

1 Fermentation innovation through complex hybridization of wild and domesticated yeasts

2

3 Quinn K. Langdon<sup>1</sup>, David Peris<sup>1,2,3</sup>, EmilyClare P. Baker<sup>1,4</sup>, Dana A. Opulente<sup>1,2</sup>, Huu-Vang

4 Nguyen<sup>5</sup>, Ursula Bond<sup>6</sup>, Paula Gonçalves<sup>7</sup>, José Paulo Sampaio<sup>7</sup>, Diego Libkind<sup>8</sup>, Chris Todd

5 Hittinger<sup>1,2,4,@</sup>

6

7 <sup>1</sup>Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy

8 Institute, Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI 53706,

9 USA

10 <sup>2</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI

11 53706, USA

12 <sup>3</sup>Department of Food Biotechnology, Institute of Agrochemistry and Food Technology (IATA),

13 CSIC, Valencia, Spain

14 <sup>4</sup>Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI

15 53706, USA

16 <sup>5</sup>Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

17 <sup>6</sup>Department of Microbiology, School of Genetics and Microbiology, Trinity College Dublin,

18 Ireland

19 <sup>7</sup>UCIBIO-REQUIMTE, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia,

20 Universidade Nova de Lisboa, Caparica, Portugal

21 <sup>8</sup>Laboratorio de Microbiología Aplicada, Biotecnología y Bioinformática de Levaduras, Instituto

22 Andino Patagónico de Tecnologías Biológicas y Geoambientales (IPATEC), Consejo Nacional

DOI: [10.1038/s41559-019-0998-8](https://doi.org/10.1038/s41559-019-0998-8)

23 de Investigaciones, Científicas y Técnicas (CONICET)-Universidad Nacional del Comahue,  
24 8400 Bariloche, Argentina

25 @Corresponding author: cthittinger@wisc.edu

26

27 Abstract

28         The most common fermented beverage, lager beer, is produced by interspecies hybrids of  
29 the brewing yeast *Saccharomyces cerevisiae* and its wild relative *Saccharomyces eubayanus*.  
30 Lager-brewing yeasts are not the only example of hybrid vigor or heterosis in yeasts, but the full  
31 breadth of interspecies hybrids associated with human fermentations has received less attention.  
32 Here we present a comprehensive genomic analysis of 122 *Saccharomyces* hybrids and  
33 introgressed strains. These strains arose from hybridization events between two to four species.  
34 Hybrids with *S. cerevisiae* contributions originated from three lineages of domesticated *S.*  
35 *cerevisiae*, including the major wine-making lineage and two distinct brewing lineages. In  
36 contrast, the undomesticated parents of these interspecies hybrids were all from wild Holarctic or  
37 European lineages. Most hybrids have inherited a mitochondrial genome from a parent other than  
38 *S. cerevisiae*, which recent functional studies suggest could confer adaptation to colder  
39 temperatures. A subset of hybrids associated with crisp flavor profiles, including both lineages of  
40 lager-brewing yeasts, have inherited inactivated *S. cerevisiae* alleles of critical phenolic off-  
41 flavor genes and/or lost functional copies from the wild parent through multiple genetic  
42 mechanisms. These complex hybrids shed light on the convergent and divergent evolutionary  
43 trajectories of interspecies hybrids and their impact on innovation in lager-brewing and other  
44 diverse fermentation industries.

45

46 Introduction

47 Humans have been producing and consuming fermented beverages for thousands of years  
48 <sup>1</sup>. During this process, they have unwittingly shaped the evolutionary history of the microbes that  
49 are responsible for fermented products. The star of fermented beverage production is often  
50 *Saccharomyces cerevisiae*. Many studies have investigated the evolutionary impact of  
51 domestication in fermentation environments on the genomes of different lineages of this species  
52 <sup>2-13</sup>. These human-associated fermentation environments have also led to innovation through the  
53 hybridization of distantly related species.

54 Lager beers are made with hybrids between the distantly related species *S. cerevisiae* and  
55 *Saccharomyces eubayanus* <sup>14-16</sup>. These hybrids combine unique properties from each; *S.*  
56 *cerevisiae*'s carbon utilization and fermentation capabilities combined with *S. eubayanus*'s  
57 cryotolerance to produce yeasts that could ferment well in the cold <sup>17-22</sup>. Other interspecies  
58 hybrids of *Saccharomyces* have been associated, both favorably and unfavorably, with diverse  
59 fermentations. *S. cerevisiae* × *Saccharomyces kudriavzevii* hybrids are prized for their unique  
60 flavor profiles in beer and wine <sup>23</sup>. Conversely, hybrids and introgressed strains with large  
61 genomic contributions from *S. eubayanus* and *Saccharomyces uvarum*, are viewed as  
62 contaminants in breweries due to the production of off-flavors, while other strains have been  
63 associated with sparkling wine and cider fermentation <sup>16,24,25</sup>. Although these previous studies  
64 have hinted at the complexity of fermentation hybrids, their focus on a handful of strains or a  
65 handful of loci has only given us a fleeting glimpse of the diversity *Saccharomyces* hybrids, their  
66 total genomic compositions, and their evolution.

67 Here we identified, sequenced, and analyzed the genomes of 122 interspecies hybrids and  
68 introgressed strains in the genus *Saccharomyces* to understand their origins and evolutionary

69 innovations. This collection contains pairwise hybrids, as well as more complex hybrids and  
70 introgressed strains with three or four parent species. We show that all genomic contributions  
71 from *S. cerevisiae* have arisen out of three domesticated lineages of *S. cerevisiae*, while all other  
72 parents belonged to Holarctic or European wild lineages of their respective species. We also  
73 analyzed inheritance of the mitochondrial genome and the genetic events generating functional  
74 diversity in genes relevant to fermented beverages. The genomic complexity of these hybrids  
75 provides insight into their origins and evolutionary successes in human-associated fermentation  
76 environments.

77

## 78 Results

### 79 Summary of Interspecies Hybrid Types

80 Here, we analyzed the genome sequences of 122 interspecies hybrids and introgressed  
81 strains of *Saccharomyces*, 63 strains of which are newly sequenced here, more than doubling the  
82 number of previously published hybrid genomes. Collectively, industrial settings dominated the  
83 isolation origins of all hybrids; 86% (n=105) were from beer, wine, cider, a distillery, or other  
84 beverages (Figure 1b, Table S1, Supplementary Text). We identified four types of hybrids: 1)  
85 lager-like (*S. cerevisiae* (*Scer*) × *S. eubayanus* (*Seub*)) (n=56); 2) *S. cerevisiae* × *S. kudriavzevii*  
86 (*Skud*) (n=15); 3) *S. eubayanus* × *S. uvarum* (*Suva*) (n=41); and 4) more complex hybrids, with  
87 three or four parent species (n=11 more than doubling those previously identified<sup>26</sup>) (Figure 1a,  
88 Table S1, Supplementary Text). These more complex hybrids fell into three groups: 4A) *S.*  
89 *cerevisiae* × *S. kudriavzevii* × *S. eubayanus* × *S. uvarum* (n=5), 4B) *S. cerevisiae* × *S. eubayanus*  
90 × *S. uvarum* (n=4), and 4C) one *S. cerevisiae* × *S. kudriavzevii* × *S. eubayanus* (Table S1). The  
91 lager-like hybrids were almost exclusively associated with beer (Figure 1b) and have genomic

92 contributions that were consistent with previous observations in the two lineages (Saaz and  
93 Frohberg)<sup>27</sup>. The *S. cerevisiae* × *S. kudriavzevii* strains were associated with beer and wine  
94 (Figure 1b). They had considerable differences in *S. kudriavzevii* genomic content, suggesting  
95 that these hybrids are of variable ages and evolutionary histories. The *S. eubayanus* × *S. uvarum*  
96 hybrids and introgressed strains were the most variable, both in isolation environment and  
97 genomic contributions (Figure 1, Table S1). The wide range in genomic contributions in these  
98 strains was likely influenced by their ability to backcross due to the low, but non-zero, spore  
99 viability of hybrids of these sister species<sup>16</sup>. These *S. eubayanus* × *S. uvarum* strains had the  
100 most total number of translocations ( $\chi^2 = 1250.1$ ,  $p_{\text{adj}} = 2.64 \text{ E-}15$ ), as well as the most  
101 translocations shared with other hybrid types ( $\chi^2 = 15.964$ ,  $p_{\text{adj}} = 0.0138$ ) (Figure S2). The  
102 shared nature of some of these translocations in hybrids with more than two parents suggests that  
103 *S. eubayanus* × *S. uvarum* introgressed strains further hybridized to produce some of the  
104 complex three or four parent species hybrids. Thus, these four types of hybrids each show unique  
105 dynamics in genome evolution and are used for different products that range from several  
106 regional niche beverages to the globally dominant beer style, lagers.

107

#### 108 Wild Parent Populations

109 Three out of four of the species contributing to these hybrids (*S. kudriavzevii*, *S. uvarum*,  
110 and *S. eubayanus*) have primarily been isolated from wild settings and have global distributions  
111 with populations that reflect their geography<sup>28,29</sup>. We used these established populations and  
112 phylogenomic and PCA approaches to evaluate the origins of these hybrids (Supplementary  
113 Text).

114 *S. kudriavzevii* has been isolated in Europe and Asia and consists of three described  
115 populations: Asia A, Asia B, and Europe<sup>23,30,31</sup>. The *S. kudriavzevii* sub-genomes of the hybrids  
116 all clustered with the European population as a monophyletic clade (Figure 2a, Figure S3, Table  
117 S2, File S1, Supplementary Text). These findings show that these hybrids were drawn from a  
118 closely related lineage of the European population of *S. kudriavzevii*.

119 In *S. eubayanus*, analysis of both large and small contributions, showed that these hybrids  
120 and introgressed strains clustered with the Holarctic lineage of *S. eubayanus* (Figure 2b, Figure  
121 S5, Table S2, File S3, Supplementary Text). Our vastly expanded dataset suggests that the  
122 Holarctic lineage is the closest known relative of all industrially relevant *S. eubayanus* hybrids  
123 and introgressed strains. The array of hybrids observed here requires that multiple hybridization  
124 events occurred between this lineage and other species. We also analyzed genetic diversity of the  
125 *S. eubayanus* contributions to industrial hybrids and introgressed strains (Supplementary Text).  
126 We found low nucleotide diversity in lager-like hybrids that shows that these widely used  
127 interspecies hybrids arose out of a narrow swath of *S. eubayanus* diversity, while the less  
128 frequently used hybrids and introgressed strains retained more nucleotide diversity.

129 *S. uvarum* has a parallel population structure to *S. eubayanus*<sup>26,32</sup>, with the exception of  
130 its increased isolation frequency in the Northern Hemisphere and the presence of pure strains  
131 isolated from Europe. Here we found that all contributions from *S. uvarum* arose out of the *S.*  
132 *uvarum* Holarctic lineage<sup>26</sup>. In contrast to our *S. eubayanus* findings, the *S. uvarum* sub-  
133 genomes of these hybrids and introgressed strains were interspersed with pure wild strains  
134 (Figure 2c, Figure S7 & S7, Table S2, File S5 & S6). These findings suggest that there have been  
135 multiple hybridization events and extensive backcrossing with wild lineages of *S. uvarum*,  
136 integrating wild diversity into these hybrids and leading to a diverse set of introgressed strains.

137

138 Domesticated *S. cerevisiae* Parent Lineages

139         Of the species contributing to domesticated interspecies hybrids, *S. cerevisiae* has the  
140 most extensive datasets, including industrial yeasts<sup>5,8-11</sup>. Through both phylogenomic and PCA  
141 approaches, we recapitulated the previously described domesticated *S. cerevisiae* clades<sup>8,9</sup>, and  
142 our 81 interspecies hybrids with *S. cerevisiae* contributions fell into three domesticated lineages:  
143 Wine, Ale/Beer1, and Beer2 (Figure 2d, Figure S9, Table S2, File S7).

144         The *S. cerevisiae* × *S. kudriavzevii* hybrids grouped with both Beer2 and Wine. Strains  
145 with contributions from three or four parent species fell into both clades (Beer2 and Wine),  
146 suggesting that these complex hybrids originated stepwise through iterative hybridization  
147 (Supplementary Text).

148         Interestingly, the only hybrids we detected in the Ale/Beer1 group were the lager-  
149 brewing yeasts (Figure 2d). The *S. cerevisiae* sub-genomes of the Saaz and Frohberg lager-  
150 brewing lineages formed distinct clades, and although we identified more Frohberg strains,  
151 Frohberg genetic diversity was lower (Supplementary Text). To determine if there was a  
152 particular clade of Ale/Beer1 that was the closest known relative to lager-brewing hybrids, we  
153 performed a targeted analysis of just the Ale/Beer1 *S. cerevisiae* strains and lager-brewing  
154 hybrids, (Figure S10 & S10, Table S2, File S8, Supplementary Text). Our concatenated  
155 phylogenomic analyses did not strongly support any recognized geographical clade of Ale/Beer1  
156 *S. cerevisiae* strains as the closest outgroup to the lager-brewing yeasts. Our PCA analyses,  
157 which make no assumptions about consistent genome-wide signals, suggested several Stout beer,  
158 Wheat beer, and mosaic strains as sharing the most ancestry with lager-brewing yeasts, rather  
159 than any clade affiliated with a geographic style (Figure S9). Overall, our analyses clearly show

160 that lager strains belong to the Ale/Beer1 lineage of *S. cerevisiae* and suggest affinity with a  
161 novel set of diverse beer yeasts, but they do not support any known extant strain as the sole  
162 closest relative.

