



## Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages

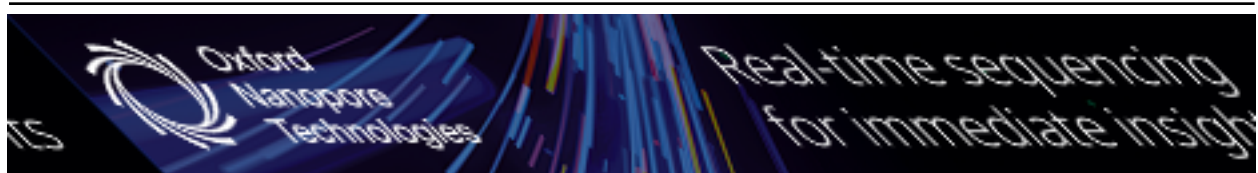
Tracy J Lee, Yuching Liu, Wei-An Liu, et al.

*Genome Res.* published online March 31, 2022

Access the most recent version at doi:[10.1101/gr.276286.121](https://doi.org/10.1101/gr.276286.121)

---

<b>P&lt;P</b>	Published online March 31, 2022 in advance of the print journal.
<b>Accepted Manuscript</b>	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
<b>Open Access</b>	Freely available online through the <i>Genome Research</i> Open Access option.
<b>Creative Commons License</b>	This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at <a href="http://creativecommons.org/licenses/by-nc/4.0/">http://creativecommons.org/licenses/by-nc/4.0/</a> .
<b>Email Alerting Service</b>	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or <a href="#">click here</a> .



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Published by Cold Spring Harbor Laboratory Press

1 **Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals**  
2 **ecology and evolution of pre-domesticated lineages**

3  
4 Tracy J. Lee<sup>1,2,3</sup>, Yu-Ching Liu<sup>1</sup>, Wei-An Liu<sup>1</sup>, Yu-Fei Lin<sup>1</sup>, Hsin-Han Lee<sup>1,4,5</sup>, Hwei-  
5 Mien Ke<sup>1</sup>, Jen-Pen Huang<sup>1</sup>, Mei-Yeh Jade Lu<sup>1</sup>, Chia-Lun Hsieh<sup>1</sup>, Kuo-Fang Chung<sup>1</sup>,  
6 Gianni Liti<sup>6</sup> and Isheng Jason Tsai<sup>1,2,3,4,5</sup>

7  
8 <sup>1</sup>Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

9 <sup>2</sup>Biodiversity Program, Taiwan International Graduate Program, Academia Sinica and National  
10 Taiwan Normal University, Taipei, Taiwan.

11 <sup>3</sup>Department of Life Science, National Taiwan Normal University, 116 Wenshan, Taipei,  
12 Taiwan

13 <sup>4</sup>Bioinformatics Program, Taiwan International Graduate Program, National Taiwan University,  
14 Taipei, Taiwan

15 <sup>5</sup>Bioinformatics Program, Institute of Information Science, Taiwan International Graduate  
16 Program, Academia Sinica, Taipei, Taiwan

17 <sup>6</sup>Université Côte d'Azur, CNRS, INSERM, IRCAN, Nice, France

18  
19  
20 Correspondence: Isheng Jason Tsai [ijtsai@sinica.edu.tw](mailto:ijtsai@sinica.edu.tw)

36 **ABSTRACT**

37       The ecology and genetic diversity of model yeast *Saccharomyces cerevisiae* prior  
38 to human domestication remain poorly understood. Taiwan is regarded as part of this  
39 yeast's geographic birthplace where the most divergent natural lineage was discovered.  
40 Here, we extensively sampled the broad-leaf forests across this continental island to  
41 probe the ancestral species diversity. We found that *S. cerevisiae* is distributed  
42 ubiquitously at low abundance in the forests. Whole-genome sequencing of 121 isolates  
43 revealed nine distinct lineages that diverged from Asian lineages during the Pleistocene,  
44 when a transient continental shelf land bridge connected Taiwan to other major  
45 landmasses. Three lineages are endemic to Taiwan and six are widespread in Asia,  
46 making this region a focal biodiversity hotspot. Both ancient and recent admixture  
47 events were detected between natural lineages and a genetic ancestry component  
48 associated with isolates from fruits was detected in most admixed isolates. Collectively,  
49 Taiwanese isolates harbor genetic diversity comparable to that of the whole Asia  
50 continent, and different lineages have coexisted at a fine spatial scale even on the same  
51 tree. Patterns of variations within each lineage revealed that *S. cerevisiae* is highly  
52 clonal and predominantly reproduces asexually in nature. We identified different  
53 selection patterns shaping the coding sequences of natural lineages and found fewer  
54 gene family expansion and contractions which contrast with domesticated lineages.  
55 This study establishes that *S. cerevisiae* has rich natural diversity sheltered from human  
56 influences, making it a powerful model system in microbial ecology.

57

58

59

60

61

62

63

64

65

66

67

68

69

## 70 INTRODUCTION

71  
72 The yeast genus *Saccharomyces*, which includes *S. cerevisiae*, is a powerful model  
73 system for revealing patterns of genomic variation underlying reproductive isolation  
74 and adaptation in eukaryotic microorganisms. Surveys of population genetic data have  
75 been used in *S. cerevisiae* to date the origin of key domestication events (Gallone et al.  
76 2016; Duan et al. 2018; Peter et al. 2018), life cycle frequencies in nature (Tsai et al.  
77 2008), the genomic basis of adaptation at continental scale (Duan et al. 2018; Peter et  
78 al. 2018), and more recently to establish its geographical origin and dispersal history  
79 (Xia et al. 2017). Phylogenomic analyses of the *Saccharomyces sensu stricto* complex  
80 and extensive sequencing of collections across the world suggest that *S. cerevisiae*  
81 originated in East Asia (Duan et al. 2018; Peter et al. 2018). The 1,011 genome  
82 project—the most broad large scale yeast population genomic study—discovered that  
83 three wild isolates from Taiwan showed an unprecedented high genetic diversity  
84 compared to populations from the rest of the world (Peter et al. 2018). Population  
85 genomics of 266 domestic and wild isolates in China revealed six wild lineages from  
86 primeval forests. The newly identified CHN-IX group represents the most diverged  
87 lineage (Duan et al. 2018). Isolates from this group and the three Taiwanese isolates  
88 grouped into a single lineage that exhibited a disjunct geographic distribution  
89 (Bendixsen et al. 2021). Although considerable knowledge is available on the  
90 biogeography and population genetics of plants and animals across continents  
91 (Whittaker et al. 2017), little is known about how eukaryotic microorganisms such as  
92 *S. cerevisiae* disperse, establish, reproduce and persist in nature (Liti 2015).

93  
94 Most *S. cerevisiae* biology has been based on experiments on a handful of  
95 laboratory domesticated strains, but comprehensive analyses of the ecology and  
96 evolutionary biology of *S. cerevisiae* in the wild are still unavailable. In nature, *S.*  
97 *cerevisiae* have been isolated from the bark, fruits, surrounding soil, and leaves of  
98 plants belonging to several different families (Naumov et al. 2013), with early reports  
99 suggesting that the yeast is most successfully isolated from the oak Family Fagaceae  
100 (Sniegowski et al. 2002; Sampaio and Gonçalves 2008; Wang et al. 2012). *S. cerevisiae*  
101 contains high genetic diversity in certain populations, including lineage-specific  
102 variants that display clear population structures (Barnett 1992; Wang et al. 2012;  
103 Cromie et al. 2013; Strobe et al. 2015; Gallone et al. 2016; Gonçalves et al. 2016; Zhu  
104 et al. 2016; Duan et al. 2018; Legras et al. 2018; Peter et al. 2018) and explain

105 phenotypic variance similar to common variants (Fournier et al. 2019). Samples from  
106 natural habitats tend to be homozygous diploids forming unique populations with  
107 minimal genetic admixture, while lineages associated with human activities were likely  
108 heterozygous, containing higher ploidy and greater genetic admixture leading to a  
109 mosaic genome makeup (Diezmann and Dietrich 2009; Liti et al. 2009; Wang et al.  
110 2012; Almeida et al. 2015). The diverse natural lineages of *S. cerevisiae* present in East  
111 Asia provide an excellent opportunity to study the natural diversity of this species,  
112 which was previously believed to be fully domesticated (Fay and Benavides 2005).

113

114 Taiwan is a continental shelf island with the fifth highest tree density in the world  
115 (Crowther et al. 2015). Among the 13 climate-related forests types in Taiwan, five are  
116 Fagaceae-dominated natural forests on low- and mid-elevation mountains (Li et al.  
117 2013), thus a potentially ideal natural habitat for *S. cerevisiae*. Taiwan also harbors a  
118 high phylogenetic diversity of flowering plants (53 out of 64 angiosperm orders present  
119 under the APG IV classification system (Lin and Chung 2017) and endemism compared  
120 to other oceanic islands (Hsieh 2002), raising the possibility that the associated  
121 microbial populations are genetically different from their continental counterparts.  
122 Here, we set out to characterize the intra genetic diversity, relative abundance and  
123 distribution of *S. cerevisiae* in Taiwanese forests over four years of broad-sampling.  
124 Our study provides novel insights of the pre-domestication phase of *S. cerevisiae* and  
125 broaden our understanding of the ecological and biogeographic implications prior to  
126 anthropogenic impacts.

127

128

129

130

131

132

133

134

135

136

137

138

## 139 RESULTS

140

### 141 Deep sampling of natural *S. cerevisiae* from Taiwanese forests

142 From July 2016 to October 2020, our sampling strategy consisted of maximizing  
143 the number of localities associated with Fagaceae hosts and sampling a broad range of  
144 plant families present in Taiwanese broad-leaved forests (**Fig. 1A, Supplemental**  
145 **Table S1**). We surveyed 693 plant hosts belonging to 43 orders, 86 families and 156  
146 genera (**Supplemental Table S2**) collected over 113 non-overlapping 1 km<sup>2</sup> grids.  
147 Various substrates (twigs, bark, leaves, flowers, fruits and topsoil around trees) were  
148 collected from each tree and subject to selective media enrichments resulting in 5,526  
149 independent incubations (**Supplemental Table S3**). The successful isolation rates of *S.*  
150 *cerevisiae* per sample and per tree host was 1.9 and 10.8%, respectively, higher than  
151 from Brazilian forests (Barbosa et al. 2016) and Slovenia oak forests (Dashko et al.  
152 2016), but lower than from North American oaks (Sniegowski et al. 2002) and Chinese  
153 wild niches (Wang et al. 2012). These isolates were recovered across altitudes of 0–  
154 2,100 meters from 18 plant families (**Fig. 1B**), with a majority from Fagaceae including  
155 four genera (27 *Quercus*, nine *Lithocarpus*, eight *Castanopsis*, and one *Fagus* species).  
156 Ten plant genera had higher isolation rates than *Quercus*, ranging from 40 to 100% per  
157 plant, albeit this recovery rate applied for as few as one tree (**Supplemental Table S2**).  
158 Among Fagaceae, *Quercus pachyloma* showed the highest isolation rate (75%; three  
159 out of four trees). Of the 339 lichen samples, four yielded successful isolations. Among  
160 the types of substrates, litter had the highest isolation rate (8.1%), providing the  
161 majority of recovered *S. cerevisiae* isolates (26.2%), followed by fruit, soil, bark and  
162 leaves (around 4–5% each). In general, the majority of samples were collected from  
163 July to December, and we found the isolation rate to be highest in July (18.9% per host  
164 tree), followed by September and October (17.5 and 11.3%, respectively). Isolation  
165 rates in other months remained around 0–11% (**Supplemental Table S3**).

