



HAL
open science

Late Pleistocene-dated divergence between South Hemisphere populations of the non-conventional yeast *L. cidri*

Pablo Villarreal, Carlos Villarroel, Sam O'Donnell, Nicolas Agier, Julian Quintero-Galvis, Tomas Peña, Roberto Nespolo, Gilles Fischer, Cristian Varela, Francisco Cubillos

► **To cite this version:**

Pablo Villarreal, Carlos Villarroel, Sam O'Donnell, Nicolas Agier, Julian Quintero-Galvis, et al.. Late Pleistocene-dated divergence between South Hemisphere populations of the non-conventional yeast *L. cidri*. *Environmental Microbiology*, 2022, 10.1111/1462-2920.16103 . hal-03873794

HAL Id: hal-03873794

<https://hal.science/hal-03873794>

Submitted on 4 Dec 2022

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

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

1 **Late Pleistocene-dated divergence between South Hemisphere**
2 **populations of the non-conventional yeast *L. cidri*.**

3 Pablo Villarreal^{1,2}, Carlos A. Villarroel^{2,3,4}, Sam O'Donnell⁵, Nicolas Agier⁵, Julian F.
4 Quintero-Galvis^{2,6}, Tomas A. Peña^{1,2}, Roberto F. Nespolo^{2,6,7,8}, Gilles Fischer⁵, Cristian
5 Varela^{9,10}, and Francisco A. Cubillos^{1,2,8*}.

6 ¹*Facultad de Química y Biología, Departamento de Biología, Universidad de Santiago de Chile, Santiago,*
7 *9170022, Chile.*

8 ²*Millennium Institute for Integrative Biology (iBio), Santiago, 7500574, Chile.*

9 ³*Instituto de Ciencias Biológicas, Universidad de Talca, 1 Poniente 1141, Talca 3460000, Chile.*

10 ⁴*Instituto de Investigación Interdisciplinaria (I3), Universidad de Talca, Talca 3460000, Chile.*

11 ⁵*Laboratory of Computational and Quantitative Biology, CNRS, Institut de Biologie Paris-Seine, Sorbonne*
12 *Université, F-75005, Paris, France.*

13 ⁶*Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia, Chile.*

14 ⁷*Center of Applied Ecology and Sustainability (CAPES), Facultad de Ciencias Biológicas, Universidad*
15 *Católica de Chile, Santiago, Chile.*

16 ⁸*Millennium Nucleus of Patagonian Limit of Life (LiLi)*

17 ⁹*The Australian Wine Research Institute, Glen Osmond, PO Box 197, Adelaide, SA 5064, Australia.*

18 ¹⁰*Department of Wine and Food Science, University of Adelaide, Glen Osmond, Adelaide SA 5064, Australia.*

19 *Corresponding author. Email address: francisco.cubillos.r@usach.cl

20 Running title: Phylogenomics of wild *L. cidri* strains.

21 **ABSTRACT**

22 Most organisms belonging to the *Saccharomycotina* subphylum have high genetic diversity and a vast
23 repertoire of metabolisms and lifestyles. *Lachancea cidri* is an ideal yeast model for exploring the interplay
24 between genetics, ecological function, and evolution. *L. cidri* diverged from the *Saccharomyces* lineage
25 before the whole-genome duplication and is distributed across the South Hemisphere, displaying an important
26 ecological success. We applied phylogenomics to investigate the genetic variation of *L. cidri* isolates obtained
27 from Australia and South America. Our approach revealed the presence of two main lineages according to
28 their geographic distribution (Aus and SoAm). Estimation of the divergence time suggest that SoAm and Aus
29 lineages diverged near the last glacial maximum event during the Pleistocene (64-8 KYA). Interestingly, we
30 found that the French reference strain is closely related to the Australian strains, with a recent divergence
31 (405-51 YA), likely associated to human movements. Additionally, we identified different lineages within the
32 South American population, revealing that Patagonia contains a similar genetic diversity comparable to that of
33 other lineages in *S. cerevisiae*. These findings support the idea of a Pleistocene-dated divergence between
34 South Hemisphere lineages, where the *Nothofagus* and *Araucaria* ecological niches likely favored the
35 extensive distribution of *L. cidri* in Patagonia.

36

37 **Keywords:** yeasts, phylogenomics, genetic diversity, population structure, *Lachancea cidri*, South
38 Hemisphere

39

40

41

42

43

44

45

46

47

48

49

50

51 INTRODUCTION

52 Evaluating the genetic diversity of non-model species is essential for obtaining a thorough understanding of
53 biodiversity and for analyzing the complexity of genotype-phenotype relationships in nature (Fournier et al.,
54 2017). In this sense, microorganisms are particularly important since they represent a large proportion of the
55 ecosystem's biomass and provide several ecosystem services such as foods, fuel, and nitrogen fixation
56 capacity (Chen et al., 2019; Mirza et al., 2014). Particularly important are organisms belonging to the
57 *Saccharomycotina* subphylum (yeasts), many of which have been domesticated for human benefit,
58 representing an ideal group of model and non-model organisms for phylogenomics and population genomics
59 studies (Peter and Schacherer, 2016). However, most phylogenomic studies have been focused on
60 *Saccharomyces* species and to a lesser extent on other genera like *Candida* or *Lachancea*, ignoring the
61 overwhelming biodiversity in the *Saccharomycotina* subphylum (Almeida et al., 2015; Gallone et al., 2016;
62 Goncalves et al., 2016; Liti et al., 2009). In this context, different phylogenomic studies have provided new
63 evidence on the evolutionary history of non-conventional yeast species and the genetic basis of their
64 phenotypic diversity (Cadez et al., 2019; Friedrich et al., 2012; Vakirlis et al., 2018; Vakirlis et al., 2016).

65 A particularly interesting group of species are the *Lachancea* cluster (Porter et al., 2019a), initially proposed
66 in 2003 by Kurtzman P. (Kurtzman, 2003). This genus diverged from the *Saccharomyces* lineage prior to the
67 ancestral whole-genome duplication (WGD), 100 or more MYA (Hranilovic et al., 2017). *Lachancea* species
68 harbor eight chromosomes and exhibit high levels of synteny in the coding regions (Porter et al., 2019b).
69 However, chromosome sizes and nucleotide diversity differ significantly between species (Lachance and
70 Kurtzman, 2011) and there is higher genetic diversity and genetic distance between *Lachancea* species
71 compared to *Saccharomyces* (Vakirlis et al., 2016). *Lachancea* species have been recovered from a wide
72 variety of ecological niches, ranging from plants (Esteve-Zarzoso et al., 2001; Gonzalez et al., 2007; Romano
73 and Suzzi, 1993a, b), tree barks (Nespolo et al., 2020; Villarreal et al., 2021), tree exudates (Varela et al.,
74 2020), insects (Phaff et al., 1956), soil (Lee et al., 2009; Mesquita et al., 2013), water (Kodama and Kyono,
75 1974) as well as food and beverages (Magalhães et al., 2011; Marsh et al., 2014; Nova et al., 2009; Pereira et
76 al., 2011; Tzanetakis et al., 1998; Wojtatowicz et al., 2001). Despite the large number of ecosystems where

77 *Lachancea* spp. are present, we still lack information concerning the genetic and genomic adaptations needed
78 to survive in such a range of ecological niches (Porter et al., 2019b). In this sense, the number of currently
79 available *Lachancea* genomes is still low, and therefore most species remain unexplored.

80 The availability of complete genomes of many individuals is a prerequisite for exploring the diversity and
81 genomic structure of a species. Thus far, genomic studies have been carried out in *Lachancea thermotolerans*,
82 *L. fermentati*, and *L. kluyveri*, contributing to a better understanding of the population structure, ecology, and
83 evolution of these species (Porter et al., 2019b). *L. thermotolerans* is the best-studied species of the genus
84 (Banilas et al., 2016; Hranilovic et al., 2017; Hranilovic et al., 2018), and it has been reported that the
85 evolution of *L. thermotolerans* was driven by geographic isolation and local adaptation (Hranilovic et al.,
86 2017). Nonetheless, studies related to the natural population diversity, evolution, and biotechnological
87 potential of other species of the genus are still scarce, mainly due to the low number of isolates sampled to
88 date.

89 Recently, *L. cidri* was recovered from cider fermentations in Europe, eucalyptus tree sap in Australia and
90 from *Nothofagus* forests in Patagonia. The isolates exhibit an interesting phenotypic diversity and
91 biotechnological potential for wine and mead fermentations (Nespolo et al., 2020; Villarreal et al., 2021).
92 Patagonian *L. cidri* isolates showed a greater fitness in high throughput microcultivation assays for
93 fermentative-related conditions. Interestingly, microfermentations in different musts demonstrated the
94 potential of *L. cidri* in mead, surpassing the fermentative capacity of commercial *S. cerevisiae* strains,
95 together with providing a distinctive organoleptic profile to the final ferment, mostly because of the
96 production of acetic and succinic acid (Villarreal et al., 2021). However, the phenotypic diversity (and the
97 underpinning genetic variation) of this species has not been determined. The distribution of *L. cidri* in the
98 South Hemisphere represents an excellent opportunity to establish the ecological and geographic determinants
99 of the actual distribution of a non-conventional yeast (Porter et al., 2019a, b; Villarreal et al., 2021).

100 In order to address this problem, we characterized the genomic and phenotypic profiles of 55 *L. cidri* strains.
101 The isolates were collected from the bark of *Nothofagus* and *Araucaria* trees in Chile and from tree sap
102 samples of *Eucalyptus gunni* in Australia, together with a single isolate from cider fermentations in France.
103 Overall, we suggest that Patagonia represents a natural and unique reservoir of genetic and phenotypic

104 diversity for the species. This is explained by the broad and varied environments in the altitudinal and
105 latitudinal gradients provided by this biome. Then, our results increased the number of lineages of *Lachancea*
106 *cidri*, extending our current knowledge of genetics, ecology, and biotechnological potential of this novel
107 natural resource from Patagonia.

108 **EXPERIMENTAL PROCEDURES**

109 **Sample area, yeast isolation, and *L. cidri* identification.**

110 Sampling was performed as previously described (Sampaio and Goncalves, 2008) in twelve locations in
111 primary Forests in Chile (Table S1). The sampling strategy included the identification of the tree species and
112 obtaining a single bark sample from each tree. Each sample was obtained from trees separated for, at least, 5-
113 10 meters. Overall, we obtained between 25-100 samples per sampling site. The 5 g bark samples were
114 collected aseptically from different *Nothofagus pumilio*, *Nothofagus dombeyi*, *Nothofagus antartica* and
115 *Araucaria araucana* trees, and immediately incubated in 10 mL enrichment media (2% yeast nitrogen base,
116 1% raffinose, 2% peptone and 8% ethanol) (Sampaio and Goncalves, 2008). After 10 days, 5 μ L were
117 streaked onto YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose) agar and isolated colonies were
118 stored in 20% v/v glycerol at -80°C. *L. cidri* isolates were determined by amplification and sequencing of the
119 internal transcribed spacer region (*ITS1-5.8rRNA-ITS2*) using universal fungal primers, ITS-1 5'-
120 TCCGTAGGTGAACCTGCGG-3' and ITS-4 5'-TCCTCCGCTTATTGATATGC-3'. Genomic DNA was
121 extracted from 5-mL cultures grown overnight in YPD media with shaking at 20°C, using 50 mg mL⁻¹
122 Zymolyase 20T Concentrate and the Wizard® Genomic DNA Purification Kit according to the manufacturer's
123 instructions. Polymerase Chain Reactions (PCR) were performed in a final volume of 10 μ L containing 0.2
124 μ L 10 mg mL⁻¹ BSA (Promega), 3.5 μ L 2X GoTaq® Colorless Master Mix, 0.6 μ L 10 μ M ITS1 primer, 0.6
125 μ L 10 μ M ITS4 primer, 1 μ L diluted DNA (10 ng μ L⁻¹) and 4.1 μ L nuclease-free water (Fujita et al., 2001).
126 PCR amplifications were performed in a T100™ Thermal Cycler thermocycler (BioRad) as follows: initial
127 denaturation at 94°C for 5 min, then 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 2 min,
128 and extension at 72°C for 2 min; followed by a final extension at 72°C for 5 min. To determine the *ITS1-*
129 *5.8SrRNA-ITS2* PCR amplicon size, reaction products were analyzed in a 2% (w/v) agarose gel by
130 electrophoresis. ITS sequencing was performed in an Applied Biosystems™ 3500 instrument (AustralOmics,

131 Valdivia, Chile). Additionally, *L. cidri* isolates were obtained from tree sap samples of *Eucalyptus gunnii*
132 from the Central Plateau of Tasmania, Australia, as previously described (Varela et al., 2020). Briefly, the
133 Australian isolates originated from three different locations in the Central Plateau, Trawtha Makuminya,
134 Skullbone Plains (7 kms away) and Five Rivers/Serpentine (18 kms away from the first location). From a
135 total number of 1,029 yeast cultures isolated from 47 bark, soil and/or sap samples, 138 cultures were initially
136 identified as *L. cidri* by ITS sequencing. All 138 *L. cidri* isolates were screened under different environmental
137 conditions (different carbon sources and stress conditions) in solid media and 25 were selected based on
138 phenotype differences for further analysis. The isolates obtained and their geographical locations are listed in
139 Table S1.