163 Collectively, our data and analyses conclusively show that there have been multiple  
164 interspecies hybridization events between different domesticated lineages of *S. cerevisiae* and  
165 wild strains from three other *Saccharomyces* species (Figure 2d). The sheer number and diversity  
166 of hybrids analyzed here shows that evolutionary and industrial innovation through hybridization  
167 has happened on a scale and with a complexity beyond what previous smaller scale studies have  
168 suggested. In these diverse hybrids, the domesticated *S. cerevisiae* sub-genomes were likely  
169 preadapted with general industrial fermentation traits, while the wild parent likely contributed  
170 one or more traits advantageous in the specific new industrial fermentation niche being explored.

171

#### 172 Mitochondrial Genome Inheritance

173 The classic example of yeast hybrid vigor comes from the cryotolerance of lager-brewing  
174 yeasts. *S. eubayanus*, *S. kudriavzevii*, and *S. uvarum* are all known to tolerate much colder  
175 temperatures<sup>33,34</sup>, and recent functional experiments have shown that the mitochondrial genome  
176 (mtDNA) plays a pivotal role in the cryotolerance of interspecies hybrids<sup>17,35</sup>. Strikingly, in our  
177 comprehensive dataset, a majority (94%) of the hybrids inherited a mtDNA from another  
178 species, rather than the *S. cerevisiae* mtDNA (Figure 3a).

179 We tested if the parent that donated the mtDNA was also the parent that contributed the  
180 most nuclear gene content. We used a logistic regression to determine if the same parent species  
181 contributed both the mtDNA and the most complete set of orthologs. We found that this trend  
182 was generally true ( $p=8.0E-6$ , AIC= 83.75), but there were informative outliers (Figure 3b). In

183 particular, more than half of the hybrids with *S. kudriavzevii* nuclear contributions inherited the  
184 *S. kudriavzevii* mtDNA, despite the fact that the *S. kudriavzevii* nuclear contribution was never in  
185 the majority. This discrepancy could be due to a fitness advantage conferred by the *S.*  
186 *kudriavzevii* mtDNA in colder fermentations, or it could be due to a fitness advantage conferred  
187 by the *S. cerevisiae* or other nuclear genomes<sup>36,37</sup>. Indeed, all outliers in our logistic regression  
188 analysis were in the direction of inheriting a cryotolerant parent's mtDNA. These findings  
189 suggest that the inheritance of a cryotolerant mtDNA allowed these hybrids to thrive in colder  
190 environments where pure *S. cerevisiae* strains struggle, providing evolutionary and genetic  
191 innovation that enabled new fermentation techniques, such as lager brewing.

192         Hundreds of nuclear-encoded proteins localize to the mitochondria<sup>38</sup>. This interaction  
193 can be a source of genetic incompatibilities between the nuclear and mtDNAs, several of which  
194 have been characterized in *Saccharomyces* interspecies hybrids<sup>39-41</sup>. Therefore, we tested  
195 whether mitochondrially localized, nuclear-encoded genes were retained more often than other  
196 genes encoded in the nuclear genome matching the mtDNA parent. We found that more  
197 mitochondrially localized genes were retained in the same ratio as all other orthologs ( $p =$   
198 0.8612, odds ratio = 0.9653) (Table S3, Figure 3c). Although these results suggest that  
199 mitochondrial localization is not the main cause of the correlation between nuclear and mtDNA  
200 content, some nuance is warranted. First, only a small number of mitochondrially localized genes  
201 have been implicated in mito-nuclear incompatibilities<sup>39-41</sup>, and other factors that do not rely on  
202 protein localization could also play a role (e.g. metabolite exchange between the mitochondria  
203 and cytoplasm). Perhaps more importantly, these hybrids have often lost whole chromosomes or  
204 regions containing hundreds of genes at a time through chromosome mis-segregation or mitotic  
205 recombination events<sup>15</sup>; this restriction imposed by genetic linkage may prevent fine-scale

206 retention or loss and obscure any signal driven by specific genes. Finally, some yet unmapped  
207 cryotolerant nuclear alleles might also be favored independently from the cryotolerant mtDNA.  
208 Overall, from this dataset, we conclude that there is a strong correlation between the amount of  
209 nuclear and mitochondrial DNA contributed by each parent species, but mitochondrially  
210 localized genes are not more affected than other genes.

211

#### 212 Pan-Genome Analyses:

213 To characterize the core genome of these hybrids, we first analyzed the retention of  
214 1:1:1:1 orthologs conserved in all four parent species and determined which parents contributed  
215 the least and most coding sequences to each hybrid. As few as 12 genes were retained in one  
216 strain, whereas some hybrids have retained almost complete sets of orthologs from all their  
217 parents (Figure S12, and Table S4). On average, these hybrids retained 56.2% of orthologs from  
218 the parent who contributed the least genomic material.

219 We performed de novo genome assemblies to analyze the genomic content that was not  
220 present in the parent reference genomes (Figure S13). On average, these hybrids had 47.7 kbp of  
221 novel genomic content; the minimum was 2.2 kbp, and the maximum was 363.3 kbp. In addition  
222 to novel content that may come from the pan-genomes of other the *Saccharomyces* species, we  
223 detected previously characterized content from prior *S. cerevisiae* pan-genome analyses,  
224 including horizontally transferred genes (Supplemental Text)<sup>5,12,42</sup>. When we searched this  
225 material for *Saccharomyces*-like genes for which we could assign a function, we found an  
226 enrichment in genes associated with sugar transport, including the Gene Ontology<sup>43,44</sup> terms:  
227 transporter activity (corrected p-val = 4.67E-08), sugar:proton symporter activity (corrected p-val  
228 = 6.04E-08), cation:sugar symporter activity (corrected p-val = 6.04E-08), and sugar

229 transmembrane transporter activity (corrected p-val = 6.04E-08) (Table S5). The enrichment of  
230 sugar transport genes in the novel content of these hybrids and introgressed strains is consistent  
231 with strong selection for these activities in industrial fermentation environments.

232

### 233 Maltotriose Utilization Genes

234 We took a more detailed look at maltotriose utilizing genes because maltotriose is  
235 generally the second most abundant sugar in beer wort or malt extract, and *Saccharomyces*  
236 strains that utilize it are relatively rare outside of domesticated ale-brewing strains<sup>45-48</sup>. Our  
237 analyses of lager-brewing yeasts suggest that both *S. cerevisiae* and *S. eubayanus* contributed  
238 genes encoding functional maltotriose transporters to the hybrids, including alleles of *S.*  
239 *cerevisiae* *MTT1* and *S. eubayanus* *AGT1* previously shown to be functional<sup>18</sup> (Figure 5b,  
240 Supplementary Text). We also recovered other predicted maltose/maltotriose transporter  
241 homologs in other interspecies hybrids and their parent species, which have yet to be explored  
242 functionally (Table S6). We conclude that the complexity and diversity of maltose transporter  
243 genes across *Saccharomyces* species is extensive and may have provided a source of functional  
244 diversity to fermentation hybrids.

245

### 246 Phenolic Off-Flavor Genes

247 The introduction of genes from wild strains, especially the mitochondrial genome and *S.*  
248 *eubayanus* *AGT1*, may have been key to cold fermentations, but other genes likely negatively  
249 impacted products. 4-vinyl guaiacol (4VG) is perceived as a clove-like, phenolic, or smoky  
250 flavor and considered an undesirable off-flavor in most beers. Lager beers are known for their  
251 crisp flavor profiles that lack appreciable 4VG, while wild strains of *S. eubayanus* and other

252 species produce 4VG<sup>49</sup>. Two genes, *PADI* and *FDCI*, are essential for the production of 4VG  
253<sup>50</sup>. Studies in ale-brewing yeast show that this trait is under strong domestication selection  
254 (Supplementary Text), but the genotypes of *PADI* and *FDCI* across diverse interspecies hybrids  
255 already in use by industry have not been investigated, nor have the evolutionary genetic events  
256 leading to these genotypes. In our large hybrid dataset, we analyzed both retention and predicted  
257 functionality of *PADI* and *FDCI* alleles from their parent species (Figure 4).

258 In both *S. cerevisiae* × *S. kudriavzevii* and *S. eubayanus* × *S. uvarum* hybrids and  
259 introgressed strains, we found both *FDCI* and *PADI* alleles that were predicted to be functional  
260 (Supplementary Text). These findings may reflect selection for diverse flavors, which are  
261 desirable in niche Trappist-style beers made with *S. cerevisiae* × *S. kudriavzevii*. In contrast *S.*  
262 *eubayanus* × *S. uvarum* are often viewed as contaminants in industrial brewing environments,  
263 and production of 4VG could contribute to this perception.

264 In the lager-brewing hybrids, we found that all strains have lost the ability to produce  
265 4VG, but mechanism of this loss differed between Saaz and Frohberg (Supplementary Text). The  
266 Frohberg lager strains likely inherited a loss-of-function *FDCI* allele from their domesticated *S.*  
267 *cerevisiae* parent and functional *PADI* and *FDCI* alleles from their *S. eubayanus* parent. These  
268 functional wild alleles were then lost through translocations, likely due to break-induced  
269 replication. In contrast, the Saaz lineage has completely lost both the *S. cerevisiae* and *S.*  
270 *eubayanus* alleles of these genes through aneuploidy, an evolutionary trajectory facilitated by the  
271 fact that these subtelomeric genes reside on different chromosomes in these two species. The end  
272 result is that both Saaz and Frohberg lagers lack substantial phenolic off-flavors and have a crisp  
273 flavor profile. Even though Saaz and Frohberg strains evolved this trait through different final  
274 mutations that removed functional *S. eubayanus* alleles, the pre-adaptation of the domesticated *S.*

275 *cerevisiae* parent, which already lacked functional genes, played a critical role by limiting the  
276 number of mutations needed. The contrast between Saaz and Froberg strains highlights that  
277 there are many potential evolutionary trajectories open to interspecies hybrids to achieve a  
278 domestication trait.

279

## 280 Conclusions

281 Here, we characterized the genomes of 122 interspecies yeast hybrids and introgressed  
282 strains, the largest dataset of its kind to date. These hybrids have complex genomes with  
283 contributions from two to four species: *S. cerevisiae*, *S. kudriavzevii*, *S. uvarum*, and *S.*  
284 *eubayanus* (Figure 5a). The hybrids with *S. cerevisiae* contributions all arose out of three  
285 domesticated *S. cerevisiae* lineages: the wine lineage and two distinct beer clades. In contrast, all  
286 the *S. kudriavzevii*, *S. uvarum*, and *S. eubayanus* parents belonged to Holarctic or European wild  
287 lineages. Our results show how hybrid vigor also applies to microbes, with the domesticated *S.*  
288 *cerevisiae* parents providing genes and traits pre-adapted for industrial fermentations and the  
289 divergent species of *Saccharomyces* contributing new genes and traits that led to the successes of  
290 these hybrids in specific products. First, the frequent retention of mitochondrial genomes from  
291 cryotolerant parents likely conferred a fitness advantage during cold fermentation (Figure 5b).  
292 Second, although the *S. cerevisiae* genome is required for maltotriose utilization by hybrids, both  
293 *S. eubayanus* and *S. cerevisiae* contributed functional maltotriose transporter genes to lager-  
294 brewing yeasts. Third, phenolic off-flavor genes have been inactivated or eliminated from lager-  
295 brewing yeasts by multiple types of mutations (Figure 5b), while these genes have been retained  
296 in yeasts that ferment products where phenolic off-flavor is prized.

297           Hundreds of years ago, a *S. cerevisiae* strain meeting a *S. eubayanus* strain sparked the  
298 cold-brewing revolution, and crisp refreshing lagers eventually overtook the global beer market.  
299 This extensive genomic dataset reveals the genetic mechanisms and distinct evolutionary  
300 trajectories followed by hybrid and introgressed strains associated with fermentation products.  
301 These diverse hybrids and introgressed strains highlight how dynamic and complex fermentation  
302 innovation has cascaded down divergent and convergent evolutionary trajectories.

303

304 Methods

305 Strain Selection and Sequencing

306           The strains newly published here are from wild or beverage isolations, the Agricultural  
307 Research Service (ARS) NRRL collection (<https://nrml.ncaur.usda.gov>), and commercially  
308 available sources. Table S7 contains the full metadata for strains. Whole genome Illumina  
309 paired-end sequencing was done as previously described using either 2X100 or 2X250 reads<sup>32,51</sup>.  
310 This short-read data is available through the NCBI SRA database under the accession number  
311 PRJNA522928. Short-read data for published genomes were downloaded from NCBI; Table S8  
312 contains a full list of accession numbers and citations<sup>8,9,11,16,26,30,32,42,52–72</sup>.

313

314 Hybrid Identification

315           We used **sppIDer**<sup>73</sup>, a hybrid detection and analysis pipeline, to identify new hybrids,  
316 pure species, and reconfirm the species and hybrid identities of published data. For **sppIDer**,  
317 we used a combination reference genome that included all published genomes for all the  
318 *Saccharomyces* species<sup>63,72,74,75</sup> (<https://www.yeastgenome.org/>,  
319 [www.saccharomycessensustricto.org](http://www.saccharomycessensustricto.org)). For *S. kudriavzevii*, we used the genome from the

320 Portuguese strain ZP591. As previously noted <sup>72</sup>, the published *S. uvarum* genome has the labels  
321 for chromosomes X and XII swapped, so we manually corrected them. We ran `sppIDer` with  
322 parameters set to identify genomic contributions >1% of the total genome. As `sppIDer` is  
323 reference genome-based, inheritance of regions not in the reference genome was not analyzed.  
324 Therefore, interspecies hybrids with only minor or subtelomeric introgressions were missed with  
325 this method. We also detected some smaller introgressions through the pan-genome analyses (see  
326 below).