166

167

168

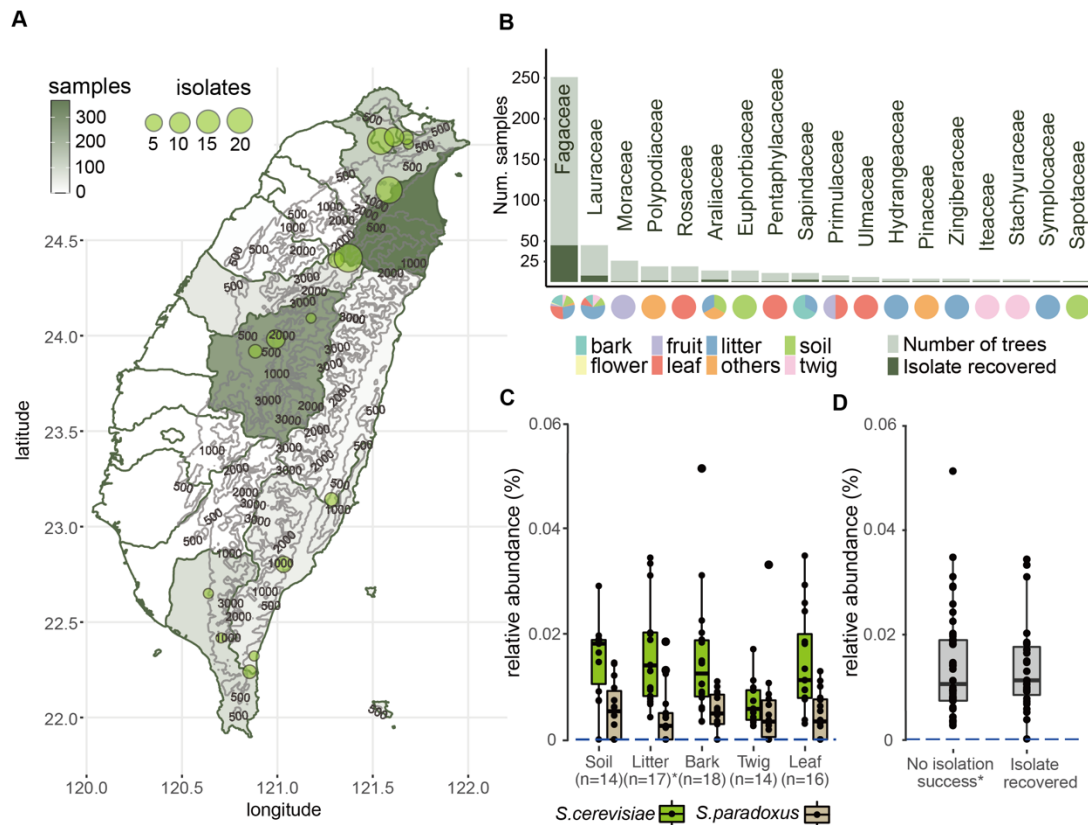
169

170

171

172

173



174

175 **Figure 1. Sampling and isolation of *S. cerevisiae* in Taiwan.**

176 (A) Map of Taiwan showing sampling efforts in each county, with darker shades  
 177 representing areas with higher number of samples collected and circles denoting the  
 178 locations where *S. cerevisiae* was successfully isolated. One isolate found on Dongsha  
 179 Island is not shown on this map. (B) Eighteen plant families from which *S. cerevisiae*  
 180 was isolated. The darker color on each bar corresponds to the number of plants that  
 181 yielded a successful isolation. Another 73 plant families from which we did not obtain  
 182 any *S. cerevisiae* isolates are not shown. Pie charts below each bar represent the  
 183 substrate surrounding plants from which samples were recovered. (C) Pairwise  
 184 comparisons found no differences in the relative abundances of *S. cerevisiae* among  
 185 bark, leaf, twig (Wilcoxon-rank with Bonferroni correction, bark-leaf,  $P=1.0$ ; bark-  
 186 twig,  $P=0.118$ , leaf-twig,  $P=0.461$ ) and (D) between samples with or without isolation  
 187 success.

188

189

190

191

192

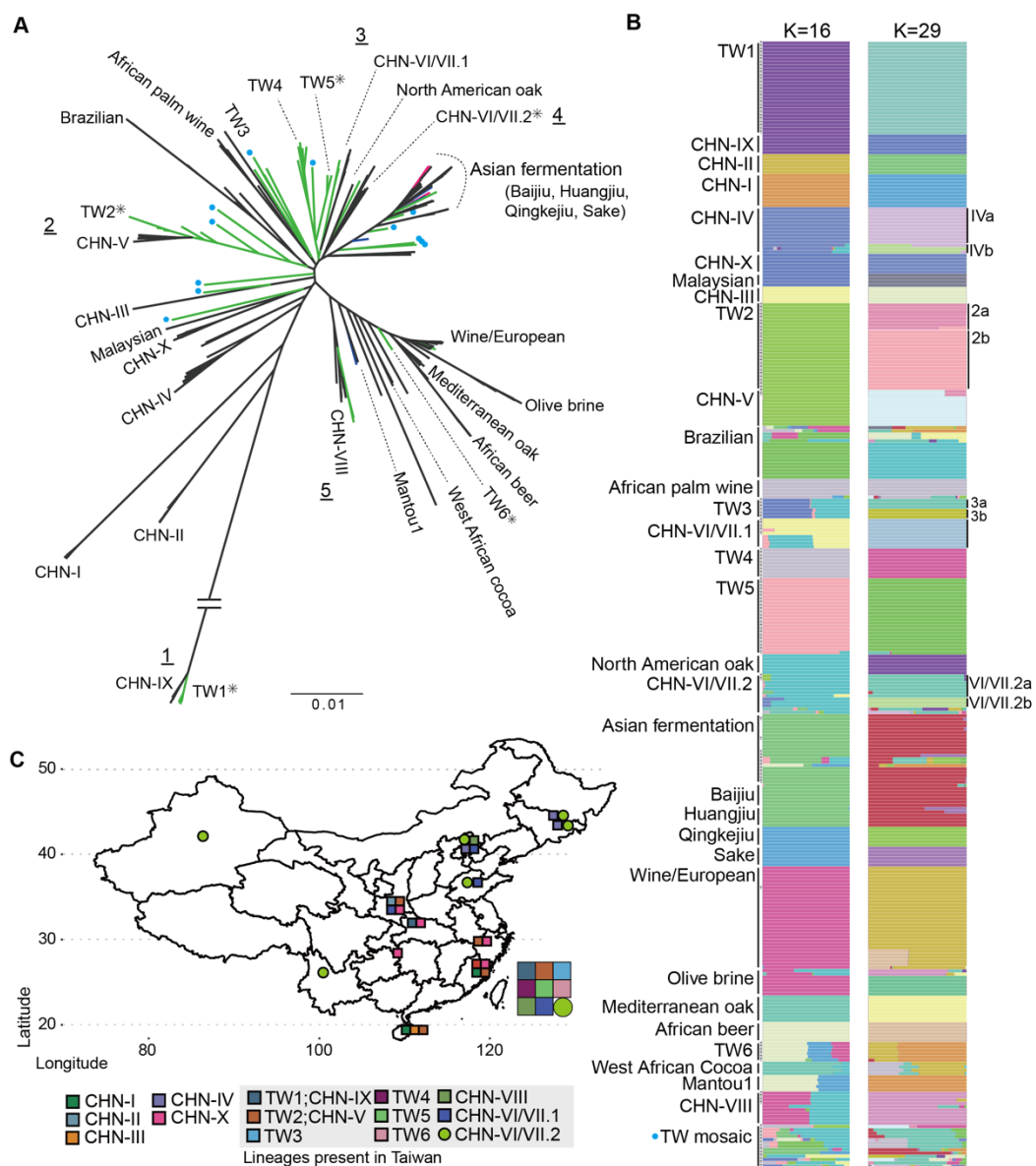
193 Recurrent sampling of eight trees over two years showed differential isolation  
194 successes (**Supplemental Table S4**), suggesting that *S. cerevisiae* had different  
195 abundances in different parts or trees. Focusing on a total of five substrates from 18  
196 trees within ~100 m<sup>2</sup> of this forest (**Supplemental Fig S1 and Table 4**), ITS amplicon  
197 sequencing succeeded in detecting just two amplicon sequence variants (ASVs)  
198 belonging to the *Saccharomyces* genus—*S. cerevisiae* and *S. paradoxus*. In contrast to  
199 surveys in temperate and boreal forests (Charron et al. 2014; Kowallik and Greig 2016;  
200 Brysch-Herzberg and Seidel 2017), *S. cerevisiae* had a higher relative abundance  
201 calculated as the percentage of the total taxa-classified reads than *S. paradoxus* in the  
202 subtropics (**Fig. 1C**). The sequence relative abundance of *S. cerevisiae* was on average  
203 0.012% in these trees belonging to seven families regardless of substrates sampled; this  
204 suggested that, despite being ubiquitous in nature, *S. cerevisiae* lives in small  
205 populations. The relative abundances of *S. cerevisiae* were found to be constant  
206 between pairwise comparisons of bark, leaves and twigs (**Fig. 1C**; Wilcoxon-rank with  
207 Bonferroni correction, bark-leaf P=1.0; bark-twig P=0.118; leaf-twig P=0.461); among  
208 tree families (**Supplemental Fig S2**, P=1.0); and whether a *S. cerevisiae* isolate was  
209 recovered (**Fig. 1D**; P=0.89). In addition, bioclimatic variables extracted from GPS  
210 coordinates also exhibited no difference between sites at which isolates were and were  
211 not recovered (**Supplementary Info and Supplemental Table S5**). Together, these  
212 results imply that the primary habitat of *S. cerevisiae* is unlikely associated with a single  
213 tree host.

214

### 215 **Multiple natural *S. cerevisiae* lineages in Taiwan**

216 We sequenced the genomes of 121 isolates with a median coverage of 91× depth  
217 (**Supplemental Table S6**). All isolates were primarily homozygous (average  
218 heterozygosity: 0.01%) diploids, with the exception of isolate PD36A, which was a  
219 triploid (**Supplemental Fig S3**) estimated by flow cytometry (**Supplementary Info**).  
220 We constructed a maximum likelihood phylogeny based on 765,169 SNPs segregating  
221 in 340 isolates (**Fig. 2A**) by including of 219 representative isolates previously studied  
222 from multiple habitats (Barbosa et al. 2016; Duan et al. 2018; Peter et al. 2018; Pontes  
223 et al. 2019) that sampled all the major world-wide wild and domesticated lineages. The  
224 topology of the isolate phylogeny is largely consistent with a previous neighbor joining  
225 tree from the 1,011 *S. cerevisiae* genome project (Peter et al. 2018): the natural isolates  
226 were mostly grouped according to sampling locations, while industrial isolates were  
227 grouped according to fermentation sources. In particular, the Wine/European lineage

228 and Asian Fermentation lineage were separated by a suite of natural isolates, suggesting  
 229 independent domestication events (Fay and Benavides 2005; Liti et al. 2009; Gonçalves  
 230 et al. 2016; Gallone et al. 2018). The African Palm Wine lineage was separated from  
 231 the West African Cocoa lineage and placed near the branch leading to the Asian  
 232 Fermentation lineage. Furthermore, the CHN-VI/VII lineage, which was collected from  
 233 fruits, was further separated into two lineages consistently with geographical proximity  
 234 of its members (designated as CHN-VI/VII.1 and CHN-VI/VII.2 in **Fig. 2A and 2C**,  
 235 **Supplemental Table S6**).  
 236



237  
 238 **Figure 2. Phylogeny and population structures of 340 *S. cerevisiae* isolates.**  
 239 **(A)** Unrooted phylogeny based on 765,169 genome-wide SNPs. Bootstrap support was

240 >90% in all major lineages except inner nodes within some lineages indicated by  
241 asterisks. Natural, industrial and fermentation-related isolates discovered in Taiwan are  
242 in colored in green, blue and magenta, respectively. Mosaic Taiwanese isolates from  
243 ADMIXTURE analyses are labelled with blue dots on branch tips. Five cases in which  
244 Taiwanese and Chinese isolates were found to be monophyletic are indicated with  
245 underscored numbers. The Asian fermentation lineage includes Baijiu-, Huangjiu-,  
246 Qingkejiu-, Sake- and fermentation-related isolates from Taiwan, as shown in (b). **(B)**  
247 Population structure from ADMIXTURE analysis at K=16 and 29. Labels on the left  
248 side of the bars indicate each group from K=16 and some were further separated in  
249 K=29 which are annotated on the right side. Natural Taiwanese isolates with admixed  
250 genome makeup are shown together in the TW mosaic group. **(C)** Map of China and  
251 Taiwan indicating where the *S. cerevisiae* natural lineages were found (colored squares  
252 and circle). CHN-IV isolates that were sampled from Japan are not shown on this map.  
253

254 Previous studies of natural *S. cerevisiae* revealed that most lineages comprise isolates  
255 from neighboring geographic origins (Duan et al. 2018; Peter et al. 2018); however,  
256 natural Taiwanese isolates are found throughout the phylogeny despite the small size  
257 of the island (**Fig. 2A**). The population structure of the 340 isolates used for the  
258 phylogeny were analyzed using ADMIXTURE (Alexander et al. 2009) with K from 2  
259 to 30. Cross validation (CV) error was lowest at K=29 (CV error = 0.09025), though it  
260 only differed <1% between K=16 and 30 (**Fig. 2B, Supplemental Fig S4**).  
261 ADMIXTURE at K=16 was largely consistent with the phylogenetic lineages such as  
262 placing CHN-VI/VII into two genetic groups. ADMIXTURE at K=29 further separated  
263 two instances in which a group was split into solely either Chinese or Taiwanese  
264 isolates, suggesting the presence of lineage-specific segregating sites as a result of  
265 geographical isolation (**Fig. 2B and Supplemental Table S7**). Some groups  
266 comprising isolates from a proximate geographical origin were further split into smaller  
267 groups, suggesting ongoing genetic differentiation. Based on ADMIXTURE K=29, we  
268 reused previously assigned group names (Duan et al. 2018; Peter et al. 2018) and  
269 designated these differentiated groups as well as new lineages exclusively found in  
270 Taiwan TW1 to TW6 (the most diverged lineage was TW1 and progressively labelled  
271 clockwise; **Fig. 2**). Examples include the recovery of 28 TW1 isolates clustered with  
272 CHN-IX (Duan et al. 2018; Bendixsen et al. 2021), together representing the most  
273 divergent lineage to date, and a new TW4 lineage that did not contain any Chinese  
274 strains (**Fig. 2**). This new lineage included isolates sampled from lichens and four

275 isolates sampled from mushrooms that were previously placed in an undefined lineage  
276 (Peter et al. 2018), suggesting a possible association with other fungi (Spribille et al.  
277 2016). In other instances, Taiwanese isolates were found in three previously assigned  
278 groups such as CHN-VI/VII.1, CHN-VI/VII.2 and CHN-VIII. Isolates of the most  
279 diverged TW1/CHN-IX lineage were separated by approximately 1,400 km, with four  
280 other natural lineages (CHN-I, V, VI/VII, and X) in between. Twenty-three isolates  
281 from northern Taiwan (TW2) clustered with the CHN-V population sampled as far as  
282 1500 km apart. Together, these results suggest that Taiwan harbors the highest number  
283 of lineages that exhibit disjunct distributions followed by the Hubei-Shanxi region (nine  
284 and five, respectively; **Fig. 2C**)

285

### 286 **Evidence of admixture in natural lineages**

287 Both inter- and intra-species spontaneous hybridizations have been documented  
288 in *Saccharomyces* species. For instance, the wild *S. paradoxus* SpC\* lineage present in  
289 North America (Eberlein et al. 2019) and domesticated *S. cerevisiae* Alpechin lineage  
290 (D'Angiolo et al. 2020) are classic examples of past hybridizations that played genomic  
291 and phenotypic diversities (Barbosa et al. 2016; Duan et al. 2018; Peter et al. 2018;  
292 Eberlein et al. 2019). Most Taiwanese isolates tend to have little admixture, with 20  
293 and 5% (27/137, 7/137) of isolates containing at least 10% of the genetic component  
294 from two and at least three genetic ancestries (**Fig. 2B, Supplemental Table S7**),  
295 respectively. We confirmed the genetic components of domesticated strains' origins in  
296 wild isolates from African cocoa (Peter et al. 2018), olive brines and Brazilian  
297 forests (Barbosa et al. 2016), and identified an additional TW4 group sharing major  
298 genetic components with the steamed buns (Mantou) and Wine/European lineages,  
299 albeit recovered from nature. Other Taiwanese admixed isolates were apparent on the  
300 phylogenetic tree as isolated branches and had different levels of admixture from  
301 domesticated lineages (**Fig. 2A**). Additionally, all Taiwan isolates recovered from fruits  
302 contain CHNVI/VII-2a genetic component (**Supplemental Fig S5**); this coincides with  
303 the non-admixed CHNVI/VII-2a isolates, which have the widest geographically  
304 distribution in Asia (**Fig. 2C**).

