CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL
INSTITUTO POLITÉCNICO NACIONAL

Unidad de Genómica Avanzada
Unidad Irapuato

Population genomics and ecology of yeast associated with traditional Agave fermentation in Mexico

Tesis que presenta

Claudio Emilio López Gallegos

para obtener el Grado de

Maestro en Ciencias

en Biología Integrativa

Codirectores de la Tesis:

Dr. Alexander de Luna Fors

Dr. Eugenio Mancera Ramos

Irapuato, Gto

Marzo 2023

El siguiente trabajo de Tesis de Maestría se desarrolló en el laboratorio de Sistemas Genéticos a cargo del Dr. Alexander de Luna Fors en la Unidad de Genómica Avanzada y en el laboratorio de Evolución Molecular a cargo del Dr. Eugenio Mancera Ramos en la Unidad Irapuato, ambos pertenecientes al Centro de Investigación y de Estudios Avanzados del IPN (Cinvestav), Irapuato, Guanajuato, México. Se contó con el apoyo del CONACYT a través del programa de becas para estudios de posgrado (CVU): 1076441.

Index

Index....2
Acknowledgments....4
Abstract....5
Resumen....6
Introduction....8
Saccharomyces cerevisiae is a model organism with a complex natural history....8
The population structure of *S. cerevisiae* is determined by its ecology and biogeography....9
S. paradoxus is the sister species of *S. cerevisiae* and is commonly found in natural environments....10
Insects work as vectors for yeast dispersal....11
Recurrent hybridization in *Saccharomyces* yeasts....12
Spontaneous *Agave* fermentations are an open and diverse ecological niche for wild *Saccharomyces* yeasts....13
Objectives....15
General aim....15
Specific aims....15
Material and Methods....15
Sampling....15
Locations....15
Insects Sampling....19
Plants Sampling....19
Fermentation tank Sampling....19
Sampling of objects in distilleries....20
Data collection of field samples....20
Yeast isolation from field samples....20
Taxonomic identification of isolates....21
Mass spectrometry analysis....21
Association analyses....21
DNA extraction and genome sequencing....21
Bioinformatic analyses....22
Quality control and genome mapping....22
Variant Calling and Genotyping....22

Phylogenetic reconstruction of the sequenced strains.....23
Results.....23
A large collection of yeast isolates from agave fermentation and their surrounding natural environments throughout different geographical regions.....23
The composition of *Saccharomyces* species is different between distilleries and natural environments.....24
S. cerevisiae is the dominant *Saccharomyces* species in distilleries.....26
S. paradoxus is more frequently isolated from the bark of Fagaceae trees.....26
S. cerevisiae has a ubiquitous distribution in natural substrates.....27
Genomic and phylogenetic analysis of *S. cerevisiae* genomes.....29
S. cerevisiae phylogeny mostly reproduces its geographical origin.....29
S. cerevisiae isolates from objects, insects, and plants in distilleries are the same genetic pool as from fermentation tanks.....29
Three different sympatric populations of *S. paradoxus* were isolated from natural substrates in Mexico.....32
A novel *S. paradoxus* population associated with inter-species hybridization with *S. cerevisiae*.....34
Discussion.....37
S. cerevisiae is scarce but ubiquitous in the wild and insects may work as vectors for its dispersion.....37
Domesticated and “feral” *Saccharomyces* strains are involved in *Agave* fermentations in Mexico.....39
The *Agave* fermentation environment could promote the formation of hybrids.....41
Conclusions.....43
References.....45

Figures and Tables

Figure 1.....17
Figure 2.....18
Figure 3.....25
Figure 4.....28
Figure 5.....31
Figure 6.....33
Figure 7.....36
Table 1.....16
Table 2.....24

Acknowledgments

En primera instancia, quiero agradecer a las personas que han hecho posible el desarrollo de esta tesis. Más que a las instituciones públicas, que si bien han canalizado los recursos para que esta investigación se lleve acabo, quiero resaltar al pueblo que con sus impuestos hace esto posible. En segunda instancia y no menos importante, agradecer a todos los productores mexicanos de mezcal, raicilla, bacanora, sotol y destilados de agave en general. Gracias por abrirnos las puertas a sus fábricas de mezcal, por compartir su conocimiento, darnos la confianza y por brindar con nosotros y hacer esta investigación posible. Espero que el conocimiento aquí generado encuentre su cause para retribuirles su valiosa aportación a este proyecto.

También quiero agradecer a todos los colegas que ayudaron en la planeación y ejecución del trabajo de campo, de laboratorio y en los análisis bioinformáticos para obtener los resultados. Mención especial a Joanna Serwatowska, Susana Ruiz y Mayra Flores por la gestión administrativa para conseguir materiales e insumos necesarios para los muestreos y experimentos; a Abraham Avelar por sus aportaciones y retroalimentación en la planeación y ejecución de prácticamente todas y cada una de las fases del proyecto; a Karen, Marco, Vanessa, Toño Hernández y Antonio Basurto, por su ayuda en los muestreos y la genial convivencia durante los mismos; a Martín y Mau por su apoyo en el laboratorio; a Manuel Kirchmayer, por facilitarnos el uso de su laboratorio en el CIATEJ Zapopan; a Lucía Morales, por su apoyo y recursos brindados; a Luis Fer por su asesoría bioinformática; a mi comité de asesoras, Laila Partida, Kasia Oktaba y Xitlali Aguirre, por su retroalimentación; a mis tutores Alex y Eugenio, a quienes les tengo gran admiración y aprecio, por guiarme en este proyecto.

De igual manera, agradezco a aquellas personas que han hecho mi estancia en Irapuato una experiencia inesperadamente agradable, y que en general han aportado momentos de goce y alegría en mi vida, durante los últimos dos años y medio. En especial a TaV, por todo su apoyo y porque Irapuato no hubiera sido lo mismo sin ella. A mis amigos de la generación y a todas las personas con las que conviví y que compartieron parte de su transcurrir en estas tierras freseras conmigo.

Finalmente, quiero agradecer a mi familia, que en el ámbito personal me ha apoyado y motivado a cumplir mis sueños. Particularmente mi madre, por siempre estar ahí, a mi hermana por la sana convivencia y momentos de ocio y a mi hermano, que siempre llevaré en mi memoria, por haber sido mi persona favorita en este planeta.

Abstract

Saccharomyces yeasts are fundamental in alcoholic fermentations, such as the production of *Agave* spirits in Mexico. We know that *S. cerevisiae*, the baker's yeast, is the dominant species in *Agave* fermentations. However, in artisanal production it is unknown how and from where these microorganisms reach the fermentation tanks since producers do not use an inoculum. *Agave* fermentation occur "spontaneously" by microorganisms from the environment. Outside of the tanks, *S. cerevisiae* and its sister species *S. paradoxus* are believed to inhabit the bark of oaks (*Quercus* spp.) and other trees, and that some insects may function as vectors for its dispersal. The current research aims to understand the interaction between populations of *Saccharomyces* yeasts from natural environments and agave fermentation in Mexico, to unravel the ecological origin of the yeasts involved in the production of agave spirits. For this, we sampled distilleries, their surrounding natural environments, and areas isolated from humans, in four different states of Mexico: Oaxaca, Durango, Tamaulipas, and Nuevo Leon. Overall, we collected 876 samples from tree barks, insects, objects in distilleries, and fermentation tanks. Through selective enrichment for species of the genus *Saccharomyces*, we obtained more than 4,000 isolates and identified 73% of these at the species level using mass spectrometry. Strains of *S. cerevisiae* represented 86% of all the isolates identified as *Scharomyces,* while 14% were *S. paradoxus.* Remarkably, no other *Saccharomyces* species was isolated. *S. paradoxus* was found more frequently in natural environments, particularly in the bark of *Quercus* trees and was over represented in the Northwest of the country. Of these *Saccharomyces* strains we sequenced the genomes of 114 isolates and identified 24 that are hybrids between *S. cerevisiae* and *S. paradoxus* with varied degree of genomic contribution from each parental species. Phylogenetic analyzes revealed that strains of *S. cerevisiae* isolated from fermentation vats are the same genetic population as those isolated from insects, plants, and objects sampled within the distilleries. We also found two known populations of *S. paradoxus* in Mexico (SpA and SpB), and a previously unknown one that is probably a new subpopulation of the SpB clade. To our knowledge, our study is the first survey of wild *Saccharomyces* strains in Mexico at the genomic level, offering evidence that there are natural reservoirs inside distilleries and possible associations between yeasts and insects in the production of mezcal. This research contributes to better understand the biodiversity of microorganisms involved in the artisanal production of distilled beverages in Mexico and their natural reservoirs, with possible implications for their conservation and biotechnological use.

Resumen

Las levaduras del género *Saccharomyces* son fundamentales en las fermentaciones alcohólicas, como en la producción de destilados de agave en México. Sabemos que *S. cerevisiae*, la levadura común, es la especie dominante en las fermentaciones de agave. Sin embargo, en la producción artesanal se desconoce cómo y de dónde llegan estos microorganismos a los tanques de fermentación ya que los productores no utilizan un inóculo. Las fermentaciones de agave ocurren "espontáneamente" por microorganismos del ambiente. Fuera de los tanques, se cree que *S. cerevisiae* y su especie hermana *S. paradoxus* habitan en la corteza de los encinos o robles (*Quercus* spp.) y otros árboles, y que algunos insectos pueden funcionar como vectores en su dispersión. La presente investigación tiene como objetivo comprender la interacción entre las poblaciones de levaduras *Saccharomyces* de ambientes naturales y las fermentaciones de agave en México, para desentrañar el origen ecológico de las levaduras involucradas en la producción de destilados de agave. Para ello, muestreamos destilerías, sus alrededores y áreas naturales lejos de actividades humanas, en cuatro estados de México diferentes: Oaxaca, Durango, Tamaulipas y Nuevo León. En total, recolectamos 876 muestras de cortezas de árboles, insectos, objetos en destilerías y tanques de fermentación. Mediante enriquecimiento selectivo para especies del género *Saccharomyces*, obtuvimos más de 4000 aislados e identificamos el 73% de estos a nivel de especie mediante espectrometría de masas. El 86% de los aislados identificados como *Saccharomyces* fueron *S. cerevisiae* y el 14% *S. paradoxus*, no se aisló ninguna otra especie de *Saccharomyces. S. paradoxus* se encontró con mayor frecuencia en ambientes naturales, particularmente en la corteza de los árboles *Quercus* y estuvo sobre representada en el Noroeste del país. De estas cepas de *Saccharomyces*, secuenciamos los genomas de 114 aislados e identificamos 24 que son híbridos entre *S. cerevisiae* y *S. paradoxus* con diferentes proporciones del genoma de las especies parentales. Los análisis filogenéticos revelaron que las cepas de *S. cerevisiae* aisladas de tinas de fermentación pertenecen al mismo clado filogenético que las *S. cerevisiae* aisladas de insectos, plantas y objetos muestreados dentro de las destilerías. También encontramos dos poblaciones de *S. paradoxus* en México, que habían sido previamente descritas en Norte América (SpA y SpB) y una previamente desconocida que probablemente sea una nueva subpoblación del clado SpB. Hasta donde sabemos, nuestro estudio es la primera caracterización a nivel genómico de cepas silvestres de *Saccharomyces* en México, que ofrece evidencia sobre la existencia de reservorios naturales dentro de las destilerías y posibles asociaciones entre levaduras e insectos en la producción de mezcal. Esta investigación

contribuye a conocer mejor la biodiversidad de microorganismos involucrados en la producción artesanal de bebidas destiladas en México y sus reservorios naturales, con posibles implicaciones para su conservación y uso biotecnológico.

Introduction

Saccharomyces cerevisiae is a model organism with a complex natural history

Saccharomyces cerevisiae, the baker's yeast, has been used by humans since the dawn of civilizations. This organism was probably responsible for wine fermentation already by 3,150 B.C. (Cavalieri *et al.*, 2003). During the last century, this yeast has become a model organism in cell and molecular biology; it was the first whole-genome sequenced eukaryotic organism and it is probably the best annotated genome to date (Goffeau *et al.*, 1996; Goddard, 2015). Nevertheless, the natural history of *S. cerevisiae* is poorly understood. Therefore, many researchers have begun studies on population genomics and ecology of *S. cerevisiae* and its sister species *S. paradoxus* (Landry *et al.,* 2019; Da-Yong *et al.,* 2021; Madden *et al.,* 2018, Charron *et al.*, 2014; Allik, Miller & Greig, 2015). Since the 1990s wild isolates of *S. cerevisiae* have been reported from different natural environments (Sniegowsky *et al.*, 2002), but these isolates were thought to be migrants from human environments. This comes from the idea that *S. cerevisiae* is a domesticated species only found in human environments, and as a domesticated species, its evolution was associated specifically to fermentations of alcoholic beverages. Nevertheless, as scientists investigated the evolutionary origin of *S. cerevisiae*, domesticated and wild populations were found (Fay & Benavides, 2005).

