

# Fermentation innovation through complex hybridization of wild and domesticated yeasts

Quinn K. Langdon<sup>1</sup>, David Peris<sup>1,2,3</sup>, Emily Clare P. Baker<sup>1,4</sup>, Dana A. Opulente<sup>1,2</sup>,  
Huu-Vang 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> and  
Chris Todd Hittinger<sup>1,2,4\*</sup>

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

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

Lager beers are made with hybrids between the distantly related species *S. cerevisiae* and *S. eubayanus*<sup>14–16</sup>. These hybrids combine unique properties from each: *S. cerevisiae*'s carbon use and fermentation capabilities combined with *S. eubayanus*'s cryotolerance to produce yeasts that could ferment well in the cold<sup>17–22</sup>. Other interspecies hybrids of *Saccharomyces* have been associated, both favourably and unfavourably, with diverse fermentations. *S. cerevisiae* × *S. kudriavzevii* hybrids are prized for their unique flavour profiles in beer and wine<sup>23</sup>. Conversely, hybrids and introgressed strains with large genomic contributions from *S. eubayanus* and *S. uvarum*, are viewed as contaminants in breweries due to the production of off-flavours, while other strains have been associated with sparkling wine and cider fermentation<sup>16,24–26</sup>. Although these previous studies have hinted at the complexity of fermentation hybrids, their focus on a handful of strains or a handful of loci has only given us a fleeting

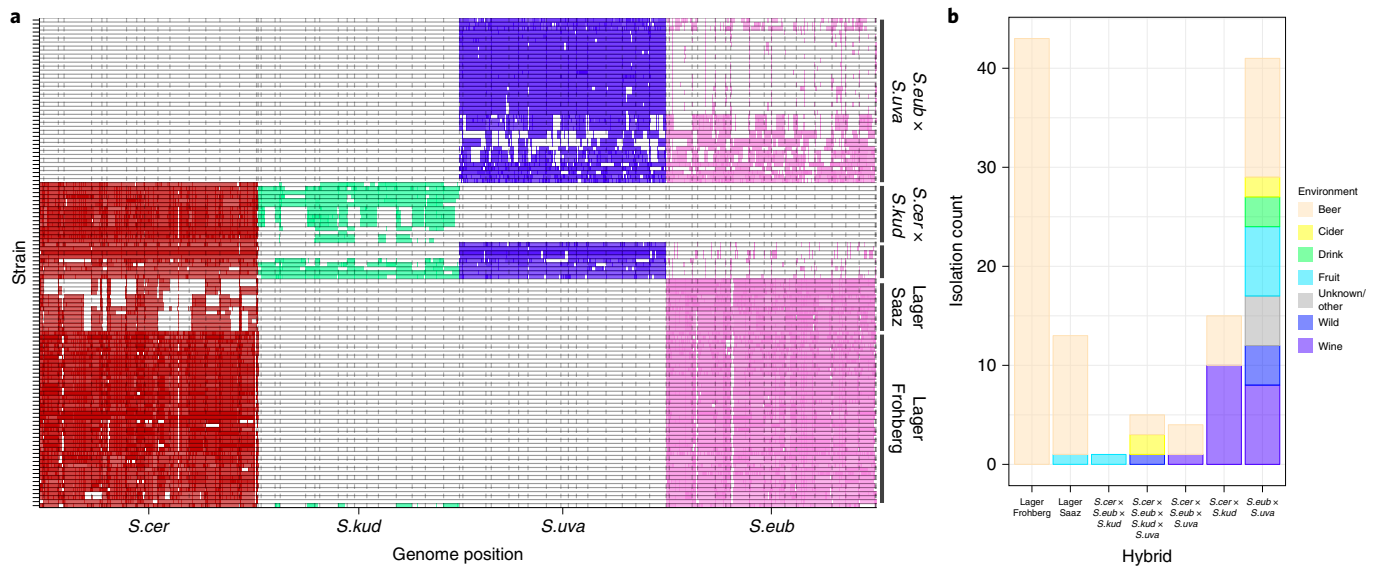
glimpse of the diversity of *Saccharomyces* hybrids, their total genomic compositions and their evolution.

Here we identified, sequenced and analysed the genomes of 122 interspecies hybrids and introgressed strains in the genus *Saccharomyces* to understand their origins and evolutionary innovations. This collection contains pairwise hybrids, as well as more complex hybrids and introgressed strains with three or four parent species. We show that all genomic contributions from *S. cerevisiae* have arisen out of three domesticated lineages of *S. cerevisiae*, while all other parents belonged to Holarctic or European wild lineages of their respective species. We also analysed inheritance of the mitochondrial genome (mtDNA) and the genetic events generating functional diversity in genes relevant to fermented beverages. The genomic complexity of these hybrids provides insight into their origins and evolutionary successes in human-associated fermentation environments.

## Results

**Summary of interspecies hybrid types.** Here, we analysed the genome sequences of 122 interspecies hybrids and introgressed strains of *Saccharomyces*, 63 strains of which are newly sequenced here, more than doubling the number of previously published hybrid genomes (Supplementary Dataset 1). Collectively, industrial settings dominated the isolation origins of all hybrids; 86% ( $n = 105$ ) were from beer, wine, cider, a distillery or other beverages (Fig. 1b,

<sup>1</sup>Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI, USA. <sup>2</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA. <sup>3</sup>Department of Food Biotechnology, Institute of Agrochemistry and Food Technology, CSIC, Valencia, Spain. <sup>4</sup>Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, USA. <sup>5</sup>Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France. <sup>6</sup>Department of Microbiology, School of Genetics and Microbiology, Trinity College Dublin, Dublin, Ireland. <sup>7</sup>UCIBIO-REQUIMTE, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal. <sup>8</sup>Laboratorio de Microbiología Aplicada, Biotecnología y Bioinformática de Levaduras, Instituto Andino Patagónico de Tecnologías Biológicas y Geoambientales, Consejo Nacional de Investigaciones Científicas y Técnicas-Universidad Nacional del Comahue, Bariloche, Argentina. \*e-mail: [cthittinger@wisc.edu](mailto:cthittinger@wisc.edu)



**Fig. 1 | Summary of genomic contributions and isolation environments for interspecies hybrids. a**, Hybrids were clustered by genomic contributions. Lager strains are in the bottom half, *S. uvarum* × *S. eubayanus* strains are at the top and most complex hybrids are in the middle, except for the single *S. cerevisiae* × *S. eubayanus* × *S. kudriavzevii* hybrid (very bottom). Individual hybrid strains are along the y axis and the genomes of the species contributing to hybrids are along the x axis. *S. cerevisiae* (*S.cer*) is in red, *S. kudriavzevii* (*S.kud*) is in green, *S. uvarum* (*S.uva*) is in purple and *S. eubayanus* (*S.eub*) is in pink. Dotted lines indicate chromosomes. Ploidy estimates are indicated by opacity, where darker regions are higher ploidy. **b**, Counts of hybrids isolated from different environments. The lagers have been split into Saaz and Froberg lineages. 'Other' is grouped with 'unknown' and represents one isolate from a distillery. Supplementary Datasets 1 and 11 include all isolation information and metadata.

Supplementary Dataset 1 and Supplementary Text). We identified four types of hybrids: (1) lager-like (*S. cerevisiae* × *S. eubayanus*) ( $n=56$ ); (2) *S. cerevisiae* × *S. kudriavzevii* ( $n=15$ ); (3) *S. eubayanus* × *S. uvarum* ( $n=41$ ); and (4) more complex hybrids, with three or four parent species ( $n=11$  more than doubling those previously identified<sup>26</sup>) (Fig. 1a, Supplementary Dataset 1 and Supplementary Text). These more complex hybrids fell into three groups: (4A) *S. cerevisiae* × *S. kudriavzevii* × *S. eubayanus* × *S. uvarum* ( $n=5$ ), (4B) *S. cerevisiae* × *S. eubayanus* × *S. uvarum* ( $n=4$ ) and (4C) one *S. cerevisiae* × *S. kudriavzevii* × *S. eubayanus* (Supplementary Dataset 1). The lager-like hybrids were almost exclusively associated with beer (Fig. 1b) and have genomic contributions that were consistent with previous observations in the two lineages, Saaz and Froberg<sup>27</sup>. The *S. cerevisiae* × *S. kudriavzevii* strains were associated with beer and wine (Fig. 1b). They had considerable differences in *S. kudriavzevii* genomic content, suggesting that these hybrids are of variable ages and evolutionary histories. The *S. eubayanus* × *S. uvarum* hybrids and introgressed strains were the most variable, both in isolation environment and genomic contributions (Fig. 1 and Supplementary Dataset 1). The wide range in genomic contributions in these strains was probably influenced by their ability to backcross due to the low but non-zero, spore viability of hybrids of these sister species<sup>16</sup>. These *S. eubayanus* × *S. uvarum* strains had the most total translocations ( $\chi^2=1,250.1$ , adjusted  $P=2.64 \times 10^{-15}$ ), as well as the most translocations shared with other hybrid types ( $\chi^2=15.964$ , adjusted  $P=0.0138$ ) (Supplementary Fig. 2). The shared nature of some of these translocations in hybrids with more than two parents suggests that *S. eubayanus* × *S. uvarum* introgressed strains further hybridized to produce some of the complex three or four parent species hybrids. Thus, these four types of hybrids each show unique dynamics in genome evolution and are used for different products that range from several regional niche beverages to the globally dominant beer style, lagers.

