RESEARCH ARTICLE SUMMARY

ECOLOGY

Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts

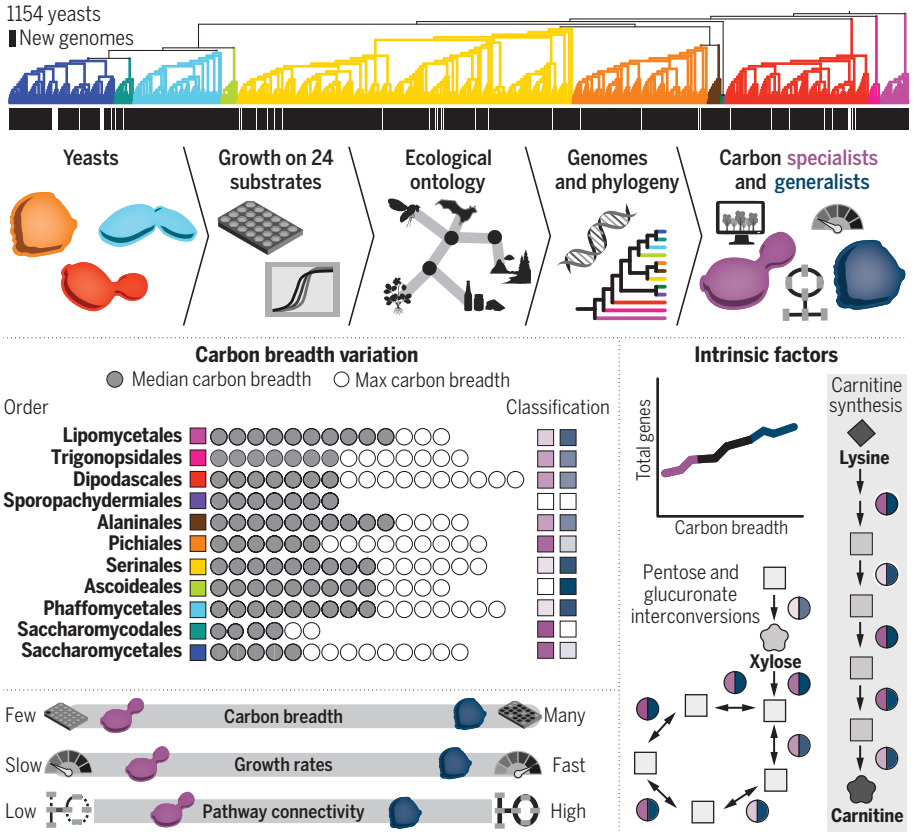
Dana A. Opulente†, Abigail Leavitt LaBella†, Marie-Claire Harrison‡, John F. Wolters‡, Chao Liu, Yonglin Li, Jacek Kominek, Jacob L. Steenwyk, Hayley R. Stoneman, Jenna VanDenAvond, Caroline R. Miller, Quinn K. Langdon, Margarida Silva, Carla Gonçalves, Emily J. Ubbelohde, Yuanning Li, Kelly V. Buh, Martin Jarzyna, Max A. B. Haase, Carlos A. Rosa, Neža Čadež, Diego Libkind, Jeremy H. DeVirgilio, Amanda Beth Hulfachor, Cletus P. Kurtzman, José Paulo Sampaio, Paula Gonçalves, Xiaofan Zhou, Xing-Xing Shen, Marizeth Groenewald, Antonis Rokas*, Chris Todd Hittinger*

INTRODUCTION: It is often said that the jack-of-all-trades is the master of none. Niche breadth varies widely across the tree of life, from narrow in specialists to broad in generalists. One ecological paradigm explains this variation by invoking trade-offs between niche breadth and performance efficiency. Generalists perform moderately well in many niches, whereas each specialist has an advantage in its own niche. A second paradigm explains niche breadth variation through extrinsic and intrinsic factors. Extrinsic factors are ecological variables that include nutrient availability, temperature, organism interactions, and heterogeneity. Intrinsic factors are encoded by organisms' genomes and affect how they access and process nutrients and tolerate stresses.

RATIONALE: To study niche breadth macroevolution, we deployed an ancient model subphylum uniquely poised for studies at [illegible]nomic, metabolic, and ecological scales. The yeast subphylum Saccharomycotina of kingdom Fungi is best known for the model baker's yeast *Saccharomyces cerevisiae* and the major human pathogen *Candida albicans*, but more than 1000 species have radiated during more than 400 million years into diverse ecological niches. Yeasts harbor gene sequence divergence comparable to that of animals and plants and are found in environments ranging from bats to cadaver tanks and from cheese caves to biofuel factories.

RESULTS: We generated a vast dataset of genome sequences of 1154 yeasts from nearly every known species, quantitative metabolic growth data in 24 conditions, and a hierarchical ecological ontology of isolation environments. Using evolutionary, machine learning, and network analyses, we found that yeast metabolic niche breadth is largely shaped by intrinsic factors. Generalist genomes encoded more genes and metabolic reactions, and our machine learning algorithm distinguished generalists from specialists using genome content with high accuracy. The most predictive features in our dataset pointed to specific genes in four pathways or complexes that are directly involved in carbon and energy metabolism, often by enhancing metabolic flexibility and robustness. Through ancestral trait reconstruction and coevolution analyses, we further demonstrated that generalists were more likely to have retained or gained traits, whereas specialists repeatedly arose through pervasive gene and trait loss. We did not find evidence for trade-offs between carbon niche breadth and growth rates; compared with specialists, carbon generalists grew faster in laboratory conditions and on more nitrogen sources. These results suggest that intrinsic genetic factors are a major driver of microbial diversity and niche breadth variation.

CONCLUSION: We generated a genomic, metabolic, and ecological dataset to show how metabolic diversity and niche breadth are encoded in yeast genomes and how these traits have evolved over deep time. Coupling a comprehensive dataset with a robust analytical framework paints a rich portrait of a diverse eukaryotic subphylum with immense impacts on human health, agriculture, and biotechnology that provides a roadmap connecting DNA to diversity. ■



A comprehensive initiative capturing genomic, metabolic, and ecological diversity among 1154 yeasts of the fungal subphylum Saccharomycotina. We built a robust phylogeny and generated extensive genomic, phenotypic, and ecological data. We identified carbon niche breadth variation and used machine leaning to identify several intrinsic factors that contribute to carbon generalism.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: antonis.rokas@vanderbilt.edu (A.R.); cthittinger@wisc.edu (C.T.H.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as D. A. Opulente *et al.*, *Science* **384**, eadj4503 (2024). DOI: 10.1126/science.adj4503

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adj4503

RESEARCH ARTICLE

ECOLOGY

Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts

Dana A. Opulente[1,2,3]†, Abigail Leavitt LaBella[4,5,6]†, Marie-Claire Harrison[4,5]‡, John F. Wolters[1,2]‡, Chao Liu[7], Yonglin Li[8], Jacek Kominek[1,2,9], Jacob L. Steenwyk[4,5,10], Hayley R. Stoneman[1,2,11], Jenna VanDenAvond[1,2], Caroline R. Miller[1,2], Quinn K. Langdon[1], Margarida Silva[12,13], Carla Gonçalves[1,4,5,12,13], Emily J. Ubbelohde[1,2], Yuanning Li[4,14,15], Kelly V. Buh[1], Martin Jarzyna[1,16], Max A. B. Haase[1,2,17,18], Carlos A. Rosa[19], Neža Čadež[20], Diego Libkind[21], Jeremy H. DeVirgilio[22], Amanda Beth Hulfachor[1,2], Cletus P. Kurtzman[22]§, José Paulo Sampaio[12,13], Paula Gonçalves[12,13], Xiaofan Zhou[4,8], Xing-Xing Shen[4,7], Marizeth Groenewald[23], Antonis Rokas[4,5]*, Chris Todd Hittinger[1,2]*

Organisms exhibit extensive variation in ecological niche breadth, from very narrow (specialists) to very broad (generalists). Two general paradigms have been proposed to explain this variation: (i) trade-offs between performance efficiency and breadth and (ii) the joint influence of extrinsic (environmental) and intrinsic (genomic) factors. We assembled genomic, metabolic, and ecological data from nearly all known species of the ancient fungal subphylum Saccharomycotina (1154 yeast strains from 1051 species), grown in 24 different environmental conditions, to examine niche breadth evolution. We found that large differences in the breadth of carbon utilization traits between yeasts stem from intrinsic differences in genes encoding specific metabolic pathways, but we found limited evidence for trade-offs. These comprehensive data argue that intrinsic factors shape niche breadth variation in microbes.

The ecological niche is a fundamental concept in ecology and evolutionary biology that explains the diversity and resource use of organisms through space and time. Species with broad niche breadths are defined as generalists, whereas those with narrow ones are specialists. There are many biotic and abiotic dimensions of the niche that can and do vary among organisms (*1*–*3*), which begs the question: What factors contribute to niche breadth variation?

Two broad paradigms have been offered as answers across a variety of taxa. The first paradigm postulates that both niche generalism and specialism are governed by trade-offs between performance efficiency and niche breadth (*4*–*9*). In the context of metabolic niche breadth, selection for increased efficiency in using a specific food source will be coupled to selection against using other food sources and vice versa. Over the long term, such selection produces generalists that use more substrates less efficiently and specialists that use fewer substrates more efficiently. Consistent with these expectations, selection for specialization in using a single food source in replicate populations of the bacterium *Escherichia coli* was coupled to a reduction in their ability to catabolize other food sources (*10*).

The second paradigm postulates that generalist and specialist phenotypes are the outcome of the joint influence of diverse extrinsic (environmental) and intrinsic (genomic) factors (*11*–*16*). Generalists and specialists are shaped by the environments in which they occur and the evolvability of their metabolic pathways rather than by trade-offs. These specific conditions will result in a unifying set of extrinsic and intrinsic features that govern the evolution of generalist and specialist phenotypes.

Extrinsic factors are the environments in which species live. They can vary with respect to numerous abiotic and biotic factors, such as spatial and temporal heterogeneity, temperature, and carbon and nitrogen availability. For example, carbon sources have been shown to be limited within endothermic hosts (*17*, *18*); temperatures and soil moisture can vary between woodland and meadow habitats as a result of canopy cover (*19*); and the availability of nitrogen sources (*20*, *21*), carbon sources (*22*–*24*), and growth-inhibiting specialized metabolites can differ because of the activities of other organisms in the environment (*25*, *26*). Variation in one or more of these extrinsic factors could exert selective pressure on traits, resulting in generalism and specialism (*27*).

Intrinsic factors that may influence niche breadth include the evolution of promiscuous enzymes responsible for the utilization of multiple resources (*17*, *28*–*31*), as well as overlapping biochemical, developmental, and genetic pathways (*15*, *16*). For example, yeast *MAL* and *IMA* genes are promiscuous enzymes associated with the utilization of multiple carbon sources in yeasts; that is, they can increase niche breadth by enabling broader consumption (*17*, *28*). Conversely, gene loss due to drift or relaxed selection, which is likely in environments with lower nutrient diversity, could lead to narrower niche breadths (*32*). The diversity of traits and the genes that control them leads to the hypothesis that niche breadth variation may reflect the interplay between evolutionary and ecological forces acting on intrinsic factors.