327         Hybrid isolation environment was classified based on marketed product type for  
328 commercial strains; for published strains or strains from the ARS NRRL collection, we used  
329 available metadata supplied by the authors or depositors. Full details on hybrid isolation  
330 environment classification can be found in Table S1. To determine if there was an association  
331 between hybrid type and isolation environment, we completed  $\chi^2$  analyses of hybrid by  
332 environment and of environment by hybrid with a Bonferroni multiple test correction in R. We  
333 limited this test to our most common (n>15) hybrid types (*S. cerevisiae* × *S. eubayanus*, *S.*  
334 *cerevisiae* × *S. kudriavzevii*, and *S. eubayanus* × *S. uvarum*) and the most common (n>8) origins  
335 (beer, wine, and fruit).

336

### 337 Whole Genome Sequence Assembly Pipeline

338         Alignment and single nucleotide polymorphism (SNP) calling were done as described  
339 previously <sup>32</sup>. Briefly, short reads were mapped with `bwa` “mem” to a concatenated reference  
340 genome of just the contributing parents. Reference genomes used for concatenation were the  
341 same as used for `sppIDer`. `Samtools` “view” and “sort” were then used to prepare the  
342 mapped reads with a mapping quality greater than 20 for SNP calling. PCR duplicates were

343 removed with `picard` “MarkDuplicates”, and read groups were set with `picard`  
344 “AddOrReplaceReadGroups”. SNPs were called with GATK’s haplotype caller. Genome  
345 coverage per base pair was assessed with `bedtools` “genomeCoverageBed”. Strain-specific  
346 FASTA files were created by replacing called SNPs in repeat-masked concatenated reference  
347 genomes. Variants called as indels were replaced with Ns. Regions of extremely high coverage,  
348 (i.e. the 99.9th percentile of genome-wide coverage) were masked as Ns. Regions that do not  
349 exist in hybrids were masked as Ns, and regions at low coverage (i.e. between 3X-10X,  
350 depending on where the 10<sup>th</sup> percentile of the distribution of depth of coverage across the  
351 concatenated genomes fell) were masked as Ns. The strain-specific FASTAs for hybrid genomes  
352 were split into their component sub-genomes to be analyzed with pure strains.

353         Genomic completeness was estimated as the percent of the reference genome with  
354 coverage above the low-coverage masking threshold. Ploidy was estimated across the  
355 combination genome in 10-kbp windows. We used the R package *modes* (version 0.7.0) to  
356 analyze the distribution of depth of coverage and determine the antimodes, which correspond to  
357 a change in ploidy state. Some manual curation was needed for strains with “smiley patterns”, a  
358 pattern of increased coverage at chromosome ends that has been noted in other depth-of-  
359 coverage analyses<sup>8,76</sup> and may be due chromatin structure<sup>77</sup>. For these strains, we used only the  
360 coverages that fell below the 95<sup>th</sup> percentile to estimate the antimodes and then assigned the  
361 distal ends to the largest ploidy estimated. We also visually checked and corrected rare instances  
362 when a “smiley pattern” lowered the ploidy estimate for the middle of the chromosome. From  
363 this antimode analysis, we were able to assign each 10-kbp window a ploidy value. The total  
364 DNA base-pair content contributed by each parent could then be estimated as the sum of each  
365 ploidy value multiplied by 10k and the number of windows with that ploidy value. Correcting

366 this total DNA content per species by the total sum of all contributing species gave us a measure  
367 of total genomic content per species. Genomic contribution to a hybrid genome can be viewed as  
368 genomic content and genomic completeness. To estimate genomic completeness, we determined  
369 what percent of a total parent sub-genome had at least one haploid copy. To estimate genomic  
370 content, we took into account both completeness and ploidy across the combination of sub-  
371 genomes. Full details on hybrid genome contributions can be found in Table S1. For  
372 visualizations, we clustered the strains based on ploidy estimated across the combination genome  
373 using Ward's method in the R package *pvclust* (v. 2.0-0) <sup>78</sup>.

374 For each strain, we calculated the number of sites called as heterozygous with GATK for  
375 each sub-genome. Strains with more than 20,000 heterozygous sites in any sub-genome were  
376 phased with GATK's "ReadBackedPhasing" command <sup>79</sup>, which can phase short regions of the  
377 genome based on overlapping reads. We then split the output into two phases, one that retains  
378 more reference variants and one that contains more alternative variants in phased regions. This  
379 pseudo-phasing allowed us to investigate regions that are less similar to the published reference.  
380 We converted these phases into two strain-specific FASTA files and masked them for coverage  
381 as above. Both phases were included in all downstream analyses involving phased genomes,  
382 which are noted as "strainID 1" or "strainID 2".

383

#### 384 1:1:1:1 Orthologs

385 We identified genes that are orthologous across all parent genomes based on the  
386 annotations in the published gff files for each reference genome, which yielded a list of 3,856  
387 genes. We used the coordinates to determine the coverage for each ortholog. Gene presence was  
388 noted if the mean coverage for that ortholog was >3X.

389

## 390 De Novo Genome Assembly and Pan-Genome Analyses

391 We assembled the hybrid genomes with the meta-assembler `iWGS`<sup>80</sup> and choose the best  
392 assembly based on the largest N50 score. All hybrids, except DBVPG6257, were successfully  
393 assembled and are available under GenBank BioProject PRJNA522928.

394 We mapped the short-read data back to these assembled genomes and used the `sppIDer`  
395 output to classify to which parent reference genome each short read mapped. With this analysis,  
396 we determined which reads did not map to a parent reference genome but did assemble de novo  
397 into a contig of 1.5-kbp or greater. We classified these regions as “unmapped” and used a  
398 `tBLASTx` to search for *S. cerevisiae*-like genes using S288C ORFs and retaining hits with e-  
399 value  $< 10^{-10}$ . To determine if this set of genes identified in these novel assembled regions were  
400 enriched for any functions, we used `GO Term Finder` (Version 0.86)<sup>43,44</sup>. To determine the  
401 potential origin of these novel regions, we used a `BLASTn` search of the NCBI nucleotide  
402 database (v5). The output of this was then parsed for number of hits with an e-value  $< 10^{-10}$ . To  
403 determine the number of hits to different species, we completed  $\chi^2$  analyses with a Bonferroni  
404 multiple test correction in R.

405

## 406 Translocation Identification

407 To detect shared breakpoints and translocations, we use `LUMPY`<sup>81</sup> with the mapped short-  
408 read data. We masked for repetitive regions by excluding regions with coverage above twice the  
409 genome-wide mean. Each breakpoint call had to be supported by at least 4 reads to be included  
410 in downstream analyses. We parsed this output for species sub-genome, hybrid type, and the  
411 species pair between which the translocation was detected. We calculated the total number of

412 called breakpoints, breakpoints that were shared in at least two hybrids of the same type, and  
413 breakpoints that were shared in multiple hybrid types. We compared these different categories  
414 with  $\chi^2$  analyses and a Bonferroni multiple test correction in R.

415 We also identified translocations from the de novo assemblies. For this analysis, we used  
416 `sppIDeR` results to assign regions of the de novo assemblies to a parent species. Some regions  
417 were unmapped with `sppIDeR`, as noted above. Additionally, some regions had high coverage  
418 from multiple parents in the de novo assembly, where the donor species could not be  
419 unambiguously assigned; these regions are likely repetitive and difficult to assemble.  
420 Translocations were identified when regions that were >2-kbp came from different donor species  
421 and were assembled with <100-bp of unmapped or ambiguous data separating them. On average,  
422 we identified 17 translocations per strain. From this output, we counted the number of  
423 translocations identified in each hybrid type, the donor species, and the pair of species between  
424 which the translocations occurred. We compared hybrid type, species pair, and individual species  
425 with a  $\chi^2$  analyses with a Bonferroni multiple test correction in R.

426

#### 427 Mitochondrial Genome Analysis Pipeline

428 We use `mitoSppIDeR`<sup>73</sup> to determine the mitochondrial genome (mtDNA) parent for  
429 the hybrids. This analysis was done in a similar manner to the whole genome `sppIDeR` analysis,  
430 except that mtDNAs for each *Saccharomyces* species were used<sup>72,82,83</sup>, except *Saccharomyces*  
431 *jurei*. GenBank accessions lacking full manuscripts included *S. mikatae* (KX707788) and *S.*  
432 *kudriavzevii* (KX707787).

433 To determine if the mtDNA parent was associated with retention of the nuclear genes, we  
434 performed a logistic regression in R. We used the set of 1:1:1:1 orthologs to determine which

435 parent contributed the most complete set of orthologous genes. To determine if there was an  
436 enrichment for the retention of nuclear-encoded, mitochondrially interacting proteins, we used  
437 the set of genes products identified as localize to the mitochondria through the Yeast GFP Fusion  
438 Localization Database <sup>38</sup>. When we filtered for genes that were also 1:1:1:1 orthologs, our final  
439 list consisted of 459 genes. To determine if there was a linear relationship between retention of  
440 mitochondrially localized genes and all other orthologs, we performed a linear regression and to  
441 determine if there were more mitochondrially localized genes retained compared to all other  
442 genes, we used a Fisher's Exact Test with a Bonferroni correction. Tests were performed in R.

443         Since past work has shown that reticulate evolution, introgression, and horizontal gene  
444 transfers are widespread in *Saccharomyces* mtDNAs <sup>84</sup>, we wanted to explore the inheritance of  
445 mitochondrially encoded genes in more depth. Due in part to their high AT content (~85%),  
446 mtDNAs are often poorly covered using Illumina sequencing. In particular, intergenic regions  
447 and coding sequencing with transposable elements (introns, homing endonucleases, and GC  
448 clusters) can be difficult to assemble. To explore the phylogenetic relationships of these  
449 mtDNAs, we used a bait-prey bioinformatic method to pull out the read sequences of coding  
450 sequences. We used *HybPiper* <sup>85</sup> to pull out reads from the hybrid Illumina libraries that  
451 mapped to those mitochondrial genes using gene sequences from reference strains used in  
452 *mitoSpIDer* as baits. These extracted Illumina reads were aligned to the reference genes in  
453 *Geneious* (v. 6.1.6) <sup>86</sup> and manually assembled. We successfully covered six mitochondrial  
454 genes (*COX2*, *COX3*, *ATP6*, *ATP8*, *ATP9*, and *15S rRNA*), which were used to construct the  
455 mitochondrial phylogenetic haplotype network. This unique set of unambiguously completed  
456 genes was concatenated (4.7-kbp) by strain to produce the haplotype for each pure  
457 *Saccharomyces* or hybrid strain (Figure S14). Haplotypes and haplotype frequencies for each

458 strain were encoded as a nexus-formatted file for PopART v1.7.2<sup>87</sup>. The haplotype network was  
459 reconstructed using the TCS method<sup>88</sup>. Strains were assigned to each haplotype using DnaSP v5  
460<sup>89</sup>. For some strains, we could not assemble the *15S rRNA* gene because of low-coverage data.  
461 For these strains, we inferred their haplotype designation based on an analysis where we  
462 removed the *15S rRNA* gene. This information is not included in Figure S14 but can be found in  
463 Table S9.

464

#### 465 Genes of Functional Interest Analysis Pipeline

466 To assemble the sequences of genes relevant to brewing, we again used HybPiper<sup>85</sup>.  
467 To be included for further analyses, the assembled length had to be at least as long as the bait  
468 gene and had to have a minimum 10X depth of coverage. For the baits, we used either gene  
469 sequences from the *S. cerevisiae* strain S288C found on the *Saccharomyces* Genome Database  
470 (<https://www.yeastgenome.org>); from the *S. eubayanus* type strain, CBS12357<sup>T</sup><sup>72</sup>; or the lager  
471 strain W34/70<sup>90</sup>. For the *PADI* analysis in *S. eubayanus* × *S. uvarum* hybrids, we used the *PADI*  
472 gene sequence from the *S. uvarum* reference genome, CBS7001<sup>63</sup>. To get precise gene locations  
473 for *PADI* and *FDCI*, we used a tBLASTn search of the *S. eubayanus*, *S. kudriavzevii*, and *S.*  
474 *uvarum* reference genomes with the *S. cerevisiae* sequences for these genes as the query.

475 The assembled genes were aligned with MAFFT v.7<sup>91</sup>, allowing for reverse  
476 complementation. The alignments were manually trimmed to the protein-coding sequences. For  
477 *PADI* and *FDCI*, the alignments were conceptually translated to amino acid sequences, and  
478 haplotype networks were built with a modified minimum-spanning network and visualized with  
479 *iGraph*<sup>92</sup> in R. The haplotype networks were split into communities as previously described<sup>93</sup>.

480 Pairwise distances between sequences were calculated using the trimmed MAFFT  
481 nucleotide sequence alignments and the p-distance method as implemented in MEGA-X<sup>94</sup> with  
482 the following parameters: substitutions to include Transitions + Transversions, assuming uniform  
483 rates among sites, and using pairwise deletion of gaps. The percent identity of hits to the bait  
484 sequence was organized by species, and hybrid status was recorded in Table S6, along with the  
485 origin of the bait gene and tallies of sequences whose translations were visually identified as  
486 being incomplete or containing premature stop codons.

487

#### 488 Phylogenomic and Population Structure Analyses

489 We masked regions with no coverage as Ns, which is interpreted as missing data by most  
490 tools; therefore, for downstream whole genome analyses, we only included sub-genomes that  
491 were >50% complete (i.e. major contributions). To include the minor contribution hybrids in the  
492 non-*S. cerevisiae* analyses, we used reduced genomes that were concatenations of the regions of  
493 the genome that existed in at least one minor hybrid (Table S10). This procedure allowed us  
494 include strains with minor introgressions and only use regions of the genome that had been  
495 contributed by the minor parent. To balance some of our analyses for Saaz and Frohberg lager  
496 strains, we used a random subset of Frohberg strains to match the number of Saaz strains.