305

306 To confirm that gene flow occurred between genetic groups, we applied TreeMix  
307 (Pickrell and Pritchard 2012) to designated groups from ADMIXTURE K=16 (**Fig. 3A,**  
308 **Supplementary Info and Fig S6**). The TreeMix phylogeny first indicated extensive  
309 gene flow among domesticated lineages such as solid- and liquid- state fermentation

310 products and between natural lineages sister to domesticated lineages. Examples  
311 include isolates from steamed buns (Mantou) and Asian alcoholic beverages (Sake and  
312 Qingkejiu), as well as TW6 forest isolates. Second, the phylogeny also identified gene  
313 flow between natural lineages sister to the Wine/European and Asian Fermentation  
314 lineages. The CHN-VIII group emerged from both the Wine/European and fruit  
315 enriched CHN-VI/VII-2 lineages, which contain isolates from fruits and the natural  
316 environment across the Asian continent, including Taiwan. We also recovered hybrids  
317 between natural lineages that coexisted in proximity. Two isolates each belonging to  
318 TW4 or TW2 lineage came from fallen fruit, while PD38A was isolated from fruit  
319 growing on a *Castanopsis fargesii* tree (**Fig. 3a**). This PD38A hybridization timing was  
320 likely to be recent, given the presence of large haplotype blocks not extensively broken  
321 down by recombination containing variants identical to each parental lineage  
322 (**Supplemental Fig S7**). Overall, these results suggest that hybridizations were  
323 common in *S. cerevisiae* and that some admixed lineages have persisted in nature.  
324 Reanalysis of the TreeMix phylogeny based on ADMIXTURE group K=29 shows  
325 consistent results—recurrent migrations occurred between lineages, leading to the  
326 Wine/European and Asian Fermentation lineages (**Supplementary Info,**  
327 **Supplemental Fig S8 and S9**). To incorporate these findings into a comparative  
328 resource, we further sequenced the genomes of 24 Taiwanese isolates representing all  
329 the natural lineages discovered in Taiwan using Oxford Nanopore reads  
330 (**Supplemental Table S8**).

331

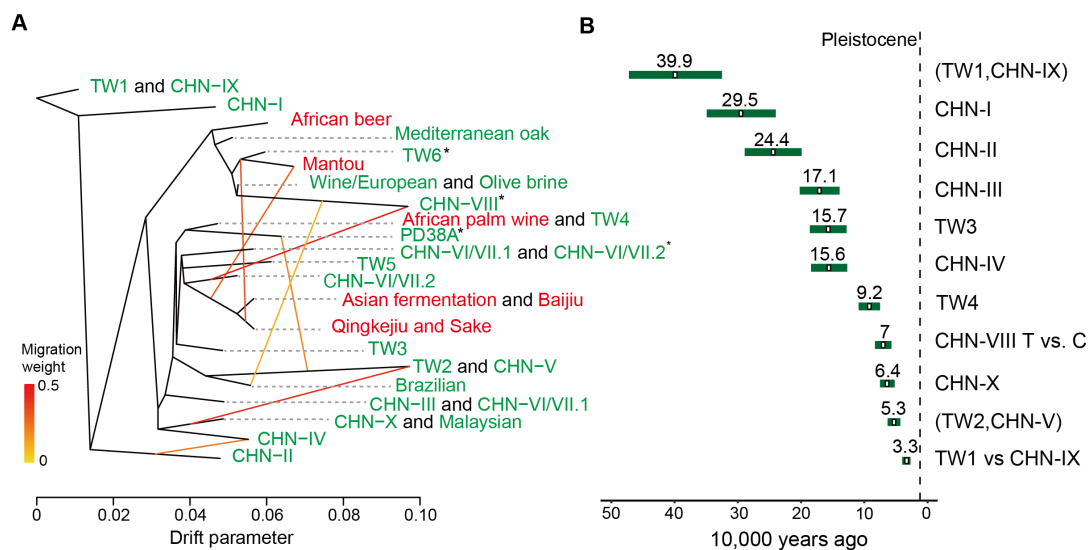
332 Using molecular calibrations, divergence between different natural lineages as  
333 well as Chinese/Taiwanese split was inferred using either pairwise divergence or a  
334 phylogenomic approach (**Supplementary Info, Supplemental Figure S10 and Table**  
335 **S9**). A more recent divergence was estimated from the former approach which the  
336 lineages were on average diverged 0.03–0.07 million years ago (Ma) (**Fig. 3B**)  
337 compared to 0.54–1.11 Ma inferred from the phylogeny (**Supplemental Table S9**).  
338 Together, these estimates fell during the Pleistocene epoch suggesting that the split may  
339 represent a vicariant event resulting from the submergence of the Taiwan Strait Land  
340 Bridge during interglacial periods and/or uplift of Taiwanese mountains (Teng 1990)  
341 during this period.

342

343

344

345



346

347 **Figure 3. Migration and divergence time between lineages**

348 **(A)** Migration edges (yellow to red colored lines) estimated by TreeMix showing  
 349 seven migration edges on the phylogeny. Different edge colors indicate the strength of  
 350 migration. Lineages were colored according to isolation sources (red and green denote  
 351 domesticated and wild environment, respectively). Asterisk denotes lineages that  
 352 contain multiple genetic components from different K from the ADMIXTURE  
 353 analyses. **(B)** Molecular estimate of time to the most recent common ancestor in  
 354 different *S. cerevisiae* lineages. The estimates is shown in **Supplemental Table S9A**.

355

### 356 Biogeography of wild *S. cerevisiae* lineages

357 In nature, single genetically homogenous fungal populations are generally found  
 358 in distinct geographical regions as a result of isolation by distance (Branco et al. 2017;  
 359 Chung et al. 2017; He et al. 2022). In contrast, the presence of multiple *S. cerevisiae*  
 360 lineages at the same locality in Taiwan, even on the same tree, is striking (**Fig. 4A**,  
 361 **Supplemental Fig S11 and Table 6**). In one sampling area, four lineages were  
 362 recovered less than 35 km apart in central Taiwan (TW1-TW4 and mosaics, n=10;  
 363 **Supplemental Fig S11**). In another sampling site—Fushan Botanical Garden, where  
 364 we obtained 23 isolates comprising three lineages and admixed isolates were recovered  
 365 (**Fig. 4A**). Both significant negative and positive correlation between genetic and  
 366 geographical distance were observed in isolate pairwise comparisons in close distances  
 367 ( $P < 0.05$  with 1,000 permutations; **Supplemental Fig S12**). However, no such  
 368 association was found of the whole region (Mantel's  $r = 0.07$ ,  $P = 0.23$ ; **Fig. 4B**),  
 369 suggesting that in a given region the relationships between isolates was less determined

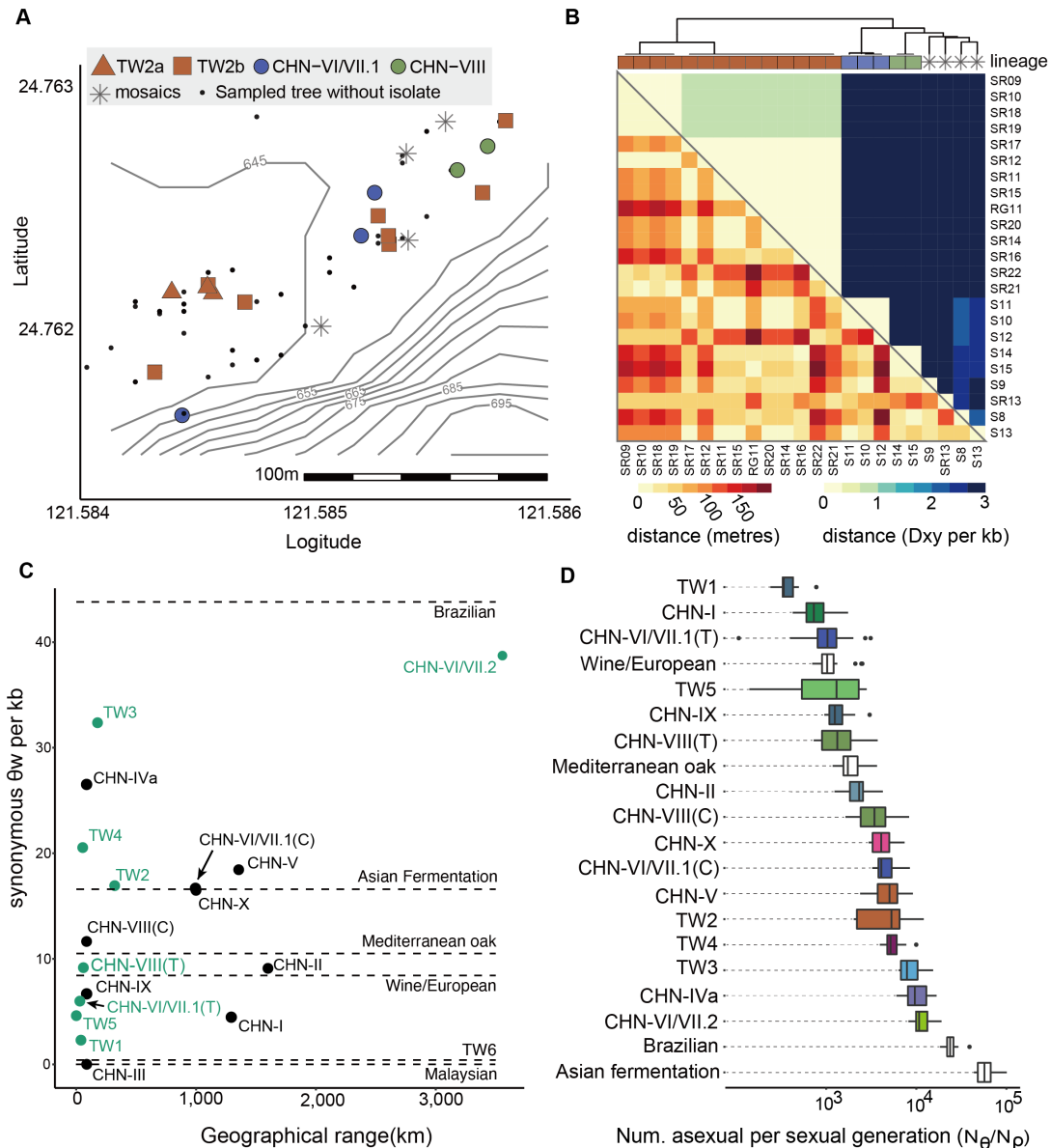
370 by population structure of single lineages but dictated by the heterogeneity of multiple  
371 lineages coexisted at small spatial scale. The admixed isolates did not contain genetic  
372 components from adjacent isolates, but instead from CHN-VI/VIII.2a and others  
373 (**Supplemental Fig S13**). In addition, these combinations of coexisting lineages were  
374 not present in a similar locality range in China (**Fig. 2C**), suggesting coexisting of  
375 lineages was established by independent dispersal events.