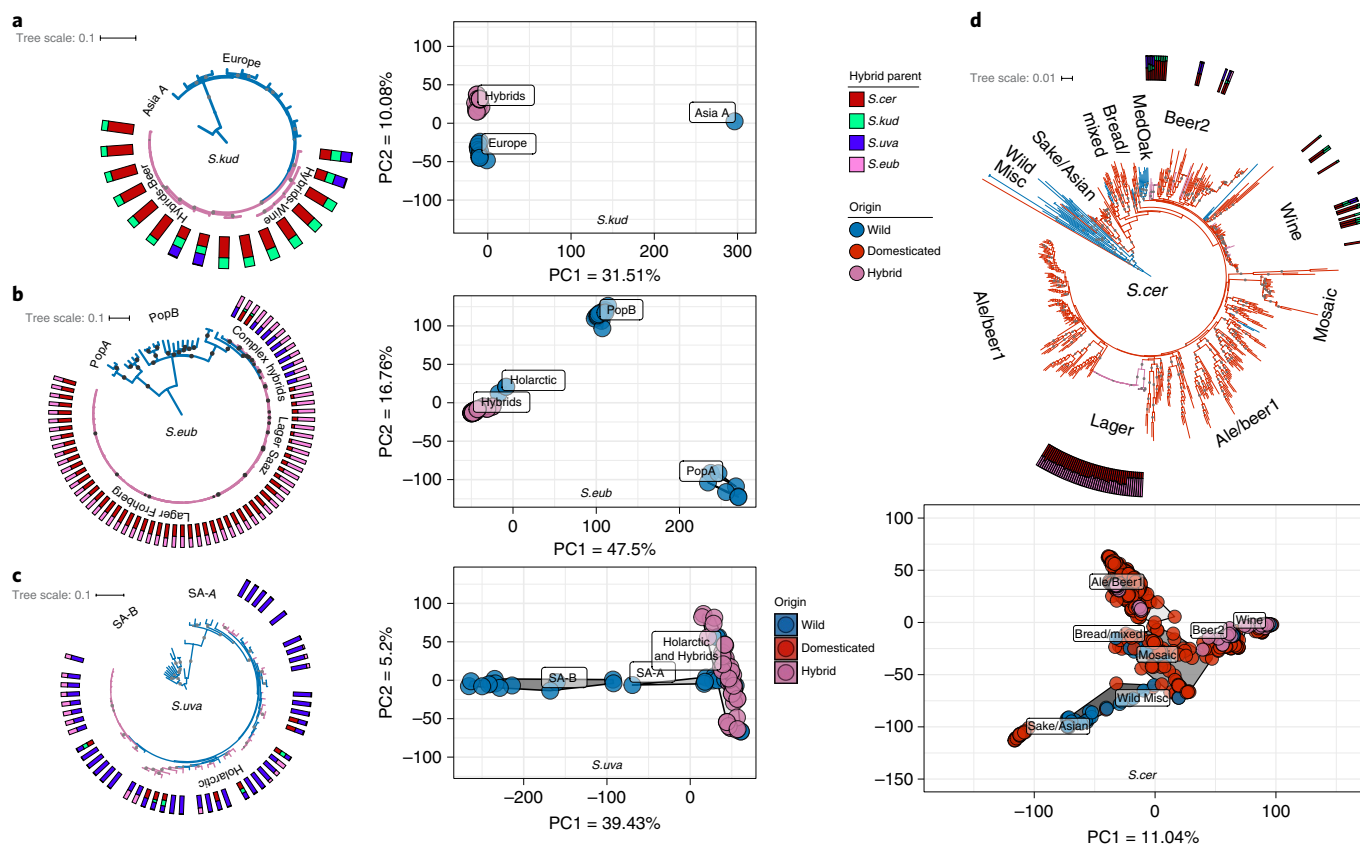
**Wild parent populations.** Three out of four of the species contributing to these hybrids (*S. kudriavzevii*, *S. uvarum* and *S. eubayanus*)

have primarily been isolated from wild settings and have global distributions with populations that reflect their geography<sup>28,29</sup>. We used these established populations and phylogenomic and principal component analysis (PCA) approaches to evaluate the origins of these hybrids (Supplementary Text).

*S. kudriavzevii* has been isolated in Europe and Asia and consists of three described populations: Asia A, Asia B and Europe<sup>23,30,31</sup>. The *S. kudriavzevii* subgenomes of the hybrids all clustered with the European population as a monophyletic clade (Fig. 2a, Supplementary Fig. 3, Supplementary Datasets 2 and 3 and Supplementary Text). These findings show that these hybrids were drawn from a closely related lineage of the European population of *S. kudriavzevii*.

In *S. eubayanus*, analysis of both large and small contributions, showed that these hybrids and introgressed strains clustered with the Holarctic lineage of *S. eubayanus* (Fig. 2b, Supplementary Fig. 5, Supplementary Datasets 2 and 5 and Supplementary Text). Our vastly expanded dataset suggests that the Holarctic lineage is the closest known relative of all industrially relevant *S. eubayanus* hybrids and introgressed strains. The array of hybrids observed here requires that multiple hybridization events occurred between this lineage and other species. We also analysed genetic diversity of the *S. eubayanus* contributions to industrial hybrids and introgressed strains (Supplementary Text). We found low nucleotide diversity in lager-like hybrids that shows that these widely used interspecies hybrids arose out of a narrow swath of *S. eubayanus* diversity, while the less frequently used hybrids and introgressed strains retained more nucleotide diversity.

*S. uvarum* has a parallel population structure to *S. eubayanus*<sup>26,32</sup>, with the exception of its increased isolation frequency in the Northern Hemisphere and the presence of pure strains isolated from Europe. Here we found that all contributions from *S. uvarum* arose out of the *S. uvarum* Holarctic lineage<sup>26</sup>. In contrast to our *S. eubayanus* findings, the *S. uvarum* subgenomes of these hybrids and introgressed strains were interspersed with pure wild strains (Fig. 2c, Supplementary Figs. 7



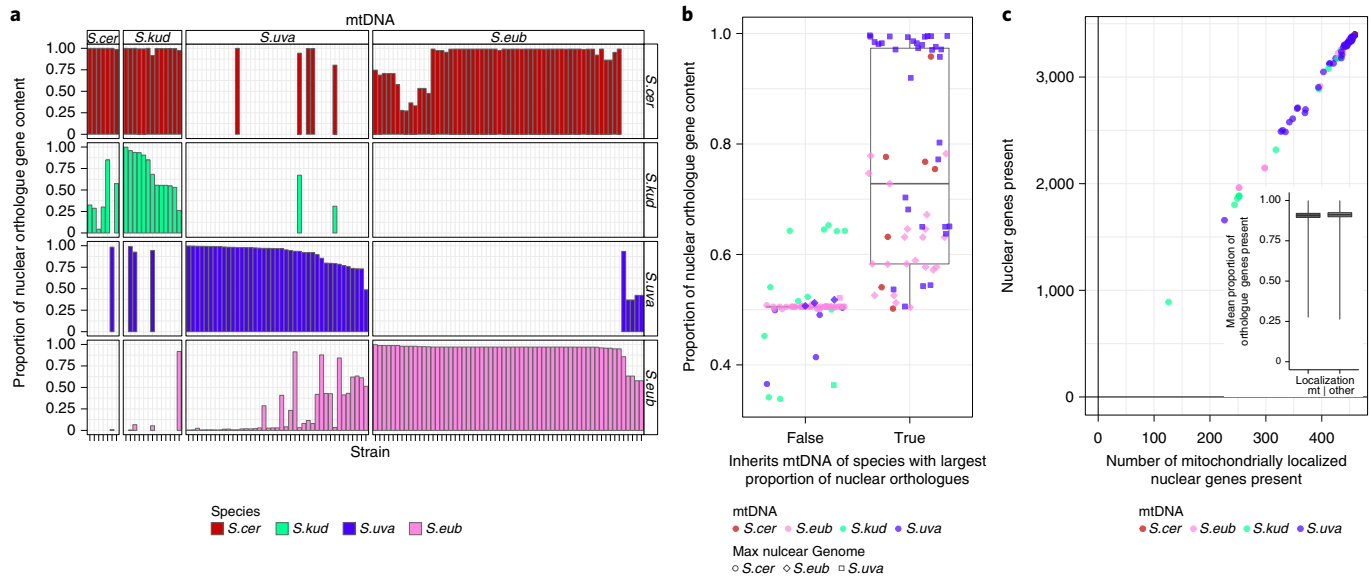
**Fig. 2 | Population and phylogenomic analyses of *S. cerevisiae*, *S. kudriavzevii*, *S. uvarum*, *S. eubayanus* and their hybrid subgenomes.** All phylogenies were built with RAxML with pure strains of a species and any hybrids with >50% complete subgenome for given species. Bootstrap support values >70% are shown as grey dots. Branches are coloured by the origin of isolation for each strain. Each hybrid has a stacked bar plot showing the genomic content for each species contributing to their genome; species colours are the same as in Fig. 1a. For the PCA, dot colours represent strains' origins and colour clouds represent populations or lineages. The axes of all PCAs are scaled to the same range. Phylogenies with strain names, Newick formatted files and data frames used to build PCAs are available as Supplementary Figs. 2, 4, 6 and 8 and Supplementary Datasets 2, 3, 5, 7 and 9. **a**, Left: Phylogeny of *S. kudriavzevii* with 30 strains and 38,992 SNPs from across the genome and rooted with an Asia B strain, IFO1803 (removed for clarity). Right: Principal component projection for PC1 and PC2, excluding Asia B. **b**, Phylogeny of *S. eubayanus* with 92 strains and 18,878 SNPs from across the genome, rooted with Population A (PopA). Right: Principal component projection of PC1 and PC2. **c**, Phylogeny for *S. uvarum* with 82 strains and 18,652 SNPs from across the genome, rooted with the Australasian lineage (removed for clarity). Right: Principal component projection for PC1 and PC2, excluding the Australasian lineage. SA-A, South America A; SA-B, South America B. **d**, Top: Phylogeny for *S. cerevisiae* with 612 phased (for strains with >20,000 heterozygous sites) or unphased haplotypes and 21,222 SNPs from across the genome, rooted with the Taiwanese strain EN14S01 (removed for clarity). Previously identified wild lineages from West Africa, Malaysia, North America, Japan and the Philippines are included in the Wild Misc group<sup>11,74</sup>. The other lineages are named in a similar manner to previous studies on ale-brewing and Mediterranean Oak (MedOak) strains<sup>8,9,71</sup>. Bottom: Principal component projection for PC1 and PC2 (including EN14S01, which groups with Sake/Asian).

and 8 and Supplementary Datasets 2, 7 and 8). These findings suggest that there have been multiple hybridization events and extensive backcrossing with wild lineages of *S. uvarum*, integrating wild diversity into these hybrids and leading to a diverse set of introgressed strains.