The subphylum Saccharomycotina (phylum Ascomycota, kingdom Fungi)—which includes the baker's yeast *Saccharomyces cerevisiae*,

[1]Laboratory of Genetics, Wisconsin Energy Institute, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, University of Wisconsin–Madison, Madison, WI 53726, USA. [2]DOE Great Lakes Bioenergy Research Center, University of Wisconsin–Madison, Madison, WI 53726, USA. [3]Biology Department, Villanova University, Villanova, PA 19085, USA. [4]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA. [5]Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA. [6]North Carolina Research Center (NCRC), Department of Bioinformatics and Genomics, The University of North Carolina at Charlotte, Kannapolis, NC 28081, USA. [7]College of Agriculture and Biotechnology and Centre for Evolutionary and Organismal Biology, Zhejiang University, Hangzhou 310058, China. [8]Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Center, South China Agricultural University, Guangzhou 510642, China. [9]LifeMine Therapeutics, Inc., Cambridge, MA 02140, USA. [10]Howard Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA. [11]University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA. [12]UCIBIO, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal. [13]Associate Laboratory i4HB, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal. [14]Institute of Marine Science and Technology, Shandong University, Qingdao 266237, China. [15]Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao 266237, China. [16]Graduate Program in Neuroscience and Department of Biology, Washington University School of Medicine, St. Louis, MO 63130, USA. [17]Vilcek Institute of Graduate Biomedical Sciences and Institute for Systems Genetics, NYU Langone Health, New York, NY 10016, USA. [18]Department of Mechanistic Cell Biology, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. [19]Departamento de Microbiologia, ICB, C.P. 486, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil. [20]Food Science and Technology Department, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia. [21]Centro de Referencia en Levaduras y Tecnología Cervecera (CRELTEC), Instituto Andino Patagónico de Tecnologías Biológicas y Geoambientales (IPATEC), Universidad Nacional del Comahue, CONICET, CRUB, Quintral 1250, San Carlos de Bariloche, 8400, Río Negro, Argentina. [22]Mycotoxin Prevention and Applied Microbiology Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, US Department of Agriculture, Peoria, IL 61604, USA. [23]Westerdijk Fungal Biodiversity Institute, 3584 CT Utrecht, Netherlands.

*Corresponding author. Email: antonis.rokas@vanderbilt.edu (A.R.); cthittinger@wisc.edu (C.T.H.)

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

§Deceased.

the opportunistic pathogen *Candida albicans*, and the oleochemical cell factory *Yarrowia lipolytica*—exhibits extensive ecological, genomic, and metabolic diversity. Thus, it is a superb system for testing paradigms for the evolution of metabolic niche breadth (Fig. 1). The genomes of Saccharomycotina species, commonly referred to as yeasts, are highly diverse; levels of gene sequence divergence across yeasts are comparable to levels observed across plants and animals, and the subphylum also harbors considerable variation in gene content, including metabolic genes (*28*). Additionally, extensive experimental work in model yeasts, such as *S. cerevisiae* (*33*) and *C. albicans* (*34*), provides validated functional genetic information.

Yeast growth profiles have been characterized across many carbon and nitrogen sources and environmental conditions (e.g., temperature), and they are highly variable across species (*17*, *28*, *35*). This phenotypic diversity is coupled to their ecological diversity. Yeasts are found in almost every biome on a wide array of substrates, and the isolation environments (defined as the specific environmental location where a strain was originally isolated) of these yeasts are associated with specific phenotypic traits. For example, both glucose and sucrose fermentation are positively associated with living on fruits, fermented substrates, and juices (*17*), particularly among multiple yeast genera that have been linked to wine production and food spoilage (*17*, *36*, *37*). Opportunistic fungal pathogens have also evolved metabolic strategies that allow them to colonize the complex ecosystem of the human body, where carbon availability varies spatially and temporally (*17*, *38*, *39*). This treasure trove of genomic, metabolic, and

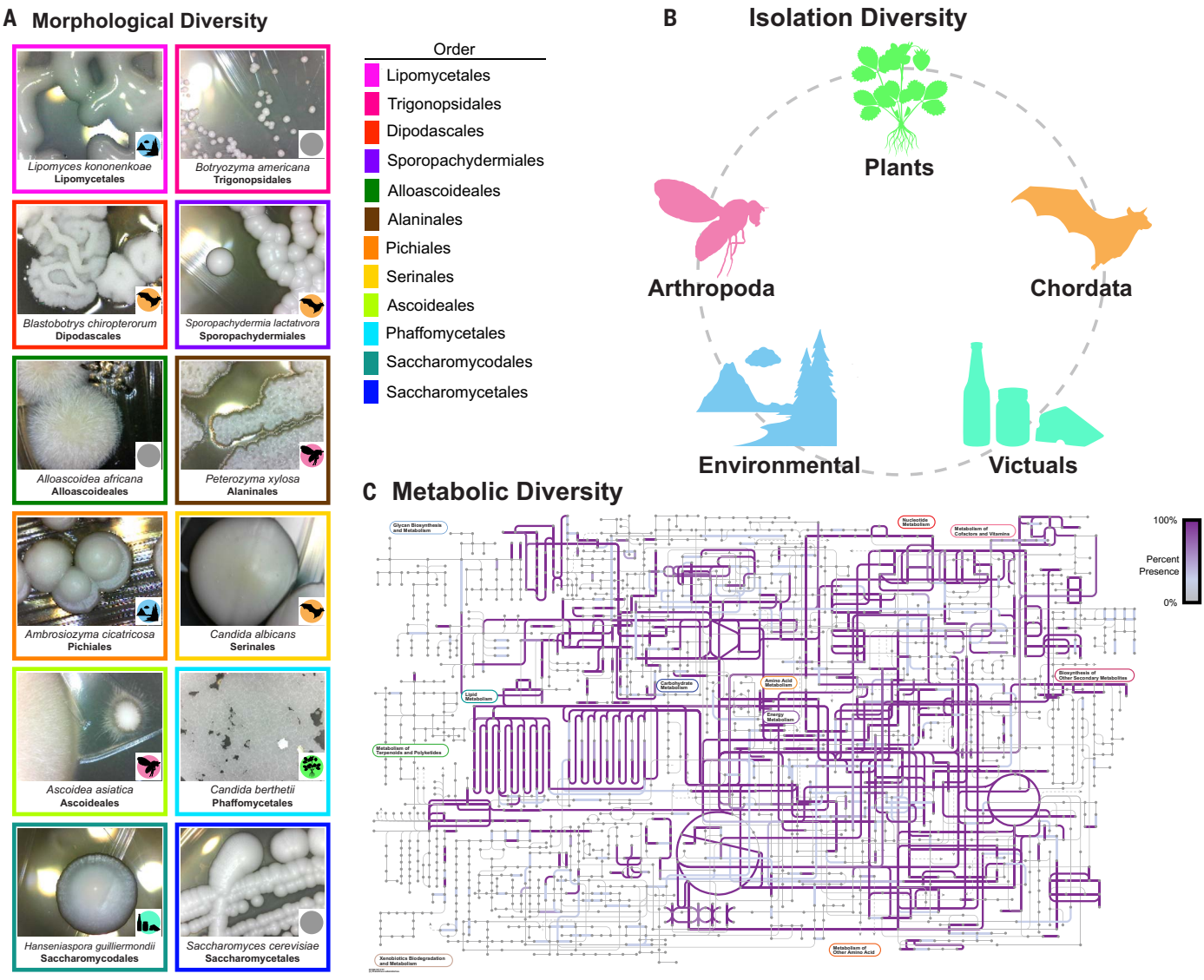


Fig. 1. Yeasts are morphologically, ecologically, and metabolically diverse. (**A**) Images of yeasts from different orders. The color of the box surrounding the image indicates the species' order. The color of the circle in the bottom right-hand corner of the image represents the isolation environment for the strain of the species sequenced and phenotyped during this study. Yeast colonies are morphologically diverse; they can vary in shape, color, size, dullness, etc. (**B**) Yeasts have been isolated from every biome and continent. Strains studied were found on plants, in animals, in soil, and in many other environments. Strain-level isolation data were placed into an ecological ontology to allow for the identification of yeasts that shared higher-level ontological classes. (**C**) Yeasts are metabolically diverse. The image represents the KOs present across Saccharomycotina metabolic networks. Any pathway that is highlighted in purple is present across a subset of yeasts; the saturation of the purple represents the proportion of yeasts with the pathway.

environmental diversity across a subphylum makes Saccharomycotina an attractive and highly tractable system for studying niche breadth evolution.

To gain insight into the factors that contribute to metabolic niche breadth variation, we quantified variation in genome content, isolation environment, and carbon and nitrogen metabolism for 1154 yeast strains, which represent nearly all known species in the subphylum Saccharomycotina. This dataset enabled us to evaluate the evidence for the two niche breadth evolution paradigms (trade-offs versus underlying intrinsic and extrinsic factors) across species with broad (generalists) and narrow (specialists) carbon niche breadths. Our evolutionary, machine learning, and network analyses uncovered a unifying set of intrinsic factors among generalists that were largely absent in specialists and pinpointed specific genetic differences between generalists and specialists, including previously unidentified associations between carbon generalism and specific metabolic pathways. By contrast, we found limited evidence for trade-offs between carbon generalism and growth rate. Through ancestral trait reconstruction and coevolution analyses, we further demonstrated that generalists were more likely to have retained or gained traits, whereas specialists repeatedly arose through pervasive gene and trait loss. The genomic, metabolic, evolutionary, and ecological data for nearly all known species of the 400-million-year-old yeast subphylum Saccharomycotina provided in this work, coupled with the availability of multiple genetic models in the subphylum, present an inimitable resource and framework for linking genomic variation to phenotypic and ecological variation.