376

377 The overall genetic diversity of Taiwanese isolates was comparable to that of  
378 Chinese isolates (Taiwan  $\theta_{\pi}=5\times 10^{-3}$  versus China  $\theta_{\pi}=6\times 10^{-3}$ ), even though the samples  
379 were only meters to tens of kilometers apart (**Supplemental Fig S14**). This reinforced  
380 that the pattern of *S. cerevisiae* diversity in a geographical region was shaped by the  
381 presence of multiple lineages and heterogeneity of metapopulations in the same habitat.  
382 Up to a two-fold difference was observed in genetic diversity between lineages, with  
383 the aforementioned most widespread CHN-VI/VII.2a group harboring the greatest  
384 diversity (**Fig. 4C and Supplemental Table S10**). In contrast, when comparing isolates  
385 on the same tree at an extreme microgeographic scale, we found instances of all isolates  
386 being clonal or from different lineages with pairwise differences differed by  $\sim 35,000$ -  
387 fold (1–35,922 maximum number of pairwise mismatches of isolates recovered on the  
388 same tree;  $\theta_{\pi}=8.3\times 10^{-8}$ – $2.9\times 10^{-3}$ ; **Supplemental Table S11**). Three out of seven  
389 lineages have exhibited a linear isolation-by-distance (IBD)(Meirmans 2012) signature  
390 including the aforementioned TW2 lineage ( $p<0.05$ ; **Supplemental Fig S15**). TW2  
391 lineage exhibited a central-southern Taiwan discontinuous distribution, where isolates  
392 are found as much as 194 km apart. This suggests that the greater the geographical  
393 range, the higher the likelihood of genetic differentiation. Indeed, greater sequence  
394 divergence was shown when intra-lineage isolates between lineages were  $>10$  km apart  
395 ( $P<0.001$ ; Wilcoxon rank-sum test; **Supplemental Fig S16**), which supported genetic  
396 differentiation as a result of geographical isolation (Liti et al. 2006).

397

398



399

400 **Figure 4. Patterns of genetic variations and geographical distribution. (A)** Fine-  
 401 scale geographic sampling at Fushan Botanical Garden in Taiwan. A total of 106 tree  
 402 sites constituting 286 substrates were sampled in this region. Different colors  
 403 represent different lineages and filled circle denote sampled trees from which *S.*  
 404 *cerevisiae* was not successfully isolated. **(B)** Genetic and geographic distance of  
 405 isolate pairs identified in (A). **(C)** Lack of correlation between genetic diversity  $\theta_w$  at  
 406 synonymous site and geographical range across lineages. Diversity for lineages where  
 407 geographical range is unavailable is indicated with dashed lines. **(D)** Frequency of  
 408 asexual per sexual generations across lineages.

409

410

## 411 Population genomic across lineages

412 Patterns of segregating sites can be used to infer the relative contributions and  
413 frequencies of reproduction modes in nature (Tsai et al. 2008). Wild *S. cerevisiae*  
414 isolates were highly inbred: Wright's inbreeding coefficient  $F$  was an average of 0.99  
415 and clones made up 16–100% of each lineage (**Supplemental Table S6**), suggesting  
416 that most generations were mitotic regardless of lineage. We estimated that the effective  
417 population size of mutational ( $N_e$ ) and recombinational ( $N_\rho$ ) diversity for all  
418 chromosomes was  $4.1 \times 10^6$ – $7.7 \times 10^7$  and 197–12,821, respectively, averaging across  
419 chromosomes (**Supplemental Table S12**) of selected lineages (**Supplemental Table**  
420 **S13**). The differences between both  $N_e$  estimates equates to approximately 382–61,264  
421 mitotic cell divisions for every meiosis event (**Fig. 4D**). Such estimates overlap with  
422 previous estimates of 12,500–62,500 clonal generations based on the decay of  
423 heterozygosity during mitosis (Magwene et al. 2011), 1,000–3,000 in two  
424 genealogically independent populations of *S. paradoxus* (Tsai et al. 2008) and less in  
425 800,000 generations in the fission yeast *Schizosaccharomyces pombe* (Farlow et al.  
426 2015).

427

428 We calculated the mean neutrality index  $NI_{TG}$  (Stoletzki and Eyre-Walker 2010)  
429 for each lineage using polymorphism data from each lineage and *S. paradoxus* as an  
430 outgroup (**Fig. 5A**).  $NI_{TG}$  was higher in the domesticated lineages such as  
431 Wine/European as well as the most diverged TW1/CHN-IX amongst the natural  
432 lineages, suggesting more selection in purging the deleterious alleles in these lineages.  
433 We found that variations in  $NI_{TG}$  in natural lineages were not due the differences in  
434 effective population size inferred from mutational diversity (Kendall's  $\tau = -0.26$ ;  $P =$   
435  $0.11$ , **Supplemental Fig S17**) but from recombinational size (Kendall's  $\tau = 0.33$ ;  
436  $P = 0.047$ , **Fig. 5B**), suggesting that the selection efficacy was greater when  
437 recombination occurred during sexual reproduction, consistent with the results of  
438 experimental evolution in a laboratory setting (Goddard et al. 2005). Such a relationship  
439 was more significant when lineages with low recombination were removed (Brazilian  
440 and the Asian fermentation lineage was excluded; Kendall's  $\tau = 0.52$ ;  $P = 0.002$ , **Fig. 5B**),  
441 indicating similar efficacy of selection in the absence or low recombination.

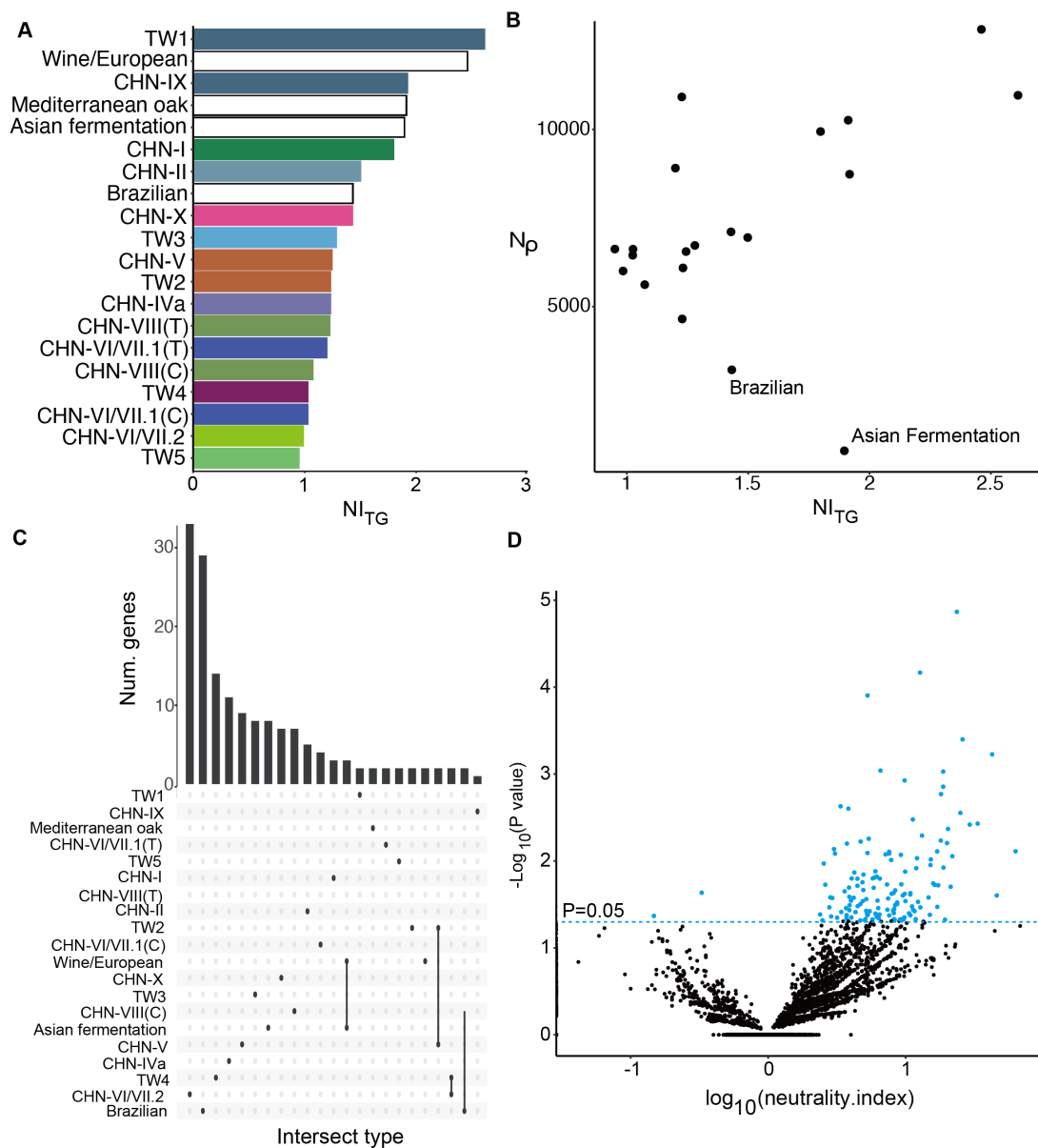
442

443 We next investigated the extent of selection at the gene level within each lineage  
444 by conducting the McDonald-Kreitman test (McDonald and Kreitman 1991). Overall,  
445 we found 18–503 genes with a neutrality index ( $NI$ )  $> 1$  in each lineage (**Fisher's exact**

446 **test  $P < 0.05$ ; Supplemental Table S14**), compared to 1–38 genes with  $NI < 1$   
447 (**Fisher’s exact test  $P < 0.05$ ; Fig. 5C and Supplemental Table S15**), indicating that  
448 more genes had an excess of amino acid polymorphisms than were under positive  
449 selection (**Fig. 5C and Supplemental Fig S18**). The most genes with  $NI > 1$  in the Asia  
450 lineages belonged to the most diverse TW3 (144 genes, **Fig. 5D**). The majority of these  
451 genes indicative of departure from neutrality were observed in only one lineage  
452 emphasizing the lineages’ independent evolutionary history (**Supplemental Fig S19**).  
453 These negatively selected genes together were found to be enriched in biological  
454 process such as response to stimulus, cell communication and intracellular signal  
455 transduction (**Supplemental Table S16**). Within each lineage, no genes under either  
456 purifying or positive selection were enriched in any particular biological processes,  
457 except for the Brazilian lineage, which contained a sufficient number of genes showing  
458  $NI > 1$ . Three genes (*CDC10*, *CIT2* and *SAT4*) in both the Wine/European and Asian  
459 Fermentation lineages showed the largest overlap of genes with  $NI < 1$  among the  
460 lineages (**Fig. 5c**). *CIT2* encodes a citrate synthase was involved in ethanol tolerance  
461 (Kasavi et al. 2014). Similarly, 25 genes under negative selection in these two lineages  
462 were the only overlap category found to be enriched in biological processes including  
463 Ras protein signal transduction (**Supplemental Table S17**), which were also targets of  
464 adaptation across different experimental evolution experiments (Long et al. 2015).  
465 Together these results suggest that the common selective pressure from domestication  
466 may have driven the adaptations of these genes. It is unlikely that this overlap was the  
467 result of stronger divergent selection with the *S. paradoxus* outgroup because the  
468 pattern was consistent when we used the McDonald-Kreitman with TW3 and CHN-  
469 VIII as outgroups, as they were sister to each of the domesticated lineages  
470 (**Supplemental Fig S20**).

471

472



473

474 **Figure 5. Population genomics across lineages.** (A)  $NI_{TG}$  estimates in natural and  
 475 two domesticated lineages (B) Relationship between  $NI_{TG}$  and  $N_p$  across natural  
 476 lineages (C) Lineage-specific and shared genes with  $NI < 1$ . (D)  $NI$  from the  
 477 McDonald-Kreitman test for each gene in the TW3 lineage with *S. paradoxus* as the  
 478 outgroup. Genes that were significantly different from  $NI = 0$  were highlighted in blue.

479

#### 480 Population differentiation dynamics between lineages

481 The presence of different levels of shared genetic components observed between  
 482 Chinese and Taiwanese isolates among the five shared lineages suggested a distinct  
 483 differentiation between the disjunct populations. The average ratio of nonsynonymous  
 484 to synonymous substitution rates ( $d_N/d_S$ ) between China and Taiwan isolates across

485 lineages was 0.21 (**Fig. 6A**), suggesting that there was pervasive negative selection  
486 acting on the coding sequences of *S. cerevisiae*, with only 40–303 out of 6,572 genes  
487 showing signals of positive or balancing selection ( $d_N/d_S > 1$ ) across the Taiwanese  
488 lineages. Consistent with observations from NI<sub>TG</sub>, most of these genes were lineage-  
489 specific, with only *AIM21*, involved in mitochondrial inheritance, detected in four out  
490 of five lineages (**Supplemental Fig S21 and Table S18**) suggesting that selection acted  
491 independently in these lineages.

492

493 Gene duplication played an important role in the evolution of domesticated *S.*  
494 *cerevisiae* strains exhibiting more rapid copy number variation than wild strains  
495 (Bergström et al. 2014; Yue et al. 2017; Duan et al. 2018). To investigate the extent to  
496 which gene families differed between sister natural lineages, we *de novo* assembled,  
497 annotated and inferred orthogroup of non-clonal isolates using OrthoFinder (Emms and  
498 Kelly 2015). Compared to domesticated lineages (116 CHN-VIII vs Wine/European  
499 and 111 TW3 vs Asian fermentation; **Supplemental Fig S22**), only 17-49 orthogroup  
500 (OGs) were found to differ between the Chinese and Taiwanese lineages since their  
501 split (Wilcoxon rank sum test  $P < 0.05$ ; **Fig. 6B and Supplemental Figure S23**). A  
502 large fraction (36.7-94.7%) were single copy expansion or contractions (**Supplemental**  
503 **Table S19**), lineage specific and enriched in subtelomeres (**Supplemental Table S20**).  
504 The category that overlapped the most comprised seven OGs that were significantly  
505 different in two coexisting lineages: CHN-VIII and TW2 (**Fig. 4A and Supplemental**  
506 **Figure S24**). In addition, the largest orthogroup inferred was made up of hexose  
507 transporter genes (HXT), which are involved in polyol transport; this orthogroup was  
508 significant in four out of seven lineage comparisons (**Fig. 6C**). Copy numbers differed  
509 both between domesticated and natural lineages and among the natural lineages  
510 (**Supplemental Figure S25**). The Taiwanese lineages typically exhibited expanded  
511 HXT copies compared to Chinese or domesticated lineages and inspecting isolates with  
512 long read assemblies revealed these copies were colinear regardless of lineage (**Fig. 6D**  
513 **and Supplemental Figure S26**). Together these results suggest that the Taiwanese  
514 isolates may have maintained a larger HXT repertoire, perhaps allowing them to utilise  
515 different sugar types or concentrations.