**Domesticated *S. cerevisiae* parent lineages.** Of the species contributing to domesticated interspecies hybrids, *S. cerevisiae* has the most extensive datasets, including industrial yeasts<sup>5,8–11</sup>. We recapitulated the previously described domesticated *S. cerevisiae* clades<sup>8,9</sup> and our 81 interspecies hybrids with *S. cerevisiae* contributions fell into three domesticated lineages: Wine, Ale/Beer1 and Beer2 (Fig. 2d, Supplementary Fig. 9 and Supplementary Datasets 2 and 9).

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

Interestingly, the only hybrids we detected in the Ale/Beer1 group were the lager-brewing yeasts (Fig. 2d). The *S. cerevisiae* subgenomes of the Saaz and Froberg lager-brewing lineages formed distinct clades and, although we identified more Froberg strains, Froberg genetic diversity was lower (Supplementary Text). To determine if there was a particular clade of Ale/Beer1 that was the closest known relative to lager-brewing yeasts, we performed a targeted analysis of just the Ale/Beer1 *S. cerevisiae* strains and lager-brewing hybrids, (Supplementary Figs. 10 and 11, Supplementary Datasets 2 and 10 and Supplementary Text). Our concatenated phylogenomic analyses did not strongly support any recognized geographical clade of Ale/Beer1 *S. cerevisiae* strains as the closest outgroup to the lager-brewing yeasts. Our PCA analyses, which make no assumptions about consistent genome-wide signals, suggested several Stout beer, Wheat beer and mosaic strains as sharing the most ancestry with lager-brewing yeasts, rather than any clade affiliated with a geographic style (Supplementary Fig. 9). Overall,



**Fig. 3 | Mitochondrial genome inheritance in interspecies hybrids.** **a**, The bar plots show proportion of 1:1:1 orthologue content for each subgenome for each hybrid grouped by the mtDNA parent, which are labelled across the top. Colours represent different parent species and are the same as in Fig. 1a. **b**, Analysis of concordance between which mtDNA was inherited and which parent contributed the most complete set of orthologous genes. ‘True’ includes hybrids that inherited the most nuclear gene content from the same species as the mtDNA. ‘False’ includes hybrids with mtDNA that did not match the species that contributed the most nuclear gene content. Colours represent the mtDNA parent and shapes represent the largest nuclear genome contributor. The middle of the box plot corresponds to the median, the upper and lower limits are the 75th and 25th percentiles respectively and the whiskers extend to the largest or smallest value no greater than 1.5 $\times$  the differences between the 75th and 25th percentiles. There was a significant correlation between the mtDNA parent and the largest nuclear genomic contributor (logistic regression  $P=3.58 \times 10^{-8}$ , AIC=118.21). Notably, the *S. eubayanus*  $\times$  *S. uvarum* hybrids, which have often undergone many backcrossing events, follow this trend and are both cryotolerant species. **c**, Linear relationship of the number of 1:1:1 orthologues versus the number of nuclear-encoded, mitochondrially localized genes present in the subgenome that matches the mtDNA (linear regression  $P=2.0 \times 10^{-16}$ , AIC=1151.5). The inset shows the mean proportion of mitochondrially localized versus all other nuclear genes present in the subgenome that matches the mitochondrial parent ( $P=0.8612$ , odds ratio=0.9653). Here, the middle of the box plot corresponds to the mean, the upper and lower limits are one standard error from the mean and the whiskers extend to the maximum and minimum values.

our analyses clearly show that lager strains belong to the Ale/Beer1 lineage of *S. cerevisiae* and suggest affinity with a new set of diverse beer yeasts but they do not support any known extant strain as the sole closest relative.

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

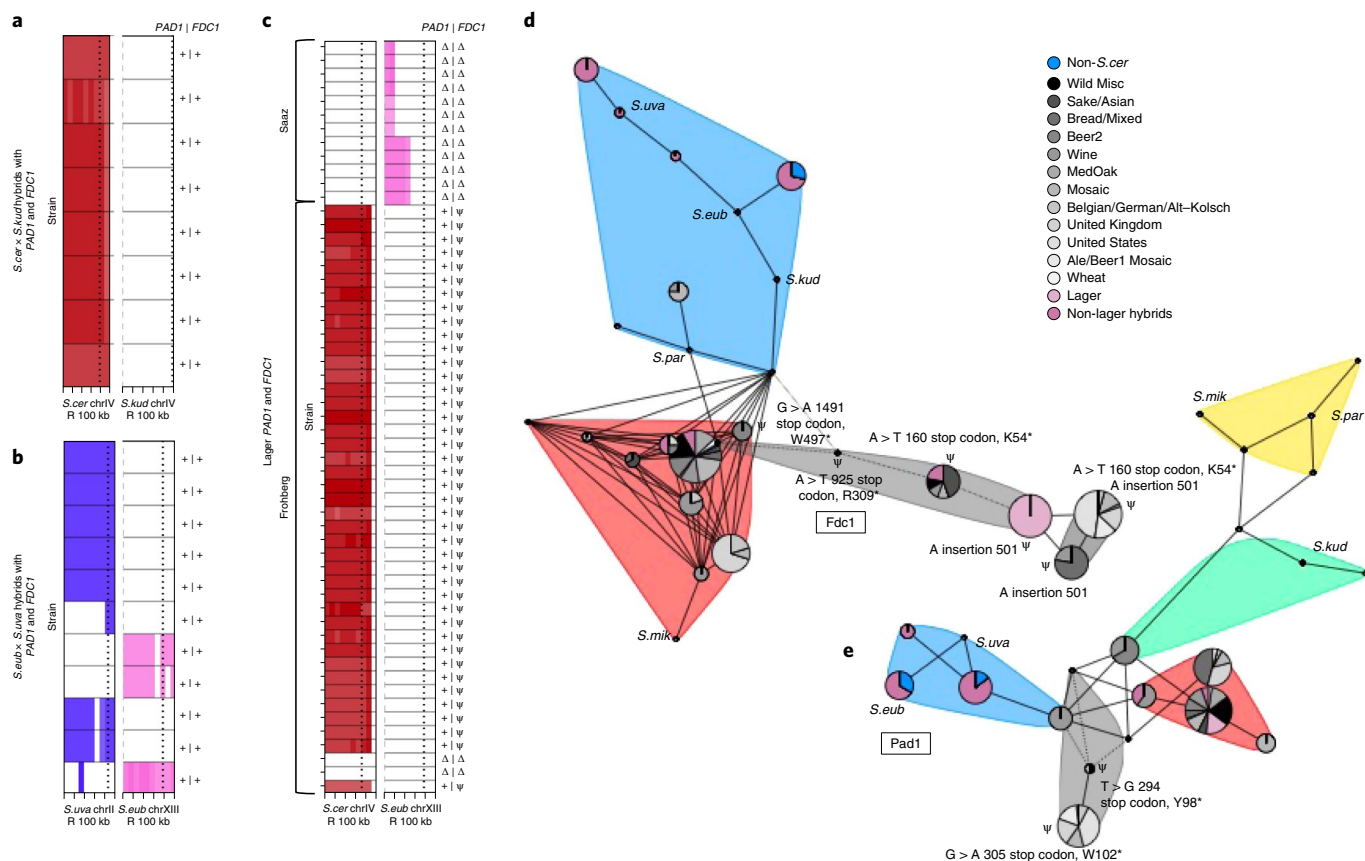
**Mitochondrial genome inheritance.** The classic example of yeast hybrid vigour comes from the cryotolerance of lager-brewing yeasts. *S. eubayanus*, *S. kudriavzevii* and *S. uvarum* are all known to tolerate much colder temperatures<sup>33,34</sup> and recent functional experiments have shown that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

We tested if the parent that donated the mtDNA was also the parent that contributed the most nuclear gene content. We used a logistic regression to determine if the same parent species contributed both the mtDNA and the most complete set of orthologues.

We found that this trend was generally true ( $P=8.0 \times 10^{-6}$ ; Akaike information criterion, AIC=83.75) but there were informative outliers (Fig. 3b). In particular, more than half of the hybrids with *S. kudriavzevii* nuclear contributions inherited the *S. kudriavzevii* mtDNA, despite the fact that the *S. kudriavzevii* nuclear contribution was never in the majority. This discrepancy could be due to a fitness advantage conferred by the *S. kudriavzevii* mtDNA in colder fermentations or it could be due to a fitness advantage conferred by the *S. cerevisiae* or other nuclear genomes<sup>36,37</sup>. Indeed, all outliers in our logistic regression analysis were in the direction of inheriting a cryotolerant parent’s mtDNA. These findings suggest that the inheritance of a cryotolerant mtDNA allowed these hybrids to thrive in colder environments where pure *S. cerevisiae* strains struggle, providing evolutionary and genetic innovation that enabled new fermentation techniques, such as lager brewing.

