Yeast adapts to diverse ecological niches driven by genomics and metabolic reprogramming

(article starts on next page)

Yeast adapts to diverse ecological niches driven by genomics and metabolic reprogramming

Haoyu Wang (王浩宇)[a,b,c] [ID], Jens Nielsen[d,e] [ID], Yongjin J. Zhou (周雍进)[b,f,g,1], and Hongzhong Lu (鲁洪中)[a,1] [ID]

Affiliations are included on p. 10.

The famous model organism *Saccharomyces cerevisiae* is widely present in a variety of natural and human-associated habitats. Despite extensive studies of this organism, the metabolic mechanisms driving its adaptation to varying niches remain elusive. We here gathered genomic resources from 1,807 *S. cerevisiae* strains and assembled them into a high-quality pangenome, facilitating the comprehensive characterization of genetic diversity across isolates. Utilizing the pangenome, 1,807 strain-specific genome-scale metabolic models (ssGEMs) were generated, which performed well in quantitative predictions of cellular phenotypes, thus helping to examine the metabolic disparities among all *S. cerevisiae* strains. Integrative analyses of fluxomics and transcriptomics with ssGEMs showcased ubiquitous transcriptional regulation of metabolic flux in specific pathways (i.e., amino acid synthesis) at a population level. Additionally, the gene/reaction inactivation analysis through the ssGEMs refined by transcriptomics showed that *S. cerevisiae* strains from various ecological niches had undergone reductive evolution at both the genomic and metabolic network levels when compared to wild isolates. Finally, the compiled analysis of the pangenome, transcriptome, and metabolic fluxome revealed remarkable metabolic differences among *S. cerevisiae* strains originating from distinct oxygen-limited niches, including human gut and cheese environments, and identified convergent metabolic evolution, such as downregulation of oxidative phosphorylation pathways. Together, these results illustrate how yeast adapts to distinct niches modulated by genomic and metabolic reprogramming, and provide computational resources for translating yeast genotype to fitness in future studies.

Saccharomyces cerevisiae | pangenome | strain-specific genome-scale metabolic model | environmental adaptation | metabolic reprogramming

As one of the most famous eukaryotic model organisms, the budding yeast *Saccharomyces cerevisiae* has been widely used in areas of synthetic and systems biology (1–5). The first whole genome sequence of *S. cerevisiae* was released in 1996, followed by numerous molecular studies using its laboratory mutants (3, 6–8). Well adapted to laboratory environments, *S. cerevisiae* is also ubiquitous in both natural and human-associated niches. The wide distribution around the earth has shaped the genetic and phenotypic diversity of distinct *S. cerevisiae* isolates (9). Elucidating how *S. cerevisiae* strains adapt to varied environments could provide insights into the metabolic rewiring of eukaryotic organisms driven by natural and artificial selection (10).

The advancement of high-throughput sequencing technology has made it possible to profile the genomic evolution of *S. cerevisiae* on a larger scale. Over the past decade, amounts of genome sequencing have been performed to characterize the genetic diversity of *S. cerevisiae* strains sampled from a wide range of habitats, such as wine, plants, and human-guts (11–17). Notably, Peter et al. reported the whole genome sequencing for 1,011 *S. cerevisiae* strains worldwide, successfully delineating the global evolutionary portrait of the species (9). More recently, the same group performed pantranscriptome analysis of ~1,000 *S. cerevisiae* strains and illustrated the gene expression patterns across the major evolutionary clades (18). Together, these omics datasets provide valuable resources to study the evolution of *S. cerevisiae* at system and population levels.

The pangenome, defined as the collection of all genes encompassed by a group of individuals from a certain species, has become a powerful tool to characterize the genomic and metabolic diversity of studied species (19). Currently, there are two representative pangenomes for *S. cerevisiae*, here named Sce-pan1011 and Sce-pan1364, which were built for 1,011 and 1,364 isolates, respectively. By contrast, the Sce-pan1011 contains a total of 7,796 ORFs, while the Sce-pan1364 has only 7,078 ORFs (9, 20). With the accumulation of newly sequenced *S. cerevisiae* strains, it is thus critical to rebuild the pangenome of *S. cerevisiae* considering all sequenced isolates, as well as refining its

Significance

This study establishes a framework for understanding how genetic diversity shapes metabolic adaptation across *Saccharomyces cerevisiae* strains. By integrating pangenomics, transcriptomics, and fluxomics, we find that niche specialization leads to streamlined genomes and metabolism, gene expression exerts potential transcriptional control over some core biochemical pathways on a population level, and strains in various oxygen-limited environments independently evolve similar metabolic reprogramming. The newly reconstructed pangenome and high-quality 1,807 ssGEMs enhance the predictions of strain fitness and serve as valuable computational resources for the scientific community. These resources support future research on yeast evolution, ecology, and metabolic engineering, facilitating the engineering of yeasts with enhanced traits for industrial biotechnology and the understanding of the evolutionary principles occurred in other nonmodel yeast species.

architecture to solve the existing conflict between the aforementioned two versions of pangenomes.

Genome-scale metabolic models (GEMs) provide a mathematical framework representing the entire metabolism of an organism (21). It could serve as a tool to mechanistically link genotype to phenotype. It has been effectively applied to study the metabolic rewiring and hybridization during the adaptive evolution of the *Saccharomyces* genus (22, 23). Since the publication of the first GEM in 2003, *S. cerevisiae* GEMs have been continuously improved through iterative updates by the community (24, 25). During GEMs reconstruction, the genomic information of *S. cerevisiae* S288c was intensively leveraged, while omitting the gene gain and loss existing in other isolates. Therefore, it is challenging to directly apply the current *S. cerevisiae* GEMs-Yeast8 to simulate the metabolic diversity of other isolates. In this regard, Lu et al. have built pan-GEMs for 1,011 *S. cerevisiae* strains and, for the first time, developed strain-specific GEMs for all the studied strains, illustrating metabolic conservation and variation at single-strain resolution (26). However, the limitation of omics datasets, i.e., transcriptomics and phenomics, to some extent, hinders the applications of GEMs in reflecting the in vivo metabolic diversity of those strains (26).

To systematically elucidate the molecular mechanisms underpinning the adaptation of *S. cerevisiae* strains across a variety of environments, we reconstructed a new version of pangenome for *S. cerevisiae* utilizing the high-quality genome sequences from 1,807 isolates, which helps to comprehensively describe genetic diversity within the species. Subsequently, inspired by the detailed functional annotation of the newly built pangenome, 1,807 strain-specific genome-scale metabolic models (ssGEMs) were established to quantitatively assess metabolic differences between strains. We further integrated large-scale transcriptomics data with ssGEMs to investigate the transcriptional regulation of the metabolic network at the population level. Finally, a multidimensional integration analysis of the pangenome, transcriptome, and fluxome was performed to characterize the genetic and metabolic features of *S. cerevisiae* strains that originated from anaerobic environments, including the human gut and dairy niches. Overall, with advanced genomic and modeling analysis, our work shows how the metabolism of *S. cerevisiae* was dynamically reprogrammed to adapt to diverse ecological niches around the world.