516

517

518

519



529 **DISCUSSION**

530

531 A comprehensive understanding of the natural history of the budding yeast *S.*  
532 *cerevisiae* is key to further utilizing one of the most human-exploited microorganisms.  
533 In this study, we leveraged a four-year extensive sampling in Taiwan and combined  
534 metabarcoding approach to uncover *S. cerevisiae*'s ubiquitous presence but low  
535 abundance in broadleaf forests. We isolated and whole-genome sequenced 121 isolates  
536 to confirm the presence of the most diverged lineage TW1 (Bendixsen et al. 2021) and  
537 uncover five additional lineages that shared ancestries with lineages found in China as  
538 well as four new lineages exclusively found in Taiwan. We show that sympatric  
539 lineages coexist in different parts of Taiwan and identified introgressions between  
540 lineages. We found that the population structure of *S. cerevisiae* can be explained by a  
541 markup of different lineages that each outcrossed on average once in every 382–61,264  
542 mitotic generations. These differences resulted in different selection efficacies across  
543 the lineages. The availability of high-quality *S. cerevisiae* assemblies presented here,  
544 in addition to genetic, molecular tools and genome resources such as the 1,011 genomes  
545 collection (Peter et al. 2018) already available in this model organism provides an  
546 exciting new platform to study microbial ecology.

547

548 Although *S. cerevisiae* has repeatedly been recovered from oak bark in the  
549 Northern Hemisphere (Sniegowski et al. 2002; Robinson et al. 2016) and being the only  
550 substrate of isolation in recent studies (Goddard and Greig 2015; Duan et al. 2018), our  
551 findings shows that *S. cerevisiae* is present as a generalist occurring at low abundance  
552 in a variety of broadleaf forest substrates. In addition to temperature, we speculate that  
553 isolation success for *S. cerevisiae* was shaped by co-existing microbial communities  
554 (Kowallik et al. 2015) competing with *S. cerevisiae* in the enrichment media. In  
555 addition, at a lineage level *S. cerevisiae* was found to be associated with particular  
556 environments, suggesting that it may have had an ecological niche (Goddard and Greig  
557 2015): TW4 was isolated only from fungal fruiting bodies and lichens though further  
558 work is needed to conclude a possible symbiotic relationship, and a CHN-VI/VII.2  
559 genetic component was present in many lineages and enriched in isolates recovered  
560 from the tree fruit substrate. Higher frequencies of admixed isolates observed in fruits  
561 may simply be a result of increased contacts with other lineages. Alternatively, fruits  
562 and organisms associated with those fruits such as frugivorous animals and vectors may  
563 represent niches that promote hybridization; for instance, sporulation has been

564 suggested to be an adaptation that allows cells to survive in nutrient-depleted conditions  
565 such as insects' intestines during experimental passaging (Thomasson et al. 2021).  
566 Notably, the presence of CHN-VI/VII.2 genetic components in many natural lineages  
567 across the world, as well as in admixed isolates found in fruits, raises the possibility  
568 that the common ancestors dispersed from East Asia were from this lineage. In addition  
569 to abiotic factors, we speculate that such dispersal events of fruits may be aided by  
570 insects and human foraging.

571

572 We found that, unlike the general expectation in biogeographic studies that an  
573 island only contains a subset of genetic diversity from the mainland population, the  
574 genetic diversity of *S. cerevisiae* populations from Taiwan can be as diverse as those  
575 found in the Asia continent. The persistence of ancestral lineages may be a result of  
576 Taiwan being a high environmentally heterogeneous region (Ali 2018; Lin et al. 2020)  
577 and its prolonged bioclimatic stability (Tsukada 1966) than that of nearby eastern China.  
578 Alternatively, the geographic scale for distinguishing island and mainland populations  
579 and the importance of habitat diversity may differ between microorganisms (Davison  
580 et al. 2018) and other macroorganisms, such as animals and plants. The biogeography  
581 of *S. cerevisiae* appears to be similar to that of its associated flora in East Asia. Disjunct  
582 distributions of plants between Taiwan and different parts of China are common (Jianfei  
583 et al. 2013). The phylogeography of representative herbaceous and woody plants  
584 indicates that these representatives originated in mainland China, then migrated to  
585 Taiwan and the Ryukyu Archipelago during the Pleistocene as sea level fluctuations  
586 yielded recurring landbridges (Chiang and Schaal 2006; Niu et al. 2018; Jiang et al.  
587 2019). We note that the Pleistocene was also the period when several tree species  
588 extinctions first took place across both the Americas (Seersholm et al. 2020) and  
589 Europe (Magri et al. 2017); this was followed by a rapid migration of *Quercus* that  
590 made it the dominant tree genus (Magri et al. 2017), which may have played a role in  
591 the restricted *S. cerevisiae* lineages observed outside East Asia. A systematic sampling  
592 of *S. cerevisiae* in the mainland continent—especially regions containing flora records  
593 exhibiting a disjunct distribution like in Taiwan, e.g., the Himalaya-Hengduan  
594 mountains (Niu et al. 2018), as well as plate boundaries—may help us better understand  
595 the biogeography of *S. cerevisiae*.

596

597 Our findings of rampant hybridization events between wild, wild with  
598 domesticated and domesticated lineages bring new perspectives to the ongoing debate

599 over whether *S. cerevisiae* domestication happened once (Duan et al. 2018; Han et al.  
600 2021) or multiple times (Almeida et al. 2015; Peter et al. 2018). By revealing frequent  
601 hybridizations between natural lineages, we show that isolates used in Asian and  
602 European fermentations may have been domesticated independently from the lineage  
603 CHN-VI/VII.2, and the single-domestication-event notion may be confounded by  
604 admixed isolates. Isolates from Asian fermentations were sister to the CHN-VI/VII.2  
605 lineage, and subsequent genetic differentiations of this group have led to independent  
606 lineages such as the North American oak group, or the Mediterranean oaks group which  
607 is sister to the European/Wine isolates (**Fig. 2A and 3A**). Isolates outside of East Asia  
608 likely bear genetic components of this group. This may result in the placement of these  
609 isolates in or close to this group in a phylogeny. Ongoing hybridizations also complicate  
610 the inference; for instance the Brazilian rum population is a result of hybridization  
611 between European/Wine and North American groups (Almeida et al. 2015). Efforts to  
612 identify signatures of domestication environments (Han et al. 2021) may also be  
613 challenging when admixture is detected between these lineages. Isolation and recording  
614 the frequencies of these admixed isolates in nature could provide further insights into  
615 the conditions in which new lineages emerge.

616

617 Inferring population history in *S. cerevisiae* with different frequencies of asexual  
618 and sexual generations (Tsai et al. 2008) is challenging when using population genetics  
619 methods designed around human heterozygosity and recombination rates (Li and  
620 Durbin 2011). Disagreement in the divergences estimated from the phylogeny and  
621 pairwise divergence between isolates were observed. The phylogenomic method  
622 assume no gene flow and recombination with different lineages, and although *S.*  
623 *cerevisiae* is a predominantly asexual organism, recombination  $\rho$  was still detected and  
624 thus inflate divergence (Schierup and Hein 2000; Li et al. 2019). Conversely, estimates  
625 from pairwise divergences were more consistent to other reports (Leducq et al. 2016)  
626 but may underestimate the true divergence as we do not know the extent of quiescence  
627 of different *S. cerevisiae* lineages (Gray et al. 2004). Recent advances in directly  
628 tracking genotype evolution across natural habitats (Xia et al. 2017; Rudman et al. 2022)  
629 may lead to more accurate inferences once some of the fundamental parameters such  
630 as average generation time can be obtained in nature.

631

632

633 To conclude, we combined deep sampling, metabarcoding, isolate collection and  
634 whole-genome resequencing to illuminate the pre-domestication phase of  
635 *Saccharomyces cerevisiae* at an unprecedented resolution. The roles of *S. cerevisiae* in  
636 temperate forest environment have been studied in details (Mozzachiodi et al. 2022),  
637 and we reveal that multiple natural lineages of *S. cerevisiae* persist in the subtropic and  
638 tropical broadleaf forests in Taiwan, indicating the species is found everywhere but  
639 some genetically differentiated lineages prefer certain substrates. These observations  
640 help us to revisit our understanding of eukaryotic microorganism evolution—for  
641 instance, an alternating life cycle seems to be a convenient life history trait when  
642 genetically diverged partners are around. As more and more ecosystems—e.g., tropical  
643 cloud forests (Karger et al. 2021)—and biodiversity are lost, actions should be taken to  
644 conserve and reveal the ecology and evolution of not just *S. cerevisiae* but species with  
645 a proposed geographical origin. The availability and gene flow between these lineages  
646 also allow future experiments such as on hybrid fitness to be designed to resemble the  
647 subject’s natural scenarios rather than relying on domesticated strains.

648

649

## 650 METHODS

651

### 652 Sampling and isolating *Saccharomyces cerevisiae*

653 From September 2016 to October 2020, we collected a total of 2,461  
654 environmental samples from various substrates (bark  $n=340$ , twigs  $n=328$ , leaf  
655  $n=528$ , litter  $n=320$  and fruit  $n=78$ ) surrounding 693 plant hosts (**Supplemental**  
656 **Table S2**). A total of 339 lichen samples, aliquots from six fermentation practices,  
657 and 68 from other sources (insect corpse  $n=43$ , fruiting body  $n=14$ , industrial strains  
658  $n=5$  and others where biomaterial was sampled only once  $n=6$ ) were also collected.  
659 Collection time and GPS coordinate in gpx format of host plants was recorded on the  
660 day of collection. Leaves, flowers of host plants were photographed. Bioclimatic  
661 variables of sampling sites were retrieved from CHELSA(Karger et al. 2017) database  
662 (v1.2) using recorded GPS coordinates. Digital terrain models (DTM) of sampling  
663 sites were retrieved from Taiwan’s Open Government Data website  
664 (<https://data.gov.tw/dataset/35430>). Environmental samples were collected using  
665 alcohol sterilized tweezers or spoons and stored in zip bags. Whenever possible  
666 during the sampling trips metadata such as the identity of the host plant, lichens and  
667 altitude were recorded. Samples were redistributed into 50ml falcon tubes and stored

668 at room temperature. Each sample was divided into two proportions and immersed in  
669 two enrichment media: a liquid medium made up of either i) 3 g/L yeast extract, 3 g/L  
670 malt extract, 5 g/L peptone, 10 g/L sucrose, 7.6% EtOH, 1 mg/L chloramphenicol,  
671 and 0.1 % of 1-M HCl as used in ref(Sniegowski et al. 2002) or ii) YPD containing  
672 10% dextrose and 5% ethanol adjusted to pH 5.3 as used in (Hyma and Fay 2013).  
673 Samples were incubated at 30°C until signs of microbial growth and fermentation  
674 were detected, such as white sediment and effervescence. Sediments were then  
675 streaked onto YPD agar plates. Single colonies were picked out and incubated in  
676 potassium acetate medium 23°C for 7-10 days(Liti et al. 2017). Single colonies with  
677 ascus-like (four spores) structures under microscope were picked out and streaked  
678 onto YPD agar plates. Sanger sequencing and gel electrophoresis of ITS1-5.8S-ITS2  
679 region PCR amplified with ITS1F/ITS4 primer set were performed to identify the  
680 species of isolates(White et al. 1990; Gardes and Bruns 1993). Pilot sampling,  
681 modification and rationale during the course of sampling strategies are further  
682 provided in **Supplementary Info**. Sampling efforts were visualized using the R's  
683 package ggplot2 (v.3.3.5) and annotated with metR (v.0.10.0;  
684 <https://github.com/eliocamp/metR>) and ggspatial (v.1.1.5;  
685 <https://paleolimbot.github.io/ggspatial/>). In order to determine ploidy levels for our  
686 isolates, we carried out flow cytometry analysis for the 105 Taiwanese isolates from  
687 this study using propidium iodide (PI) staining assay using previously established  
688 protocols (Todd et al. 2018) (**Supplementary Info**).

689

### 690 **DNA extraction**

691 Field-collected environmental samples can vary, so we preprocessed these  
692 samples and extracted their DNA differently (see **Supplementary Info** for details). For  
693 whole genome sequencing of *Saccharomyces cerevisiae*, isolates taken from frozen  
694 stocks were streaked out onto YPD plates and incubated in 30°C until colonies became  
695 visible. Single colonies were then incubated in 5ml YPD liquid medium at 30°C in a  
696 shaker at 200 rpm overnight. High molecular weight genomic DNA was extracted using  
697 protocol described in (Denis et al. 2018). DNA quality was determined by Qubit  
698 readings, A260, A280, A260/280 ratios on NanoDrop and gel electrophoresis.