140 **FACS analysis**

141 DNA content was analysed using the propidium iodide (PI) staining assay. Cells were first recovered from the
142 glycerol stocks on YPD solid media and incubated overnight at 30 °C. The following day, a fraction of each
143 patch was taken with a pipette tip, transferred into liquid YPD media, and incubated overnight at 30 °C. Then,
144 1 mL of each culture was taken and resuspended in 2.3 mL cold 70% ethanol. Cells were fixed overnight at
145 4 °C, washed twice with PBS, resuspended in 100 µL staining solution (15 µM PI, 100 µg mL⁻¹ RNase A,
146 0.1% v/v Triton-X, in PBS) and finally incubated for 3 h at 37 °C in the dark. Ten thousand cells for each
147 sample were analysed on a FACS-Calibur flow cytometer. Cells were excited at 488 nm and fluorescence was
148 collected with an FL2-A filter. Three strains with known ploidy were used as a control: Two *S. euabaynus* (n
149 and 2n) and one *S. cerevisiae* (4n).

150 **Whole-genome sequencing and Variant Calling.**

151 Genomic DNA from *L. cidri* was prepared for whole-genome sequencing using a Qiagen Genomic-tip 20/G
152 kit (Qiagen, Hilden, Germany) as previously described (Nespolo et al., 2020) and sent for sequencing using
153 DNBseq technology (BGISEQ-G400 platform) (Liu et al., 2018). Read quality was checked using FastQC
154 0.11.8 (Andrews, 2010). Reads were processed with fastp 0.19.4 (low quality 3' end trimming, 40 bp
155 minimum read size) (Brickwedde et al., 2018; Chen et al., 2018). We also obtained publicly available
156 sequencing reads of *L. cidri* CBS2950 (Agier et al., 2018), which were processed identically. Reads were
157 aligned against the *L. cidri* CBS2950 reference genome (Vakirlis et al., 2016) using BWA-mem (options: -M -

158 R) (Li, 2013). Mapping quality and overall statistics were collected and examined with Qualimap (García-
159 Alcalde et al., 2012). Sorting and indexing of output bam files were performed using SAMTOOLS 1.9 (Li et
160 al., 2009). A *L. fermentai* isolate (CBS770) was also mapped against the *L. cidri* CBS2950 genome for
161 phylogenetic analysis (Bellut et al., 2020). Mapping files were tagged for duplicates using MarkDuplicates of
162 Picard tools 2.18.14 (<http://broadinstitute.github.io/picard/>). Variant calling and filtering were done with
163 GATK version 4.0.10.1 (DePristo et al., 2011). More specifically, variants were called per sample and
164 chromosome using HaplotypeCaller (default settings), after which variant databases were built using
165 GenomicsDBImport. Genotypes for each chromosome were called using GenotypeGVCFs (-G
166 StandardAnnotation). Variant files were merged into one genome-wide file using MergeVcfs. This file was
167 divided into SNP calls and INDEL calls using SelectVariants. We applied recommended filters for coverage
168 (> 10 mapping reads = "FORMAT/DP>10") and quality (--minQ 30) (Van der Auwera et al., 2013a). This
169 VCF file was further filtered, depending on the requirements of the given analysis, using vcftools (Van der
170 Auwera et al., 2013a, b). For all datasets, we only considered SNPs that had no missing data using vcftools
171 option --max-missing 1.

172 **Phylogeographic reconstruction**

173 To perform phylogeny on our SNP dataset, a VCF file containing 218,266 SNPs obtained from 55 strains
174 was converted to phylip format using vcf2phylip (Ortiz., 2019) and used as input for IQ-TREE (Nguyen et al.,
175 2015) to generate a maximum likelihood phylogeny with the ultrafast bootstrap option and ascertainment bias
176 correction (-st DNA -o CBS707 -m GTR+ASC -nt 8 -bb 1000) (Hoang et al., 2018). The number of
177 parsimony informative sites was 10,680 SNPs. Trees were visualized on the iTOL website
178 (<http://itol.embl.de>).

179 **Population structure analyses**

180 Given that different tools can provide inconsistent results (Stift et al., 2019), we used STRUCTURE,
181 fineSTRUCTURE and SMARTPCA to determine the population structure in *L. cidri*. For STRUCTURE
182 analysis, a thinned version of the VCF file was generated with vcftools 0.1.15 (--thin 500) (Danecek et al.,
183 2011), containing 11,447 similarly-spaced SNPs, while including only *L. cidri* strains. STRUCTURE was run
184 on this dataset five times for K values ranging from 1 to 7, with 10,000 burn-in and 100,000 replications for

185 each run, using admixture model, infer alpha, lambda =1, fpriormean =1, unifprioralpha 1, and alpha max 10.
186 The structure-selector website was employed to obtain the optimal K values
187 (<http://lme.qdio.ac.cn/StructureSelector/>) (Li and Liu, 2018) according to the Evanno method (Evanno et al.,
188 2005). The results for each K were plotted using CLUMPAK (Kopelman et al., 2015) and visualized using a
189 structure plot (<http://omicspeaks.com/strplot2/>) (Ramasamy et al., 2014). In addition, we performed
190 clustering analyses of the same samples by using SMARTPCA without outlier removal (Patterson et al.,
191 2006). For fineSTRUCTURE analysis (Lawson et al., 2012), a VCF file that included all SNPs among *L. cidri*
192 strains was phased using BEAGLE 3.0.4 (Browning and Browning, 2007). Since we lacked a *L. cidri*
193 recombination map, we used a constant recombination rate between consecutive SNPs based on the average
194 recombination rate of *L. kluyveri* (0.4 cM kbp⁻¹, (Brion et al., 2017)). All versus all chromosomal painting was
195 performed with Chromopainter V2, and the output was further analyzed with fineSTRUCTURE (-x 100000 -y
196 100000 -z 1000). Plotting of the ancestry matrix was done using fineSTRUCTURE R scripts.

197 **Population genetics**

198 We estimated the pairwise of nucleotide diversity, π and the proportion of segregant sites, θ_w using the R
199 package PopGenome (Pfeifer et al., 2014). F_{ST} Values across the genome were calculated with StAMPP 1.5.1
200 Weir and Cockerham's unbiased estimator to obtain 95% confidence intervals by performing 5,000 bootstraps
201 (Pembleton et al., 2013; Weir and Cockerham, 1984). The IBD (Isolation by Distance) was performed using
202 the seven locations from Chile. The significance of correlations between distance matrices in the Mantel test
203 was evaluated using 10,000 permutations. The geographic distances were calculated using latitude/longitude
204 of all relevant sampling points using the function of GenAlEx v 6.5 (Peakall and Smouse, 2012). Population
205 pairwise F_{ST} values were used to build the genetic distance matrices for this test. We predicted the number of
206 generations since the most recent common ancestor of any pair of lineages using the mutation parameter, as
207 previously described (Ruderfer et al., 2006).

$$208 \quad \theta_w = 2N_0\mu, \text{ where } N_0 \text{ represents the number of generations}$$

209 Given that the single-base mutation rate is similar in different yeasts species (Farlow et al., 2015; Fijarczyk et
210 al., 2021), we used the mutation rate ($\mu = 1.84 \times 10^{-10}$ mutations per nucleotide per generation) previously
211 reported from laboratory estimates on *S. cerevisiae* (Lynch et al., 2008) for calculations.

212 **Species delimitation analysis**

213 We used the SNAPP software (Bryant et al., 2012) to test if lineages of *L. cidri* could be delimited at the
214 molecular level using the Bayes Factors criteria implemented in BEAST v2.6 (Bouckaert et al., 2019). We
215 evaluated three competing lineages models with the BFD* method (Leache et al., 2014), using as an outgroup
216 the Australian strain (group A, Table S2), and as the interest group we used the strains from B: Huilo-Huilo,
217 C: Coyhaique, D: Altos de Lircay. SNAPP analyses were only performed with a subset of individuals per
218 lineages for computational feasibility; the total number of SNPs in the dataset was 218,266. To change the
219 VCF haploid to diploid we used the java code VcfFilterJdk (Lindenbaum and Redon, 2018). We used the
220 Marginal likelihood estimation with 12 steps for the dataset to steppingstone sampling. We used default
221 settings (e.g., Yule prior for species tree and branch length estimations; gamma distributions for ancestral
222 theta), with five million burn-in iterations. To ensure convergence, we confirmed that our runs reached
223 effective sample size values >500 after burn-in using the program Tracer v1.4 (Rambaut et al., 2018).
224 Maximum credibility trees were generated using TreeAnnotator in BEAST v2.6, and the posterior
225 distributions of species trees were visualized using DensiTree (Bouckaert, 2010).

226 **Phenotypic diversity among isolates**

227 To determine the phenotypic diversity across the *L. cidri* isolates, we estimated growth and biomass
228 production in several environmental conditions that represent different yeast habitats, either in nature or in
229 industrial settings. For this, we performed high-throughput phenotyping in 96-well microculture plates as
230 previously described (Kessi-Perez et al., 2016). Briefly, cells were pre-cultivated in 200 μ L YNB medium
231 without agitation at 20°C for 48 hours. For the experimental run, each well was inoculated with ten μ L of pre-
232 inoculum to an optical density (OD) of 0.03–0.1 in 200 μ L of different media compositions (Table S3).
233 Culture OD was measured at 620 nm every 30 minutes for 96 hours. From these data, three parameters were
234 estimated: lag phase, growth rate (μ_{max}), and maximum OD using the Growth Rates software with default
235 parameters (Hall et al., 2014).

236 **Statistical analysis**

237 All statistical analyses were performed using biological triplicates. The differences were considered
238 statistically significant at p -values <0.05 . The comparison of growth rate between populations and isolation
239 locations was undertaken with a paired t -test. The heat map and phenotypic principal component analysis
240 (PCA) were created using R software (Team, 2008), the "pheatmap" and "prcomp" packages stats 3.6.0, and
241 plotted using the "ggplot2" and "ggbiplot" packages.