A genomic, evolutionary, and metabolic portrait of Saccharomycotina

We sequenced and assembled 953 genomes in this study and combined them with 140 genomes previously sequenced by the Y1000+ Project (*40*) and 61 publicly available genomes (data S1). Our dataset contained 1154 genomes from 1051 species, including 1037 taxonomic type (i.e., ex-type) strains. Multiple strains were sequenced from 41 species, including a total of 19 recognized varieties distributed across nine species (i.e., two to three varieties per species). Sixty-one of the strains whose genomes were sequenced could not be assigned to any of the known species; thus, they are candidates for novel species. The genomic dataset spans 96 yeast genera, which is ~90% of currently described genera (*41*). Excluded genera were typically those for which no living culture was available or those described after our last round of genome sequencing in February 2021. Our genome sequencing added between 1 and 336 species to each order, most notably expanding the order Serinales (previously major clade CUG-Ser1), which contains the human pathogens *C. albicans* and *Candida auris*, from 94 genomes to 430. All genome assemblies totaled ~15 billion base pairs. The assemblies had a mean N50 (the sequence length of the shortest contig at 50% of the total assembly length) of 387.5 kb, which was comparable to our previous smaller-scale dataset of 332 genomes (417.2 kb) (fig. S1A and data S1) (*28*). All genomes were annotated to identify putative coding sequences. On average, 5908 ± 1069 (mean ± SD) protein-coding sequences were identified per genome with a range from 3775 (*Starmerella lactis-condensi*) to 20,704 (*Magnusiomyces magnusii*) (fig. S1B) (*42*). Functional annotations were conducted using Kyoto Encyclopedia of Genes and Genomes (KEGG) and InterPro. GC content (subphylum mean = 41.1 ± 6.61%) ranged from 23.9% (*Candida bohioensis*) to 66.8% (*Candida pseudocylindracea*), and genome size (subphylum mean = 13.2 ± 3.5 Mb) ranged from 7.2 Mb (*Starmerella lactis-condensi*) to 41.3 Mb (*M. magnusii*) (fig. S1, C and D, and data S1). Of the 1154 yeast genomes, 1000 (~87%) had ≥90% of the 2137 predefined single-copy orthologs defined by OrthoDB v10 (data S1) (*43*, *44*).

At least three independent nuclear codon reassignments are known to have occurred during the evolution of the subphylum (*45*). Given the large number of newly added genomes, we inferred codon tables and tRNA genes to confirm the known reassignments and test for potential new reassignments (data S2). These results were consistent with the previously observed codon reassignments. Notably, genomes of the order Ascoideales had a diversity of tRNAs with CAG anticodons predicted to decode CUG codons, which is consistent with previous findings that these yeasts may stochastically decode CUG as both leucine and serine (*46*).

To infer the genome-scale phylogeny of the Saccharomycotina, we used 1403 orthologous groups (OGs) from 1154 Saccharomycotina genomes and 21 outgroups. Nearly all internodes in both concatenation-based (1136/1153, 99%) and coalescent-based (1123/1153, 97%) phylogenies received strong (≥95%) support (Fig. 2 and figs. S2 and S3). The two phylogenies were highly congruent, with only 60/1153 (5%) conflicting internodes (fig. S3). Moreover, relationships among the 12 recently circumscribed taxonomic orders (*41*) (previously major clades) were congruent with previous studies (*28*, *47*, *48*), including the placement of the Ascoideales (previously CUG-Ser2) and Alaninales (previously CUG-Ala).

To examine the evolution of metabolic niche breadth across Saccharomycotina, we quantified the growth rates of 853 yeast strains on 18 carbon sources, 6 nitrogen sources, and a no-carbon control (data S3). We found that yeasts displayed variation in growth rates across carbon (fig. S4A) and nitrogen sources (fig. S4B); on average, each yeast strain could metabolize eight carbon (Fig. 3A) and two nitrogen sources (fig. S5). Comparison of growth rates on different carbon sources revealed that 65.22% of yeasts (n = 557) grew fastest on glucose, whereas the remaining 34.78% (n = 297) grew faster on another carbon source (fig. S6). Mannose, an epimer of glucose not typically tested in yeast growth experiments, was the carbon source on which yeasts grew fastest, on average, after glucose (n = 112). We also found that 77 yeasts grew faster on fructose than glucose, including cases where their maximum growth rate was on a third carbon source. Several of these yeasts (n = 7) were in Dipodascales, which contains many known fructophilic yeasts (*49*). The ability to grow faster on fructose was independently verified in a second laboratory on a subset of yeasts (data S4).

A lack of evidence for trade-offs between carbon niche breadth and growth rates

We statistically classified yeasts into three categories for both carbon and nitrogen utilization niche breadths—specialist, standard, and generalist (data S3). We found that, for both carbon and nitrogen metabolism, most yeasts were classified as standard yeasts (i.e., yeasts that did not fall into the extremes for carbon niche breadth) (76.0%, 648/853, and 78.4%, 669/853, respectively) (data S3 and Fig. 3A). Of the remaining 24.0% (n = 205/853), 53.7% (n = 110/205) were specialists and 46.3% (n = 95/205) were generalists for carbon sources (Fig. 3A). The median numbers of carbon sources used by specialist, standard, and generalist yeasts were 4, 8, and 12, respectively. Carbon generalists and specialists were widely distributed across the subphylum (Fig. 2), and all orders with more than 15 phenotyped strains (n = 8) featured both generalists and specialists. However, the relative proportion of generalists and specialists within orders varied greatly. For example, the order Saccharomycetales (n = 82) had 3 generalists and 33 specialists, whereas the order Serinales (n = 347) had 53 generalists and 9 specialists. This result suggests that yeast orders exhibit distinct eco-evolutionary trajectories.

First, we tested for a trade-off between growth rate and carbon niche breadth by investigating whether specialists had a growth rate advantage over other yeasts in some conditions. We compared all growth rates within each carbon source by classifying growth into three categories: slow (growth rate in the lower quartile), intermediate, and fast (growth rate in the upper quartile). We found a statistically significant interaction between carbon classification and growth rate ($P < 2.2 \times 10^{-16}$); specialists were more often slow growers (38%, 146/381 growth rates) than fast growers (15%, 54/381), whereas generalists were more often fast

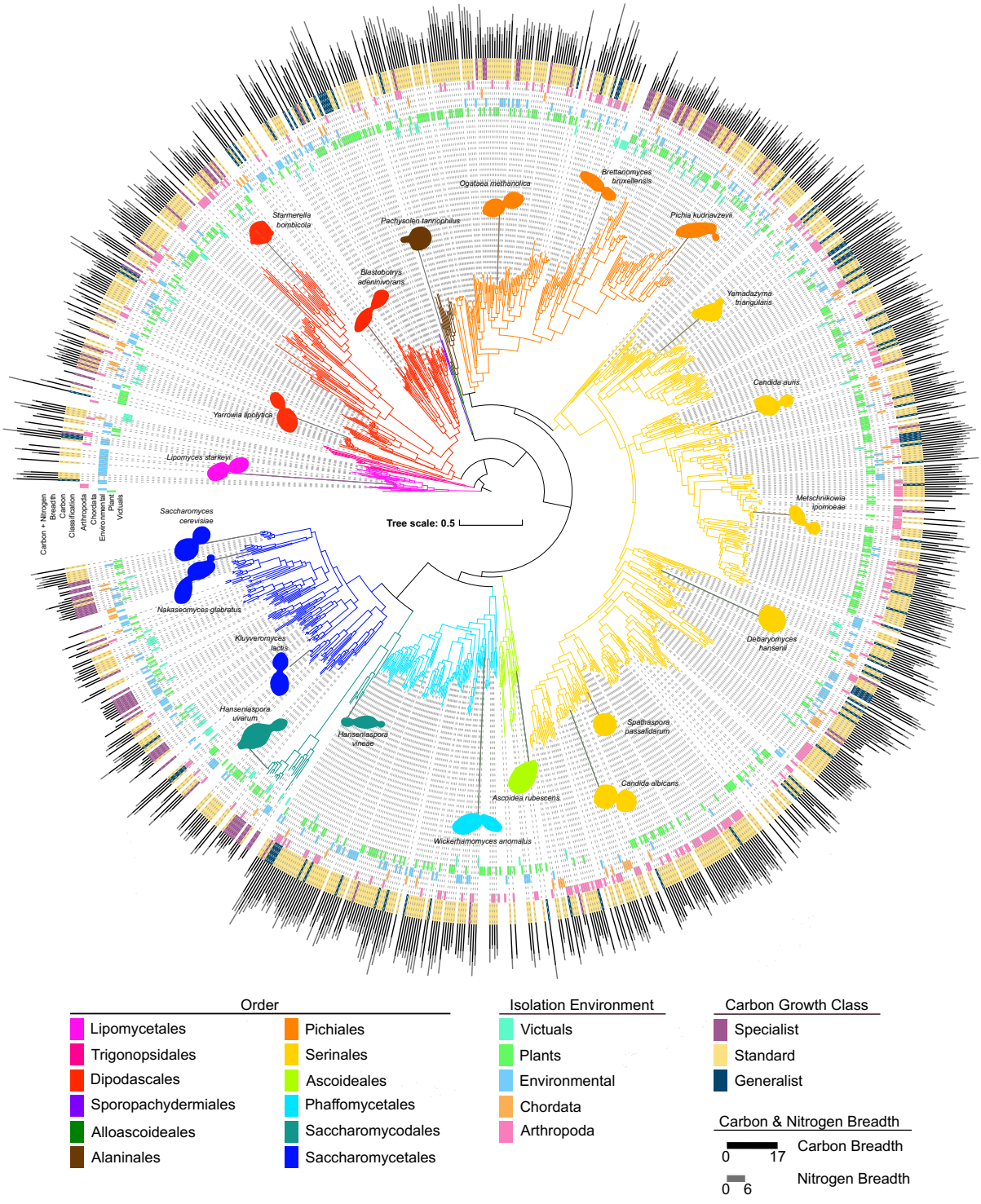


Fig. 2. Yeast traits are widely distributed across the phylogeny. The phylogeny of 1154 yeasts and fungal outgroups built from 1403 OGs of genes. Branches are colored according to their taxonomic assignment to an order of Saccharomycotina (*41*). The innermost rings are colored by the top-level type of isolation environment in which each specific strain was isolated. The purple, yellow, and blue rings identify the carbon growth classification for each strain. This classification is based on the carbon niche breadth, which is represented by the bar graph on the exterior of the tree, along with nitrogen breadth. All traits illustrated (isolation environment, carbon growth class, nitrogen breadth, and carbon niche breadth) are widely distributed across the tree; no order has one trait exclusively.

growers (33%, 403/1222 growth rates) than slow growers (20%, 238/1222 growth rates) (Fig. 3B). Moreover, there were fewer specialists than generalists in the fast category across all tested carbon sources (data S5). We also examined linear phylogenetically corrected correlations between growth rates and carbon niche breadth. We found that growth rates on five carbon sources were positively correlated with carbon niche breadth when accounting for phylogeny and multiple-testing correction (glucose $P = 0.0028$, mannose $P = 0.0056$,