Results

Pangenome of 1,807 *S. cerevisiae* Strains. To comprehensively investigate the genetic diversity of *S. cerevisiae*, we compiled 1,913 *S. cerevisiae* genomes from NCBI and published datasets (9, 11, 13, 14, 16, 17, 20). Following unified genome annotation and quality control, 1,807 high-quality genomes were retained to construct the pangenome (*SI Appendix*, Fig. S1). We also refined the sampled environments for each strain used in the subsequent analysis. Our collection, curation, and refinement of each isolate's genome laid a solid foundation for the reconstruction of the *S. cerevisiae* pangenome from scratch. Consequently, a total of 7,514 distinct ORFs were identified throughout the 1,807 genomes to construct the new version of the *S. cerevisiae* pangenome, which was named Sce-pan1807 (Fig. 1*A* and *SI Appendix*, Fig. S1). Compared to the previous study (9), this analysis incorporated a broader range of ecological niches and geological contexts (Fig. 1*B* and *SI Appendix*, Fig. S1 *C* and *D*). While previous pangenome
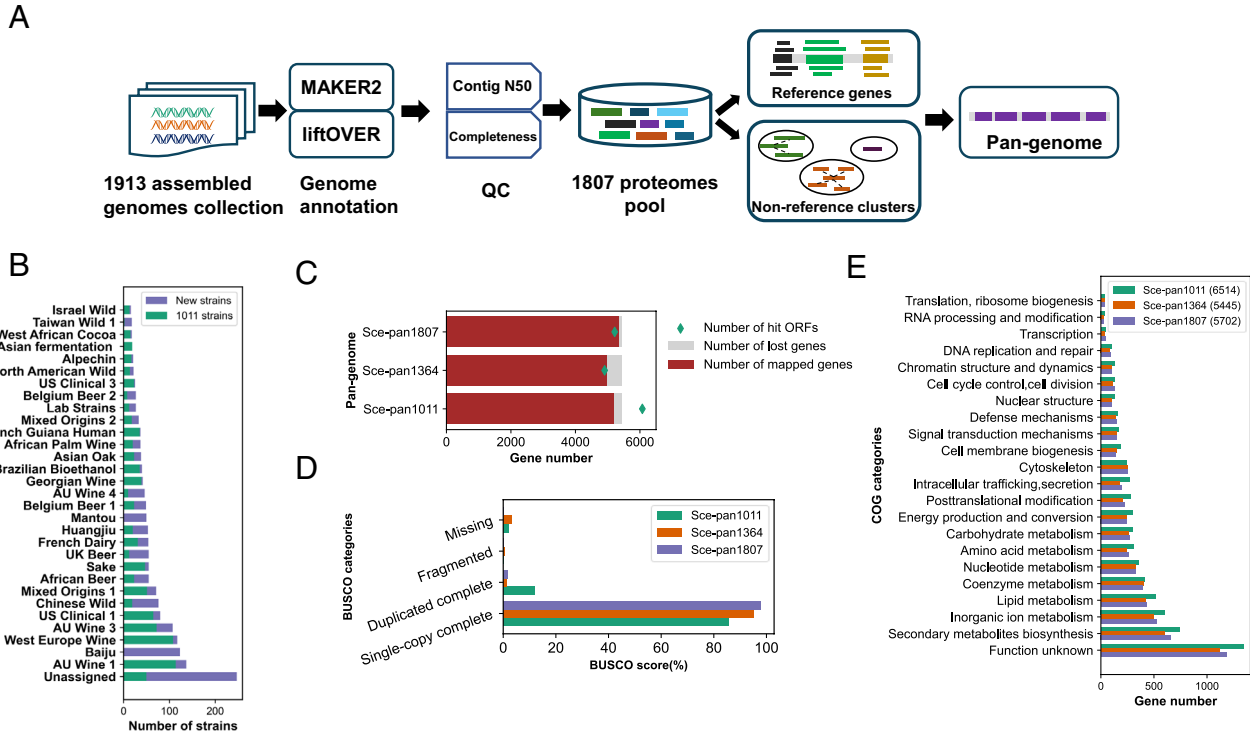
Fig. 1. Construction and quality analysis of the *S. cerevisiae* pangenome. Schematic overview of the workflow employed for constructing the *S. cerevisiae* pangenome (*A*). Distribution of strain number across different ecological origins, comparing isolates from the Sce-pan1011 reference dataset (9) with newly incorporated strains (*B*). Comparative analysis in gene representation between the newly built pangenome (Sce-pan1807) and the previously two published ones (Sce-pan1011, Sce-pan1364). During the mapping analysis, the genome of *S. cerevisiae* CEN.PK was set as the reference genome. "Number of hit ORFs" indicates the number of pan genes that have been detected in the reference genome. "Number of lost genes" indicates the number of the reference strain's genes that have not been detected in the pangenome. "Number of mapped genes" indicates the number of the reference strain's genes that have been detected in pangenome (*C*). Assessment of pangenome completeness using the BUSCO method and the saccharomycetes_odb10 database (*D*). Functional categorization of pangenomes using COG (*E*). The number in brackets in the legend indicates the number of the functionally annotated genes in the pangenomes.

Table 1. Comparison of all pangenomes used in this study

pangenome	Strains/ species	Total ORFs	Core ORFs	Accessory ORFs	Reference
Sce-pan1807	1,807	7,514	4,766	2,748	This study
Sce-pan1011	1,101	7,796	4,942	2,854	Peter et al. (9)
Sce-pan1364	1,364	7,078	5,293	1,785	Li et al. (20)
Yeast-pan	8	9,232	2,974	6,258	This study

analysis revealed that diploid strains generally exhibit growth advantages in natural environments (9), we found that among the newly added strains, certain specialized clades (such as Baijiu, Mantou) with high ploidy (ploidy > 2n) demonstrated natural selection advantages comparable to diploid strains (*SI Appendix,* Fig. S2). Additionally, we performed phylogenetic analysis on the unique genes identified in the newly added strains and found a potential horizontal gene transfer (HGT) from *Oomycota* species in China (*SI Appendix,* Fig. S2).

At the first glance, the size of Sce-pan1807 (7,972 ORFs) is comparable with two previously pangenomes—Sce-pan1011 (7,796 ORFs) and Sce-pan1364 (7,078 ORFs) (9, 20) (Table 1). To further evaluate the quality of Sce-pan1807, we compared its performance against these prior pangenomes in representing genomic information. In this aspect, genomes from two well-studied strains—*S. cerevisiae* CEN.PK and S288c were employed to map against each pangenome's representative sequences by sequence alignment (Fig. 1*C* and *SI Appendix,* Fig. S1*E*). All 3 pangenomes could cover most genes in CEN.PK (~98% for Sce-pan1807, ~95% for Sce-pan1011 and ~91% for Sce-pan1364); however, each pangenome still exhibited some degree of gene missing during the above blast analysis. Particularly, the prior pangenome Sce-pan1364, containing 7,078 genes, was accompanied by the highest number of missing genes. Further, 249 genes from CEN.PK could not be mapped onto the Sce-pan1364, indicating that 7,078 ORFs in the Sce-pan1364 are insufficient to represent all genes in the species. By comparison, for Sce-pan1011, the hit pangene number (6,075) is larger than the total gene number of CEN.PK (5,451), implying that Sce-pan1011 may contain some redundant genes (Fig. 1*C*). The same analysis for S288c showed consistent results (*SI Appendix,* Fig. S1*E*). Furthermore, an all-vs.-all BLAST analysis confirmed the presence of redundancy in Sce-pan1011 (*SI Appendix,* Fig. S3). The quality of the three pangenomes was further assessed through the genome completeness evaluation using BUSCO (27). All 3 pangenomes exhibited nearly no fragmented genes. However, the pangenome Sce-pan1364 has the highest missing score (1.9% for Sce-pan1011, 3.2% for Sce-pan1364 vs. 0.3% for Sce-pan1807), while Sce-pan1011 was accompanied by the highest duplicated score (12.1% for Sce-pan1011, 1.3% for Sce-pan1364 vs. 1.6% for Sce-pan1807) (Fig. 1*D*). Collectively, these analyses indicate that Sce-pan1807 performs well in reflecting the genomic information from different *S. cerevisiae* strains.

To illustrate the metabolic functions encoded in the pangenome, we performed a functional annotation for all 3 pangenomes using the eggNOG tool (28), as well as a functional enrichment analysis based on the annotation from Clusters of Orthologous Groups (COGs). We found that the functional distribution of COG categories throughout the 3 pangenomes displays a similar tendency (Fig. 1*E*), albeit there were some slight deviations across all COG categories, which may be due to the differences in the coverage and quality scores of the 3 pangenomes.

Accurate Definition of Core Genome. The core genome consists of genes shared across most strains of a species. To accurately define the core genome of *S. cerevisiae*, we performed a thorough analysis of the core genome size considering the frequency of ORFs occurring in different strains (Fig. 2*A*). We found that as the threshold for defining core genes increased from 95 to 100% strain coverage, the core genome size decreased from 5,496 to 939. Notably, when the threshold exceeded 99%, the core genome size experienced a precipitous decline, which likely due to errors or incompleteness during the sequencing and assembly processes, which is a common issue in many population genomics studies (29–31). Based on this sensitivity analysis, we selected 99% as the threshold to define the core genome in this study, yielding 4,766 core genes and 2,748 accessory genes within our *S. cerevisiae* pangenome. By comparison, the core genome size is smaller in our work than that in the previous pangenome assemblies (Dataset S1). The slightly larger core genome size in Sce-pan1011 (4,942) is primarily due to the smaller number of strains, as well as the existing gene redundancy discussed above (Fig. 1 *C* and *D*), while the significantly larger core genome size in Sce-pan1364 (5,293) can be explained by its more relaxed threshold (95%) used to define core genes (20).

Next, the metabolic functions of core genomes from three *S. cerevisiae* pangenome studies were compared thoroughly (Fig. 2*B*). For each core genome, the gene functions were classified based on the COG categories. Overall, the 3 core genomes exhibit highly similar functional distributions, highlighting the functional conservation encoded by the core genome of *S. cerevisiae*. Notably, a portion of the core genes in all 3 core genomes belong to the group of unknown functions (920 for Sce-pan1011, 945 for Sce-pan1364, and 950 for this study), highlighting substantial gaps in our understanding of the *S. cerevisiae* genome.

Utilizing the same approach, we assembled a pangenome for eight distinct *Saccharomyces* species, including *S. cerevisiae* (Fig. 2*C* and *SI Appendix,* Table S1). There are 9,232 ORFs in the pangenomes of the 8 yeast species, comprising 2,974 core genes and 6,258 accessory genes. When compared to the pangenome of *S. cerevisiae*, the larger pangenome size (9,232 vs. 7,514) and the smaller core genome size (2,974 vs. 4,766) for those 8 yeast species showcased the high quality of the pan- and core-genome constructed in this study.