699

### 700 **Library construction and whole-genome sequencing**

701 For Illumina sequencing, paired-end libraries were constructed using the Illumina  
702 Nextera or NEB Next Ultra DNA library preparation kit with the manufacturer's

703 protocol. The first 91 isolates were sequenced by Illumina HiSeq 2500 and the  
704 remaining 30 were sequenced by Novaseq to produce 125- and 150-bp paired-end reads,  
705 respectively. Oxford Nanopore libraries were prepared using SQK-LSK109 with 12  
706 isolates multiplexed by EXP-NBD104 and EXP-NBD114 barcoding kit (ver  
707 NBE\_9065\_v109\_revV\_14Aug2019) and sequenced by a R9.4.1 flow cell on a  
708 GridION instrument. A total of 24 isolates were run on two flow cells. Nanopore fast5  
709 files were basecalled using Guppy (v4.0.11).

710

### 711 **Amplicon sequencing and analysis**

712 Amplicon libraries were constructed as previously described(Tedersoo et al.  
713 2014) from 89 environmental samples (18 bark, 18 twig, 18 leaf, 18 litter, 17 soil),  
714 three positive controls (*Saccharomyces cerevisiae* S288C, *S. paradoxus* YDG197 and  
715 lab isolate *Pseudocercospora fraxinii*); and DNA from two *Escherichia coli* as a  
716 template to confirm primer specificity towards only fungal species. The ITS3ngs (5'-  
717 CANGGATGAAGAACGYRG-3') and ITS4ngsUni (5'-  
718 CCTSCSCTTANTDATATGC-3') primer pair(Tedersoo et al. 2015) was used. Two  
719 no template controls were included during the PCR step to confirm that amplicon  
720 generation was free of contaminating DNA. To determine the background amplicon  
721 noise from experimental pipeline, a sterile filter was treated and processed as one the  
722 field samples. Amplicons were normalized using the SequalPrep™ Normalization  
723 Plate Kit (Thermo Fisher Scientific, ID: A1051001), then pooled and concentrated  
724 using AMPure XP (Beckman Coulter, ID: A63881). Finished DNA libraries were  
725 sequenced on the Illumina MiSeq platform using 2x300 bp pair-end sequencing  
726 chemistry.

727

728 Raw sequencing reads containing the Illumina sequencing index were  
729 demultiplexed using *sabre* (v1.0; <https://github.com/najoshi/sabre>). Sequencing  
730 quality was determined using *FastQC* (v0.11.7; [https://github.com/s-](https://github.com/s-andrews/FastQC)  
731 [andrews/FastQC](https://github.com/s-andrews/FastQC)). Reads were quality filtered based on a Qscore >20 and 50 base  
732 pairs were trimmed from the 3' end end using *usearch*(Edgar 2010) (v11.0.667).  
733 Filtered reads were processed following the UPARSE (Edgar 2013) pipeline. In brief,  
734 paired reads were merged and dereplicated into unique sequences. Unique sequences  
735 were filtered using *usearch* default settings. Filtered sequences were denoised into  
736 zero-radius operational taxonomic unit (zOTUs) using the *unoise2* (Edgar 2016b)  
737 algorithm. The taxonomy of zOTUs was classified using the *SINTAX* (Edgar 2016a)

738 algorithm (Edgar 2016) against the UNITE (Nilsson et al. 2018) Fungal database  
739 (v8.2). Merged reads were assigned into zOTUs with 100% sequence identity and  
740 tabulated using the *usearch\_global* function. Processed reads were analyzed in the R-  
741 Studio environment (v 1.2.5033). Sequencing data were analyzed with  
742 phyloseq(McMurdie and Holmes 2013) (v1.34). Statistical significance was tested for  
743 using *kruskal.test* from the *stats* package in R (R Core Team 2021).

744

## 745 **Variant calling**

746 To determine the evolutionary history of new Taiwanese isolates, we collected a  
747 total of 219 published genomes representing established *S. cerevisiae* industrial and  
748 natural populations: 102 isolates from the 1,011 genome project(Peter et al. 2018) (31  
749 Wine/European, 8 Mediterranean oak, 6 African beer, 6 African palm wine, 4 West  
750 African cocoa, 4 Malaysian bertam palm nectar, 6 North American oak, 6 Sake, 11  
751 Asian fermentation, 1 CHN-I, 1 CHN-III, 4 CHN-IV, 1 CHN-V, 6 Mixed origin  
752 groups and 7 other isolates of Taiwanese origin), 93 isolates from the Chinese  
753 population(Duan et al. 2018) (69 CHN-I to CHN-X isolates excluding those  
754 previously sequenced in the 1,011 genome project, 5 isolate from Mantou1, 6  
755 Huaugjiu, 7 Baijiu and 6 Qingkejiu), 16 isolates from the Brazilian wild  
756 lineage(Barbosa et al. 2016) and eight isolates from olive brine(Pontes et al. 2019).  
757 This combined with the 121 isolates from this study yielded a total of 340 individuals,  
758 30% of which originated from industrial sources and 70% from the natural  
759 environment (**Supplemental Table S6**). Read quality was examined with FastQC  
760 (v.0.11.9; <https://github.com/s-andrews/FastQC>). Read quality and adaptor trimming  
761 was performed using Trimmomatic(Bolger et al. 2014) (v0.36; Pair end mode,  
762 ILLUMINACLIP;LEADING:20;TRAILING:20;SLIDINGWINDOW:4:20;MINLEN:  
763 150). For the 340 samples, 64-95% of raw paired reads from the 340 samples were  
764 kept after trimming. Trimmed reads were each mapped to the S288C reference  
765 genome version R64-2-1 using Burrows-Wheeler Aligner(Li and Durbin 2009) (v  
766 0.7.17-r1188) and the mapping rate was 91-99%. Duplicate reads were marked using  
767 GATK (McKenna et al. 2010) MarkDuplicates (v.4.1.9.0) . Variants were first called  
768 in multi-sample manner and filtered using BCFtools(Danecek et al. 2021) v1.8 (-d  
769 1332; QUAL 30, MQ 30, AC >=2 and 50% missingness; genotype-filtered with  
770 minDP 3). 88% (1,150,658/1,306,082) of variants were retained. Second, variants  
771 were also called and filtered with FreeBayes (Garrison and Marth 2012) and  
772 VCFtools (Danecek et al. 2011) (v. 1.3.2 and v 0.1.15, respectively; minDP 3, QUAL

773 30, MQ 30, AC  $\geq 2$  and 50% missingness, sites with  $0.25 < AB < 0.75$  and  
774  $0.9 < MQM/MQMR < 1.05$  were retained). 56% of sites were retained based on these  
775 criteria (818,025/1,443,685). Finally, 808,864 intersecting variants discovered from  
776 both callers were used for further analysis. The functional effects of variants were  
777 annotated with SnpEff(Cingolani et al. 2012) (v.4.3t).

778

### 779 **Assembly, annotation and ortholog identification**

780 Nanopore reads of each isolate were assembled using Canu (Koren et al. 2017)  
781 (v1.9). For isolates without long reads, Illumina paired-end reads were assembled using  
782 SPAdes (Bankevich et al. 2012) (v. 3.14.1, options  $k$ -mer size 21, 33, 55, 77 and --  
783 careful). Consensus sequences of the assemblies were polished with four rounds of  
784 Racon (Vaser et al. 2017) (v1.4.11), one round of Medaka (v1.0.1) using nanopore raw  
785 reads and five rounds of Pilon using Illumina reads. The assemblies were further  
786 scaffolded using RagTag (Alonge et al. 2019) against the S288C genome reference.  
787 Annotations were then transferred using Liftoff (Shumate et al. 2021), with additional  
788 *de novo* annotations using AUGUSTUS (Stanke et al. 2006) on regions without any  
789 transferred annotations. Orthogroup was inferred using OrthoFinder (Emms and Kelly  
790 2015) (v.2.5.4). Orthogroup that differentially abundant between assemblies produced  
791 using different sequencing technologies were excluded from further analyses. The  
792 assembly metrics and description of the nanopore assemblies are shown in  
793 **Supplemental Table S8**.

794

### 795 **Phylogenomic analyses**

796 After removing 43,695 invariant sites resulting from ambiguous nucleotide codes  
797 among all isolates, the remaining 765,169 variable sites were used to construct a  
798 phylogeny for the 340 isolates. The resulting best-fit model was indicated by BIC to be  
799 TVMe+R3 first with IQ-TREE. In addition, a maximum likelihood phylogeny was  
800 inferred using IQ-TREE with the TVMe+R3+ASC model and a 1,000 ultrafast  
801 bootstrap approximation (Hoang et al. 2018; Minh et al. 2020). A separate *S. cerevisiae*  
802 lineage phylogeny was inferred and used in MCMCtree method of the PAML (Yang  
803 2007) package to estimate the divergence time among the *S. cerevisiae* lineages  
804 **(Supplementary Info)**.

805

### 806 **Diversity, population structure and demography estimates**

807 For the population structure estimate, biallelic SNPs were kept and filtered based  
808 on linkage disequilibrium. Sites that are linked were filtered out using PLINK(Chang  
809 et al. 2015) (v1.90b4), excluding pairs of loci with  $r^2 > 0.5$  (--indep-pairwise 50 10 0.5  
810 --r2). The remaining 482,161 sites were used for ancestry estimation by ADMIXTURE  
811 (Alexander et al. 2009) using K=2 to K=30 with five-fold cross validation from five  
812 runs of different seed numbers. CV errors for each K value in five runs were compared  
813 to choose the representative number of clusters. Migration signals on the phylogeny  
814 were estimated with TreeMix using 1000 bootstrap for natural populations according  
815 to clusters in K=16. The numbers of migration edges were estimated, aided by the optM  
816 (v. 0.1.5) package (Fitak 2021) and presented in **Supplementary Info**.

817 A consensus genome sequence containing variants for each isolate was generated  
818 from the SNPs matrix using BCFtools (Danecek et al. 2021) consensus (v.R64-2-1)  
819 with the S288C reference genome sequence. For the IBD analysis, geographical  
820 distance between isolates was measured using the sf package in R for Taiwanese  
821 isolates with GPS records. For Chinese isolates, since GPS records were not available,  
822 we used approximate coordinates for each sample site (Duan et al. 2018) as  
823 recommended by the authors. To estimate the maximum geographical distance within  
824 the Chinese lineage, we chose sample sites that were the furthest apart. For instance, in  
825 CHN-V, the distance between Shanxi and Hainan was used. For lineages sampled from  
826 only one site (CHN-II, CHN-IX), the largest range of the site was used. Diversity  
827 estimates for 16 nuclear chromosomes and corresponding coding/noncoding regions  
828 were examined by VariScan (Vilella et al. 2005) with RunMode 11 ( $n < 4$ ) and 12 ( $n \geq 4$ ).  
829 These diversity estimates were used to infer frequency of sex according to (Tsai et al.  
830 2008) and detailed in **Supplementary Info**.

831

## 832 DATA ACCESS

833 The sequencing data of the 121 *S. cerevisiae* isolates and ITS amplicon sequences  
834 of 89 samples generated in this study have been submitted to the NCBI BioProject  
835 database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number  
836 PRJNA755173. The accession numbers of the isolates are also shown in **Supplemental**  
837 **Table S6**. The zOTU table for the amplicon data and all of the scripts written to carry  
838 out this study was deposited in <https://github.com/tjleez/popgen.methods> and are also  
839 available as **Supplemental Code**.

840

841

842 **COMPETING INTEREST STATEMENT**

843 The authors declare no competing interests.

844

845 **ACKNOWLEDGMENTS**

846 We thank Cheng-Ruei Lee for the insightful comments on the ADMIXTURE  
847 analyses. We thank Mao-Ning Tuamu for the helpful suggestions on how to deal  
848 with bioclimatic variable data. We thank Nguyen Huu-Vang, Cheng-Ruei Lee, Jun-Yi  
849 Leu, Ben-Yang Liao, Dang Liu and John Wang for commenting on earlier versions of  
850 the manuscript. We thank Jun-Yi Leu for the experimental advice. We thank Bo-Fei  
851 Chen and Ling-Ting Kao for helping with the initial sampling trips. We thank Tze-Fu  
852 Hsu, Yi-Hsiu Kuan, H. Thorsten Lumbsch, Matthew Nelsen, for collecting/providing  
853 some of the biomaterials. We thank Shou-Fu Duan and Feng-Yan Bai for providing  
854 early access to sequencing data and recommendations on how to approximate the  
855 geographical positions of the Chinese isolates. We are grateful to the National Center  
856 for High-Performance Computing for its computer time and for letting us use its  
857 facilities. I.J.T. was supported by the Ministry of Science and Technology, Taiwan  
858 under grant 110-2628-B-001-027 and Career Development Award AS-CDA-107-L01,  
859 Academia Sinica.

860

861 **AUTHORS CONTRIBUTION**

862 I.J.T. conceived and led the study. T.J.L, Y.C.L and W.A.L carried out the  
863 sampling and isolation of *Saccharomyces cerevisiae*. J.P.H, C.L.H and K.F.C helped  
864 with the sampling and identified the lichen and plant samples. T.J.L, W.A.L, Y.F.L and  
865 H.M.K conducted the experiments. Y.F.L carried out the amplicon analyses. H.H.L.,  
866 H.M.K and I.J.T. performed the sequencing and assemblies of the *S. cerevisiae*  
867 genomes. T.J.L, Y.C.L, H.H.L and I.J.T. carried out the population genomic analyses.  
868 Y.C.L, H.H.L and I.J.T. carried out the comparative genomics, phylogenomics analyses  
869 and the divergence time estimation. T.J.L and I.J.T. wrote the manuscript with  
870 substantial input from J.P.H, K.F.C and G.L.