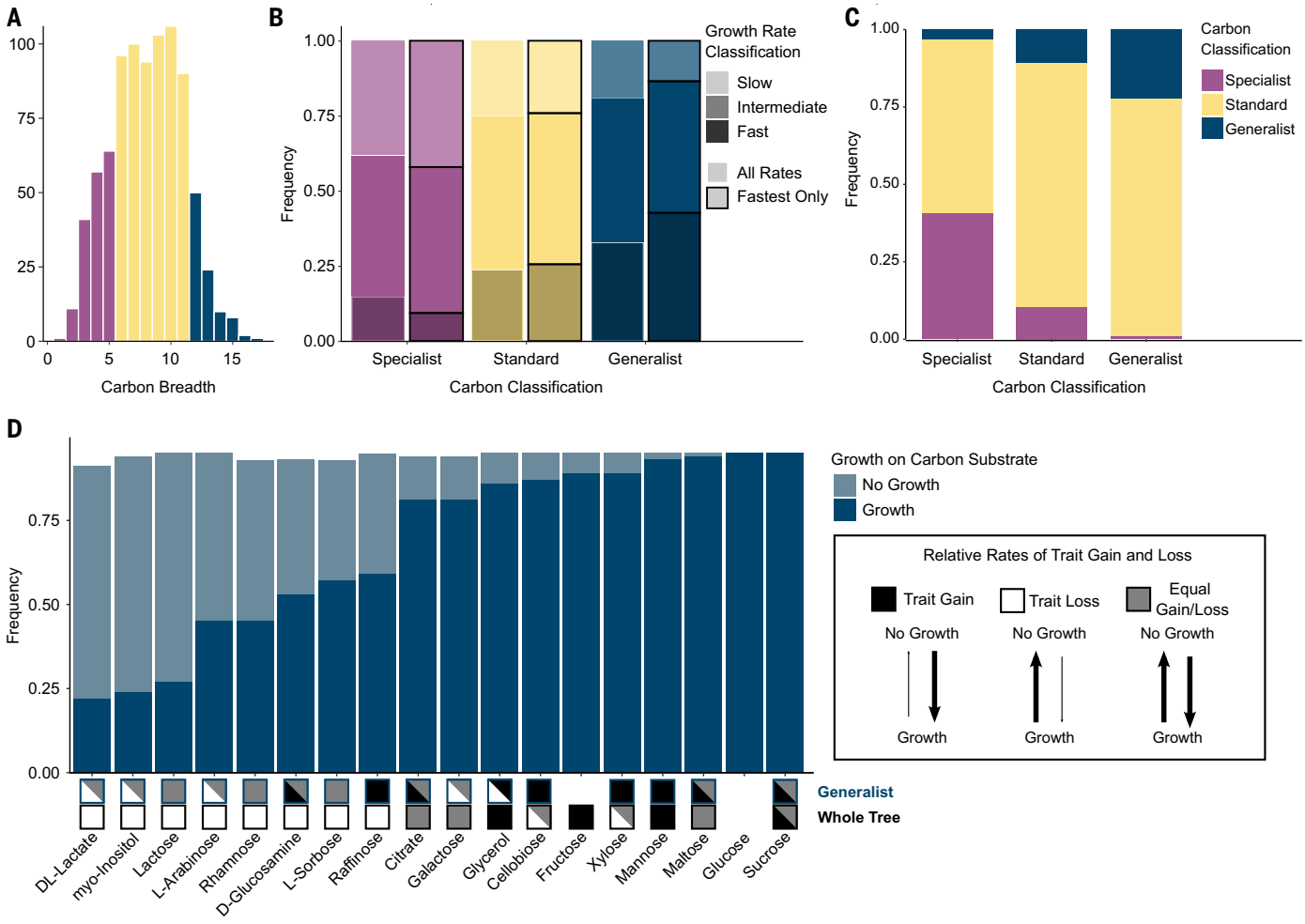


Fig. 3. Carbon specialists and generalists differ in nitrogen breadth, growth rate, and evolutionary history. (**A**) Histogram of carbon niche breadth across yeasts (n = 853). The colors of the bars represent the ranges for the different carbon classifications. Metabolic classifications were determined by permuting the binary carbon growth matrix (n = 1000 permutations). To determine the metabolic strategy of a yeast, we calculated the observed and expected (permuted) breadth for each yeast and calculated the binomial confidence intervals to determine significant differences in breadth. Generalists had a significantly larger carbon niche breadth than expected by chance, and specialists had a significantly smaller carbon niche breadth. If a yeast was not classified as either a generalist or a specialist, it was classified as standard. (**B**) The growth rates for each yeast on each of the 18 carbon sources were categorized as slow (bottom 25%), intermediate (median 50%), or fast (top 25%) using either all the rates per yeast (white outline) or only the highest rate per yeast (black outline). Carbon generalists had the highest proportion of fast growth rates (33% of all rates, 43% of fastest rates), whereas specialists had the smallest proportion (15% of all rates, 9% of fastest rates) The inverse was also true, with carbon generalists having the smallest proportion of slow growth rates (19% of all rates, 14% of fastest rates) and carbon specialists having the highest proportion of slow growth rates (38% of all rates, 42% of fastest rates). (**C**) Stacked bar graph of carbon metabolic strategies within each nitrogen metabolic strategy. (**D**) Carbon generalists shared many of the same growth traits: 10 of 18 growth traits were found in more than 75% of generalists. Many of the carbon sources had different evolutionary trends in a generalist background as compared with across the whole tree. Three different evolutionary models are shown: trait gain (black), trait loss (white), and equal rates of trait gain and loss (gray). No box indicates that the trait was not coevolving with background or across the tree. More than one evolutionary model is shown in cases where the reverse jump model spent 75% or less of the time on a single model. For example, the model testing correlated evolution between growth on D-glucosamine and generalist carbon classification reported a model string with a greater rate of gain in 55% of the run and a model string with equal rates of gain and loss in 29% of the run; therefore, we reported both the trait gain and equal gain-loss model in the generalist analysis.

myo-inositol P = 0.0083, galactose P = 0.0024, and fructose P = 0.0111; all slopes between 0.001 and 0.002) (table S1 and fig. S7A). No significant negative correlations were identified, which would have indicated that specialists were faster growers.

Second, we repeated these analyses using only the fastest growth rate for each yeast because specialists might outperform other yeasts only in the environment in which they are specialized. We found that the proportion of fast-growing specialists was 9% (10/107), a decrease from the 15% of fast-growing specialists found when we compared all growth rates across all substrates, whereas the proportion of fast-growing generalists was 43% (38/89), an increase from 33% (Fig. 3B). Thus, the strong interaction between carbon classification and growth rates persisted when only the fastest rates were considered ($P = 7.8 \times 10^{-11}$). In this case, carbon niche breadth was significantly and positively correlated with growth rates on glucose (P = 0.0002, slope = 0.002), sucrose (P = 0.0032, slope = 0.001), and fructose (P = 0.0062, slope = 0.001) after accounting for multiple testing and phylogeny (table S1 and fig. S7B).

A third analysis using the fastest growth rate for each specialist compared with all other growth rates yielded similar results (table S1 and fig. S7C). In this analysis, the growth rate for a carbon source included only specialists whose growth rate was highest on that carbon source and any growth rates for standard and generalist yeasts. Moreover, specialists were not the fastest-growing yeast in any of the carbon sources tested, including glucose. Our findings suggest that generalists grow faster on more substrates compared with specialists, including under conditions preferred by specialists.

We next tested whether there was a tradeoff between carbon and nitrogen breadth. We found significantly fewer carbon generalists that were also nitrogen specialists (n = 1) and carbon specialists that were also nitrogen generalists (n = 2) than expected by chance (P = 3.26×10^{-14}) (Fig. 3C). Moreover, trait-trait coevolutionary analysis found that carbon generalists tended to also be nitrogen generalists (Bayes factor > 2). Furthermore, our analyses of coevolution between carbon and nitrogen generalism showed that nitrogen generalism arises almost exclusively in a genetic background of carbon generalism (i.e., in carbon generalism lineages; table S2). In other words, carbon generalism mainly arises before and may facilitate nitrogen generalism. Additionally, phylogenetic regression analysis showed a strong positive correlation between carbon and nitrogen niche breadth (reported P = 0.000, slope of correlation = 0.92; table S2). These results suggest that there is an evolutionarily conserved functional connection between carbon and nitrogen metabolism in yeasts. Consistent with our finding, it is well known that certain amino acids can serve as both a carbon and nitrogen source and, as such, are dually regulated by both carbon and nitrogen signaling systems (*50*, *51*). Additionally, many metabolic pathways are known to be controlled by signals from other compounds or nutrients. In bacteria, nitrogen, sulfur, phosphorus, and iron metabolism can even be controlled by carbon metabolism (*50*, *52*).

Our previous analysis of 332 yeasts identified a pervasive pattern of trait loss (*28*), which suggests that generalists have either retained carbon-acquisition traits over long evolutionary timescales or gained traits, unlike their nongeneralist relatives. To test these hypotheses, we compared the relative rates of carbon trait gain or loss, either across all yeasts or specifically within generalist lineages, while taking phylogeny into account (Fig. 3D and table S3). For the eight carbon traits found in less than 75% of generalists, we identified a strong trend of trait loss across the entire phylogeny but some evidence of trait gain in the generalist background. Therefore, carbon generalists appear to have both gained and retained carbon traits that were otherwise lost broadly across the rest of the subphylum.

Intrinsic factors shape carbon niche breadth variation in yeasts

Given the extreme carbon niche breadths of generalists and specialists, we next tested whether these two groups have independent factors favoring generalist and specialist phenotypes. Extrinsic factors, such as carbon availability in an isolation environment, could shape variation in metabolic niche breadth. Similar environments, which are likely to share extrinsic factors, may favor the evolution of generalists or specialists. To explore the possibility that some environments contain extrinsic factors that shape carbon niche breadth, we identified the precise isolation environment for each possible yeast strain (1088 total). We then grouped strains by similar environments using a formal hierarchical ontology of isolation environments. This ontology contained 1597 classes (specific environments) (fig. S8 and data S6). Environment classifications at the highest level of our ontology generally contained similar numbers of generalists and specialists: Arthropoda (24 generalists and 16 specialists), Chordata (7 and 8), plants (25 and 31), and food or drink (5 and 16). Furthermore, generalists and specialists shared environments. For example, *Hyphopichia homilentoma* (generalist) and *Wickerhamomyces sydowiorum* (specialist) were both isolated from tunnels of the woodboring beetle *Sinoxylon ruficorne* in the red bushwillow *Combretum apiculatus*. Given the limited number of generalists and specialists within an environment and the fact that we only had a single environment per strain, we were unable to rigorously test for extrinsic factors that favor generalists or specialists. We anticipate that incorporation of improved characterizations of yeast habitats and the addition of isolation environment data into our formal ontology will enable future investigations of the environmental factors shaping carbon niche breadth evolution.

We next hypothesized that the genomes of generalists may contain a larger number of metabolic genes, which are intrinsic factors, compared with those of specialists. We found that both the total number of genes and the number of KEGG ortholog groups (KOs) were both positively and significantly associated with carbon niche breadth (Fig. 4A and fig. S10, A and B). Notably, we found that, for every additional carbon source a yeast could metabolize, its genome contained, on average, an additional 36 genes and 2 KOs.

Metabolic networks, including the carbon metabolism network, are more complex than just the total number of genes because they are highly interconnected owing to shared enzymes and pathways. To examine whether metabolic network structure varied between generalists and specialists, we used KOs to build metabolic networks for all yeasts and tested for a correlation between carbon niche breadth and six common network properties that reflect biological complexity (Fig. 4B; fig. S10, C to F; and data S7) (*53*, *54*). Relative to carbon specialists, carbon generalists had a higher edge count, or more connections between nodes of the network (Fig. 4B) (*55*). Both carbon generalists and specialists had disassortative networks, or networks with high levels of connection between nodes with dissimilar properties—a property of all biological networks (*56*). However, relative to specialists, the generalist networks were less disassortative, or had more highly interconnected nodes (Fig. 4B). There were no significant correlations between carbon niche breadth and the other network properties (fig. S10, C to F). Despite the extreme difference in carbon metabolism capabilities, carbon generalists and specialists had only slight differences in the size and shape of their global KEGG metabolic networks. These results suggest that generalist and specialist networks are overall similar in size and shape but differ in how they are wired.