497 Phylogenomic trees were built with RAxML v8.1<sup>95</sup> using SNPs from the whole genome for the  
498 major analyses or the reduced genome for the minor analyses. Trees were visualized with iTOL

499 <sup>96</sup>. The PCA analyses were done with the *ade4* package in R<sup>97</sup> and visualized with *ggPlot2*

500 <sup>98</sup>. Estimates of adjusted  $\pi$  ( $\pi * 100$ ) were calculated with the *PopGenome* package in R<sup>99</sup>.

501

502 Data and Code Availability

503           References and accession numbers for the published data used can be found in Table S8.  
504 Short-read data newly published here is available through the NCBI SRA database under the  
505 accession number PRJNA522928. Custom R and Python scripts used for this publication can be  
506 found on GitHub (<https://github.com/qlangdon/hybrid-ferment-invent>).

507

#### 508 Author Contributions

509 QKL performed most analyses with assistance from DAO; DP and QKL performed  
510 mitochondrial genome analyses and drafted text; EPB and QKL analyzed genes of functional  
511 interest and drafted text; QKL, EPB, and DAO sequenced genomes; HVN, UB, PG, and JPS  
512 contributed key strains to study design; QKL, DP, EPB, DL, and CTH designed the study; and  
513 QKL and CTH wrote the manuscript with editorial input from all co-authors.

514

#### 515 Acknowledgments

516 We thank Kevin J. Verstrepen for coordinating publication with their study; Amanda B.  
517 Hulfachor and Martin Bontrager for preparing a subset of Illumina libraries; the University of  
518 Wisconsin Biotechnology Center DNA Sequencing Facility for providing Illumina sequencing  
519 facilities and services; Marc-André Lachance, Ashley Kinart, Drew T. Doering, Randy Thiel,  
520 and Dan Carey for strains; and Margaret Langdon, Amanda B. Hulfachor, and Kayla Sylvester  
521 for collecting fermentation samples and/or isolating strains. This material is based upon work  
522 supported by the National Science Foundation under Grant Nos. DEB-1253634 (to CTH) and  
523 DGE-1256259 (Graduate Research Fellowship to QKL), the USDA National Institute of Food  
524 and Agriculture Hatch Project No. 1003258 to CTH, and in part by the DOE Great Lakes  
525 Bioenergy Research Center (DOE BER Office of Science Nos. DE-SC0018409 and DE-FC02-

526 07ER64494 to Timothy J. Donohue). QKL was also supported by the Predoctoral Training  
527 Program in Genetics, funded by the National Institutes of Health (5T32GM007133). DP is a  
528 Marie Sklodowska-Curie fellow of the European Union's Horizon 2020 research and innovation  
529 program (Grant Agreement No. 747775). EPB was supported by a Louis and Elsa Thomsen  
530 Wisconsin Distinguished Graduate Fellowship. DL was supported by CONICET (PIP 392),  
531 FONCyT (PICT 3677), and Universidad Nacional del Comahue (B199). CTH is a Pew Scholar  
532 in the Biomedical Sciences, Vilas Faculty Early Career Investigator, and H. I. Romnes Faculty  
533 Fellow, supported by the Pew Charitable Trusts, Vilas Trust Estate, and Office of the Vice  
534 Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni  
535 Research Foundation (WARF), respectively.

536

#### 537 References

- 538 1. Hornsey, I. S. *Alcohol and Its Role in the Evolution of Human Society*. (RSC Publishing,  
539 2012).
- 540 2. Fay, J. C. & Benavides, J. A. Evidence for Domesticated and Wild Populations of  
541 *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, e5 (2005).
- 542 3. Liti, G., Peruffo, A., James, S. A., Roberts, I. N. & Louis, E. J. Inferences of evolutionary  
543 relationships from a population survey of LTR-retrotransposons and telomeric-associated  
544 sequences in the *Saccharomyces sensu stricto* complex. *Yeast* **22**, 177–192 (2005).
- 545 4. Gallone, B. *et al.* Origins, evolution, domestication and diversity of *Saccharomyces* beer  
546 yeasts. *Curr. Opin. Biotechnol.* **49**, 148–155 (2018).
- 547 5. Legras, J. L. *et al.* Adaptation of *S. cerevisiae* to fermented food environments reveals  
548 remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* **35**,

- 549 1712–1727 (2018).
- 550 6. Rodríguez, M. E. *et al.* *Saccharomyces uvarum* is responsible for the traditional  
551 fermentation of apple chicha in Patagonia. *FEMS Yeast Res.* **17**, fow109 (2017).
- 552 7. Barbosa, R. *et al.* Multiple Rounds of Artificial Selection Promote Microbe Secondary  
553 Domestication—The Case of Cachaça Yeasts. *Genome Biol. Evol.* **10**, 1939–1955 (2018).
- 554 8. Gallone, B. *et al.* Domestication and Divergence of *Saccharomyces cerevisiae* Beer  
555 Yeasts. *Cell* **166**, 1397-1410.e16 (2016).
- 556 9. Gonçalves, M. *et al.* Distinct Domestication Trajectories in Top- Fermenting Beer Yeasts  
557 and Wine Yeasts. *Curr. Biol.* **26**, 1–12 (2016).
- 558 10. Duan, S. F. *et al.* The origin and adaptive evolution of domesticated populations of yeast  
559 from Far East Asia. *Nat. Commun.* **9**, (2018).
- 560 11. Peter, J. *et al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*  
561 **556**, 339–344 (2018).
- 562 12. Marsit, S. & Dequin, S. Diversity and adaptive evolution of *Saccharomyces* wine yeast: a  
563 review. *FEMS Yeast Res.* **15**, 1–12 (2015).
- 564 13. Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P. & Sampaio, J. P. Adaptive  
565 divergence in wine yeasts and their wild relatives suggests a prominent role for  
566 introgressions and rapid evolution at noncoding sites. *Mol. Ecol.* **26**, 2167–2182 (2017).
- 567 14. Hittinger, C. T., Steele, J. L. & Ryder, D. S. Diverse yeasts for diverse fermented  
568 beverages and foods. *Curr. Opin. Biotechnol.* **49**, 199–206 (2018).
- 569 15. Gibson, B. & Liti, G. *Saccharomyces pastorianus*: genomic insights inspiring innovation  
570 for industry. *Yeast* **32**, 17–27 (2015).
- 571 16. Libkind, D. *et al.* Microbe domestication and the identification of the wild genetic stock of

- 572 lager-brewing yeast. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14539–44 (2011).
- 573 17. Baker, E. P. *et al.* Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci. Adv.*  
574 **5**, eaav1869 (2019).
- 575 18. Baker, E. P. & Hittinger, C. T. Evolution of a novel chimeric maltotriose transporter in  
576 *Saccharomyces eubayanus* from parent proteins unable to perform this function. *PLOS*  
577 *Genet.* **15**, e1007786 (2019).
- 578 19. Hebly, M. *et al.* *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds  
579 and beyond. *FEMS Yeast Res.* **15**, 1–14 (2015).
- 580 20. Gibson, B. R., Storgårds, E., Krogerus, K. & Vidgren, V. Comparative physiology and  
581 fermentation performance of Saaz and Froberg lager yeast strains and the parental  
582 species *Saccharomyces eubayanus*. *Yeast* **30**, 255–266 (2013).
- 583 21. Gorter de Vries, A. *et al.* Laboratory evolution of a *Saccharomyces cerevisiae* × *S.*  
584 *eubayanus* hybrid under simulated lager-brewing conditions: genetic diversity and  
585 phenotypic convergence. *bioRxiv* **31**, 1–43 (2018).
- 586 22. Monerawela, C. & Bond, U. Brewing up a storm: The genomes of lager yeasts and how  
587 they evolved. *Biotechnol. Adv.* **35**, 512–519 (2017).
- 588 23. Peris, D., Pérez-Torrado, R., Hittinger, C. T., Barrio, E. & Querol, A. On the origins and  
589 industrial applications of *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii*  
590 hybrids. *Yeast* **35**, 51–69 (2018).
- 591 24. Nguyen, H. V. & Boekhout, T. Characterization of *Saccharomyces uvarum* (Beijerinck,  
592 1898) and related hybrids: Assessment of molecular markers that predict the parent and  
593 hybrid genomes and a proposal to name yeast hybrids. *FEMS Yeast Res.* **17**, 1–19 (2017).
- 594 25. Nguyen, H. V., Legras, J. L., Neuvéglise, C. & Gaillardin, C. Deciphering the

- 595 hybridisation history leading to the lager lineage based on the mosaic genomes of  
596 *Saccharomyces bayanus* strains NBRC1948 and CBS380 T. *PLoS One* **6**, (2011).
- 597 26. Almeida, P. *et al.* A Gondwanan imprint on global diversity and domestication of wine  
598 and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* **5**, 4044 (2014).
- 599 27. Dunn, B. & Sherlock, G. Reconstruction of the genome origins and evolution of the  
600 hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res.* **18**, 1610–1623 (2008).
- 601 28. Hittinger, C. T. *Saccharomyces* diversity and evolution: a budding model genus. *Trends*  
602 *Genet.* **29**, 309–17 (2013).
- 603 29. Boynton, P. J. & Greig, D. The ecology and evolution of non-domesticated  
604 *Saccharomyces* species. *Yeast* **31**, 449–462 (2014).
- 605 30. Hittinger, C. T. *et al.* Remarkably ancient balanced polymorphisms in a multi-locus gene  
606 network. *Nature* **464**, 54–58 (2010).
- 607 31. Sampaio, J. P. & Gonçalves, P. Natural populations of *Saccharomyces kudriavzevii* in  
608 Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S.*  
609 *paradoxus*. *Appl. Environ. Microbiol.* **74**, 2144–52 (2008).
- 610 32. Peris, D. *et al.* Complex Ancestries of Lager-Brewing Hybrids Were Shaped by Standing  
611 Variation in the Wild Yeast *Saccharomyces eubayanus*. *PLoS Genet.* **12**, (2016).
- 612 33. Salvadó, Z., Arroyo-López, F. N., Barrio, E., Querol, A. & Guillamón, J. M. Quantifying  
613 the individual effects of ethanol and temperature on the fitness advantage of  
614 *Saccharomyces cerevisiae*. *Food Microbiol.* **28**, 1155–61 (2011).
- 615 34. Gonçalves, P., Valério, E., Correia, C., de Almeida, J. M. G. C. F. & Sampaio, J. P.  
616 Evidence for divergent evolution of growth temperature preference in sympatric  
617 *Saccharomyces* species. *PLoS One* **6**, e20739 (2011).

- 618 35. Li, X. C., Peris, D., Hittinger, C. T., Sia, E. A. & Fay, J. C. Mitochondria-encoded genes  
619 contribute to evolution of heat and cold tolerance in yeast. *Sci. Adv.* **5**, eaav1848 (2019).
- 620 36. Ortiz-Tovar, G., Pérez-Torrado, R., Adam, A. C., Barrio, E. & Querol, A. A comparison  
621 of the performance of natural hybrids *Saccharomyces cerevisiae* × *Saccharomyces*  
622 *kudriavzevii* at low temperatures reveals the crucial role of their *S. kudriavzevii* genomic  
623 contribution. *Int. J. Food Microbiol.* **274**, 12–19 (2018).
- 624 37. Tronchoni, J., Medina, V., Guillamón, J. M., Querol, A. & Pérez-Torrado, R.  
625 Transcriptomics of cryophilic *Saccharomyces kudriavzevii* reveals the key role of gene  
626 translation efficiency in cold stress adaptations. *BMC Genomics* **15**, 1–10 (2014).
- 627 38. Huh, K. *et al.* *Global analysis of protein localization in budding yeast.* (2003).
- 628 39. Chou, J. Y., Hung, Y. S., Lin, K. H., Lee, H. Y. & Leu, J. Y. Multiple molecular  
629 mechanisms cause reproductive isolation between three yeast species. *PLoS Biol.* **8**,  
630 (2010).
- 631 40. Lee, H. Y. *et al.* Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid  
632 Sterility between Two Yeast Species. *Cell* **135**, 1065–1073 (2008).
- 633 41. Hou, J. & Schacherer, J. Negative epistasis: a route to intraspecific reproductive isolation  
634 in yeast? *Curr. Genet.* **62**, 25–29 (2016).
- 635 42. Novo, M. *et al.* Eukaryote-to-eukaryote gene transfer events revealed by the genome  
636 sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl. Acad. Sci.* **106**,  
637 16333–16338 (2009).
- 638 43. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**,  
639 25–29 (2000).
- 640 44. Consortium, T. G. O. The Gene Ontology Resource: 20 years and still GOing strong.