Evolution of the Core Genome within and across Yeast Species. Through sequence alignment between the core genomes of *S. cerevisiae* and 8 yeast species, the genes in *S. cerevisiae* core genome can be further categorized into yeast core genes (common in 8 yeast species) and *S. cerevisiae*-specific core genes (only found in a portion of 8 yeast species, Fig. 2*D*). The average gene copy number differed relatively slightly among *S. cerevisiae* strains (mean gene copy number: 1.05 for the core genome vs. 1.09 for the accessory genome, *P* value = 2.06×10^{-14}, *SI Appendix,* Fig. S4*A*). However, the gene copy number of core genes varied more than that of the accessory genome (mean coefficient of variation: 0.17 for core genome vs. 0.13 for accessory genome, *P* value = 2.10×10^{-28}, Fig. 2*D*). In line with these findings, the comparison between yeast core and *S. cerevisiae* specific core genes revealed a similar pattern, with yeast core genes exhibiting a marginally higher level of variation in gene copy number compared to *S. cerevisiae* specific core genes (mean coefficient of variation: 0.17 for core genome vs. 0.16 for accessory genome, *P* value = 8.94×10^{-4}, Fig. 2*D*). These findings reveal subtle differences in gene copy number between the core and accessory genomes, as well as between the yeast core and *S. cerevisiae*-specific core genes.
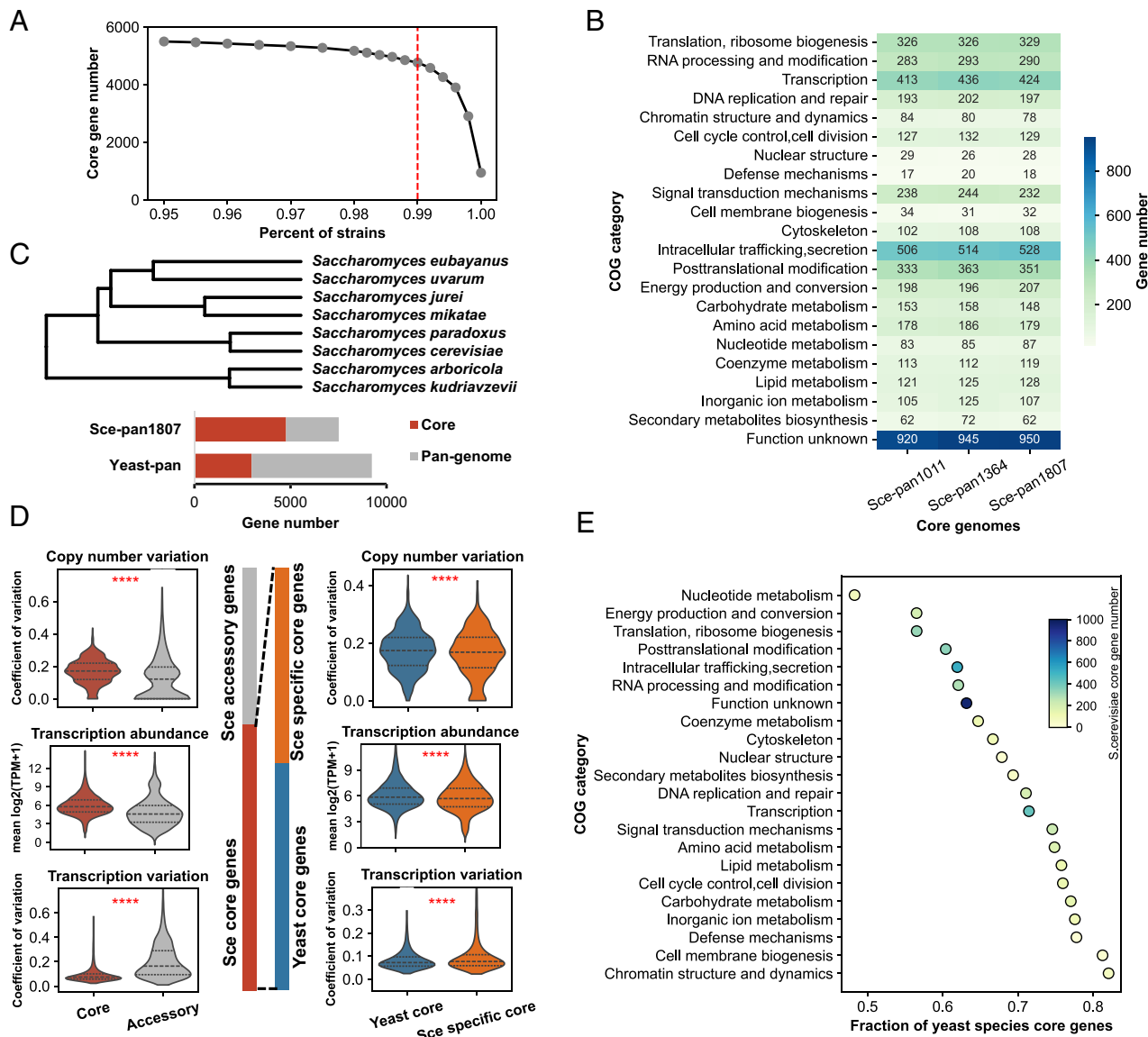
Fig. 2. Core genome analysis of 1,807 *S. cerevisiae* strains. Sensitivity analysis for core genome definition across *S. cerevisiae* isolates (*A*). Functional comparison of core genomes derived from three distinct pangenomes (*B*). A phylogenetic relationship among 7 *Saccharomyces* species (32), as well as the comparison of the pangenomes between yeast species and *S. cerevisiae* (*C*). Comparative analysis in gene copy number, gene expression/variation between *S. cerevisiae* core and accessory genes, as well as between yeast core and nonyeast core genes. Here, the core genes from *S. cerevisiae* were further divided into yeast core and nonyeast core genes, the latter only existing in *S. cerevisiae*, according to the newly defined core genomes from 8 yeast species (*D*). Conservation analysis of yeast core genes across different functional categories. The X-axis represents the proportion of yeast species' core genes in all *S. cerevisiae* core genes for each COG category (*E*). The statistical analysis for two group comparison was based on a two-sided Welch's *t* test. Sce is an abbreviation of *S. cerevisiae*. ****P value < 0.0001.

Subsequently, we compared the functional distribution of yeast core genes and *S. cerevisiae*-specific core genes within the *S. cerevisiae* core genome. For each COG, the fraction of yeast core genes per *S. cerevisiae* core genes was calculated, respectively. The higher the fraction of yeast core genes, the greater the conservation in the evolution of metabolism from specific COGs. The top 5 most conserved COGs are chromatin structure and dynamic, inorganic ion metabolism, carbohydrate metabolism, cell membrane biogenesis, and amino acid biogenesis (Fig. 2*E*), with 4 of them involved in enzyme-driven cellular metabolism. In contrast, some metabolism-related categories, such as nucleotide metabolism, RNA processing and modification, and energy production and conversion, show lower conservation, indicating the loss of *S. cerevisiae* core genes in other yeast species. Collectively, those differences in metabolism conservation during the evolution of

the yeast core genome may contribute to phenotypic diversity across yeast species (32–34).

To investigate gene expression differences, we analyzed large-scale transcriptome data of 969 *S. cerevisiae* strains from a recent study (18), most of which were included in our pangenome analysis. The comparative analysis of gene expression profiles between *S. cerevisiae* core and accessory genomes revealed that core genes are more abundant but have less variance than accessory genes (*P* value = 2.14×10^{-10}, Fig. 2*D*), which is consistent with a previous study (18). A parallel comparison between yeast core and *S. cerevisiae* specific core genes displayed a subtle but consistent trend (mean relative gene expression abundance: 6.07 for yeast core genes vs. 5.81 for *S. cerevisiae* specific core genes, *P* value = 1.06×10^{-7}; mean coefficient of variation: 0.086 for yeast core genes vs. 0.099 for *S. cerevisiae* specific core genes, *P* value = 1.70×10^{-10}, Fig. 2*D*).

Overall, our findings show that *S. cerevisiae* core genes are prone to being highly and stably expressed across strains.

Pangenome Infers Strain-Specific-GEMs. To further explore the metabolic diversity in *S. cerevisiae* isolates, the GEM for each strain was reconstructed based on Sce-pan1807 (Fig. 3*A*). Initially, the pan-GEMs of *S. cerevisiae* (pan-GEMs-1011) from our previous work was updated to reflect the corrected pangenome and recent metabolic network updates of *S. cerevisiae* (*SI Appendix*, Fig. S5*A*) (26, 35). The updated pan-GEMs encompassed 4,176 reactions, 1,259 genes, and 2,887 metabolites (Fig. 3*B*), which was named pan-GEMs-1807. Then, using the pan-GEMs-1807 and gene presence matrix, we reconstructed 1,807 strain-specific GEMs (ssGEMs) via an automated pipeline. The reaction sizes among these models varied from 3,794 to 4,025, with the number of genes ranging from 1,053 to 1,186. Furthermore, strains from different niches exhibited distinct distributions in metabolic network size (Fig. 3*C*). For instance, strains in specific clades such as French Dairy and African Beer displayed significantly fewer metabolic genes and reactions compared to strains from other clades, although the two clades were not similar in metabolic network composition. A positive correlation was observed between metabolic gene number and reaction number (Pearson's correlation coefficient = 0.64, *P* value = 2.8×10^{-208}, *SI Appendix*, Fig. S5*B*). According to our predictive analysis, 85% of the ssGEMs could simulate the strain growth under glucose minimal media conditions. However, the theoretical maximum biomass yields did not differ significantly between strains from different habitats (*SI Appendix*, Fig. S5*D*). The existence of 15% nonviable strains may be related to auxotroph caused by long-term adaptation or domestication. Then, 87% of the nonviable ssGEMs could successfully predict the growth after the gap-filling step and the numbers of gap-filled reactions are fewer than seven (*SI Appendix*, Fig. S5 *C* and *E*). Interestingly, we found that the auxotroph predicted by the model was mainly due to the loss of crucial metabolic genes and the corresponding reactions in pathways important for precursor biosynthesis (*SI Appendix*, Table S2). For example, the reaction r_1838, catalyzed by homocitrate synthase (HCS), which is essential for lysine biosynthesis (36, 37), has been found to be absent in 97 ssGEMs. Note that the incomplete genome sequencing may contribute to these losses, so future physiological studies are needed to validate the predicted auxotrophies (Table 2).