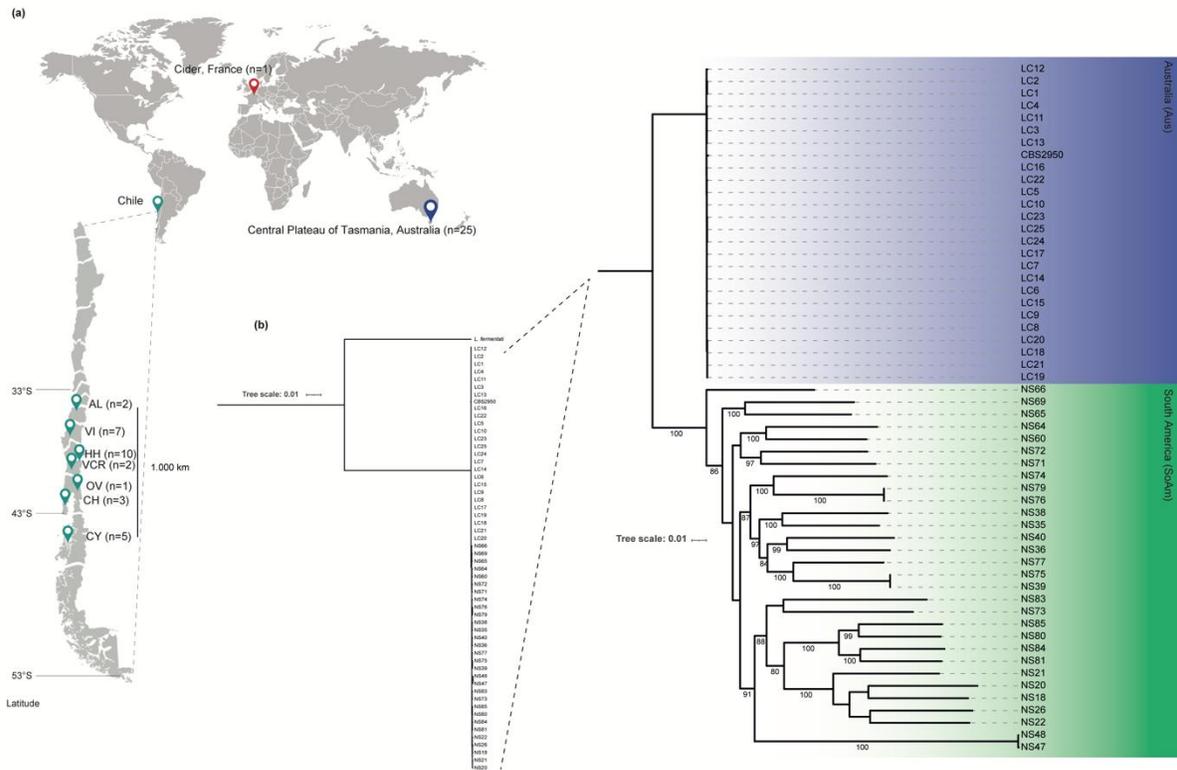
242

243 **RESULTS**

244 **High prevalence of *L. cidri* among non-*Saccharomyces* yeasts in natural environments**

245 To explore the distribution of *L. cidri* in central and southern Chile, we obtained yeast colonies from different
246 tree bark samples collected between 2017 and 2019 that had been previously identified as non-*Saccharomyces*
247 (Nespolo et al., 2020) using the internal transcribed spacer region (*ITS1-5.8rRNA-ITS2*). Cultures identified as
248 *L. cidri* were found in 7 of the 13 sampling regions, between Altos de Lircay National Park (VII Maule
249 Region, Chile) in the north and Coyhaique National Reserve (XI Aysén Region, Chile) in the south.
250 Moreover, the distribution of *L. cidri* were widely distributed in South Chile, covering regions from the
251 Pacific coast up to the Andes Mountains (approximately 1,000 km long and 300 km wide) (Fig 1a, Table S4).
252 Overall, we identified 30 *L. cidri* strains in Patagonia, meaning that *L. cidri* is the most-represented non-
253 conventional ethanol-tolerant species in our survey. The *L. cidri* isolation frequency varied between 1.33 to
254 11.11 % depending on the national park, independent of latitude (Pearson $r = -0.338$, p -value > 0.05) (Fig
255 S1a, Table S4). In addition, when only non-*Saccharomyces* yeasts were considered, the isolation frequency
256 values varied between 9.72-100 % (Fig. S1b, Table S4), indicating a high prevalence of this non-conventional
257 yeast in *Nothofagus* bark samples. The samples from Chile were obtained in four different tree species, *N.*
258 *pumilio* (representing 26.7 % of the samples), *N. dombeyi* (46.7 %), *N. antartica* (3.3 %), and *A. araucana*
259 (23.3 %) (Fig. S1c). The frequency of *L. cidri* in Patagonia was determined by the host tree (hypergeometric
260 test, Table S5); *N. dombeyi* was the most prevalent host. Additionally, cultures identified as *L. cidri* were
261 obtained in a previous study from eucalyptus tree sap samples collected in the Central Plateau of Tasmania,
262 Australia (Varela et al., 2020) (Fig. 1a). In this case, *L. cidri* represented 13% of the total non-*Saccharomyces*
263 isolates found. Our results demonstrate a high prevalence of *L. cidri* in Patagonia, distributed from high

264 altitude (Andes Mountains) to low altitude regions (near the Pacific Ocean), covering an extensive range of
 265 environmental conditions.



266

267 **Figure 1. Geographic distribution and phylogeny of *L. cidri* isolates.** (a) World map depicting the different
 268 locations where *L. cidri* were isolated. In Chile, strains were obtained from seven sampling locations. (b)
 269 Maximum likelihood tree depicting the genetic relationships between 56 strains using 218,266 biallelic SNPs
 270 (substitution model GTR+F+ASC) and manually-rooted with *L. fermentati* as the outgroup.

271

272 **Phylogenomic studies in *L. cidri***

273 To determine the phylogenetic relationships between strains and the genomic variations in *L. cidri*, we
 274 initially estimated the ploidy levels of the isolates. FACs analysis revealed that all *L. cidri* isolates were
 275 haploids (Fig. S2). Subsequently, we sequenced the complete genomes of 30 strains from South America, and
 276 included an extra set of 25 strains from Australia obtained from sap samples of *Eucalyptus gunnii* from the
 277 Central Plateau of Tasmania, Australia. In addition, we incorporated previously published data for the
 278 reference strain *L. cidri* CBS2950 (isolated from cider in France) and the *L. fermentati* strain CBS707, which
 279 we used as an outgroup. On average, across the 30 South American genomes, we obtained 2,670 SNPs per

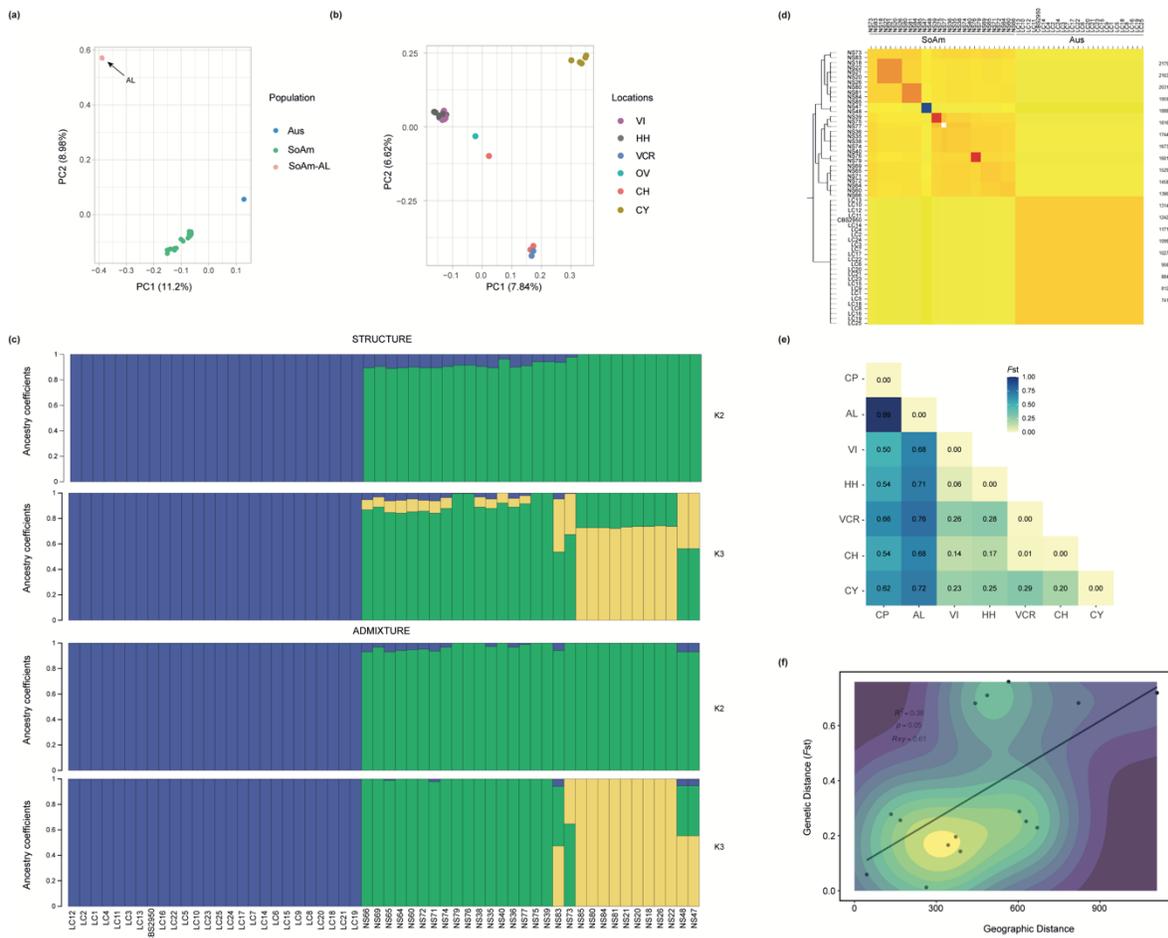
280 strain relative to the reference genome (SNPs were found on average every ~ 3.8 kb between two strains). On
281 the other hand, across the 25 Australian genomes, we obtained on average 36 SNPs per strain relative to the
282 reference genome (SNPs were found on average every ~ 282 kb), indicating apparent differences compared to
283 the South American group of strains. In parallel, we found different numbers of insertions and deletions
284 depending on the strain relative to the reference genome, ranging from 79 in the Australian strains to 124 in
285 the South American strains (the Bioinformatic Summary statistics are shown in Table S6). Interestingly, the
286 high number of INDELs are unique to the reference strain, rather than a general trend between any two strains
287 (Table S6).

288 The phylogeographic result with maximum-likelihood phylogeny revealed a topology with two well-
289 supported main clades separating South American and Australian strains (Fig. 1b). Australian strains clustered
290 into a single clade (hereafter referred to as Aus), together with the French reference strain *L. cidri* CBS2950.
291 This clustering of Australian and the CBS2950 strain, together with the low number of polymorphisms found
292 between them (Table S7), suggest a recent migration event between Australia and France (or another closely-
293 located region). In contrast, the South American strains were separated in a more complex clade distribution,
294 with substantial differences in branch lengths, as well as different subclades with phylogeographic structures
295 (Fig. 1b). The South American clade (hereafter referred to as SoAm) harbored 15,706 unique SNPs compared
296 to the reference strain (Table S7). Interestingly, the two isolates from Altos de Lircay National Park (AL) (the
297 northernmost locality) (Fig. 1a) cluster together and harbor the greatest genetic divergence within the SoAm
298 group (Fig 1b). Overall, these results suggest a broader genetic diversity in SoAm lineages, where the AL
299 branch showed the highest genetic divergence in the latitudinal gradient (Fig. 1b).

300 **Population structure analyses reveal two main genetic groups in *L. cidri*.**

301 To explore the connection between strains and localities, we performed an analysis of the distribution of the
302 genetic variants found in the 55 *L. cidri* strains (Fig. 2a and b). The principal component analysis (PCA)
303 showed a clear separation of the strains depending on their geographical origin, where according to the first
304 two components, PC1 (11.2 %) and PC2 (8.98%), the Aus group is completely separated from the other set of
305 strains (Fig. 2a), illustrating again the genetic differentiation between Aus and SoAm *L. cidri*. Interestingly,
306 SoAm exhibited a separation into two additional groups, one corresponding to AL (most divergent clade)

307 located in the upper part of the PCA plot and the other in the lower part corresponding to the remaining SoAm
 308 strains (Fig. 2a). A second PCA considering just the latter group, clearly revealed that the strains distributed
 309 based on their geographic origin of isolation, including latitude and longitude. According to the first two
 310 principal components, PC1 (7.84 %) and PC2 (6.62 %), those strains collected near to the Pacific Ocean
 311 (Valdivian Coastal Reserve and Chiloé National Park) separated from those isolates obtained in mountainous
 312 regions (color chart, Fig. 2b).