- 641 *Nucleic Acids Res.* **47**, D330–D338 (2019).
- 642 45. Han, E.-K., Cotty, F., Sottas, C., Jiang, H. & Michels, C. A. Characterization of AGT1  
643 encoding a general alpha-glucoside transporter from *Saccharomyces*. *Mol. Microbiol.* **17**,  
644 1093–1107 (1995).
- 645 46. Salema-Oom, M., Pinto, V. V., Gonçalves, P. & Spencer-Martins, I. Maltotriose  
646 Utilization by Industrial. *Society* **71**, 5044–5049 (2005).
- 647 47. Horák, J. Regulations of sugar transporters: insights from yeast. *Curr. Genet.* **59**, 1–31  
648 (2013).
- 649 48. Dietvorst, J., Londesborough, J. & Steensma, H. Y. Maltotriose utilization in lager yeast  
650 strains: MTTI encodes a maltotriose transporter. *Yeast* **22**, 775–788 (2005).
- 651 49. Diderich, J. A., Weening, S. M., van den Broek, M., Pronk, J. T. & Daran, J.-M. G.  
652 Selection of Pof-*Saccharomyces eubayanus* Variants for the Construction of *S. cerevisiae*  
653 × *S. eubayanus* Hybrids With Reduced 4-Vinyl Guaiacol Formation. *Front. Microbiol.* **9**,  
654 1640 (2018).
- 655 50. Mukai, N., Masaki, K., Fujii, T., Kawamukai, M. & Iefuji, H. PAD1 and FDC1 are  
656 essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J.*  
657 *Biosci. Bioeng.* **109**, 564–569 (2010).
- 658 51. Shen, X.-X. *et al.* Tempo and Mode of Genome Evolution in the Budding Yeast  
659 Subphylum. *Cell* **175**, 1533-1545.e20 (2018).
- 660 52. Bing, J., Han, P.-J., Liu, W.-Q., Wang, Q.-M. & Bai, F.-Y. Evidence for a Far East Asian  
661 origin of lager beer yeast. *Curr. Biol.* **24**, R380-1 (2014).
- 662 53. Borneman, A. R., Forgan, A. H., Pretorius, I. S. & Chambers, P. J. Comparative genome  
663 analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Res.* **8**, 1185–1195

- 664 (2008).
- 665 54. Borneman, A. R. *et al.* Whole-Genome Comparison Reveals Novel Genetic Elements  
666 That Characterize the Genome of Industrial Strains of *Saccharomyces cerevisiae*. *PLoS*  
667 *Genet.* **7**, e1001287 (2011).
- 668 55. Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A. & Schmidt, S. A. Whole  
669 Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy Across  
670 the Spectrum of Commercial Wine Strains of *Saccharomyces cerevisiae*. *G3* **6**, 957–971  
671 (2016).
- 672 56. Dunn, B., Richter, C., Kvittek, D. J., Pugh, T. & Sherlock, G. Analysis of the  
673 *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed  
674 in diverse yeast strains from differing industrial environments. *Genome Res.* **22**, 908–924  
675 (2012).
- 676 57. Gayevskiy, V. & Goddard, M. R. *Saccharomyces eubayanus* and *Saccharomyces*  
677 *arboricola* reside in North Island native New Zealand forests. *Environ. Microbiol.* **18**,  
678 1137–1147 (2016).
- 679 58. Gayevskiy, V., Lee, S. & Goddard, M. R. European derived *Saccharomyces cerevisiae*  
680 colonisation of New Zealand vineyards aided by humans. *FEMS Yeast Res.* **16**, 1–12  
681 (2016).
- 682 59. Hewitt, S. K., Donaldson, I. J., Lovell, S. C. & Delneri, D. Sequencing and  
683 characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of  
684 chimeric genes and gives evidence of breakpoint reuse. *PLoS One* **9**, e92203 (2014).
- 685 60. Hose, J. *et al.* Dosage compensation can buffer copynumber variation in wild yeast. *Elife*  
686 **4**, 1–28 (2015).

- 687 61. Krogerus, K., Preiss, R. & Gibson, B. A unique *Saccharomyces cerevisiae* ×  
688 *Saccharomyces uvarum* hybrid isolated from norwegian farmhouse beer: Characterization  
689 and reconstruction. *Front. Microbiol.* **9**, 1–15 (2018).
- 690 62. Okuno, M. *et al.* Next-generation sequencing analysis of lager brewing yeast strains  
691 reveals the evolutionary history of interspecies hybridization. *DNA Res.* **1**, 1–14 (2016).
- 692 63. Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome  
693 Sequences and Strain Resources for the *Saccharomyces sensu stricto* Genus. *G3* **1**, 11–25  
694 (2011).
- 695 64. Skelly, D. A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic  
696 diversity in budding yeast. *Genome Res.* **23**, 1496–1504 (2013).
- 697 65. Strobe, P. K. *et al.* The 100-genomes strains, an *S. cerevisiae* resource that illuminates its  
698 natural phenotypic and genotypic variation and emergence as an opportunistic pathogen.  
699 *Genome Res.* **125**, 762–774 (2015).
- 700 66. van den Broek, M. *et al.* Chromosomal copy number variation in *Saccharomyces*  
701 *pastorianus* is evidence for extensive genome dynamics in industrial lager brewing strains.  
702 *Appl. Environ. Microbiol.* **81**, 6253–6267 (2015).
- 703 67. Yue, J. X. *et al.* Contrasting evolutionary genome dynamics between domesticated and  
704 wild yeasts. *Nat. Genet.* **49**, 913–924 (2017).
- 705 68. Zheng, D. Q. *et al.* Genome sequencing and genetic breeding of a bioethanol  
706 *Saccharomyces cerevisiae* strain YJS329. *BMC Genomics* **13**, (2012).
- 707 69. Bergström, A. *et al.* A high-definition view of functional genetic variation from natural  
708 yeast genomes. *Mol. Biol. Evol.* **31**, 872–88 (2014).
- 709 70. Akao, T. *et al.* Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai

710 no. 7. *DNA Res.* **18**, 423–434 (2011).

711 71. Almeida, P. *et al.* A population genomics insight into the Mediterranean origins of wine  
712 yeast domestication. *Mol. Ecol.* **24**, 5412–5427 (2015).

713 72. Baker, E. *et al.* The genome sequence of *Saccharomyces eubayanus* and the domestication  
714 of lager-brewing yeasts. *Mol. Biol. Evol.* **32**, 2818–2831 (2015).

715 73. Langdon, Q. K., Peris, D., Kyle, B. & Hittinger, C. T. sppIDer: A Species Identification  
716 Tool to Investigate Hybrid Genomes with High-Throughput Sequencing. *Mol. Biol. Evol.*  
717 **35**, 2835–2849 (2018).

718 74. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341  
719 (2009).

720 75. Liti, G. *et al.* High quality de novo sequencing and assembly of the *Saccharomyces*  
721 *arboricolus* genome. *BMC Genomics* **14**, (2013).

722 76. Peris, D. *et al.* Biotechnology for Biofuels Hybridization and adaptive evolution of diverse  
723 *Saccharomyces* species for cellulosic biofuel production. *Biotechnol. Biofuels* **10**, 1–19  
724 (2017).

725 77. Teytelman, L. *et al.* Impact of Chromatin Structures on DNA Processing for Genomic  
726 Analyses. *PLoS One* **4**, e6700 (2009).

727 78. Suzuki, R. & Shimodaira, H. Pvclust: An R package for assessing the uncertainty in  
728 hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

729 79. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for  
730 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

731 80. Zhou, X. *et al.* In Silico Whole Genome Sequencer and Analyzer ( iWGS ): a  
732 Computational Pipeline to Guide the Design and Analysis of de novo Genome Sequencing

- 733 Studies. *G3* **6**, 3655–3662 (2016).
- 734 81. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic  
735 framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
- 736 82. Foury, F., Roganti, T., Lecrenier, N. & Purnelle, B. The complete sequence of the  
737 mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* **440**, 325–331 (1998).
- 738 83. Sulo, P. *et al.* The evolutionary history of *Saccharomyces* species inferred from completed  
739 mitochondrial genomes and revision in the ‘yeast mitochondrial genetic code’. *DNA Res.*  
740 **24**, 571–583 (2017).
- 741 84. Peris, D. *et al.* Molecular Phylogenetics and Evolution Mitochondrial introgression  
742 suggests extensive ancestral hybridization events among *Saccharomyces* species. *Mol.*  
743 *Phylogenet. Evol.* **108**, 49–60 (2017).
- 744 85. Johnson, M. G. *et al.* HybPiper: Extracting Coding Sequence and Introns for  
745 Phylogenetics from High- Throughput Sequencing Reads Using Target Enrichment. *Appl.*  
746 *Plant Sci.* **4**, (2016).
- 747 86. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform  
748 for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
- 749 87. Leigh, J. W. & Bryant, D. POPART: Full-feature software for haplotype network  
750 construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
- 751 88. Clement, M., Snell, Q., Walke, P., Posada, D. & Crandall, K. TCS: estimating gene  
752 genealogies. in *Proceedings 16th International Parallel and Distributed Processing*  
753 *Symposium 7* pp (IEEE, 2002). doi:10.1109/IPDPS.2002.1016585
- 754 89. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA  
755 polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).

- 756 90. Walther, A., Hesselbart, A. & Wendland, J. Genome Sequence of *Saccharomyces*  
757 *carlsbergensis*, the World's First Pure Culture Lager Yeast. *G3* **4**, 783–793 (2014).
- 758 91. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:  
759 Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- 760 92. Csardi, G. & Nepusz, T. The igraph software package for complex network research.  
761 *InterJournal* **1695**, 1–9 (2006).
- 762 93. Opuente, D. A. *et al.* Factors driving metabolic diversity in the budding yeast subphylum.  
763 *BMC Biol.* **16**, 1–15 (2018).
- 764 94. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular  
765 evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–  
766 1549 (2018).
- 767 95. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of  
768 large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- 769 96. Letunic, I. & Bork, P. Interactive tree of life ( iTOL ) v3: an online tool for the display and  
770 annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
- 771 97. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers.  
772 *Bioinformatics* **24**, 1403–1405 (2008).
- 773 98. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York,  
774 2009).
- 775 99. Pfeifer, B. & Wittelsbueger, U. Package 'PopGenome '. (2015). doi:10.1111/rssb.12200  
776
- 777 Figure Legends

778 Figure 1. Summary of genomic contributions and isolation environments for interspecies  
779 hybrids. (a) Hybrids were clustered by genomic contributions. Lager strains are in the bottom  
780 half, *S. uvarum* × *S. eubayanus* strains are at the top, and most complex hybrids are in the  
781 middle, except for the single *S. cerevisiae* × *S. eubayanus* × *S. kudriavzevii* hybrid (very bottom).  
782 Individual hybrid strains are along the y-axis, and the genomes of the species contributing to  
783 hybrids are along the x-axis. *S. cerevisiae* (*Scer*) is in red, *S. kudriavzevii* (*Skud*) is in green, *S.*  
784 *uvarum* (*Suva*) is in purple, and *S. eubayanus* (*Seub*) is in pink. Dotted lines indicate  
785 chromosomes. Ploidy estimates are indicated by opacity, where darker regions are higher ploidy.  
786 (b) Counts of hybrids isolated from different environments. The lagers have been split into Saaz  
787 and Frohberg lineages. Other is grouped with Unknown and represents one isolate from a  
788 distillery. Tables S1 & S3 includes all isolation information and metadata.

789

790 Figure 2. Population and phylogenomic analyses of *S. cerevisiae*, *S. kudriavzevii*, *S. uvarum*, *S.*  
791 *eubayanus*, and their hybrid sub-genomes.

792 All phylogenies were built with RAxML with pure strains of a species and any hybrids with  
793 >50% complete sub-genome for given species. Bootstrap support values >70% are shown as  
794 gray dots. Branches are colored by the origin of isolation for each strain. Each hybrid has a  
795 stacked bar plot showing the genomic content for each species contributing to their genome;  
796 species colors are the same as in Figure 1a. For the Principal Component Analyses (PCA), dot  
797 colors represent strains' origins, and color clouds represent populations or lineages. The axes of  
798 all PCAs are scaled to the same range. Phylogenies with strain names, Newick formatted files,  
799 and data frames used to build PCAs are available as Figures S2, S4, S6, and S8; Table S2; and  
800 Files S1, S3, S5, and S7. (a) Left: Phylogeny of *S. kudriavzevii* with 30 strains and 38,992 SNPs

801 from across the genome and rooted with an Asia B strain, IFO1803 (removed for clarity). Right:  
802 Principal component projection for PC1 and PC2, excluding Asia B. (b) Phylogeny of *S.*  
803 *eubayanus* with 92 strains and 18,878 SNPs from across the genome, rooted with Population A  
804 (PopA). Right: Principal component projection of PC1 and PC2. (c) Phylogeny for *S. uvarum*  
805 with 82 strains and 18,652 SNPs from across the genome, rooted with the Australasian lineage  
806 (removed for clarity). Right: Principal component projection for PC1 and PC2, excluding the  
807 Australasian lineage. (d) Top: Phylogeny for *S. cerevisiae* with 612 phased (for strains with  
808 >20K heterozygous sites) or unphased haplotypes and 21,222 SNPs from across the genome,  
809 rooted with the Taiwanese strain EN14S01 (removed for clarity). Previously identified wild  
810 lineages from West Africa, Malaysia, North America, Japan, and the Philippines are included in  
811 the Wild Misc group<sup>11,74</sup>. The other lineages are named in a similar manner to previous studies  
812 on ale-brewing and Mediterranean Oak (MedOak) strains<sup>8,9,71</sup>. Bottom: Principal component  
813 projection for PC1 and PC2 (including EN14S01, which groups with Sake/Asian).

814

815 Figure 3. Mitochondrial genome inheritance in interspecies hybrids.

816 (a) The bar plots show proportion of 1:1:1:1 ortholog content for each sub-genome for each  
817 hybrid grouped by the mitochondrial genome (mtDNA) parent, which are labeled across the top.  
818 Colors represent different parent species and are that same as in of Figure 1a. (b) Analysis of  
819 concordance between which mtDNA was inherited and which parent contributed the most  
820 complete set of orthologous genes. “True” includes hybrids that inherited the most nuclear gene  
821 content from the same species as the mtDNA. “False” includes hybrids with mtDNA that did not  
822 match the species that contributed the most nuclear gene content. Colors represent the mtDNA  
823 parent, and shapes represent the largest nuclear genome contributor. The middle of the box plot

824 corresponds to the median, the upper and lower limits are the 75<sup>th</sup> and 25<sup>th</sup> percentiles  
825 respectively, and the whiskers extend to the largest or smallest value no greater than 1.5 × the  
826 differences between the 75<sup>th</sup> and 25<sup>th</sup> percentiles. There was a significant correlation between the  
827 mtDNA parent and the largest nuclear genomic contributor (logistic regression  $p=3.58E-8$ , AIC=  
828 118.21). Notably, the *S. eubayanus* × *S. uvarum* hybrids, which have often undergone many  
829 backcrossing events, follow this trend and are both cryotolerant species. (c) Linear relationship  
830 of the number of 1:1:1:1 orthologs versus the number of nuclear-encoded, mitochondrially  
831 localized genes present in the sub-genome that matches the mtDNA (linear regression  $p=2.0E-$   
832 16, AIC= 1151.5). The inset shows the mean proportion of mitochondrially localized versus all  
833 other nuclear genes present in the sub-genome that matches the mitochondrial parent ( $p =$   
834 0.8612, odds ratio = 0.9653).