Fermentation is the most common habitat where *S. cerevisiae* is found, in a wide range of beverages that can vary in chemical composition, selecting for different yeast strains. But this is not the only habitat of the common yeast. Natural environments like insect guts and bark trees host *S. cerevisiae.* In addition, this organism can be part of the human microbiota, and there are also pathogenic strains for humans (Liti, 2015). Wang *et al.* (2012) performed an extensive field survey of *S. cerevisiae* in China revealing its ubiquitous distribution in natural environments. In this survey, the general success isolation rate of *S. cerevisiae* was 10.6 % from different sample types. Nevertheless, the success rate was higher in bark of oak trees (16.5 %), suggesting an association of *S. cerevisiae* with Fagaceae trees (Sampaio & Goncalves, 2016) as its sister species *S. paradoxus.* Despite these associations, the overall ubiquitous distribution of *S. cerevisiae* is consistent with the neutral model of ecology for species distribution, being a nomadic microbe rather than living in a specific niche (Goddard, 2015).

How yeast survives in the wild remains unclear. Perhaps the environmental stress-

resistant meiotic spores resulting from the sexual cycle help yeasts to survive in natural environments with limited nutrients, remaining latent until more favorable conditions occur. Duan *et al.* (2018) have found that isolates from natural environments are mostly homozygous while almost all the domesticated isolates from fermentation environments are heterozygous, and heterozygosity is negatively correlated with sporulation and spore viability rates. In fact, homozigocity in wild strains could be because enrichment procedures may cause that rare spore germinate and become diploid after mate-type switching (Knight & Goddard, 2016). On the other hand, industrial strains of *S. cerevisiae* have lower sporulation efficency, partially because of aneuploidies (Fischer, Liti & Llorente, 2021).

The population structure of S. cerevisiae is determined by its ecology and biogeography

The pattern of genetic differentiation between lineages of *S. cerevisiae* is correlated with their degree of human association, geography, and ecology (Peter *et al*, 2018). According to their ecological conditions of growth, domesticated yeasts from liquid state fermentation (LSF), like wine, beer, and mezcal; and solid state fermentation (SSF), like bread and cheese, form two different major clades in the species phylogeny. In turn, these two domesticated groups of *S. cerevisiae* form a monophyletic group different from wild strains (Han *et al.,* 2021). An out-of-China origin for the species is mainly supported by two evidence. First, domesticated populations cluster in a monophyletic clade and are distinct from wild populations, which have the longest branches of the phylogenetic tree. Second, there is more genetic divergence and diversity in the wild than in domesticated strains. Actually, there are lineages in the wild clade with three times the genetic diversity of the entire domesticated populations (Duan *et al*, 2018; Fay & Benavides, 2005; Han *et al.,* 2016). The Taiwanese lineage described by Peter *et al.* (2018) is the most divergent population with an average of 1.1% sequence divergence to non-Taiwanese strains. Whether the domesticated lineages originated from single or multiple domestication events is still under debate and the mechanism causing the diversification of the wild lineages also remains to be clarified (Bai *et al.*, 2022).

Into the LSF, SSF, and Wild major clades of the *S. cerevisiae* species phylogeny, there are 24 clades, a mosaic group, and several subclades, which correlates with its isolation region and fermentation type (Liti *et al.*, 2009, Peter *et al.*, 2018). China is the region with more lineages and with more genetic diversity of the species (Alasmmar & Delneri, 2020). About the origin of domesticated strains, Duan *et al.* (2018) suggest a unique bottleneck event in the evolutionary history of the domesticated population from the wild

population of *S. cerevisiae* in China. On the other hand, Almeida *et al.* (2015) have suggested that the wild Mediterranean Oak population of *S. cerevisiae* is the ancestral lineage of the Wine/Europe domesticated clade, and Ludlow *et al.* (2016) also suggested that strains involved in coffee and cacao fermentation have independent origins. Wild strains have been isolated mostly from Asia. Nevertheless, there are reports of a wild lineage isolated from North America (Peter *et al.,* 2018), and Barbosa *et al.* (2016) reported a distinct South American wild population isolated from natural environments in Brazil. In the South American continent Peter *et al.* (2018) and Han *et al.* (2021) have reported domesticated and human-related isolates from the French Guyana and Mexican Agave fermentations as distinct lineages of *S. cerevisiae* that belong to the major clade of LSF.

It is possible that the habitat of *S. cerevisiae* 8,000 years ago was restricted to forests and the species consisted of biogeographically well-defined populations around the world, all derived from the Asian lineage, the most ancestral population. Then, after the discovery of fermentation by humans, the history of the common baker's yeast was profoundly marked by domestication, both genetically and biogeographically (Eberlein *et al.*, 2015).

S. paradoxus is the sister species of S. cerevisiae and is commonly found in natural environments

Some of the species of the *Saccharomyces* genus have experienced a long history of domestication while others remain in natural environments as wild species (Eberlein, 2019). This is the case of *S. paradoxus*, the sister species of *S. cerevisiae,* considered to be mainly undomesticated. Althougth *S. paradoxus* is rarely found associated with fermentations, there are recent investigations on its beer brewing potential (Nikulin *et al.,* 2020). However, *S. paradoxus* is an ideal subject for studies about the ecology and natural history of yeast*,* being also the genetically closest species to *S. cerevisiae*. *S. paradoxus* has been isolated from natural environments since the last century, most commonly from oak trees (Sniegowski, 2002, Phaff *et al.,* 1955), although it was identified as S*accharomyces douglasii* or just as *Saccharomyces* sp. In recent decades some researchers have named the South American isolates of *S. paradoxus* as *S. cariocanus,* however, this is not commonly accepted (Boynton and Greig, 2014).

The first *S. paradoxus* population described was the clade from Europe (SpA), and later reports showed that *S. paradoxus* isolated from North America belong to a different population (SpB) than European strains (Naumov *et al.,* 1998). Nowadays, *S. paradoxus*

strains are classified into two major subpopulations: European, East Russian and Japanese (SpA), and North American (SpB). However, recent studies also place together *S. cariocanus* (synn. *S paradoxus* from South America) with the North American subpopulation (Liti, Barton & Louis, 2008). Furthermore, most recent analyses, split the North American clade into subpopulations SpB, SpC, SpC*, and SpD. Being the subpopulations SpB and SpC native from America, and SpC* and SpD the result of hybridization process between different subpopulations. The SpB clade has a population substructure that is consistent with its broad geographic distribution in North America (Charron, Leducq & Landry, 2014; Leducq *et al*., 2016; Eberlein *et al.*, 2019).

The ecological interaction between *S. paradoxus* and oak trees is not fully understood. *S. paradoxus* grew well on a sterile medium made from oak bark and has negative and positive interaction with *Pseudomona* spp. and *Mucilaginibacter* spp. respectively, both of which are part of the microbiome of oak trees. Interestingly, the density of *S. paradoxus* on the oak trees is about 1.87 cells per cm^2. It is a rare species and its survival depends on the interactions with the microbial community of the oak tree (Kowallik *et al.,* 2015). Leducq *et al.* (2014) reviewed the climatic adaptations of *Saccharomyces* species and proposed that *S. paradoxus* can grow at lower temperatures than its sister species *S. cerevisiae*. That said, there are differences in temperature preference between the *S. paradoxus* subpopulations (Leducq *et al.*, 2016). However, *S paradoxus* is not a cryotolerant species, and is sympatric in nature with its sibling cryotolerant species *S. uvarum* (Goncalves *et al.*, 2011) and also with *S. cerevisiae* (Naumov *et al.*, 1998). The same authors proposed that temperature plays a key role in the coexistence of these two species in North America by contributing to niche divergence and therefore enabling opposing competitive exclusion. Another survey found a season-dependent isolation success in North America, with an increased isolation rate from August to September and from decaying fruits (Charron *et al.* 2014). Dashko *et al.* (2016) sampled vineyards and non-vineyard locations in Slovenia to characterize the distribution and abundance of *Saccharomyces* yeasts and concluded that there is not a clear-cut difference in the abundance and distribution of *S. cerevisiae* and *S. paradoxus* within Slovenian vineyards and forests.

Insects work as vectors for yeast dispersal

It has been proposed that insects work as vectors for yeast dispersal (Gilbert, 1980), and this is because yeast cells are unable to disperse by themselves (Mortimer & Polsinelli, 1999). There are reports in the literature about fruit flies (Drosophilidae), bees and

wasps (Hymenoptera), and even beetles (Coleoptera) hosting *Saccharomyces* yeasts (Meriggi *et al.*, 2020).

Insects may also influence the propensity of sexual reproduction in yeasts. For example, Reuter, Bell & Greig (2007) found that yeast dispersal by *Drosophila melanogaster* increased more than ten-fold the outbreeding rates of *S. cerevisiae*. The sexual cycle of *Saccharomyces* yeasts is triggered by adverse environmental conditions, like starvation. Meiosis produces a tetrad of four stress resistance haploid spores (2 a, and 2 α), covered by an ascus. Each spore has a spore wall with four layers: two inner polysaccharide layers composed of *β*-glucan and α-mannan, a central chitosan layer, and an outermost layer of cross-linked dityrosine (Feldmann, 2012). The inner layer of the spore is related to the vegetative cell wall, and the two outer layers confer much of the spore's resistance to environmental damage, like insect digestion. It is believed that the outer spore layers of *S. cerevisiae* are specifically adapted to survive the digestion of their vectors, like fruit flies. Digestion of the ascus sac may provide nutritional value to the insect host, promoting yeast dispersal from one substrate to another (Collucio *et al.*, 2004; Coluccio *et al.*, 2008). It is thought that flies are attracted to substrates inhabited by yeasts by volatile metabolites produced by yeasts marking specific carbon sources. Some authors have suggested that yeast cells eaten by flies die in the digestive tract, and that yeasts are rather dispersed by cells added to the legs of fruit flies (Christiaens *et al*., 2014).

The observation that the passage through the insect gut increases yeast outbreeding has also been observed with wasps; these insects have been shown to be an important reservoir for yeast during the winter period in wine fermentation. *Polistes* wasps not only can host *S. cerevisiae* cells over more than three months, but they’re also able to transfer these microorganisms to their progenie by the habit of feeding their larvae through regurgitation of the content of a small part of their digestive tract, the crop (Stefanini *et al.*, 2012; Stefanini *et al.*, 2016). It has been observed that the hibernation of yeasts in wasps also improves interspecific mating between *S. cerevisiae* and *S. paradoxus*. The frequency of interspecific yeast hybrids seem to increase during long periods (4 months) of wasp hibernation (Stefanini *et al.,* 2016).

Recurrent hybridization in Saccharomyces yeasts

Hybridization between species is a common process with drastic evolutionary outcomes. Through gene flow, hybridization can avoid divergence between species, but on the other side, it can lead to speciation by combining independently evolved genomes and

conferring advantageous phenotypes and reproductive isolation from the parents (Leducq *et al.,* 2016). There are many examples of hybrid vigor or heterosis in yeast interspecific hybrids. The best understood is the case of *S. pastorianus*, a hybrid between *S. cerevisiae* and *S. eubayanus* that emerged 500-600 years ago and whose most important trait is cryotolerance in lager beer fermentations (Monerawela & Bond, 2017). In addition, five other cases of hybridizations apart from *S. pastorianus* are known in *Saccharomyces* yeasts from fermentative environments and some involve more than two parental species: *S. cerevisiae* × *S. kudriavzevii*; S. *eubayanus* × *S. uvarum*; *S. cerevisiae* × *S. kudriavzevii* × *S. eubayanus* × *S. uvarum*; *S. cerevisiae* × *S. eubayanus* × *S. uvarum*; and *S. cerevisiae* × *S. kudriavzevii* × *S. eubayanus*. All hybrid and introgressed strains that involve *S. eubayanus* and *S. uvarum* were formed between strains from the Holarctic lineage of each species. The genomes of *S. kudriavzevii* in hybrids also cluster specifically in a monophyletic clade with the European population (Langdon *et al.*, 2019).

Natural hybridization is common in wild yeast. The subpopulations of *S. paradoxus* SpC* and SpD are derived from natural hybridization events, but within the *S. paradoxus* species. The SpC* clade is derived from the hybridization between strains from the SpC and SpB subpopulations and presents partial postzygotic reproductive isolation with its parents. The SpD linage originated from a recent hybridization between SpC* and SpB, showing partial reproductive isolation with other lineages (Leducq *et al.*, 2016; Eberlein *et al.*, 2019).

Spontaneous Agave fermentations are an open and diverse ecological niche for wild Saccharomyces yeasts

In Mesoamerica, native people use to prepered and drink intoxicating beverage from different plants like cactus, cornstalks, mesquite pods, sap from agaves, and spruted maize (Bruman, 2000). Fermentation of several species of *Agave* has been carried on since 200-550 A.D. with the production of pulque, an alcoholic beverage from the fermented sap of mature *Agave* plants (Correa-Ascencio *et al.*, 2014). However, there is no evidence of distillation in Mesoamerica before Europeans invaded America. It is believed that agave distillation started in Colima, using *A. angustifolia* and an adaptation of the coconut spirit distillation established in Mexico by Philippine people around the 16th century (Zizumbo-Villarreal & Colunga-GarcíaMarín, 2007). There is great variety of agave spirits produced in Mexico, and at least four of them have adquired denomination of origin, which are raicilla, bacanora, tequila, and mezcal. The NOM-070-SCFI-2016, established the regulation for Mezcal production in Mexico, and the regions from where it

can be named like that. Currently, the authorized areas include the entire states of Oaxaca, Zacatecas, Durango, Guerrero and San Luis Potosi and some municipalities of Michoacán, Tamaulipas, Guanajuato, Puebla, Morelos, Estado de Mexico, and, very recently, Sinaloa (Arellano-Plaza, 2022).