871

872

873

874

875

876 **REFERENCES**

877

- 878 Alexander DH, Novembre J, Lange K. 2009. Fast Model-Based Estimation of  
879 Ancestry in Unrelated Individuals. *Genome research* **19**: 1655-1664.
- 880 Ali JR. 2018. Islands as biological substrates: Continental. *Journal of Biogeography*  
881 **45**: 1003-1018.
- 882 Almeida P, Barbosa R, Zalar P, Imanishi Y, Shimizu K, Turchetti B, Legras J-L, Serra  
883 M, Dequin S, Couloux A et al. 2015. A population genomics insight into the  
884 Mediterranean origins of wine yeast domestication. *Molecular Ecology* **24**:  
885 5412-5427.
- 886 Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman  
887 ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding  
888 of draft genomes. *Genome Biology* **20**.
- 889 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,  
890 Nikolenko SI, Pham S, Prjibelski AD et al. 2012. SPAdes: A New Genome  
891 Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal*  
892 *of Computational Biology* **19**: 455-477.
- 893 Barbosa R, Almeida P, Safar SV, Santos RO, Morais PB, Nielly-Thibault L, Leducq  
894 JB, Landry CR, Goncalves P, Rosa CA et al. 2016. Evidence of Natural  
895 Hybridization in Brazilian Wild Lineages of *Saccharomyces cerevisiae*.  
896 *Genome Biol Evol* **8**: 317-329.
- 897 Barnett JA. 1992. The taxonomy of the genus *Saccharomyces meyenex reess*: A short  
898 review for non-taxonomists. *Yeast* **8**: 1-23.
- 899 Bendixsen DP, Gettle N, Gilchrist C, Zhang Z, Stelkens R. 2021. Genomic Evidence  
900 of an Ancient East Asian Divergence Event in Wild *Saccharomyces cerevisiae*.  
901 *Genome Biol Evol* **13**.
- 902 Bergström A, Simpson JT, Salinas F, Barré B, Parts L, Zia A, Nguyen Ba AN, Moses  
903 AM, Louis EJ, Mustonen V et al. 2014. A High-Definition View of Functional  
904 Genetic Variation from Natural Yeast Genomes. *Molecular Biology and*  
905 *Evolution* **31**: 872-888.
- 906 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina  
907 Sequence Data. *Bioinformatics (Oxford, England)*  
908 doi:10.1093/bioinformatics/btu170: 1-7.
- 909 Branco S, Bi K, Liao HL, Gladieux P, Badouin H, Ellison CE, Nguyen NH, Vilgalys  
910 R, Peay KG, Taylor JW et al. 2017. Continental-level population

- 911 differentiation and environmental adaptation in the mushroom *Suillus*  
912 *brevipes*. *Mol Ecol* **26**: 2063-2076.
- 913 Brysch-Herzberg M, Seidel M. 2017. Distribution patterns of *Saccharomyces* species  
914 in cultural landscapes of Germany. *FEMS Yeast Res* **17**.
- 915 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-  
916 generation PLINK: rising to the challenge of larger and richer datasets.  
917 *Gigascience* **4**: 7.
- 918 Charron G, Leducq JB, Bertin C, Dube AK, Landry CR. 2014. Exploring the northern  
919 limit of the distribution of *Saccharomyces cerevisiae* and *Saccharomyces*  
920 *paradoxus* in North America. *FEMS Yeast Res* **14**: 281-288.
- 921 Chiang T-Y, Schaal BA. 2006. Phylogeography of plants in Taiwan and the Ryukyu  
922 Archipelago. *Taxon* **55**: 31-41.
- 923 Chung CL, Lee TJ, Akiba M, Lee HH, Kuo TH, Liu D, Ke HM, Yokoi T, Roa MB, Lu  
924 MJ et al. 2017. Comparative and population genomic landscape of *Phellinus*  
925 *noxius*: A hypervariable fungus causing root rot in trees. *Mol Ecol*  
926 doi:10.1111/mec.14359.
- 927 Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden  
928 DM. 2012. A program for annotating and predicting the effects of single  
929 nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila*  
930 *melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**: 80-92.
- 931 Cromie GA, Hyma KE, Ludlow CL, Garmendia-torres C, Gilbert TL, May P, Huang  
932 AA, Dudley AM, Fay JC. 2013. Genomic Sequence Diversity and Population  
933 Structure of *Saccharomyces cerevisiae* Assessed by RAD-seq. *G3 (Bethesda,*  
934 *Md)* **3**: 2163-2171.
- 935 Crowther TW, Glick HB, Covey KR, Bettigole C, Maynard DS, Thomas SM, Smith  
936 JR, Hintler G, Duguid MC, Amatulli G et al. 2015. Mapping tree density at a  
937 global scale. *Nature* **525**: 201-205.
- 938 D'Angiolo M, De Chiara M, Yue J-X, Irizar A, Stenberg S, Persson K, Llored A,  
939 Barré B, Schacherer J, Marangoni R et al. 2020. A yeast living ancestor  
940 reveals the origin of genomic introgressions. *Nature* **587**: 420-425.
- 941 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,  
942 Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and  
943 VCFtools. *Bioinformatics* **27**: 2156-2158.
- 944 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A,  
945 Keane T, McCarthy SA, Davies RM et al. 2021. Twelve years of SAMtools

- 946 and BCFtools. *Gigascience* **10**.
- 947 Dashko S, Liu P, Volk H, Butinar L, Piskur J, Fay JC. 2016. Changes in the Relative  
948 Abundance of Two *Saccharomyces* Species from Oak Forests to Wine  
949 Fermentations. *Front Microbiol* **7**: 215.
- 950 Davison J, Moora M, Öpik M, Ainsaar L, Ducouso M, Hiiesalu I, Jairus T, Johnson  
951 N, Jourand P, Kalamees R et al. 2018. Microbial island biogeography:  
952 isolation shapes the life history characteristics but not diversity of root-  
953 symbiotic fungal communities. *The ISME Journal* **12**: 2211-2224.
- 954 Denis E, Denis E, Sanchez S, Mairey B, Beluche O, Cruaud C, Lemainque A,  
955 Wincker P, Barbe V. 2018. Extracting high molecular weight genomic DNA  
956 from *Saccharomyces cerevisiae*. *Protocol Exchange*  
957 doi:10.1038/protex.2018.076.
- 958 Diezmann S, Dietrich FS. 2009. *Saccharomyces cerevisiae*: population divergence  
959 and resistance to oxidative stress in clinical, domesticated and wild isolates.  
960 *PloS one* **4**: e5317.
- 961 Duan SF, Han PJ, Wang QM, Liu WQ, Shi JY, Li K, Zhang XL, Bai FY. 2018. The  
962 origin and adaptive evolution of domesticated populations of yeast from Far  
963 East Asia. *Nat Communications* **9**: 2690.
- 964 Eberlein C, Henault M, Fijarczyk A, Charron G, Bouvier M, Kohn LM, Anderson JB,  
965 Landry CR. 2019. Hybridization is a recurrent evolutionary stimulus in wild  
966 yeast speciation. *Nat Commun* **10**: 923.
- 967 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.  
968 *Bioinformatics (Oxford, England)* **26**: 2460-2461.
- 969 Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon  
970 reads. *Nature Methods* **647**: 1-5.
- 971 Edgar RC. 2016a. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and  
972 ITS sequences. *bioRxiv* doi:10.1101/074161.
- 973 Edgar RC. 2016b. UNOISE2: improved error-correction for Illumina 16S and ITS  
974 amplicon sequencing. *bioRxiv* doi:10.1101/081257.
- 975 Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome  
976 comparisons dramatically improves orthogroup inference accuracy. *Genome*  
977 *Biol* **16**: 157.
- 978 Farlow A, Long H, Arnoux S, Sung W, Doak TG, Nordborg M, Lynch M. 2015. The  
979 Spontaneous Mutation Rate in the Fission Yeast *Schizosaccharomyces pombe*.  
980 *Genetics* **201**: 737-744.

- 981 Fay JC, Benavides Ja. 2005. Evidence for domesticated and wild populations of  
982 *Saccharomyces cerevisiae*. *PLoS genetics* **1**: 66-71.
- 983 Fitak RR. 2021. OptM: estimating the optimal number of migration edges on  
984 population trees using Treemix. *Biology Methods and Protocols* **6**.
- 985 Fournier T, Abou Saada O, Hou J, Peter J, Caudal E, Schacherer J. 2019. Extensive  
986 impact of low-frequency variants on the phenotypic landscape at population-  
987 scale. *eLife* **8**.
- 988 Gallone B, Mertens S, Gordon JL, Maere S, Verstrepen KJ, Steensels J. 2018. Origins,  
989 evolution, domestication and diversity of *Saccharomyces* beer yeasts. *Current*  
990 *Opinion in Biotechnology* **49**: 148-155.
- 991 Gallone B, Steensels J, Baele G, Maere S, Verstrepen KJ, Prah T, Soriaga L, Saels V,  
992 Herrera-Malaver B, Merlevede A et al. 2016. Domestication and Divergence  
993 of *Saccharomyces cerevisiae* Beer Yeasts. *Cell* **166**: 1397-1410.e1316.
- 994 Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes  
995 - application to the identification of mycorrhizae and rusts. *Molecular Ecology*  
996 **2**: 113-118.
- 997 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read  
998 sequencing. *arXiv* **1207**: 3907.
- 999 Goddard MR, Godfray HCJ, Burt A. 2005. Sex increases the efficacy of natural  
1000 selection in experimental yeast populations. *Nature* **434**: 636-640.
- 1001 Goddard MR, Greig D. 2015. *Saccharomyces cerevisiae*: a nomadic yeast with no  
1002 niche? *FEMS Yeast Res* **15**.
- 1003 Gonçalves M, Pontes A, Almeida P, Barbosa R, Serra M, Libkind D, Hutzler M,  
1004 Gonçalves P, Sampaio JP. 2016. Distinct Domestication Trajectories in Top-  
1005 Fermenting Beer Yeasts and Wine Yeasts. *Current Biology* **26**: 2750-2761.
- 1006 Gray JV, Petsko GA, Johnston GC, Ringe D, Singer RA, Werner-Washburne M. 2004.  
1007 “Sleeping Beauty”: Quiescence in *Saccharomyces cerevisiae*. *Microbiology*  
1008 *and Molecular Biology Reviews* **68**: 187-206.
- 1009 Han D-Y, Han P-J, Rumbold K, Koricha AD, Duan S-F, Song L, Shi J-Y, Li K, Wang  
1010 Q-M, Bai F-Y. 2021. Adaptive Gene Content and Allele Distribution  
1011 Variations in the Wild and Domesticated Populations of *Saccharomyces*  
1012 *cerevisiae*. *Frontiers in Microbiology* **12**.
- 1013 He PY, Shao XQ, Duan SF, Han DY, Li K, Shi JY, Zhang RP, Han PJ, Wang QM, Bai  
1014 FY. 2022. Highly diverged lineages of *Saccharomyces paradoxus* in temperate  
1015 to subtropical climate zones in China. *Yeast* doi:10.1002/yea.3688.

- 1016 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2:  
1017 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**: 518-522.
- 1018 Hsieh C. 2002. Composition, Endemism and Phytogeographical Affinities of the  
1019 Taiwan Flora. *TAIWANIA* **47**: 298-310.
- 1020 Hyma KE, Fay JC. 2013. Mixing of vineyard and oak-tree ecotypes of  
1021 *Saccharomyces cerevisiae* in North American vineyards. *Mol Ecol* **22**: 2917-  
1022 2930.
- 1023 Jianfei Y, Zhiduan C, Bing L, Haining Q, Yong Y. 2013. Disjunct distribution of  
1024 vascular plants between southwestern area and Taiwan area in China.  
1025 *Biodiversity Science* **20**: 482-494.
- 1026 Jiang X-L, Gardner EM, Meng H-H, Deng M, Xu G-B. 2019. Land bridges in the  
1027 Pleistocene contributed to flora assembly on the continental islands of South  
1028 China: Insights from the evolutionary history of *Quercus championii*.  
1029 *Molecular Phylogenetics and Evolution* **132**: 36-45.
- 1030 Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann  
1031 NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the  
1032 earth's land surface areas. *Scientific Data* **4**.
- 1033 Karger DN, Kessler M, Lehnert M, Jetz W. 2021. Limited protection and ongoing loss  
1034 of tropical cloud forest biodiversity and ecosystems worldwide. *Nature*  
1035 *Ecology & Evolution* doi:10.1038/s41559-021-01450-y.
- 1036 Kasavi C, Eraslan S, Arga KY, Oner ET, Kirdar B. 2014. A system based network  
1037 approach to ethanol tolerance in *Saccharomyces cerevisiae*. *BMC Syst Biol* **8**:  
1038 90.
- 1039 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu:  
1040 scalable and accurate long-read assembly via adaptive k-mer weighting and  
1041 repeat separation. *Genome Res* **27**: 722-736.
- 1042 Kowallik V, Greig D. 2016. A systematic forest survey showing an association of  
1043 *Saccharomyces paradoxus* with oak leaf litter. *Environ Microbiol Rep* **8**: 833-  
1044 841.
- 1045 Kowallik V, Miller E, Greig D. 2015. The interaction of *Saccharomyces paradoxus*  
1046 with its natural competitors on oak bark. *Mol Ecol* **24**: 1596-1610.
- 1047 Leducq JB, Nielly-Thibault L, Charron G, Eberlein C, Verta JP, Samani P, Sylvester  
1048 K, Hittinger CT, Bell G, Landry CR. 2016. Speciation driven by hybridization  
1049 and chromosomal plasticity in a wild yeast. *Nat Microbiol* **1**: 15003.
- 1050 Legras J-I, Galeote V, Bigey F, Camarasa C, Marsit S, Nidelet T, Sanchez I, Couloux