Hundreds of nuclear-encoded proteins localize to the mitochondria<sup>38</sup>. This interaction can be a source of genetic incompatibilities between the nuclear and mtDNAs, several of which have been characterized in *Saccharomyces* interspecies hybrids<sup>39–41</sup>. Therefore, we tested whether mitochondrially localized, nuclear-encoded genes were retained more often than other genes encoded in the nuclear genome matching the mtDNA parent. We found that mitochondrially localized genes were retained in the same ratio as all other orthologues ( $P=0.8612$ , odds ratio=0.9653) (Supplementary Dataset 11 and Fig. 3c). Although these results suggest that mitochondrial localization is not the main cause of the correlation between nuclear and mtDNA content, some nuance is warranted. First, only a few mitochondrially localized genes have been implicated in mito-nuclear incompatibilities<sup>39–41</sup> and other factors that do



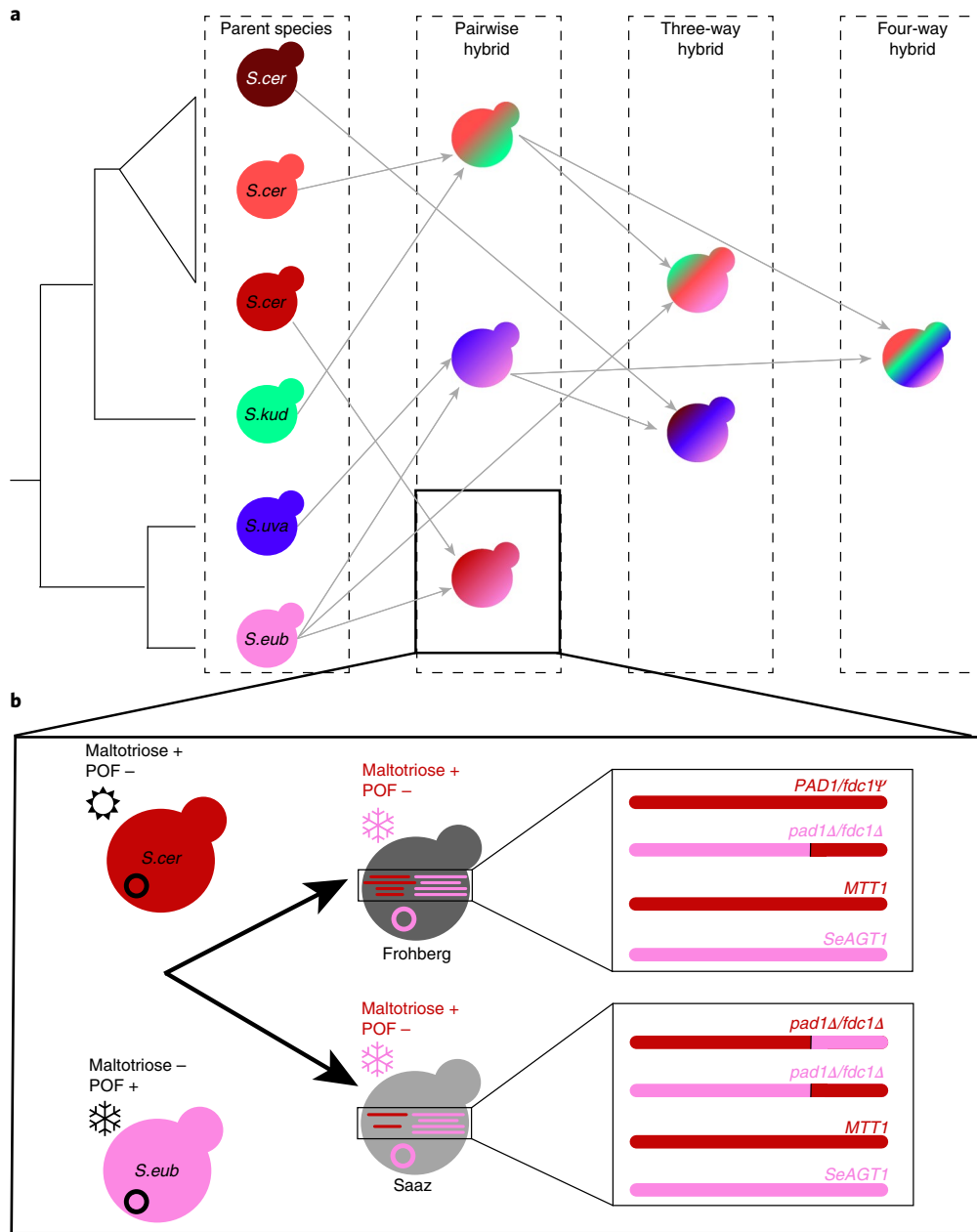


**Fig. 4 | Hybrid inheritance and functionality of genes responsible for 4VG production.** **a–c**, Retention of the regions where the adjacent *PAD1* and *FDC1* genes, which are both required for 4VG production, are located in each parent species, shown as 10-kb windows of ploidy estimates over last 100-kb of the chromosome. Gene locations are represented by black dotted lines. Higher opacity represents higher ploidy. Species colours are that same as in Fig. 1a. *S.par*, *S. paradoxus*; *S.mik*, *S. mikatae*. **a**, *S.cer* × *S.kud* hybrids: all strains inherited versions of both *PAD1* and *FDC1* from *S.cer* that are predicted to be functional, +|+ but they have lost the *S.kud* alleles. **b**, *S.uva* × *S.eub* hybrids: all strains inherited versions of *PAD1* and *FDC1*, from either *S.uva* or *S.eub*, that are predicted to be functional, +|+. **c**, All lager strains have completely lost the region in the *S.eub* genome where these genes reside. Additionally, all Saaz strains have also completely lost the *S.cer* versions of these genes, Δ | Δ. All but two Froberg strains have retained versions of *PAD1* from *S.cer* that are predicted to be functional but inherited *S.cer* alleles of *FDC1* that are predicted to be inactive due to a frameshift mutation, +|Ψ. **d, e**, Haplotype networks were built for the amino acid sequences for *Fdc1* (**d**) and *Pad1* (**e**). Coloured pies correspond to *S.cer* lineages, hybrids or wild species with size representing the number of strains with that haplotype. Non-*S.cer* nodes or groups of nodes are labelled by the species to which they correspond. Coloured clouds correspond to communities: red is mostly *S.cer*, blue is mostly non-*S.cer* (including *S.eub* and *S.uva*), yellow is mostly *S.par* and *S.mik*, green is mostly *S.kud* and grey is mostly loss-of-function alleles. Pseudogenes are marked as Ψ with additional information about the loss-of-function nucleotide and amino-acid changes. Dotted connections represent >100 amino acid differences.

not rely on protein localization could also play a role (for example, metabolite exchange between the mitochondria and cytoplasm). Perhaps more importantly, these hybrids have often lost whole chromosomes or regions containing hundreds of genes at a time through chromosome mis-segregation or mitotic recombination events<sup>15</sup>; this restriction imposed by genetic linkage may prevent fine-scale retention or loss and obscure any signal driven by specific genes. Overall, from this dataset, we conclude that there is a strong correlation between the amount of nuclear and mitochondrial DNA contributed by each parent species but mitochondrially localized genes are not more affected than other genes.

**Pan-genome analyses.** In analysis of the core genome, we found that retention of 1:1:1 orthologues from the contributing parent genomes ranged from as few as 12 genes to almost complete sets of orthologues from all their parents (Supplementary Fig. 12 and Supplementary Dataset 12). On average, hybrids retained 56.2% of orthologues from the parent who contributed the least genomic material.

We performed de novo genome assemblies to analyse the genomic content that was not present in the parent reference genomes (Supplementary Fig. 13). On average, hybrids had 47.7 kilobases (kb) of novel genomic content; the minimum was 2.2 kb and the maximum was 363.3 kb. In addition to novel content that may come from the pan-genomes of other the *Saccharomyces* species, we detected previously characterized content from prior *S. cerevisiae* pan-genome analyses, including horizontally transferred genes (Supplementary Text)<sup>5,12,42</sup>. When we searched this material for *Saccharomyces*-like genes for which we could assign a function, we found an enrichment in genes associated with sugar transport, including the Gene Ontology<sup>43,44</sup> terms: transporter activity (corrected  $P = 4.67 \times 10^{-8}$ ), sugar:proton symporter activity (corrected  $P = 6.04 \times 10^{-8}$ ), cation:sugar symporter activity (corrected  $P = 6.04 \times 10^{-8}$ ) and sugar transmembrane transporter activity (corrected  $P = 6.04 \times 10^{-8}$ ) (Supplementary Dataset 7). The enrichment of sugar transport genes in the novel content of these hybrids and introgressed strains is consistent with strong selection for these activities in industrial fermentation environments.