835

836 Figure 4. Hybrid inheritance and functionality of genes responsible for 4-vinyl guaiacol (4-VG)  
837 production.

838 Retention of the regions where the adjacent *PADI* and *FDCI* genes, which are both required for  
839 4-VG production, are located in each parent species (a-c), shown as 10-kbp windows of ploidy  
840 estimates over last 100-kbp of the chromosome. Gene locations are represented by black dotted  
841 lines. Higher opacity represents higher ploidy. Species colors are that same as in Figure 1a. *Scer*  
842 = *S. cerevisiae*, *Spar* = *Saccharomyces paradoxus*, *Smik* = *Saccharomyces mikatae*, *Skud* = *S.*  
843 *kudriavzevii*, *Suva* = *S. uvarum*, and *Seub* = *S. eubayanus*. (a) *Scer* × *Skud* hybrids: all strains  
844 inherited versions of both *PADI* and *FDCI* from *Scer* that are predicted to be functional, + | +,  
845 but they have lost the *Skud* alleles. (b) *Suva* × *Seub* hybrids: all strains inherited versions of  
846 *PADI* and *FDCI*, from either *Suva* or *Seub*, that are predicted to be functional, + | +. (c) All

847 lager strains have completely lost the region in the *Seub* genome where these genes reside.  
848 Additionally, all Saaz strains have also completely lost the *Scer* versions of these genes,  $\Delta$  |  $\Delta$ .  
849 All but two Frohberg strains have retained versions of *PADI* from *Scer* that are predicted to be  
850 functional, but inherited *Scer* alleles of *FDCI* that are predicted to be inactive due to a frameshift  
851 mutation, + |  $\Psi$ . Haplotype networks were built for the amino acid sequences for Fdc1 (d) and  
852 Pad1 (e). Colored pies correspond to *Scer* lineages, hybrids, or wild species with size  
853 representing the number of strains with that haplotype. Non-*Scer* nodes or groups of nodes are  
854 labeled by the species to which they correspond. Colored clouds correspond to communities: red  
855 is mostly *Scer*, blue is mostly non-*Scer* (including *Seub* and *Suva*), yellow is mostly *Spar* and  
856 *Smik*, green is mostly *Skud*, and gray is mostly loss-of-function alleles. Pseudogenes are marked  
857 as  $\Psi$  with additional information about the loss-of-function nucleotide and amino-acid changes.  
858 Dotted connections represent >100 amino acid differences.

859

860 Figure 5. Summary of hybrids and origin of lager traits.

861 (a) Simplified summary of parents and resulting hybrids. On the left is a cladogram of just the  
862 *Saccharomyces* species that have contributed to fermented beverage hybrids. Three distinct  
863 lineages of *S. cerevisiae* (*Scer*) have contributed to hybrids; for the wild parents (*S. kudriavzevii*  
864 (*Skud*), *S. uvarum* (*Suva*), and *S. eubayanus* (*Seub*)), Holarctic or European lineages gave rise to  
865 the hybrids. Gray lines point from each parent to the resulting hybrid. The order of secondary or  
866 tertiary hybridization events was inferred from genome composition. This simplified view does  
867 not show when multiple lineages of *Scer* have contributed to different hybrid types (e.g. *Scer*  $\times$   
868 *Skud* hybrids), backcrossing (e.g. *Seub*  $\times$  *Suva* hybrids), or minor subtelomeric contributions (e.g.  
869 small *Scer* contributions to some *Seub*  $\times$  *Suva* hybrids). (b) Summary of how lager-brewing

870 yeasts acquired their unique trait profile. The two lager-brewing lineages, Saaz and Frohberg,  
871 arose out of hybridizations between domesticated *Scer* ale strains and wild *Seub* strains. The *Scer*  
872 strains could utilize maltotriose (+), did not produce phenolic-off-flavor (POF<sup>-</sup>), and preferred  
873 warmer temperatures (☀), while the *Seub* strains tolerated colder temperatures (\*), could not  
874 use maltotriose (-), and produced phenolic-off-flavors (POF<sup>+</sup>). The two lager-brewing lineages  
875 inherited the *Seub* mitochondrial genome (pink circle), which partly conferred cryotolerance.  
876 Both lineages also inherited maltotriose transporter genes from both parents (*MTT1* from *Scer*  
877 and *SeAGT1* from *Seub*). Finally, both lineages convergently became POF<sup>-</sup> through multiple  
878 distinct mechanisms, including pre-adaptation in the *S. cerevisiae* ale-brewing parent due to a  
879 mutated pseudogene (*PADI* | *fdc1Ψ* in red), aneuploidy removing functional *S. eubayanus* genes  
880 (*pad1Δ* | *fdc1Δ* in pink), and translocations in all Saaz strains and some Frohberg strains (*pad1Δ* |  
881 *fdc1Δ* in red).

882

883 Figure S1. Genomic contribution comparison of Muri and WLP351.

884 Modified sppIDer plot, where the y-axis is estimated ploidy, rather than coverage, for the *S.*  
885 *cerevisiae* (50%) × *S. eubayanus* (5%) × *S. uvarum* (45%) strains Muri<sup>61</sup> and WLP351.

886

887 Figure S2. Summary of total genomic coverage and shared translocations.

888 The minimum and maximum normalized coverage of all strains that contain each chromosome  
889 are shown as colored bars. Darker chromosomes mean that chromosome is present in more  
890 strains. Vertical dotted lines represent translocations that are shared in at least four strains,  
891 including between hybrid types. The color of the line represents the reciprocal species. (a) Only

892 lager strains and translocations found only in lagers. (b) All 122 hybrids and interspecies  
893 translocations.

894

895 Figure S3. Phylogenomic trees for *S. kudriavzevii* with strains labeled.

896 (a) Phylogeny identical to Figure 2a with strains labeled. (b) Phylogeny identical to Figure S4  
897 with strains labeled. Newick files are available as Files S1 & S2.

898

899 Figure S4. Phylogenomic and population placement of hybrids with minor *S. kudriavzevii*  
900 contributions.

901 (a) Phylogenomic tree built with 36 strains and 12,424 SNPs from regions of the genome that  
902 exist in at least one minor contributing hybrid. Bootstrap support values >70% are shown as gray  
903 dots. Branch colors represent origin of isolation. The inner colors correspond to origin or  
904 population. Outer stacked bar plots show the genomic content for each of the hybrids; species  
905 colors match Figure 1a. (b) PCA using whole genome data for European *S. kudriavzevii* strains  
906 and all major contributor hybrids. (c) PCA using a reduced genome (67%) but including  
907 additional minor hybrids. Phylogenies with strain names, Newick formatted files, and data  
908 frames used to build PCAs are available as Figure S3, Table S2, and File S2.

909

910 Figure S5. Phylogenomic trees for *S. eubayanus* with strains labeled

911 (a) Phylogeny identical to Figure 2b with strains labeled. (b) Phylogeny identical to Figure S6  
912 with strains labeled. Newick files available as Files S3 & S4.

913

914 Figure S6. Phylogenomic and population placement of hybrids with minor *S. eubayanus*  
915 contributions.  
916 (a) Phylogenomic tree built with 112 strains and 69,631 SNPs from regions of the genome that  
917 exist in at least one minor contributing hybrid. Bootstrap support values >70% are shown as gray  
918 dots. Branch colors represent origin of isolation. The inner colors correspond to origin or  
919 population. Outer stacked bar plots show the genomic content for each of the hybrids; species  
920 colors match Figure 1a. Long branches are biased by the extensive missing data in hybrids with  
921 very small contributions from *S. eubayanus*. (b) PCA using whole genome data for Holarctic *S.*  
922 *eubayanus* strains and all major contributor hybrids. (c) PCA using a reduced genome (25%) but  
923 including additional minor hybrids. Phylogenies with strain names, Newick formatted files, and  
924 data frames used to build PCAs are available as Figure S5, Table S2, and File S4.

925

926 Figure S7. Phylogenomic trees for *S. uvarum* with strains labeled.

927 (a) Phylogeny identical to Figure 2c with strains labeled. (b) Phylogeny identical to Figure S8  
928 with strains labeled. Newick files are available as Files S5 & S6.

929

930 Figure S8. Phylogenomic and population placement of hybrids with minor *S. uvarum*

931 contributions.

932 (a) Phylogenomic tree built with 69 strains and 36,541 SNPs from regions of the genome that  
933 exist in at least one minor contributing hybrid. Bootstrap support values >70% are shown as gray  
934 dots. Branch colors represent origin of isolation. The inner colors correspond to origin or  
935 population. Outer stacked bar plots show the genomic content for each of the hybrids; species  
936 colors match Figure 1A. (b) PCA using whole genome data for Holarctic *S. uvarum* strains and

937 all major contributor hybrids. (c) PCA using a reduced genome (84%) but including additional  
938 minor hybrids. Phylogenies with strain names, Newick formatted files, and data frames used to  
939 build PCAs are available as Figure S7, Table S2, and File S6.

940

941 Figure S9. Phylogenomic tree for full *S. cerevisiae* analysis with strains labeled.

942 Phylogeny identical to Figure 2d with strains labeled. A Newick file is available as File S7.

943

944 Figure S10. Phylogenomic and population placement of lagers within the Ale/Beer1 clade.

945 (a) Phylogenomic tree built with 267 strains and 21,953 SNPs from the whole genome. The total  
946 number of Frohberg strains was down-sampled to match the same number of Saaz strains. The  
947 tree was rooted with the Wine strain DBVPG1106. Bootstrap support values >70% are shown as  
948 gray dots. Branch colors represent origin of isolation. The inner colors correspond to origin or  
949 population. Outer stacked bar plots show the genomic content for each of the hybrids; species  
950 colors match Figure 1a. (b) PCA using whole genome data for Ale/Beer1 strains, all Saaz strains,  
951 and the down-sampled set of Frohberg strains. The two lineages of lager strains form separate  
952 groups, but they do not cluster with any described geographical lineage of the Ale/Beer1 clade.  
953 Pure *S. cerevisiae* Ale/Beer1 strains outside of the labeled lineages are unplaced, including a  
954 cluster of Stout strains, Wheat strains, and mosaic strains that our analyses suggest share the  
955 most ancestry with lager-brewing yeasts. (c) PCA using all lager strains. The low diversity in the  
956 Frohberg lager strains drives PC1, which led us to balance the dataset by down-sampling this  
957 lineage. Phylogenies with strain names, Newick formatted files, and data frames used to build  
958 PCAs are available as Figure S11, Table S2, and File S8.

959

960 Figure S11. Phylogenomic tree for Ale/Beer1 *S. cerevisiae* analysis with strains labeled.  
961 Phylogeny identical to Figure S10 with strains labeled. A Newick file is available as File S8.  
962  
963 Figure S12. 1:1:1:1 orthologs present in hybrid genomes.  
964 (a) Stacked bar chart of all 1:1:1:1 orthologs present in hybrids. Strains are sorted from most to  
965 least ortholog content. Completeness of the ortholog set from the species that contributed the  
966 most (b) or least (c) orthologs to the strains. Strains are ordered independently in all panels.  
967  
968 Figures S13. Complete de novo genome assembly for all strains.  
969 Total assembled genome for each strain. Regions are colored by which parent could be assigned  
970 in the de novo assembly based on the sppIDer results. “Multi” are regions where reads from  
971 many species mapped at high coverage. “Unmapped” are novel regions assembled from reads  
972 that do not map to parent reference genomes. For each assembly, contigs are ordered from  
973 largest to smallest from left to right.  
974  
975 Figure S14. Mitochondrial genome haplotype network.  
976 Six mitochondrial genes were concatenated in 364 wild *Saccharomyces* strains and interspecies  
977 hybrids and used to build a TCS<sup>88</sup> phylogenetic network. Haplotype classification is provided in  
978 Table S9. Haplotypes are represented by circles, and circle size is scaled according to the  
979 haplotype frequency. Pie charts show the frequency of haplotypes based on species or hybrid  
980 designation. The number of mutations separating each haplotype are indicated by lines on the  
981 edges connecting the haplotype circles.  
982

983 Figure S15.

984 Labeled (in turquoise) haplotype networks for *PADI* and *FDCI*. Edge numbers are the number  
985 of amino acid changes. Networks correspond to those used in Figure 4 for the amino acid  
986 sequences of (a) Fdc1 and (b) Pad1. (b) A different haplotype network orientation of Figure 4E  
987 that increases the visibility of each community and haplotype. Table S9 contains the key to  
988 which strains belong to which haplotype.

989

990 Table S1. All hybrids and their parent contributions.

991 Table S2. PCA analyses.

992       Percent explained by each principal component included in column headers.

993 Table S3. Results of Fisher’s Exact Test and Bonferroni correction of mitochondrially localized  
994 genes.

995       mtInteracting = nuclear-encoded but mitochondrially localized gene.

996 Table S4. Summary of number of 1:1:1:1 orthologs present in each sub-genome.

997 Table S5. GO term results of genes found in novel regions of the de novo assembled genomes.

998 Table S6. Brewing relevant gene summaries.

999       “-“ Indicates when HybPiper failed to recover and assemble genes for this group or that  
1000 these assemblies failed our length and coverage cutoffs.

1001 Table S7. Metadata for all strains newly sequenced in this study.

1002       The “New hybrid” column denotes hybrid genome sequences that are newly published in  
1003 this study.

1004       *Scer* = *S. cerevisiae*, *Spar* = *Saccharomyces paradoxus*, *Smik* = *Saccharomyces mikatae*,  
1005       *Skud* = *S. kudriavzevii*, *Suva* = *S. uvarum*, and *Seub* = *S. eubayanus*.