We next investigated differences in the composition of generalist and specialist networks. Generalists and specialists largely showed similar compositions across KOs, but a small set of KOs was depleted (presence < 20%) in specialists and enriched (presence > 85%) in generalists (table S4). Generalist-enriched KOs were related to nitrogen, fructose, mannose, and galactose metabolisms. Enrichment of these terms suggests that differences in gene content contribute to the overall carbon metabolism trait differences observed between generalists and specialists.

Unifying genetic features of carbon niche breadth generalists

To gain further insight into the genes and pathways contributing to the observed carbon niche breadth variation across the yeast subphylum, we used machine learning. Specifically, we trained a supervised random forest classifier to use KO presence and absence as predictive features for carbon niche breadth classification. Niche breadth classification of generalists and specialists was used instead of the actual number of carbon sources because there were insufficient numbers of yeasts for some values to adequately train our model (e.g., there was only one yeast that grew on 17/18 carbon sources, but there were 64 yeasts that grew on five carbon sources). The resulting classifier was both highly sensitive and specific, correctly classifying 88% of specialists and 89% of generalists [area under the curve (AUC) = 0.93] (Fig. 4C). The high accuracy suggests that generalist and specialist KEGG networks differ in ways that were not detected in the KO enrichment analysis.

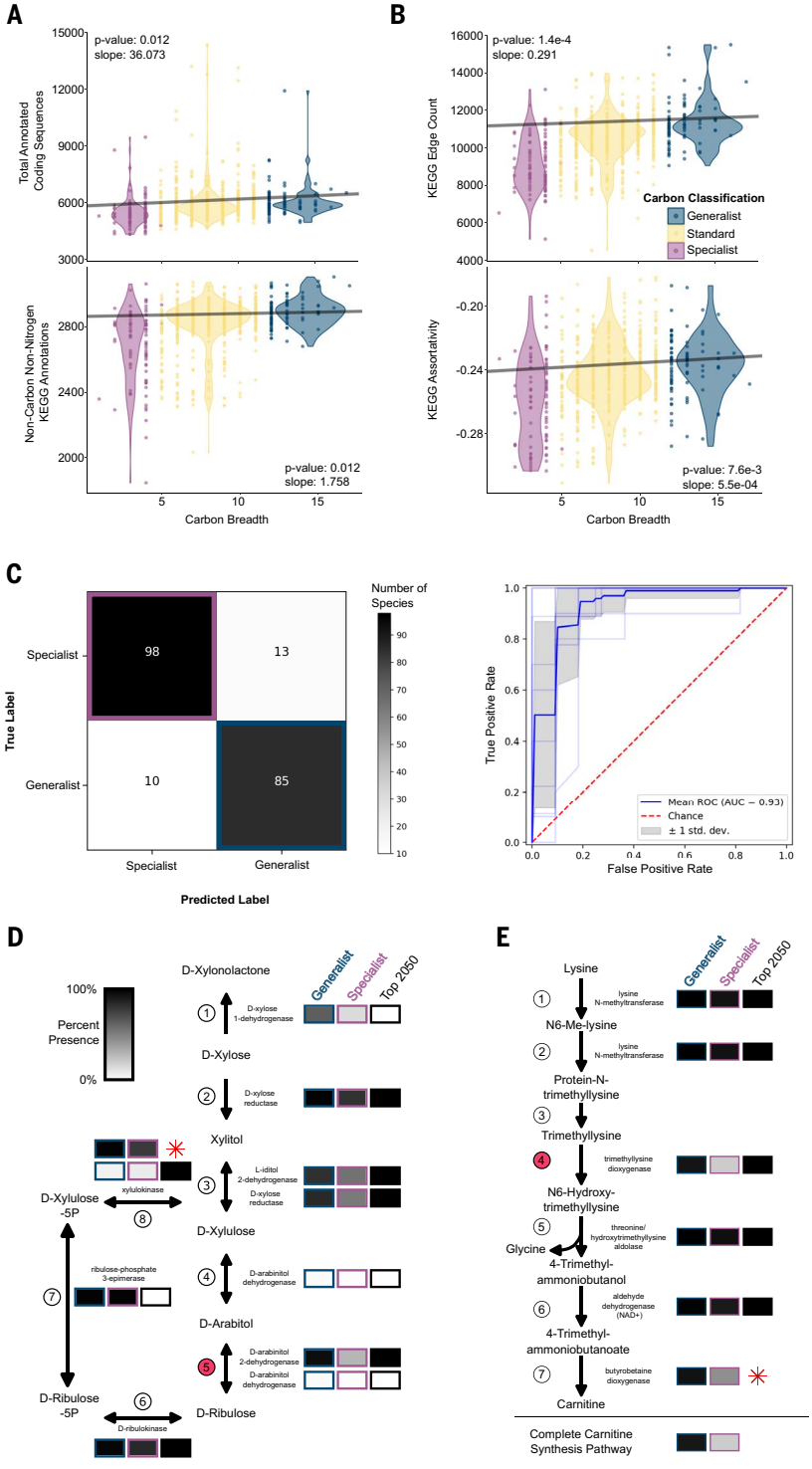


Fig. 4. Generalist and specialist metabolism differs in expected and unexpected ways. (**A**) Total annotated coding sequences (top) and total number of annotated KOs (bottom) were both positively and significantly correlated with carbon niche breadth using a phylogenetic generalized least squares (PGLS) analysis. One outlier with a predicted number of coding sequences is not visualized but was included in the analysis (*M. magnusii*; number of protein-coding genes = 20,704; carbon niche breadth = 9). (**B**) Two KEGG network statistics were significantly and positively correlated with carbon niche breadth when taking into account phylogenetic relatedness (PGLS). KEGG edge count (top) and KEGG assortativity (bottom) were both elevated in carbon generalists. (**C**) Yeasts were classified into generalists and specialists using a machine learning algorithm trained on the KOs. The correct classification occurred in 88% of specialists and 89% of generalists. The receiver operating characteristic (ROC) analysis suggests that both the sensitivity and specificity of our model are excellent (AUC = 0.93). (**D**) Multiple reactions in the pentose and glucuronate interconversions pathway were important in classifying yeasts into generalists and specialists as determined by the leave-out analysis, which identified 2050 informative KOs (black boxes). Boxes are shaded as the percent of each carbon classification with at least one enzyme in that step of the reaction. The reaction with the third-highest relative importance in the machine learning analysis is shown in step 5 and is facilitated by D-arabinitol 2-dehydrogenase. Notably, experimental studies suggest that yeast D-arabinitol 2-dehydrogenase is also capable of completing the reaction in step 4 (*91*). Step 8 was among the top features used in the machine learning analysis, despite the fact that KEGG only partially annotated this gene. The xylulokinase encoded by yeast *XYL3* has been well studied (*58*). Therefore, we reannotated the *XYL3* gene and have shown its relative abundance (red star). (**E**) The carnitine biosynthesis pathway includes multiple reactions that are important for classifying carbon generalists and specialists. The reaction in step 4 had the fourth-highest relative importance in the machine learning classification of carbon classification. Step 7 was not annotated by KEGG in any of our yeasts, but this step had been previously characterized in *C. albicans* as being facilitated by the trimethyllysine dioxygenase enzyme encoded by *BBH2* (*64*). We reannotated *BBH2* using this reference sequence and calculated the relative abundance in each carbon classification (red star). Finally, we determined the number of yeasts that could hypothetically complete the lysine-to-carnitine biosynthesis pathway.

Examination of the features on which the classifier relied using dropout analysis identified 2050 KOs that significantly contributed to classification accuracy. Approximately 5000 individual yeast KOs were used to train the algorithm, which suggests that many KOs contributed some information to niche breadth classification. We further examined the top four features because the fifth feature had only half the relative importance score of each of the top four. Two of the top four features had direct links to the catabolism of specific carbon substrates, demonstrating the power and precision of our algorithm. The KO for *manB* (K01192), which encodes a β-mannosidase, had the second-highest relative importance (relative importance, 0.048). This KO was identified in 7% of specialists (8/111) and 80% of generalists (76/95). β-mannosidases are known to have a role in microbial utilization of *N*-glycans as a carbon source (*57*). Almost all the carbon generalists (93/95) can use mannose, which leads to the hypothesis that generalists likely use the mannose moieties present in *N*-glycans as a carbon and energy source.

The KO with the third-highest importance was K17738 (relative importance, 0.043),

which is the *ARD* gene encoding D-arabinitol 2-dehydrogenase, an important component of the pentose and glucuronate interconversions pathway (Fig. 4D, step 5). This KO was more frequently present in the genomes of generalists (96%, 91/95) than in the genomes of specialists (71%, 79/111). In a portion of this pathway, five of the eight reactions were among the 2050 KOs (with two falling in the top 100 KOs) that contributed to the classification of carbon generalists and specialists (Fig. 4D, black boxes). Notably, growth on xylose was included in our carbon classification, and the xylose metabolism genes *XYL1* (Fig. 4D, step 2), *XYL2* (step 3), and *XYL3* (step 8) were all identified as important features (with *XYL1* falling within the top 100), which suggests that xylose metabolism genes may be promiscuous and have multiple metabolic capabilities (*58*). This result also supports the hypothesis that intrinsic genetic factors contribute to niche breadth by connecting pathways.

The feature with the highest relative importance was K03940 (relative importance, 0.062), which encodes an NADH (the reduced form of nicotinamide adenine dinucleotide) ubiquinone oxidoreductase core subunit (NDUFS7 in humans) of complex I of the mitochondrial electron transport chain. This KO was identified in 29% of specialists (32/111) and 95% of generalists (90/95). Complex I is known to vary widely in presence and makeup, including the presence of an alternative pathway in some yeasts (*59*). For example, in *S. cerevisiae*, the NADH oxidoreductase function of complex I is conducted by three single-subunit enzymes (Ndi1p, Nde1p, or Nde2p) (*60*). Conversely, in *Y. lipolytica*, complex I is composed of 42 subunits, including the NADH ubiquinone oxidoreductase NUKM (K03940) (*61*). Thirty additional complex I enzymes were within the top 2050 KOs, and two fell within the top 10%—K03941 and K03966, which are both NADH ubiquinone oxidoreductases in the β subcomplex (KEGG map00190). The Saccharomycetales and Saccharomycodales have both completely lost the canonical complex I and contain many specialist yeasts (*59*). The relatively high importance of K03940, however, is not solely due to these orders because the effect is widespread. For example, within the Pichiales, 100% (5/5) of generalist genomes encode K03940, in contrast to only 18% (6/33) of specialists. Complex I has been implicated in *C. albicans* growth and virulence (*62*), as a global regulator of fungal secondary metabolism in *Aspergillus* (*63*), and results in a higher proton motive force compared with the alternative pathway in *S. cerevisiae*. The presence of complex I in generalists, therefore, may support increased carbon niche breadth and elevated growth rates.