**Fig. 5 | Summary of hybrids and origin of lager traits. a**, Simplified summary of parents and resulting hybrids. On the left is a cladogram of just the *Saccharomyces* species that have contributed to fermented beverage hybrids. Three distinct lineages of *S.cer* have contributed to hybrids; for the wild parents, Holarctic or European lineages gave rise to the hybrids. Grey lines point from each parent to the resulting hybrid. The order of secondary or tertiary hybridization events was inferred from genome composition. This simplified view does not show when multiple lineages of *S.cer* have contributed to different hybrid types (for example, *S.cer* × *S.kud* hybrids), backcrossing (for example, *S.eub* × *S.uva* hybrids) or minor subtelomeric contributions (for example, small *S.cer* contributions to some *S.eub* × *S.uva* hybrids). **b**, Summary of how lager-brewing yeasts acquired their unique trait profile. The two lager-brewing lineages, Saaz and Frohberg, arose out of hybridizations between domesticated *S.cer* ale strains and wild *S.eub* strains. The *S.cer* strains could use maltotriose (+), did not produce phenolic off-flavours (POF<sup>-</sup>) and preferred warmer temperatures (Sun symbol), while the *S.eub* strains tolerated colder temperatures (snowflake symbol), could not use maltotriose (-) and produced phenolic off-flavours (POF<sup>+</sup>). The two lager-brewing lineages inherited the *S.eub* mitochondrial genome (pink circle), which partly conferred cryotolerance. Both lineages also inherited maltotriose transporter genes from both parents (*MTT1* from *S.cer* and *SeAGT1* from *S.eub*). Finally, both lineages convergently became POF<sup>-</sup> through multiple distinct mechanisms, including preadaptation in the *S. cerevisiae* ale-brewing parent due to a mutated pseudogene (*PAD1|fdc1Y* in red), aneuploidy removing functional *S. eubayanus* genes (*pad1Δ|fdc1Δ* in pink) and translocations in all Saaz strains and some Frohberg strains (*pad1Δ|fdc1Δ* in red).

**Genes for maltotriose use.** We took a more detailed look at the genes for maltotriose use because maltotriose is generally the second most abundant sugar in beer wort or malt extract and *Saccharomyces* strains that use it are relatively rare outside of domesticated ale-brewing strains<sup>45–48</sup>. Our analyses of lager-brewing

yeasts suggest that both *S. cerevisiae* and *S. eubayanus* contributed genes encoding functional maltotriose transporters to the hybrids, including alleles of *S. cerevisiae* *MTT1* and *S. eubayanus* *AGT1* previously shown to be functional<sup>18</sup> (Fig. 5b and Supplementary Text). We also recovered other predicted maltose/maltotriose transporter

homologues in other interspecies hybrids and their parent species, which have yet to be explored functionally (Supplementary Dataset 14). We conclude that the complexity and diversity of maltose transporter genes across *Saccharomyces* species is extensive and may have provided a source of functional diversity to fermentation hybrids.

**Phenolic off-flavour genes.** The introduction of genes from wild strains, especially the mitochondrial genome and *S. eubayanus* *AGT1*, may have been key to cold fermentations but other genes probably negatively affected products. The aromatic 4-vinyl guaiacol (4VG) is perceived as a clove-like, phenolic or smoky flavour and considered an undesirable off-flavour in most beers. Lager beers are known for their crisp flavour profiles that lack appreciable 4VG, while wild strains of *S. eubayanus* and other species produce 4VG<sup>49</sup>. Two genes, *PAD1* and *FDC1*, are essential for the production of 4VG<sup>50</sup>. Studies in ale-brewing yeast show that this trait is under strong domestication selection (Supplementary Text) but the genotypes of *PAD1* and *FDC1* across diverse interspecies hybrids already in use by industry have not been investigated, nor have the evolutionary genetic events leading to these genotypes. In our large hybrid dataset, we analysed both retention and predicted functionality of *PAD1* and *FDC1* alleles from their parent species (Fig. 4).

In both *S. cerevisiae* × *S. kudriavzevii* and *S. eubayanus* × *S. uvarum* hybrids and introgressed strains, we found both *FDC1* and *PAD1* alleles that were predicted to be functional (Supplementary Text). These findings may reflect selection for diverse flavours, which are desirable in niche Trappist-style beers made with *S. cerevisiae* × *S. kudriavzevii*. In contrast, *S. eubayanus* × *S. uvarum* are often viewed as contaminants in industrial brewing environments and production of 4VG could contribute to this perception.

In the lager-brewing hybrids, we found that all strains have lost the ability to produce 4VG but the mechanism of this loss differed between Saaz and Froberg (Supplementary Text). The Froberg lager strains probably inherited a loss-of-function *FDC1* allele from their domesticated *S. cerevisiae* parent and functional *PAD1* and *FDC1* alleles from their *S. eubayanus* parent. These functional wild alleles were then lost through translocations, probably due to break-induced replication. In contrast, the Saaz lineage has completely lost both the *S. cerevisiae* and *S. eubayanus* alleles of these genes through aneuploidy, an evolutionary trajectory facilitated by the fact that these subtelomeric genes reside on different chromosomes in these two species. The end result is that both Saaz and Froberg lagers lack substantial phenolic off-flavours and have a crisp flavour profile. Even though Saaz and Froberg strains evolved this trait through different final mutations that removed functional *S. eubayanus* alleles, the pre-adaptation of the domesticated *S. cerevisiae* parent, which already lacked functional genes, played a critical role by limiting the number of mutations needed. The contrast between Saaz and Froberg strains highlights that there are many potential evolutionary trajectories open to interspecies hybrids to achieve a domestication trait.

## Conclusions

Here, we characterized the genomes of 122 interspecies yeast hybrids and introgressed strains, the largest dataset of its kind to date. These hybrids have complex genomes with contributions from two to four species: *S. cerevisiae*, *S. kudriavzevii*, *S. uvarum* and *S. eubayanus* (Fig. 5a). The hybrids with *S. cerevisiae* contributions all arose out of three domesticated *S. cerevisiae* lineages: the wine lineage and two distinct beer clades. In contrast, all the *S. kudriavzevii*, *S. uvarum* and *S. eubayanus* parents belonged to Holarctic or European wild lineages. Our results show how hybrid vigour also applies to microbes, with the domesticated *S. cerevisiae* parents providing genes and traits preadapted for industrial fermentations and the divergent species of *Saccharomyces* contributing new genes and traits that led to the successes of these hybrids in specific products.

First, the frequent retention of mitochondrial genomes from cryotolerant parents probably conferred a fitness advantage during cold fermentation (Fig. 5b). Second, although the *S. cerevisiae* genome is required for maltotriose use by hybrids, both *S. eubayanus* and *S. cerevisiae* contributed functional maltotriose transporter genes to lager-brewing yeasts. Third, phenolic off-flavour genes have been inactivated or eliminated from lager-brewing yeasts by multiple types of mutations (Fig. 5b), while these genes have been retained in yeasts that ferment products where phenolic off-flavour is prized.

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

## Methods

**Strain selection and sequencing.** The strains newly published here are from wild or beverage isolations, the Agricultural Research Service (ARS) NRRL collection (<https://nrrl.ncaur.usda.gov>) and commercially available sources. Supplementary Dataset 15 contains the full metadata for strains. Whole-genome Illumina paired-end sequencing was done as previously described using either 2 × 100 or 2 × 250 reads<sup>32,51</sup>. These short-read data are available through the NCBI SRA database under the accession number PRJNA522928. Short-read data for published genomes were downloaded from NCBI; Supplementary Dataset 16 contains a full list of accession numbers and citations<sup>8,9,11,16,26,30,32,42,52–72</sup>.

**Hybrid identification.** We used sppIDer<sup>73</sup>, a hybrid detection and analysis pipeline, to identify new hybrids, pure species and reconfirm the species and hybrid identities of published data. For sppIDer, we used a combination reference genome that included all published genomes for all the *Saccharomyces* species<sup>63,72,74,75</sup> (<https://www.yeastgenome.org> and [www.saccharomycesgenetics.org](http://www.saccharomycesgenetics.org)). For *S. kudriavzevii*, we used the genome from the Portuguese strain ZP591. As previously noted<sup>72</sup>, the published *S. uvarum* genome has the labels for chromosomes X and XII swapped, so we manually corrected them. We ran sppIDer with parameters set to identify genomic contributions >1% of the total genome. As sppIDer is reference genome-based, inheritance of regions not in the reference genome was not analysed. Therefore, interspecies hybrids with only minor or subtelomeric introgressions were missed with this method. We also detected some smaller introgressions through the pan-genome analyses (see below).

Hybrid isolation environment was classified on the basis of marketed product type for commercial strains; for published strains or strains from the ARS NRRL collection, we used available metadata supplied by the authors or depositors. Full details of hybrid isolation environment classification can be found in Supplementary Dataset 1. To determine if there was an association between hybrid type and isolation environment, we completed  $\chi^2$  analyses of hybrid by environment and of environment by hybrid with a Bonferroni multiple test correction in R. We limited this test to our most common ( $n > 15$ ) hybrid types (*S. cerevisiae* × *S. eubayanus*, *S. cerevisiae* × *S. kudriavzevii* and *S. eubayanus* × *S. uvarum*) and the most common ( $n > 8$ ) origins (beer, wine and fruit).

**Whole-genome sequence assembly pipeline.** Alignment and single nucleotide polymorphism (SNP) calling were done as described previously<sup>32</sup>. Briefly, short reads were mapped with bwa 'mem' to a concatenated reference genome of just the contributing parents. Reference genomes used for concatenation were the same as used for sppIDer. Samtools 'view' and 'sort' were then used to prepare the mapped reads with a mapping quality greater than 20 for SNP calling. PCR duplicates were removed with picard 'MarkDuplicates' and read groups were set with picard 'AddOrReplaceReadGroups'. SNPs were called with GATK's haplotype caller. Genome coverage per base pair was assessed with bedtools 'genomeCoverageBed'. Strain-specific FASTA files were created by replacing called SNPs in repeat-masked concatenated reference genomes. Variants called as indels were replaced with Ns. Regions of extremely high coverage (the 99.9th percentile of genome-wide coverage) were masked as Ns. Regions that do not exist in hybrids were masked as Ns and regions at low coverage (between 3× and 10×, depending on where the 10th percentile of the distribution of depth of coverage across the concatenated genomes fell) were masked as Ns. The strain-specific FASTAs for hybrid genomes were split into their component subgenomes to be analysed with pure strains.