1006 Table S8. Published data accession information.

1007 Table S9. Haplotype key for mitochondrial genomes, *PADI*, and *FDCI*.

1008 Dataset A only includes strains where *15S rRNA* could be assembled, while Dataset B has

1009 *15S rRNA* removed.

1010 Table S10. Regions used for minor contribution analyses.

1011

1012 File S1. Newick formatted file of the *S. kudriavzevii* phylogeny with major hybrids.

1013 File S2. Newick formatted file of the *S. kudriavzevii* phylogeny with minor hybrids.

1014 File S3. Newick formatted file of the *S. eubayanus* phylogeny with major hybrids.

1015 File S4. Newick formatted file of the *S. eubayanus* phylogeny with minor hybrids.

1016 File S5. Newick formatted file of the *S. uvarum* phylogeny with major hybrids.

1017 File S5. Newick formatted file of the *S. uvarum* phylogeny with minor hybrids.

1018 File S7. Newick formatted file of the *S. cerevisiae* phylogeny with all strains analyzed.

1019 File S8. Newick formatted file of the *S. cerevisiae* phylogeny of just the Ale/Beer1 clade.

1020

Figure 1

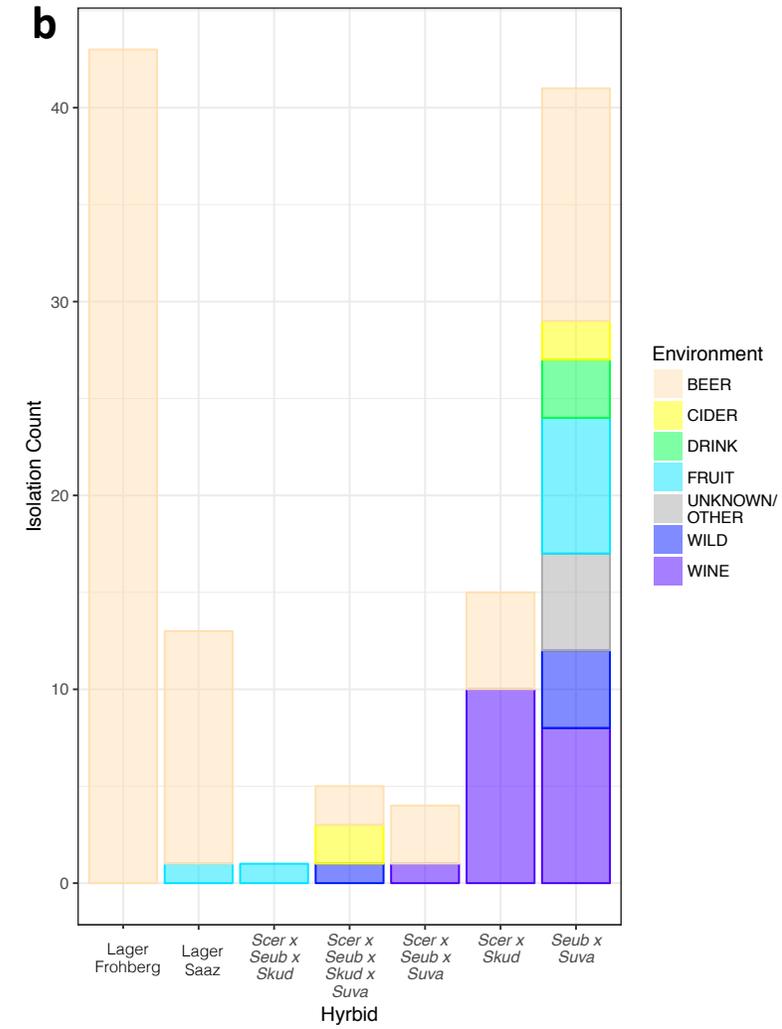
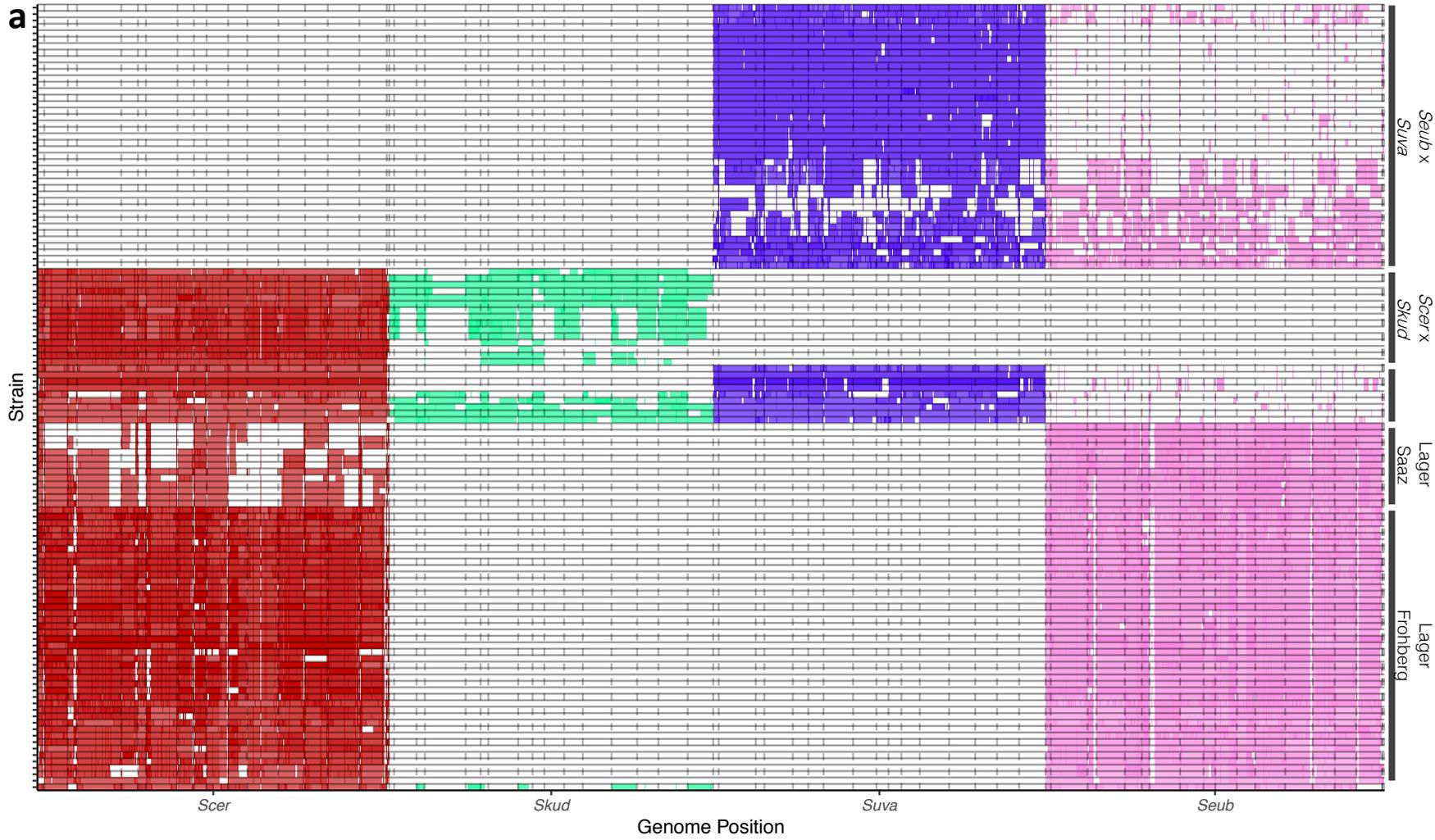