313

314 **Figure 2. Population structure of *L. cidri*.** (a) Plot of the genomic variation distribution of 56 *L. cidri* strains
 315 based on the first two components of a PCA performed using 13,275 unlinked SNPs. Each dot represents a
 316 single strain. The black arrow denotes strains from Altos de Lircay National Park (AL). (b) PCA only
 317 considering strains from Patagonia. The color reflects the grouping by location. Villarrica National Park (VI),
 318 Huilo-Huilo Biological Reserve (HH), Valdivian Coastal Reserve (VCR), Osorno Volcano (OV), Chiloé
 319 National Park (CH), and Coyhaique National Reserve (CY). (c) STRUCTURE and ADMIXTURE. An
 320 optimum K = 2 group is shown in both cases. (d) Heat map of a Co-ancestry matrix obtained using
 321 fineSTRUCTURE chunk counts. Each row and column represent an isolated color indicating genetic sharing
 322 (yellow = low sharing, blue = high sharing). The tree shows the clusters inferred from the co-ancestry matrix.
 323 (e) F_{ST} values between localities, using the same abbreviations shown in (b) as well as Central Plateau of

324 Tasmania (CP), Australia. (f) IBD (Isolation by distance) estimates considering seven localities of southern
325 Chile. The density of data points is indicated by colors.

326

327 To determine the population genetic structure of *L. cidri*, we performed several complementary approaches,
328 including STRUCTURE, ADMIXTURE, and fineSTRUCTURE clustering (Fig. 2c and d). The three
329 analyses revealed a high degree of differentiation between strains isolated from South America and Australia.
330 STRUCTURE and ADMIXTURE analyses indicated an optimum $K = 2$ groups ($\Delta K_2 = 301$ for
331 STRUCTURE, Fig. 2c, Table S8), showing the presence of two genetic groups, SoAm and Aus. Further
332 analysis using the fineSTRUCTURE Co-ancestry matrix indicated a clear separation between the SoAm and
333 Aus isolates (Fig. 2d), confirming the presence of two large populations. Additionally, we identified a series
334 of subgroups within the SoAm clade, suggesting sub lineage in central and southern Chile confirming the
335 great differentiation and divergence of the group of isolates obtained from AL. To corroborate the results
336 obtained, we calculated the genetic differentiation (F_{ST}) between each of the localities from which *L. cidri*
337 isolates were obtained (Fig. 2e, Fig. S3a, Table S9). In this case, we found low to high F_{ST} values ranging
338 from 0.04 (p -value < 0.001) between Villarrica National Park and the Huilo-Huilo Biological Reserve, up to
339 0.99 (p -value < 0.001) between Altos de Lircay National Park (AL) and Central Plateau of Tasmania (CP),
340 Australia (Fig. 2e). Although the STRUCTURE and ADMIXTURE analyses did not suggest AL as a third
341 lineage, we found the highest F_{ST} values between AL and the other localities from South America, suggesting
342 that this could represent a third lineage in our study (Fig. 2e, Table S9). Furthermore, a genome wide F_{ST}
343 analysis clearly demonstrates an even distribution of values across the genome between AL and the other
344 South American strains, exhibiting a differentiation throughout the genome (Fig S3a). Intermediate and low
345 F_{ST} values were found between closer localities in southern Chile (Fig. 2e, Table S9). Under this scenario,
346 with IBD we found a moderately significant correlation ($R^2 = 0.39$, p -value = 0.05) (Fig. 2f) between genetic
347 differentiation and geographic distance in the localities of southern Chile, suggesting that the genetic
348 differences found in Patagonian *L. cidri* could result from the geographical distribution (distance in
349 kilometers).

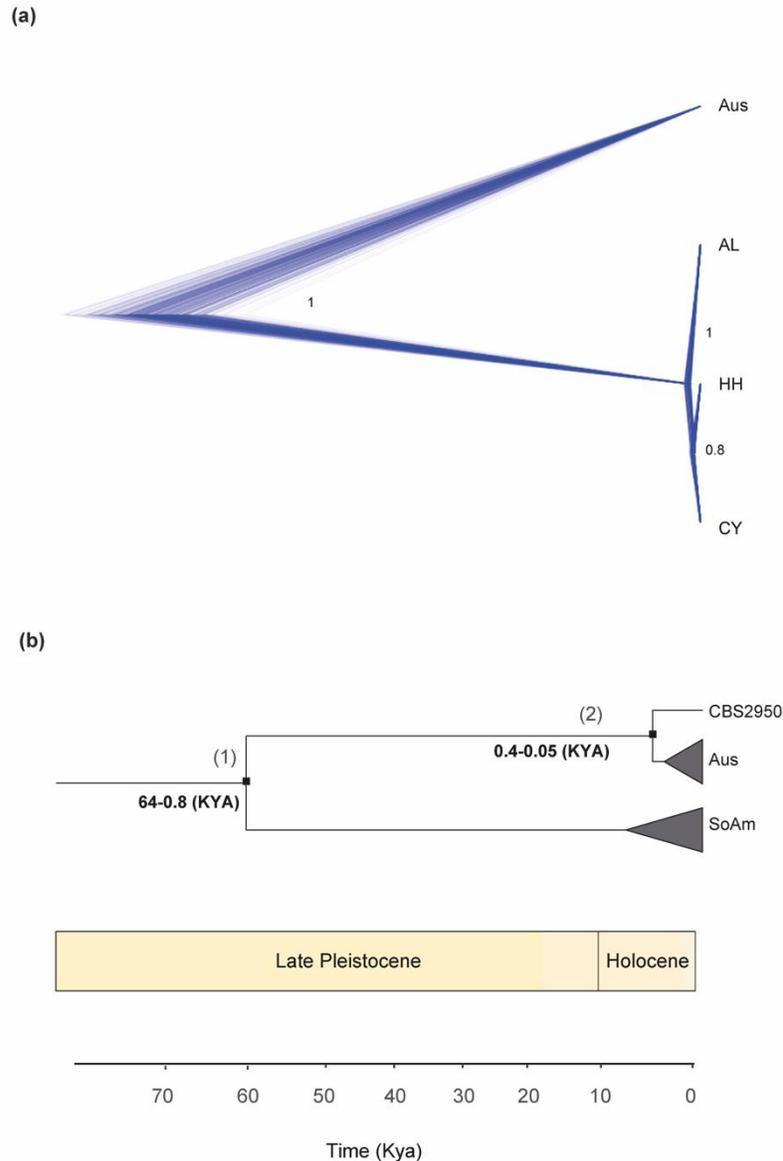
350 Southern SoAm isolates exhibited high genetic diversity. In this case, we calculated nucleotide diversity (π)
351 (Fig. S3b, Table S10), and observed a high nucleotide diversity in the SoAm population, most likely due to

352 the number of different locations from which the *L. cidri* strains were isolated (Fig. S3b). When analyzing the
353 diversity at each locality, higher π values were observed in sampling sites from southern Chile (e.g., Villarrica
354 National Park, Huilo-Huilo Biological Reserve, and Chiloe National Park) compared to central Chile (e.g.,
355 Altos de Lircay) or Australia, which exhibited the lowest π values (Fig. S3c), despite being the location with
356 the highest number of isolates (n=25). These results likely reflect a clonal population in Australia, possibly
357 affected by the sampling strategy or the small territory covered in Tasmania. Furthermore, the low number of
358 SNPs between the French reference strain and the Australian population, also suggests a likely low genetic
359 diversity in Australia. The low nucleotide divergence between Australian strains and the French reference
360 strain suggests a recent introduction of the species in Australia or Europe (Fig. S3c, Table S10).

361

362 **Pleistocene dated divergence between the main lineages in *L. cidri***

363 Given the genetic divergence of the AL clade and considering the distribution of SNPs according to the PCA,
364 the clustering in the co-ancestry matrix in fineSTRUCTURE and the high F_{ST} values compared to the rest of
365 the SoAm populations, we evaluated the possibility that the group of isolates from AL could represent a
366 different lineage in the species. For this, we performed a species delimitation analyses according to the
367 different models generated, and hypothesized the presence of one to three genetically-different lineages or
368 groups utilizing strains of Aus, AL, HH and CY as representatives of the different lineages (Fig. 3a). The
369 Bayes factor species delimitation (SNAPP) analysis supported the scenario where AL represents a different
370 lineage in the species (model 3, Fig 3a), exhibiting the highest Bayes factor (43421.01) (Table S2) and a high
371 Bayesian support (IB = 1, Fig 3a). In addition, this analysis also predicts the presence of four different groups
372 separating HH and CY as a different lineage (Table S2); however, the Bayesian support in the species tree
373 was low and therefore these could not be considered as different lineages (IB= 0.8, Fig 3a). Altogether, this
374 suggests the presence of three different lineages in *L. cidri*: Aus, AL and SoAm (HH and CY).



375

376 **Figure 3. Species delimitation and divergence time analysis in *L. cidri* populations.** (a) DensiTrees
 377 gathered in the Bayesian analyses, where nodes values correspond to the Bayesian inference (IB) result of the
 378 species tree. Strains from Australia (Aus), Altos de Lircay National Park (AL), Huilo-Huilo Biological
 379 Reserve (HH), and Coyhaique National Reserve (CY) are included in the analysis. The blue traces indicate
 380 the most frequently occurring topology; red indicates the second most likely topology and the third in light
 381 blue. (b) Estimation of the divergence time since the last common ancestor between groups. Time between (1)
 382 Australia lineage and SoAm lineage; Time between (2) Australia lineage and CBS2950. Timeline in Million's
 383 years (Ky).

384

385 Considering the results obtained, we estimated the divergence time since the most recent common ancestor
 386 between lineages. Using the population genomic data and following the same strategy as previously used for
 387 *L. kluyveri* and *S. cerevisiae* (Friedrich et al., 2015; Ruderfer et al., 2006) we predicted the number of