Genomic completeness was estimated as the percentage of the reference genome with coverage above the low-coverage masking threshold. Ploidy was estimated across the combination genome in 10-kb windows. We used the R package modes (v.0.7.0) to analyse the distribution of depth of coverage and

determine the antimodes, which correspond to a change in ploidy state. Some manual curation was needed for strains with ‘smiley patterns’, a pattern of increased coverage at chromosome ends that has been noted in other depth-of-coverage analyses<sup>8,76</sup> and may be due to chromatin structure<sup>77</sup>. For these strains, we used only the coverages that fell below the 95th percentile to estimate the antimodes and then assigned the distal ends to the largest ploidy estimated. We also visually checked and corrected rare instances when a ‘smiley pattern’ lowered the ploidy estimate for the middle of the chromosome. From this antimode analysis, we were able to assign each 10-kb window a ploidy value. The total DNA base-pair content contributed by each parent could then be estimated as the sum of each ploidy value multiplied by 10,000 and the number of windows with that ploidy value. Correcting this total DNA content per species by the total sum of all contributing species gave us a measure of total genomic content per species. Genomic contribution to a hybrid genome can be viewed as genomic content and genomic completeness. To estimate genomic completeness, we determined what percentage of a total parent subgenome had at least one haploid copy. To estimate genomic content, we took into account both completeness and ploidy across the combination of subgenomes. Full details on hybrid genome contributions can be found in Supplementary Dataset 1. For visualizations, we clustered the strains on the basis of ploidy estimated across the combination genome using Ward’s method in the R package *pvclust* (v.2.0-0; ref. <sup>78</sup>).

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

**1:1:1 orthologues.** To characterize the core genome of these hybrids, we first analysed the retention of 1:1:1 orthologues conserved in all four parent species and determined which parents contributed the least and most coding sequences to each hybrid. We identified genes that are orthologous across all parent genomes based on the annotations in the published gff files for each reference genome, which yielded a list of 3,856 genes. We used the coordinates to determine the coverage for each orthologue. Gene presence was noted if the mean coverage for that orthologue was  $>3\times$ .

**De novo genome assembly and pan-genome analyses.** We assembled the hybrid genomes with the meta-assembler *iwgms*<sup>30</sup> and chose the best assembly based on the largest N50 score. All hybrids, except DBVPG6257, were successfully assembled and are available under GenBank BioProject PRJNA522928.

We mapped the short-read data back to these assembled genomes and used the *sppIDer* output to classify to which parent reference genome each short-read mapped. With this analysis, we determined which reads did not map to a parent reference genome but did assemble de novo into a contig of 1.5 kb or greater. We classified these regions as ‘unmapped’ and used a tBLASTx to search for *S. cerevisiae*-like genes using S288C open reading frames and retaining hits with *e*-value  $<10^{-10}$ . To determine if this set of genes identified in these novel assembled regions were enriched for any functions, we used GO Term Finder (v.0.86; refs. <sup>43,44</sup>). To determine the potential origin of these novel regions, we used a BLASTn search of the NCBI nucleotide database (v.5). The output of this was then parsed for number of hits with an *e*-value  $<10^{-10}$ . To determine the number of hits to different species, we completed  $\chi^2$  analyses with a Bonferroni multiple test correction in R.

**Translocation identification.** To detect shared breakpoints and translocations, we use *LUMPY*<sup>81</sup> with the mapped short-read data. We masked for repetitive regions by excluding regions with coverage above twice the genome-wide mean. Each breakpoint call had to be supported by at least four reads to be included in downstream analyses. We parsed this output for species subgenome, hybrid type and the species pair between which the translocation was detected. We calculated the total number of called breakpoints, breakpoints that were shared in at least two hybrids of the same type and breakpoints that were shared in multiple hybrid types. We compared these different categories with  $\chi^2$  analyses and a Bonferroni multiple test correction in R.

We also identified translocations from the de novo assemblies. For this analysis, we used *sppIDer* results to assign regions of the de novo assemblies to a parent species. Some regions were unmapped with *sppIDer*, as noted above. Additionally, some regions had high coverage from multiple parents in the de novo assembly, where the donor species could not be unambiguously assigned; these regions are probably repetitive and difficult to assemble. Translocations were identified when regions that were  $>2$  kb came from different donor species and were assembled with  $<100$  base pairs (bp) of unmapped or ambiguous data separating them. On average, we identified 17 translocations per strain. From this output, we counted the number of translocations identified in each hybrid type, the

donor species and the pair of species between which the translocations occurred. We compared hybrid type, species pair and individual species with a  $\chi^2$  analyses with a Bonferroni multiple test correction in R.

**Mitochondrial genome analysis pipeline.** We use *mitoSppIDer*<sup>73</sup> to determine the mtDNA parent for the hybrids. This analysis was done in a similar manner to the whole-genome *sppIDer* analysis, except that mtDNAs for each *Saccharomyces* species were used<sup>72,82,83</sup>, except *S. jurei*. GenBank accessions lacking full manuscripts included *S. mikatae* (KX707788) and *S. kudriavzevii* (KX707787).

To determine if the mtDNA parent was associated with retention of the nuclear genes, we performed a logistic regression in R. We used the set of 1:1:1 orthologues to determine which parent contributed the most complete set of orthologous genes. To determine if there was an enrichment for the retention of nuclear-encoded, mitochondrially interacting proteins, we used the set of genes products identified as localized to the mitochondria through the Yeast GFP Fusion Localization Database<sup>88</sup>. When we filtered for genes that were also 1:1:1 orthologues, our final list consisted of 459 genes. To determine if there was a linear relationship between retention of mitochondrially localized genes and all other orthologues, we performed a linear regression and, to determine if there were more mitochondrially localized genes retained compared to all other genes, we used a Fisher’s exact test with a Bonferroni correction. Tests were performed in R.

Since past work has shown that reticulate evolution, introgression and horizontal gene transfers are widespread in *Saccharomyces* mtDNAs<sup>84</sup>, we wanted to explore the inheritance of mitochondrially encoded genes in more depth. Due in part to their high AT content ( $\sim 85\%$ ), mtDNAs are often poorly covered using Illumina sequencing. In particular, intergenic regions and coding sequences with transposable elements (introns, homing endonucleases and GC clusters) can be difficult to assemble. To explore the phylogenetic relationships of these mtDNAs, we used a bait–prey bioinformatic method to pull out the read sequences of coding sequences. We used *HybPiper*<sup>85</sup> to pull out reads from the hybrid Illumina libraries that mapped to those mitochondrial genes using gene sequences from reference strains used in *mitoSppIDer* as baits. These extracted Illumina reads were aligned to the reference genes in *Geneious* (v.6.1.6; ref. <sup>86</sup>) and manually assembled. We successfully covered six mitochondrial genes (*COX2*, *COX3*, *ATP6*, *ATP8*, *ATP9* and 15S ribosomal RNA), which were used to construct the mitochondrial phylogenetic haplotype network. This unique set of unambiguously completed genes was concatenated (4.7 kb) by strain to produce the haplotype for each pure *Saccharomyces* or hybrid strain (Supplementary Fig. 14). Haplotypes and haplotype frequencies for each strain were encoded as a nexus-formatted file for *PopART* v.1.7.2 (ref. <sup>87</sup>). The haplotype network was reconstructed using the TCS method<sup>88</sup>. Strains were assigned to each haplotype using *DnaSP* v.5 (ref. <sup>89</sup>). For some strains, we could not assemble the 15S rRNA gene because of low-coverage data. For these strains, we inferred their haplotype designation based on an analysis where we removed the 15S rRNA gene. This information is not included in Supplementary Fig. 14 but can be found in Supplementary Dataset 17.

**Genes of functional interest analysis pipeline.** To assemble the sequences of genes relevant to brewing, we again used *HybPiper*<sup>85</sup>. To be included for further analyses, the assembled length had to be at least as long as the bait gene and had to have a minimum 10 $\times$  depth of coverage. For the baits, we used either gene sequences from the *S. cerevisiae* strain S288C found on the *Saccharomyces* Genome Database (<https://www.yeastgenome.org>) from the *S. eubayanus* type strain, CBS12357<sup>T</sup> (ref. <sup>72</sup>); or the lager strain W34/70 (ref. <sup>90</sup>). For the *PAD1* analysis in *S. eubayanus*  $\times$  *S. uvarum* hybrids, we used the *PAD1* gene sequence from the *S. uvarum* reference genome, CBS7001 (ref. <sup>63</sup>). To get precise gene locations for *PAD1* and *FDC1*, we used a tBLASTn search of the *S. eubayanus*, *S. kudriavzevii* and *S. uvarum* reference genomes with the *S. cerevisiae* sequences for these genes as the query.

The assembled genes were aligned with *MAFFT* v.7 (ref. <sup>91</sup>), allowing for reverse complementation. The alignments were manually trimmed to the protein-coding sequences. For *PAD1* and *FDC1*, the alignments were conceptually translated to amino acid sequences and haplotype networks were built with a modified minimum-spanning network and visualized with *iGraph*<sup>92</sup> in R. The haplotype networks were split into communities as previously described<sup>93</sup>.

Pairwise distances between sequences were calculated using the trimmed *MAFFT* nucleotide sequence alignments and the *P*-distance method as implemented in *MEGA-X*<sup>94</sup> with the following parameters: substitutions to include Transitions+Transversions, assuming uniform rates among sites and using pairwise deletion of gaps. The percentage identity of hits to the bait sequence was organized by species and hybrid status was recorded in Supplementary Dataset 14, along with the origin of the bait gene and tallies of sequences whose translations were visually identified as being incomplete or containing premature stop codons.