To further assess the quality of the 1807 ssGEMs, the utilization of 65 carbon sources by two strains CEN.PK and Ethanol Red were predicted, for which experimental results were already available (16). The predictive accuracy was 0.64 for CEN.PK and 0.70 for Ethanol Red, which was comparable to that of the well-curated model -Yeast9 (35) (Fig. 3*D*), thus lending confidence to our model prediction capability.
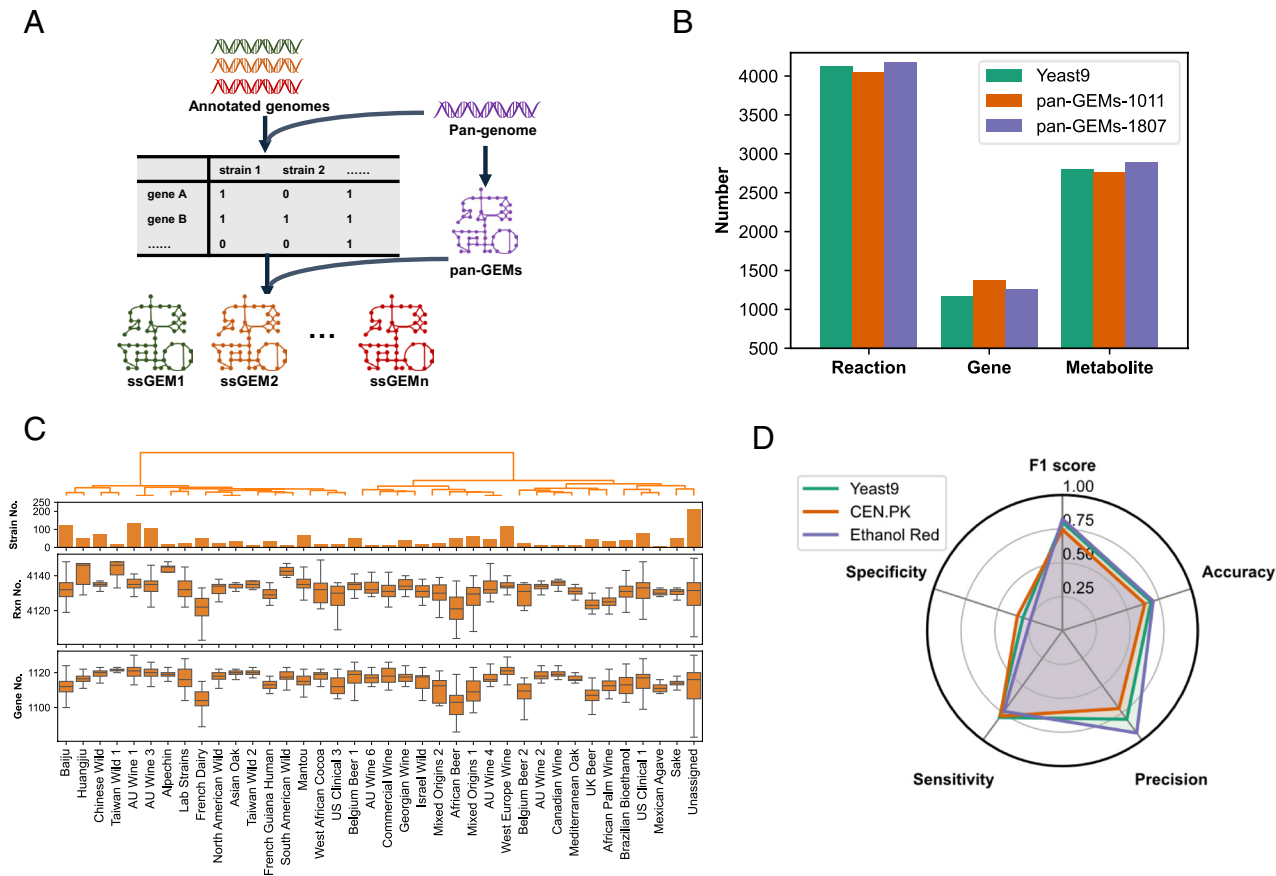
Fig. 3. Construction and evaluation of 1,807 ssGEMs. Workflow for the reconstruction of ssGEMs (*A*). Numbers of reactions, genes, and metabolites in pan-GEMs built in this work (*B*). Distribution of metabolic reactions and gene content across 1,807 ssGEMs, categorized by ecological niche (*C*). Evaluation of the prediction capability of ssGEMs through in silico utilization analysis of 61 different carbon sources by three different *S. cerevisiae* strains—S288c, CEN.PK, and Ethanol red. Here, the newly built ssGEMs for CEN.PK and Ethanol red were used in the comparative analysis, and Yeast9 (35), the consesus GEMs for S288c, was used as the reference (*D*). The experiment dataset used for graph *D* is from a previous study (16).

Table 2. Classification of *S. cerevisiae* clades by ecological source

Ecological source	Included clade IDs
Wild	Chinese Wild, South American Wild, Taiwan Wild 2, Asian Oak
Human	French Guiana human
Dairy	French dairy
Bioethanol	Brazilian bioethanol

Model-Based Analysis Explores the Evolution of Metabolic Network at the Population Level. GEMs are widely used in omics integrative analysis (38, 39). To explore how the genome-wide gene expression variations shape the metabolic rewiring at a population level, the transcriptome data from 969 *S. cerevisiae* isolates under the identical culture condition in synthetic complete medium with glucose as the sole carbon source (18) was used to reconstruct condition-specific GEMs base on the above ssGEMs.

Using a well-established algorithm called GIMME (40), we constructed 907 transcriptome-pruned ssGEMs by removing reactions with low gene expression levels (Fig. 4A). Compared to wild *S. cerevisiae* strains, a pattern of reductive evolution was observed in numerous human-associated clades, such as African Palm Wine and French Guiana Human. This reduction is characterized by a smaller active metabolic network, potentially attributable to the more stable ecological niches where these strains inhabit relative to the fluctuating environments of wild strains (41, 42). However, this trend was less pronounced at the gene level than at the reaction level, with some clades possessing a higher number of metabolic-related genes. This discrepancy may arise from gene copy number variations and the presence of isoenzymes (Fig. 4B). To validate the methodological robustness, we employed an alternative algorithm, tINIT (43), the output of which displayed a similar tendency (*SI Appendix,* Fig. S6). The observed slight discrepancies between the outputs from the above two algorithms can be attributed to the difference in principles and parameter setting during the condition-specific GEMs reconstruction (44). Examining the conservation of core reactions owned by the wild strains in other clades, some reactions were frequently inactivated along with the low gene expression levels revealed by condition-specific GEMs reconstruction, which may result in variation in the cellular metabolism across strains (Fig. 4C and *SI Appendix,* Fig. S7A). For example, the frequently inactivated reaction r_0730, involving the mitochondrial formylation of Met-tRNAiMet catalyzed by mitochondrial methionyl-tRNA synthetase (MetRS), has been shown to have an incremental effect on mitochondrial translation (45, 46). Further experimental validation would be necessary to understand the biological implications of these reactions classified as inactivation in this study. Pathway enrichment analysis of the 40 most commonly inactivated reactions during model refining reveals that the majority of those reactions are related to cell permeability and sensitivity [steroid biosynthesis (47), adjusted *P* value = 4.61×10^{-8}; glycerophospholipid metabolism (48), adjusted *P* value = 0.01; inositol phosphate metabolism (49), adjusted *P* value = 0.004], cofactor metabolism [folate metabolism (50), adjusted *P* value = 0.02; thiamine metabolism (51), adjusted *P* value = 0.02], and protein synthesis [threonylcarbamoyladenosine (t6A) metabolism, adjusted *P* value = 0.02, which is involved in a universally conserved tRNA modification (52)] (Fig. 4D).

Furthermore, our findings revealed that the pathway-level reaction inactivation varied by clade, reflecting unique evolutionary adaptations to specific niches (Fig. 4E). For example, inositol

phosphate metabolism, which plays an important role in membrane synthesis (49) and stress responses (53), experienced significant reactions inactivation in domesticated clades such as French Dairy and AU Wine 1, which may be related to the adaptation to high osmotic stress conditions during domestication. As another example, the steroid biosynthesis, oxygen-dependent in *S. cerevisiae* (54), is consistently inactive in strains from French Guiana human clade (55), which mainly grow under low-oxygen conditions. The unique inactivation of specific pathways in these clades offers insights into the distinct metabolic adaptations of *S. cerevisiae* to diverse environments. However, a more profound understanding of the underlying mechanisms necessitates further investigation by incorporating additional quantitative phenotypic and omics datasets.