- 1051 A, Guy J, Franco-duarte R et al. 2018. Adaptation of *S. cerevisiae* to  
1052 fermented food environments reveals remarkable genome plasticity and the  
1053 footprints of domestication. doi:10.1093/molbev/msy066.
- 1054 Li C-F, Chytrý M, Zelený D, Chen M-Y, Chen T-Y, Chiou C-R, Hsia Y-J, Liu H-Y,  
1055 Yang S-Z, Yeh C-L et al. 2013. Classification of Taiwan forest vegetation.  
1056 *Applied Vegetation Science* **16**: 698-719.
- 1057 Li G, Figueiró HV, Eizirik E, Murphy WJ, Yoder A. 2019. Recombination-Aware  
1058 Phylogenomics Reveals the Structured Genomic Landscape of Hybridizing  
1059 Cat Species. *Molecular Biology and Evolution* **36**: 2111-2126.
- 1060 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler  
1061 transform. *Bioinformatics* **25**: 1754-1760.
- 1062 Li H, Durbin R. 2011. Inference of human population history from individual whole-  
1063 genome sequences. *Nature* **475**: 493-496.
- 1064 Lin C-T, Chung K-F. 2017. Phylogenetic Classification of Seed Plants of Taiwan.  
1065 *Botanical Studies* **58**.
- 1066 Lin HY, Li CF, Chen TY, Hsieh CF, Wang G, Wang T, Hu JM, Ohlemuller R. 2020.  
1067 Climate-based approach for modeling the distribution of montane forest  
1068 vegetation in Taiwan. *Applied Vegetation Science* **23**: 239-253.
- 1069 Liti G. 2015. The fascinating and secret wild life of the budding yeast *S. cerevisiae*.  
1070 *eLife* **4**.
- 1071 Liti G, Barton DBH, Louis EJ. 2006. Sequence diversity, reproductive isolation and  
1072 species concepts in *Saccharomyces*. *Genetics* **174**: 839-850.
- 1073 Liti G, Carter DM, Moses AM, Warringer J, Parts L, James Sa, Davey RP, Roberts IN,  
1074 Burt A, Koufopanou V et al. 2009. Population genomics of domestic and wild  
1075 yeasts. *Nature* **458**: 337-341.
- 1076 Liti G, Warringer J, Blomberg A. 2017. Isolation and Laboratory Domestication of  
1077 Natural Yeast Strains. *Cold Spring Harbor Protocols* **2017**.
- 1078 Long A, Liti G, Luptak A, Tenailon O. 2015. Elucidating the molecular architecture  
1079 of adaptation via evolve and resequence experiments. *Nature Reviews*  
1080 *Genetics* **16**: 567-582.
- 1081 Magri D, Di Rita F, Aranbarri J, Fletcher W, González-Sampériz P. 2017. Quaternary  
1082 disappearance of tree taxa from Southern Europe: Timing and trends.  
1083 *Quaternary Science Reviews* **163**: 23-55.
- 1084 Magwene PM, Kayıkçı Ö, Granek Ja, Reininga JM, Scholl Z, Murray D. 2011.  
1085 Outcrossing, mitotic recombination, and life-history trade-offs shape genome

- 1086 evolution in *Saccharomyces cerevisiae*. *Proceedings of the National Academy*  
1087 *of Sciences of the United States of America* **108**: 1987-1992.
- 1088 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in  
1089 *Drosophila*. *Nature* **351**: 652-654.
- 1090 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella  
1091 K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit:  
1092 a MapReduce framework for analyzing next-generation DNA sequencing data.  
1093 *Genome research* **20**: 1297-1303.
- 1094 McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive  
1095 analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.
- 1096 Meirans PG. 2012. The trouble with isolation by distance. *Molecular Ecology* **21**:  
1097 2839-2846.
- 1098 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Michael D, Haeseler VA,  
1099 Lanfear R. 2020. IQ-TREE 2 : New Models and Efficient Methods for  
1100 Phylogenetic Inference in the Genomic Era. doi:10.1093/molbev/msaa015: 6-  
1101 10.
- 1102 Mozzachiodi S, Bai FY, Baldrian P, Bell G, Boundy-Mills K, Buzzini P, Čadež N,  
1103 Rizzo FC, Dashko S, Dimitrov R et al. 2022. Yeasts from temperate forests.  
1104 *Yeast* doi:10.1002/yea.3699.
- 1105 Naumov GI, Lee CF, Naumova ES. 2013. Molecular genetic diversity of the  
1106 *Saccharomyces* yeasts in Taiwan: *Saccharomyces arboricola*, *Saccharomyces*  
1107 *cerevisiae* and *Saccharomyces kudriavzevii*. *Antonie van Leeuwenhoek*,  
1108 *International Journal of General and Molecular Microbiology* **103**: 217-228.
- 1109 Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D,  
1110 Kennedy P, Picard K, Glockner FO, Tedersoo L et al. 2018. The UNITE  
1111 database for molecular identification of fungi: handling dark taxa and parallel  
1112 taxonomic classifications. *Nucleic Acids Res* doi:10.1093/nar/gky1022.
- 1113 Niu Y-T, Ye J-F, Zhang J-L, Wan J-Z, Yang T, Wei X-X, Lu L-M, Li J-H, Chen Z-D.  
1114 2018. Long-distance dispersal or postglacial contraction? Insights into  
1115 disjunction between Himalaya-Hengduan Mountains and Taiwan in a cold-  
1116 adapted herbaceous genus, *Triplostegia*. *Ecology and Evolution* **8**: 1131-1146.
- 1117 Peter J, De Chiara M, Friedrich A, Yue JX, Pflieger D, Bergstrom A, Sigwalt A, Barre  
1118 B, Freel K, Llored A et al. 2018. Genome evolution across 1,011  
1119 *Saccharomyces cerevisiae* isolates. *Nature* **556**: 339-344.
- 1120 Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from

- 1121 genome-wide allele frequency data. *PLoS Genet* **8**: e1002967.
- 1122 Pontes A, Čadež N, Gonçalves P, Sampaio JP. 2019. A Quasi-Domestic Relic  
1123 Hybrid Population of *Saccharomyces cerevisiae* × *S. paradoxus* Adapted to  
1124 Olive Brine. *Frontiers in Genetics* **10**.
- 1125 R Core Team. 2021. R: A Language and Environment for Statistical Computing. R  
1126 Foundation for Statistical Computing, Vienna, Austria.
- 1127 Robinson HA, Pinharanda A, Bensasson D. 2016. Summer temperature can predict  
1128 the distribution of wild yeast populations. *Ecol Evol* **6**: 1236-1250.
- 1129 Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama  
1130 T, Petrov DA, Schmidt P. 2022. Direct observation of adaptive tracking on  
1131 ecological time scales in *Drosophila*. *Science* **375**.
- 1132 Sampaio JP, Gonçalves P. 2008. Natural populations of *Saccharomyces kudriavzevii*  
1133 in Portugal are associated with Oak bark and are sympatric with *S. cerevisiae*  
1134 and *S. paradoxus*. *Applied and Environmental Microbiology* **74**: 2144-2152.
- 1135 Schierup MH, Hein J. 2000. Consequences of recombination on traditional  
1136 phylogenetic analysis. *Genetics* **156**: 879-891.
- 1137 Seersholm FV, Werndly DJ, Grealy A, Johnson T, Keenan Early EM, Lundelius EL,  
1138 Winsborough B, Farr GE, Toomey R, Hansen AJ et al. 2020. Rapid range  
1139 shifts and megafaunal extinctions associated with late Pleistocene climate  
1140 change. *Nature Communications* **11**.
- 1141 Shumate A, Salzberg SL, Valencia A. 2021. Liftoff: accurate mapping of gene  
1142 annotations. *Bioinformatics* **37**: 1639-1643.
- 1143 Sniegowski PD, Dombrowski PG, Fingerman E. 2002. *Saccharomyces cerevisiae* and  
1144 *Saccharomyces paradoxus* coexist in a natural woodland site in North America  
1145 and display different levels of reproductive isolation from European  
1146 conspecifics. *FEMS Yeast Res* **1**: 299-306.
- 1147 Spribille T, Tuovinen V, Resl P, Vanderpool D, Wolinski H, Aime MC, Schneider K,  
1148 Stabentheiner E, Toome-Heller M, Thor G et al. 2016. Basidiomycete yeasts in  
1149 the cortex of ascomycete macrolichens. *Science* **353**: 488-492.
- 1150 Stanke M, Tzvetkova A, Morgenstern B. 2006. AUGUSTUS at EGASP: using EST,  
1151 protein and genomic alignments for improved gene prediction in the human  
1152 genome. *Genome biology* **7 Suppl 1**: S11.11-18.
- 1153 Stoletzki N, Eyre-Walker A. 2010. Estimation of the Neutrality Index. *Molecular*  
1154 *Biology and Evolution* **28**: 63-70.
- 1155 Strobe PK, Skelly Da, Kozmin SG, Mahadevan G, Stone Ea, Magwene PM, Dietrich

- 1156 FS, Mccusker JH, Carolina N, Sciences B et al. 2015. The 100-genomes  
1157 strains , an *S . cerevisiae* resource that illuminates its natural phenotypic and  
1158 genotypic variation and emergence as an opportunistic pathogen.  
1159 doi:10.1101/gr.185538.114.: 1-13.
- 1160 Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V,  
1161 Nilsson H, Hildebrand F et al. 2015. Shotgun metagenomes and multiple  
1162 primer pair-barcode combinations of amplicons reveal biases in  
1163 metabarcoding analyses of fungi. *MycoKeys* **10**: 1-43.
- 1164 Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV,  
1165 Vasco-Palacios AM, Thu PQ, Suija A et al. 2014. Global diversity and  
1166 geography of soil fungi. *Science* **346**.
- 1167 Teng LS. 1990. Geotectonic evolution of late Cenozoic arc-continent collision in  
1168 Taiwan. *Tectonophysics* **183**: 57-76.
- 1169 Thomasson KM, Franks A, Teotonio H, Proulx SR. 2021. Testing the adaptive value  
1170 of sporulation in budding yeast using experimental evolution. *Evolution*  
1171 doi:10.1111/evo.14265.
- 1172 Todd RT, Braverman AL, Selmecki A. 2018. Flow Cytometry Analysis of Fungal  
1173 Ploidy. *Current Protocols in Microbiology* **50**: 1-20.
- 1174 Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild  
1175 yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proceedings of*  
1176 *the National Academy of Sciences of the United States of America* **105**: 4957-  
1177 4962.
- 1178 Tsukada M. 1966. Late Pleistocene Vegetation and Climate in Taiwan (Formosa).  
1179 *Proceedings of the National Academy of Sciences* **55**: 543-548.
- 1180 Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome  
1181 assembly from long uncorrected reads. *Genome Res* **27**: 737-746.
- 1182 Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. 2005. VariScan: Analysis of  
1183 evolutionary patterns from large-scale DNA sequence polymorphism data.  
1184 *Bioinformatics* **21**: 2791-2793.
- 1185 Wang QM, Liu WQ, Liti G, Wang SA, Bai FY. 2012. Surprisingly diverged  
1186 populations of *Saccharomyces cerevisiae* in natural environments remote from  
1187 human activity. *Mol Ecol* **21**: 5404-5417.
- 1188 White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and Direct Sequencing of  
1189 Fungal Ribosomal Rna Genes for Phylogenetics. In *PCR Protocols*,  
1190 doi:10.1016/b978-0-12-372180-8.50042-1, pp. 315-322.

- 1191 Whittaker RJ, Fernandez-Palacios JM, Matthews TJ, Borregaard MK, Triantis KA.  
1192 2017. Island biogeography: Taking the long view of nature's laboratories.  
1193 *Science* **357**.
- 1194 Xia W, Nielly-Thibault L, Charron G, Landry CR, Kasimer D, Anderson JB, Kohn  
1195 LM. 2017. Population genomics reveals structure at the individual, host-tree  
1196 scale and persistence of genotypic variants of the undomesticated  
1197 yeast *Saccharomyces paradoxus* in a natural woodland. *Molecular Ecology* **26**:  
1198 995-1007.
- 1199 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*  
1200 **24**: 1586-1591.
- 1201 Yue J-X, Li J, Aigrain L, Hallin J, Persson K, Oliver K, Bergström A, Coupland P,  
1202 Warringer J, Lagomarsino MC et al. 2017. Contrasting evolutionary genome  
1203 dynamics between domesticated and wild yeasts. *Nature Genetics* **49**: 913-  
1204 924.
- 1205 Zhu YO, Sherlock G, Petrov DA. 2016. Whole Genome Analysis of 132 Clinical  
1206 *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation. *G3*  
1207 (*Bethesda, Md*) **6**: 2421-2434.  
1208