**Phylogenomic and population structure analyses.** We masked regions with no coverage as Ns, which is interpreted as missing data by most tools; therefore, for downstream whole-genome analyses, we only included subgenomes that were  $>50\%$  complete (major contributions). To include the minor contribution hybrids in the non-*S. cerevisiae* analyses, we used reduced genomes that were concatenations of the regions of the genome that existed in at least one minor



hybrid (Supplementary Dataset 18). This procedure allowed us include strains with minor introgressions and only use regions of the genome that had been contributed by the minor parent. To balance some of our analyses for Saaz and Froberg lager strains, we used a random subset of Froberg strains to match the number of Saaz strains.

Phylogenomic trees were built with RAxML v.8.1 (ref. <sup>95</sup>) using SNPs from the whole genome for the major analyses or the reduced genome for the minor analyses. Trees were visualized with iTOL<sup>96</sup>. The PCA analyses were done with the adegenet package in R<sup>97</sup> and visualized with ggPlot2 (ref. <sup>98</sup>). Estimates of adjusted  $\pi$  ( $\pi \times 100$ ) were calculated with the PopGenome package in R<sup>99</sup>.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

References and accession numbers for the published data used can be found in Supplementary Dataset 16. Short-read data (newly published here) are available through the NCBI SRA database under the BioProject accession number PRJNA522928. Assembled genomes published here are available under GenBank BioProject PRJNA522928.

## Code availability

Custom R and Python scripts used for this publication can be found on GitHub (<https://github.com/qlangdon/hybrid-ferment-invent>).

Received: 20 February 2019; Accepted: 2 September 2019;

Published online: 21 October 2019

## References

- Hornsey, I. S. *Alcohol and Its Role in the Evolution of Human Society* (RSC Publishing, 2012).
- Fay, J. C. & Benavides, J. A. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, e5 (2005).
- Liti, G., Peruffo, A., James, S. A., Roberts, I. N. & Louis, E. J. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. *Yeast* **22**, 177–192 (2005).
- Gallone, B. et al. Origins, evolution, domestication and diversity of *Saccharomyces* beer yeasts. *Curr. Opin. Biotechnol.* **49**, 148–155 (2018).
- Legras, J. L. et al. Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* **35**, 1712–1727 (2018).
- Rodríguez, M. E. et al. *Saccharomyces uvarum* is responsible for the traditional fermentation of apple chicha in Patagonia. *FEMS Yeast Res.* **17**, fow109 (2017).
- Barbosa, R. et al. Multiple rounds of artificial selection promote microbe secondary domestication—the case of cachaça yeasts. *Genome Biol. Evol.* **10**, 1939–1955 (2018).
- Gallone, B. et al. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166**, 1397–1410 (2016).
- Gonçalves, M. et al. Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* **26**, 1–12 (2016).
- Duan, S. F. et al. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat. Commun.* **9**, 2690 (2018).
- Peter, J. et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).
- Marsit, S. & Dequin, S. Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. *FEMS Yeast Res.* **15**, 1–12 (2015).
- Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P. & Sampaio, J. P. Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites. *Mol. Ecol.* **26**, 2167–2182 (2017).
- Hittinger, C. T., Steele, J. L. & Ryder, D. S. Diverse yeasts for diverse fermented beverages and foods. *Curr. Opin. Biotechnol.* **49**, 199–206 (2018).
- Gibson, B. & Liti, G. *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. *Yeast* **32**, 17–27 (2015).
- Libkind, D. et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl Acad. Sci. USA* **108**, 14539–14544 (2011).
- Baker, E. P. et al. Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci. Adv.* **5**, eaav1869 (2019).
- Baker, E. P. & Hittinger, C. T. Evolution of a novel chimeric maltotriose transporter in *Saccharomyces eubayanus* from parent proteins unable to perform this function. *PLoS Genet.* **15**, e1007786 (2019).
- Hebly, M. et al. *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res.* **15**, 1–14 (2015).
- Gibson, B. R., Storgårds, E., Krogerus, K. & Vidgren, V. Comparative physiology and fermentation performance of Saaz and Froberg lager yeast strains and the parental species *Saccharomyces eubayanus*. *Yeast* **30**, 255–266 (2013).
- Gorter de Vries, A. R. et al. Laboratory evolution of a *Saccharomyces cerevisiae* × *S. eubayanus* hybrid under simulated lager-brewing conditions. *Front. Genet.* **10**, 242 (2019).
- Monerawela, C. & Bond, U. Brewing up a storm: the genomes of lager yeasts and how they evolved. *Biotechnol. Adv.* **35**, 512–519 (2017).
- Peris, D., Pérez-Torrado, R., Hittinger, C. T., Barrio, E. & Querol, A. On the origins and industrial applications of *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* hybrids. *Yeast* **35**, 51–69 (2018).
- Nguyen, H. V. & Boekhout, T. Characterization of *Saccharomyces uvarum* (Beijerinck, 1898) and related hybrids: assessment of molecular markers that predict the parent and hybrid genomes and a proposal to name yeast hybrids. *FEMS Yeast Res.* **17**, 1–19 (2017).
- Nguyen, H. V., Legras, J. L., Neuvéglise, C. & Gaillardin, C. Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380 T. *PLoS ONE* **6**, e25821 (2011).
- Almeida, P. et al. A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* **5**, 4044 (2014).
- Dunn, B. & Sherlock, G. Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res.* **18**, 1610–1623 (2008).
- Hittinger, C. T. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* **29**, 309–317 (2013).
- Boynton, P. J. & Greig, D. The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast* **31**, 449–462 (2014).
- Hittinger, C. T. et al. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* **464**, 54–58 (2010).
- Sampaio, J. P. & Gonçalves, P. Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl. Environ. Microbiol.* **74**, 2144–2152 (2008).
- Peris, D. et al. Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. *PLoS Genet.* **12**, e1006155 (2016).
- Salvadó, Z., Arroyo-López, F. N., Barrio, E., Querol, A. & Guillamón, J. M. Quantifying the individual effects of ethanol and temperature on the fitness advantage of *Saccharomyces cerevisiae*. *Food Microbiol.* **28**, 1155–1161 (2011).
- Gonçalves, P., Valério, E., Correia, C., de Almeida, J. M. G. C. F. & Sampaio, J. P. Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS ONE* **6**, e20739 (2011).
- Li, X. C., Peris, D., Hittinger, C. T., Sia, E. A. & Fay, J. C. Mitochondria-encoded genes contribute to evolution of heat and cold tolerance in yeast. *Sci. Adv.* **5**, eaav1848 (2019).
- Ortiz-Tovar, G., Pérez-Torrado, R., Adam, A. C., Barrio, E. & Querol, A. A comparison of the performance of natural hybrids *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* at low temperatures reveals the crucial role of their *S. kudriavzevii* genomic contribution. *Int. J. Food Microbiol.* **274**, 12–19 (2018).
- Tronchoni, J., Medina, V., Guillamón, J. M., Querol, A. & Pérez-Torrado, R. Transcriptomics of cryophilic *Saccharomyces kudriavzevii* reveals the key role of gene translation efficiency in cold stress adaptations. *BMC Genomics* **15**, 1–10 (2014).
- Huh, K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Chou, J. Y., Hung, Y. S., Lin, K. H., Lee, H. Y. & Leu, J. Y. Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol.* **8**, e1000432 (2010).
- Lee, H. Y. et al. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**, 1065–1073 (2008).
- Hou, J. & Schacherer, J. Negative epistasis: a route to intraspecific reproductive isolation in yeast? *Curr. Genet.* **62**, 25–29 (2016).
- Novo, M. et al. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl Acad. Sci. USA* **106**, 16333–16338 (2009).
- Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
- Consortium, T. G. O. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
- Han, E.-K., Cotty, F., Sottas, C., Jiang, H. & Michels, C. A. Characterization of AGT1 encoding a general alpha-glucoside transporter from *Saccharomyces*. *Mol. Microbiol.* **17**, 1093–1107 (1995).
- Salema-Oom, M., Pinto, V. V., Gonçalves, P. & Spencer-Martins, I. Maltotriose utilization by industrial *Saccharomyces* strains: characterization of a new member of the alpha-glucoside transporter family. *Appl. Environ. Microbiol.* **71**, 5044–5049 (2005).
- Horák, J. Regulations of sugar transporters: insights from yeast. *Curr. Genet.* **59**, 1–31 (2013).

