Fungal Ecology

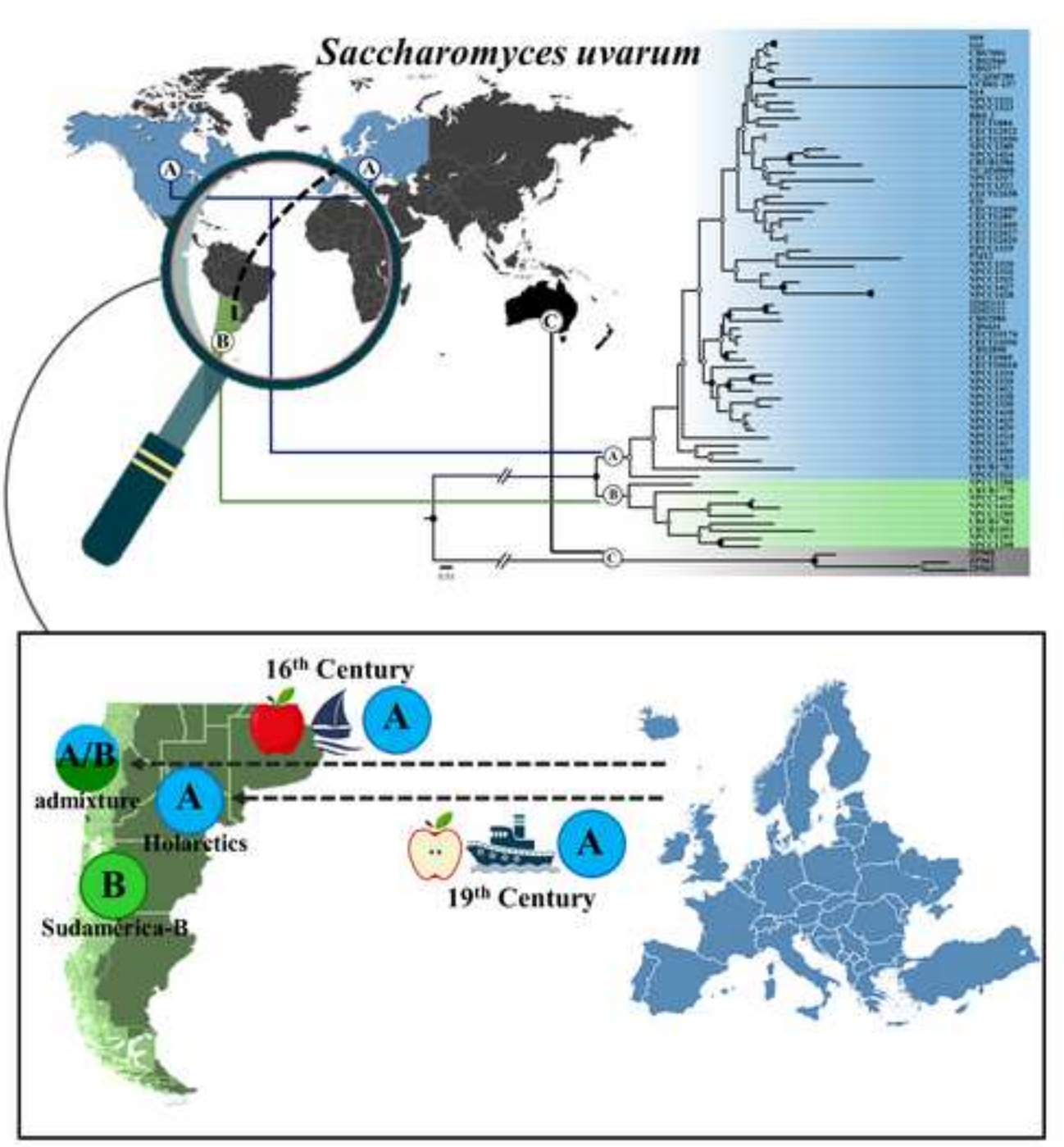
# HUMAN-ASSOCIATED MIGRATION OF HOLARCTIC Saccharomyces uvarum STRAINS TO PATAGONIA

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | FUNECO_2019_220R3 |
| **Article Type:** | Research Paper |
| **Keywords:** | Biodiversity; domestication; Saccharomyces uvarum; fermentation; natural habitat; dispersion; Patagonia |
| **Corresponding Author:** | Christian A Lopes<br>Consejo Nacional de Investigaciones Cientificas y Tecnicas<br>NEUQUEN, NEUQUEN ARGENTINA |
| **First Author:** | Melisa Gonzalez Flores |
| **Order of Authors:** | Melisa Gonzalez Flores |
| | María Eugenia Rodriguez |
| | David Peris |
| | Amparo Querol |
| | Eladio Barrio |
| | Christian A. Lopes |
| **Abstract:** | The domestication of Saccharomyces cerevisiae , as the result of its adaptation to human-manipulated processes , has been well analyzed by both phenotypic and genomic approaches. However, in other yeast species with industrial applications, such as Saccharomyces uvarum , these studies are very limited. To deepen knowledge about possible domestication in S. uvarum , an analysis of the genetic diversity of a series of S. uvarum strains isolated from different habitats was performed.<br>Our results show that the greatest S. uvarum population diversity worldwide is observed in Patagonia, where strains of this species can be isolated from industrial and traditional fermentations as well as from natural environments. This greater Patagonian diversity is due to the presence of strains belonging to two genetically differentiated populations, South America B (SA-B), and Holarctic/South America A (H/SA-A). The H/SA-A population of Patagonia is directly related to apple fermentation environments, mainly from cider fermentations but also, to a lesser extent, from traditional apple chicha. Our data suggest that strains from the Holarctic population colonized Patagonia. This is possibly associated with the introduction of apple trees by European immigrants, since the Spanish colonization of Chile in the 16 th century and the introduction of new apple tree cultivars in the upper valley of the Negro River, Argentina, during the 19 th century. During this process of colonization, Holarctic strains hybridized with SA-B yeasts, generating a series of admixed strains, mainly present in the traditional apple chicha fermentations. |
| **Suggested Reviewers:** | Rosa Pando Bedriñana<br>rpando@serida.org<br>Specialist in cider yeasts |
| | Huu-Vang Nguyen<br>vang08@yahoo.fr<br>Specialist in Saccharomyces uvarum diversity |
| | Matthias Sipiczki<br>gacela@post.sk<br>Specialist in Saccharomyces uvarum diversity |
| **Response to Reviewers:** | |

Graphical Abstract

Click here to access/download;Graphical Abstract;2.Graphical Abstract.png



# HUMAN-ASSOCIATED MIGRATION OF HOLARCTIC *Saccharomyces uvarum* STRAINS TO PATAGONIA

Melisa González Flores[1,2], María Eugenia Rodríguez[1,3], David Peris[4,5], Amparo Querol[4], Eladio Barrio[4,6] and Christian Ariel Lopes [1,2]

[1]Instituto de Investigación y Desarrollo en Ingeniería de Procesos, Biotecnología y Energías Alternativas (PROBIEN, Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina – Universidad Nacional del Comahue). Buenos Aires 1400, (8300) Neuquén, Argentina.

[2]Facultad de Ciencias Agrarias, Universidad Nacional del Comahue, Argentina.

[3]Facultad de Ciencias Médicas, Universidad Nacional del Comahue, Argentina.

[4]Instituto de Agroquímica y Tecnología de los Alimentos, IATA-CSIC. Paterna, Spain.

[5]Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, Oslo, Norway

[6]Departament de Genètica, Universitat de València, Valencia, Spain.

Keywords: *Saccharomyces uvarum*, domestication, biodiversity, fermentation, natural habitat, dispersion, Patagonia

*Correspondence: clopes@conicet.gov.ar

**ABSTRACT**

The domestication of *Saccharomyces cerevisiae*, as the result of its adaptation to human-manipulated processes, has been well analyzed by both phenotypic and genomic approaches. However, in other yeast species with industrial applications, such as *Saccharomyces uvarum*, these studies are very limited. To deepen knowledge about possible domestication in *S. uvarum*, an analysis of the genetic diversity of a series of *S. uvarum* strains isolated from different habitats was performed.

Our results show that the greatest *S. uvarum* population diversity worldwide is observed in Patagonia, where strains of this species can be isolated from industrial and traditional fermentations as well as from natural environments. This greater Patagonian diversity is due to the presence of strains belonging to two genetically differentiated populations, South America B (SA-B), and Holarctic/South America A (H/SA-A). The H/SA-A population of Patagonia is directly related to apple fermentation environments, mainly from cider fermentations but also, to a lesser extent, from traditional apple chicha. Our data suggest that strains from the Holarctic population colonized Patagonia. This is possibly associated with the introduction of apple trees by European immigrants, since the Spanish colonization of Chile in the 16$^{th}$ century and the introduction of new apple tree cultivars in the upper valley of the Negro River, Argentina, during the 19$^{th}$ century. During this process of colonization, Holarctic strains hybridized with SA-B yeasts, generating a series of admixed strains, mainly present in the traditional apple chicha fermentations.

## INTRODUCTION

Plants', animals', and microorganisms' exposure to anthropic environments has led to the emergence of adaptive divergence in response to human selection, a process known as domestication (Gibson and Liti, 2015, Goddard and Greig, 2015).

Wild populations from which domesticated species were originated are still present, constituting an interesting source of information for evolutionary studies related to human history. Among yeasts, *S. cerevisiae* species has been the most extensively studied case (Gibson and Liti, 2015; Gallone et al., 2016; Duan et al., 2018). However, domestication studies in other yeasts associated with anthropogenic environments are very limited.

A population genomic study, using whole-genome sequences of 54 strains of *S. uvarum*, suggested the existence of two differentiated populations in South America, named South America A (SA-A) and South America B (SA-B), with signs of sympatric diversification (Almeida et al., 2014). Strains from SA-A and SA-B populations were isolated from *Nothofagus* spp. trees in Patagonian forests, Argentina. A third population, Holarctic, closely related to SA-A, consisted of wild and industrial isolates from diverse geographic origins in the Northern Hemisphere. Finally, a fourth population, Australasia, contained wild isolates from New Zealand. This Australasian population was found to be the most distant lineage to South American and Holarctic populations (Almeida et al., 2014).

Bioprospecting efforts isolated additional *S. uvarum* strains from both *Araucaria araucana* forests and traditional fermentations made from feral apples (called apple chichas) in Patagonia (Rodriguez et al., 2014; 2017). Phylogenetic analyses demonstrated that most *S. uvarum* strains from apple chichas were closely related to the Holarctic strains (Rodríguez et al., 2017). Recently, a number of *S. uvarum* strains associated with ciders fermented at low temperatures in Patagonia were also identified (Gonzalez Flores et al., 2019).

Altogether, this extended *S. uvarum* strain collection suggests that Patagonia provides a diversity of *S. uvarum* strains associated with diverse habitats, from the most anthropic (cider) to the most natural habitats (*Nothofagus* and *Araucaria* forests), with interesting intermediate environments including the traditional apple chichas.

This work aimed to evaluate the phylogeographic relationships and genetic diversity of a complete set of *S. uvarum* strains obtained from different habitats at a local scale (Patagonia). Both phylogenetic and population analyses delimited the number of *S. uvarum* populations and the influence of humans on the microbial migration and diversity found in this region.

## MATERIALS AND METHODS

### 1. Yeast strains

A total of 123 *S. uvarum* Patagonian isolates, previously identified and molecularly characterized in our laboratory, were used in this study (55 isolated from cider, 47 isolates from apple chicha, and 21 from *A. araucana*) (Supplementary Table 1). Additionally, sequences from 29 European *S. uvarum* strains (Almeida et al., 2014; Perez-Tráves et al., 2014), from one North American strain (UCD61-137; Almeida et al., 2014), from five Northern Patagonian strains isolated from natural environments (strains CRUB1586; CRUB1778; CRUB1991; CRUB1783; CRUB1782), and from three New Zealand strains (strain ZP962; ZP963; ZP966; Almeida et al. 2014) were included for comparative purposes (Supplementary Table 1).

### 2. Phylogeographical analysis

All Northern Patagonian isolates were used for the phylogeographical analysis (Supplementary Table 1). The analysis was carried out with partial sequences of 10 nuclear

genes: *AHA1* and *BRE5* (located on different arms of chromosome XIVtIItIV), *CAF20* (chr. VIIItXV), *CAT8* (chr. XIII), *CYC3* (chr. I), *GAL4* (chr. XVI), *MET6* (chr. V), *EGT2* (chr. IItIItXIV), *SAE2* (chr. VII), and *VTA1* (chr. XII), and the mitochondrial gene *COX2*. The complete dataset explored 4536 bp of the nuclear genome and 512 bp of the mitochondrial genome. In both sequence data sets, heterozygous sites were not detected.

PCR-amplifications were performed as described by Rodríguez et al., (2017). PCR products were cleaned using the AccuPrep PCR purification kit (Bioneer, Inc, USA) and submitted to a sequencing service (Macrogen Korea). The nuclear gene sequences were deposited in the NCBI database, under accession numbers MK889401-MK889462, MK889432-MK889462, MK889463-MK889493, MH458339-MH458369, MH458370-MH458400 and MH458308-MH458338 and the mitochondrial *COX2* gene under accession numbers MK909807-MK909817. Each set of homologous sequences was aligned with the ClustalW program (Thompson et al., 1994) and concatenated with the Sequences Matrix 1.7.6 software (Vaidya et al., 2011).

**2.1. Geographical dispersion**

The study of the geographical dispersion was performed with the concatenated matrix of the 10 nuclear genes and the geo-references of the isolation sites of all isolates, to locate the haplotypes in proportion to the cartographic locations from each site. The haplotype pie charts representation on the map was constructed using the online software PhyloGeoWiz (http://phylogeoviz.org).

**2.2. Median Joining Network**

Haplotype classification was carried out in DnaSP v5 (Librado and Rozas, 2009). Median-joining (MJ) networks for concatenated matrix nuclear genes sequences were performed using Network 4.5 (Bandelt et al., 1999).

### 2.3. MisMatch Distribution (MMDI) analysis

The distribution of the pairwise differences between sequences (MMDI) makes it possible to describe the events of population dynamics that involve a change in the size of the population and could estimate the evolutionary forces that have occurred in the population. The distribution tends to be multimodal in populations in equilibrium and unimodal in populations that have suffered a bottleneck or the beginning of population expansion. The construction of the distribution is based on the initial theta ($\theta_0$) parameters that refer to the value before population expansion, final theta $\theta_1$ after population expansion, and tau (t) is an expansion time index expressed in units of mutational time (Rogers and Harpending, 1992).

The construction of the allele frequencies distribution was carried out using the Arlequin v3.5 software (Excoffier and Lischer, 2010). The Arlequin input file was generated by DnaSP v5 software (Librado and Rozas, 2009). Analyses were performed with and without geographical structuring. In the first case, groups correspond to sampling sites. These analyses were carried out with pairwise nucleotide differences, and including estimated parameters of demographic and spatial expansions. Subsequently, a neutrality test was performed to compare with the results obtained with MMDI by sampling site. The statistical significance of the MMDI distribution was estimated by the sum of the squares of the deviations (SSDI) (Schneider and Excoffier, 1999) where the observed distribution is compared with that expected under a null population expansion model.

### 3. Phylogenetic analysis

A set of 69 *S. uvarum* strains isolated from both natural and fermentative environments distributed in different countries of Europe, America and Oceania (Supplementary Table 1) were included for phylogenetic analyses using the 10 nuclear genes (mentioned above). The

three Australasian *S. uvarum* strains were used as outgroups in the analysis of nuclear genes. In the case of the mitochondrial gene *COX2*, 6 strains BR6-2, PM12, UCD-61-137, ZP962, ZP963, and ZP966 were not included because their mitochondrial sequences were not available. The best evolutionary models were selected with JModelTest 2.1.10 (Darriba et al., 2012), and the best-fitting model was the Tamura Nei, (1993). Phylogenetic trees were reconstructed by the Neighbor-Joining (NJ) method (Saitou and Nei, 1987) with 1000 bootstrap replicates for branch support, performed in MEGA7 (Kumar et al., 2016), as well as with the Neighbor-net (NN) method implemented in SplitsTree 4.14.3 (Huson and Bryant, 2006).

In addition, NJ trees and NN networks were obtained for each nuclear gene region to compare with the previous analyses and to determine, gene by gene, which alleles were present in each strain.

**4. Population structure analysis**

We used the individual-based Bayesian clustering methods implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000) to investigate population subdivisions with admixture in *S. uvarum*, with the concatenated sequence alignments of the nuclear gene regions. STRUCTURE was calculated for the 69 *S. uvarum* strains with 10 independent analyses carried out for each number of clusters K ($1 \leq K \leq 10$), with 500 000 MCMC interactions after a burn-in of 50 000 steps. We determined the significant number of groups (K) (Evanno et al., 2005) with the program STRUCTURE-HARVESTER version 0.6.94 (Earl and von Holdt, 2012). Allele frequencies were assumed to be correlated among populations, and the ancestry model allowed for admixture. CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) was used to compute the similarity coefficient between replicate

simulations of STRUCTURE. Structure plot V2.0 was used for graphical representations (Ramasamy et al., 2014).

**5. Genetic diversity among populations**

DnaSP v5 (Librado and Rozas, 2009) and Arlequin 3.1 (Excoffier and Lischer, 2010) were used to calculate genetic diversity statistics, such as nucleotide diversity ($\pi$), the number of segregating sites (S), the pairwise absolute divergence ($D_{XY}$) and the fixation index $F_{ST}$. Similarly, both Tajima's D (Tajima, 1989) and Fu's Fs (Fu, 1997) tests, as well as the nucleotide distances between populations, were calculated by using MEGA7 (Kumar et al., 2016).

The intraspecific haplotype variability (IHV) was calculated as the proportion between the number of haplotypes over the total of isolates from the sampling site.

**6. Recombination analysis**

Alignments of the 10 individual nuclear gene regions were analyzed with the RDPv4.97 software (Martin et al., 2015) to detect putative recombination events at the within-gene level. No significant recombination points were detected by at least two of the total methods implemented in RDP after applying Bonferroni corrections for multiple comparisons.

**7. Evolutionary relationships in *S. uvarum***

A time tree for the *S. uvarum* concatenated sequences was inferred using the Reltime method (Tamura et al., 2012) implemented in MEGA7 (Kumar et al., 2016). For this purpose, we excluded those strains that appeared as admixed in the STRUCTURE analysis, resulting in a set of 51 strains. The analysis was based only on synonymous substitution (supposedly neutral) and the distances obtained with the Li-Wu-Luo method (Li et al., 1985). The time tree was computed using 1 calibration constraint, a time of divergence of 35-80 Mya between the Australasian and the South America B populations, based on the

paleogeographic reconstructions and biogeographic patterns reported by Sanmartín and Ronquist, (2004).