Using transcriptome-pruned ssGEMs, we analyzed the correlation between metabolic flux and gene expression levels in 969 *S. cerevisiae* isolates (18). Based on the measured relative growth data (18), we fixed the specific growth rates in the model accordingly and calculated the flux distributions by minimizing glucose uptake via flux balance analysis (FBA). After that, the correlation analysis between fluxes and gene expression yielded coefficients ranging from –0.47 to 0.54 (Fig. 4F). The observed weak overall correlation between metabolic fluxes and transcription may be attributed to the fact that metabolic flux regulation occurs at multiscale levels, including but not limited to transcriptional control (56). The presence of negative correlations in some reactions could be attributed to local flux coordination (57). Reactions were ranked based on their correlation coefficients. The top 10 reactions, which are predominantly regulated at the transcriptional level, primarily participate in specific subpathways, i.e., amino acid metabolism and central carbon metabolism (*SI Appendix,* Fig. S7C). Meanwhile, pathway enrichment analysis revealed that subpathways in amino acid metabolism (e.g., cysteine and methionine metabolism, adjusted *P* value = 0.04; glycine, serine, and threonine metabolism, adjusted *P* value = 0.004; phenylalanine, tyrosine, and tryptophan biosynthesis, adjusted *P* value = 0.02) and central carbon metabolism (such as the TCA cycle, adjusted *P* value = 0.02) have stronger correlations between flux and transcription (Fig. 4G). Our findings are consistent with previous studies showing that the TCA cycle and synthesis of amino acid fluxes are predominantly regulated by transcription (58–60). Collectively, on a population scale, our integrative analysis reveals a weak correlation at the genome-wide level between fluxomics and transcriptomics, implying that changes in gene expression do not always reflect changes in fluxes (56, 58).

Multidimensional Analysis Captures Potential Adaptative Mechanism of *S. cerevisiae* Strains Sampled from Oxygen-Limited Conditions. Finally, a thorough analysis that encompassed transcriptome, pangenome, and ssGEMs was performed to understand how *S. cerevisiae* adapts to specific ecological niches. We focused on the strains belonging to the dairy, bioethanol, and human clades, which typically encounter oxygen limitation in their biological habitats (55) and have distinct metabolic networks compared to the wild strains (Fig. 4B). We first examined whether single nucleotide polymorphisms (SNPs) could explain phenotypic variation across these clades. Although SNPs distinguished the clades genetically (Fig. 5A), they accounted for only 27% of the variation in growth rate ($R^2 = 0.27$, *SI Appendix,* Fig. S8). Instead, these growth disparities may be better explained by transcriptome data, copy number variations (CNVs), and gene presence/absence (*SI Appendix,* Fig. S8). Thus, the evolutionary mechanisms underlying the phenotypic variance of those strains were further examined at both the gene and metabolic network levels.
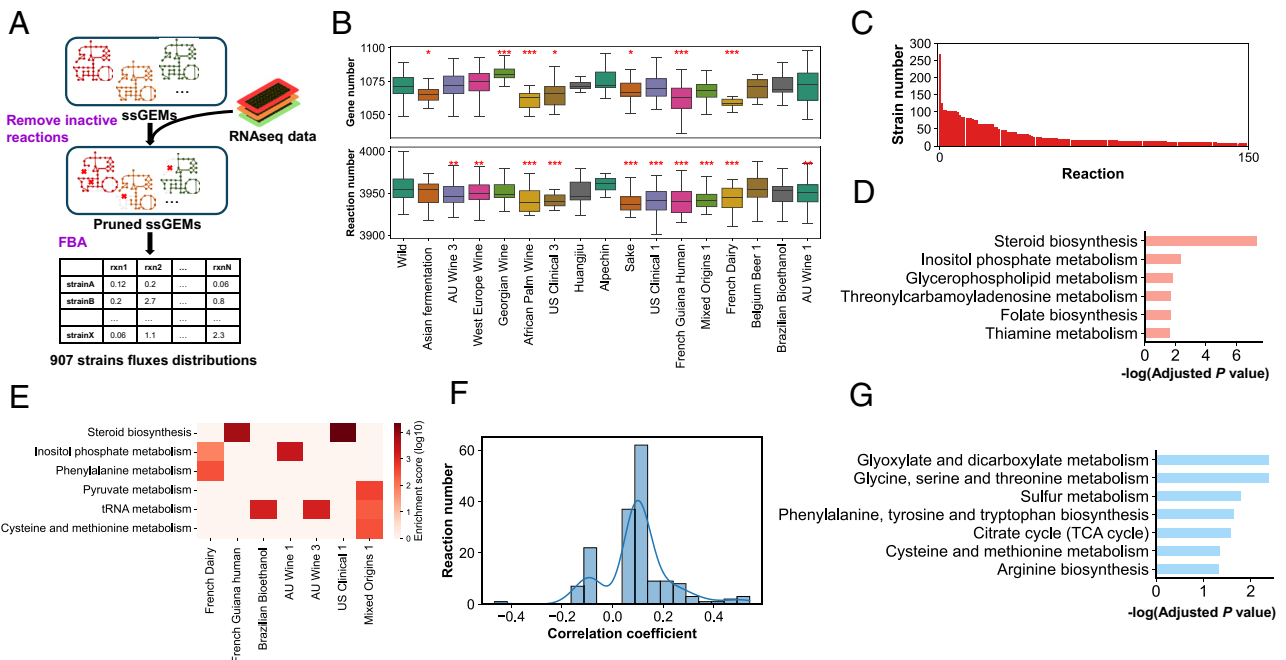
Fig. 4. Integration analysis of ssGEMs and transcriptomics helps to explore the evolutionary mechanisms of metabolic networks. Workflow for the construction and simulation of transcriptome-pruned ssGEMs (*A*). Comparative analysis of metabolic network sizes across distinct evolutionary clades of *S. cerevisiae*, based on previously established SNP-derived phylogenetic classifications (9) (*B*). The statistical analysis of wild core reactions which are lost in other strains' ssGEMs (*C*). Pathway enrichment analysis involving the top 40 wild core reactions which are lost in other strains' ssGEMs (*D*). Pathway enrichment analysis of wild core reactions which are lost in other strains' ssGEMs across different clades (*E*). Correlation analysis between metabolic fluxes and associated gene expression levels. Spearman's rank correlation coefficients were computed, and *P* values were subsequently adjusted for multiple comparisons using the Benjamini–Hochberg (FDR) procedure. Reactions exhibiting an FDR-adjusted *P*-value < 0.05 are presented. (*F*). Pathway enrichment analysis for reactions showing high correlation (*G*). The statistical analysis for two group comparison was based on a two-sided Welch's *t* test. *P value < 0.05; **P value < 0.01; ***P value < 0.001.

First, we extracted the pangenome for human, dairy, bioethanol, and wild strains, from Sce-pan1807 (Fig. 5*B*). To mitigate the impact of strain number disparities on the pangenome structure, we randomly selected 25 strains from each category for subsequent analysis. The human, dairy, and bioethanol populations have a relatively smaller pangenome size compared to the wild population, albeit with a larger number of core genes. The human clade exhibited the smallest pangenome size and the largest core genome size, potentially attributable to the stressful growth condition within the human gut (61). This suggests that the reductive genome evolution is widespread in strains under specific ecological environments, consistent with a recent study (62). Moreover, the variations in the core genomes among different strain clades highlight the genetic diversity arising from adaptive evolution (*SI Appendix*, Fig. S8*B*). PCA showed that, unlike the core genome content (Fig. 5*C* and *SI Appendix,* Fig. S8 *C* and *D*), both the accessory genome and the pangenome content can be employed to effectively classify strains from clades of human, dairy, bioethanol, and wild based on gene presence/absence or gene copy number. Next, a random forest algorithm was used to extract pivotal genetic features for classifying strains based on their accessory genome (*SI Appendix*, Fig. S8*E*). As a result, each clade displayed unique genetic features in the accessory genome in terms of gene presence/absence and copy number (Fig. 5*D* and *SI Appendix,* Fig. S8*F*). Although some of the top feature genes have unknown functions, annotations of all feature genes suggest that they are mainly related to stress response and cell membrane/cell wall synthesis (*SI Appendix*, Fig. S9 *A* and *B*) (63). Additionally, the feature genes extracted based on gene presence and copy number data are different (*SI Appendix,* Fig. S9 *A* and *B*), which hints that the variations in gene content and gene copy number may have been modulated in response to environmental stress.

Next, we investigated the metabolic flux rewiring mechanisms underlying *S. cerevisiae* strains from human, dairy, and bioethanol clades through ssGEMs simulation. To accurately capture the trends in metabolic flux variations across various strains, we integrated both relative growth data and transcriptomics into ssGEMs for simulation (64), during which the relative growth rate was just set as a soft constraint (*SI Appendix*). The predicted growth rates are consistent with the experimental data (Pearson's correlation coefficient = 0.85, *P* value = 3.68×10^{-28}, *SI Appendix,* Fig. S10*A*). Additionally, our simulation successfully captured the phenotype of bioethanol strains, which show higher ethanol yield than wild strains (65). Interestingly, the human clade also exhibits a higher ethanol secretion compared to the wild strains, potentially due to its adaptive evolution in an oxygen-limited human gut environment (*SI Appendix,* Fig. S10*B*) (55). Overall, the reliability of our pipeline in condition-specific flux simulation is demonstrated by the precise predictions for growth and ethanol secretion of *S. cerevisiae* strains from different clades.