The last KO that we investigated was K00474 (relative importance, 0.043), which encodes a trimethyllysine dioxygenase involved in lysine degradation. Every step in the pathway that degrades lysine to carnitine, except the last step, was identified as important in the machine learning classification. The last step (Fig. 4E, step 7) was not annotated by KEGG in any of our yeasts. Therefore, we annotated the *BBH2* gene, which encodes the trimethyllysine dioxygenase, directly from our predicted coding sequences using previously published reference sequences (*64*). After manual annotation of *BBH2*, we found that most carbon generalists were predicted to be able to complete the carnitine biosynthesis pathway (91%, 86/95), whereas relatively few carbon specialists were predicted to do so (20%, 22/111). Carnitine plays an important role in the transport of acetyl coenzyme A (acetyl-CoA), which in turn is a major metabolite that contributes to many metabolic pathways, including the production of adenosine 5′-triphosphate (ATP) in the mitochondrial tricarboxylic acid (TCA) cycle. Acetyl-CoA can be produced within the mitochondria when glucose is available, or, when glucose is unavailable, it can be transported into the mitochondria using the carnitine shuttle (*65*). Some yeasts, including *C. albicans*, rely solely on the carnitine shuttle for this transport (*64*), whereas other yeasts, such as *S. cerevisiae*, can use a carnitine-independent method for acetyl-CoA transport (*66*). Similarly, some yeasts, such as *C. albicans*, can synthesize carnitine; others, such as *S. cerevisiae*, cannot and rely on exogenous sources. A complete carnitine synthesis pathway may ensure acetyl-CoA transport when glucose is unavailable, especially in species that rely solely on the carnitine shuttle.

Additionally, carnitine and carnitine acetyltransferases can be essential for growth on some nonfermentable carbon sources. These include ethanol as well as glycerol in certain *S. cerevisiae* mutants with disrupted citrate metabolism (*67*). We found that 90.5% (86/95) of generalists can grow on glycerol compared with only 24.5% (27/110) of specialists (table S2). Moreover, specialists that could grow on glycerol were more likely to have the complete carnitine synthesis pathway compared with those that did not (*z*-test, $\chi^2 = 10.425$, $P = 0.0186$). These results suggest that carnitine production affords metabolic flexibility and carbon niche breadth.

Human yeast pathogens include both carbon generalists and specialists

This comprehensive dataset and analytical framework provide the opportunity to study how the observed genomic, metabolic, and environmental variation across the subphylum is associated with any complex trait of interest (*68*–*70*). To illustrate this potential, we examined the metabolic niche breadths of yeast pathogens of humans compared with those of their nonpathogenic close relatives (using a specific phylogenetic distance cutoff to standardize the clades) (Fig. 5). The World Health Organization (WHO) recently released its first-ever fungal priority pathogens list, which included six Saccharomycotina species (*71*). We defined 11 yeasts as opportunistic human pathogens because they are known to cause human infections and generally require biosafety level 2 (BSL-2) precautions in research laboratories.

Carbon sources and availability vary in vivo in humans, which suggests that carbon niche breadth may play an important role in promoting or preventing fungal pathogenesis (*72*). Yeasts are subject to diverse microenvironments characterized by varying nutrients within a host (*39*, *72*, *73*). Their capacity to survive under fluctuating carbon conditions has been closely associated with virulence. For example, lactate assimilation across the *C. albicans* clade, and in *Nakaseomyces glabratus* (syn. *Candida glabrata*), is associated with increased antifungal and osmotic stress resistance and has been shown to reduce phagocytosis within the host (*73*). Notably, these pathogens exhibit reduced resistance to the antifungal drug amphotericin B when grown in culture media containing lactate relative to culture media containing glucose (*73*). We found that pathogens spanned the range of carbon niche breadth classifications and included specialist, standard, and generalist yeasts. Carbon niche breadths within pathogenic yeasts ranged from 15 in *Meyerozyma guilliermondii* to only 2 in *N. glabratus* (*74*). Furthermore, the proportion of pathogenic yeasts classified as standard, generalist, and specialist was similar to that of their nonpathogenic relatives (Fig. 5, A and B). Collectively, these results suggest that yeast pathogenicity is not associated with carbon niche breadth.

Previous work in *C. albicans* linked its pathogenicity to its high growth rate (*75*). To examine whether this link holds across yeast pathogens, we visualized all pathogenic yeasts and their relatives on a phylogenetically corrected principal components analysis using all our growth rate data (Fig. 5C). We observed no clustering of pathogenic yeasts using carbon growth rates. Moreover, yeast pathogens within the same clade varied in their growth rate on glucose by almost threefold: *Candida parapsilosis* had a growth rate of 0.042, whereas *Candida tropicalis* had a growth rate of 0.124. Our growth rate data, however, were collected at a specific temperature in defined media and may not reflect growth rates in human infections.

We also examined the roles of temperature, gene content, and environment in yeast pathogenicity. One feature known to be necessary, but insufficient, for pathogenicity is growth at human body temperature, or 37°C (Fig. 5D) (*39*). We observed that relatives of human

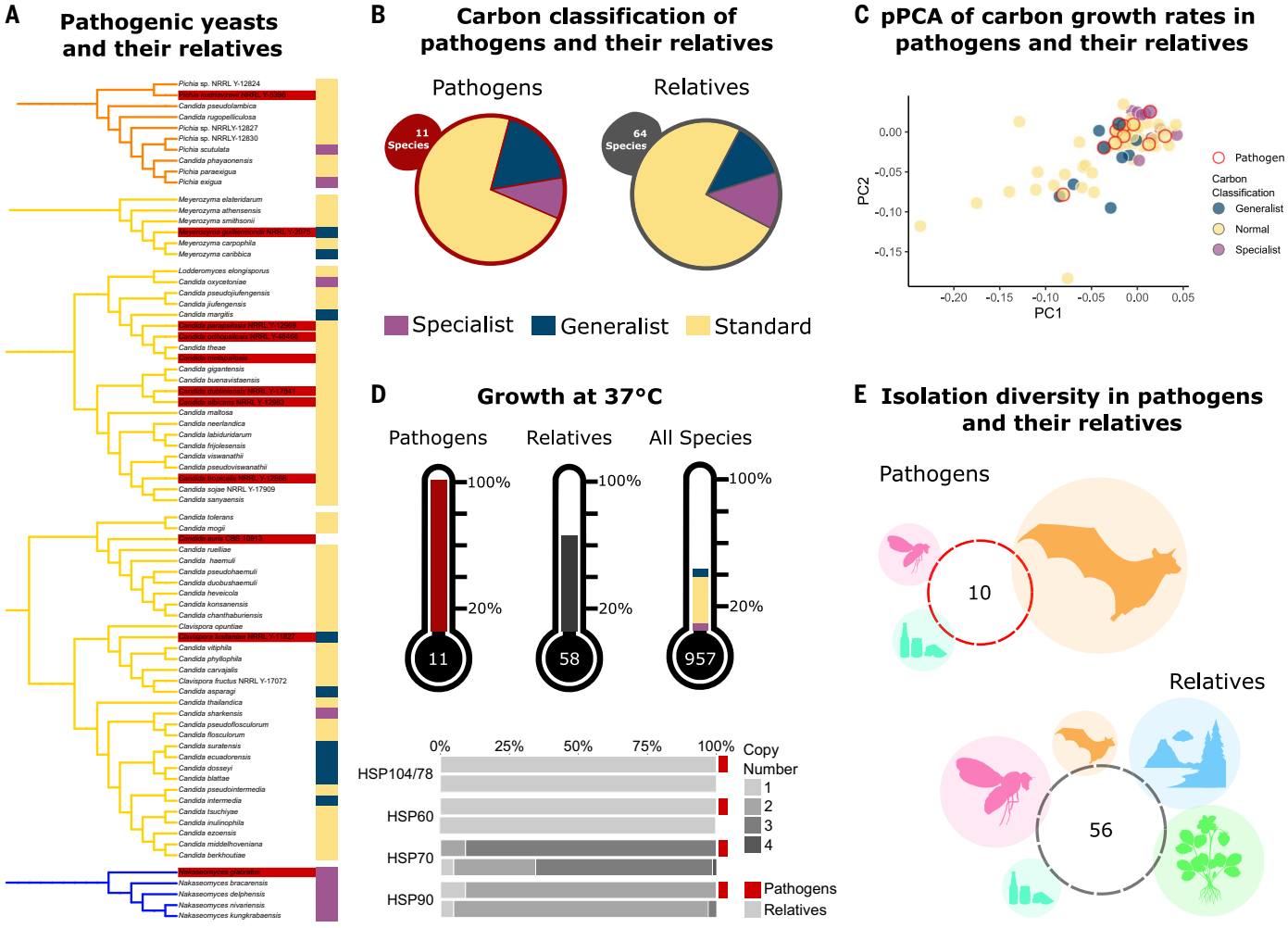


Fig. 5. Carbon generalism and specialism are not associated with yeast pathogenicity. (**A**) The phylogenetic clades containing human fungal pathogens. Clades reflect all species within a specific phylogenetic distance from the identified pathogen. Pathogens are found in three different orders, and at least one pathogen is classified in the generalist, specialist, and standard categories. (**B**) Pathogens and their relatives had nearly identical proportions of generalist, specialist, and standard yeasts. This result suggests that carbon niche breadth is not a defining or predictive factor for the potential of a species to gain the ability to infect humans. (**C**) Pathogens and their relatives did not differ substantially in their growth rates on carbon substrates. The phylogenetically corrected principal components analysis (pPCA) was constructed using growth rates on carbon substrates and projected onto the first two components (totaling 80% of the total variance). Pathogens did not cluster together, and generalists and specialists appeared further apart. This result suggests that pathogens do not have shared growth rate characteristics. (**D**) Proportion of yeasts that can grow at 37°C in pathogens, their relatives, and all sampled yeasts. All yeasts identified as pathogens can grow at 37°C. Pathogenic yeasts were significantly more likely to grow at 37°C compared with their nonpathogenic relatives (χ^2, $P = 0.042$). Heat shock protein (*HSP*) gene copy number was determined using InterPro and KEGG orthologs. *HSP* gene copy number was not significantly associated with pathogenicity. (**E**) Isolation environment for the specific strains of pathogens and their relatives. Circles are proportional to the percent of yeasts isolated from Chordata (orange), Arthropoda (pink), victuals (teal), environmental (blue), and plants (green).

pathogens had an elevated frequency of growth at 37°C (~64%) compared with all yeasts for which growth at this temperature was measured (~41%). This result likely reflects the necessity of growth at 37°C to evolve before pathogenicity. Heat shock proteins (HSPs) are also known to affect temperature tolerance (*76*). Examination of copy number variation in the genes encoding HSPs in the pathogenic species and their relatives identified a slight increase in *HSP70* gene copy number among pathogenic yeasts (Fig. 5D). Finally, we found that pathogenic yeasts and their relatives had been isolated from all examined environments (Fig. 5E). Our analyses suggest that pathogenicity can emerge in species across the spectrum of carbon metabolic breadth. Moreover, the lack of notable differences between yeast pathogens and their nonpathogenic relatives supports the hypothesis that the traits and genetic elements contributing to pathogenicity are not broadly shared across pathogens but are specific to each (*77*). The data and analyses presented in this work provide a model for the investigation of other complex traits across Saccharomycotina using our ensemble of genomic, metabolic, and environmental data.