## RESULTS

### 1. Population dynamic in Patagonian *S. uvarum* strains

To evaluate the phylogeographic relationship and detect evidence of gene flow among *S. uvarum* strains from different origins in Patagonia, a multilocus sequencing approach using 10 nuclear genes was carried out for a set of 123 isolates from natural habitats, apple chichas and ciders (Supplementary Table S1). A geographical representation of the dispersal of this species in Patagonia, including the number of isolates obtained from each sampling area, is depicted in Figure 1A.

The concatenated sequence data obtained from the 10 nuclear genes analyzed, allowed the identification of a total of 31 different haplotypes. Only one haplotype, represented by the isolate NPCC1309, was found in different sampling sites, including Villarrica, Tirúa, and Pucón. The haplotype diversity was indicated in Figure 1B and the relation between isolates and haplotypes by sampling site was included in Table 1. The lowest diversity (7 %) was observed among the isolates from cider, a well-established industrial fermentation; while the highest diversity values were found in natural habitats (50-54 %) as well as in apple chicha, a traditional fermentation process (30-50 %) (Figure 1B).

A haplotype network was obtained using the sequences of the 10 nuclear genes of all isolates from the Patagonia (Figure 1C). For this analysis, five additional strains obtained from natural habitats in a previous study (Almeida et al., 2014) were included as references. Two groups, named i and ii, were clearly differentiated in the haplotype network. The absence of subtended edges connecting a haplotype with two differentiated groups might indicate either limited or null gene flow between groups. It is interesting to note that group i

included strains from three origins: industrial processes (cider), traditional fermentation (apple chicha), and natural environments. The branched structure of this group indicates a close relationship among these strains (Figure 1C).

Haplotypes from sampling areas Cipolletti, Tromen, Villarrica, and Tirúa (indicated with dotted lines in Figure 1C) clustered together, supporting a putative common origin. In contrast, haplotypes obtained from Huechulafquen and Pucón did not show geographical structure, indicating possible migration events in these areas.

We performed a Mismatch Distribution (MMDI) analysis of the Patagonian strains. All 123 Patagonian isolates were firstly analyzed together (general analysis), and the same analysis was repeated taking into consideration each particular sampling area (Figure 2). In the general analysis, the model predicted a multimodal distribution, congruent with a population in geographical equilibrium. The same distribution was observed in most sites analyzed separately, except for the Cipolletti and Pucón areas that exhibited unimodal and bimodal distributions, respectively (Figure 2).

In all sites showing multimodal distribution, the neutrality tests indicated population expansion events (Table 1). This can be observed mainly in the Huechulafquen and Tromen areas, where both indices (Tajima's D and Fs) were significant and the $\theta_1$ values were very high with respect to the initial population $\theta_0$.

In the Cipolletti sampling area, all 55 isolates were obtained from a unique fermentation process, a cider spontaneously fermented at 13°C (González Flores et al., 2019). Although the number of isolates was relatively high, both the sampling sources and the number of haplotypes (4) were very low. In this sampling area, the SSD adjustment of MMDI and the neutrality tests (Tajima's D and Fs) with positive and significant values, suggest a possible recent selective sweep or a population contraction after a recent bottleneck (Table 1). Also,

the comparison of both $\theta_0$ and $\theta_1$ indices (Table 1), suggests a lack of population growth in relation to the other origins. In the Pucón sampling area, the adjustment of the MMDI model was not significant, indicating a population bias probably due to the low number of samples (Table 1).

**2. Patagonian *S. uvarum* strains belong to the South America A/Holarctic lineage**

The phylogenetic relationships within the complete set of *S. uvarum* strains from Patagonia was inferred through the reconstruction of both a NJ (Neighbor Joining) phylogenetic tree and a NN (Neighbor Net) phylogenetic network. A randomly selected strain for each haplotype (Figure 1C) was included in the NJ phylogenetic analysis (Supplementary Table S1). Representative strains from each lineage described by Almeida et al., (2014) were also included in this analysis (Figure 3).

Three clades, named A, B and C, were supported by high bootstrap values (Figure 3 and Supplementary Figure S1). Clade A included strains from Holarctic (H) and South America A (SA-A) populations defined by Almeida et al. (2014), and it was highly heterogeneous due to the large geographic distribution of the strains (Europe, North, and South America) and the diverse ecological niches: natural and anthropic (fermented beverages). All cider strains from Cipolletti (site F in Figure 1B) grouped with clade A.

Clade B contains the South American (SA-B) strains, and it was only composed by strains isolated from natural habitats in Patagonia, including all strains from Tromen (site A in Figure 1B) and the strain NPCC1293 isolated from Huechulafquen (site B in Figure 1B).

Finally, clade C contains the Australasian population.

In order to associate the phylogenetic clades to populations, a population structure analysis was carried out with STRUCTURE software. Population structure inference indicated the presence of two highly supported genetic clusters or populations (K=2) among the analyzed

strains. One of these clusters was assigned to the Australasian strains and the other cluster contained all remaining isolates (Figure 3). In a second analysis, we excluded the Australasian strains due to marked divergence with the rest of the strains. In this analysis, an optimal K of 2 was obtained, supporting the phylogenetic result. K1 contained Holarctic/South America A (H/SA-A) lineages defined previously by Almeida et al., 2014, and K2 contains all SA-B strains (Figure 3 and Supplementary Figure S2). Due to the impossibility of differentiating the SA-A population from the Holarctic in our analysis, we used a unique name H/SA-A to refer to strains of both populations.

The population structure analyses were represented according to the phylogenetic tree position of the strains. The same analyses ordered by Q values were depicted in Supplementary Figure S2 to differentiate the populations and the admixed strains.

In the population structure analysis carried out excluding Australasian strains, one population was coincident with clade B (SA-B) in the NJ tree. This population was composed of strains isolated from natural environments, including all those isolated from Tromen (site A in Figure 1A) as well as strains CRUB1783, 1778, and 1991, belonging to population SA-B according to Almeida et al., (2014). The second population corresponded to clade A (H/SA-A) in the NJ tree and grouped all European and North American strains, as well as all Patagonian strains isolated from apple fermentation (apple chicha and cider). Finally, 16 strains included in Clade A (H/SA-A) according to the NJ tree, showed different degrees of admixture between H/SA-A and SA-B. All of them were obtained from Patagonia, except for strains UCD61-132 and CECT12600 from North America and Spain, respectively. Seven of the Patagonian admixed strains were obtained from natural environments, and the other nine were isolated from apple fermentations carried out in different sampling locations (Cipolletti, Tirúa, Pucón, and Villarrica) (Figure 3).

Alignments of the 10 individual nuclear gene regions were analyzed with the RDPv4.97 software (Martin et al., 2015). No significant recombination points were detected by at least two of the methods implemented in RDP, after applying Bonferroni corrections for multiple comparisons.

Once intragenic recombination was discarded as the responsible for admixture genotypes, gene-by-gene NJ (data not shown) and NN phylogenetic reconstructions were obtained to determine if allele segregation after hybridization could explain the admixture patterns. Neighbor Net reconstructions are depicted in Supplementary Figure S3, A to J. The phylogenetic reconstructions of these ten genes were consistent with the results obtained in the population structure analysis. In 7 gene phylogenies, alleles present in pure strains from clades A (H/SA-A) and B (SA-B) appeared in two different groups. One group included alleles from clade A pure strains and the other group included those from clade B pure strains. Alleles from admixed strains appeared in one group or the other, depending on the gene. The exceptions were genes *BRE5, MET6,* and *SAE2*. In the case of *BRE5*, two groups of alleles were clearly separated; however, several Holarctic strains that appeared as belonging to clade A (H/SA-A) according to the population structure analysis, shared alleles with some of the pure strains from clade B (SA-B). In the other two cases (*MET6* and *SAE2* genes), alleles were intermingled and could not be separated into two clear groups. Ancestral polymorphism segregation could be responsible for the low discriminatory power of these genes, as well as the results obtained with *BRE5*.

Therefore, we can classify the alleles of the strains according to the groups in which they cluster for each gene (Supplementary Figure S4). This way, strains significantly belonging to clades A (H/SA-A) or B (SA-B), contained 9-10 alleles included in the same group, and admixed strains contained different combinations of alleles from both groups, except for

the North American strain UCD61-137, considered as admixed in the population structure analysis but containing no SA-B allele.

Groups by population structure and geographic location were considered to perform divergence and diversity comparisons among them. The Holarctic population was divided into all the pure Holarctic strains from the Northern Hemisphere (HOL-EU) and from Patagonia (HOL-PAT). Also, in Patagonia, the Admixture group (ADM)- formed by strains with different degrees of mixture of the HOL and SA-B population- and SA-B monophyletic group were observed.

The nucleotide diversity ($\pi$) calculated for all the strains used in this analysis, excluding the strains of the Australasia population (AUS), was $\pi = 0.0061$. However, the $\pi$ value obtained for all Patagonian strains (including those isolated from both natural and fermentative environments) was $\pi = 0.0057$ and the one obtained for the European Holarctic strains (HOL-EU) was $\pi = 0.0032$. This difference indicates that diversity is higher in Patagonia than in all the Northern Hemisphere. An explanation for this difference could be the great diversity observed within the strains included in SA-B population ($\pi = 0.0052$) as well as in the admixture strains (ADM) where the diversity was $\pi = 0.0055$ (17 Patagonian strains with different degrees of admixture were considered to calculate this value, as indicated in Supplementary Figure S2). When non-admixed Holarctic strains isolated from fermentative environments in Patagonia (HOL-PAT) were analyzed, the nucleotide diversity ($\pi =$ 0.0032) was similar to that observed in the HOL-EU strains.

Both the pairwise absolute divergence ($D_{XY}$) and the fixation index $F_{ST}$ were calculated among these populations (Table 2). Results evidenced the lowest $D_{XY}$ and the highest $F_{ST}$ between the HOL-EU and HOL-PAT populations, indicating a putative common origin (Table 2). The SA-B population showed the same degree of divergence (according to both

$D_{XY}$ and $F_{ST}$ indices) with both HOL-EU and HOL-PAT populations. On the other hand, the ADM strains showed the lowest $D_{XY}$ and the highest $F_{ST}$ when compared with both HOL-EU and HOL-PAT populations and SA-B population, indicating that ADM strains are more similar to the Holarctic population (Table 2). These results are congruent with the population structure analysis (Supplementary Figure S2) and with the position these strains occupied in the phylogenetic tree (Clade A, Figure 3). Finally, the AUS population is the most divergent compared to any of the other populations (Table 2).

**3. Time tree in *S. uvarum***

A divergence time tree analysis was performed to estimate the most ancestral population in Patagonia (Figure 4). All 18 admixed strains were removed from this analysis and only the synonymous sites were used. With this new alignment, a tree topology similar to that depicted in Figure 3 was obtained. The calibration of the time tree was based on the estimates of 35-80 Mya for the divergence of the Australasian (corresponding to clade C) and the Southern South American (corresponding to clade B) biogeographic areas proposed by Sanmartín and Ronquist, (2004), as shown by different paleogeographic reconstructions and biogeographic evidence. According to these approximate divergence times, and in agreement with the nucleotide diversity estimates, the SA-B population is the early divergent Patagonian lineage and it is autochthonous to Patagonia (Figures 3 and 4).

**4. Phylogeny of *S. uvarum* at the mitochondrial level**

To determine the level of congruence between nuclear and mitochondrial genomes the sequences of the mitochondrial gene *COX2* were analyzed for the 63 *S. uvarum* strains. Strains ZP962, ZP963, ZP966, PM12, BR6-2, and UCD61-137 were not included in this analysis because their mitochondrial DNA sequences were not available. *COX2* gene was selected because of its high discriminatory power to differentiate *S. uvarum* strains (Pérez-

Través et al., 2014; Rodríguez et al., 2017). The Neighbor Net phylogenetic analysis (Figure 5A), evidenced two big groups named i and ii. Group i included only strains from Patagonian natural habitats. This group was formed mainly by strains from Huechulafquen including admixture strains (according to the population structure analysis performed with the nuclear genetic structure) as well as the strain NPCC1293 belonging to the SA-B population. Group ii was more heterogeneous and included all the European strains (from both natural and fermentative environments), all Patagonian strains from fermentation as well as some Patagonian strains from natural environment (Figure 5A). Within this large group, three haplotypes (named H1-H3) included all European strains and most fermentative strains from Patagonia. A distant subclade appears in the group ii that includes admixed strains (according to their nuclear gene sequences) isolated from apple chicha in Tirúa. Interestingly, these strains showed a small portion of 40 bp (from the 472 bp position of the amplicon) in their *COX2* sequences that showed 100% similarity with *S. cerevisiae* S288c *COX2* (Supplementary Table S2). This could be due to an introgression event at the mitochondrial level as previously described for many *Saccharomyces* yeasts (Peris et al., 2017). By excluding this region from the analysis, the new NN phylogenetic network was less branched (Figure 5B), with two large groups of similar characteristics to those observed previously (Figure 5A). However, in this new analysis, the small group of strains with the *S. cerevisiae* introgression is located within the group i, together with strains from natural habitats (Figure 5B).

To compare the mitochondrial information with the nuclear data, NJ trees were also obtained with the *COX2* sequences, including and excluding the *S. cerevisiae* introgressed region. These NJ trees showed topologies consistent with the NN networks, although the NJ trees were less informative (Supplementary Figure S5). Both the structure and topology

observed at the nuclear level were not the same observed at the mitochondrial level, even using the same methodology of analysis. These results suggest that the evolutionary histories of both nuclear and mitochondrial genomes were not coincidental, probably due to hybridization and segregation between the SA-B and H/SA-A populations, which seems more evident at the mitochondrial level.

Regarding the nucleotide diversity in the mitochondrial gene *COX2*, the $\pi$ values calculated with and without the DNA region from *S. cerevisiae* were 0.137845 and 0.10964, respectively. Comparing this index with the nucleotide diversity calculated at the nuclear level, similar, and even higher values were observed (considering the data with the *S. cerevisiae* region), indicating that this mitochondrial gene is highly polymorphic. Finally, the Neutrality Test indicates a population expansion model similar to that observed at the nuclear level, although the results were very different when considering the data with or without the *S. cerevisiae* portion (Fs: -25.6915 and D: -1.91995 vs Fs: -1.48692 and D: -2.04246, respectively).

## DISCUSSION

### 1. Introduction of domesticated Holarctic *S. uvarum* strains into Patagonia

Besides the well-reported presence of *S. uvarum* in natural habitats in Patagonia (Libkind et al., 2011; Rodríguez et al., 2014), the recent discovery of this species from apple chichas (Rodríguez et al., 2017) and ciders (Gonzalez Flores et al., 2019), fermented at low temperature in the same region, bring out a new interesting picture about the diversity of this species in South America. In the present work, phylogenetic analysis carried out using multilocus-sequence data evidenced the presence of three clades that seem to correspond with those named A, B, and C obtained using a whole-genome sequencing approach by Almeida et al., (2014). Additionally, population structure analysis supports the existence of

two populations in Patagonia named SA-B and H/SA-A, the last one including those identified as SA-A and Holarctic by Almeida et al., (2014). These authors proposed the differentiation of both SA-A and the Holarctic populations based on the geographic origin of the strains and the basal position of the SA-A strains in the clade A of the phylogenomic tree. They also suggested that SA-A is an autochthonous population from Patagonia that originated the Holarctic population (in Europe), manifesting a migratory flow direction from Patagonia to Europe. The authors proposed that the differentiation between populations SA-A and SA-B in sympatry in Patagonia could be a result of geographical barriers that separated *Nothofagus* species during the glaciation periods (Mathiansen et al., 2010). However, the presence of admixed strains (between SA-B and Holarctic population) inside SA-A population was puzzling.

The inclusion of our new set of strains from fermentative environments in Patagonia supports the previous phylogenetic structure (Almeida et al., 2014); however, SA-A and HOL populations were not displayed as two independent clades. It might suggest that both lineages are subpopulations of a unique H/SA-A population. Additionally, the SA-A population (that now included all *S. uvarum* strains isolated from Patagonian apple chicha and cider) does not appear to be pure, but rather a mixture between SA-B and HOL populations.

Taking into consideration that SA-B diverged earlier than the HOL population and that it is the most biodiverse group, it seems probable that the common ancestor of both populations was autochthonous from Patagonia, a conclusion supported by both the divergence tree and the highest diversity in the SA-B population. The discovery of several HOL strains in Patagonia and the absence of SA-B strains in Europe, support the hypothesis of a later introduction of European Holarctic strains in Patagonia (colonization hypothesis). This

phenomenon, previously suggested by Rodriguez et al., (2017), is reinforced in this work according to our new results. The colonization hypothesis is also supported by the presence of clearly Holarctic strains in Patagonian fermentations. The arrival of these "immigrant" strains can also explain the presence of admixed strains only in Patagonia (and not in Europe, which would confirm the direction of this migratory flow) as a result of hybridization events with the autochthonous population SA-B.

The most likely scenario for the introduction of the Holarctic population in Patagonia is through apple introduction because the purely Holarctic populations of Patagonia were found only in fermentative environments related to this fruit (both chicha and cider). Additionally, *S. uvarum* has frequently been isolated from European ciders, reaching similar or even higher population levels than the ones of *S. cerevisiae* (Morrissey et al., 2004; Coton et al., 2006; Suárez Valles et al., 2007). The history of the introduction of apple in Chilean Patagonia (where apple chicha strains were isolated) is different from the Argentinian one (strains isolated from cider). Apple trees were planted in Chile by Spanish colonizers during the 16$^{th}$ century and later dispersed by nomadic Mapuche communities in the Andean Northern Patagonia (Calvo et al., 2010; Rodríguez et al., 2017). During the 19th century, new European immigrants established in the upper valley of the Negro River (Cipolletti area, Argentina), and introduced new domesticated cultivars of apple trees and vines (Bendini, 2005). Based on these historical facts, we could suggest different migration events of the *S. uvarum* Holarctic population detected in apple chichas (Chile) and ciders (Argentina). This phenomenon might explain the higher genetic variability in apple chichas than in ciders (purely Holarctic), mainly due to the presence of mixed strains with the SA-B population, and the possible event of contraction in the population size detected in the MMDI analysis in the last strains.