We investigated whether the metabolic variations among the selected strains from the above 4 different clades could be captured by flux analysis. We found that, unlike the bioethanol clade, the human and dairy clades could be clearly distinguished from the wild strain through PCA only based on flux distributions (Fig. 5*E* and *SI Appendix,* Fig. S10*C*). However, the differentiation of strains within the four clades based on metabolic fluxes did not achieve the same resolution as clustering based on gene presence or gene copy numbers (Fig. 5*E* and *SI Appendix,* Fig. S10*C*), which is likely due to the fact that all RNA-seq datasets used in our analysis were collected under the same growth condition (18). This also hints that the metabolic fluxes, compared to the genomic variation, may be more conserved in maintaining cellular fitness. Subsequently, with the strains from the wild clade as the reference,
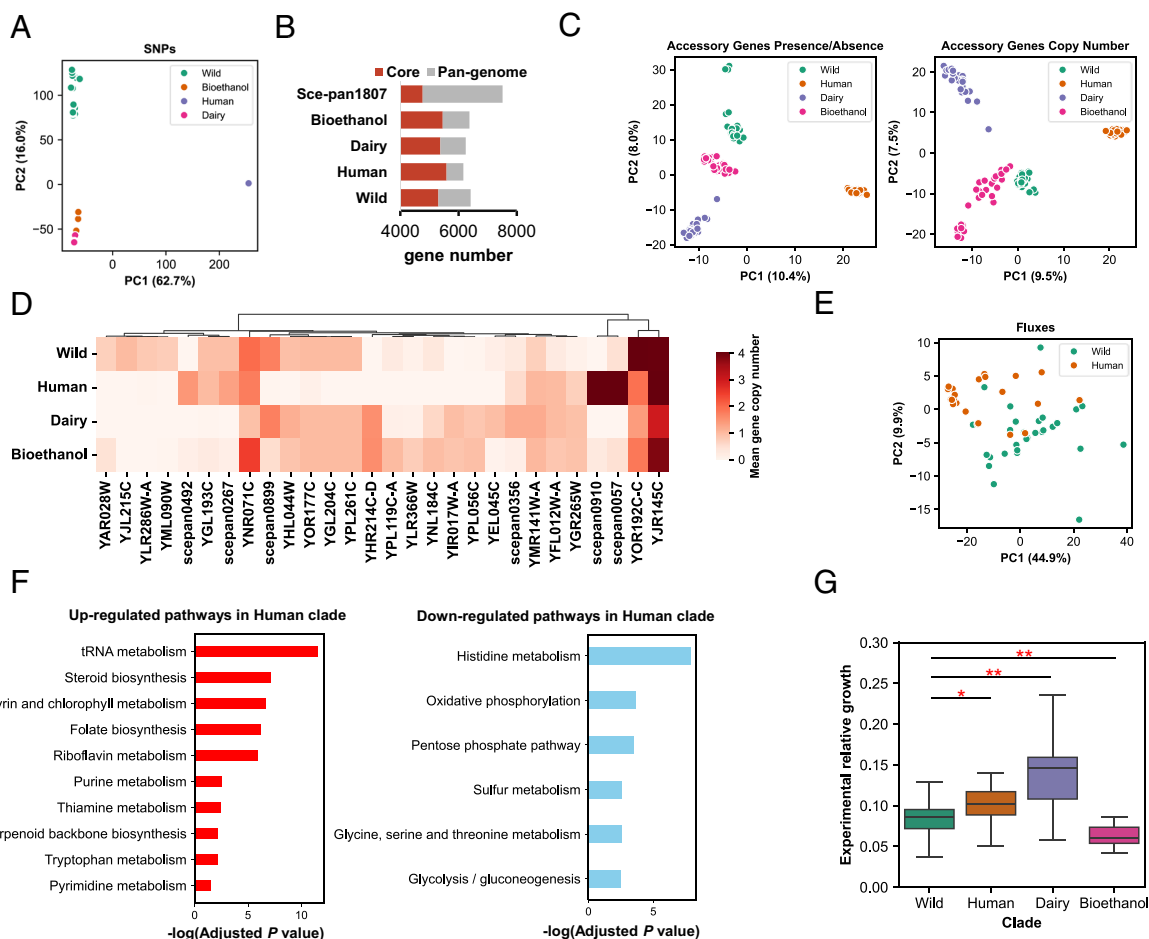
Fig. 5. Multidimensional analysis from pangenomes to metabolic networks helps to uncover the potential adaptive mechanisms of *S. cerevisiae* from human, dairy, and bioethanol-related niches. PCA clustering analysis of strains from different clades based on the SNPs (*A*). A comparison of the pangenome among strains from the clades of bioethanol, human, and dairy and wild (*B*). PCA clustering analysis of strains from different clades based on the gene presence/absence matrix (*Left*) and the gene copy number matrix of the accessory genome (*Right*) (*C*). Top genes with copy number variation used to classify strains from different clades by the random forest algorithm (*D*). PCA clustering analysis of strains from the wild and human gut environment using genome-scale fluxes obtained from ssGEMs simulation (*E*). Enrichment analysis of reactions with differential metabolic fluxes between human-related strains and the wild (*F*). Growth rate comparison across human, dairy, bioethanol, and wild clades based on experimental growth rate with glucose as the main carbon source (estimated based on the measured t-mid time dataset) (*G*).

the key metabolic features were delineated for the human, dairy, and bioethanol clades through machine learning, respectively (*SI Appendix*, Fig. S10 *D–F*). Specifically, within the human clade, dCMP deaminase was identified as a pivotal metabolic feature. The down-regulation of flux through the reaction catalyzed by dCMP deaminase might play a role in maintaining the balanced dNTP pool required for proper DNA metabolism (66). We further evaluated the metabolic variations at the pathway level (Fig. 5*F* and *SI Appendix*, Fig. S10 *G* and *H*). Interestingly, we found that there are many similar metabolic signatures among strains from the human, dairy, and bioethanol clades. For example, due to the long-term adaptation to oxygen-limited environments, both human and bioethanol strains exhibited a weak activity in the oxidative phosphorylation pathway even under the same batch cultivation. Intriguingly, dairy and human strains showed nearly identical patterns in metabolic variation, with upregulated pathways predominantly associated with growth, including tRNA metabolism (67), steroid biosynthesis (68), terpenoid biosynthesis (69), purine metabolism (70), and pyrimidine metabolism (71) (Fig. 5*F*). This result is consistent with the observed faster growth rates of both human and dairy strains compared to wild strains

(18) (Fig. 5*G*). It may be because the increased activity in ethanol secretion from niches specific adaptation occurred in those strains strengthens Crabtree effect when growing under the same condition (*SI Appendix*, Fig. S10*B*), thus promoting the cellular growth accordingly. Additionally, a significant downregulation of histidine metabolism in both human and dairy clades is consistent with findings from a recent comparative metabolomics study of *S. cerevisiae* (72) (Fig. 5*F* and *SI Appendix*, Fig. S10*G*). These shared metabolic variations among human, dairy, and bioethanol clades underscore the pervasive signatures of convergent evolution within the metabolic networks of independently adaptive subpopulations. As for strains within the human clade, we found that the upregulation of subpathways involved in cofactor metabolism, particularly thiamin metabolism (*P* value = 0.003), which generates the indispensable cofactor thiamine pyrophosphate (TPP) for enzymes participating in central carbon and energy metabolism (73), has been documented at both metabolic flux and transcriptional levels (Fig. 5*G* and *SI Appendix*, Fig. S11*A*). This upregulation may coincide with the adaptive evolution of strains inside the human gut environment, which is possibly driven by the resource competition among microbial species within the human gut (61, 74).

Discussion

Microorganisms, including *S. cerevisiae,* undergo adaptive evolution to improve cellular fitness with their genome compositions shaped by environmental pressures to promote trait diversity (9, 75, 76). Our study introduces a high-quality *S. cerevisiae* pangenome, Sce-pan1807 and put forward a unified framework integrating pangenome analysis, metabolic model simulation and transcriptomic profiling to elucidate the principles of metabolic reprogramming underlying yeast adaptation across diverse environments. Compared to prior studies (9, 20), our pangenome offers greater comprehensiveness due to a refined workflow and an extensive genomic dataset (Fig. 1*B* and *SI Appendix,* Fig. S1*D*). Our pangenome resolves inconsistencies in earlier versions by reducing gene redundancy and improving strain coverage (Fig. 1 *C* and *D* and *SI Appendix,* Fig. S1*E*), providing a chance to probe the genetic diversity of *S. cerevisiae* strains worldwide. By incorporating 796 additional strains, our pangenome significantly enhances representation across geographical regions and ecological niches. Notably, previously underrepresented niches, such as the Mantou and Baijiu clades, exhibit unique ploidy characteristics and novel genes, which offer new insight into genetic adaptation (*SI Appendix,* Fig. S2). Given that 1,223 new genes, absent in the *S. cerevisiae* model strains S288c and CEN.PK 113.7D, were identified through our pangenome analysis, the functional studies of those genes will certainly deepen our understanding of how the gain of genes impacts cellular fitness. Currently, our pangenome analysis is primarily focused on gene existence/copy analysis across diverse strains. In further comparative genomic analysis, the structural variants (SVs) and single-nucleotide polymorphisms (SNPs), which widely exist in yeast genomes (9, 77), should also be taken into account to illustrate how multidimensional evolutionary signatures contribute to trait diversity.