In the last 10 years mezcal production has increased its yield 8 times officially, however, estimations considering other destilled agave beverages and personal communication from producers have calculated that agave spirits production is twice the officially reported (Arellano-Plaza *et a*l., 2022). The increase in the production of agave spirits and its profits has caused an excessive use of natural resources (Hernández-López, 2020), and changes in the social interactions between producers, distribution intermediaries, and global markets. Some producers make efforts to maintain their cultural identity, natural resources and traditional techniques of production while growing their enterprises and placing their products into the global markets (Arellano-Plaza *et a*l., 2022). Most of the conservation effort is focused on replanting agaves and trees used as firewood, but microorganisms are usually not in consideration, despite their importance in *Agave* fermentation.

In general terms, agave spirits production consists of five stages: Harvesting of the agave cores, cooking the cores, ground and juice extraction, fermentation of the must, and distillation. In traditional production processes, fermentation is open, occuring “spontaneously” by the activity of native microorganisms and without an inoculum provided by the producer (Arellano-Plaza *et al.*, 2022). Cooked agave juices and must are very rich in carbohydrates allowing growth of a variety of microorganisms. However, they are poor in nitrogen and contain a variety of secondary compounds such as saponins, 5-(hydroxymethyl)furfural (HMF), and furfural, which are inhibitory for yeasts and other microorganisms (Alcazar, 2017). Preliminary analyses have shown that *S. cerevisiae* strains isolated from agave fermentations in Mexico are resistant to HMF, and furfural, growing better than strains from the other regions of the world in the presence of these chemical compounds (Gallegos-Casillas, 2020).

Considering that yeast associated with the agave fermentation have unique phenotyipical traits, like the HMF resistence, it is possible that these microorganisms are adapted to the agave fermentation environment. However, it remains cryptic where do they live when there is no active fermentation and how do they arrive to the new batches. One possibility, which we will adress in this project, is that insects may work as vector for yeast dispersal from natural reservoirs to fermentation tanks. On the other

hand, it is also possible that yeast survive in the fermentation tanks between fermentation batches. Both are valid hypotheses that are not mutually exclusive.

Objectives

General aim

- To identify the vectors and natural reservoirs of *Saccharomyces* yeasts associated with traditional agave fermentation.

Specific aims

- To identify if there are families of plants working as natural reservoirs of the *Saccharomyces* yeasts associated with traditional agave fermentation.
- To determine the presence of *Saccharomyces* yeasts associated with traditional agave fermentation in insects in the distilleries and natural environments in their surroundings.
- To determine if *Saccharomyces* yeasts isolated from traditional agave fermentation and those isolated from natural environments around the distilleries are the same genetic population.

Material and Methods

Sampling

Locations

We performed sampling in Mexico during 2021, in the states of Oaxaca, Durango, Tamaulipas, and Nuevo Leon (fig. 1, and table 1). We defined two location types, the *distilleries, or fermenting places* (Ferm) where the production process is carried on, and the *natural environments* (NatEnv) close to the distilleries, and those far from human activities.

We visited a total of 40 locations; 25 Natnv, where we collected insects and plants, and 15 Ferm, where we also collected different objects used in the production process and agave must ferment beside the insects and plants in the factory. These objects include

tools, cooked agave stems, bottom and walls of tanks without active fermentation, waste of fermentation. The NatEnv visited cover three different geographical regions of Mexico (fig. 1A). Oaxaca state is in the south of the Mexican Transvolcanic Belt and has a complex orography, with mountains and valleys formed by the convergence of mountain ranges. Tamaulipas state is on the northeast side of the mountain range *Sierra Madre Oriental*, a natural barrier for many living organisms and where many endemic species can be found. There we collected samples in the *Sierra de San Carlos*, an isolated mountain range with a maximum elevation of 1,786 meters above the level of the sea. The places we visited in Nuevo Leon are right in the middle of the Sierra Madre Oriental. On the northwest is the state of Durango, on the border of *Sierra Madre Occidental*, the other main mountain range in Mexico (fig. 1B). In the NatEnv locations, we defined a 50x10m transect with bark trees as the dominant vegetation. We choose locations where individuals of the tree family Fagaceae were present since this is a known natural reservoir for yeasts, as mentioned in the introduction. Nevertheless, the bark from other plant families in the transect was also collected.

Table 1. Description of the geographic regions where sampling was performed

State	Geographic region	Average Altitude (msnm)	Average Anual Precipitation (mm)	Average Anual Temperature (°C)
Oaxaca	South	1668	1555	22
Durango	Northwest	2038	500	17
Tamaulipas	Northeast	694	780	23.5
Nuevo Leon	Northeast	2512	650	20

Figure 1. A) Map illustrating the location of the 40 sites that were sampled throughout 2021. The color of the points corresponds to Oaxaca (purple), Durango (blue), Tamaulipas (red), and Nuevo Leon (yellow). B) Photos of some of the locations in natural environments from three of the states we went, each photo has the name of the municipality and the state.

Figure 2. Photos of distilleries visited in Durango (A), Oaxaca (C), Nuevo Leon (E), and Tamaulipas (F). B) Sampling of the agave must in cryovials. D) Stem of cooked agave plant as an example of an object collected inside distilleries.

Insects Sampling

Insects were actively trapped with an entomological net previously sterilized in the laboratory with UV radiation for two hours in a laminar flow hood. The net was swiped 1 m above the ground until 10-20 insects were trapped, then put in a conical sterile 50 mL tube with an autoclaved custom-made aspirator. In the field, after each use, nets were sterilized by soaking them in 70% ethanol. A sterile net was used for each different location to avoid horizontal contamination between transects. Although with this method, contamination between tubes with insects of the same transect is possible, not all the tubes of the same location fermented, so the net does not seem to be a vector for horizontal contamination between samples.

Insects were identified at the level of Order in the field and in some cases at the level of Family. Afterward, they were stored in a cooler with ice until further processing in the laboratory.

Plants Sampling

Trees in the transect with a minimum trunk width of 10 cm were sampled. We collected eigth to 20 g of bark in 50 mL sterile conical tubes with a cork borer or knife and a tweezer. The borer, knife, and tweezers were autoclaved before sampling and cleaned with 10% Cl followed by 70% ethanol solutions between each sample to avoid horizontal contamination. The solutions of ethanol and Cl were stored in aliquots of 50 mL in conical tubes where the tool was submerged for 1 min before use, and new aliquots were used in each transect. Samples collected in the conical tubes were stored in a cooler with ice for 3-5 days until further processing in the laboratory. We also took photos of the leaves, bark, flowers, and fruits, if present, for taxonomic identification. For each morphospecies in the transect, a branch with leaves and flowers was collected and was preserved in a botanical press.

Fermentation tank Sampling

To collect samples from fermentation tanks in distilleries, an aliquot was taken with a sterile serological pipette and four ml were stored in a sterile cryovial (Corning Scientific) with glycerol to a final concentration of 25% (fig. 2B). Two other cryovials were also filled but without glycerol. All cryovials were stored in liquid nitrogen until arrival at the laboratory where they were stored at -70°C until further processing.

Sampling of objects in distilleries

To identify potential reservoirs of yeasts associated with agave fermentation and uncover where inoculation of fermentation occurs, samples of different kinds of objects used in agave spirit elaboration were taken in distilleries. Even when distilleries were closed or without production at that moment, producers allowed us to take samples of tools, agave residues, wood, etc. In active distilleries we collected agave plants recently cooked, waste of agave grinding, wood pounders, solid remains of previous fermentations in the tanks, firewood, and other objects (fig. 2D). Depending on the object sampled we used different tools to collect it, yet all tools were previously sterilized with 10% Cl and 70% ethanol.

Data collection of field samples

All metadata and samples were recorded on the KoboToolbox server using the KoboCollect Android application or equivalent in iPhones. In the Kobo server, we created the templates, which were filled for each sample in the field using the mobile apps. Data cleaning was performed in python v3.8.8 and OpenRefine v3.4.1.

Yeast isolation from field samples

To enrich yeasts in samples we implemented the protocol of Liti, Warringer & Blomberg (2017), with modifications. The 50 mL conical tube containing the sample was directly used for enrichment. 20 mL of Saccharomyces Sensu Stricto Enrichment Medium were added and incubated for up to three weeks at 25°C regularly inspecting them for signs of fermentation, like sedimentation, turbidity, and CO_2 production. The actively fermenting cultures were diluted 10^{-3} or 10^{-4} depending on the level of saturation, and plated using glass beads in petri dishes with 20 mL of WL Nutrient Agar (SIGMA and Difco). This medium allows morphology differentiation of yeast colonies due to their pH indicators. Petri dishes were incubated at 25°C for 2-5 days until colonies were well defined, and the agar medium changed from green to blue and then refrigerated at 4°C until strains selection. For strain selection, I selected 12 colonies with *Saccharomyces*-like morphology and 12 with varied morphologies, and placed them in 150 uL of liquid YPD in a 96 well plate. Plates were then incubated at 30°C for two days, 100 uL of 50% glycerol was the added and strains were stored at -70°C.

Taxonomic identification of isolates

Mass spectrometry analysis

For taxonomic identification of isolates, we used matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) that determines the unique proteomic fingerprint of an organism and matches its characteristic mass spectrum pattern with an extensive reference library to determine the organism identity at the species level. We used a MALDI Biotyper® instrument at the laboratory of Manuel Kirchmayer at CIATEJ. The output of the instrument is a score that ranges from 0 to 3 and that reflects whether the tested isolate belongs to a given species in the reference database. Isolates with a score below 1.5 cannot be identified reliably (Bruker). We assessed direct bulk biomass from colonies growth in solid YPD. In some cases, we carried out a protein extraction to improve the quality of the mass spectrum. The MALDI-TOF output also ranks a list of ten probable taxonomic identification identities of each analyte according to their score value. Both parameters, rank identity list and score value, were used to define the most interesting strains for genome sequencing, besides their geography and substrate from which were isolated.

Association analyses

We first group by the location type from where yeast were isolated. Then, we performed Fisher's exact tests to analyze if *S. Cerevisiae* or *S. Paradoxus* were isolated more frequently than expected from any location type (Spurley *et al.*, 2022). Statistical analyses were performed in R v4.2.2.

DNA extraction and genome sequencing

For DNA extraction we implemented a protocol modified from the MasterPure™ Yeast DNA Purification Kit. In brief, we used our own lysis buffer for enzymatic lysis with zymolyase and lyticase, samples were also treated with RNAse A, and precipitated with isopropanol. DNA yield was quantified with a Qubit™ fluorescence spectrophotometer and the 260/280 and 260/230 ratios were determined with a ND-1000 Spectrophotometer (NanoDrop®). For short-read DNA sequencing library preparation and sequencing was performed at the BGI using the DNBSeq platform (China) with 150 bp paired-end reads for each strain.

Bioinformatic analyses

Once the reads were available, we downloaded them to the National Laboratory of Advanced Scientific Visualization supercluster (Lavis) of the National Autonomous University of Mexico (UNAM). For reproducibility, python scripts were used to run the genomic analyses.

Quality control and genome mapping

BGI reported a quality assessment of the raw reads delivered with SOAPnuke. However, adapters and low-quality reads were further removed with fastp v0.20.0 (Chen, 2018) for each of the 114 genomes sequenced. Afterwards, raw reads were mapped to a concatenated genome of the eight species of the *Saccharomyces* genus, plus the reference genomes of *Kluyveromyces marxianus* and of *Pichia kudriavzevii*, with bwa v0.7.4 (Li & Durbin, 2009). Plots with the coverage depth to the concatenated reference genomes were made using samtools v1.9 and R v3.6.1. Then I used a python script to assign the species identity to each sequenced genome according to the proportion of reads mapped to each reference genome. If >90% of the reads mapped to just *S. cerevisiae* or *S. paradoxus* it was considered as either one of these two species. If reads mapped two both species at least in a 10/90% ratio it was considered as a hybrid genome.

Variant Calling and Genotyping

Variant calling of mapped genomes was performed with GATK v4.1.1.0 (DePristo *et al.*, 2011) following its best practice workflow manual for quality control of Single Nucleotide Polymorphisms (SNPs) calling. The total number of SNPs called from the 114 genomic sequences ranges from 28,748 to 447,879. Afterwards, the SNPs of all strains that were identified by genome coverage as either *S. cerevisiae* or hybrids were concatenated in a matrix containing all the SNPs mapped only to the *S. cerevisiae* reference genome.

At this point, genomes from strains from fermentation tanks previously sequenced as part of the Yeast Genomes MX project, as well as genomes from hybrid strains isolated from Brazil (Barbosa *et al.*, 2016) and *S. cerevisiae* isolates from French Guyana (Peter *et al.*, 2018) were also included. At the end, the matrix contained a total of 286 *S. cerevisiae* genomes. A similar SNP matrix was generated with 115 *S. paradoxus* genomes, including 46 from *S. cerevisiae x S. paradoxus* hybrid strains, 24 of which were isolated in this work, two from hybrids from Brazil (Barbosa *et al.*, 2016), and the rest from previous collection efforts in Mexico (54 identified only as *S. paradoxus*), and

three genomes from each one of the recognized subpopulations of the species, SpA, SpB, SpC, SpC*, and SpD (Eberlein *et al.*, 2019; Xia *et al.*, 2017). GATK v4.1.1.0 (DePristo *et al.*, 2011) was also used for genotyping samples, and produce a matrix containing only those sites that were found to be variant in at least one of the samples. This step was done for either the *S. cerevisiae* or *S. paradoxus* genome. Finally, low-quality SNPs were filtered using default GATK best practice recommendations, and 5% of SNPs missing per sample was allowed.

Phylogenetic reconstruction of the sequenced strains

The matrix with the genotyped SNPs was aligned and converted from the default GATK vcf format to phy format with a python script (Ortiz, 2019). Afterwards, the phylogenetic trees for both, *S. cerevisiae* and *S. paradoxus*, were made with RAxML v8.2.12, that only supports GTR substitution model (Stamatakis, 2014). I used the algorithm for maximum likelihood and 100 bootstraps. To visualize the phylogenetic trees, I used Microreact (Argimon *et al.*, 2016).

Results

A large collection of yeast isolates from agave fermentation and their surrounding natural environments throughout different geographical regions

We chose the states of Oaxaca, Durango, and Tamaulipas for strains isolation because they represent three different ecological and geographical regions in Mexico (fig. 1A). In addition, there are differences between these regions in the procedures used to produce agave spirits. For example, in Tamaulipas, the fermentation is performed only with agave juice while in Durango and Oaxaca the whole grounded agave plant, named *bagazo* in Spanish, is fermented. There are substantially differences in the duration of the fermentations as well. In Oaxaca, it ranges from five to 21 days while in Durango it usually only lasts around three days, according to the testimony of producers. Throughout 2021 we made three field trips and collected 861 samples from insects, plants, objects, and fermentation tanks (table 2), we tried to collect as many samples as our work capacity, and resources allowed us from each sample. After implementation of the enrichment protocol, we generated a collection of 4,006 microorganism isolates from 290 different samples and 40 locations throughout the country. 2,903 of the 4006

isolates (73.4%) were identified at the species level by (score >1.49), from which 940 were *S. cerevisiae* (32.4%), and 154 were *S. paradoxus* (5.3%). Other non-*Saccharomyces* species were isolated as well (fig. 3A). Actually, Natural Environments have more diversity considering the Shanon index (2.47) than distilleries (1.85). Considering the Shanon index of different substrates from which isolates were collected, insects (2.18) and plants (2.4) have more diversity than Objects or Tanks (1.76 and 1.3, respectively). Alpha diversity is also greater in NatEnv (39 different species) than in distilleries (12 species). In addition to *Saccharomyces* yeasts, the most commonly isolates species were *Lachancea thermotolerance*, *Pichia manshurica, Pichia kudriavzevii,* and *Kodamaea ohmeri*.

Table 2.

Number of samples collected from which *Saccharomyces* strains were isolated and from which genomes were sequenced.

Sample Type	Samples collected	Samples with ***Saccharomyces*** Isolates	Samples with ***Saccharomyces*** genomes sequenced
Insects	341	32	17
Plants	382	34	31
Objects	98	39	19
Fermentation Tanks	40	34	15
Total	861	139	82

The composition of Saccharomyces species is different between distilleries and natural environments

The fact that most of the isolates were *S. cerevisiae* is not surprising since our enrichment method is specifically designed to isolate *Saccharomyces sensu stricto* species. However, besides this bias, we found that enrichments of samples from distilleries have a different community composition than those coming from natural locations according to the identification of microorganisms by MALDI-ToF (fig. 3A). This difference does not seem to be due to the different substrates that we collected in the distilleries, but rather due to the location type since different substrates from which microorganisms were isolated inside the distilleries have similar community composition (fig. 3B). *L. thermotolerans*, *L. fermentati*, *K. ohmeri,* and *S. paradoxus* are more frequently found in enrichments from natural locations than in distilleries, while *S. cerevisiae* is more frequently isolated from substrates sampled within the distilleries (fig.

3B, and 4B). ANOVA analysis indicates that the location type affects significantly the number of isolates obtained of *Saccharomyces* yeasts ($p<0.01$).

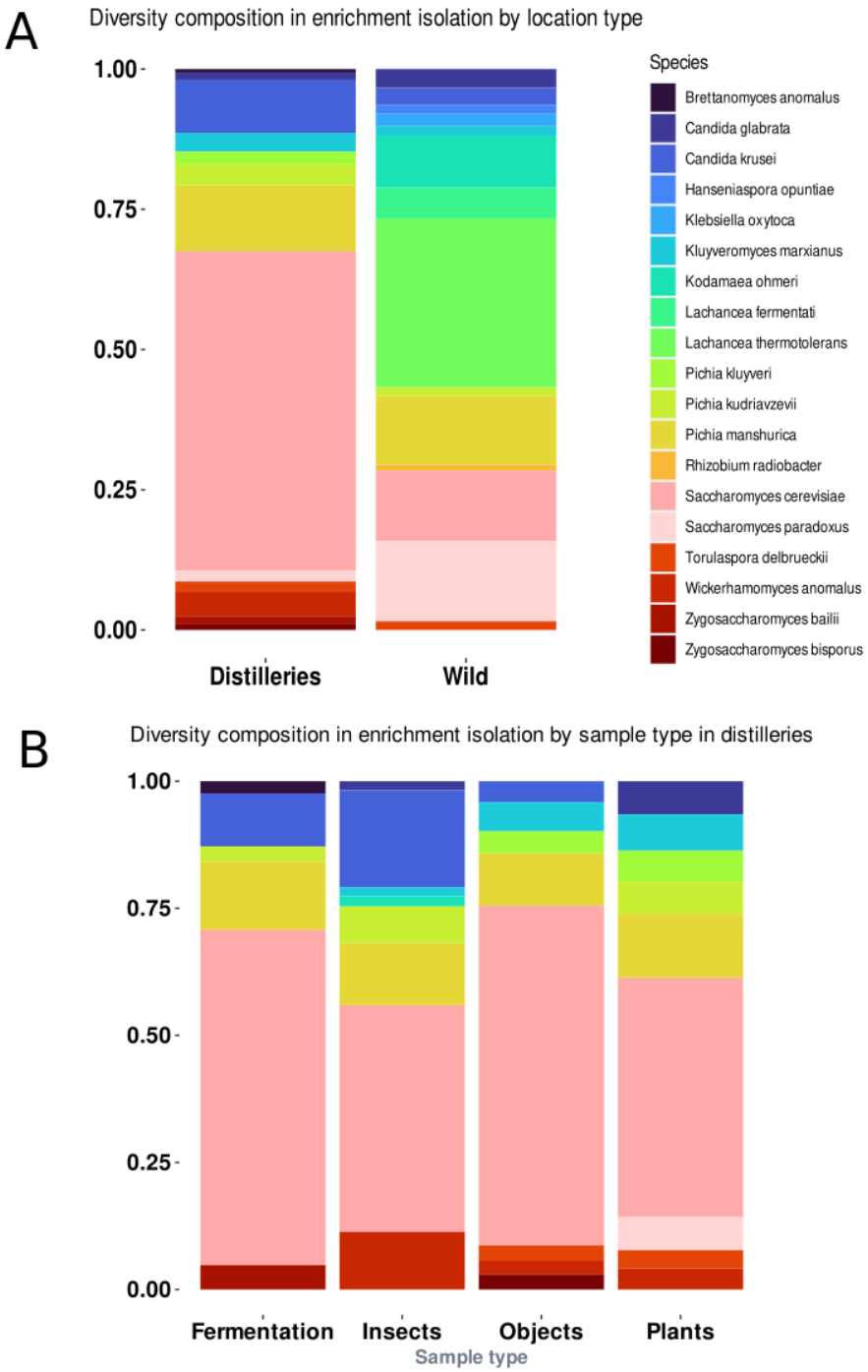

Figure 3. Microorganisms diversity in different sample and location types
A) Diversity of microorganisms in enrichments from different locations. Distilleries involve samples from plants, insects, and objects collected inside the distillery. Wild refers to natural environments not related to human activities. B) Diversity of microorganisms accordI will try to sleep before 3aming to the sample type collected only inside the distilleries. The color code is the same as in A.

S. cerevisiae is the dominant Saccharomyces species in distilleries

We sampled insects and plants, either in distilleries or in natural environments. In distilleries, we also collected objects used in the production process, like the core of plants before and after cooking, residues in the tanks without active fermentation, and tools used to ground and handle the agave stems. We found two members of the *Saccharomyces* genus in our sampling. *S. cerevisiae* was the species most abundant in the enrichments of samples from distilleries, regardless of the sample type (fig. 4A). *S. paradoxus*, which is considered a wild yeast despite being the phylogenetically closest relative to *S. cerevisiae,* was more frequently isolated from locations far from human activities (fig. 4B). Fisher tests indicates there are significant differences ($p < 2.2e-16$) between the frequency of *S. cerevisiae* and *S. paradoxus* isolates from nature and distilleries (fig. 4B), and no significant difference was found between the different objects collected within distilleries and the fermentation tanks ($p=0.18$).

Overall, we found a smaller number of species in enrichments coming from samples collected inside distilleries. For example, species of the genus *Lachancea* were only isolated from natural environments far from distilleries. The substrate from where strains were sampled had little effect, compared to the location type (natural or distilleries), on the species composition of the enrichments, and in the number of isolates of *S. cerevisiae* obtained inside distilleries (fig. 4B). However, the proportion of *S. cerevisiae* isolates from plants and insects collected in the distilleries was less than the proportion of isolates coming from fermentation and objects used in the production process.

S. paradoxus is more frequently isolated from the bark of Fagaceae trees

S. paradoxus showed the opposite pattern than its sister species. Most isolates of *S. paradoxus* come from locations far from the distilleries and human activities (fig. 4B). Plants were the substrate from which most of the *S. paradoxus* isolates were collected. To be more precise, we found it in samples from oak barks (Fagaceae) and cactus fruits (Cactaceae) with more frequency than in any other plant family. However, we only sampled five prickly pears, and on three of them the fruit was collected, and the bark on the other two. The three cactus fruit samples hosted *Saccharomyces* yeasts, while these yeasts were not isolated from other parts of the same plants.

Overall, we collected 173 oak barks and isolated 91 *S. paradoxus* and 71 *S. cerevisiae* strains from 21 of them. It means that the isolation rate of *Saccharomyces* yeasts for this substrate is 12.13% (fig. 4C). Although the proportion of *S. paradoxus* isolates in oak

trees is much higher than in other substrates collected in our survey, the number of isolates that we obtained was not so different from *S. cerevisiae* (91 vs 71), and 66% of the samples from oaks hosted both species (14/21). Both species may exist in natural environments, but they are scarce, and the fact that *S. paradoxus* is isolated more frequently from oaks could be a bias since oaks were the most collected plant in our isolation efforts.

S. cerevisiae has a ubiquitous distribution in natural substrates

In insect and plant samples the overall isolation success rate of *Saccharomyces* yeast was around 12%. There was not a great difference on the isolation rate of *Saccharomyces* yeast from any of the specific host families we collected in our sampling. The family host with the highest isolation rate was Drosophilidae (~ 22%). Interestingly, we found a higher isolation rate of *Saccharomyces* yeasts from pine trees (Pinaceae) than in oaks (Fagaceae). However, the difference in the total number of samples collected from each could be affecting these results (fig. 4C). On the other hand, the isolation rate of *S. cerevisiae* from fermentation tanks was 85%, which is expected to be high since it is considered as the best fermentative microorganism.