In a deeper analysis, we observed that in the Tirúa region there are 70% purely Holarctic strains and 30% admixture strains bearing less than 28% of SA-B alleles. Interestingly, this sampling site is the furthest from the places where representatives of the population SA-B are located. In Villarrica and Pucón regions, closer to the locations where SA-B population strains are frequently isolated, the percentage of admixture is higher (between 50 and 100% of the strains). Finally, in Cipolletti region, the percentage of purely Holarctic strains was 75% and only one admixed strain bearing less than 10% of SA-B alleles was detected. Interestingly, in this cider producer region from Argentina, apple cultivars were introduced 300 years later than the first introduction event in Chile.

The phylogeographic events that are directly related to the different habitat and the complex population dynamics observed in *S. uvarum* in Patagonia have also been described for other species. In general, it has been assumed that neutral selection events are the main forces governing in natural habitats, contrarily to the directed selective pressures observed in the anthropic environments. In this sense, Arias, (2008) evaluated the effect of neutral evolution in wine and non-wine populations of *S. cerevisiae*. The author evidenced that the wine strains were subjected to a positive (or directional) selection force favouring a certain type of genotypes, while neutral evolution was observed within the non-wine strains. In the most anthropic environment evaluated in this work (cider), a unimodal model shifting to the left was observed indicating a marked effect of directional selection and a contraction with a drastic selective sweep of the population. This model evidences an artificial selection (domestication) event occurring in this anthropic environment. This type of selection is likely to occur in the most anthropic environment possessing a series of yeast stressful practices that are not present in traditional fermentations, such as the use of sulfite as an antimicrobial agent (Pérez-Ortín et al., 2002; González Flores et al., 2019).

Regarding migratory events, human-associated microbial dispersions have already been described for *Saccharomyces paradoxus* (Zhang et al., 2010). *S. paradoxus* strains isolated from New Zealand oaks were genetically related to European strains of the same species. The authors suggested that these strains may have been introduced recently in association with the oak acorns (*Quercus robur*) brought in by European immigrants.

Also, it has already been described that one important factor contributing to the adaptive potential of introduced species is the generation of intraspecific hybrids (admixis). The benefits of admixture include increased genetic variation as well as the formation of individuals with novel trait combinations through segregation and recombination (Dlugosch et al., 2008). The advantage of these admixed individuals may be modest in the native environment where local adaptations are well established (Verhoeven, 2010). However, they can become relevant in new environments such as apple fermentation. Interestingly, most admixture strains were found in fermentative environments. Moreover, the low number of admixture strains in natural environments denotes that the cost of maintaining hybrid genomes is greater than the adaptive benefit obtained by yeasts in these natural environments, in which ancestral genomes are well adapted.

**2. Complex ancestries in the *S. uvarum* strains from Patagonia are also evidenced in the mitochondrial DNA**

The analysis of the sequences of the mitochondrial gene *COX2* revealed a much more complex history than that resolved by the analysis of the nuclear genome. This complexity included the existence of gene flow between both H/SA-A and SA-B populations detected in Patagonia. This flow is particularly evidenced in a group of strains included in Group i of the Neighbor-net phylogenetic network of the *S. uvarum COX2* mitochondrial gene. This group, genetically distant from the rest, included strains belonging to population SA-B, as

well as strains belonging to population H/SA-A, all of them associated with natural habitats in Patagonia.

Although the analysis of the whole mitochondrial genome, or at least several genes, could yield more accurate results, the situation observed in the *COX2* gene also evidences the existence of a certain degree of contact between fermentative and natural populations of this species in Patagonia.

Finally, it is interesting to note the presence of introgression of *S. cerevisiae COX2* among the analyzed strains. These are typical events associated with a reticulated evolution, which can rapidly confer adaptive advantages to the organisms during the colonization of a new niche, as described by Peris et al., (2014; 2017). Several examples of introgressions have been described in *S. eubayanus* (Peris et al., 2017), *S. uvarum* (Almeida et al., 2014), and *S. cerevisiae* (Legras et al., 2018). In this work, the *S. cerevisiae* fragment was observed in four strains obtained from apple chicha (Tirúa area). This fragment could confer an adaptive advantage since the same recombination site was previously reported by Peris et al., (2017) as a hot spot of recombination associated with a GC cluster.

## CONCLUSIONS

Patagonia is an ideal place to study the evolutionary dynamics and natural history of *S. uvarum*. Results obtained in this work added to those generated in a few previous works about this species (Almeida et al., 2014; Rodriguez et al., 2014), depicting a more complete scenario of the *S. uvarum* population dynamics. Some of the most important conclusions obtained from our work are: i) there is no SA-A population as such, but rather it is a paraphylic group within the Holarctic population, represented by admixture strains between the SA-B and Holarctic populations described previously; ii) SA-B and H/SA-A

populations, which separated a long time ago, together with the low diversity detected in the H/SA-A population, are evidence of a founding effect from an ancestral population localized in Patagonia which caused a clear differentiation of both SA-B and H/SA-A populations; iii) there has been a recent introduction or migratory invasion event of the Holarctic population into Patagonia and this population has made contact with the ancestral population (SA-B) allowing the generation of admixture strains in Patagonia associated with fermentative environments; iv) an invasion phenomenon has been detected in fermentative environments, which promotes the selection of Holarctic alleles or the generation of admixture events and restricts the scenario in Patagonia to strains from fermentative habitats (purely Holarctic or admixed) and natural strains (purely SA-B). This implies that both populations are being selected against directions given the adaptation to the different environments, which promotes a barrier not so much geographic but rather adaptive, and even allows a sympatric evolution in Patagonia, encouraging the preservation of the great diversity reported for the species in South America.

## ACKNOWLEDGMENTS

This work was supported by grants PICT 2015-1198 from ANPCyT, PIP 2015-555 from CONICET and PI04-A128 from Universidad Nacional del Comahue (Argentina) to C.A.L, as well as Grants RTI2018-093744-B-C31 and RTI2018-093744-B-C32 from the Spanish Government and FEDER to A.Q. and E.B. respectively. M.G.F thanks CONICET for her Postdoctoral fellowship and BEC.ar (Argentina) program for financing her stay at the University of Valencia, Spain.

DP thanks the Marie Sklodowska-Curie fellowship of the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 747775).

## REFERENCES

Almeida, P., Goncalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masenuf-Pomarede,

I., Albertin, W., Durrens, P., Sherman, D.J., Marullo, P., Hittinger, C.T., Sampaio, J.P., 2014. A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. Nat. Commun. 5, 4044.

Arias, J.A., 2008. Diversidad genética en las especies del complejo *Saccharomyces sensu stricto* de fermentaciones tradicionales. Doctorate Thesis, Universitat de València, Spain. https://www.tdx.cat/handle/10803/9940.

Bandelt, H.J., Forster, P., Arne Rohl., 1999. Median-Joining Networks for Inferring Intraspecific Phylogenies. Mol. Biol. Evol. 16, 37–48.

Bendini, M.I., 2005. Fruticultura en el norte de la Patagonia: procesos de cambio y tramas sociales. Press. Univ. du Mirail 131–148.

Calvo, P., Carrera, A., Sánchez, E., Poverene, M., 2010. Genetic Diversity in Wild Apple (*Malus sp*.) Populations in Argentina. Am. J. Plant Sci. Biotechnol. 3, 99–105.

Coton, E., Coton, M., Levert, D., Casaregola, S., Sohier, D., 2006. Yeast ecology in French cider and black olive natural fermentations. Int. J. Food Microbiol. 108, 130–135.

Darriba,D., Taboada, G.L., Doallo, R., Posadas, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. 9 (8) 1-772.

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. Nat. Commun. 9, 1–13.

Dlugosch, K.M., Parker, I.M., 2008. Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. Mol. Ecol. 17, 431–449.

Earl, D., and Vonholdt, B., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet. Resour. 4, 359-361.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14**,** 2611-2620.

Excoffier, L., and Lischer, H.E.L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564-567.

Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147, 915–925.

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., Verstrepen, K.J., 2016. Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. Cell 166, 1397–1410.

Gibson, B., Liti, G., 2015. *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. Yeast 32, 17–27.

Goddard, M.R., Greig, D., 2015. *Saccharomyces cerevisiae*: a nomadic yeast with no niche? FEMS Yeast Res. 15, 1–6.

González Flores, M., Rodríguez, M.E., Origone, A.C., Oteiza, J.M., Querol, A., Lopes, 2019. *Saccharomyces uvarum* isolated from Patagonian ciders shows excellent fermentative performance for low temperature cidermaking. Food Res. Int. 126, 1-11.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Jakobsson, M., Rosenberg, N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801-1806.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Legras, J.L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sánchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldón, T., Schuller, D., Sampaio, J.P., Dequin, S., 2018. Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. Mol. Biol. Evol. 35, 1712–1727.

Li, W.-H., Wu, C.-I., Luo C.-C., 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2, 150-174.

Libkind, D., Hittinger, C.T., C., Valério, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P., Sampaio, J.P., 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. PNAS 108, 1–6.

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Mathiasen, P., Premoli, A.C., 2010. Out in the cold : genetic variation of *Nothofagus pumilio* (*Nothofagaceae*) provides evidence for latitudinally distinct evolutionary histories in austral South America. Mol. Ecol. 19, 371–385.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, 1-5.

Morrissey, W.F., Davenport, B., Querol, A., Dobson, A.D.W., 2004. The role of indigenous yeasts in traditional Irish cider fermentations. J. Appl. Microbiol. 97, 647–655.

Pérez-Ortín, J.E., Querol, A., Puig, S., Barrio, E., 2002. Molecular Characterization of a Chromosomal Rearrangement Involved in the Adaptive Evolution of Yeast Strains. Genome Res. 12, 1533–1539.

Pérez-Través, L., Lopes, C. A., Barrio, E., & Querol, A. (2014). On the Complexity of the Saccharomyces bayanus Taxon: Hybridization and Potential Hybrid Speciation. PLoS ONE. 9(4), 1-14.

Peris, D., Arias, J.A., Orlić, S., Belloch, C., Pérez-Través, L., Querol, A., Barrio, E., 2017 Mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species. Mol. Phylogenet. Evol. 108, 49-60

Peris, D., Sylvester, K., Libkind, D., Gonçalves, P., Sampaio, J.P., Alexander,W.G., Hittinger,C.T., 2014. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. Mol. Ecol. 23, 2031-2045.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics. 155, 945-959.

Ramasamy, R.K., Ramasamy, S., Bindroo, B.B., Naik, V.G., 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3, 431.

Rodríguez, M.E., Pérez-Tráves, L., Sangorrín, M.P., Barrio, E., Querol, A., Lopes, C.A., 2017. *Saccharomyces uvarum* is responsible for the traditional fermentation of apple chicha in Patagonia. FEMS Yeast Res. 17, 1–11.

Rodríguez, M.E., Pérez-Tráves, L., Sangorrín, M.P., Barrio, E., Lopes, C.A., 2014. *Saccharomyces eubayanus* and *Saccharomyces uvarum* associated with the

fermentation of *Araucaria araucana* seeds in Patagonia. FEMS Yeast Res. 14, 948–965.

Rogers, A.R., Harpending, H.C., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9, 552-569.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evo. 4, 406–425.

Sanmartín, I., Ronquist, F., 2004. Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. Syst. Biol. 53, 278-298.

Schneider, S., Excoffier, L., 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. Genetics. 152, 1079–1089.

Suárez-Vallés, B., Pando Bedriñana, R., Fernández-Tascón, N., Querol, A., Rodríguez-Madrera, R., 2007. Yeast species associated with the spontaneous fermentation of cider. Food Microbiol. 24, 25–31.

Tajima, F., 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA polymorphism. Genetics 123, 585–595.

Tamura, K., Battistuzzi, F.U., Billing-Ross, P., Murillo, O., Filipski, A., Kumar, S., 2012. Estimating divergence times in large molecular phylogenies. PNAS. 109, 19333–19338.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171–180.

Verhoeven, K.J.F., Macel, M., Wolfe, L.M., Biere, A., 2010. Population admixture , biological invasions and the balance between local adaptation and inbreeding depression. Proc. R. Soc. B Biol. Sci. 1–7.

Zhang, H., Skelton, A., Gardner, R.C., Goddard, M.R., 2010. Saccharomycesparadoxus and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: evidence for migration from Europe and interspecies hybrids. FEMS Yeast Res. 10, 941–947.

## FIGURE LEGENDS

**Figure 1: A.** Geographic distribution of *S. uvarum* isolates. The circle size indicates the number of isolates per sampling area. **B.** Bar plot of the frequency of haplotypes for each sampling area. A: Tromen, Argentina; B: Huechulafquen, Argentina; C: Tirúa, Chile; D: Villarrica, Chile; E: Pucón, Chile; F: Cipolletti, Argentina. **IHV**: The intraspecific haplotype variability represents the proportion between the number of haplotypes over the total of isolates from the sampling site. The arrows indicate the only haplotype that was isolated in different sampling areas. **C.** Median Joining (MJ) network net reconstructed using the concatenated partial sequences of ten nuclear genes (*AHA1*, *BRE5*, CAF20, *CAT8*, *CYC3*, *GAL4*, *MET6, EGT2, SAE2*, and *VTA1*). All the strains belong to North Patagonian Culture Collection NPCC (Neuquén, Argentina), except for the strains CRUB1778, CRUB1586, CURB1991, CRUB1782, CRUB1783 used as external reference yeasts. Circles represent individual haplotypes, and the circle size is correlated with the number of isolates of each haplotype, scaled according to the legend.

**Figure 2:** Mismatch distribution (MMDI) analysis of all the haplotypes of *S. uvarum* of Northern Patagonia. Dotted line, the observed pairwise difference data are represented, and continuous line, those predicted by the model.

**Figure 3:** Neighbor-Joining (NJ) phylogeny of 69 strains based on concatenated (4536 positions in the final dataset) nuclear genes (*AHA1*, *BRE5*, *CAF20*, *CAT8*, *CYC3*, *GAL4*, *MET6*, *EGT2*, *SAE2*, and *VTA1*). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura Nei model (Tamura and Nei, 1993) method and expressed as the number of base substitutions per site. The three main clades are marked by letters A, B, and C. P The population structure analysis is represented laterally to the right of the phylogenetic tree. The optimal K=2 was determined according to the Evanno et al. (2005) Test. Wo AU: Population structure without the Australasian strains. W AU: Population structure including Australasian strains. Both panels to the right of the tree indicate the origin and source of isolation of the strains. Origin: South America in white, Europe in black, North America in plaid and Oceania striped. Source: Wild environments in green, apple chicha in pink, cider in blue, wine in purple, beer in brown and unknown in grey.

**Figure 4.** Time tree inferred using the Reltime method and estimates of branch lengths inferred using the Neighbor-Joining method. The time tree was computed using 1 calibration constraints. The analysis involved 51 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1494 positions in the final dataset. In blue, green, and black, all the pure Holarctics, SA-B and Australasian strains were collapsed. The numbers in bold located above each clade represent the millions of years (Myr) in which the clade diverged. Italic numbers under each clade indicate bootstraps values. Values below H/SA-A, SA-B, and Australasian labels indicate the nucleotide distance ± sd inside the clade. The values in the middle of the arrows indicate the nucleotide distance between clades.

**Figure 5.** Neighbor-net phylogenetic network of the *S. uvarum COX2* mitochondrial gene. The scale represents the number of nucleotide substitutions per site. **A:** Neighbor-net including the introgressed region (512pb). **B:** Neighbor-net phylogenetic network excluding the introgressed region (472pb). Both panels to the left of the label indicate the origin and source of the strains. Origin: South America in white, Europe in black. Source: Wild environments in green, apple chicha in pink, cider in blue, wine in purple, beer in brown and unknown origin in grey.

**Supplementary Figure 1**. Neighbor-net phylogenetic network of the *S. uvarum* using ten nuclear genes (*AHA1*, *BRE5*, *CAF20*, *CAT8*, *CYC3*, *GAL4*, *MET6*, *EGT2*, *SAE2*, and *VTA1*). The scale represents the number of nucleotide substitutions per site. In blue, green, and black all strains corresponding to clade A B and C of the phylogenetic tree (NJ Figure 3) were represented respectively. The names of each culture collection were abbreviated with their first two letters.

**Supplementary Figure 2.** Structure population order by Q value. Both panels below the labels strains indicate the origin and source of isolation of the strains. Origin: South America in white, Europe in black and North America in plaid. Source: Wild environments in green, apple chicha in pink, cider in blue, wine in purple, beer in brown and unknown in grey.

**Supplementary Figure 3.** Neighbor-net phylogenetic network of the *S. uvarum* of different nuclear genes. **S3.A-J** corresponds to *AHA1*, *BRE5*, *CAF20*, *CAT8*, *CYC3*, *GAL4*, *MET6*, *EGT2*, *SAE2*, and *VTA1* nuclear genes respectively.

**Supplementary Figure 4.** Summary of the allele contribution of each gene as observed in Supplementary Figures 3.A-J. In blue and green, the strains that have H/SA-A and SA-B alleles for each gene are represented respectively. In orange, those alleles that are located in an intermediate position of the Neighbor-net are represented and the origin of the allele (indeterminate) is not clear. Both panels above the strains labels indicate the origin and source of the strains. Origin: South America in white, Europe in black and North America in plaid. Source: Wild environments in green, apple chicha in pink, cider in blue, wine in purple, beer in brown and unknown origin in grey. The bar plot represents the amount of H/SA-A or SA-B alleles that each strain contains.

**Supplementary Figure 5.** Neighbor-Joining (NJ) phylogeny of 63 strains based on concatenated *COX2* gene. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura Nei model (Tamura and Nei, 1993) method and expressed as the number of base substitutions per site. **S5.A:** Phylogenetic tree including the introgressed region (512pb). **S5.B:** Phylogenetic tree without introgressed region (472pb).

Figure1

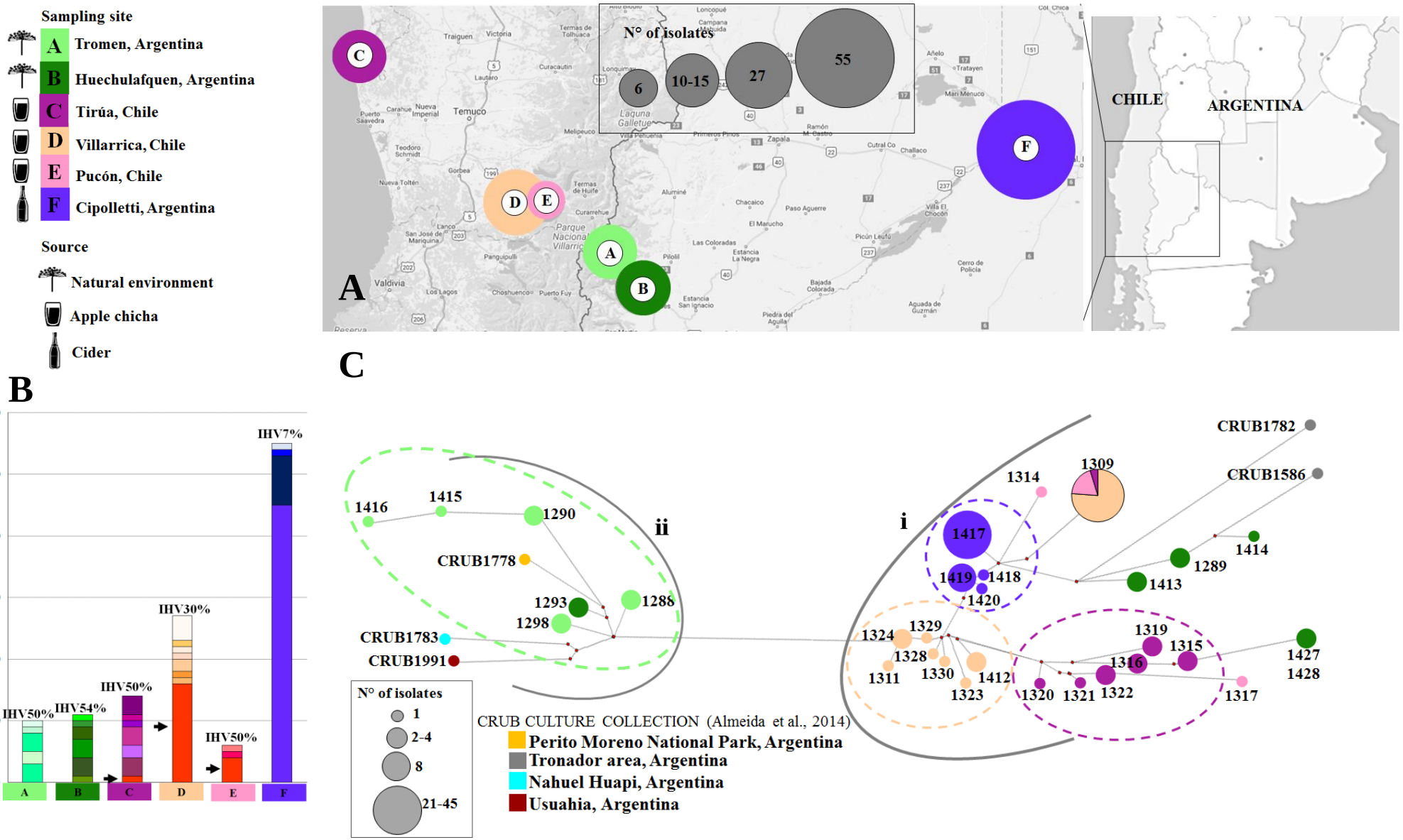

Sampling site
A Tromen, Argentina
B Huechulafquen, Argentina
C Tirúa, Chile
D Villarrica, Chile
E Pucón, Chile
F Cipolletti, Argentina
Source
Natural environment
Apple chicha
Cider
A
N° of isolates
6
10-15
27
55
CHILE
ARGENTINA
B
N° of isolates
IHV50%
IHV54%
IHV50%
IHV30%
IHV50%
IHV7%
A
B
C
D
E
F
C
1416
1415
1290
CRUB1778
1293
1298
1288
CRUB1783
CRUB1991
ii
i
1314
1309
1417
1419
1418
1420
CRUB1782
CRUB1586
1414
1289
1413
1324
1329
1328
1311
1330
1412
1323
1319
1315
1316
1320
1321
1322
1317
1427
1428
N° of isolates
1
2-4
8
21-45
CRUB CULTURE COLLECTION (Almeida et al., 2014)
Perito Moreno National Park, Argentina
Tronador area, Argentina
Nahuel Huapi, Argentina
Usuahia, Argentina

Figure2

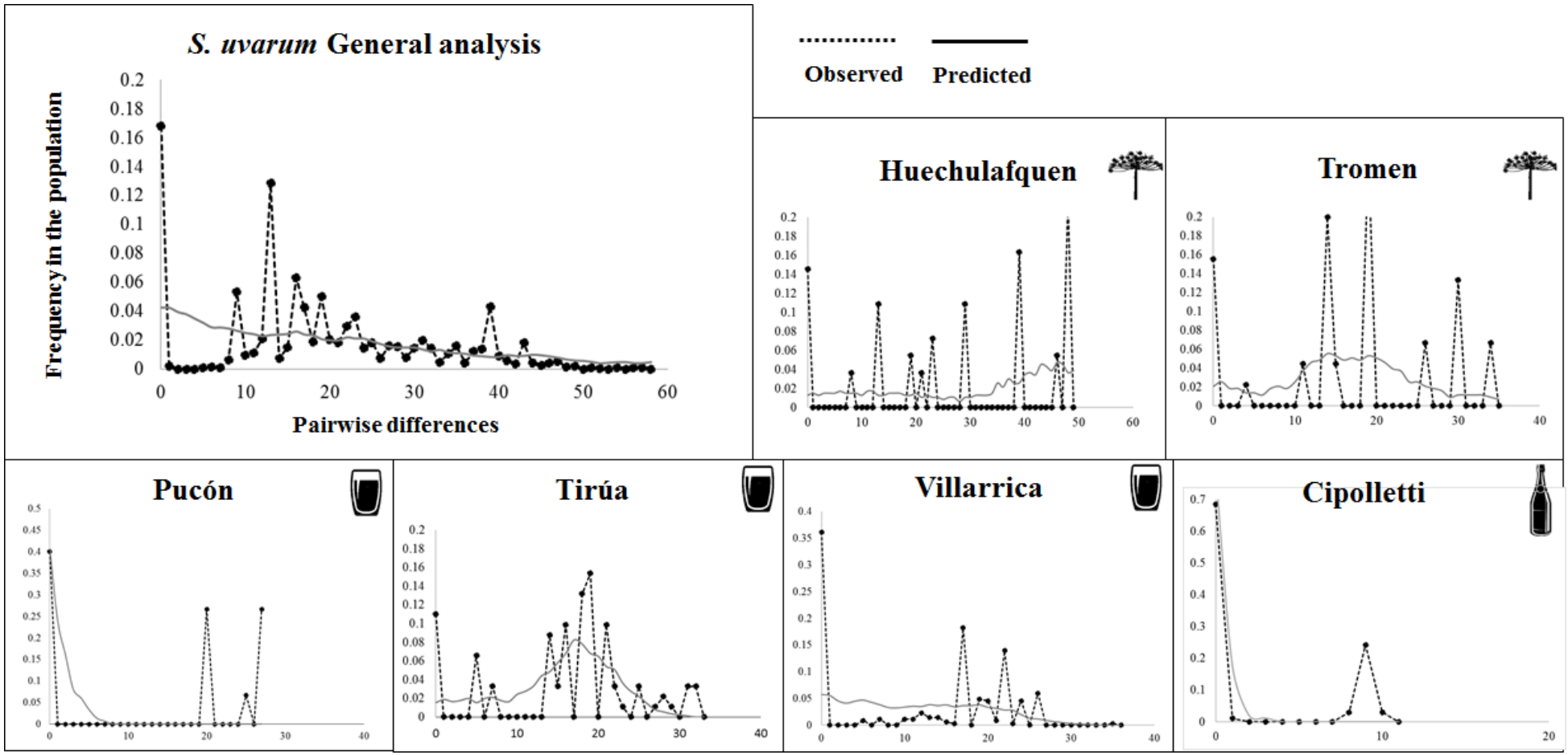

S. uvarum General analysis
Frequency in the population
Pairwise differences
Observed
Predicted
Huechulafquen
Tromen
Pucón
Tirúa
Villarrica
Cipolletti

Figure3

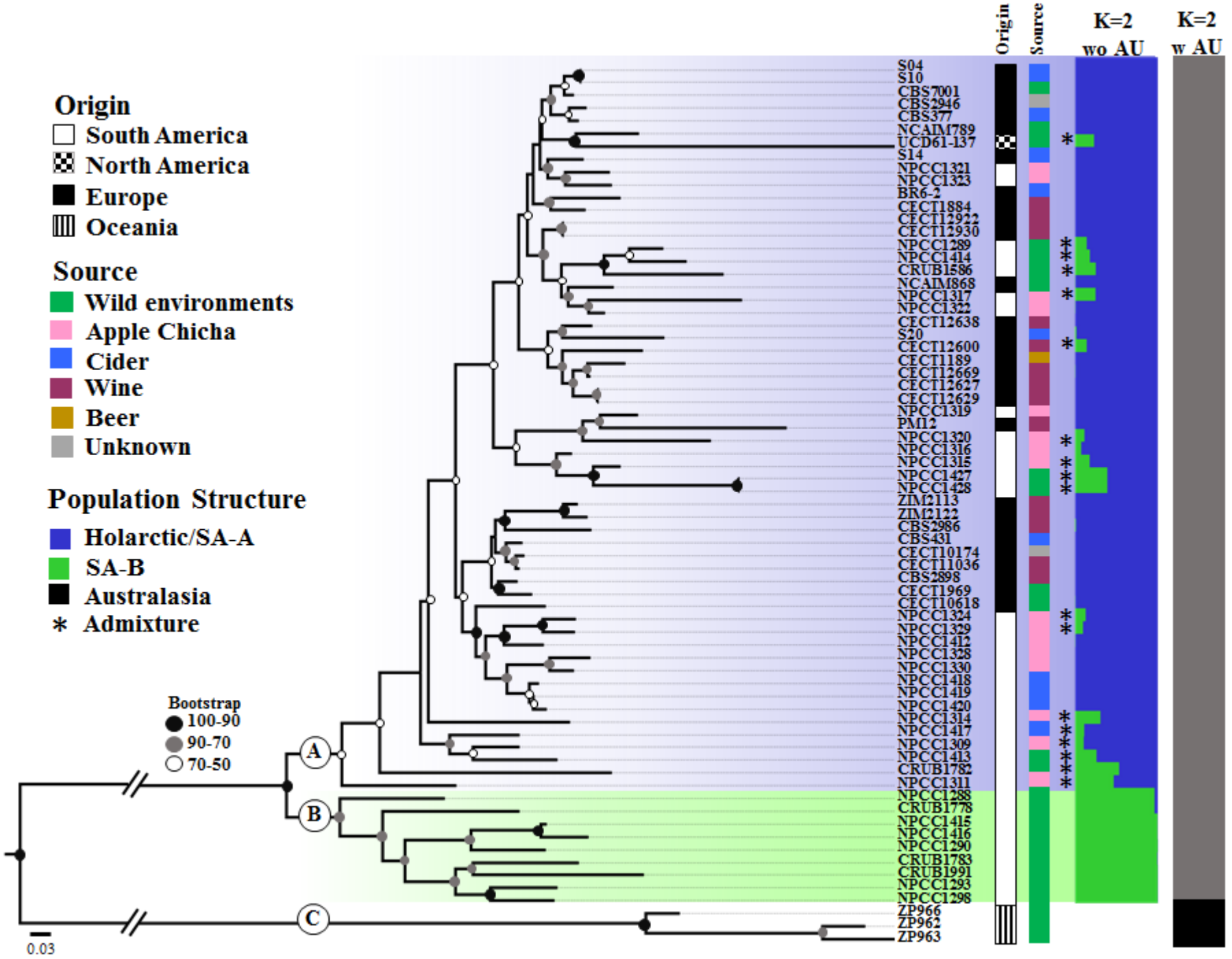

Origin
South America
North America
Europe
Oceania
Source
Wild environments
Apple Chicha
Cider
Wine
Beer
Unknown
Population Structure
Holarctic/SA-A
SA-B
Australasia
* Admixture
Bootstrap
100-90
90-70
70-50
A
B
C
0.03
Origin
Source
K=2 wo AU
K=2 w AU
S04
S10
CBS7001
CBS2946
CBS377
NCAIM789
UCD61-137
S14
NPCC1321
NPCC1323
BR6-2
CECT1884
CECT12922
CECT12930
NPCC1289
NPCC1414
CRUB1586
NCAIM868
NPCC1317
NPCC1322
CECT12638
S20
CECT12600
CECT1189
CECT12669
CECT12627
CECT12629
NPCC1319
PM12
NPCC1320
NPCC1316
NPCC1315
NPCC1427
NPCC1428
ZIM2113
ZIM2122
CBS2986
CBS431
CECT10174
CECT11036
CBS2898
CECT1969
CECT10618
NPCC1324
NPCC1329
NPCC1412
NPCC1328
NPCC1330
NPCC1418
NPCC1419
NPCC1420
NPCC1314
NPCC1417
NPCC1309
NPCC1413
CRUB1782
NPCC1311
NPCC1288
CRUB1778
NPCC1415
NPCC1416
NPCC1290
CRUB1783
CRUB1991
NPCC1293
NPCC1298
ZP966
ZP962
ZP963

Figure4

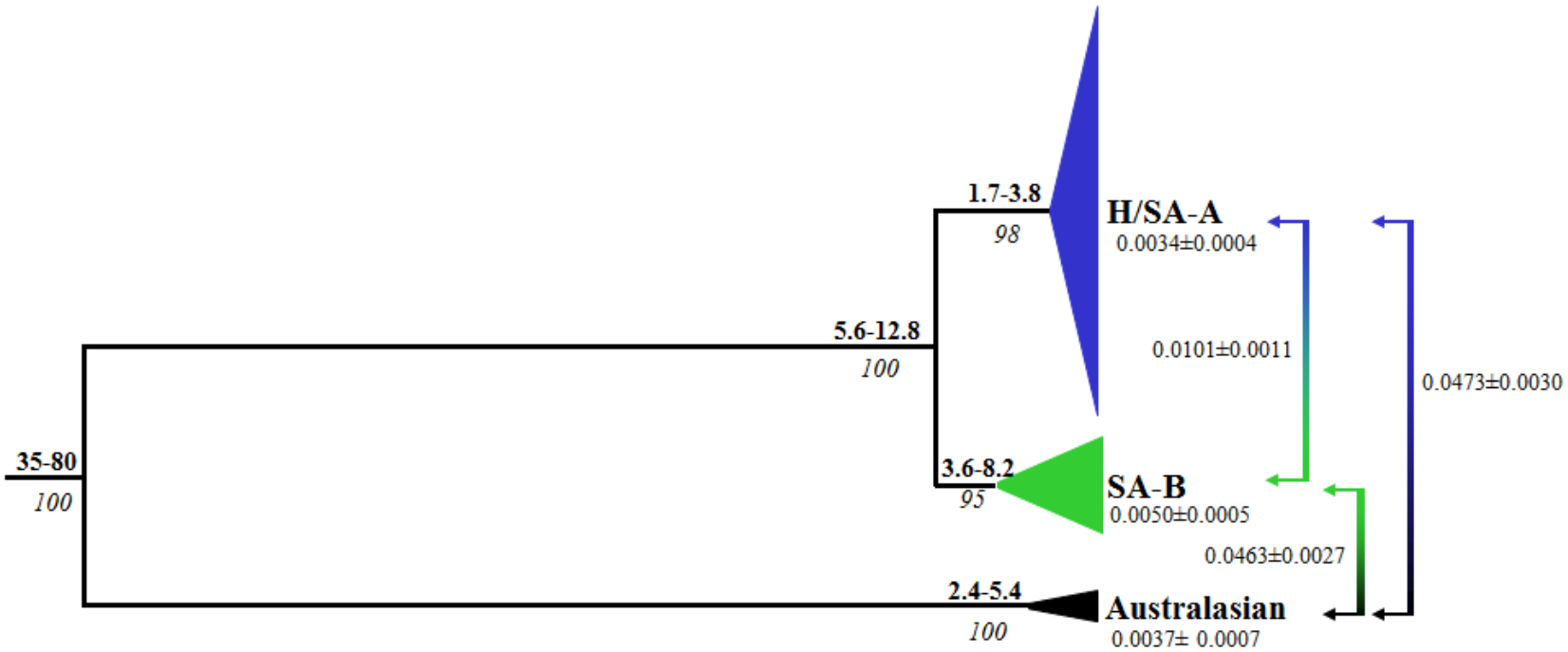
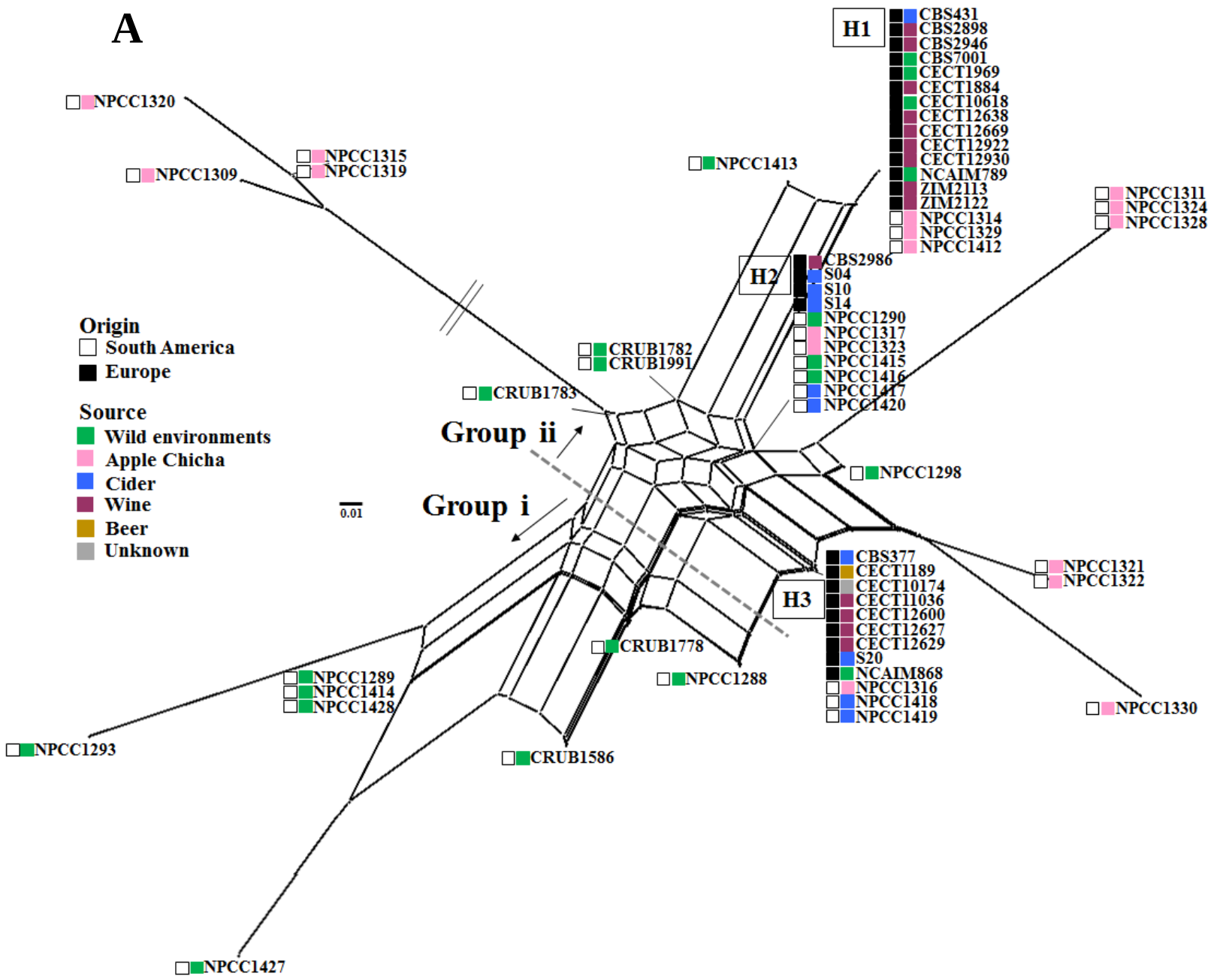