48. Dietvorst, J., Londesborough, J. & Steensma, H. Y. Maltotriose utilization in lager yeast strains: MTTI encodes a maltotriose transporter. *Yeast* **22**, 775–788 (2005).
49. Diderich, J. A., Weening, S. M., van den Broek, M., Pronk, J. T. & Daran, J.-M. G. Selection of Pof-*Saccharomyces eubayanus* variants for the construction of *S. cerevisiae* × *S. eubayanus* hybrids with reduced 4-vinyl guaiacol formation. *Front. Microbiol.* **9**, 1640 (2018).
50. Mukai, N., Masaki, K., Fujii, T., Kawamukai, M. & Iefuji, H. PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* **109**, 564–569 (2010).
51. Shen, X.-X. et al. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**, 1533–1545 (2018).
52. Bing, J., Han, P.-J., Liu, W.-Q., Wang, Q.-M. & Bai, F.-Y. Evidence for a Far East Asian origin of lager beer yeast. *Curr. Biol.* **24**, R380–R381 (2014).
53. Borneman, A. R., Forgan, A. H., Pretorius, I. S. & Chambers, P. J. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Res.* **8**, 1185–1195 (2008).
54. Borneman, A. R. et al. Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.* **7**, e1001287 (2011).
55. Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A. & Schmidt, S. A. Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3* **6**, 957–971 (2016).
56. Dunn, B., Richter, C., Kvitkic, D. J., Pugh, T. & Sherlock, G. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res.* **22**, 908–924 (2012).
57. Gayevskiy, V. & Goddard, M. R. *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ. Microbiol.* **18**, 1137–1147 (2016).
58. Gayevskiy, V., Lee, S. & Goddard, M. R. European derived *Saccharomyces cerevisiae* colonisation of New Zealand vineyards aided by humans. *FEMS Yeast Res.* **16**, 1–12 (2016).
59. Hewitt, S. K., Donaldson, I. J., Lovell, S. C. & Delneri, D. Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS ONE* **9**, e92203 (2014).
60. Hose, J. et al. Dosage compensation can buffer copy number variation in wild yeast. *eLife* **4**, 1–28 (2015).
61. Krogerus, K., Preiss, R. & Gibson, B. A unique *Saccharomyces cerevisiae* × *Saccharomyces uvarum* hybrid isolated from Norwegian farmhouse beer: characterization and reconstruction. *Front. Microbiol.* **9**, 1–15 (2018).
62. Okuno, M. et al. Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res.* **1**, 1–14 (2016).
63. Scannell, D. R. et al. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3* **1**, 11–25 (2011).
64. Skelly, D. A. et al. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res.* **23**, 1496–1504 (2013).
65. Strobe, P. K. et al. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* **125**, 762–774 (2015).
66. van den Broek, M. et al. Chromosomal copy number variation in *Saccharomyces pastorianus* is evidence for extensive genome dynamics in industrial lager brewing strains. *Appl. Environ. Microbiol.* **81**, 6253–6267 (2015).
67. Yue, J. X. et al. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* **49**, 913–924 (2017).
68. Zheng, D. Q. et al. Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJ5329. *BMC Genomics* **13**, 479 (2012).
69. Bergström, A. et al. A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* **31**, 872–888 (2014).
70. Akao, T. et al. Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. *DNA Res.* **18**, 423–434 (2011).
71. Almeida, P. et al. A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol. Ecol.* **24**, 5412–5427 (2015).
72. Baker, E. et al. The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* **32**, 2818–2831 (2015).
73. Langdon, Q. K., Peris, D., Kyle, B. & Hittinger, C. T. sppIDer: a species identification tool to investigate hybrid genomes with high-throughput sequencing. *Mol. Biol. Evol.* **35**, 2835–2849 (2018).
74. Liti, G. et al. Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
75. Liti, G. et al. High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC Genomics* **14**, 69 (2013).
76. Peris, D. et al. Biotechnology for biofuels hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. *Biotechnol. Biofuels* **10**, 1–19 (2017).
77. Teytelman, L. et al. Impact of chromatin structures on DNA processing for genomic analyses. *PLoS ONE* **4**, e6700 (2009).
78. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
79. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
80. Zhou, X. et al. In silico whole genome sequencer and analyzer (iWGS): a computational pipeline to guide the design and analysis of de novo genome sequencing studies. *G3* **6**, 3655–3662 (2016).
81. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
82. Foury, F., Roganti, T., Lecrenier, N. & Purnelle, B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* **440**, 325–331 (1998).
83. Sulo, P. et al. The evolutionary history of *Saccharomyces* species inferred from completed mitochondrial genomes and revision in the 'yeast mitochondrial genetic code'. *DNA Res.* **24**, 571–583 (2017).
84. Peris, D. et al. Molecular phylogenetics and evolution mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species. *Mol. Phylogenet. Evol.* **108**, 49–60 (2017).
85. Johnson, M. G. et al. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* <https://doi.org/10.3732/apps.1600016> (2016).
86. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
87. Leigh, J. W. & Bryant, D. POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
88. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659 (2000).
89. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
90. Walther, A., Hesselbart, A. & Wendland, J. Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3* **4**, 783–793 (2014).
91. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
92. Cserdi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **1695**, 1–9 (2006).
93. Opulente, D. A. et al. Factors driving metabolic diversity in the budding yeast subphylum. *BMC Biol.* **16**, 1–15 (2018).
94. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
95. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
96. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
97. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
98. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2009).
99. Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).

## Acknowledgements

We thank K.J. Verstrepen for coordinating publication with their study; A.B. Hulfachor and M. Bontrager for preparing a subset of Illumina libraries; the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing Illumina sequencing facilities and services; M.-A. Lachance, A. Kinart, D.T. Doering, R. Thiel and D. Carey for strains; and M. Langdon, A.B. Hulfachor and K. Sylvester for collecting fermentation samples and/or isolating strains. This material is based upon work supported by the National Science Foundation (grant nos. DEB-1253634 to C.T.H. and DGE-1256259 (Graduate Research Fellowship to Q.K.L.)), the USDA National Institute of Food and Agriculture Hatch Project (nos. 1003258 and 1020204 to C.T.H.) and in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science nos. DE-SC0018409 and DE-FC02-07ER64494). Q.K.L. was also supported by the Predoctoral Training Program in Genetics, funded by the National Institutes of Health (grant no. 5T32GM007133). D.P. is a Marie Skłodowska-Curie fellow of the European Union's Horizon 2020 research and innovation programme (grant no. 747775). E.P.B. was supported by a Louis and Elsa Thomsen Wisconsin Distinguished Graduate Fellowship. U.B. is funded by Horizon 2020 MSCA-ITN grant no. 764364. D.L. was supported by CONICET (grant no. PIP 392), FONCYT (grant no. PICT 3677)

and Universidad Nacional del Comahue (grant no. B199). C.T.H. is a Pew Scholar in the Biomedical Sciences, Vilas Faculty Early Career Investigator and H.I. Romnes Faculty Fellow, supported by the Pew Charitable Trusts, Vilas Trust Estate and Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, respectively.

### Author contributions

Q.K.L. performed most analyses with assistance from D.A.O. D.P. and Q.K.L. performed mitochondrial genome analyses and drafted text. E.P.B. and Q.K.L. analysed genes of functional interest and drafted text. Q.K.L., E.P.B. and D.A.O. sequenced genomes. H.-V.N., U.B., P.G. and J.P.S. contributed key strains to study design. Q.K.L., D.P., E.P.B., D.L. and C.T.H. designed the study. Q.K.L. and C.T.H. wrote the manuscript with editorial input from all coauthors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41559-019-0998-8>.

**Correspondence and requests for materials** should be addressed to C.T.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

N/A

Data analysis

For hybrid classification, we used sppIDer, an open source pipeline to classify hybrids based on short read data. For read mapping and SNP calling, the following software was used: bwa (v0.7.12), samtools (v1.7), picard (v1.98), bedtools (v2.27.0), and GATK (v3). To determine the ploidy based on depth of coverage, we used the R package modes (v0.7.0); and to cluster the strains by ploidy, we used pvclust (v. 2.0-0). Many visualizations were done with the R package ggplot2 (2.2.1). For phylogeny analyses, we used RAxML (v8.1) and visualized with iTol (v4.3). For principal component analyses, we used the R package adegenet (v2.0.1). To determine pi values, we used the R package PopGenome (v2.6.1). To assemble genes of interest, we used HybPiper (v1.2). For the mitochondrial work, we used Geneious (v. 6.1.6), PopART (v1.7.2), and DnaSP (v5). To determine the location of a gene in a reference genome, we used tBLASTn. To align functionally relevant genes, we used MAFFT (v.7), to compare pairwise differences of genes we used MEGA-X. To make haplotype networks for PAD1 and FDC1, we used a custom R script and visualized with iGraph (1.2.3). We assembled the hybrid genomes with the meta-assembler iWGS (v1.1). To determine if this set of genes identified in these novel assembled regions were enriched for any functions, we used GO Term Finder (Version 0.86). To detect shared breakpoints and translocations, we use LUMPY. The versions of R used were 3.3.0, 3.5.0, and 3.5.2; the version of python used was 2.6.6, and we used BioPython (1.66). Custom R and python scripts have been deposited on gitHub.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.



## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

References and accession numbers for the published data used can be found in Supplementary File 16. Short-read data newly published here is available through the NCBI SRA database under the BioProject accession number PRJNA522928. Assembled genomes published here are available under GenBank BioProject PRJNA522928. Custom R and Python scripts used for this publication can be found on GitHub (<https://github.com/qlangdon/hybrid-ferment-invent>).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Population and phylogneomic study of Saccharomcyes yeast hybrids from fermentation and wild sources.
Research sample	Whole genome data for pure Saccharomcyes cerevisiae, Saccharomcyes kudriavzevii, Saccharomcyes uvarum, and Saccharomcyes eubayanus and hybrids of these species. Yeast strains came from commercial sources, strain collections, and from isolations of fermented products. We analyzed both newly sequenced whole genome data and existing whole genome data for these species of hybrids of these species. With this dataset we recapitulated the know population structure of these species.
Sampling strategy	For new strains, we targeted commercial strains that are used for fermentations associated with known hybrids, including lagers and Belgian-style beers. For strains collections, we acquired any strains which were classified just to the genus level or with a hybrid species name "Saccharomcyes bayanus" and "Saccharomcyes pastorianus". We included all strains we acquired for this study; the sample size represents all available whole genome data for hybrids of these species.
Data collection	N/A
Timing and spatial scale	N/A
Data exclusions	In a subset of our analyses, we included fewer Frohberg lager strains because the low diversity in this group biased the PCA approaches. However, we also present the analyses with the full data set.
Reproducibility	N/A
Randomization	N/A
Blinding	N/A
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involvement in the study                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data               |

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involvement in the study                        |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |