN	-
CRO2d	UK	EUROPEAN	-
OAK11c	UK	EUROPEAN	-
CRO12d	UK	EUROPEAN	-
Q59	UK	EUROPEAN	-
Q89	UK	EUROPEAN	-
DBF01	UK	EUROPEAN	-
CBS 406	Netherlands	EUROPEAN	-
ZP 614	Germany	EUROPEAN	-
ZP 665	Germany	EUROPEAN	-
EM11	Germany	EUROPEAN	-
EM13	Germany	EUROPEAN	-
EM55	Germany	EUROPEAN	-
EM213	Germany	EUROPEAN	-
NBRC 102004	Finland	EUROPEAN	-
ALKO2639	Finland	EUROPEAN	-
NBRC 102003	Lithuania	EUROPEAN	-
CBS 5829	Denmark	EUROPEAN	-
CBS 432	Unknown	EUROPEAN	-
EXF 6618	Slovenia	EUROPEAN	-
EXF 6641	Slovenia	EUROPEAN	-
EXF 6729	Slovenia	EUROPEAN	-
EXF 6635	Slovenia	EUROPEAN	-
INMIV11	Ukraine	EUROPEAN	-

STRAIN	ORIGIN	POPULATION	MEL
INMIV544	Ukraine	EUROPEAN	-
ZP 703	USA	NORTH AMERICAN	-
YPS 125	USA	NORTH AMERICAN	-
YPS 138	USA	NORTH AMERICAN	-
YPS 644	USA	NORTH AMERICAN	-
UCD-FST 61-248	USA	NORTH AMERICAN	+
UCD 51-137	USA	NORTH AMERICAN	-
UCD-72-129	USA	NORTH AMERICAN	-
UWOPS 1979-129	Canada	NORTH AMERICAN	-
UWOPS 1979-527	Canada	NORTH AMERICAN	-
UFRJ 50816	Brazil	NORTH AMERICAN	-
NRLLY2047	USA	NORTH AMERICAN	-
UWOPS 91-917-1	Hawaii	HAWAIIAN	+
CBS 8440	Russia	FAR EAST	-
CBS 8443	Russia	FAR EAST	-
KPN 3829	Siberia	FAR EAST	-
IFO 1804	Japan	FAR EAST	-
UCD-67-570W	Japan	FAR EAST	-
N16	Russia	FAR EAST	-
N43	Russia	FAR EAST	-
N44	Russia	FAR EAST	-
N12	Azerbeijan	FAR EAST	-
ZP 811			-
ZP 751			-
ZP 838			-
DBH58			-

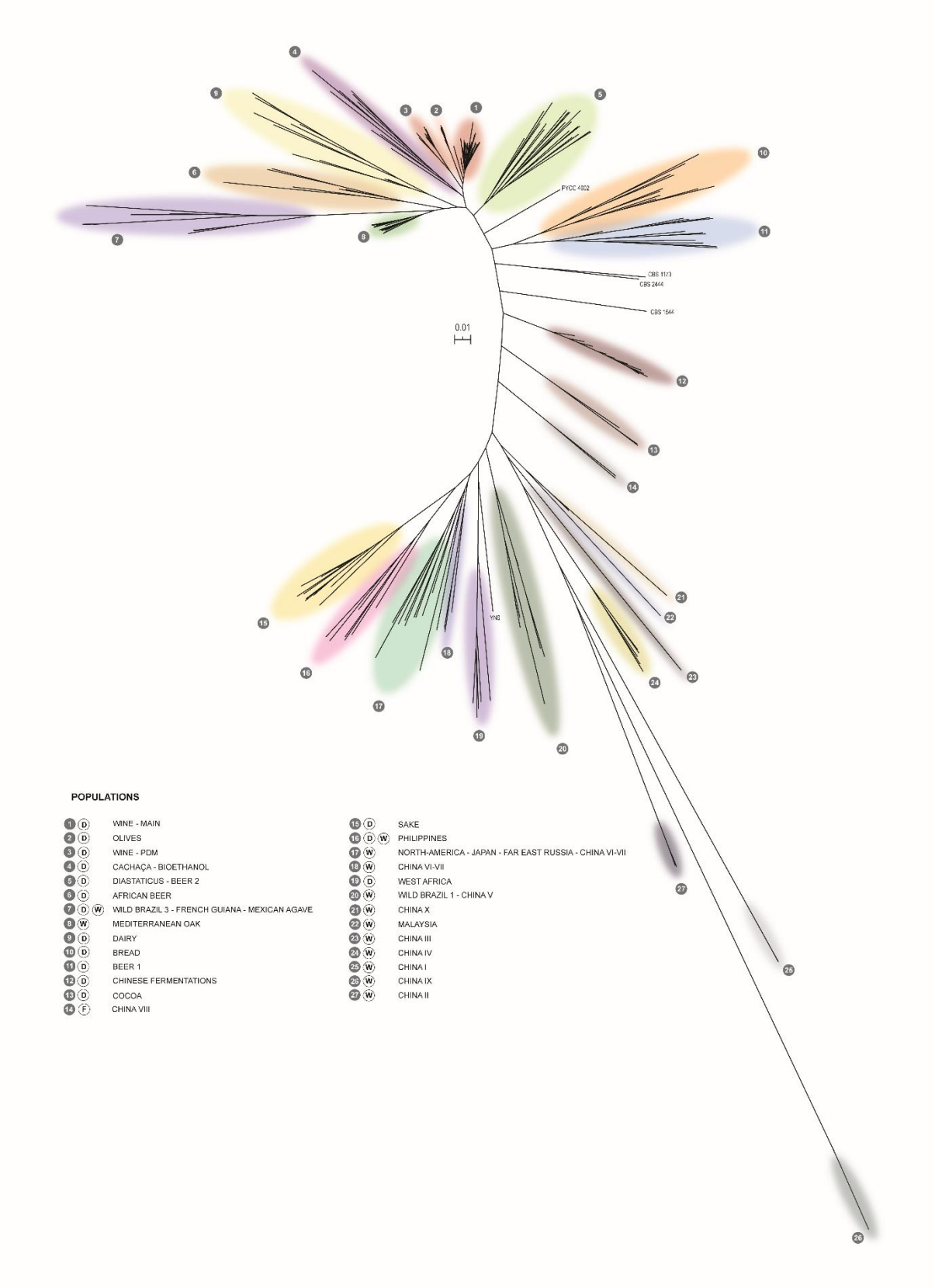


Figure A3.1 - Unrooted phylogeny *S. cerevisiae* genomes representing the known diversity of the species. The phylogeny was inferred from 247 sequences and 983,701 SNPs using TVM+F+G4 model of sequence evolution and the maximum likelihood method as implemented in IQ-TREE. Branch lengths correspond to the expected number of substitutions per site. The 27 clades detected are numbered in gray circles.

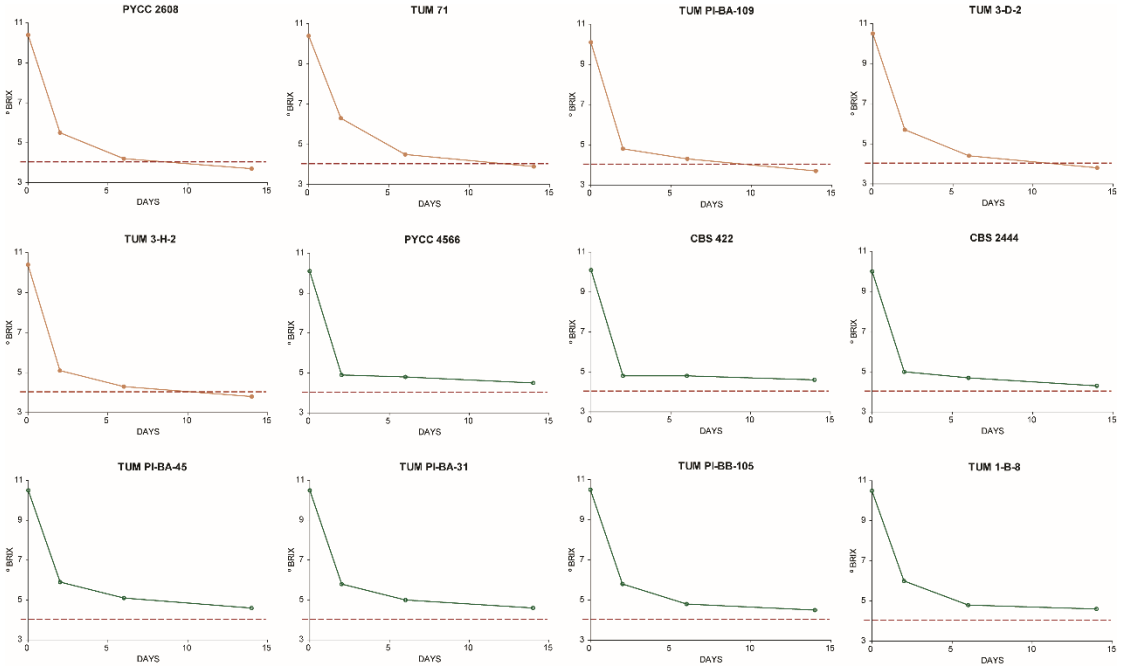


Figure A3.2 - °Brix measurements during growth in beer wort for 15 days at 25 °C. Strains able to attenuate the initial 11 °Brix of the medium to 4 °Brix or lower were considered as diastase positive.