Figure 2

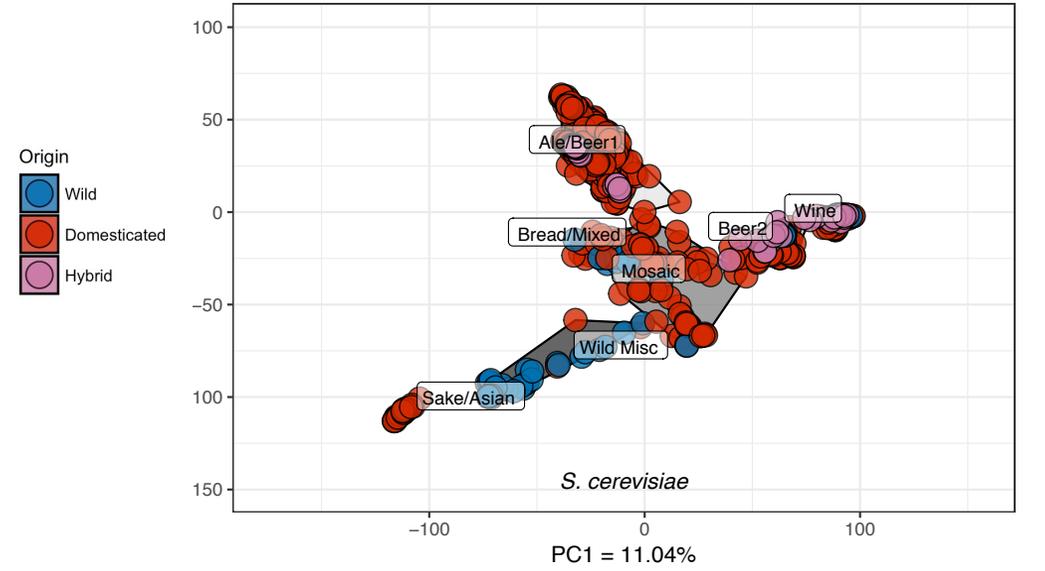
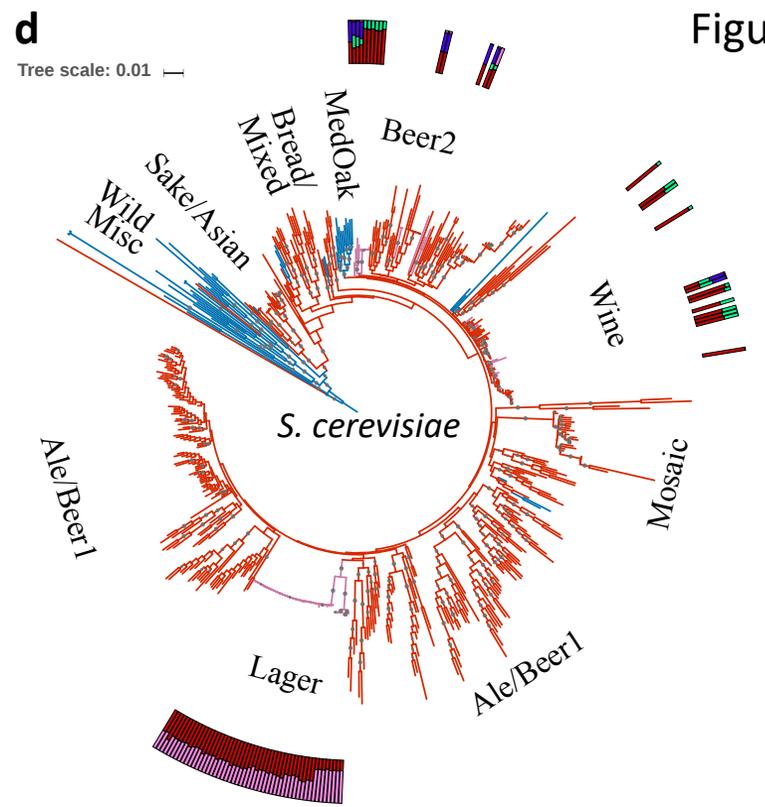
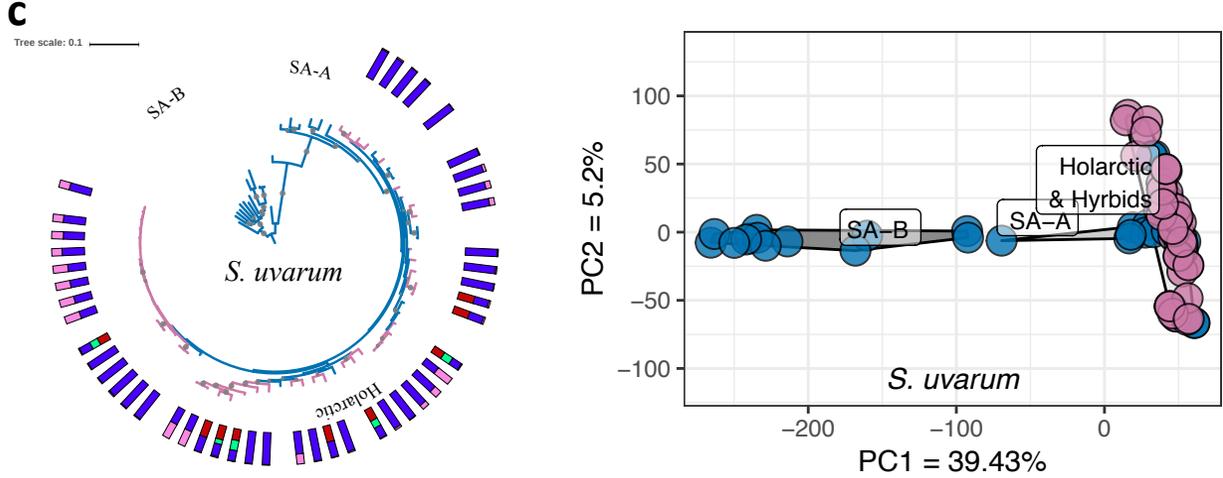
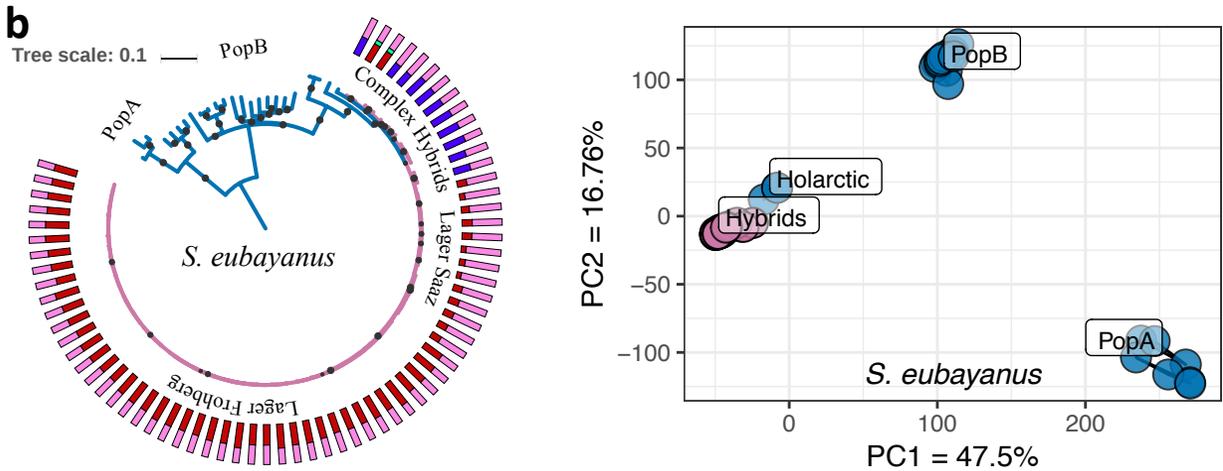
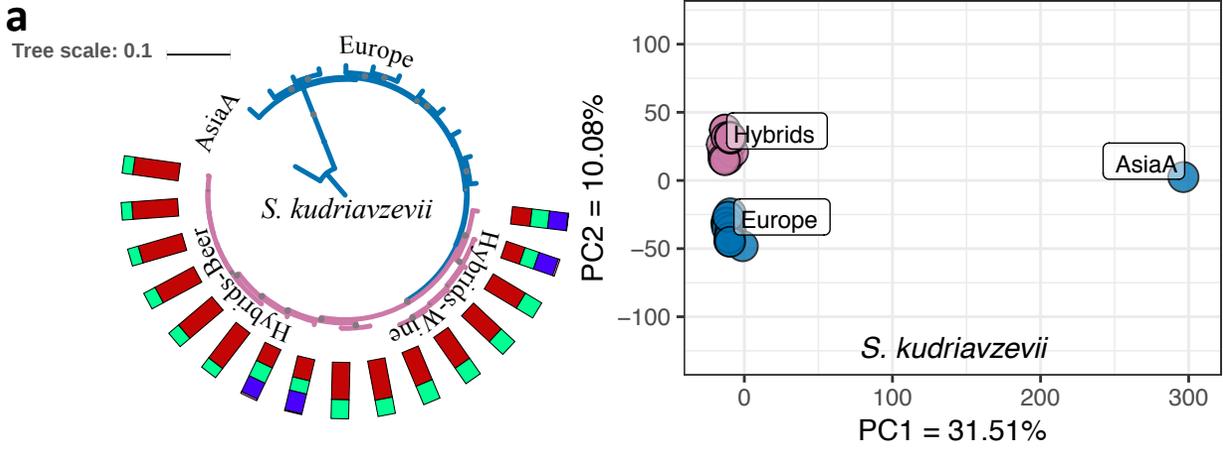
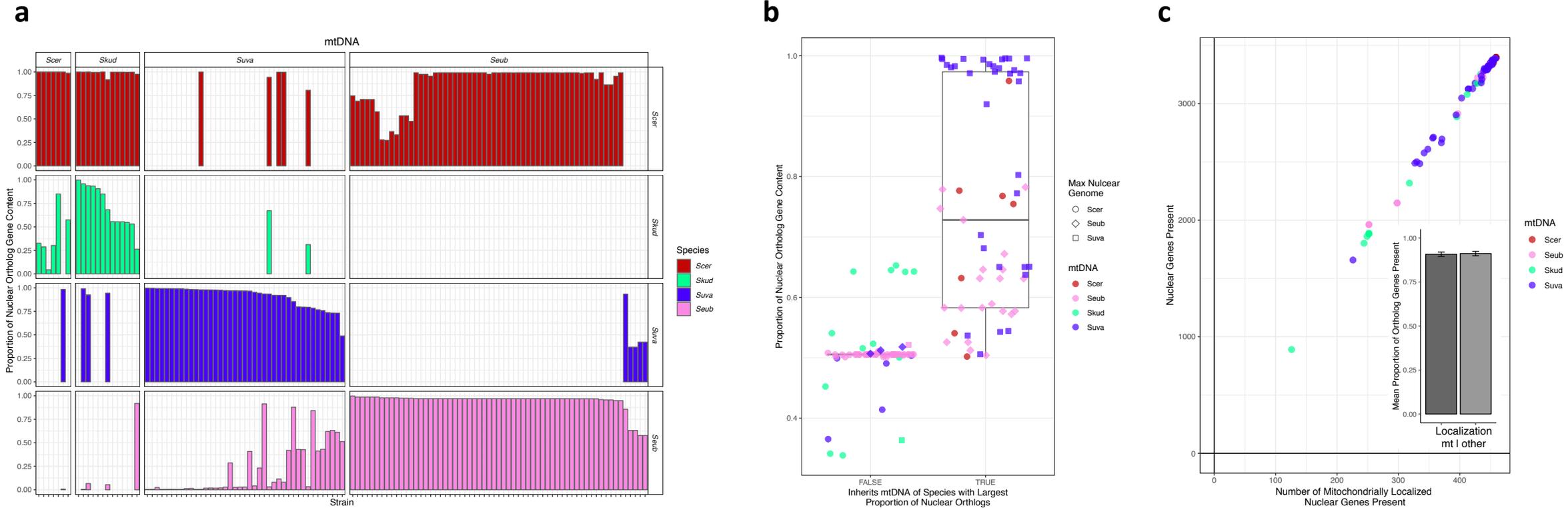


Figure 3



**Figure 4**

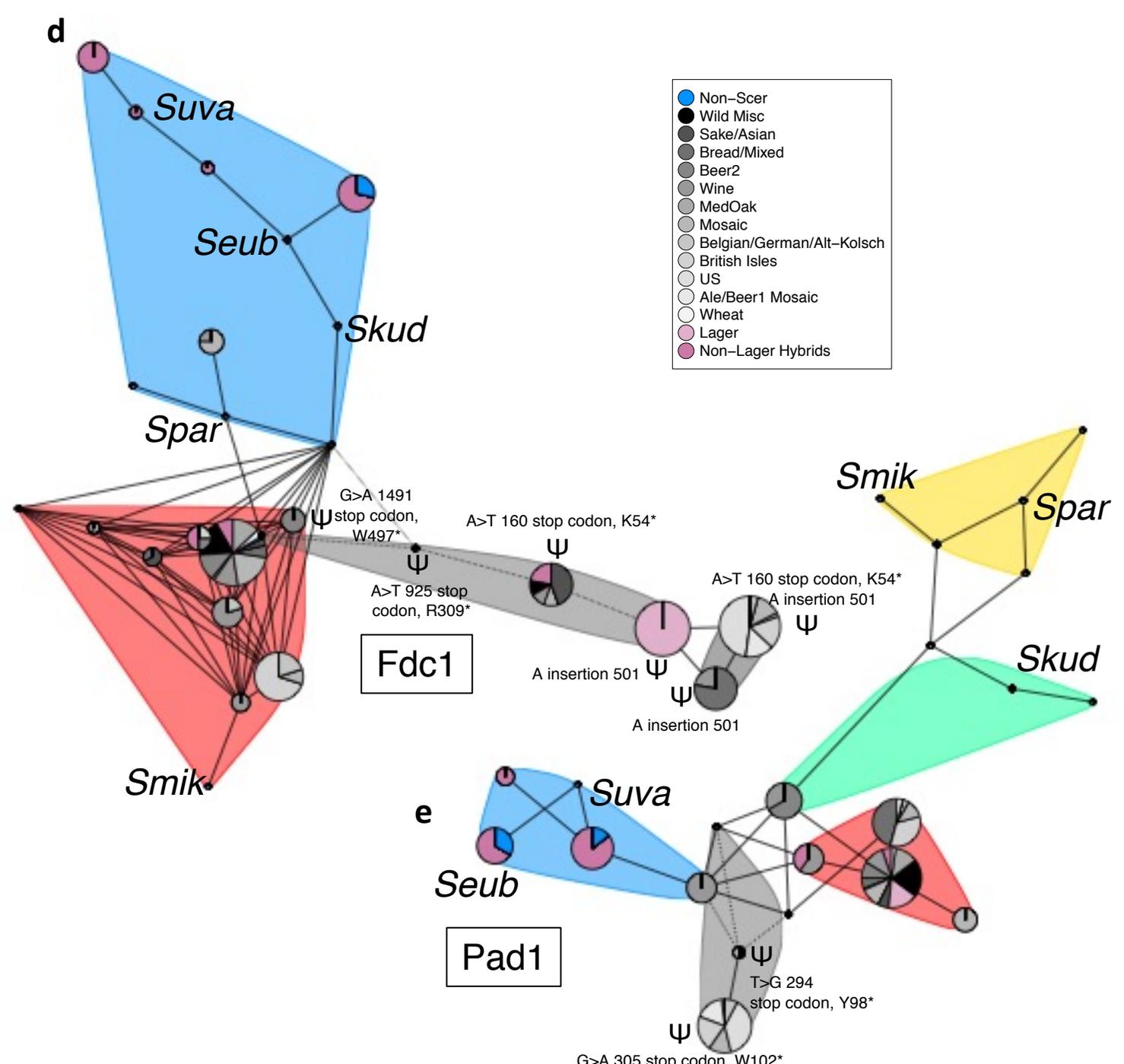
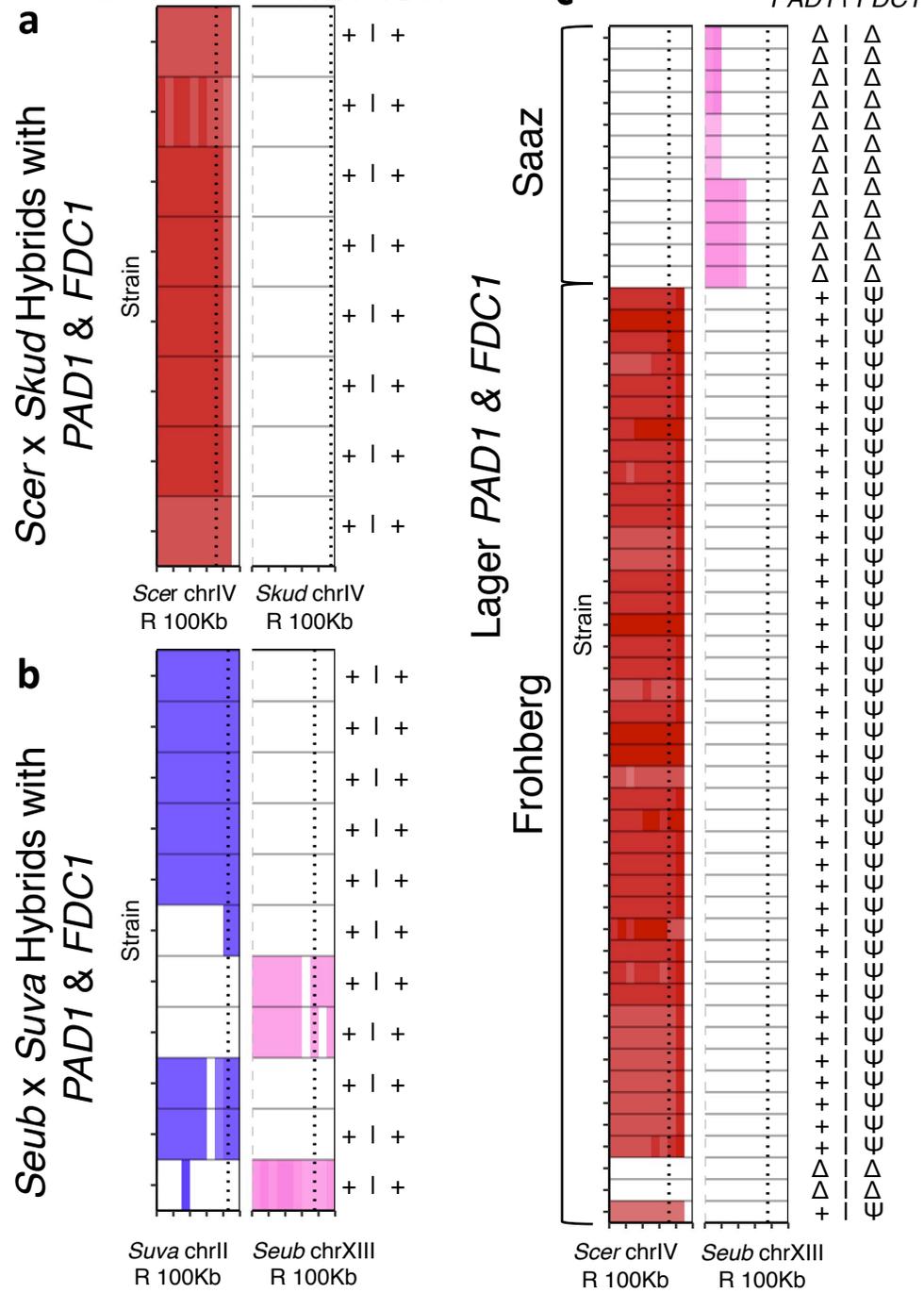


Figure 5