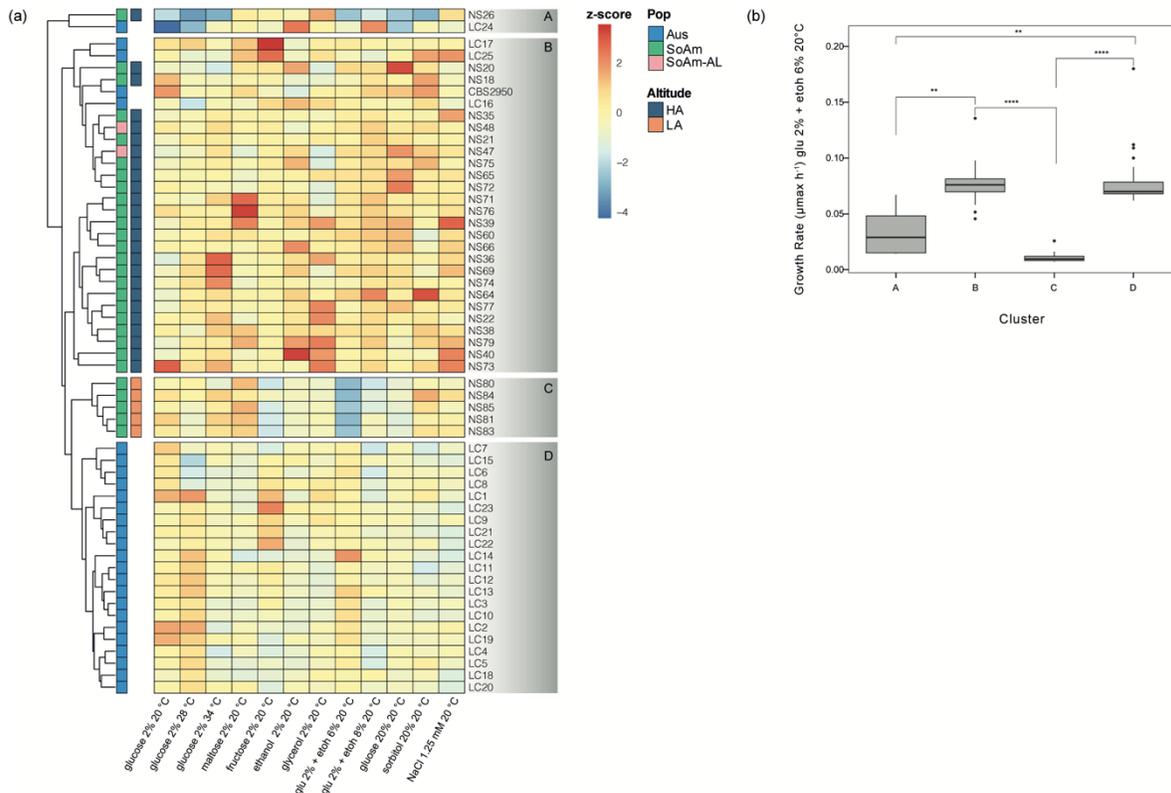
388 generations that have passed since the divergence of any two groups. Our data lead to an estimation of $23.9 \times$
389 10^6 generations since the most recent common ancestor in *L. cidri*. In the Australian lineage, we obtained an
390 estimation of 1.5×10^5 generations since the most recent common ancestor of the Australian strains and the
391 reference CBS2950 strain. In contrast to this, when we analyzed the number of generations of the AL most
392 recent common ancestor, we obtain 23×10^6 generations. Considering that cell division of yeast in the wild
393 ranges from 1 to 8 generations per day (Fay and Benavides, 2005), the divergence time of the Australia and
394 South America lineage correspond to approximately 64-8 KYA (Fig 3b), dating back to the late Pleistocene.
395 On the other hand, the divergence of the Australian strains and the CBS2950 strain is very recent and
396 corresponds to approximately 405-51 YA (Fig 3b). Finally, the coalescence time estimated with population
397 genomic data of the AL lineage would correspond to approximately 62-7.8 KYA.

398

399 **Phenotypic diversity across *L. cidri* populations**

400 To estimate the impact of genetic diversity upon phenotypic diversity, we assessed the phenotypic diversity of
401 *L. cidri* in the complete set of 56 isolates. For this, we performed microcultures and estimated growth rates
402 (μ_{max} , corresponding to the maximum specific growth rate) from curves obtained under 12 conditions
403 comprising different carbon sources, stresses, and environmental parameters. A clustered heat map of the
404 phenotypic correlations between strains showed a separation of the isolates according to their geographic
405 origin, either from SoAm or from Aus, except for three Australian isolates that clustered together with strains
406 from Patagonia (Fig. 4a, Table S11). In this sense, hierarchical clustering of the phenotypic data distributed
407 the strains across four main clades (Fig. 4a). Cluster A contained two isolates (one from SoAm and one from
408 Aus), showing the lowest growth rates for most conditions (Fig. 4a). The different SoAm isolates were
409 distributed in Clusters B and C. Cluster B showed the greatest growth rates compared to the other three
410 clusters, while Cluster C showed intermediate values and exhibited low growth rates in fructose and ethanol.
411 Interestingly, all isolates in Cluster C belong to localities close to the Pacific Ocean (low altitude, Valdivian
412 Coastal Reserve, and Chiloé National Park) (Fig. 4a). Most strains from Australia were distributed in Cluster
413 D ($n=21$). However, three Australian strains grouped in cluster B, likely because of their greater growth rates
414 using maltose and fructose as carbon sources. The main difference between clusters is the ability of the strains

415 to grow in ethanol-containing media. In this case, we observed significant differences when we compared the
 416 growth rate between all clusters (p -value < 0.001, t-test). For example, in Clusters B and D we found isolates
 417 with the highest growth rates under 6% ethanol. Cluster B contains SoAm isolates from high altitude localities
 418 together with strain CBS2950 and three strains of the Aus lineage. On the other hand, Cluster D consists
 419 entirely of Aus lineage isolates.



420

421 **Fig. 4. Phenotypic diversity across *L. cidri* populations.** (a) Heat map depicting the phenotypic diversity in
 422 *L. cidri* obtained from growth parameter using 12 different microculture conditions. Strains are grouped by
 423 hierarchical clustering; colors indicate the lineages (Australia (Aus) and South America (SoAm)), locality
 424 (Altos de Lircay National Park (AL), Villarrica National Park (VI), Huilo-Huilo Biological Reserve (HH),
 425 Valdivian Coastal Reserve (VCR), Osorno Volcano (OV), Chiloé National Park (CH), Coyhaique National
 426 Reserve (CY), Central Plateau of Tasmania (CP), Australia, and France (FR)) and isolation altitude (high
 427 altitude (HA) and low altitude (LA)). The heat map was elaborated based on the z-score of μ_{max} (h⁻¹). (b)
 428 Fitness (μ_{max} h⁻¹) variation in 6% v/v ethanol. Asterisks indicate statistically-significant differences
 429 according to the t-test (* p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001).
 430

431 We observed a clear separation between the SoAm and Aus populations (Fig. S4) using PCA, allowing us to
 432 further dissect the phenotypic data (Fig. S4). When analyzing the distribution of each isolate, we also
 433 observed a separation based on locality (Fig. S4b). Nevertheless, the SoAm populations exhibited the greatest

434 growth rates across the different conditions compared to the Aus isolates (p -value < 0.001 , t-test) (Fig. S5a).
435 The SoAm populations exhibited phenotypic differences based on longitude. We found two phenotypic
436 groups, termed high and low altitude locations (HA and LA, respectively), depending on whether they were
437 collected near the Andes or the Pacific Coast, respectively (Fig. 4a). In general, strains isolated from HA
438 locations show greater growth rates than those isolated from LA locations (p -value < 0.001 , t-test) (Fig. S4b).
439 Interestingly, LA strains showed a higher growth rate than HA only in 2% maltose (p -value < 0.05 , t-test)
440 (Fig. S5c). The longitude comparison showed significant differences for 8 of the 12 conditions evaluated (p -
441 value < 0.05 , t-test) (Fig. S5c), with the most notable differences (p -value < 0.001 , t-test) in ethanol (6 and
442 8% v/v), glucose (20% w/v), sorbitol (20% w/v) and NaCl (1.25 mM) (Fig. S5c).

443 **DISCUSSION**

444 Large-scale population genomics studies performed on *Saccharomyces* spp. allowed us to understand
445 yeast dynamics in nature and their association with anthropogenic environments (Peter and Schacherer, 2016).
446 In this sense, the analysis of complete genomes and the use of molecular markers aided the understanding of
447 the natural genetic diversity in *Saccharomyces* populations and highlights how they adapt to the environments
448 which they inhabit (Almeida et al., 2015; Almeida et al., 2014; Bendixsen et al., 2021; Parts et al., 2021; Peter
449 and Schacherer, 2016). In the last decade, advances in sequencing technologies have made it possible to
450 extrapolate this type of study to non-model species that have long been overlooked (Bansal and Boucher,
451 2019; Dujon and Louis, 2017). *L. cidri*, a non-model species, is an attractive model to study since it's a
452 Crabtree-positive yeast located phylogenetically before the whole-genome duplication (Porter et al., 2019b;
453 Vakirlis et al., 2016), and which exhibits substantial biotechnological potential for fermentation (Varela et al.,
454 2020; Villarreal et al., 2021). In contrast to *Saccharomyces* spp., which are mostly diploid in nature (Almeida
455 et al., 2014; Nespolo et al., 2020; Peter et al., 2018), *L. cidri* isolates are haploids, which might be indicative
456 of a different reproductive cycle. It has been reported that even with identical genomes, ploidy itself is known
457 to have different effects on yeasts. Therefore, understanding how ploidy affects the ecology and evolution of
458 organisms has long been a topic of interest (Gerstein et al., 2011; Gerstein and Otto, 2009). Haploid yeasts
459 have a competitive advantage over diploid yeasts. For example, one of the main differences between the two
460 ploidy levels is cell size, where differences in volume have been shown to directly affect relative fitness in
461 some environments. Under nutrient stress, where a limited concentration of nutrients diffuses across the cell

462 membrane, haploid organisms (which have a smaller size than diploids) would have a greater advantage,
463 likely favoring their survival under extreme environmental conditions (Gerstein et al., 2011).

464 Our results demonstrate that *L. cidri* is widely-distributed in Patagonia. In fact, *L. cidri* is found
465 between latitudes 35 °S and 45 °S in Patagonia, which in general are considered non-Mediterranean (cold
466 regions). However, *L. cidri* was not found in extreme southern regions such as Tierra del Fuego (54 °S),
467 probably due to southern Patagonia's extreme cold climatic conditions, where the temperature is below
468 freezing for a large fraction of the year (Ponce JF and M., 2014). Similar observations were reported in other
469 organisms such as *Dromiciops* species (Quintero-Galvis et al., 2022), and *Petaurus* species (Malekian et al.,
470 2010) where geography and climate conditions determined the organism's distribution. *L. cidri* samples
471 obtained in South America were collected from tree species belonging to the genera *Nothofagus* and
472 *Araucaria* in National Parks from Chile, covering approximately 1,000 km of territory. Then, the
473 biogeographical history of the South American population of *L. cidri* correlates well with the history and
474 distribution of the *Nothofagus* forest, comparable to reports for other species isolated in Patagonia
475 (*Saccharomyces* and non-*Saccharomyces*) (Nespolo et al., 2020). Our results agree with previous findings,
476 where it has been demonstrated that *L. cidri* exhibits significant host preference, being more frequently
477 isolated from *N. dombeyi* in Patagonia (Nespolo et al., 2020; Villarreal et al., 2021). In this way, this study,
478 together with previous studies in *Saccharomyces* yeasts from forests ecosystems (Almeida et al., 2014; Cadez
479 et al., 2021; Langdon et al., 2020; Libkind et al., 2011; Nespolo et al., 2020) demonstrate the impact of the
480 *Nothofagus* biome (Woodward et al., 2004) across the southern hemisphere on the current distribution of the
481 *Lachancea* genus throughout Patagonia.

482 The aim of the current study was to compare the genomic and phenotypic landscape across *L. cidri*
483 isolates, regardless of how these cultures were isolated. The phylogenomic analysis of wild *L. cidri* (SoAm)
484 strains demonstrates a SoAm genetic diversity ($\pi = 0,000278$) comparable to that of other *Saccharomyces*
485 natural populations, such as *S. eubayanus* (PB-2 and PB-3, $\pi = 0,0009$ and, $\pi = 0,0005$ respectively) and *S.*
486 *cerevisiae* lineages, such as Alpechin, Ecuadorean, North America oak, and Sake clades ($\pi = 0,00095$;
487 $0,00015$; $0,00010$; and $0,00080$, respectively) (Peter et al., 2018). In contrast, we observed a lower nucleotide
488 diversity in the SoAm populations compared to other *S. cerevisiae* lineages from Taiwan ($\pi = 0,005$) or China
489 ($\pi = 0,006$) (Lee et al., 2022). These differences are likely explained by the number of isolation sites where

490 yeasts identified as *L. cidri* were obtained, and further sampling efforts could improve our understanding on
491 the genetic diversity in *L. cidri*. Since the strains from Australia were collected in a smaller territory and
492 showed relatively low genetic variation (Varela et al., 2020), we were not able to assess comparable
493 parameters to South America. In this sense, sampling in Patagonia a broader geographic distribution with
494 varied seasonal and spatial variation generated an enormous genetic differentiation and nucleotide diversity
495 between localities. Therefore, our population structure analysis, developed with different strategies, generated
496 a consistent phylogenetic picture: two populations depending on geographic origin (SoAm and Aus), with the
497 Australian isolates forming a tight cluster with the reference strain from France (CBS2950). The utilization of
498 STRUCTURE, together with fineSTRUCTURE and PCA are appropriate for this type of analysis, since
499 altogether, these analyses can provide further insights into the population history within a species (Stift et al.,
500 2019). The Australian population showed a low genetic diversity, and a low genetic differentiation to the
501 reference strain, suggesting a recent exchange between Australia and Europe, likely coalescing between 405
502 to 51 years ago. In contrast, the variability found within the South American population is in accordance with
503 the different localities from which the isolates were obtained. The different environmental conditions and the
504 extensive distribution of the *Nothofagus* forest in Patagonia likely facilitated the presence of a unique genetic
505 diversity (Cadez et al., 2019; Cubillos et al., 2019; Nespolo et al., 2020; Villarreal et al., 2021), not only in
506 yeasts, but also in higher mammals like *Dromiciops gliroides* and *D. bozinovici* (a recently proposed species
507 with higher nucleotide diversity in South America) (Quintero-Galvis et al., 2022). In the same way, the
508 phenotypic data corroborates the large genetic diversity observed in Patagonia. Although it has been reported
509 that genetic diversity is not always a reflection of diversity at the physiological level (Banilas et al., 2016;
510 Pfliegler et al., 2014), we were able to demonstrate how greater diversity at the genetic level is determinant
511 when performing phenotypic analyses under natural and stressful conditions.