Based on Sce-pan1807, we developed the most comprehensive panGEM for *S. cerevisiae* to date, encompassing all genetically supported reactions potentially present in *S. cerevisiae*. Compared to existing *S. cerevisiae* panmodels, our panGEM features significant expansions including 29 additional genes, 202 new reactions, and 139 new metabolites (Fig. 3*B*). More importantly, we successfully constructed high-quality ssGEMs for 1,807 *S. cerevisiae* strains based on Sce-pan1807 and panGEM. These ssGEMs accurately reflect the metabolic potential of each strain by accounting for gene gains and losses. Compared to models derived from other *S. cerevisiae* pangenomes, our ssGEMs demonstrate superior performance, highlighting the advantages of our pangenome (*SI Appendix,* Fig. S12). However, considering the large number of ssGEMs developed in this work, more physiological datasets are required to further evaluate the prediction capability of those ssGEMs in order to apply these models in more detailed studies. GEMs may not accurately reflect in vivo cellular metabolic activities (78). Here, guided by large-scale transcriptomics from a recent study (18), ssGEMs for 907 isolates were refined. The pruned genes/reactions from ssGEMs reflected the reductive evolution in most domesticated clades, indicating that human-related activities significantly influence *S. cerevisiae* evolution (Fig. 4 *D* and *E*).

Further, our transcriptomics-refined ssGEMs help to elucidate the relationship between transcriptional regulation and metabolic flux across different metabolic pathways on a population level. Previous pantranscriptome studies were often limited to exploring transcriptional differences across strains due to the availability of only transcriptome and phenotype data (18). However, these transcriptional differences do not always directly correspond to variations in metabolic flux. A major advancement in our study is the integration of fluxomics with transcriptomics dataset, allowing us to investigate how transcriptional levels influence metabolic flux across different reactions and metabolic pathways in *S. cerevisiae*. We found a weak genome-wide correlation between fluxes and transcription across the studied strains, implying that changes in gene expression do not always reflect changes in metabolic fluxes (Fig. 4*F*). Notably, we observed that metabolic fluxes in pathways related to amino acid metabolism and central carbon metabolism are under tighter transcriptional control (Fig. 4*G*).

By integrating pangenome, transcriptome, and metabolic modeling analyses, our study provides a comprehensive framework for understanding how *S. cerevisiae* strains adapt to diverse stressful conditions. Here, we selected strains from human gut, dairy, and bioethanol clades (9) as representative examples to investigate adaptive strategies in specific ecological niches. Compared to wild strains, the strains from the human gut environment exhibit remarkable differences at multiscale levels, including gene presence/absence, gene copy number, and fluxes of the metabolic network, all of which may contribute to the adaptation underlying the human gut environment with limited oxygen supply and intense resource competition (Fig. 5). Thus, our analysis here provides a universal paradigm for profiling the metabolic rewiring of *S. cerevisiae* at a holistic level, which could be applied to examine the long-term evolution of other valuable microorganisms. It should be noted that our current omics integrative analysis mainly examines part of strains from diverse ecological niches and the RNA-seq datasets were obtained under the same condition with the glucose as the sole substrate (18). To achieve a more comprehensive understanding of the evolutionary metabolic reprogramming in *S. cerevisiae*, omics and phenotypic datasets under diverse conditions are needed to incorporate with existing models.

In conclusion, we have recreated the new version of pangenome and large-scale strain-specific genome-scale models for 1,807 strains of *S. cerevisiae*. These resources together systematically illustrate how *S. cerevisiae* strains flexibly adapt to various stressful conditions through multidimensional evolution at both genomic and metabolic levels. As additional omics and phenotypic data become available, these computational resources can be further updated and eventually accelerate the systematic evolutionary studies of yeast species.

Materials and Methods

All the materials and methods are detailed in *SI Appendix*: Genome sequences collection and annotation; Genome assessment; Pangenome reconstruction of *S. cerevisiae*; Phylogenetic analysis of 1,807 *S. cerevisiae* strains; Clade categorization of 1,807 *S. cerevisiae* strains; Pangenome reconstruction of 8 different yeast species; ssGEMs reconstruction; ssGEMs gap-filling; ssGEMs simulation of different substrate utilizing; Transcriptome-pruned ssGEMs construction and simulation; Genome-wide SNPs analysis for strains from specific evolutionary clades; Pangenome analysis for strains from specific evolutionary clades; Model simulation for strains from clade of bioethanol, dairy, human, and wild; Exploration of the potential genetic/metabolic features with machine learning; Differential expression and KEGG pathway enrichment analyses; Statistical analysis.

Data, Materials, and Software Availability. All scripts are accessible at https://github.com/hongzhonglu/Unified_Yeast_GEMs_Database (79). All large files, including ssGEMs, annotated genome data, and transcriptome-constrained ssGEMs are accessible at https://figshare.com/s/9c2faecc9d79d4825d0d (80). All other data are included in the manuscript and/or supporting information.

body has no role in the design of the study, analysis and interpretation of the data, preparation of the manuscript, and decision to submit the manuscript for publication.

Author affiliations: [a]State Key Laboratory of Microbial Metabolism, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China; [b]Division of Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, People's Republic of China; [c]University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China; [d]Department of Life Sciences, Chalmers University of Technology, Gothenburg SE412 96, Sweden; [e]BioInnovation Institute, Copenhagen N DK2200, Denmark; [f]Chinese Academy of Sciences Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, People's Republic of China; and [g]Dalian Key Laboratory of Energy Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, People's Republic of China

Author contributions: Y.J.Z. and H.L. designed research; H.W. performed research; H.W. and H.L. analyzed data; and H.W., J.N., Y.J.Z., and H.L. wrote the paper.