1.7-3.8
98
H/SA-A
0.0034±0.0004
5.6-12.8
100
0.0101±0.0011
0.0473±0.0030
35-80
100
3.6-8.2
95
SA-B
0.0050±0.0005
0.0463±0.0027
2.4-5.4
100
Australasian
0.0037± 0.0007

Figure5


A
H1
CBS431
CBS2898
CBS2946
CBS7001
CECT1969
CECT1884
CECT10618
CECT12638
CECT12669
CECT12922
CECT12930
NCAIM789
ZIM2113
ZIM2122
NPCC1314
NPCC1329
NPCC1412
H2
CBS2986
S04
S10
S14
NPCC1290
NPCC1317
NPCC1323
NPCC1415
NPCC1416
NPCC1417
NPCC1420
H3
CBS377
CECT1189
CECT10174
CECT11036
CECT12600
CECT12627
CECT12629
S20
NCAIM868
NPCC1316
NPCC1418
NPCC1419
NPCC1320
NPCC1315
NPCC1319
NPCC1309
NPCC1413
NPCC1311
NPCC1324
NPCC1328
CRUB1782
CRUB1991
CRUB1783
NPCC1298
NPCC1321
NPCC1322
NPCC1330
CRUB1778
NPCC1288
NPCC1289
NPCC1414
NPCC1428
NPCC1293
CRUB1586
NPCC1427
Origin
South America
Europe
Source
Wild environments
Apple Chicha
Cider
Wine
Beer
Unknown
0.01
Group ii
Group i

**B**

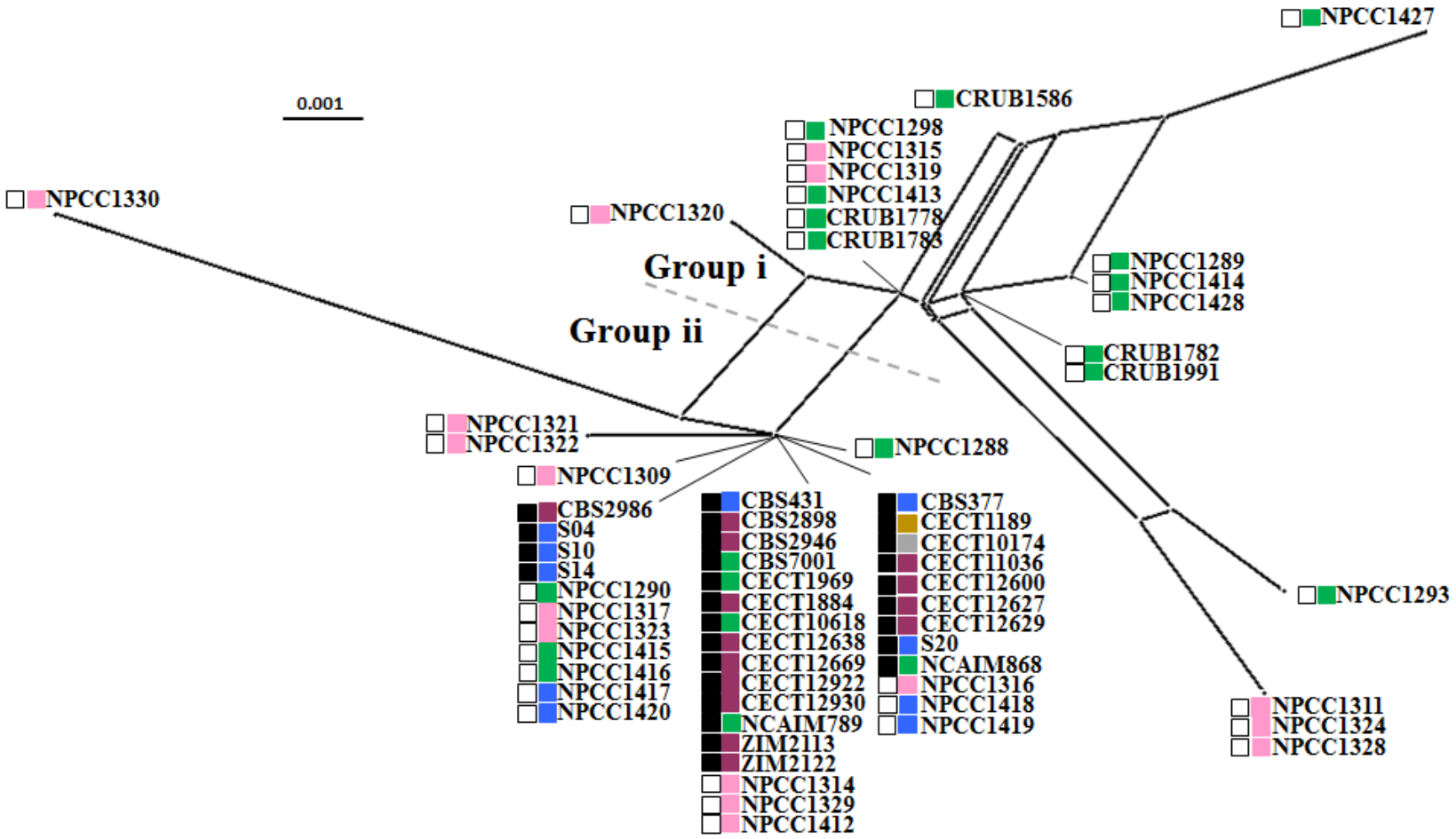
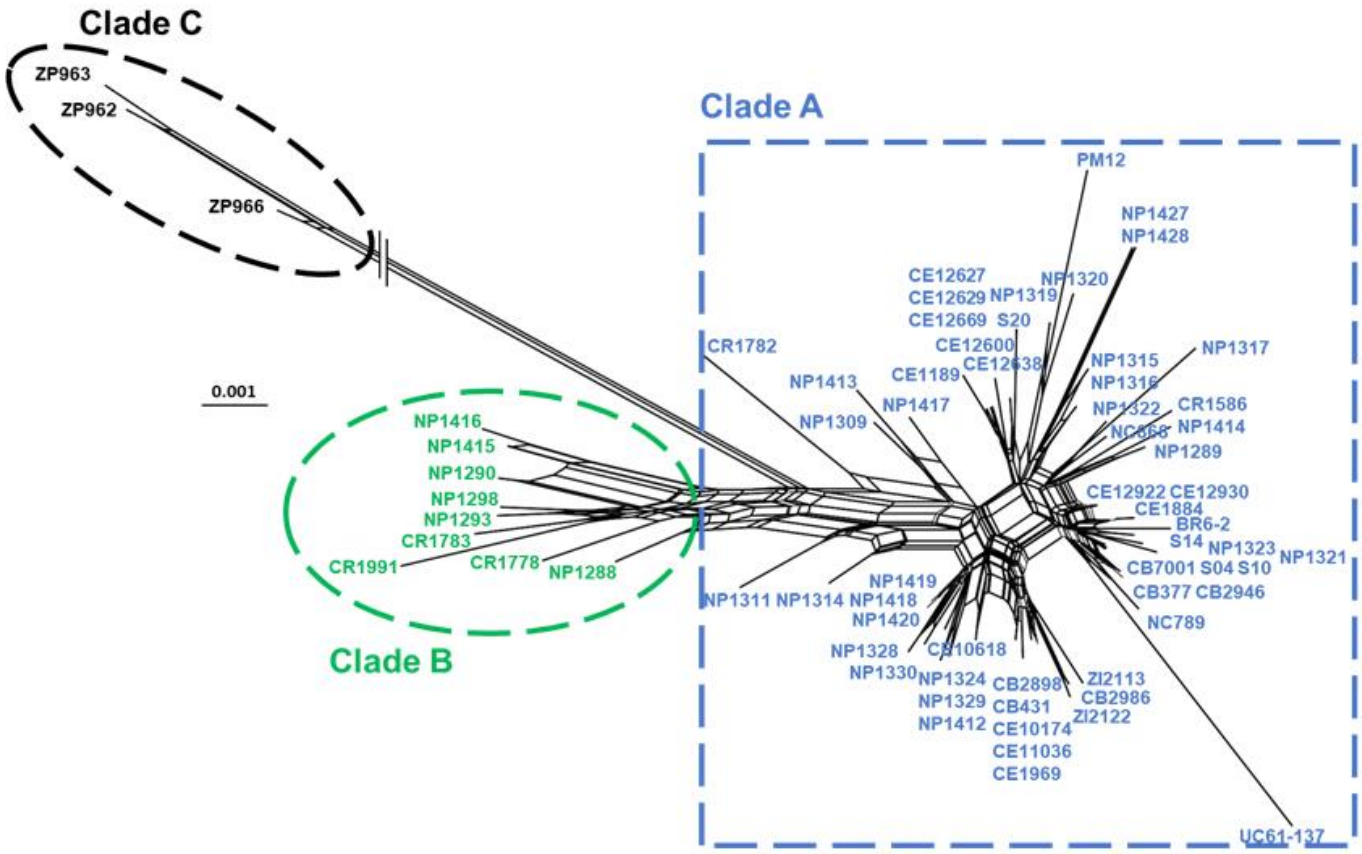

0.001
NPCC1427
CRUB1586
NPCC1298
NPCC1315
NPCC1319
NPCC1413
CRUB1778
CRUB1783
NPCC1330
NPCC1320
Group i
Group ii
NPCC1289
NPCC1414
NPCC1428
CRUB1782
CRUB1991
NPCC1321
NPCC1322
NPCC1288
NPCC1309
CBS2986
S04
S10
S14
NPCC1290
NPCC1317
NPCC1323
NPCC1415
NPCC1416
NPCC1417
NPCC1420
CBS431
CBS2898
CBS2946
CBS7001
CECT1969
CECT1884
CECT10618
CECT12638
CECT12669
CECT12922
CECT12930
NCAIM789
ZIM2113
ZIM2122
NPCC1314
NPCC1329
NPCC1412
CBS377
CECT1189
CECT10174
CECT11036
CECT12600
CECT12627
CECT12629
S20
NCAIM868
NPCC1316
NPCC1418
NPCC1419
NPCC1293
NPCC1311
NPCC1324
NPCC1328

**Figure S1**


Clade C
ZP963
ZP962
ZP966
0.001
Clade B
NP1416
NP1415
NP1290
NP1298
NP1293
CR1783
CR1991
CR1778
NP1288
Clade A
PM12
NP1427
NP1428
CE12627
CE12629
CE12669
NP1320
NP1319
S20
CR1782
CE12600
CE12638
NP1315
NP1317
NP1413
CE1189
NP1316
NP1417
NP1309
NP1322
CR1586
NP1414
NP1289
CE12922
CE12930
CE1884
BR6-2
S14
NP1323
NP1321
CB7001
S04
S10
CB377
CB2946
NC789
NP1419
NP1311
NP1314
NP1418
NP1420
NP1328
CB10618
NP1330
NP1324
CB2898
ZI2113
CB2986
NP1329
CB431
ZI2122
NP1412
CE10174
CE11036
CE1969
UC61-137

## Figure S2

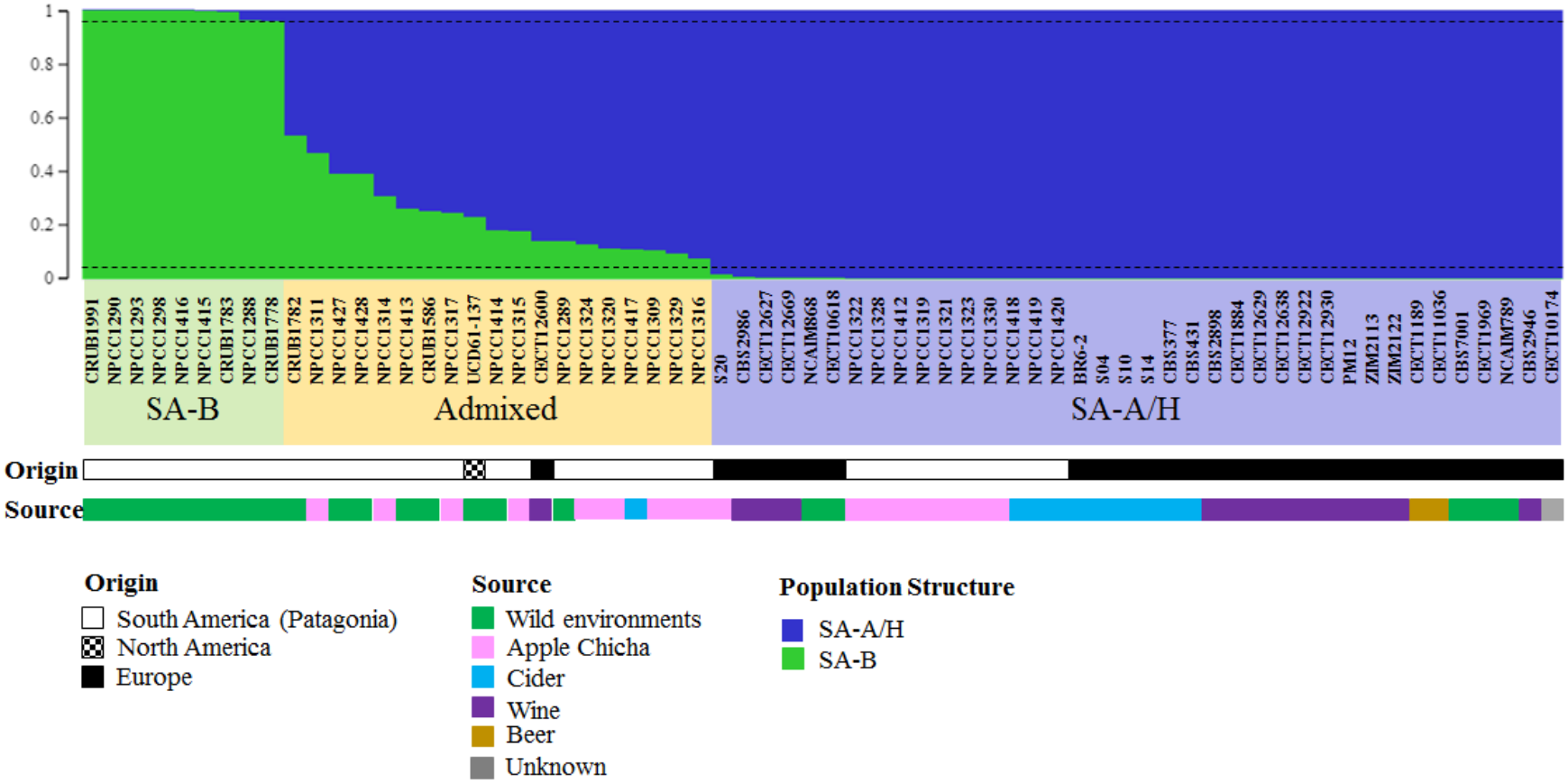

1
0.8
0.6
0.4
0.2
0
CRUB1991 NPCC1290 NPCC1293 NPCC1298 NPCC1416 NPCC1415 CRUB1783 NPCC1288 CRUB1778
CRUB1782 NPCC1311 NPCC1427 NPCC1428 NPCC1314 NPCC1413 CRUB1586 NPCC1317 UCD61-137 NPCC1414 NPCC1315 CECT12600 NPCC1289 NPCC1324 NPCC1320 NPCC1417 NPCC1309 NPCC1329 NPCC1316
S20 CBS2986 CECT12627 CECT12669 NCAIM868 CECT10618 NPCC1322 NPCC1328 NPCC1412 NPCC1319 NPCC1321 NPCC1323 NPCC1330 NPCC1418 NPCC1419 NPCC1420 BR6-2 S04 S10 S14 CBS377 CBS431 CBS2898 CECT1884 CECT12629 CECT12638 CECT12922 CECT12930 PM12 ZIM2113 ZIM2122 CECT1189 CECT11036 CBS7001 CECT1969 NCAIM789 CBS2946 CECT10174
SA-B
Admixed
SA-A/H
Origin
Source
Origin
South America (Patagonia)
North America
Europe
Source
Wild environments
Apple Chicha
Cider
Wine
Beer
Unknown
Population Structure
SA-A/H
SA-B