512 The divergence between Australian and South American strains is interesting and deserves attention.
513 The split could have originated in the Late Pleistocene, near the last glacial maximum (LGM), an event that
514 has modified the abundance and distribution of the biota in South America (Sérsic et al., 2011). The absence
515 of fossil records for yeasts and the lack of precise estimates on the number of generations complicates the
516 interpretation, but the geographic origin of our samples, the high association with native forests with a well-
517 documented phylogeographic history (Acosta et al., 2014; Hinojosa et al., 2016), and the genetic differences

518 found in *L. cidri* from South America permit to delineate some hypotheses. The LGM during the Late
519 Pleistocene (~35,000 years ago)(Davies et al., 2020) resulted in an ice cover that expanded through the south
520 of South America from latitude 53°S to 38°S, leaving only a few areas free of ice, such as Altos de Lircay
521 (Hewitt, 2000; Hinojosa, 1997). For example, despite the short distance between HH and VCR (~300 km),
522 these two sites exhibit a significant genetic distance between *L. cidri* populations, likely due the ice cover
523 separation. Thus, most areas including Coyhaique were covered by the ice, and present-day plants and
524 animals are young populations that colonized the area either from Argentina or from coastal refuges, during
525 the last ~10,000 years. This large isolation barrier would explain the divergence observed between the South
526 American *L. cidri* populations, unlike the French and Australian strains that likely migrated because of human
527 movements.

528 In summary, our results demonstrate the presence of two genetically-different populations in *L. cidri*
529 and a high phylogeographic structure among the localities of South America. The species delimitation method
530 and genetic differences between Altos de Lircay populations and southern populations in Patagonia suggest
531 that the former represents a different lineage. This is a striking result since the unique conditions of Patagonia
532 during the Late Pleistocene could have contributed to the differentiation of *L. cidri* populations in South
533 America. Furthermore, we propose that there was a recent exchange between Australia and France due to the
534 low genetic differentiation between strains from both regions. In conclusion, this work provides valuable
535 insight into the genetic and phenotypic diversity of *L. cidri*, contributing to a better understanding of the
536 phylogeography, population structure, ecology, and divergence time of this underexplored yeast species, with
537 remarkable biotechnological potential.

538

539 **ACKNOWLEDGMENTS**

540 FC acknowledges the Comisión Nacional de Investigación Científica y Tecnológica CONICYT FONDECYT
541 [1180161] and ANID - Programa Iniciativa Científica Milenio - ICN17_022 and NCN2021_050. PV is
542 supported by ANID FONDECYT POSTDOCTORADO [grant 3200575]. RN is supported by FIC
543 ‘Transferencia Levaduras Nativas para Cerveza Artesanal’ and FONDECYT grant [1180917]. FC, RN, PV,
544 TP, SO, and GF were funded by PROGRAMA DE COOPERACIÓN CIENTÍFICA ECOS-CONICYT

545 ECOS180003. We thank Dr. Vladimir Jiranek from the University of Adelaide and the AWRI Culture
546 Collection for access to *L. cidri* isolates, Michael Handford (Universidad de Chile) for language support and
547 Fundación Huilo Huilo for the assistance during the fieldwork. We also acknowledge Fundación Ciencia &
548 Vida for providing infrastructure, laboratory space and equipment for experiments.

549

550 **Author Contributions**

551 **Conceptualization:** P.V., C.V., F.A.C. **Formal analysis:** P.V. **Investigation:** P.V., C.A.V., S.O., N.A., T.P.,
552 J.F.Q-G, C.V. **Writing – original draft:** P.V., F.A.C. **Writing – review & editing:** P.V., R.F.N., C.V., G.F.,
553 J.F.Q-G, F.A.C.

554 **Data accessibility**

555 All new *fastq* sequences were deposited in the National Center for Biotechnology Information (NCBI).
556 Sequence Read Archive under the BioProject accession number PRJNA787570
557 (www.ncbi.nlm.nih.gov/bioproject/787570).

558

559 **Conflicts of interest**

560 The authors declare that there are no conflicts of interest.

561

562 **References**

563

- 564 Acosta, M.C., Mathiasen, P., and Premoli, A.C. (2014). Retracing the evolutionary history of
565 Nothofagus in its geo-climatic context: new developments in the emerging field of phylogeology.
566 *Geobiology* 12, 497-510.
- 567 Agier, N., Delmas, S., Zhang, Q., Fleiss, A., Jaszczyszyn, Y., van Dijk, E., Thermes, C., Weigt, M.,
568 Cosentino-Lagomarsino, M., and Fischer, G. (2018). The evolution of the temporal program of
569 genome replication. *Nat Commun* 9, 2199.
- 570 Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J.L., Serra, M.,
571 Dequin, S., Couloux, A., *et al.* (2015). A population genomics insight into the Mediterranean origins
572 of wine yeast domestication. *Mol Ecol* 24, 5412-5427.
- 573 Almeida, P., Goncalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarede, I., Albertin,
574 W., Durrrens, P., Sherman, D.J., Marullo, P., *et al.* (2014). A Gondwanan imprint on global diversity
575 and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat Commun* 5, 4044.
- 576 Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available
577 online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed December 5, 2019).
- 578 Banilas, G., Sgouros, G., and Nisiotou, A. (2016). Development of microsatellite markers for
579 *Lachancea thermotolerans* typing and population structure of wine-associated isolates. *Microbiol*
580 *Res* 193, 1-10.

581 Bansal, V., and Boucher, C. (2019). Sequencing Technologies and Analyses: Where Have We Been
582 and Where Are We Going? *iScience* 18, 37-41.

583 Bellut, K., Krogerus, K., and Arendt, E.K. (2020). *Lachancea fermentati* Strains Isolated From
584 Kombucha: Fundamental Insights, and Practical Application in Low Alcohol Beer Brewing. *Front*
585 *Microbiol* 11, 764.

586 Bendixsen, D.P., Gettle, N., Gilchrist, C., Zhang, Z., and Stelkens, R. (2021). Genomic Evidence of an
587 Ancient East Asian Divergence Event in Wild *Saccharomyces cerevisiae*. *Genome Biol Evol* 13.

588 Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A.,
589 Heled, J., Jones, G., Kuhnert, D., De Maio, N., *et al.* (2019). BEAST 2.5: An advanced software
590 platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 15, e1006650.

591 Bouckaert, R.R. (2010). DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* 26,
592 1372-1373.

593 Brickwedde, A., Brouwers, N., van den Broek, M., Gallego Murillo, J.S., Fraiture, J.L., Pronk, J.T.,
594 and Daran, J.G. (2018). Structural, Physiological and Regulatory Analysis of Maltose Transporter
595 Genes in *Saccharomyces eubayanus* CBS 12357(T). *Front Microbiol* 9, 1786.

596 Brion, C., Legrand, S., Peter, J., Caradec, C., Pflieger, D., Hou, J., Friedrich, A., Llorente, B., and
597 Schacherer, J. (2017). Variation of the meiotic recombination landscape and properties over a
598 broad evolutionary distance in yeasts. *PLoS Genet* 13, e1006917.

599 Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data
600 inference for whole-genome association studies by use of localized haplotype clustering. *Am J*
601 *Hum Genet* 81, 1084-1097.

602 Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., and RoyChoudhury, A. (2012). Inferring
603 species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent
604 analysis. *Mol Biol Evol* 29, 1917-1932.

605 Cadez, N., Bellora, N., Ulloa, R., Hittinger, C.T., and Libkind, D. (2019). Genomic content of a novel
606 yeast species *Hanseniaspora gamundiae* sp. nov. from fungal stromata (Cyttaria) associated with a
607 unique fermented beverage in Andean Patagonia, Argentina. *PLoS One* 14, e0210792.

608 Cadez, N., Bellora, N., Ulloa, R., Tome, M., Petkovic, H., Groenewald, M., Hittinger, C.T., and
609 Libkind, D. (2021). *Hanseniaspora smithiae* sp. nov., a Novel Apiculate Yeast Species From
610 Patagonian Forests That Lacks the Typical Genomic Domestication Signatures for Fermentative
611 Environments. *Front Microbiol* 12, 679894.

612 Chen, J., Shen, W., Xu, H., Li, Y., and Luo, T. (2019). The Composition of Nitrogen-Fixing
613 Microorganisms Correlates With Soil Nitrogen Content During Reforestation: A Comparison
614 Between Legume and Non-legume Plantations. *Front Microbiol* 10, 508.

615 Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor.
616 *Bioinformatics* 34, i884-i890.

617 Cubillos, F.A., Gibson, B., Grijalva-Vallejos, N., Krogerus, K., and Nikulin, J. (2019). Bioprospecting
618 for brewers: Exploiting natural diversity for naturally diverse beers. *Yeast* 36, 383-398.

619 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter,
620 G., Marth, G.T., Sherry, S.T., *et al.* (2011). The variant call format and VCFtools. *Bioinformatics*
621 (Oxford, England) 27, 2156-2158.

622 Davies, B.J., Darvill, C.M., Lovell, H., Bendle, J.M., Dowdeswell, J.A., Fabel, D., Garcia, J.L., Geiger,
623 A., Glasser, N.F., Gheorghiu, D.M., *et al.* (2020). The evolution of the Patagonian Ice Sheet from 35
624 ka to the present day (PATICE). *Earth-Sci Rev* 204.

625 DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del
626 Angel, G., Rivas, M.A., Hanna, M., *et al.* (2011). A framework for variation discovery and
627 genotyping using next-generation DNA sequencing data. *Nat Genet* 43, 491-498.

628 Dujon, B.A., and Louis, E.J. (2017). Genome Diversity and Evolution in the Budding Yeasts
629 (*Saccharomycotina*). *Genetics* 206, 717-750.

630 Esteve-Zarzoso, B., Peris-Toran, M.J., Garcia-Maiquez, E., Uruburu, F., and Querol, A. (2001). Yeast
631 population dynamics during the fermentation and biological aging of sherry wines. *Appl Environ*
632 *Microbiol* 67, 2056-2061.

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

635 Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T.G., Nordborg, M., and Lynch, M. (2015). The
636 Spontaneous Mutation Rate in the Fission Yeast *Schizosaccharomyces pombe*. *Genetics* 201, 737-
637 744.

638 Fay, J.C., and Benavides, J.A. (2005). Evidence for domesticated and wild populations of
639 *Saccharomyces cerevisiae*. *PLoS Genet* 1, 66-71.

640 Fijarczyk, A., Henault, M., Marsit, S., Charron, G., and Landry, C.R. (2021). Heterogeneous Mutation
641 Rates and Spectra in Yeast Hybrids. *Genome Biol Evol* 13.