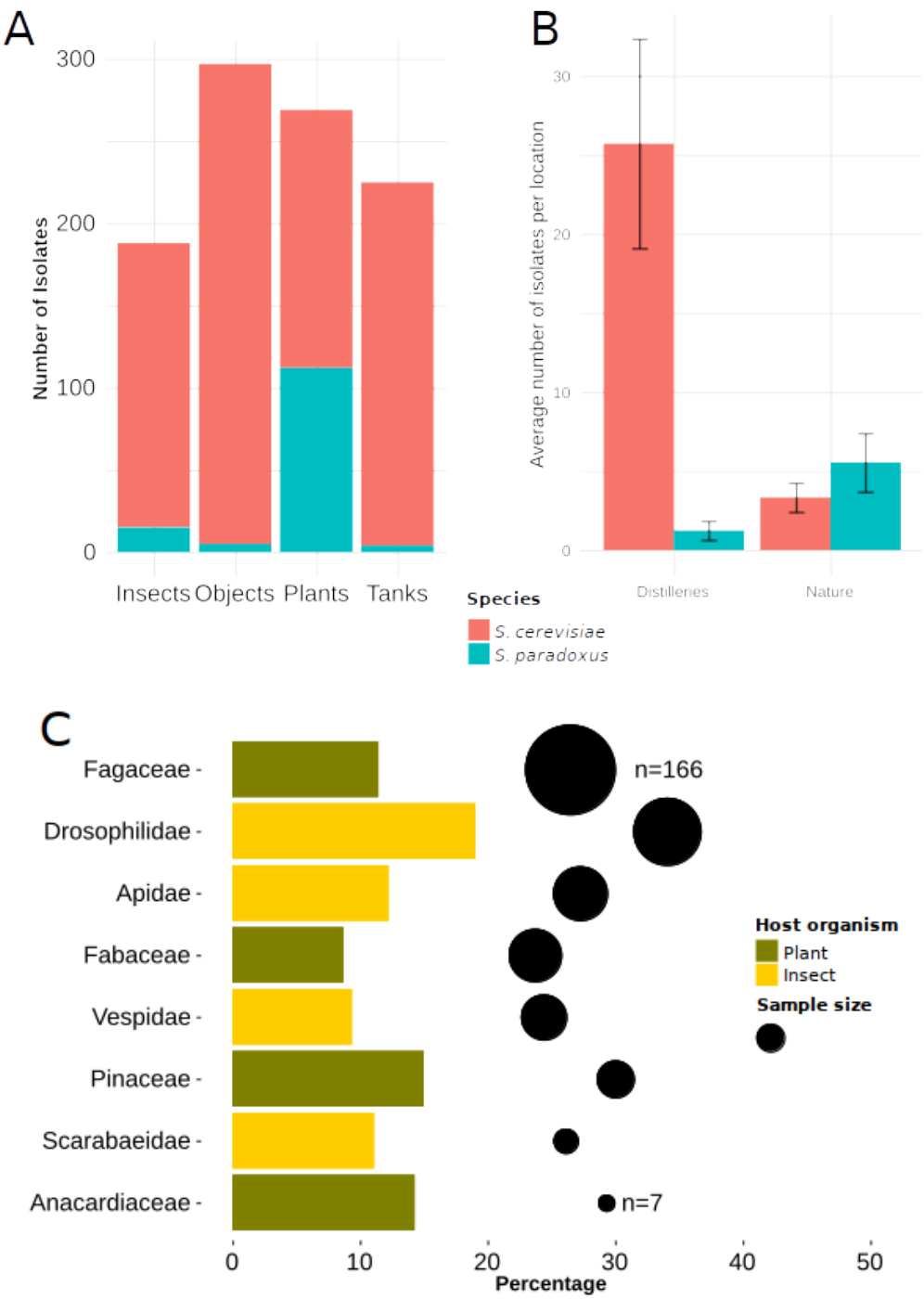

Figure 4. *S. cerevisiae* and *S. paradoxus* have different patterns of distribution in natural and artifical environments

A) proportion of *S. cerevisiae* (pink) and *S paradoxus* (blue) isolates from the different sample types collected. Samples from distilleries and natural sites are grouped. B) Average number of isolates coming from different location types; isolates coming from plants, insects, and objects within distilery are included in the same distillery category, error bars indicates standard error. C) Isolation rate of *Saccharomyces* yeasts from each of the taxonomical families sampled. Black bubbles represent the number of samples collected and bars indicate the percentage of samples from which *Saccharomyces* yeasts were isolated, in green from plants and in yellow from insects.

Genomic and phylogenetic analysis of S. cerevisiae genomes

Overall, we sequenced 106 genomes of strains that were collected in our 2021 sampling, and seven strains collected 27 years ago from the Tequila Herradura distillery in Jalisco, Mexico by Lachance (1995). To ensure wide representation, we chose the isolates for genome sequencing based on their ecological and geographical origin, and the accuracy of MALDI-ToF species identification. At least one genome from each of the 40 sampled locations was sequenced. After species identification by read mapping, we obtained 44 genomes of *S. cerevisiae*, 45 of *S. paradoxus*, and 24 *S. cerevisiae x S. paradoxus* hybrid strains. To create the phylogenetic tree of the *S. cerevisiae* isolates, as mentioned in the methodology, we also used 188 genomes from previously sequenced strains from distilleries throughout the country that were isolated between 2018 and 2021, and 30 genomes from the French Guyana subclade as an outgroup (Peter *et al.*, 2018) (fig. 5A).

S. cerevisiae phylogeny mostly reproduces its geographical origin.

As previously reported by (Urbán-Aragón, 2021), the phylogenetic structure of *S. cerevisiae*, mostly reflect the geographical origin of the strains (fig. 5A,B). Most of the strains group together as one Mexican clade, apart from strains from the outgroup of the French Guyana. The strains coming from the Northeast region of the country are the most similar to the world reference strains from French Guyana. The isolates collected from the Tequila Herradura distillery by Lachance (1995) cluster together with the strains from the same region isolated in 2021.

S. cerevisiae isolates from objects, insects, and plants in distilleries are the same genetic pool as from fermentation tanks

In accordance with the fact that *S. cerevisiae* was rarely isolated from locations far from human activities, we only obtained five sequenced genomes from natural environments (fig. 5A); three from insect samples, one from the bark of a Fabaceae tree, and another one from a Fagaceae. Of these five strains, YMX506C02 and YMX506B02 were collected in the Southern state of Oaxaca from an oak tree and a Diptera insect, respectively. Both samples, cluster together in the phylogeny with the strains isolated from insects and active fermentations inside the Monte Lobo distillery, that is located approximately 10 km apart from the site where the two bark and insect samples were collected.

In the Northwest region of the country, in the state of Durango, we observed a similar scenario with the strain YMX506C03. This strain was collected from an insect (Coleoptera) in the surrounding mountains of the Nombre De Dios municipality, and its closest relative in the phylogeny was isolated 10.3 km from there in a distillery. These observations support the idea that plants and insects outside distilleries and far from fermentation activities host the same population of *S. cerevisiae* that has been associated with the traditional fermentation of *Agave*. However, the low isolation rate of *S. cerevisiae* from natural substrates (insects and bark of trees) suggests these might not be its preferred niche.

The phylogeny of *S. cerevisiae* genomes reveals that the substrate from which yeasts were isolated has no relevance in its phylogenetic structure. Insects, plants, and objects used in the agave fermentation production process collected inside distilleries host the same genetic pool of *S. cerevisiae* isolated from fermentation tanks (fig. 5A). In addition, even when the strains isolated from plants and insects in natural environments are not genetically identical to isolates from the fermentation tanks, in most of the cases they group within the clade of the same geographical region. It is also interesting that we did not find any phylogenetic relation between the strains from the same sample type. The strains from insects cluster with strains from plants, from objects used in the production and from the fermentation itself, as long as they all are from the same geographical region. This suggests that *S. cerevisiae* strains associated to agave fermentation might be adapted to inhabit different habitats inside the distilleries.

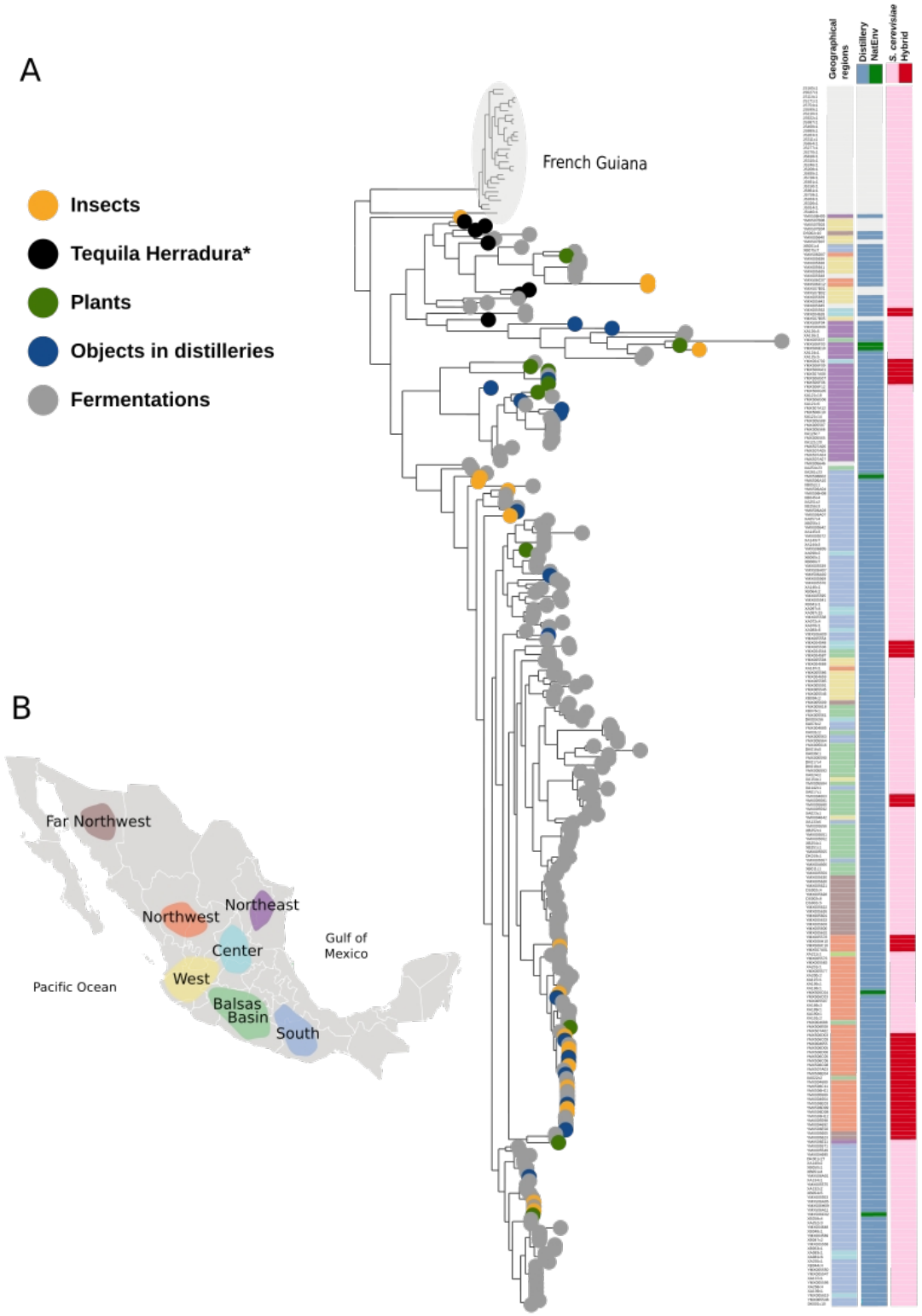

Figure 5. Phylogeny of *S. cerevisiae* A) Maximum likelihood phylogenetic tree of *S. cerevisiae* genomes using 389,761 biallelic SNPs. The color of the nodes represents the sample type from which they were isolated. Clades mainly reproduce their geographical origin, which is represented in B; blue and green bars at the bottom represent the location type from where samples were collected, and red and light red if that genome is from a *S. cerevisiae x S. paradoxus hybrid*. *Strains from Tequila Herradura were collected by Lachance (1995).

Three different sympatric populations of S. paradoxus were isolated from natural substrates in Mexico

We isolated and sequenced the genome of 44 *S. paradoxus* isolates coming from different samples and locations (table 1). The genomes mapped to the concatenated genomic reference grouped with two of the five previously reported subpopulations of *S. paradoxus* in the phylogenetic tree (fig. 6A). 11 of these strains clustered with the *S. paradoxus* subpopulation A (SpA). The rest of the isolates formed two different closely related subgroups, the previously described *S. paradoxus* subpopulation B (SpB) and another clade formed by all the *S. paradoxus* hybrid subgenomes and some 'pure' *S. paradoxus* strains. This last different clade will be called the SpB* lineage onwards (fig. 6B).

Most strains belonging to the SpB* clade were isolated from active agave fermentations and objects involved in the production process, and few from insects and plants inside the distillery. On the other hand, all but seven of the SpB genomes were isolated from plants and insects in natural locations far from distilleries. The seven strains that group with the SpB lineage and come from distilleries were isolated from objects, plants, and fermentation tanks. In the case of the SpA clade, all except one of the strains were isolated from natural environments, far from human activities. These finding show that all *S. paradoxus* subpopulations can be found sympatrically with its sibling species *S. cerevisiae.* However, there are differences in the frequency with which they are isolated from the different environments.

Our results also showed that the different linages of *S. paradoxus* coexist in sympatry in the same geographical regions, and sometimes even in the same substrate. For example, strains YMX506D10 and YMX506D11 were isolated from the same oak in the protected natural area "*Reserva de la Biósfera La Michilía"*, and the first strain grouped with the SpB clade of *S. paradoxus,* while YMX506D11 clustered with the SpA clade. We also found hybrids and "pure" *S. paradoxus* strains cohabiting in the same substrate. The isolate YMX506G01 is a hybrid between *S. cerevisiae* and *S. paradoxus*, while YMX506G01 is a *S. paradoxus* strain that belongs to the SpA clade, and both strains were isolated from the same location and substrate, a prickly pear (*Opuntia* spp.). There is a similar situation for the strains YMX506F09 and YMX506F11, which are a hybrids and an *S. paradoxus* from the SpB clade, respectively, both isolated from another prickly pear too.

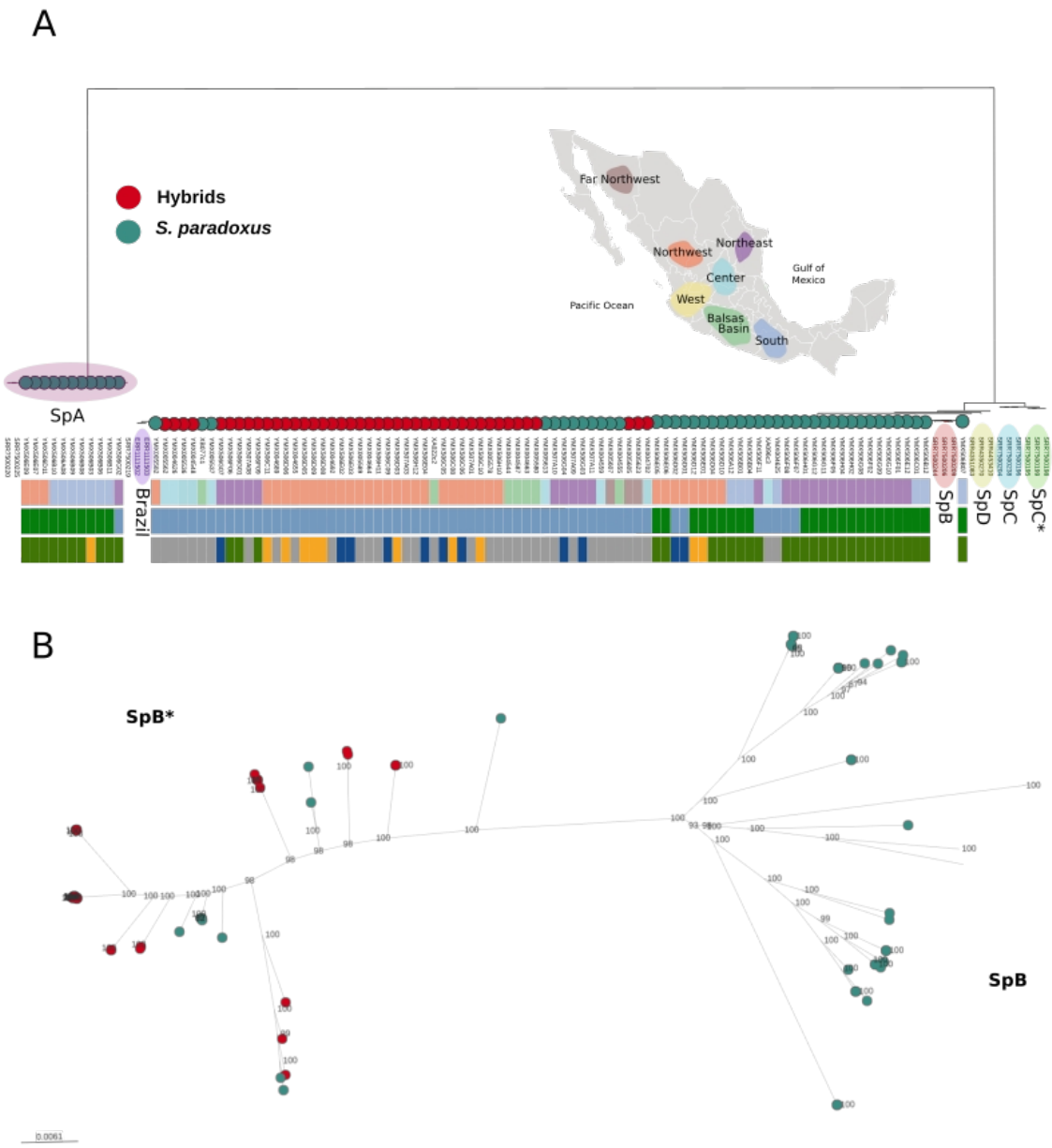

Figure 6. Phylogeny of *S. paradoxus*

A) Maximum likelihood phylogenetic tree of *S. paradoxus* genomes using 270,097 biallelic SNPs. The color of the nodes represents if the strain is *S. paradoxus* or a hybrid strain. The first row corresponds to the geographical origin of each strain, and has the same color code as in the map. The second and third row illustrates the location type, and the the sample type, respectively, using the same color code as in 5A. B) Same phylogenetic tree as A, unrooted and only with the SpB and SpB* clades.

A novel S. paradoxus population associated with inter-species hybridization with S. cerevisiae

As previously mentioned, the genomes sequenced were mapped to a concatenated reference genome composed of the eight species of the *Saccharomyces* genus to uncover hybrid strains. We found 24 hybrids coming from the states of Durango and Nuevo Leon, while no hybrids were isolated from Tamaulipas or Oaxaca. The hybrid strains had different proportions of reads mapped to the parental genomes. Interestingly, hybrids having less than ~50% of the *S. cerevisiae* parental subgenome were not found. Besides, in samplings carried out in 2019, *S. cerevisiae x S. paradoxus* hybrids strains were also found. The phylogenetic analyses suggest they share the same ancestry as the hybrid strains we isolated in 2022. There are many *S. cerevisiae* strains with a range between 0 and 5% of reads mapped to *S. paradoxus*, while the maximum percentage of *S. cerevisiae* reads introgressed into *S. paradoxus* isolates is aproximately 1.1%. Based on their genomic content there were three different groups of hybrid strains isolated. The first one with approximately 50 % (n=5) of each parental genome, the second group with a range of 63 to 69 % (n=15) of introgressions, and the third group with 77-79 % (n=4) of reads mapped to *S. paradoxus* (fig. 7A). *S. cerevisiae* strains isolated in this survey also have introgressed genes from *S. paradoxus*, but these introgressed *S. cerevisiae* strains were isolated from different geographical locations (fig. 7C). Analyzing the number of *S. paradoxus* introgressed genes in the *S. cerevisiae* we found that in Durango there is a bimodal distribution, strains with around 150 introgressed genes and a second group without introgressions. In Oaxaca all the strains have a range of between 125 and 175 genes from *S. paradoxus*, and in Tamaulipas there is a wide range of the number of introgressed genes in the different strains, from zero to more than 350 genes.

Interestingly, all the hybrids identified were only isolated from distilleries. However, the number of hybrid genomes was constant between samples from insects, objects used during production, and fermentation tanks; hybrids were found with the same frequency in these three substrates within distilleries (fig. 7B). Although we also found hybrids in plant samples from distilleries, they were less frequent than in the other sample types and no hybrids were isolated from oak trees.

Surprisingly, all the hybrid subgenomes cluster together in the here reported SpB* clade (fig 6B). This observation suggests that there is a previously not reported *S. paradoxus* lineage that is associated with hybridization of this species with *S. cerevisiae*. In all the regions from where hybrids were isolated, ‘pure’ SpB* strains were also found, and in

regions in which no hybrids were isolated, the SpB* ‘pure’ linage was also not found. For example, in the state of Oaxaca, where two subpopulations of *S. paradoxus*, SpA and SpB, were isolated, no hybrids strains were found. We also did not find *S. paradoxus* in distilleries from this state. These observation suggest an association between the newly described SpB* lineage and the presence of hybrids. Furthermore, since SpB* strains were found only inside distilleries, it is possible that this lineage could be restricted to the agave fermentation environment.

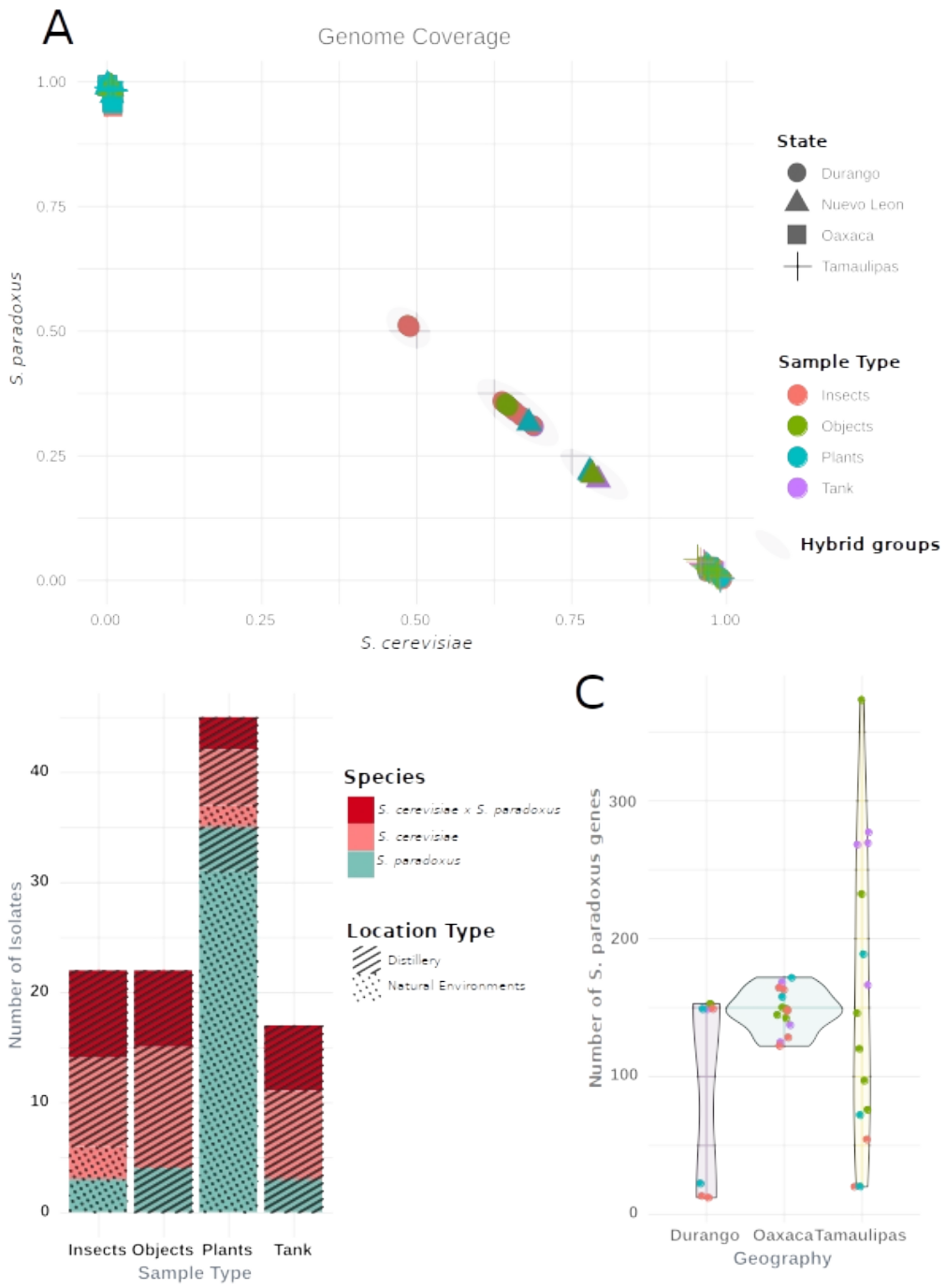

Figure 7. Hybrids were found only in distilleries and have different genomic proportion of parentals
A) There were three main groups of hybrid strains: 1) with near 50/50 parental genome proportion, 2) with 70/30, and 3) with 80/20 proportion of reads that mapped to *S. cerevisiae* and *S paradoxus*, respectively. There were *S. cerevisiae* strains with 1% to 5% of *S. paradoxus* introgressions. The shape represents its geographical origin and the color, the sample type. B) Number of strains of each species in the different substrates collected. Hybrids (purple) were isolated from all the sample types but only from distilleries (striped pattern). *S. cerevisiae* (green) was isolated mostly from distilleries, while *S. paradoxus* (yellow) was isolated more frequently from natural environments (doted pattern). C) Number of genes introgressed in S. cerevisiae strains from *S. paradoxus* by geographical region. The color code of points is the same as in A.

Discussion

S. cerevisiae is scarce but ubiquitous in the wild and insects may work as vectors for its dispersion

The overall isolation success rate of *Saccharomyces* yeasts from natural substrates (plants and insects) in our study was around 12%, which is close to the previously reported rate by Sniegowski *et al.* (2002), who isolated *Saccharomyces* yeasts from the 14% of oak samples they collected. Nevertheless, our results contrast with those of Barbosa *et al.* (2016) since they reported an overall isolation rate of ~22%, and even higher than 70% for samples coming from the bark of Fagaceae trees. This last report and other observations of *S. cerevisiae* isolated from Fagaceae trees led to the idea of the "oak niche" for *Saccharomyces* yeasts (Naumov *et al.* 1998; Sniegowski *et al.* 2002; Sampaio & Goncalves 2008; Wang *et al.* 2012; Hyma & Fay 2013). However, we collected 166 samples of Fagaceae trees, and the isolation rate was close to the rate for any other sample type. It has been also suggested that the enrichment protocols and sample bias overestimated the abundance of *Saccharomyces* yeasts in natural environments, like the bark of oak trees (Godard & Greig, 2015). In fact, a survey on *S. paradoxus,* the 'wild' sibling species of *S. cerevisiae,* has demonstrated that it is far from being a dominant species in the microbial community of the oak bark, where it has been estimated that there are approximately two *S. paradoxus* cells by square centimeter of bark (Kowallik, Miller & Greig, 2015).

We also found that *S. cerevisiae* was isolated less frequently from natural environments, where there are less sugar rich ecological substrates, than from distilleries. This is congruent with the nomadic yeast hypothesis which argues that the low isolation rate of *S. cerevisiae* on natural substrates is probably because it is a generalistic species, able of inhabiting several niche but none of them especially well (Godard & Greig, 2015). It is also possible that yeasts persist in a dormant state in these environments, such as spores. Cells in this state do not propagate and therefore low cell numbers are maintained in natural populations of yeasts. Spores also enable yeast cells to travel towards new niches by insect dispersion, which is known as the dispersal-encounter hypothesis (Madden *et al.*, 2018). In this way, *S. cerevisiae* strains associated with agave fermentation could be adapted to the itinerant production of agave spirits, since most of the producers do not use an inoculum and yeast survive in plants, insects, and objects inside distillery until they are dispersed to new fermentation batches by insects.

We also showed that *S. cerevisiae* and *S. paradoxus* are present in insects in the distilleries where agave spirits are produced and from locations far from human activities. Besides, in accordance with our phylogenetic analysis, *S. cerevisiae* strains from insects collected inside the distileries group in the same clades than the strains collected from fermentation tanks in the same geographical region. Previous reports have found that *S. cerevisiae* spores are unable to disperse by air (Mortimer & Polsinelli, 1999), hence they need a vector to reach new carbon sources. There are several reports of the isolation of *S. cerevisiae* from Coleoptera, Orthoptera, Diptera, Megaloptera, and Hymenoptera insects (Meriggi *et al.*, 2020). We isolated *S. cerevisiae* strains from all the previously mentioned taxa, except Orthoptera and Megaloptera. Actually, the taxonomical family from which we obtained more *Saccharomyces* strains was Drosophilidae (fig. 4C) with an isolations success rate of 20%. *Drosophila melanogaster* particularly, prefers to feed from substrates with yeasts because of the volatiles they produce during fermentation (Becher *et. al*, 2016). However, *D. melanogaster* seems to be a general vector for yeast, without specificity for *S. cerevisiae* or any other particular species (Quan & Eisen, 2018). Hence, *Saccharomyces* strains living in the natural environments might be transported from one substrate to another with the entire microbial community from where they were picked up. Nevertheless, transportation by fruit flies has some challenges since not all yeast species are able to survive digestion. It is known that *S. cerevisiae* vegetative cells are not able to survive the digestive tract of *D. melanogaster* but spores do. In addition, the outbreeding rate increases when *Saccharomyces* yeasts passage through the digestive tract of fruit flies due to degradation of the ascospore that encloses the spores (Reuter, Bell & Greig, 2007).

Lachance (1995) collected *S. cerevisiae* strains from fruit flies in a tequila distillery in Mexico and suggested that these insects may serve as vectors for yeast dispersal. He also reported that inoculation occurs in the early stages of the agave fermentation process, shortly after agave is cooked. We isolated yeast strains from the cooked agave steams, supporting the notion that inoculation occurs at early stages of agave spirits production. Some of the *Saccharomyces* strains were also isolated from the *bagazo*, which is a residue of agave fermentation, it consist basically on the fiber of grounded stems of *Agave* plants, producers use to piled up at the edge of the distillery, and is usually visited by flies and other insects. Hence, *bagazo* could work as a yeast reservoir for the inoculation of the next fermentations by insects in the distillery.

Domesticated and "feral" Saccharomyces strains are involved in agave fermentations in Mexico

Phylogenetic analyses of *S. cerevisiae* population support the idea that geography is an important factor determining its genetic structure (Da-Yong *et al.*, 2021). On the other side, Duan *et al.* (2018) have also argued that ecology is a determinant factor in the baker's yeast population structure, and have proposed three distinct major clades in the whole species. Wild strains isolated from natural substrates, mainly from oak trees and usually far from human activities, strains from Liquid State Fermentations (LSF), including from Mexican agave fermentations, and Solid-State Fermentations (SSF). In addition, previous analyses have found that *S. cerevisiae* genomes isolated from Mexican agave fermentation strains cluster in a different phylogenetic group from the rest of the world (Peter *et al.,* 2018). The closest relatives to the Mexican linage are strains coming from French Guyana isolated from human-related environments (Peter *et al.*, 2018). We found that regardless of the ecological substrate from which isolates were collected, strains clustered within the general Mexican Agave clade. Furthermore, most of the Mexican isolates from plants and insects collected inside distilleries were grouped together with the strains isolated from fermentation tanks in the same geographical region (fig. 5A).

We sequenced the genome of 18 strains identified as *S. cerevisiae* by MALDI-TOF and collected from natural environments, far from human activities. However, only five of these strains were actually *S. cerevisiae* when their genome was sequenced and analyzed. The rest of them were *S. paradoxus*, which has been considered as the phylogenetically closest wild relative to *S. cerevisiae* (Boynton & Greig, 2014). Therefore it was not surprising to find it more frequently in natural environments than *S. cerevisiae* that is thought to be a mostly domesticated species. Nevertheless, lineages of *S. cerevisiae* have been previously isolated from natural environments and the phylogenetic evidence suggests a wild origin of the species (Peter *et al.,* 2018).

In a population genomic study of Mediterranean yeasts, Viel *et al.* (2017) conclude that geographical distance from the vineyard is one of the key factors governing the differentiation of yeast populations. We found that three of the five strains isolated from natural environments around 10 km away from any distillery clustered together in the phylogeny with the ones collected inside distilleries from the same geographical region. Whether these isolates are feral (domesticated yeasts that returned to the wild) or wild strains is not clear and it also depends on the definition of a wild population.

Wild populations of *S. cerevisiae* are considered genetically different from domesticated ones, however it is still in debate if the genetic diversity of *S. cerevisiae* and its population structure is driven by genetic drift or natural selection (Bai *et al.*, 2022). However, some genetic and phenotypic differences between wild and domesticated strains of *S. cerevisiae* are correlated with its respective phylogenetic clade (Peter *et al.*, 2018). Han *et al.* (2021) reported that most of the wild strains lack the homing endonuclease VDE, but this enzyme is commonly present in the domesticated lineages. They also argued that the heterozygosity level of wild *S. cerevisiae* strains is significantly lower than that of the domesticated isolates. Differences in the sporulation rate and spore viability between wild and domesticated strains have been also reported, being wild isolates more efficient than domesticated ones in both parameters (Duan *et al.,* 2018). These characteristics might have an impact on the life strategy of the microorganism and suggest that wild and domesticated populations have evolved by adaptation to their environment (Bai *et al.*, 2022). To consider a strain as wild, it should be grouped within the wild clade of *S. cerevisiae* in the species phylogeny and should have the typical genetic and phenotypical hallmarks of wild isolates. An isolate from a natural environment, far from human activities but grouped within the domesticated clade could be rather considered as a feral yeast, as the three strains that we isolated from Durango and Oaxaca in this survey. These three strains clustered within the phylogenetic clade of domesticated strains but were isolated from a natural environment. However, further characterization is necessary to corroborate that the rest of the genetic and phenotypical domesticated traits are present in these isolates.

It is possible that strains from agave fermentations are a kind of “feral” yeasts. Most of these strains can survive in nature far from the fermentation environment, so they can be isolated from the bark of trees or insects in the forest but with less frequency since the natural environments are not sugar-rich and cells could be in latency. According with our isolation results of strains identified by mass spectrometry, the proportion of isolates from natural environments identified as *S. cerevisiae* or *S. paradoxus* by this technique was relatively the same for each species (fig. 4B). Since *S. paradoxus* is recognized as a wild species, the similar isolation rate suggests that *S. cerevisiae* isolates from natural environments have the characteristics of wild lineages that are adapted to inhabit these habitats. However, whole genomic sequencing revealed that only 28% of the isolates from natural environments were actually *S. cerevisiae*. Because *S. cerevisiae* (the domesticated species) is not as abundant as *S. paradoxus* (the wild species), agave strains might not in fact be adapted to live in natural environments as well as *S. paradoxus.* This suggests that they are probably domesticated strains.

There are two strains that were isolated from natural environments in Tamaulipas, not related with human activities, that clustered in the longest branch of the *S. cerevisiae* phylogenetic tree. This branch corresponds to an isolate from Morelos collected more than 20 years ago and 800 km away from Tamaulipas. Due to the length of the branch and the other strains that belong to this clade, this phylogenetic group may represent a different group than the Mexican agave clade. To confirm this hypothesis more analyses have to be performed.

Lachance (1995) collected yeasts from several substrates in a tequila distillery in Jalisco, Mexico 27 years ago. In this study *S. cerevisiae* was isolated from *Drosophila* species, cooked agave stems, agave molasses, and from all fermentation stages. We sequenced seven genomes of these *S. cerevisiae* strains covering all the substrates sampled by Lachance (1995) and found that they group in the clade with the rest of the strains previously isolated from Jalisco (Fig. 5A). This clade also contains few strains from Oaxaca, Tamaulipas, and Sinaloa, but is mostly represented by isolates associated with tequila fermentation, and collected from Jalisco. On the other hand, the strains from traditional agave fermentation seem to inhabit different habitats besides fermentation tanks, like the bark of plants, and insects, although they are not abundant in these substrates. However, this might change as natural environments are better sampled, since we sampled just 25 natural locations throughout the country, which is a first approach to describe the ecology and diversity of wild *Saccharomyces* yeasts in Mexico.

The phylogenetic similarity of strains isolated from non-fermentative substrates with those from fermentation tanks suggest that they could be feral strains. Whether they were domesticated from an ancestral wild population that inhabit America before humans discover *Agave* fermentation or were introduced to the new continent by humans remains unclear.

The agave fermentation environment could promote the formation of hybrids

Interspecific hybridization between eukaryotic organisms usually produces infertile progeny and many species have prezygotic barriers to avoid mating with a different species. The first barrier that yeasts must overcome for hybridization is temporal and spatial coexistence. For species to hybridize, they have to inhabit the same place at the same time (Steensels, Gallone & Verstrepen, 2022). *S. cerevisiae* and *S. paradoxus* are sympatric species in oak barks in some regions of the world (Naumov, Naumova & Sniegowski, 1998; Sniegowski, Dombrowski & Fingerman, 2002; Barbosa *et al.*, 2016).

As previously discussed, in the regions that we sampled we found that *S. paradoxus* and *S. cerevisiae* have different preferences for environmental locations. *S. cerevisiae* was isolated more frequently from distilleries than natural environments, while *S. paradoxus* was more recurrent in the bark of trees in natural environments than in distilleries. However, regardless of the frequency of isolation, both species were found in both location types. Hence these two species can coexist at some time in their life history in the environments that we sampled.

Despite the coexistence of both species in natural environments and even when we extensively sampled natural locations, all 24 hybrid isolates that we recovered were all isolated from fermentations, objects, plants, and insects in the distilleries. In these environments *S. cerevisiae* is considerably more abundant while *S. paradoxus* is only occasionally isolated. Overall, there are six phylogenetic clades with *S. cerevisiae* hybrid subgenomes, which might represent different hybridization events with *S. paradoxus*. This suggests that hybridization may be occurring frequently, but only in distilleries.

In yeasts there are also post-mating barriers that hybrid strains have to overcome to reproduce sexually. These are mainly due to the inability of hybrids to complete meiosis due to the lack of sequence similarity between homologous chromosomes. Therefore, LOH events or genomic rearrangements are needed to recover fertility (D'angiolo *et al.*, 2020). It is likely that the introgressed *S. cerevisiae* strains (fig. 7C) that we found associated to the agave fermentations are the result of an ancestral hybridization event with posterior backcrossing with one parental genome species (Martin & Jiggings, 2017). It is interesting that the proportion of *the S. paradoxus* parental genome in the hybrid strains is always greater than 50%, while the proportion of the *S. cerevisiae* subgenome ranges between 50 and 20%. In addition, no *S. paradoxus* strains with introgressions from *S. cerevisiae* were isolated. These observations suggest that backcrossing with a *S. cerevisiae* parent is preferred instead of a *S. paradoxus* one. Overall, our findings also suggest that there is recurrent hybridization in *Saccharomyces* yeast associated with agave fermentation.

Hybrid strains were only isolated from distilleries. The main difference between distilleries and natural locations is the proximity to a sugar-rich agave fermentation environment. In the industrial beer lineage, interspecific hybridization in the *Saccharomyces* genus facilitates niche adaptation (Gallone *et al.*, 2019). *S. cerevisiae x S. paradoxus* hybrids have been found in human-associated environments, like in olive oil production (D'angiolo *et al.,* 2020), but no hybrid adaptive traits to this environment have been identified. Another good example concerns the bioethanol industry, in which

an *S. cerevisiae x S. bayanus* hybrid was used to increase ethanol production and flocculation (Choi *et al.*, 2010). Agave fermentation hybrids may have phenotypical traits that favor their presence in distilleries rather than in natural environments. Several Maillard compounds are produced during agave cooking like HMF (Mancilla-Margalli & Lopez, 2002), to which *S. cerevisiae* strains from agave fermentations are more resistant than strains from the rest of the world (Gallegos-Casillas, 2020). In addition, the length of the fermentation could be associated with the hybrid vigor of agave strains. In Durango state, where most of the hybrid strains have been isolated, mezcal fermentation lasts for only three days in most of the distilleries. All these observations, suggest that distilleries may promote the presence of hybrids, however, further research is necessary to reveal the causal relation of this hypothesis.

Conclusions

Phylogenetic analysis reveals that *S. cerevisiae* isolates from objects, insects, and plants inside the distilleries are the same genetic pool as the isolates from agave fermentation tanks. This observation supports the idea that yeast associated with agave fermentations can live in the vicinities of the distillery when there aren't active fermentations. In agreement to what has been reported in the literature, objects used during production and the bark of trees could be reservoirs of *S. cerevisiae* and insects may work as a vector in yeast dispersal from these reservoirs to fermentation vats.

S. cerevisiae was more frequently isolated from distilleries than from natural environments away from the distilleries, and *S. paradoxus* is less frequent in artificial environments than in natural environments. These results suggest that there are differences between the ecological preferences of both *Saccharomyces* species that lead them to inhabit different substrates. One of the reasons could be the domestication process of *S. cerevisiae*, through which it has been artificially selected by humans, even if involuntarily, to better perform in artificial environments.

We found evidence of three different sympatric populations of *S. paradoxus* in Mexico, and one of these populations has not been reported before. In addition, this newly described subpopulation may be driving recurrent inter-species hybridizations with *S. cerevisiae* since all the *S. paradoxus* subgenomes of hybrid strains clustered in this new clade. The novel SpB* lineage was isolated mainly from distilleries, while other lineages were absent in this artificial environment. There could be genetic factors in this population that enhance the viability of interspecific hybrids with its sibling species *S.*

cerevisiae. In agreement with this hypothesis, although we performed extensive sampling in natural environments, hybrid strains were only isolated from distilleries and in the geographical regions where the SpB* lineage was isolated. Finally, the fact that hybrid strains were isolated only from distilleries even when *S. paradoxus* prefers to inhabit natural environments, far from human activities, suggests that hybrids could have adaptive traits for the agave fermentation environment.

Traditional producer practices have shaped the ecological niches of yeasts for centuries. It is important to do research on such ecological interactions since good practices can enhance yield production and microorganism diversity conservation. Besides, having a wide collection of fermentative yeasts makes it easier to identify strains with traits relevant to *Agave* fermentation, which would help to improve and develop biotechnological solutions for *Agave* spirits production, in general.

References

Alcázar, M., Kind, T., Gschaedler, A., Silveria, M., Arrizon, J., Fiehn, O., Vallejo, A., Higuera, I., Lugo, E., 2017. Effect of steroidal saponins from Agave on the polysaccharide cell wall composition of Saccharomyces cerevisiae and Kluyveromyces marxianus. LWT 77, 430–439. https://doi.org/10.1016/j.lwt.2016.11.048

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, 5412–5427. https://doi.org/10.1111/mec.13341

Alsammar, H., Delneri, D., 2020. An update on the diversity, ecology and biogeography of the *Saccharomyces* genus. FEMS Yeast Research 20. https://doi.org/10.1093/femsyr/foaa013

Arellano-Plaza, M., Paez-Lerma, J.B., Soto-Cruz, N.O., Kirchmayr, M.R., Gschaedler Mathis, A., 2022a. Mezcal Production in Mexico: Between Tradition and Commercial Exploitation. Frontiers in Sustainable Food Systems 6. https://doi.org/10.3389/fsufs.2022.832532

Arellano-Plaza, M., Paez-Lerma, J.B., Soto-Cruz, N.O., Kirchmayr, M.R., Gschaedler Mathis, A., 2022b. Mezcal Production in Mexico: Between Tradition and Commercial Exploitation. Frontiers in Sustainable Food Systems 6. https://doi.org/10.3389/fsufs.2022.832532

Argimón, S., Abudahab, K., Goater, R.J.E., Fedosejev, A., Bhai, J., Glasner, C., Feil, E.J., Holden, M.T.G., Yeats, C.A., Grundmann, H., Spratt, B.G., Aanensen, D.M., 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microbial Genomics 2. https://doi.org/10.1099/mgen.0.000093

Bai, F.-Y., Han, D.-Y., Duan, S.-F., Wang, Q.-M., 2022. The Ecology and Evolution of the Baker's Yeast Saccharomyces cerevisiae. Genes 13, 230. https://doi.org/10.3390/genes13020230

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, 317–329. https://doi.org/10.1093/gbe/evv263

Boynton, P.J., Greig, D., 2014. The ecology and evolution of non-domesticated *Saccharomyces* species. Yeast n/a-n/a. https://doi.org/10.1002/yea.3040

Cavalieri, D., McGovern, P.E., Hartl, D.L., Mortimer, R., Polsinelli, M., 2003. Evidence for S. cerevisiae Fermentation in Ancient Wine. Journal of Molecular Evolution 57, S226–S232. https://doi.org/10.1007/s00239-003-0031-2

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Choi, G.-W., Um, H.-J., Kang, H.-W., Kim, Y., Kim, M., Kim, Y.-H., 2010. Bioethanol production by a flocculent hybrid, CHFY0321 obtained by protoplast fusion between Saccharomyces cerevisiae and Saccharomyces bayanus. Biomass and Bioenergy 34, 1232–1242. https://doi.org/10.1016/j.biombioe.2010.03.018

Correa-Ascencio, M., Robertson, I.G., Cabrera-Cortés, O., Cabrera-Castro, R., Evershed, R.P., 2014. Pulque production from fermented agave sap as a dietary supplement in Prehispanic Mesoamerica. Proceedings of the National Academy of Sciences 111, 14223–14228. https://doi.org/10.1073/pnas.1408339111

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, 420–425. https://doi.org/10.1038/s41586-020-2889-1

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43, 491–498. https://doi.org/10.1038/ng.806

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, 2690. https://doi.org/10.1038/s41467-018-05106-7

Eberlein, C., Hénault, M., Fijarczyk, A., Charron, G., Bouvier, M., Kohn, L.M., Anderson, J.B., Landry, C.R., 2019. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. Nature Communications 10, 923. https://doi.org/10.1038/s41467-019-08809-7

Eberlein, C., Leducq, J.-B., Landry, C.R., 2015. The genomics of wild yeast populations sheds light on the domestication of man's best (micro) friend. Molecular Ecology 24, 5309–5311. https://doi.org/10.1111/mec.13380

Fay, J.C., Benavides, J.A., 2005. Evidence for Domesticated and Wild Populations of Saccharomyces cerevisiae. PLoS Genetics 1, e5. https://doi.org/10.1371/journal.pgen.0010005

Fischer, G., Liti, G., Llorente, B., 2021. The budding yeast life cycle: More complex than anticipated? Yeast 38, 5–11. https://doi.org/10.1002/yea.3533

Gallegos-Casillas, P., 2020. Caracterización fenotípica de cepas de Saccharomyces cerevisiae asociadas a la fermentación de agave. Centro de Investigación y Estudios Avanzados del IPN, Irapuato, Guanajuato.

Gallone, B., Steensels, J., Mertens, S., Dzialo, M.C., Gordon, J.L., Wauters, R., Theßeling, F.A., Bellinazzo, F., Saels, V., Herrera-Malaver, B., Prahl, T., White, C., Hutzler, M., Meußdoerffer, F., Malcorps, P., Souffriau, B., Daenen, L., Baele, G., Maere, S., Verstrepen, K.J., 2019. Interspecific hybridization facilitates niche adaptation in beer yeast. Nature Ecology & Evolution 3, 1562–1575. https://doi.org/10.1038/s41559-019-0997-9

Gilbert, D.G., 1980. Dispersal of Yeasts and Bacteria by Drosophila in a Temperate Forest. Oecologia 46, 135–137.

Goddard, M.R., Greig, D., 2015. Saccharomyces cerevisiae: a nomadic yeast with no niche? FEMS Yeast Research 15. https://doi.org/10.1093/femsyr/fov009

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, 546–567. https://doi.org/10.1126/science.274.5287.546

Han, D.-Y., Han, P.-J., Rumbold, K., Koricha, A.D., Duan, S.-F., Song, L., Shi, J.-Y., Li, K., Wang, Q.-M., Bai, F.-Y., 2021. Adaptive Gene Content and Allele Distribution Variations in the Wild and Domesticated Populations of Saccharomyces cerevisiae. Frontiers in Microbiology 12. https://doi.org/10.3389/fmicb.2021.631250

Henry J. Bruman, 2000. Alcohol in Ancient Mexico. University of Utah Press.

Knight, S.J., Goddard, M.R., 2016. Sporulation in soil as an overwinter survival strategy in *Saccharomyces cerevisiae*. FEMS Yeast Research 16, fov102. https://doi.org/10.1093/femsyr/fov102

Lachance, M.-A., 1995. Yeast communities in a natural tequila fermentation. Antonie van Leeuwenhoek 68, 151–160. https://doi.org/10.1007/BF00873100

Langdon, Q.K., Peris, D., Baker, E.P., Opulente, D.A., Nguyen, H.-V., Bond, U., Gonçalves, P., Sampaio, J.P., Libkind, D., Hittinger, C.T., 2019. Fermentation innovation through complex hybridization of wild and domesticated yeasts. Nature Ecology & Evolution 3, 1576–1586. https://doi.org/10.1038/s41559-019-0998-8

Leducq, J.-B., Nielly-Thibault, L., Charron, G., Eberlein, C., Verta, J.-P., Samani, P., Sylvester, K., Hittinger, C.T., Bell, G., Landry, C.R., 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. Nature Microbiology 1, 15003. https://doi.org/10.1038/nmicrobiol.2015.3

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

Liti, G., 2015. The fascinating and secret wild life of the budding yeast S. cerevisiae. eLife 4. https://doi.org/10.7554/eLife.05835

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, M.A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., Louis, E.J., 2009. Population genomics of domestic and wild yeasts. Nature 458, 337–341. https://doi.org/10.1038/nature07743

Ludlow, C.L., Cromie, G.A., Garmendia-Torres, C., Sirr, A., Hays, M., Field, C., Jeffery, E.W., Fay, J.C., Dudley, A.M., 2016. Independent Origins of Yeast Associated with Coffee and Cacao Fermentation. Current Biology 26, 965–971. https://doi.org/10.1016/j.cub.2016.02.012

Mancilla-Margalli, N.A., López, M.G., 2002. Generation of Maillard Compounds from Inulin during the Thermal Processing of *Agave tequilana* Weber Var. azul. Journal of Agricultural and Food Chemistry 50, 806–812. https://doi.org/10.1021/jf0110295

Martin, S.H., Jiggins, C.D., 2017. Interpreting the genomic landscape of introgression. Current Opinion in Genetics & Development 47, 69–74. https://doi.org/10.1016/j.gde.2017.08.007

Meriggi, N., Di Paola, M., Cavalieri, D., Stefanini, I., 2020. Saccharomyces cerevisiae – Insects Association: Impacts, Biogeography, and Extent. Frontiers in Microbiology 11. https://doi.org/10.3389/fmicb.2020.01629

Monerawela, C., Bond, U., 2017. Brewing up a storm: The genomes of lager yeasts and how they evolved. Biotechnology Advances 35, 512–519. https://doi.org/10.1016/j.biotechadv.2017.03.003

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

Naumov, G.I., Naumova, E.S., Sniegowski, P.D., 1998. <i>Saccharomyces paradoxus</i> and <i>Saccharomyces cerevisiae</i> are associated with exudates of North American oaks. Canadian Journal of Microbiology 44, 1045–1050. https://doi.org/10.1139/cjm-44-11-1045

Nikulin, J., Vidgren, V., Krogerus, K., Magalhães, F., Valkeemäki, S., Kangas-Heiska, T., Gibson, B., 2020. Brewing potential of the wild yeast species Saccharomyces paradoxus. European Food Research and Technology 246, 2283–2297. https://doi.org/10.1007/s00217-020-03572-2

Ortiz, E.M., 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. https://doi.org/10.5281/ZENODO.2540861

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5

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, 299–306. https://doi.org/10.1111/j.1567-1364.2002.tb00048.x

Spurley, W.J., Fisher, K.J., Langdon, Q.K., Buh, K.V., Jarzyna, M., Haase, M.A.B., Sylvester, K., Moriarty, R.V., Rodriguez, D., Sheddan, A., Wright, S., Sorlie, L., Hulfachor, A.B., Opulente, D.A., Hittinger, C.T., 2022. Substrate, temperature, and geographical patterns among nearly 2000 natural yeast isolates. Yeast 39, 55–68. https://doi.org/10.1002/yea.3679

Stamatakis, A., 2014a. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steensels, J., Gallone, B., Verstrepen, K.J., 2021. Interspecific hybridization as a driver of fungal evolution and adaptation. Nature Reviews Microbiology 19, 485–500. https://doi.org/10.1038/s41579-021-00537-4

Stefanini, I., Dapporto, L., Berná, L., Polsinelli, M., Turillazzi, S., Cavalieri, D., 2016. Social wasps are a *Saccharomyces* mating nest. Proceedings of the National Academy of Sciences 113, 2247–2251. https://doi.org/10.1073/pnas.1516453113

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 109, 13398–13403. https://doi.org/10.1073/pnas.1208362109

Urbán Aragón, J.A., 2021. Estructura poblacional de levaduras Saccharomyces aisladas de fermentación de agave. Universidad Nacional Autónoma de México (UNAM).

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, 5404–5417. https://doi.org/10.1111/j.1365-294X.2012.05732.x

Wolters, J.F., Chiu, K., Fiumera, H.L., 2015. Population structure of mitochondrial genomes in Saccharomyces cerevisiae. BMC Genomics 16, 451. https://doi.org/10.1186/s12864-015-1664-4

Xia, W., Nielly-Thibault, L., Charron, G., Landry, C.R., Kasimer, D., Anderson, J.B., Kohn, L.M., 2017. Population genomics reveals structure at the individual, host-tree scale and persistence of genotypic variants of the undomesticated yeast *Saccharomyces paradoxus* in a natural woodland. Molecular Ecology 26, 995–1007. https://doi.org/10.1111/mec.13954