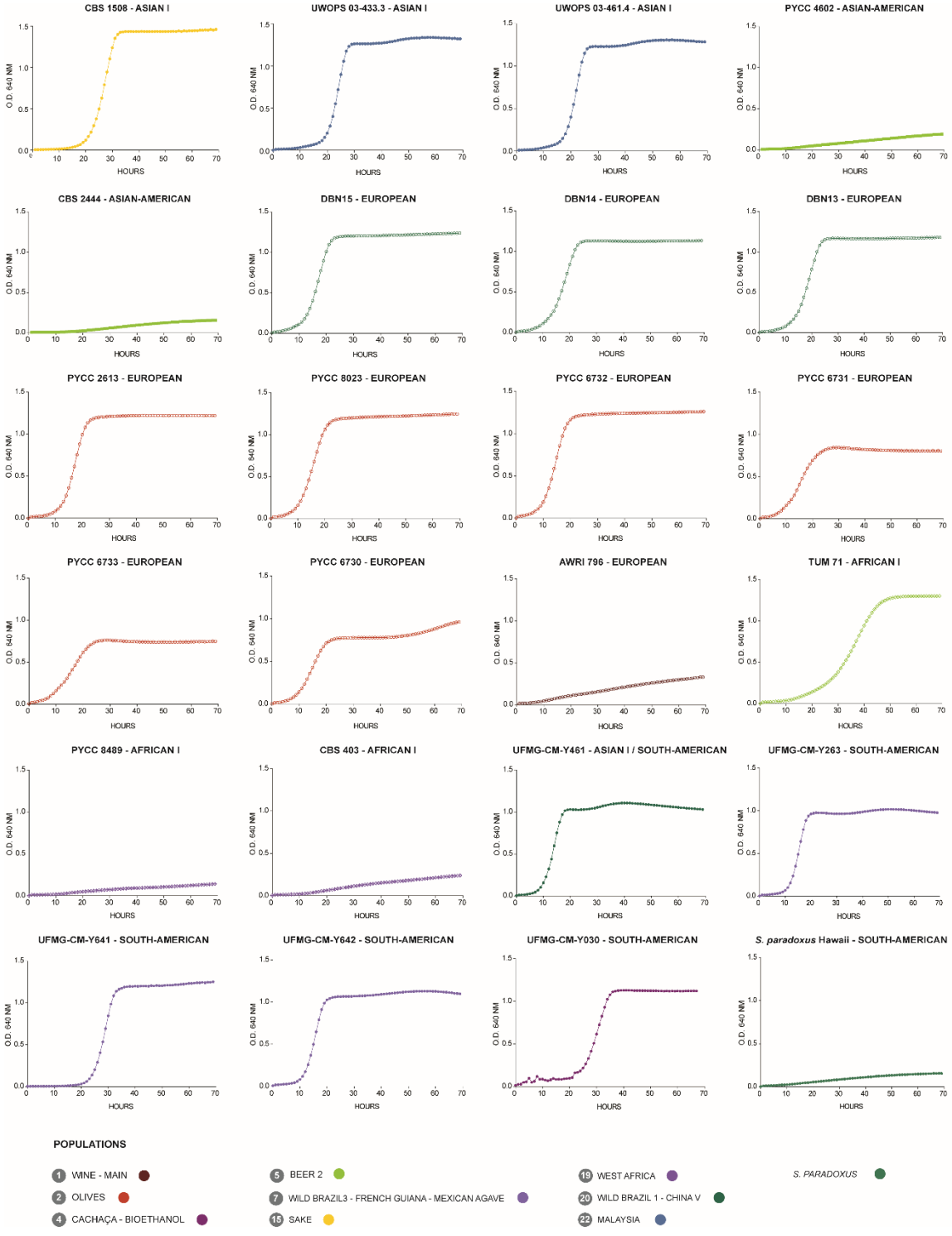


Figure A3.3 - Growth curves on 1% (w/v) melibiose medium supplemented with yeast nitrogen base and incubated at 25 °C for three days.