642 Fournier, T., Gounot, J.S., Freel, K., Cruaud, C., Lemainque, A., Aury, J.M., Wincker, P., Schacherer,
643 J., and Friedrich, A. (2017). High-Quality de Novo Genome Assembly of the *Dekkera bruxellensis*
644 Yeast Using Nanopore MinION Sequencing. *G3 (Bethesda)* 7, 3243-3250.

645 Friedrich, A., Jung, P., Reisser, C., Fischer, G., and Schacherer, J. (2015). Population genomics
646 reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol Biol Evol* 32, 184-
647 192.

648 Friedrich, A., Jung, P.P., Hou, J., Neueglise, C., and Schacherer, J. (2012). Comparative
649 mitochondrial genomics within and among yeast species of the *Lachancea* genus. *PLoS One* 7,
650 e47834.

651 Fujita, S.I., Senda, Y., Nakaguchi, S., and Hashimoto, T. (2001). Multiplex PCR using internal
652 transcribed spacer 1 and 2 regions for rapid detection and identification of yeast strains. *J Clin*
653 *Microbiol* 39, 3617-3622.

654 Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A.,
655 Roncoroni, M., Voordeckers, K., Miraglia, L., *et al.* (2016). Domestication and Divergence of
656 *Saccharomyces cerevisiae* Beer Yeasts. *Cell* 166, 1397-1410 e1316.

657 García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J.,
658 Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment
659 data. *Bioinformatics* 28, 2678-2679.

660 Gerstein, A.C., Cleathero, L.A., Mandegar, M.A., and Otto, S.P. (2011). Haploids adapt faster than
661 diploids across a range of environments. *J Evol Biol* 24, 531-540.

662 Gerstein, A.C., and Otto, S.P. (2009). Ploidy and the causes of genomic evolution. *J Hered* 100, 571-
663 581.

664 Goncalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Goncalves,
665 P., and Sampaio, J.P. (2016). Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts
666 and Wine Yeasts. *Curr Biol* 26, 2750-2761.

667 Gonzalez, S.S., Barrio, E., and Querol, A. (2007). Molecular identification and characterization of
668 wine yeasts isolated from Tenerife (Canary Island, Spain). *J Appl Microbiol* 102, 1018-1025.

669 Hall, B.G., Acar, H., Nandipati, A., and Barlow, M. (2014). Growth rates made easy. *Mol Biol Evol*
670 31, 232-238.

671 Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405, 907-913.

672 Hinojosa, L.F., Gaxiola, A., Perez, M.F., Carvajal, F., Campano, M.F., Quattrocchio, M., Nishida, H.,
673 Uemura, K., Yabe, A., Bustamante, R., *et al.* (2016). Non-congruent fossil and phylogenetic
674 evidence on the evolution of climatic niche in the Gondwana genus *Nothofagus*. *Journal of*
675 *Biogeography* 43, 555-567.

676 Hinojosa, L.F.V., C. (1997). Historia de los bosques del sur de Sudamérica, I: antecedentes
677 paleobotánicos, geológicos y climáticos del Terciario del cono sur de América. *Rev Chil Hist Nat* 70,
678 225-239.

679 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2:
680 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35, 518-522.

681 Hranilovic, A., Bely, M., Masneuf-Pomarede, I., Jiranek, V., and Albertin, W. (2017). The evolution
682 of *Lachancea thermotolerans* is driven by geographical determination, anthropisation and flux
683 between different ecosystems. *PLoS One* 12, e0184652.

684 Hranilovic, A., Gambetta, J.M., Schmidtke, L., Boss, P.K., Grbin, P.R., Masneuf-Pomarede, I., Bely,
685 M., Albertin, W., and Jiranek, V. (2018). Oenological traits of *Lachancea thermotolerans* show signs
686 of domestication and allopatric differentiation. *Sci Rep* 8, 14812.

687 Kessi-Perez, E.I., Araos, S., Garcia, V., Salinas, F., Abarca, V., Larrondo, L.F., Martinez, C., and
688 Cubillos, F.A. (2016). RIM15 antagonistic pleiotropy is responsible for differences in fermentation
689 and stress response kinetics in budding yeast. *FEMS Yeast Res* 16.

690 Kodama, K., and Kyono, T. (1974). Ascosporeogenous yeasts isolated from tree exudates in Japan.
691 v. 52.

692 Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a
693 program for identifying clustering modes and packaging population structure inferences across K.
694 *Mol Ecol Resour* 15, 1179-1191.

695 Kurtzman, C. (2003). Phylogenetic circumscription of , and other members of the
696 Saccharomycetaceae, and the proposal of the new genera , , and. *FEMS Yeast Research* 4, 233-
697 245.

698 Lachance, M.-A., and Kurtzman, C.P. (2011). *Lachancea*. In *The Yeasts*, pp. 511-519.

699 Langdon, Q.K., Peris, D., Eizaguirre, J.I., Ofulente, D.A., Buh, K.V., Sylvester, K., Jarzyna, M.,
700 Rodriguez, M.E., Lopes, C.A., Libkind, D., *et al.* (2020). Postglacial migration shaped the genomic
701 diversity and global distribution of the wild ancestor of lager-brewing hybrids. *PLoS Genet* 16,
702 e1008680.

703 Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure
704 using dense haplotype data. *PLoS Genet* 8, e1002453.

705 Leache, A.D., Fujita, M.K., Minin, V.N., and Bouckaert, R.R. (2014). Species delimitation using
706 genome-wide SNP data. *Syst Biol* 63, 534-542.

707 Lee, C.F., Yao, C.H., Liu, Y.R., Hsieh, C.W., and Young, S.S. (2009). *Lachancea dasiensis* sp. nov., an
708 ascosporeogenous yeast isolated from soil and leaves in Taiwan. *Int J Syst Evol Microbiol* 59, 1818-
709 1822.

710 Lee, T.J., Liu, Y.C., Liu, W.A., Lin, Y.F., Lee, H.H., Ke, H.M., Huang, J.P., Lu, M.J., Hsieh, C.L., Chung,
711 K.F., *et al.* (2022). Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and
712 evolution of predomesticated lineages. *Genome Res* 32, 864-877.

713 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
714 1303.

715 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and
716 Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-
717 2079.

718 Li, Y.L., and Liu, J.X. (2018). StructureSelector: A web-based software to select and visualize the
719 optimal number of clusters using multiple methods. *Mol Ecol Resour* 18, 176-177.

720 Libkind, D., Hittinger, C.T., Valerio, E., Goncalves, C., Dover, J., Johnston, M., Goncalves, P., and
721 Sampaio, J.P. (2011). Microbe domestication and the identification of the wild genetic stock of
722 lager-brewing yeast. *Proc Natl Acad Sci U S A* 108, 14539-14544.

723 Lindenbaum, P., and Redon, R. (2018). bioalcaide, samjs and vcfilterjs: object-oriented formatters
724 and filters for bioinformatics files. *Bioinformatics* 34, 1224-1225.

725 Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N.,
726 Burt, A., Koufopanou, V., *et al.* (2009). Population genomics of domestic and wild yeasts. *Nature*
727 458, 337-341.

728 Liu, W., Luo, Z., Wang, Y., Pham, N.T., Tuck, L., Pérez-Pi, I., Liu, L., Shen, Y., French, C., Auer, M., *et*
729 *al.* (2018). Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome
730 SCRaMBLE-in methods. *Nature Communications* 9, 1936.

731 Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C.R., Dopman, E.B., Dickinson, W.J., Okamoto,
732 K., Kulkarni, S., Hartl, D.L., *et al.* (2008). A genome-wide view of the spectrum of spontaneous
733 mutations in yeast. *Proc Natl Acad Sci U S A* 105, 9272-9277.

734 Magalhães, K.T., de Melo Pereira, G.V., Campos, C.R., Dragone, G., and Schwan, R.F. (2011).
735 Brazilian kefir: structure, microbial communities and chemical composition. *Braz J Microbiol* 42,
736 693-702.

737 Malekian, M., Cooper, S.J., Norman, J.A., Christidis, L., and Carthew, S.M. (2010). Molecular
738 systematics and evolutionary origins of the genus *Petaurus* (Marsupialia: Petauridae) in Australia
739 and New Guinea. *Mol Phylogenet Evol* 54, 122-135.

740 Marsh, A.J., O'Sullivan, O., Hill, C., Ross, R.P., and Cotter, P.D. (2014). Sequence-based analysis of
741 the bacterial and fungal compositions of multiple kombucha (tea fungus) samples. *Food Microbiol*
742 38, 171-178.

743 Mesquita, V.A., Magalhães-Guedes, K., Silva, C., and Schwan, R. (2013). The molecular
744 phylogenetic diversity of bacteria and fungi associated with the cerrado soil from different regions
745 of Minas Gerais, Brazil. *Int J Microbiol Res* 4, 119-131.

746 Mirza, B.S., Potisap, C., Nusslein, K., Bohannan, B.J., and Rodrigues, J.L. (2014). Response of free-
747 living nitrogen-fixing microorganisms to land use change in the Amazon rainforest. *Appl Environ*
748 *Microbiol* 80, 281-288.

749 Nespolo, R.F., Villarroel, C.A., Oporto, C.I., Tapia, S.M., Vega-Macaya, F., Urbina, K., De Chiara, M.,
750 Mozzachiodi, S., Mikhalev, E., Thompson, D., *et al.* (2020). An Out-of-Patagonia migration explains
751 the worldwide diversity and distribution of *Saccharomyces eubayanus* lineages. *PLoS Genet* 16,
752 e1008777.

753 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective
754 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32, 268-274.

755 Nova, M.X.V., Schuler, A.R.P., Brasileiro, B.T.R.V., and Morais, M.A. (2009). Yeast species involved
756 in artisanal cachaça fermentation in three stills with different technological levels in Pernambuco,
757 Brazil. *Food Microbiology* 26, 460-466.

758 Ortiz, E.M. (2019). vcf2phylip v2.0: convert a VCF matrix into several matrix formats for
759 phylogenetic analysis. (Zenodo. <https://doi.org/10.5281/zenodo.2540861>).

760 Parts, L., Batte, A., Lopes, M., Yuen, M.W., Laver, M., San Luis, B.J., Yue, J.X., Pons, C., Eray, E., Aloy,
761 P., *et al.* (2021). Natural variants suppress mutations in hundreds of essential genes. *Mol Syst Biol*
762 17, e10138.

763 Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet*
764 2, e190.

765 Peakall, R., and Smouse, P.E. (2012). GenA1Ex 6.5: genetic analysis in Excel. Population genetic
766 software for teaching and research--an update. *Bioinformatics* 28, 2537-2539.

767 Pembleton, L.W., Cogan, N.O., and Forster, J.W. (2013). StAMPP: an R package for calculation of
768 genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour* 13, 946-
769 952.

770 Pereira, L.F., Costa, C.R.L., Brasileiro, B., and de Moraes, M.A. (2011). *Lachancea mirantina* sp. nov.,
771 an ascomycetous yeast isolated from the cachaca fermentation process. *Int J Syst Evol Microbiol*
772 *61*, 989-992.

773 Peter, J., De Chiara, M., Friedrich, A., Yue, J.X., Pflieger, D., Bergstrom, A., Sigwalt, A., Barre, B.,
774 Freel, K., Llored, A., *et al.* (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae*
775 isolates. *Nature* *556*, 339-344.

776 Peter, J., and Schacherer, J. (2016). Population genomics of yeasts: towards a comprehensive view
777 across a broad evolutionary scale. *Yeast* *33*, 73-81.

778 Pfeifer, B., Wittelsburger, U., Ramos-Onsins, S.E., and Lercher, M.J. (2014). PopGenome: an
779 efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* *31*, 1929-1936.

780 Pfliegler, W.P., Horvath, E., Kallai, Z., and Sipiczki, M. (2014). Diversity of *Candida zemplinina*
781 isolates inferred from RAPD, micro/minisatellite and physiological analysis. *Microbiol Res* *169*, 402-
782 410.