Conclusions

We focused on two predominant paradigms proposed to underlie the evolution of yeast carbon niche breadth. The first paradigm, where trade-offs dominate, was not supported when we analyzed more than 10,000 growth rates measured across 853 yeasts. We found that generalists typically grew faster on carbon

sources compared with specialists, even on those carbon sources for which specialists had their maximum growth rates. Thus, the ability to metabolize additional carbon sources does not come at the cost of reduced growth rates on other carbon sources. Carbon metabolism traits found within generalists were either maintained across evolutionary time or gained, even though there was a strong overall trend for trait loss across the subphylum. Of course, trade-offs between carbon metabolism traits likely exist in natural habitats. Future experiments along gradients of different environmental conditions, such as temperature, competition, or oxygen availability, may shed additional light on condition-specific trade-offs in carbon niche breadth evolution.

By contrast, we found strong support for the second paradigm in the form of intrinsic factors that underlie the generalist phenotype. Machine learning allowed us to identify specific genes, complexes, and pathways shared by generalists but largely absent from specialists. These genes were directly involved in carbon and energy metabolism, often by enhancing metabolic flexibility and robustness. This finding supports the second paradigm because we identify a shared set of intrinsic genomic features across the generalist phenotype, even though generalists vary in the specific carbon sources that they can metabolize. This finding does not support the hypothesis of trade-offs for two reasons. First, the pathways enriched in generalists are hypothesized to increase metabolic efficiency, which is contrary to the proposed trade-off between carbon niche breadth and efficiency. Second, under the trade-off paradigm, specialists and generalists would both have specific traits that provide them with a selective advantage. However, we found that generalists, as compared with specialists, have more genes in their genomes, including those not directly associated with carbon metabolism.

Given the advantages of wide carbon niche breadth and the absence of detectable efficiency costs, the question remains: What forces are shaping specialist yeasts? In some cases, carbon specialism could be associated with rapid gene loss. For example, in the genus *Hanseniaspora* (10/14 or 71.4% of specialists), there were widespread gene losses, including of genes involved in DNA repair and carbon metabolism (*78*). Another hypothesis is that each specialist is subject to specific evolutionary pressures that would obviate unifying features. Finally, it is also possible that there are growth-associated trade-offs that we are unable to measure. Features, such as enhanced carbon sequestration, killer yeast toxins, pathogenicity, and microbial community composition, could provide specialists with advantages in highly specific environments. For example, *Hanseniaspora* species have a growth advantage over other species, including *S. cerevisiae*, on grapes at harvest and in the early stages of alcoholic fermentation (*79*). Further investigations into the evolution of yeast generalism and specialism will likely be fruitful, but a plethora of additional questions could be addressed with these data, including quantifying correlations among genes, traits, and/or ecologies; investigations of gene family evolution; research into the origins of pathogenesis; and genome-informed bioprospecting of yeasts and their genes for the sustainable production of cellulosic biofuels and bioproducts. More broadly, by coupling a comprehensive dataset with a robust analytical framework for studying macroevolutionary processes, the Y1000+ Project provides a roadmap that connects DNA to diversity.

Materials and methods summary

Detailed materials and methods can be found in the supplementary materials (*80*). All data generated as a part of the project have been deposited in a Figshare repository (*42*).

Genome sequencing, annotation, and phylogenomics

Strains were obtained primarily from the NRRL (USA) and CBS (Netherlands) culture collections. We sequenced pair-end libraries using the Illumina HiSeq 2500 platform and assembled genomes using the meta-assembler pipeline iWGS (*81*). We assessed assembly quality using Benchmarking Universal Single-Copy Orthologs (BUSCO) (*44*) and filtered the assemblies to remove mitochondrial and bacterial DNA contaminants. Genomes were functionally annotated using KEGG (*55*) and InterPro (*82*, *83*) databases. We constructed a phylogenomic data matrix from 1403 OGs (taxon occupancy for each group ≥ 50%; 719,591 amino acid sites); we inferred the phylogeny of the subphylum using both concatenation and coalescence under maximum likelihood using IQ-Tree (*84*) and ASTRAL-III (*85*), respectively, and estimated the yeast time tree using the RelTime method (*86*).

Phenotyping, niche breadth classification, and testing for trade-offs and trait coevolution

We generated quantitative growth data on 18 carbon and 6 nitrogen sources for 853 yeasts, measuring optical density every 2 hours for a week on the BMG Omega SpectroStar Plate Reader. We conducted all experiments in triplicate, and a new yeast colony was picked for each yeast across replicates. We calculated growth rates using a logistic model using the R package *grofit* (*87*). We classified yeasts as specialist, standard, or generalist for both carbon and nitrogen metabolism by calculating the binomial confidence intervals of carbon and nitrogen breadth relative to randomized growth data. We measured the correlation between carbon and nitrogen breadth and tested for trade-offs between carbon niche breadth and efficiency (by measuring the correlation between growth rates and carbon niche breadth classifications) using phylogenetic generalized least squares analyses with PGLScaper (*88*). Finally, we inferred the coevolution of carbon traits and carbon generalism or specialism using BayesTraits (http://www.evolution.reading.ac.uk).

Underlying factors driving generalist and specialist phenotypes

We identified strain-specific isolation environments for 1088 yeasts and standardized them by creating an ontology of environments and their hierarchical network using Web Protégé (https://github.com/protegeproject/webprotege). To identify underlying genomic features contributing to generalist and specialist phenotypes, we used genome annotations to build metabolic networks and quantify network variation among generalists and specialists while accounting for phylogeny. We also identified KEGG ontologies enriched in generalists and specialists using a KEGG enrichment analysis (*89*). Finally, we constructed a machine learning algorithm using the XGBoost random forest classifier (*90*), which we trained using 90% of the genomic data and using the remaining 10% for cross validation, to identify genes whose presence or absence was most strongly associated with carbon generalism and specialism.

REFERENCES AND NOTES

1. E. L. Bruns, J. Antonovics, M. E. Hood, From generalist to specialists: Variation in the host range and performance of anther-smut pathogens on *Dianthus*. *Evolution* **75**, 2494–2508 (2021). doi: 10.1111/evo.14264; pmid: 33983636
2. D. B. Preston, S. G. Johnson, Generalist grasshoppers from thermally variable sites do not have higher thermal tolerance than grasshoppers from thermally stable sites - A study of five populations. *J. Therm. Biol.* **88**, 102527 (2020). doi: 10.1016/j.jtherbio.2020.102527; pmid: 32126002
3. J. W. Wenger *et al*., Hunger artists: Yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLOS Genet.* **7**, e1002202 (2011). doi: 10.1371/journal.pgen.1002202; pmid: 21829391
4. R. H. MacArthur, *Geographical Ecology: Patterns in the Distribution of Species* (Princeton Univ. Press, 1984).
5. D. S. Wilson, J. Yoshimura, On the coexistence of specialists and generalists. *Am. Nat.* **144**, 692–707 (1994). doi: 10.1086/285702
6. A. R. Burmeister *et al*., Pleiotropy complicates a trade-off between phage resistance and antibiotic resistance. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11207–11216 (2020). doi: 10.1073/pnas.1919888117; pmid: 32424102
7. Z. Luo *et al*., Biogeographic patterns and assembly mechanisms of bacterial communities differ between habitat generalists and specialists across elevational gradients. *Front. Microbiol.* **10**, 169 (2019). doi: 10.3389/fmicb.2019.00169; pmid: 30804920
8. F. Seebacher, V. Ducret, A. G. Little, B. Adriaenssens, Generalist–specialist trade-off during thermal acclimation. *R. Soc. Open Sci.* **2**, 140251 (2015). doi: 10.1098/rsos.140251; pmid: 26064581
9. J. D. Napier *et al*., A generalist–specialist trade-off between switchgrass cytotypes impacts climate adaptation and geographic range. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2118879119 (2022). doi: 10.1073/pnas.2118879119; pmid: 35377798