Figure S3.A

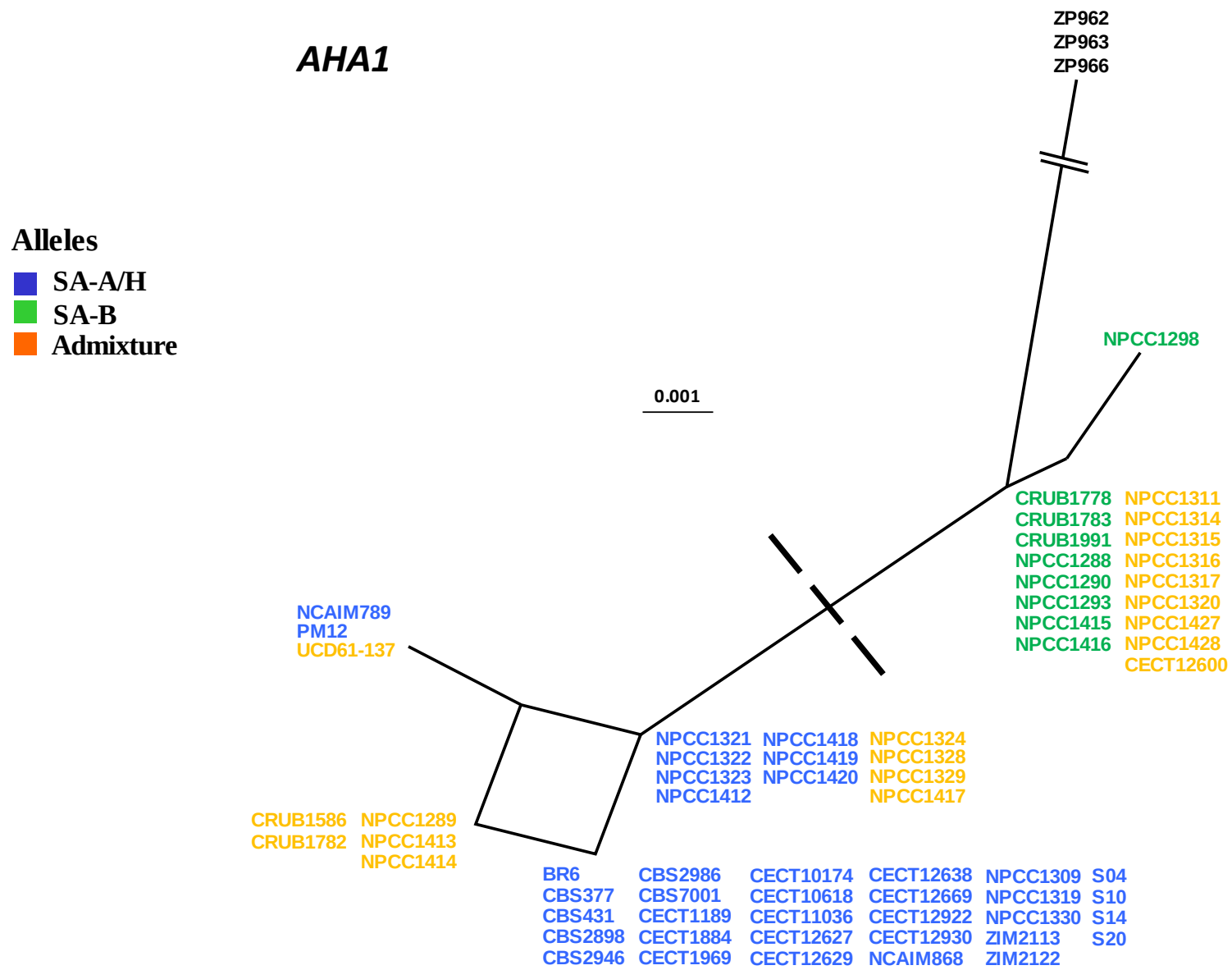

AHA1
Alleles
SA-A/H
SA-B
Admixture
ZP962
ZP963
ZP966
NPCC1298
0.001
CRUB1778 NPCC1311
CRUB1783 NPCC1314
CRUB1991 NPCC1315
NPCC1288 NPCC1316
NPCC1290 NPCC1317
NPCC1293 NPCC1320
NPCC1415 NPCC1427
NPCC1416 NPCC1428
CECT12600
NCAIM789
PM12
UCD61-137
NPCC1321 NPCC1418 NPCC1324
NPCC1322 NPCC1419 NPCC1328
NPCC1323 NPCC1420 NPCC1329
NPCC1412 NPCC1417
CRUB1586 NPCC1289
CRUB1782 NPCC1413
NPCC1414
BR6 CBS2986 CECT10174 CECT12638 NPCC1309 S04
CBS377 CBS7001 CECT10618 CECT12669 NPCC1319 S10
CBS431 CECT1189 CECT11036 CECT12922 NPCC1330 S14
CBS2898 CECT1884 CECT12627 CECT12930 ZIM2113 S20
CBS2946 CECT1969 CECT12629 NCAIM868 ZIM2122

Figure S3.B

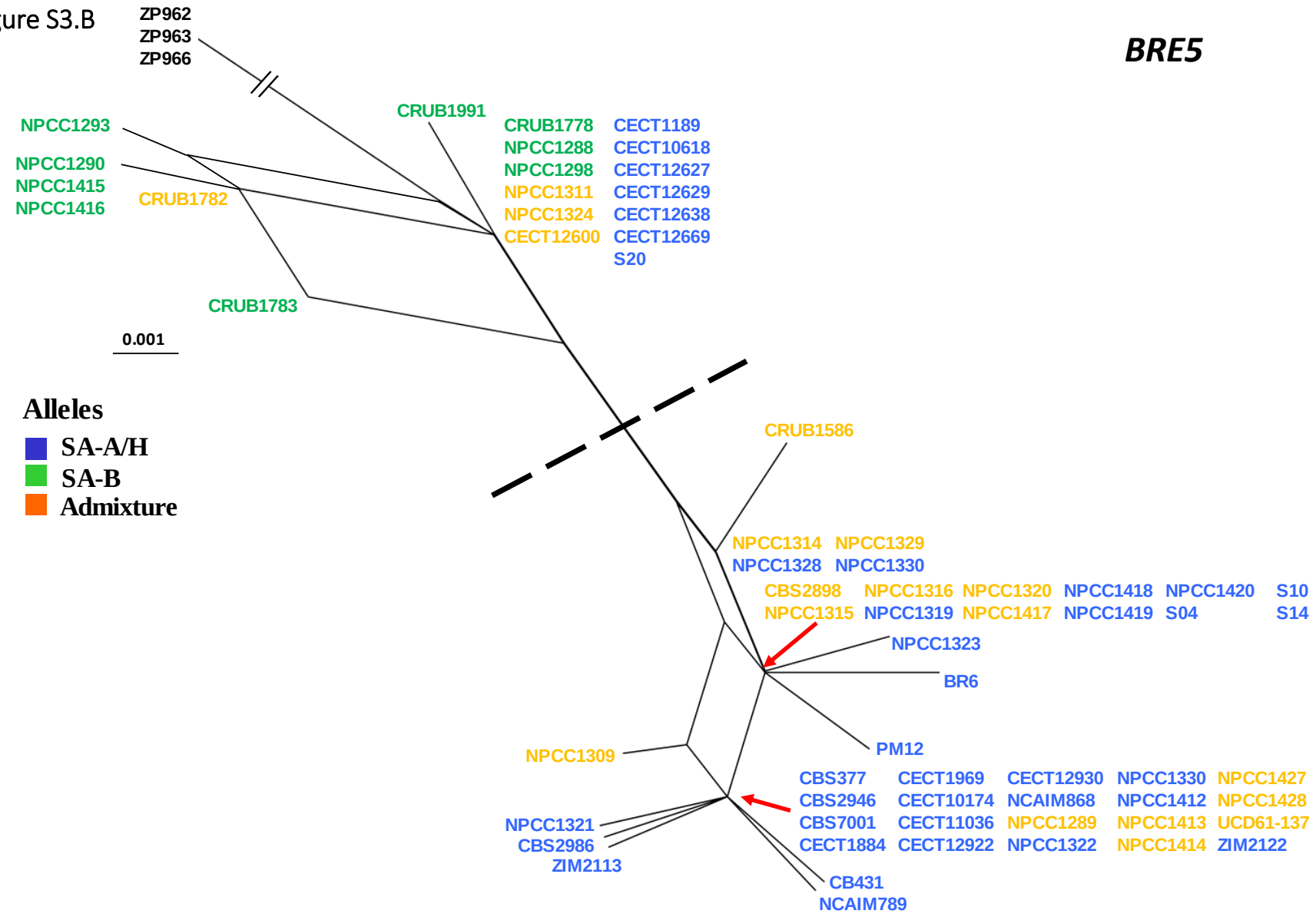

BRE5
ZP962
ZP963
ZP966
CRUB1991
NPCC1293
NPCC1290
NPCC1415
NPCC1416
CRUB1782
CRUB1783
CRUB1778 CECT1189
NPCC1288 CECT10618
NPCC1298 CECT12627
NPCC1311 CECT12629
NPCC1324 CECT12638
CECT12600 CECT12669
S20
0.001
Alleles
SA-A/H
SA-B
Admixture
CRUB1586
NPCC1314 NPCC1329
NPCC1328 NPCC1330
CBS2898 NPCC1316 NPCC1320 NPCC1418 NPCC1420 S10
NPCC1315 NPCC1319 NPCC1417 NPCC1419 S04 S14
NPCC1323
BR6
PM12
NPCC1309
CBS377 CECT1969 CECT12930 NPCC1330 NPCC1427
CBS2946 CECT10174 NCAIM868 NPCC1412 NPCC1428
CBS7001 CECT11036 NPCC1289 NPCC1413 UCD61-137
CECT1884 CECT12922 NPCC1322 NPCC1414 ZIM2122
NPCC1321
CBS2986
ZIM2113
CB431
NCAIM789

Figure S3.C

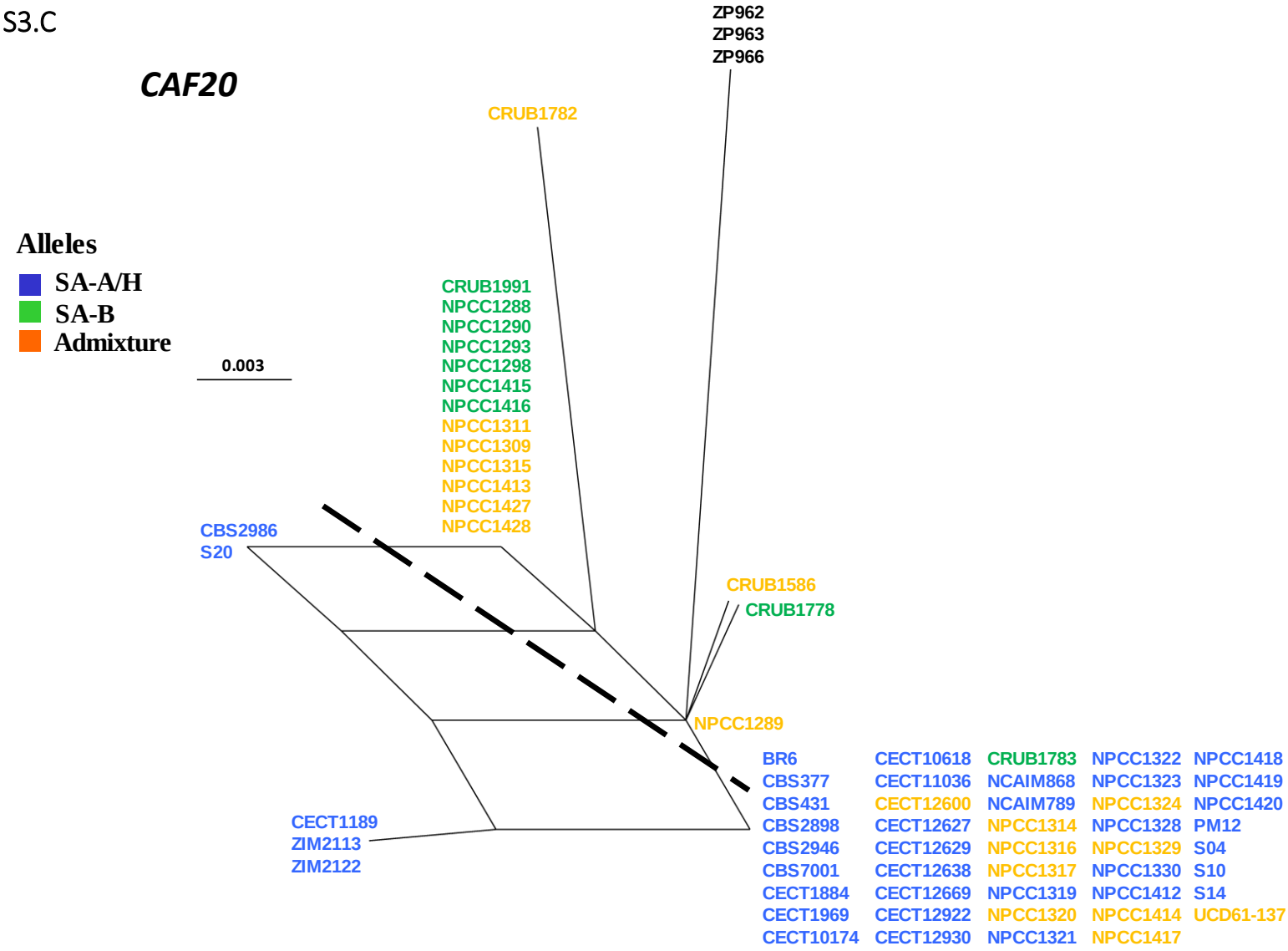

CAF20
Alleles
SA-A/H
SA-B
Admixture
0.003
ZP962
ZP963
ZP966
CRUB1782
CRUB1991
NPCC1288
NPCC1290
NPCC1293
NPCC1298
NPCC1415
NPCC1416
NPCC1311
NPCC1309
NPCC1315
NPCC1413
NPCC1427
NPCC1428
CBS2986
S20
CRUB1586
CRUB1778
NPCC1289
CECT1189
ZIM2113
ZIM2122
BR6
CBS377
CBS431
CBS2898
CBS2946
CBS7001
CECT1884
CECT1969
CECT10174
CECT10618
CECT11036
CECT12600
CECT12627
CECT12629
CECT12638
CECT12669
CECT12922
CECT12930
CRUB1783
NCAIM868
NCAIM789
NPCC1314
NPCC1316
NPCC1317
NPCC1319
NPCC1320
NPCC1321
NPCC1322
NPCC1323
NPCC1324
NPCC1328
NPCC1329
NPCC1330
NPCC1412
NPCC1414
NPCC1417
NPCC1418
NPCC1419
NPCC1420
PM12
S04
S10
S14
UCD61-137

Figure S3.D

***CAT8***

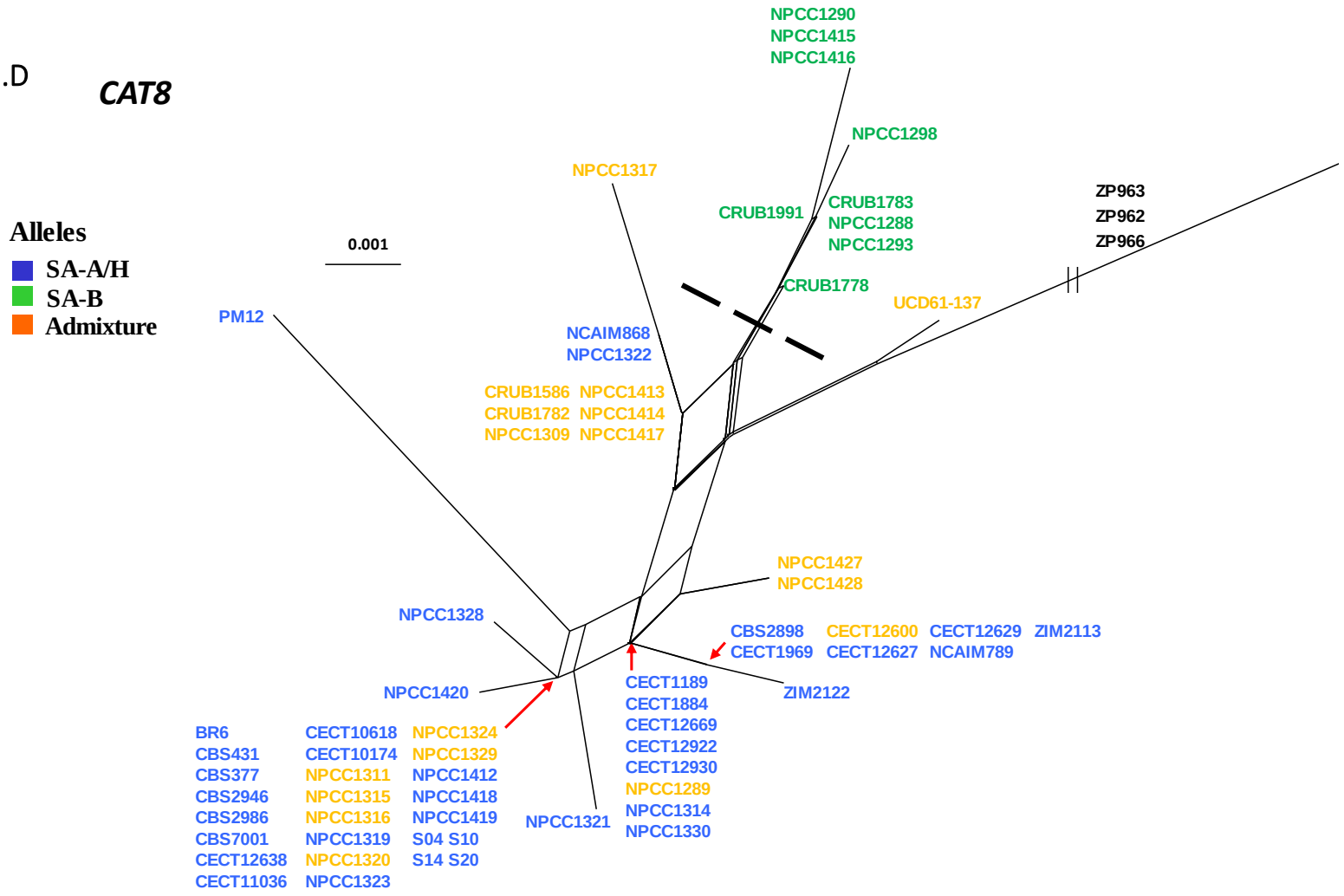


Figure S3.E ***CYC3***

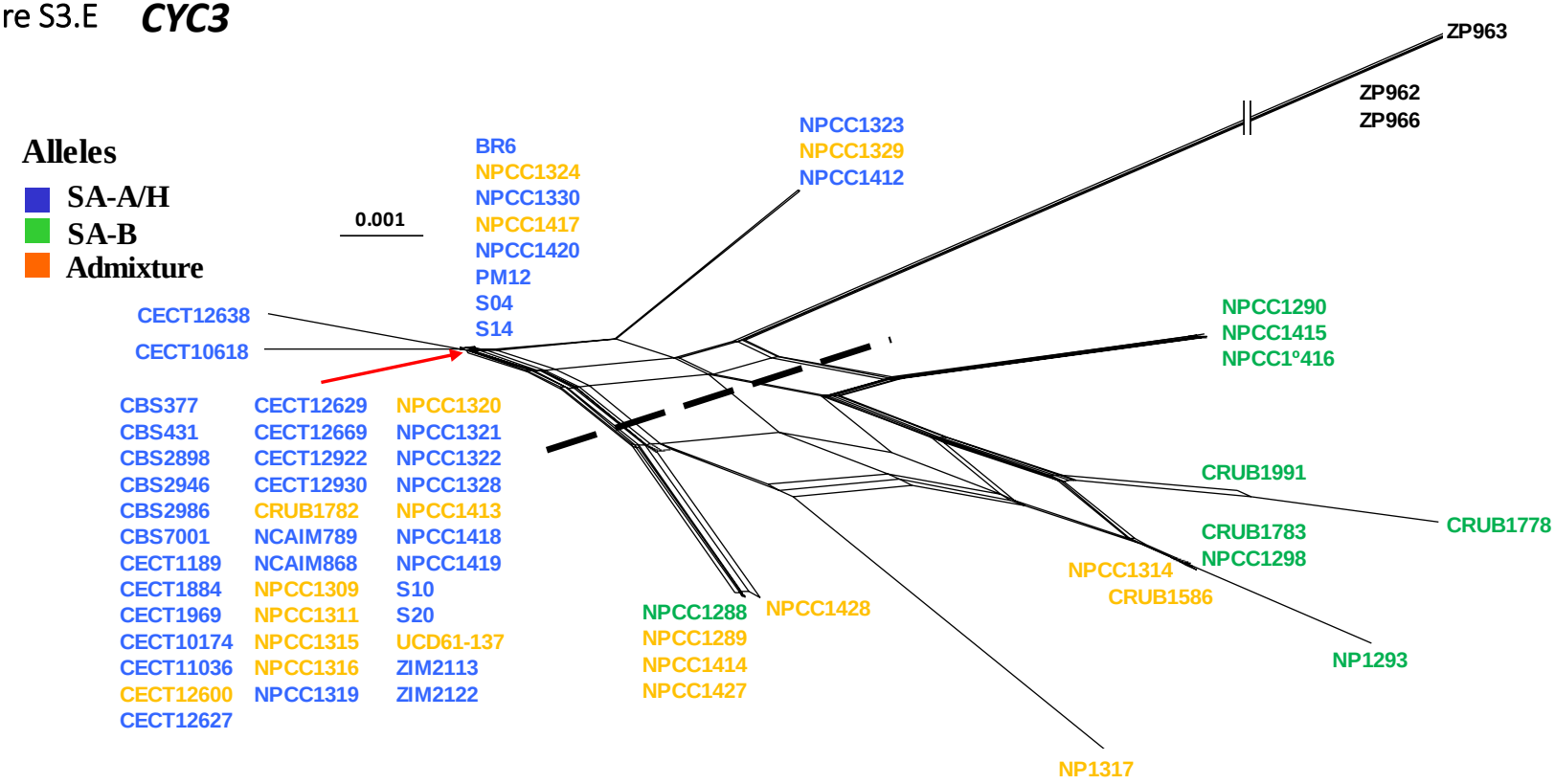


Figure S3.F

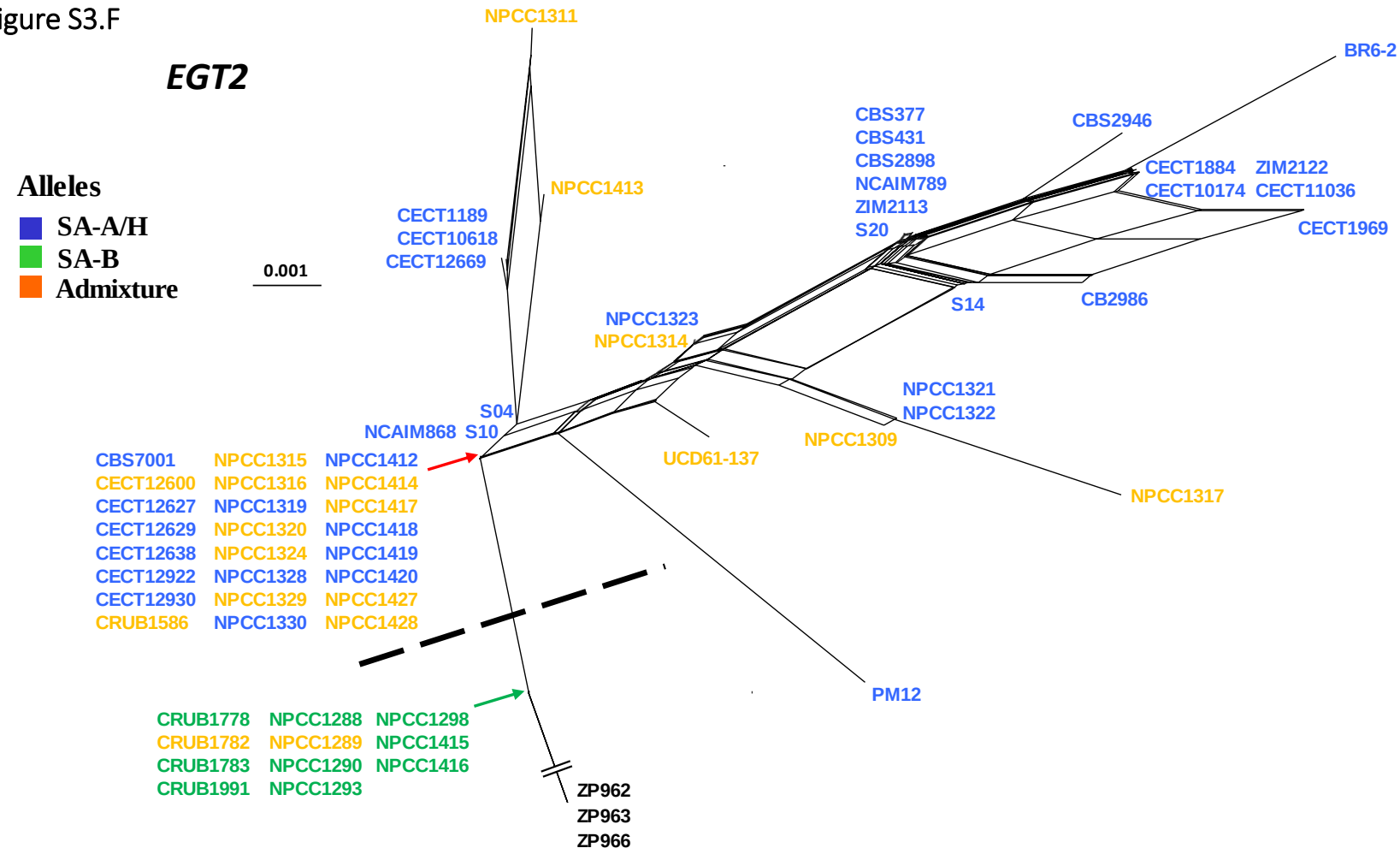

EGT2
Alleles
SA-A/H
SA-B
Admixture
0.001
NPCC1311
BR6-2
CBS377
CBS431
CBS2898
NCAIM789
ZIM2113
S20
CBS2946
CECT1884
ZIM2122
CECT10174
CECT11036
CECT1969
NPCC1413
CECT1189
CECT10618
CECT12669
S14
CB2986
NPCC1323
NPCC1314
NPCC1321
NPCC1322
S04
NCAIM868
S10
NPCC1309
UCD61-137
NPCC1317
CBS7001
CECT12600
CECT12627
CECT12629
CECT12638
CECT12922
CECT12930
CRUB1586
NPCC1315
NPCC1316
NPCC1319
NPCC1320
NPCC1324
NPCC1328
NPCC1329
NPCC1330
NPCC1412
NPCC1414
NPCC1417
NPCC1418
NPCC1419
NPCC1420
NPCC1427
NPCC1428
PM12
CRUB1778
CRUB1782
CRUB1783
CRUB1991
NPCC1288
NPCC1289
NPCC1290
NPCC1293
NPCC1298
NPCC1415
NPCC1416
ZP962
ZP963
ZP966

Figure S3.G

***GAL4***

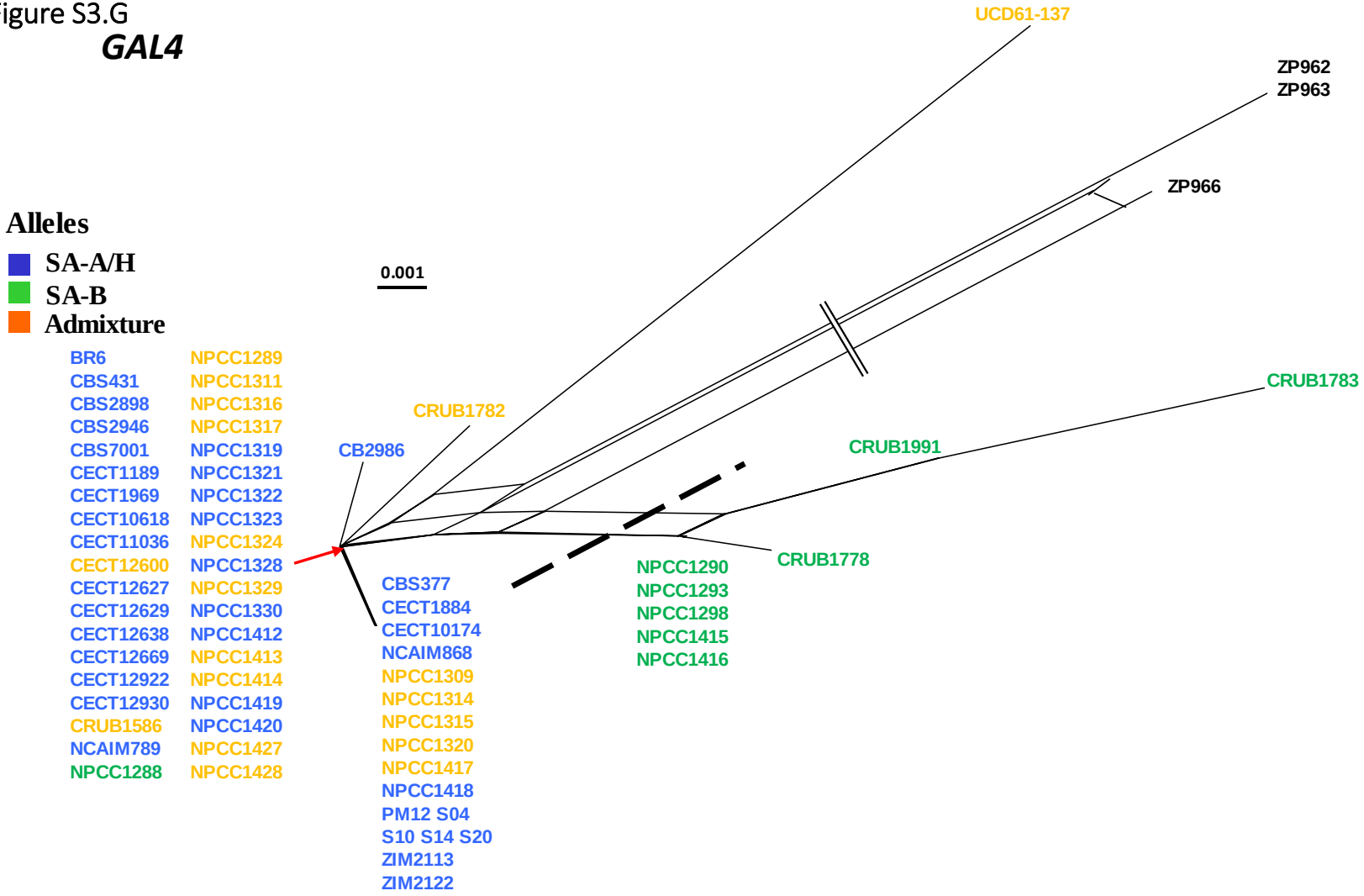
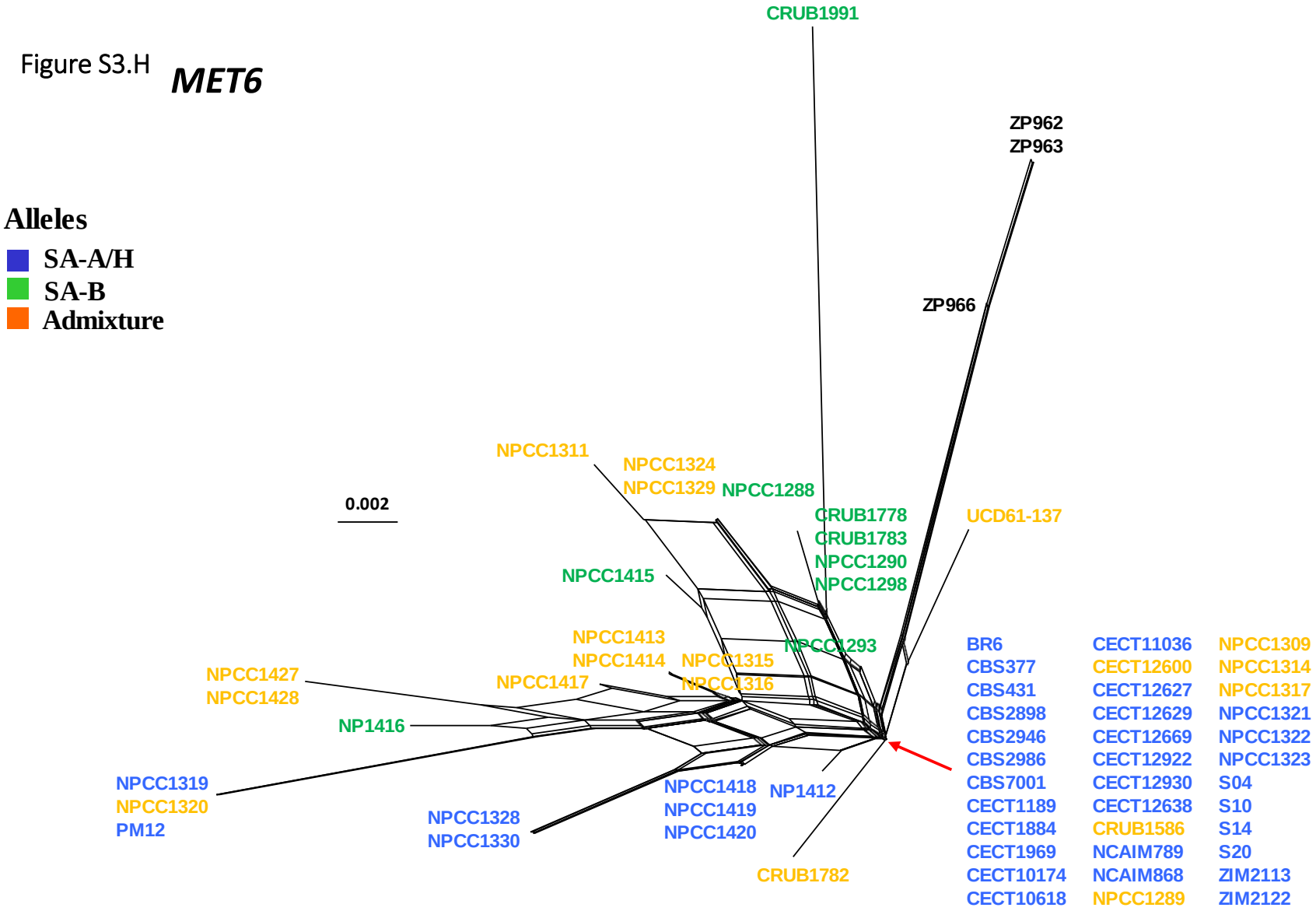


Figure S3.H ***MET6***



## *SAE2*

Figure S3.I

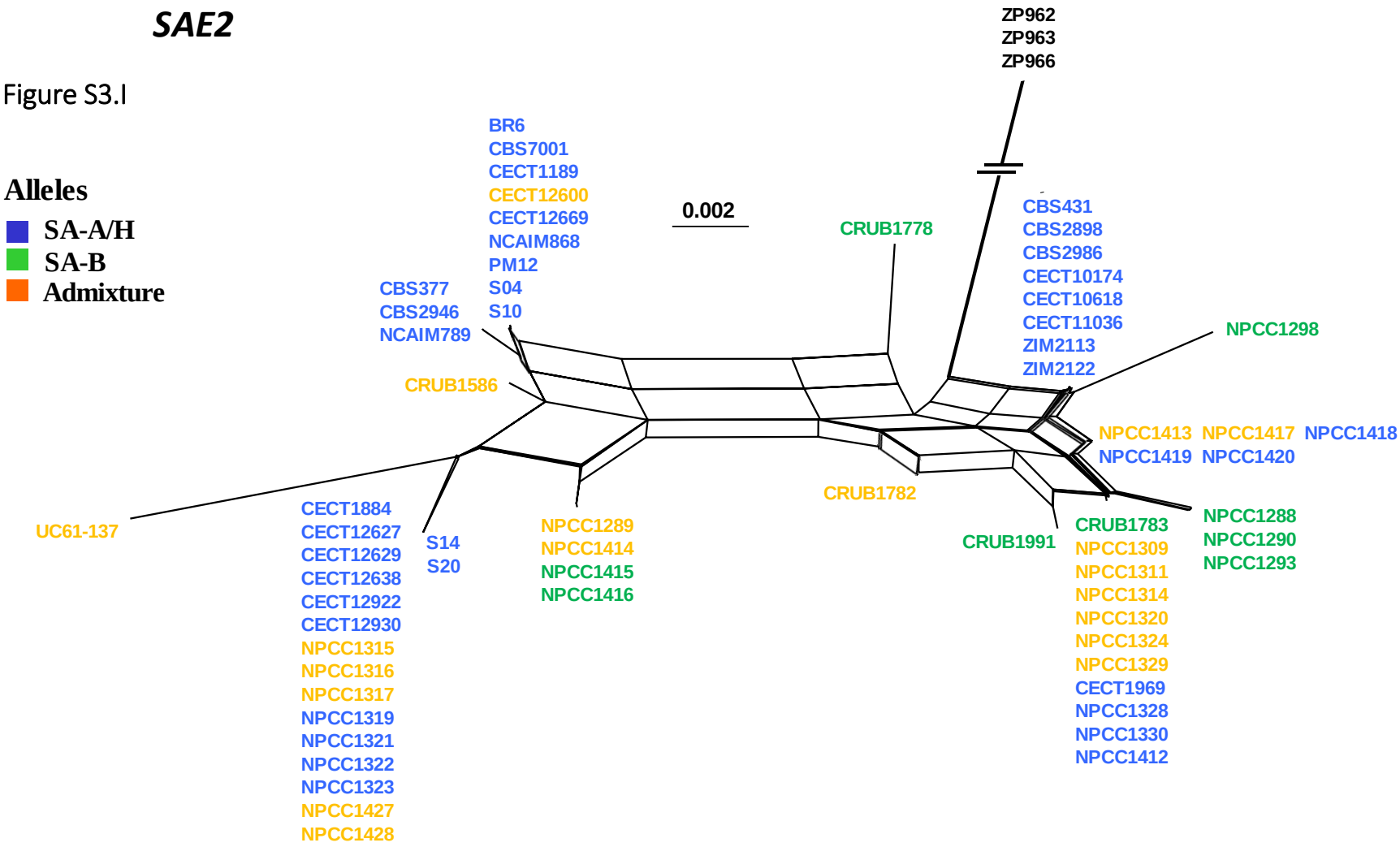


Figure S3.J ***VTA1***

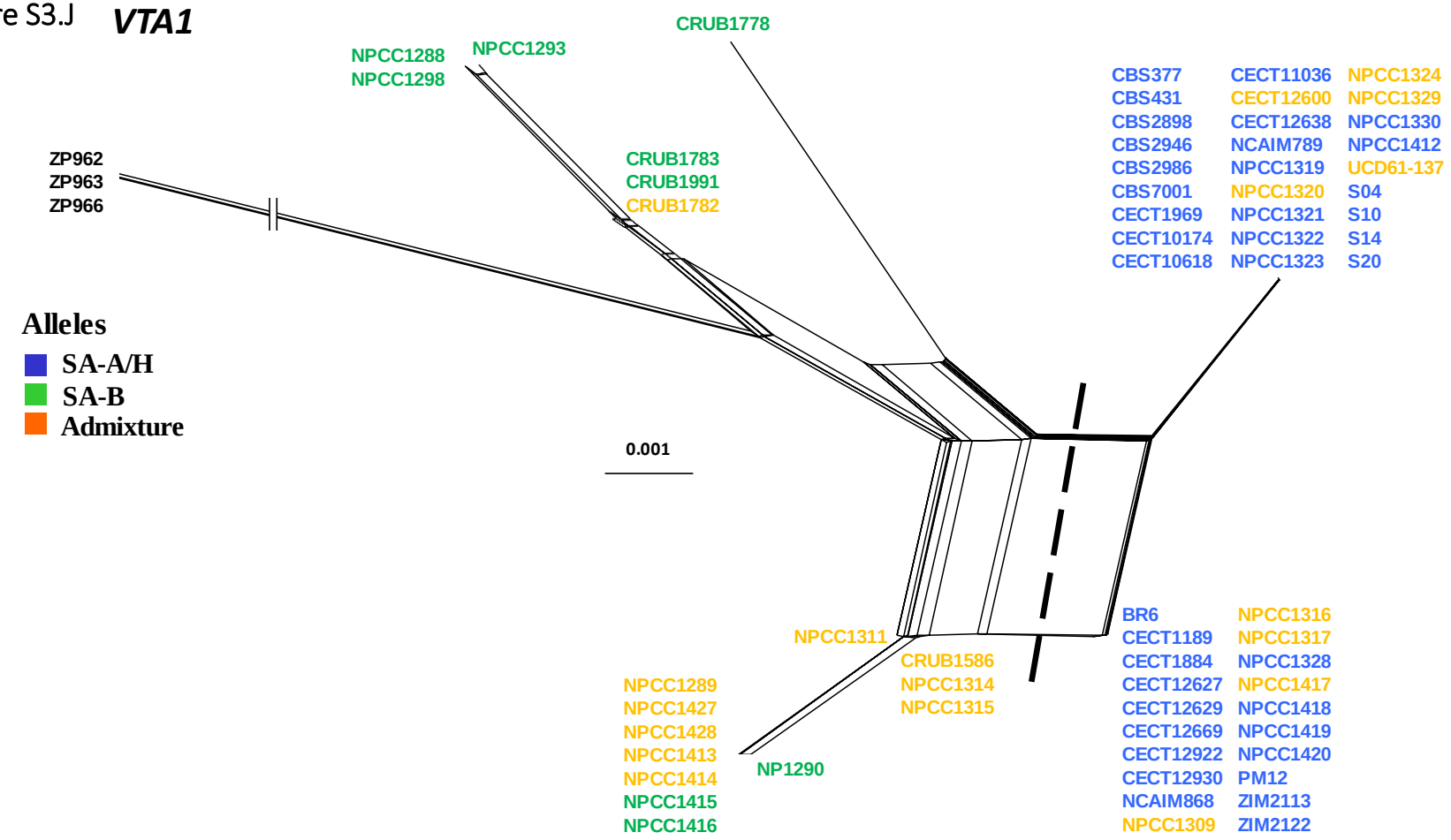


Figure S4

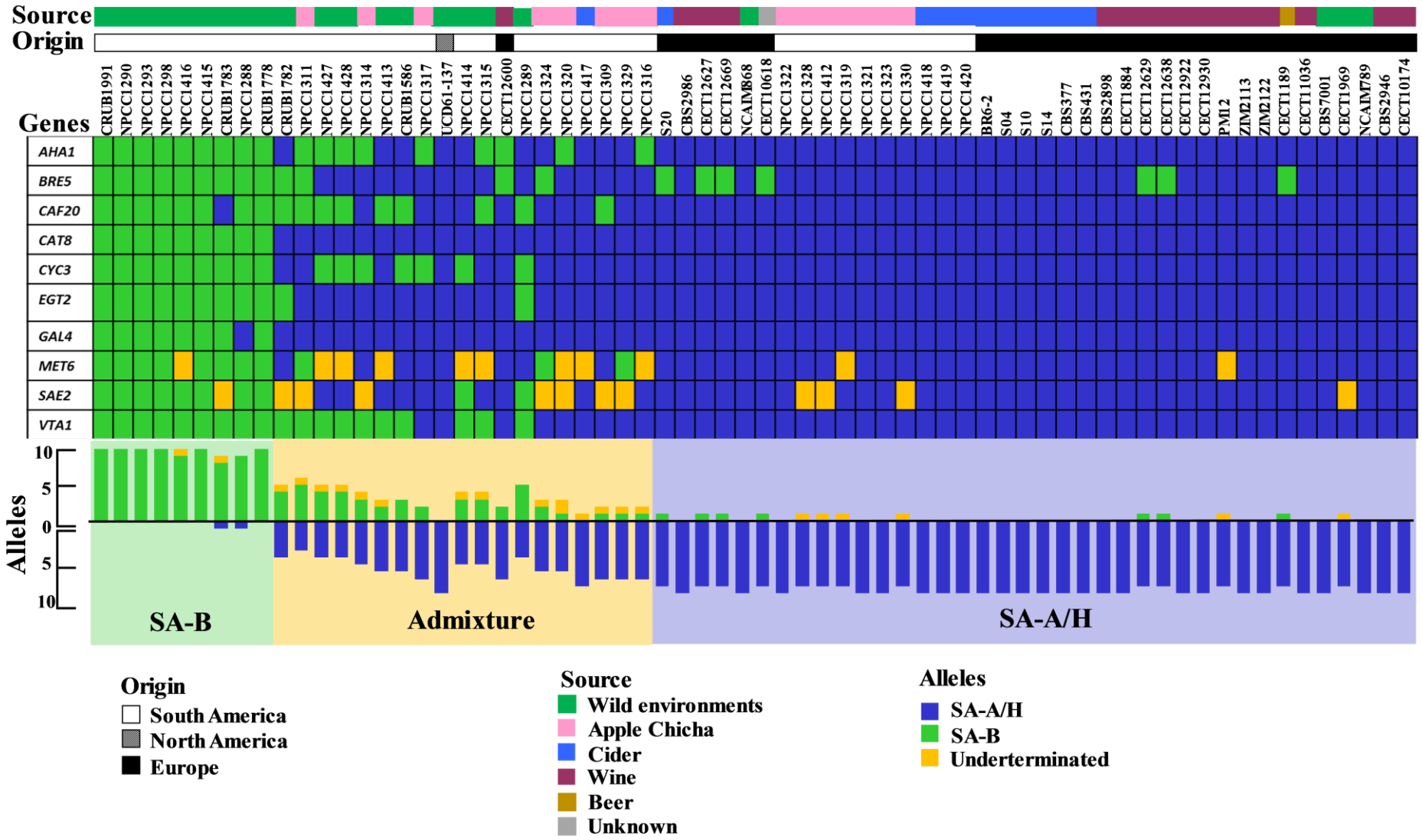

Source
Origin
Genes
AHA1
BRE5
CAF20
CAT8
CYC3
EGT2
GAL4
MET6
SAE2
VTA1
Alleles
10
5
0
5
10
SA-B
Admixture
SA-A/H
Origin
South America
North America
Europe
Source
Wild environments
Apple Chicha
Cider
Wine
Beer
Unknown
Alleles
SA-A/H
SA-B
Underterminated

**Figure S5.A**

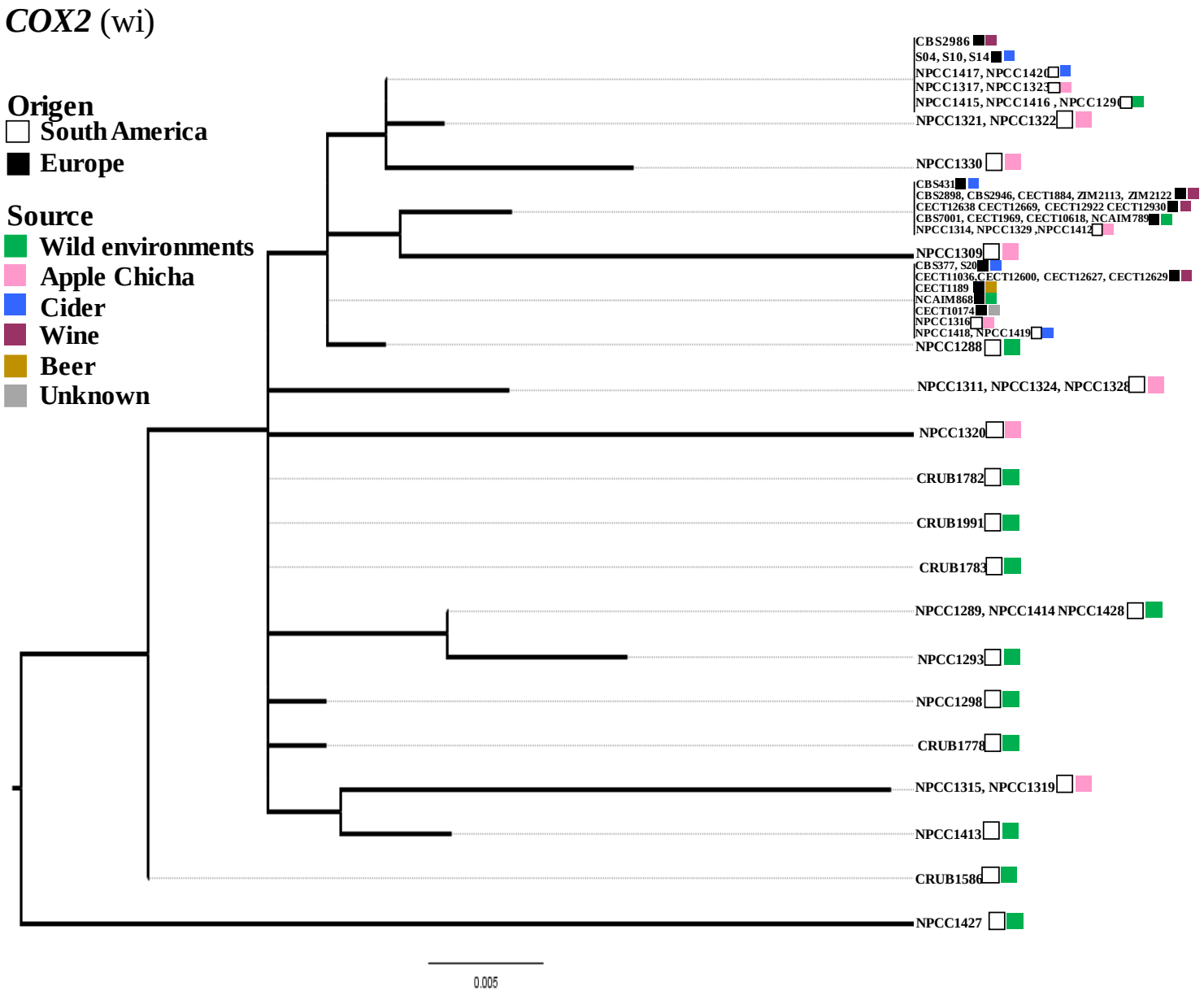

COX2 (wi)
Origen
South America
Europe
Source
Wild environments
Apple Chicha
Cider
Wine
Beer
Unknown
CBS2986
S04, S10, S14
NPCC1417, NPCC142(
NPCC1317, NPCC1325
NPCC1415, NPCC1416 , NPCC129(
NPCC1321, NPCC1322
NPCC1330
CBS431
CBS2898, CBS2946, CECT1884, ZIM2113, ZIM2122
CECT12638 CECT12669, CECT12922 CECT12930
CBS7001, CECT1969, CECT10618, NCAIM789
NPCC1314, NPCC1329 ,NPCC1412
NPCC1309
CBS377, S20
CECT11036,CECT12600, CECT12627, CECT12629
CECT1189
NCAIM868
CECT10174
NPCC1316
NPCC1418, NPCC1419
NPCC1288
NPCC1311, NPCC1324, NPCC1328
NPCC1320
CRUB1782
CRUB1991
CRUB1783
NPCC1289, NPCC1414 NPCC1428
NPCC1293
NPCC1298
CRUB1778
NPCC1315, NPCC1319
NPCC1413
CRUB1586
NPCC1427
0.005

**Figure S5.B**

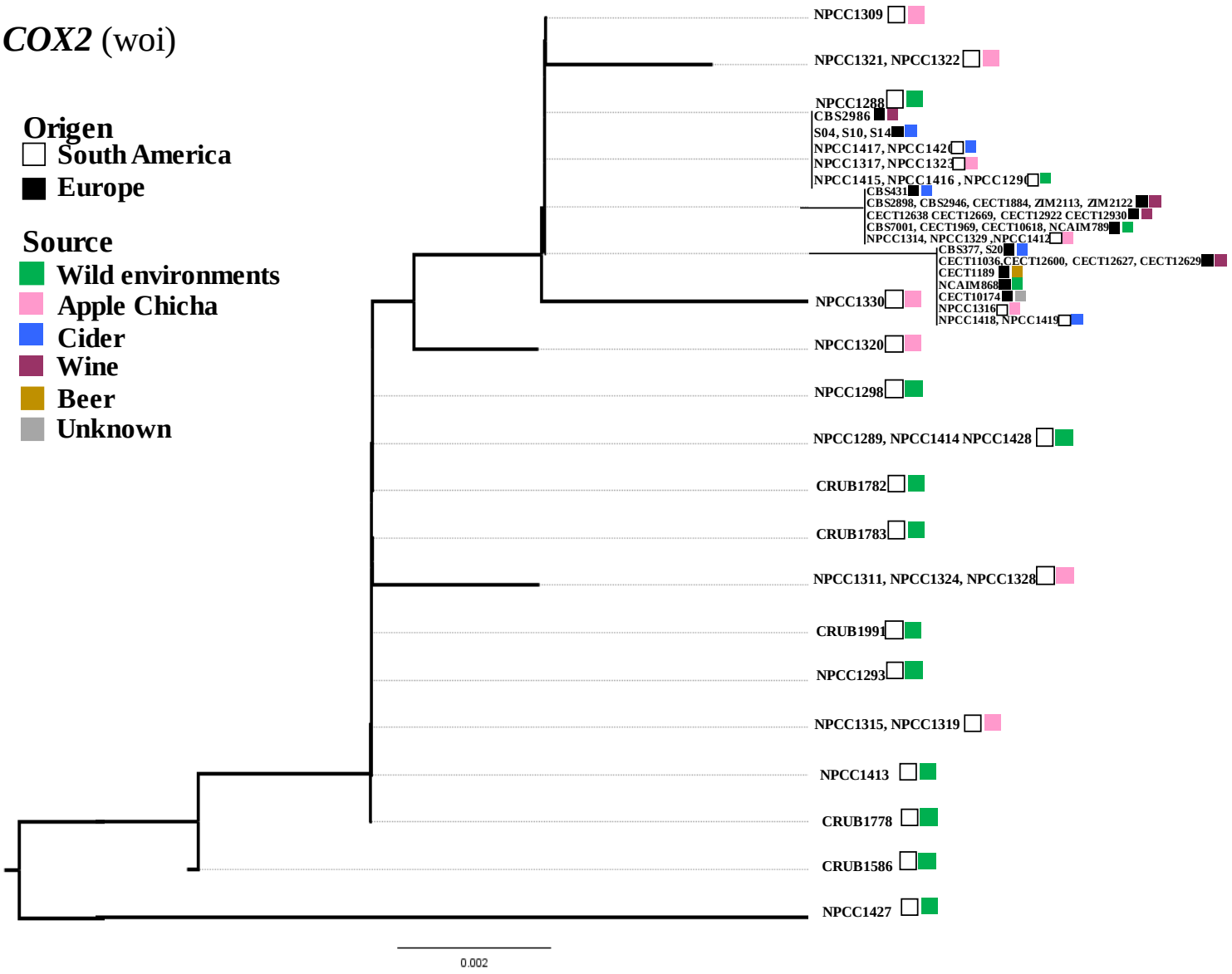

COX2 (woi)
Origen
South America
Europe
Source
Wild environments
Apple Chicha
Cider
Wine
Beer
Unknown
NPCC1309
NPCC1321, NPCC1322
NPCC1288
CBS2986
S04, S10, S14
NPCC1417, NPCC142(
NPCC1317, NPCC1323
NPCC1415, NPCC1416 , NPCC129(
CBS431
CBS2898, CBS2946, CECT1884, ZIM2113, ZIM2122
CECT12638 CECT12669, CECT12922 CECT12930
CBS7001, CECT1969, CECT10618, NCAIM789
NPCC1314, NPCC1329 ,NPCC1412
CBS377, S20
CECT11036,CECT12600, CECT12627, CECT12629
CECT1189
NCAIM868
CECT10174
NPCC1316
NPCC1418, NPCC1419
NPCC1330
NPCC1320
NPCC1298
NPCC1289, NPCC1414 NPCC1428
CRUB1782
CRUB1783
NPCC1311, NPCC1324, NPCC1328
CRUB1991
NPCC1293
NPCC1315, NPCC1319
NPCC1413
CRUB1778
CRUB1586
NPCC1427
0.002

Table1

Table2

Table Supplementary S1

Click here to access/download

**Table**

Supplementary Table 1. Strains.xlsx

Table Supplementary S2

Click here to access/download

**Table**

Supplementary Table 2.docx