783 Phaff, H.J., Miller, M.W., and Shifrine, M. (1956). The taxonomy of yeasts isolated from *Drosophila*
784 in the Yosemite region of California. *Antonie Van Leeuwenhoek* *22*, 145-161.

785 Ponce JF, and M., F. (2014). Climatic and Environmental History of Isla de los Estados, Argentina.

786 Porter, T.J., Divol, B., and Setati, M.E. (2019a). Investigating the biochemical and fermentation
787 attributes of *Lachancea* species and strains: Deciphering the potential contribution to wine
788 chemical composition. *Int J Food Microbiol* *290*, 273-287.

789 Porter, T.J., Divol, B., and Setati, M.E. (2019b). *Lachancea* yeast species: Origin, biochemical
790 characteristics and oenological significance. *Food Res Int* *119*, 378-389.

791 Quintero-Galvis, J.F., Saenz-Agudelo, P., Amico, G.C., Vazquez, S., Shafer, A.B.A., and Nespolo, R.F.
792 (2022). Genomic diversity and demographic history of the *Dromiciops* genus (Marsupialia:
793 Microbiotheriidae). *Mol Phylogenet Evol* *168*, 107405.

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

796 Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior
797 Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* *67*, 901-904.

798 Romano, P., and Suzzi, G. (1993a). Higher alcohol and acetoin production by *Zygosaccharomyces*
799 wine yeasts. *Journal of Applied Bacteriology* *75*, 541-545.

800 Romano, P., and Suzzi, G. (1993b). Potential use for *Zygosaccharomyces* species in winemaking.
801 *Journal of Wine Research* *4*, 87-94.

802 Ruderfer, D.M., Pratt, S.C., Seidel, H.S., and Kruglyak, L. (2006). Population genomic analysis of
803 outcrossing and recombination in yeast. *Nat Genet* *38*, 1077-1081.

804 Sampaio, J.P., and Goncalves, P. (2008). Natural populations of *Saccharomyces kudriavzevii* in
805 Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl*
806 *Environ Microbiol* *74*, 2144-2152.

807 Sérsic, A.N., Cosacov, A., Cocucci, A.A., Johnson, L.A., Pozner, R., Avila, L.J., Sites, J.W., Jr., and
808 Morando, M. (2011). Emerging phylogeographical patterns of plants and terrestrial vertebrates
809 from Patagonia. *Biological Journal of the Linnean Society* *103*, 475-494.

810 Stift, M., Kolar, F., and Meirns, P.G. (2019). STRUCTURE is more robust than other clustering
811 methods in simulated mixed-ploidy populations. *Heredity (Edinb)* *123*, 429-441.

812 Team, R.D.C. (2008). R: A Language and Environment for Statistical Computing. R Foundation for
813 Statistical Computing.

814 Tzanetakis, N., Hatzikamari, M., and Litopoulou-Tzanetaki, E. (1998). Yeasts of the surface
815 microflora of Feta cheese.

816 Vakirlis, N., Hebert, A.S., Ofulente, D.A., Achaz, G., Hittinger, C.T., Fischer, G., Coon, J.J., and
817 Lafontaine, I. (2018). A Molecular Portrait of De Novo Genes in Yeasts. *Mol Biol Evol* *35*, 631-645.

818 Vakirlis, N., Sarilar, V., Drillon, G., Fleiss, A., Agier, N., Meyniel, J.P., Blanpain, L., Carbone, A.,
819 Devillers, H., Dubois, K., *et al.* (2016). Reconstruction of ancestral chromosome architecture and
820 gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Res* 26,
821 918-932.

822 Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A.,
823 Jordan, T., Shakir, K., Roazen, D., Thibault, J., *et al.* (2013a). From FastQ data to high confidence
824 variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43,
825 11.10.11-11.10.33.

826 Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A.,
827 Jordan, T., Shakir, K., Roazen, D., Thibault, J., *et al.* (2013b). From FastQ data to high confidence
828 variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in*
829 *bioinformatics* 43, 11.10.11-33.

830 Varela, C., Sundstrom, J., Cuijvers, K., Jiranek, V., and Borneman, A. (2020). Discovering the
831 indigenous microbial communities associated with the natural fermentation of sap from the cider
832 gum *Eucalyptus gunnii*. *Sci Rep* 10, 14716.

833 Villarreal, P., Quintrel, P.A., Olivares-Muñoz, S., Ruiz, J.I., Nespolo, R.F., and Cubillos, F.A. (2021).
834 Identification of new ethanol-tolerant yeast strains with fermentation potential from central
835 Patagonia. *Yeast*.

836 Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population
837 Structure. *Evolution* 38, 1358-1370.

838 Wojtatowicz, M., Chrzanowska, J., Juszczak, P., Skiba, A., and Gdula, A. (2001). Identification and
839 biochemical characteristics of yeast microflora of Rokpol cheese. *International Journal of Food*
840 *Microbiology* 69, 135-140.

841 Woodward, F.I., Lomas, M.R., and Kelly, C.K. (2004). Global climate and the distribution of plant
842 biomes. *Philos Trans R Soc Lond Ser B-Biol Sci* 359, 1465-1476.

843

844 **Figures**

845 **Figure 1. Geographic distribution and phylogeny of *L. cidri* isolates.** (a) World map depicting the different
846 locations where *L. cidri* were isolated. In Chile, strains were obtained from seven sampling locations. (b)
847 Maximum likelihood tree depicting the genetic relationships between 56 strains using 218,266 biallelic SNPs
848 (substitution model GTR+F+ASC) and manually-rooted with *L. fermentati* as the outgroup.

849 **Figure 2. Population structure of *L. cidri*.** (a) Plot of the genomic variation distribution of 56 *L. cidri* strains
850 based on the first two components of a PCA performed using 13,275 unlinked SNPs. Each dot represents a
851 single strain. The black arrow denotes strains from Altos de Lircay National Park (AL). (b) PCA only
852 considering strains from Patagonia. The color reflects the grouping by location. Villarrica National Park (VI),
853 Huilo-Huilo Biological Reserve (HH), Valdivian Coastal Reserve (VCR), Osorno Volcano (OV), Chiloé
854 National Park (CH), and Coyhaique National Reserve (CY). (c) STRUCTURE and ADMIXTURE. An
855 optimum K = 2 group is shown in both cases. (d) Heat map of a Co-ancestry matrix obtained using
856 fineSTRUCTURE chunk counts. Each row and column represent an isolated color indicating genetic sharing
857 (yellow = low sharing, blue = high sharing). The tree shows the clusters inferred from the co-ancestry matrix.
858 (e) F_{ST} values between localities, using the same abbreviations shown in (b) as well as Central Plateau of
859 Tasmania (CP), Australia. (f) IBD (Isolation by distance) estimates considering seven localities of southern
860 Chile. The density of data points is indicated by colors.

861 **Figure 3. Species delimitation and divergence time analysis in *L. cidri* populations.** (a) DensiTrees
862 gathered in the Bayesian analyses, where nodes values correspond to the Bayesian inference (IB) result of the
863 species tree. Strains from Australia (Aus), Altos de Lircay National Park (AL), Huilo-Huilo Biological
864 Reserve (HH), and Coyhaique National Reserve (CY) are included in the analysis. The blue traces indicate

865 the most frequently occurring topology; red indicates the second most likely topology and the third in light
866 blue. **(b)** Estimation of the divergence time since the last common ancestor between groups. Time between (1)
867 Australia lineage and SoAm lineage; Time between (2) Australia lineage and CBS2950. Timeline in Million's
868 years (Ky).

869
870 **Fig. 4. Phenotypic diversity across *L. cidri* populations.** **(a)** Heat map depicting the phenotypic diversity in
871 *L. cidri* obtained from growth parameter using 12 different microculture conditions. Strains are grouped by
872 hierarchical clustering; colors indicate the lineages (Australia (Aus) and South America (SoAm)), locality
873 (Altos de Lircay National Park (AL), Villarrica National Park (VI), Huilo-Huilo Biological Reserve (HH),
874 Valdivian Coastal Reserve (VCR), Osorno Volcano (OV), Chiloé National Park (CH), Coyhaique National
875 Reserve (CY), Central Plateau of Tasmania (CP), Australia, and France (FR)) and isolation altitude (high
876 altitude (HA) and low altitude (LA)). The heat map was elaborated based on the z-score of μ_{max} (h^{-1}). **(b)**
877 Fitness (μ_{max} h^{-1}) variation in 6% v/v ethanol. Asterisks indicate statistically-significant differences
878 according to the t-test (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).
879

880 **Supplementary Material**

881 **Supplementary Figures**

882 **Fig. S1. (a)** Frequency of isolation of *L. cidri* compared to the total yeast found. **(b)** Frequency of successful
883 *L. cidri* isolation of the non-*Saccharomyces* yeast found. **(c)** n° of isolates per tree host.

884 **Fig. S2. FACS analysis of *L. cidri* isolates.** Number of cell vs propidium iodide intensity is shown. Haploid
885 (n), diploid (2n), and tetraploid (4n). *S. eubayanus* strains were used as a control.

886 **Figure S3. Genetic differentiation and diversity across *L. cidri* populations.** **(a)** *Fst* values across the
887 genome between lineages **(b)** Nucleotide diversity (π) in Chilean (SoAm) locations and the Australian (Aus)
888 population. **(c)** Nucleotide diversity (π) in each locality. Altos de Lircay National Park (AL), Villarrica
889 National Park (VI), Huilo-Huilo Biological Reserve (HH), Valdivian Coastal Reserve (VCR), Osorno
890 Volcano (OV), Chiloé National Park (CH), Coyhaique National Reserve (CY), and Central Plateau of
891 Tasmania (CP), Australia.

892 **Fig. S4. Principal Component Analysis of the phenotypic data.** Circles surround different populations **(a)**
893 and localities **(b)**. South America population (SoAm), Australia population (Aus), Altos de Lircay National
894 Park (AL), Villarrica National Park (VI), Huilo-Huilo Biological Reserve (HH), Valdivian Coastal Reserve
895 (VCR), Osorno Volcano (OV), Chiloé National Park (CH), Coyhaique National Reserve (CY), Central
896 Plateau of Tasmania (CP), Australia, and France (FR).
897

898 **Fig. S5. Phenotypic diversity of SoAm *L. cidri* strains isolated from high and low altitude locations.** **(a)**
899 Fitness (μ_{max} h^{-1}) variation between Aus and SoAm population. **(b)** Fitness (μ_{max} h^{-1}) variation between
900 High Altitude (HA= Altos de Lircay National Park, Villarrica National Park, Huilo-Huilo Biological Reserve,
901 Osorno Volcano, Coyhaique National Reserve) and Low Altitude (LA= Valdivian Coastal Reserve and
902 Chiloé National Park) isolation locations. **(c)** Fitness (μ_{max} h^{-1}) variation between HA and LA across
903 conditions assessed. (*) indicate statistically significant differences according to the t-test (* $p < 0.05$; ** $p <$
904 0.01 ; *** $p < 0.001$; **** $p < 0.0001$; ns = not significant).

905

906 **Supplementary Tables**

907 **Supplementary Table 1.** List of isolates employed in this study.

908 **Supplementary Table 2.** Species delimitation models.

909 **Supplementary Table 3.** Growth media utilized in this study and their composition in semi-solid and 96-well
910 plate phenotyping assays.

911 **Supplementary Table 4.** The number of samples obtained from each National Park.

912 **Supplementary Table 5.** Hypergeometric test results.

913 **Supplementary Table 6.** Summary of bioinformatics statistics.

914 **Supplementary Table 7.** Common and unique SNPs among populations and the reference strain CBS2950.

915 **Supplementary Table 8.** Structure selector output.

916 **Supplementary Table 9.** F_{ST} values by geographical location and population.

917 **Supplementary Table 10.** Summary of population genetics statistics for each clade and locality.

918 **Supplementary Table 11.** Phenotype data for *L. cidri* strains.

919