1. J. Steensels, B. Gallone, K. Voordeckers, K. J. Verstrepen, Domestication of industrial microbes. *Curr. Biol.* **29**, R381–R393 (2019).
2. J. Nielsen, Yeast cell factories on the horizon. *Science* **349**, 1050–1051 (2015).
3. A. Goffeau *et al.*, Life with 6000 genes. *Science* **274**, 546–567 (1996).
4. D. Botstein, G. R. Fink, Yeast: An experimental organism for 21st century biology. *Genetics* **189**, 695–704 (2011).
5. Y. Shao *et al.*, Creating a functional single-chromosome yeast. *Nature* **560**, 331–335 (2018).
6. E. A. Winzeler *et al.*, Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
7. G. Giaever *et al.*, Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
8. G. Turco *et al.*, Global analysis of the yeast knockout phenome. *Sci. Adv.* **9**, eadg5702 (2023).
9. J. Peter *et al.*, Genome evolution across 1, 011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).
10. S. Marsit *et al.*, Evolutionary biology through the lens of budding yeast comparative genomics. *Nat. Rev. Genet.* **18**, 581–598 (2017).
11. Y. O. Zhu, G. Sherlock, D. A. Petrov, Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3 (Bethesda)* **6**, 2421–2434 (2016).
12. T. J. Lee *et al.*, Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages. *Genome Res.* **32**, 864–877 (2022).
13. B. Gallone *et al.*, Domestication and divergence of Saccharomyces cerevisiae beer yeasts. *Cell* **166**, 1397–1410.e16 (2016).
14. S. F. Duan *et al.*, The origin and adaptive evolution of domesticated populations of yeast from far east Asia. *Nat. Commun.* **9**, 2690 (2018).
15. C. Díaz-Muñoz, M. Verce, L. Vuyst, S. Weckx, Phylogenomics of a *Saccharomyces cerevisiae* cocoa strain reveals adaptation to a West African fermented food population. *iScience* **25**, 105309 (2022).
16. K. Kang *et al.*, Linking genetic, metabolic, and phenotypic diversity among *Saccharomyces* strains using multi-omics associations. *Gigascience* **8**, giz015 (2019).
17. D. Y. Han *et al.*, Adaptive gene content and allele distribution variations in the wild and domesticated populations of *Saccharomyces cerevisiae*. *Front. Microbiol.* **12**, 631250 (2021).
18. É. Caudal *et al.*, Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. *Nat. Genet.* **56**, 1278–1287 (2024).
19. G. Vernikos, D. Medini, D. R. Riley, H. Tettelin, Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **23**, 148–154 (2015).
20. G. Li, B. Ji, J. Nielsen, The pan-genome of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **19**, foz064 (2019).
21. E. J. O'Brien, J. M. Monk, B. O. Palsson, Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987 (2015).
22. D. Henriques *et al.*, A multiphase multiobjective dynamic genome-scale model shows different redox balancing among yeast species of the *Saccharomyces* genus in fermentation. *mSystems* **6**, 00260-21 (2021).
23. S. Timouma, L. N. Balarezo-Cisneros, J.-M. Schwartz, D. Delneri, Development of a genome-scale metabolic model for the lager hybrid yeast *S. pastorianus* to understand the evolution of metabolic pathways in industrial settings. *mSystems* **9**, e00429-24 (2024).
24. J. Förster, I. Famili, P. Fu, B. Ø. Palsson, J. Nielsen, Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
25. Y. Chen, F. Li, J. Nielsen, Genome-scale modeling of yeast metabolism: Retrospectives and perspectives. *FEMS Yeast Res.* **22**, foac003 (2022).
26. H. Lu *et al.*, A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **10**, 3586 (2019).
27. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
28. C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
29. Y. Yuan *et al.*, Pan-genome analysis of transcriptional regulation in six *Salmonella enterica*Serovar *Typhimurium* strains reveals their different regulatory structures. *mSystems* **7**, e00467-22 (2022).
30. T. Yang, F. Gao, High-quality pan-genome of *Escherichia coli* generated by excluding confounding and highly similar strains reveals an association between unique gene clusters and genomic islands. *Brief. Bioinform.* **23**, bbac283 (2022).
31. C. J. Norsigian *et al.*, Systems biology approach to functionally assess the *Clostridioides difficile* pangenome reveals genetic diversity with discriminatory power. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2119396119 (2022).
32. X. X. Shen *et al.*, Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**, 1533–1545.e20 (2018).
33. H. Lu *et al.*, Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection. *Mol. Syst. Biol.* **17**, e10427 (2021).
34. R. Riley *et al.*, Comparative genomics of biotechnologically important yeasts. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9882–9887 (2016).
35. C. Zhang *et al.*, Yeast9: A consensus genome-scale metabolic model for *S. cerevisiae* curated by the community. *Mol. Syst. Biol.* **20**, 1134–1150 (2024).
36. S. Chen, J. S. Brockenbrough, J. E. Dove, J. P. Aris, Homocitrate synthase is located in the nucleus in the yeast Saccharomyces cerevisiae. *J. Biol. Chem.* **272**, 10839–10846 (1997).
37. S. Isogai *et al.*, High-level production of lysine in the yeast *Saccharomyces cerevisiae* by rational design of homocitrate synthase. *Appl. Environ. Microbiol.* **87**, e00600-21 (2021).
38. S. Opdam *et al.*, A systematic evaluation of methods for tailoring genome-scale metabolic models. *Cell Syst.* **4**, 318–329.e6 (2017).
39. C. Culley, S. Vijayakumar, G. Zampieri, C. Angione, A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18869–18879 (2020).
40. S. A. Becker, B. O. Palsson, Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* **4**, e1000082 (2008).
41. A. Rodríguez-Gijón *et al.*, A genomic perspective across Earth's microbiomes reveals that genome size in Archaea and Bacteria is linked to ecosystem type and trophic strategy. *Front. Microbiol.* **12**, 761869 (2022).
42. D. K. Ngugi *et al.*, Abiotic selection of microbial genome size in the global ocean. *Nat. Commun.* **14**, 1384 (2023).
43. R. Agren *et al.*, Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10**, 721 (2014).
44. S. Gopalakrishnan *et al.*, Guidelines for extracting biologically relevant context-specific metabolic models using gene expression data. *Metab. Eng.* **75**, 181–191 (2022).
45. Y. Li, W. B. Holmes, D. R. Appling, U. L. RajBhandary, Initiation of protein synthesis in *Saccharomyces cerevisiae* mitochondria without formylation of the initiator tRNA. *J. Bacteriol.* **182**, 2886–2892 (2000).
46. E. Wiltrout, J. M. Goodenbour, M. Fréchin, T. Pan, Misacylation of tRNA with methionine in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **40**, 10494–10506 (2012).
47. Y.-Q. Jiang, J.-P. Lin, Recent progress in strategies for steroid production in yeasts. *World J. Microbiol. Biotechnol.* **38**, 93 (2022).
48. S. A. Henry, S. D. Kohlwein, G. M. Carman, Metabolism and regulation of glycerolipids in the yeast *Saccharomyces cerevisiae*. *Genetics* **190**, 317–349 (2012).
49. C. Ye, W. M. M. S. Bandara, M. L. Greenberg, Regulation of inositol metabolism is fine-tuned by inositol pyrophosphates in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **288**, 24898–24908 (2013).
50. L. Gmelch *et al.*, Comprehensive vitamer profiling of folate mono- and polyglutamates in baker's yeast (*Saccharomyces cerevisiae*) as a function of different sample preparation procedures. *Metabolites* **10**, 301 (2020).
51. S. Hohmann, P. A. Meacock, Thiamin metabolism and thiamin diphosphate-dependent enzymes in the yeast *Saccharomyces cerevisiae*: Genetic regulation. *Biochim. Biophys. Acta* **1385**, 201–219 (1998).
52. P. C. Thiaville, D. Iwata-Reuyl, V. de Crécy-Lagard, Diversity of the biosynthesis pathway for threonylcarbamoyladenosine (t(6)A), a universal modification of tRNA. *RNA Biol.* **11**, 1529–1539 (2014).
53. D. Qiu *et al.*, Analysis of inositol phosphate metabolism by capillary electrophoresis electrospray ionization mass spectrometry. *Nat. Commun.* **11**, 6035 (2020).
54. W. J. C. Dekker, S. J. Wiersma, J. Bouwknegt, C. Mooiman, J. T. Pronk, Anaerobic growth of *Saccharomyces cerevisiae* CEN.PK113-7D does not depend on synthesis or supplementation of unsaturated fatty acids. *FEMS Yeast Res.* **19**, foz060 (2019).
55. K. Mercurio, D. Singh, E. Walden, K. Baetz, Global analysis of *Saccharomyces cerevisiae* growth in mucin. *G3 (Bethesda)* **11**, jkab294 (2021).
56. S. R. Hackett *et al.*, Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, aaf2786 (2016).
57. G. Li, L. Liu, W. Du, H. Cao, Local flux coordination and global gene expression regulation in metabolic modeling. *Nat. Commun.* **14**, 5700 (2023).
58. V. Chubukov *et al.*, Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Mol. Syst. Biol.* **9**, 709 (2013).
59. S. Bordel, R. Agren, J. Nielsen, Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput. Biol.* **6**, e1000859 (2010).
60. R. Yu, E. Vorontsov, C. Sihlbom, J. Nielsen, Non-random organization of flux control mechanisms in yeast central metabolic pathways. bioRxiv [Preprint] (2021). https://www.biorxiv.org/content/10.1101/2021.12.15.472747v1 (Accessed 11 June 2024).
61. V. Raghavan, C. F. Aquadro, E. Alani, Baker's yeast clinical isolates provide a model for how pathogenic yeasts adapt to stress. *Trends Genet.* **35**, 804–817 (2019).
62. D. M. Lane, D. L. Valentine, X. Peng, Genomic analysis of the marine yeast *Rhodotorula sphaerocarpa* ETNP2018 reveals adaptation to the open ocean. *BMC Genomics* **24**, 695 (2023).
63. S. R. Engel *et al.*, New data and collaborations at the Saccharomyces Genome Database: Updated reference genome, alleles, and the Alliance of Genome Resources. *Genetics* **220**, iyab224 (2022).
64. M. L. Jenior, T. J. Moutinho, B. V. Dougherty, J. A. Papin, Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput. Biol.* **16**, e1007099 (2020).
65. A. P. Jacobus *et al.*, Comparative genomics supports that Brazilian bioethanol *Saccharomyces cerevisiae* comprise a unified group of domesticated strains related to cachaça spirit yeasts. *Front. Microbiol.* **12**, 644089 (2021).
66. M. Niu *et al.*, ALR encoding dCMP deaminase is critical for DNA damage repair, cell cycle progression and plant development in rice. *J. Exp. Botany* **68**, 5773–5786 (2017).
67. T. Yoshihisa, C. Ohshima, K. Yunoki-Esaki, T. Endo, Cytoplasmic splicing of tRNA in *Saccharomyces cerevisiae*. *Genes Cells* **12**, 285–297 (2007).

68. T. Jordá, S. Puig, Regulation of ergosterol biosynthesis in *Saccharomyces cerevisiae*. *Genes* **11**, 795 (2020).
69. P. Yang *et al.*, Thermotolerance improvement of engineered *Saccharomyces cerevisiae* ERG5 Delta ERG4 Delta ERG3 Delta, molecular mechanism, and its application in corn ethanol production. *Biotechnol. Biofuels* **16**, 66 (2023).
70. B. Daignan-Fornier, B. Pinson, Yeast to study human purine metabolism diseases. *Cells* **8**, 67 (2019).
71. F. Lacroute, Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. *J. Bacteriol.* **95**, 824–832 (1968).
72. R. Tengölics *et al.*, The metabolic domestication syndrome of budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2313354121 (2024).
73. M. Li *et al.*, Thiamine biosynthesis in *Saccharomyces cerevisiae* is regulated by the NAD+-dependent histone deacetylase Hst1. *Mol. Cell. Biol.* **30**, 3329–3341 (2010).
74. Y. Wu, B. Li, B. Miao, C. Xie, Y.-Q. Tang, *Saccharomyces cerevisiae* employs complex regulation strategies to tolerate low pH stress during ethanol production. *Microb. Cell Fact* **21**, 247 (2022).
75. E. Bosi *et al.*, Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3801–E3809 (2016).
76. F. Rosconi *et al.*, A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat. Microbiol.* **7**, 1580–1592 (2022).
77. S. O'Donnell *et al.*, Telomere-to-telomere assemblies of 142 strains characterize the genome structural landscape in *Saccharomyces cerevisiae*. *Nat. Genet.* **55**, 1390–1399 (2023).
78. H. Lu, E. J. Kerkhoven, J. Nielsen, Multiscale models quantifying yeast physiology: Towards a whole-cell model. *Trends Biotechnol.* **40**, 291–305 (2022).
79. H. Wang, J. Zhou, H. Lu, Code from "Yeast adapts to diverse ecological niches driven by genomics and metabolic reprogramming" Github. https://github.com/hongzhonglu/Unified_Yeast_GEMs_Database. Deposited 20 December 2024.
80. H. Wang, J. Zhou, H. Lu, Data from "Yeast adapts to diverse ecological niches driven by genomics and metabolic reprogramming" Figshare. https://figshare.com/s/9c2faecc9d79d4825d0d. Deposited 20 December 2024.