10. E. A. Ostrowski, R. J. Woods, R. E. Lenski, The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. *Proc. R. Soc. B* **275**, 277–284 (2008). doi: 10.1098/rspb.2007.1244; pmid: 18029306
11. Q. Xu *et al*., Microbial generalist or specialist: Intraspecific variation and dormancy potential matter. *Mol. Ecol.* **31**, 161–173 (2022). doi: 10.1111/mec.16217; pmid: 34626522
12. R. Kassen, The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**, 173–190 (2002). doi: 10.1046/j.1420-9101.2002.00377.x
13. E. Hesse, A. Best, M. Boots, A. R. Hall, A. Buckling, Spatial heterogeneity lowers rather than increases host-parasite specialization. *J. Evol. Biol.* **28**, 1682–1690 (2015). doi: 10.1111/jeb.12689; pmid: 26135011
14. L. Büchi, S. Vuilleumier, Coexistence of specialist and generalist species is shaped by dispersal and environmental factors. *Am. Nat.* **183**, 612–624 (2014). doi: 10.1086/675756; pmid: 24739194
15. C. Bleuven, C. R. Landry, Molecular and cellular bases of adaptation to a changing environment in microorganisms. *Proc. R. Soc. B* **283**20161458 (2016). doi: 10.1098/rspb.2016.1458
16. L. Caspeta, J. Nielsen, Thermotolerant yeast strains adapted by laboratory evolution show trade-off at ancestral temperatures and preadaptation to other stresses. *mBio* **6**, e00431-15 (2015). doi: 10.1128/mBio.00431-15; pmid: 26199325
17. D. A. Opulente *et al*., Factors driving metabolic diversity in the budding yeast subphylum. *BMC Biol.* **16**, 26 (2018). doi: 10.1186/s12915-018-0498-3; pmid: 29499717
18. A. J. P. Brown, G. D. Brown, M. G. Netea, N. A. R. Gow, Metabolism impacts upon *Candida* immunogenicity and pathogenicity at multiple levels. *Trends Microbiol.* **22**, 614–622 (2014). doi: 10.1016/j.tim.2014.07.001; pmid: 25088819
19. T. Vanneste *et al*., Contrasting microclimates among hedgerows and woodlands across temperate Europe. *Agric. For. Meteorol.* **281**, 107818 (2020). doi: 10.1016/j.agrformet.2019.107818
20. A. Grzyb, A. Wolna-Maruwka, A. Niewiadomska, The significance of microbial transformation of nitrogen compounds in the light of integrated crop management. *Agronomy* **11**, 1415 (2021). doi: 10.3390/agronomy11071415
21. M. Espinosa-Urgel, J.-L. Ramos, Expression of a *Pseudomonas putida* aminotransferase involved in lysine catabolism is induced in the rhizosphere. *Appl. Environ. Microbiol.* **67**, 5219–5224 (2001). doi: 10.1128/AEM.67.11.5219-5224.2001; pmid: 11679348
22. C. Schädel, A. Blöchl, A. Richter, G. Hoch, Quantification and monosaccharide composition of hemicelluloses from different plant functional types. *Plant Physiol. Biochem.* **48**, 1–8 (2010). doi: 10.1016/j.plaphy.2009.09.008; pmid: 19926487
23. A. Ebringerová, Z. Hromádková, T. Heinze, in *Polysaccharides I: Structure, Characterisation and Use*, T. Heinze, Ed. (Springer, 2005), pp. 1–67.
24. C. H. Jaeger3rd, S. E. Lindow, W. Miller, E. Clark, M. K. Firestone, Mapping of sugar and amino acid availability in soil around roots with bacterial sensors of sucrose and tryptophan. *Appl. Environ. Microbiol.* **65**, 2685–2690 (1999). doi: 10.1128/AEM.65.6.2685-2690.1999; pmid: 10347061
25. A. M. Abdel-Mawgoud, M. M. Aboulwafa, N. A.-H. Hassouna, Characterization of surfactin produced by *Bacillus subtilis* isolate BS5. *Appl. Biochem. Biotechnol.* **150**, 289–303 (2008). doi: 10.1007/s12010-008-8153-z; pmid: 18437297
26. A. L. Demain, From natural products discovery to commercialization: A success story. *J. Ind. Microbiol. Biotechnol.* **33**, 486–495 (2006). doi: 10.1007/s10295-005-0076-x; pmid: 16402247
27. E. G. Tate, A. L. Pitt, M. D. Little, J. J. Tavano, M. A. Nickerson, Factors contributing to the range expansion and population increase of a native generalist species. *Amphib. Reptil.* **43**, 299–311 (2022). doi: 10.1163/15685381-bja10098
28. X.-X. Shen *et al*., Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175**, 1533–1545.e20 (2018). doi: 10.1016/j.cell.2018.10.023; pmid: 30415838
29. D. I. Andersson, J. Jerlström-Hultqvist, J. Näsvall, Evolution of new functions de novo and from preexisting genes. *Cold Spring Harb. Perspect. Biol.* **7**, a017996 (2015). doi: 10.1101/cshperspect.a017996; pmid: 26032716
30. K. Voordeckers *et al*., Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLOS Biol.* **10**, e1001446 (2012). doi: 10.1371/journal.pbio.1001446; pmid: 23239941
31. W. M. Patrick, E. M. Quandt, D. B. Swartzlander, I. Matsumura, Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* **24**, 2716–2722 (2007). doi: 10.1093/molbev/msm204; pmid: 17884825
32. C. S. McBride, J. R. Arguello, Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**, 1395–1416 (2007). doi: 10.1534/genetics.107.078683; pmid: 18039874
33. K. R. Christie *et al*., *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.* **32**, D311–D314 (2004). doi: 10.1093/nar/gkh033; pmid: 14681421
34. M. B. Arnaud *et al*., The *Candida* Genome Database (CGD), a community resource for *Candida albicans* gene and protein information. *Nucleic Acids Res.* **33**, D358–D363 (2005). doi: 10.1093/nar/gki003; pmid: 15608216
35. C. Kurtzman, J. W. Fell, T. Boekhout, *The Yeasts: A Taxonomic Study* (Elsevier, 2011).
36. E. Rozpędowska *et al*., Parallel evolution of the make-accumulate-consume strategy in *Saccharomyces* and *Dekkera* yeasts. *Nat. Commun.* **2**, 302 (2011). doi: 10.1038/ncomms1305; pmid: 21556056
37. G. I. Naumov, E. S. Naumova, P. D. Sniegowski, *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. *Can. J. Microbiol.* **44**, 1045–1050 (1998). doi: 10.1139/w98-104; pmid: 10029999
38. S. K. Hewitt, D. S. Foster, P. S. Dyer, S. V. Avery, Phenotypic heterogeneity in fungi: Importance and methodology. *Fungal Biol. Rev.* **30**, 176–184 (2016). doi: 10.1016/j.fbr.2016.09.002
39. I. V. Ene, S. Brunke, A. J. P. Brown, B. Hube, Metabolism in fungal pathogenesis. *Cold Spring Harb. Perspect. Med.* **4**, a019695 (2014). doi: 10.1101/cshperspect.a019695; pmid: 25190251
40. C. T. Hittinger *et al*., Genomics and the making of yeast biodiversity. *Curr. Opin. Genet. Dev.* **35**, 100–109 (2015). doi: 10.1016/j.gde.2015.10.008; pmid: 26649756
41. M. Groenewald *et al*., A genome-informed higher rank classification of the biotechnologically important fungal subphylum *Saccharomycotina*. *Stud. Mycol.* **105**, 1–22 (2023). doi: 10.3114/sim.2023.105.01
42. C. T. Hittinger *et al*., Genomic and ecological factors shaping specialism and generalism across an entire subphylum, Figshare (2024); https://doi.org/10.25452/figshare.plus.c.6714042.
43. E. V. Kriventseva *et al*., OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019). doi: 10.1093/nar/gky1053; pmid: 30395283
44. R. M. Waterhouse *et al*., BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018). doi: 10.1093/molbev/msx319; pmid: 29220515
45. T. Krassowski *et al*., Multiple Reinventions of Mating-type Switching during Budding Yeast Evolution. *Curr. Biol.* **29**, 2555–2562.e8 (2019). doi: 10.1016/j.cub.2019.06.056; pmid: 31353182
46. S. Mühlhausen *et al*., Endogenous stochastic decoding of the CUG codon by competing Ser-and Leu-tRNAs in *Ascoidea asiatica*. *Curr. Biol.* **28**, 2046–2057.e5 (2018). doi: 10.1016/j.cub.2018.04.085; pmid: 29910077
47. Y. Li *et al*., A genome-scale phylogeny of the kingdom Fungi. *Curr. Biol.* **31**, 1653–1665.e5 (2021). doi: 10.1016/j.cub.2021.01.074; pmid: 33607033
48. X.-X. Shen *et al*., Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci. Adv.* **6**, eabd0079 (2020). doi: 10.1126/sciadv.abd0079; pmid: 33148650
49. C. Gonçalves *et al*., Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. *eLife* **7**, e33034 (2018). doi: 10.7554/eLife.33034; pmid: 29648535
50. F. M. Commichau, K. Forchhammer, J. Stülke, Regulatory links between carbon and nitrogen metabolism. *Curr. Opin. Microbiol.* **9**, 167–172 (2006). doi: 10.1016/j.mib.2006.01.001; pmid: 16458044
51. J. L. Parada, B. Magasanik, Expression of the *hut* operons of *Salmonella typhimurium* in *Klebsiella aerogenes* and in *Escherichia coli*. *J. Bacteriol.* **124**, 1263–1268 (1975). doi: 10.1128/jb.124.3.1263-1268.1975; pmid: 362
52. L. Noda-Garcia, M. L. Romero Romero, L. M. Longo, I. Kolodkin-Gal, D. S. Tawfik, *Bacilli* glutamate dehydrogenases diverged via coevolution of transcription and enzyme regulation. *EMBO Rep.* **18**, 1139–1149 (2017). doi: 10.15252/embr.201743990; pmid: 28468957
53. A. Wagner, Metabolic networks and their evolution. *Adv. Exp. Med. Biol.* **751**, 29–52 (2012). doi: 10.1007/978-1-4614-3567-9_2; pmid: 22821452
54. S. R. Proulx, D. E. L. Promislow, P. C. Phillips, Network thinking in ecology and evolution. *Trends Ecol. Evol.* **20**, 345–353 (2005). doi: 10.1016/j.tree.2005.04.004; pmid: 16701391
55. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000). doi: 10.1093/nar/28.1.27
56. A.-L. Barabási, Z. N. Oltvai, Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004). doi: 10.1038/nrg1272; pmid: 14735121
57. R. L. Cordeiro *et al*., Mechanism of high-mannose *N*-glycan breakdown and metabolism by *Bifidobacterium longum*. *Nat. Chem. Biol.* **19**, 218–229 (2023). doi: 10.1038/s41589-022-01202-4; pmid: 36443572
58. R. L. Nalabothu *et al*., Codon optimization improves the prediction of xylose metabolism from gene content in budding yeasts. *Mol. Biol. Evol.* **40**, msad111 (2023). doi: 10.1093/molbev/msad111; pmid: 37154525
59. J. F. Wolters, A. L. LaBella, D. A. Opulente, A. Rokas, C. T. Hittinger, Mitochondrial Genome Diversity across the subphylum Saccharomycotina. *Front. Microbiol.* **14**, 1268944 (2023). doi: 10.3389/fmicb.2023.1268944
60. M. A. H. Luttik *et al*., The *Saccharomyces cerevisiae NDE1* and *NDE2* genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. *J. Biol. Chem.* **273**, 24529–24534 (1998). doi: 10.1074/jbc.273.38.24529; pmid: 9733747
61. C. Wirth, U. Brandt, C. Hunte, V. Zickermann, Structure and function of mitochondrial complex I. *Biochim. Biophys. Acta Bioenerg.* **1857**, 902–914 (2016). doi: 10.1016/j.bbabio.2016.02.013
62. R. She, D. F. Jarosz, Mapping causal variants with single-nucleotide resolution reveals biochemical drivers of phenotypic change. *Cell* **172**, 478–490.e15 (2018). doi: 10.1016/j.cell.2017.12.015; pmid: 29373829
63. M. Bromley *et al*., Mitochondrial complex I is a global regulator of secondary metabolism, virulence and azole sensitivity in fungi. *PLOS ONE* **11**, e0158724 (2016). doi: 10.1371/journal.pone.0158724; pmid: 27438017
64. K. Strijbis *et al*., Carnitine-dependent transport of acetyl coenzyme A in *Candida albicans* is essential for growth on nonfermentable carbon sources and contributes to biofilm formation. *Eukaryot. Cell* **7**, 610–618 (2008). doi: 10.1128/EC.00017-08; pmid: 18281597
65. K. Strijbis, B. Distel, Intracellular acetyl unit transport in fungal carbon metabolism. *Eukaryot. Cell* **9**, 1809–1815 (2010). doi: 10.1128/ec.00172-10; pmid: 20889721
66. C. W. T. van Roermund, E. H. Hettema, M. van den Berg, H. F. Tabak, R. J. A. Wanders, Molecular characterization of carnitine-dependent transport of acetyl-CoA from peroxisomes to mitochondria in *Saccharomyces cerevisiae* and identification of a plasma membrane carnitine transporter, Agp2p. *EMBO J.* **18**, 5843–5852 (1999). doi: 10.1093/emboj/18.21.5843; pmid: 10545096
67. J. H. Swiegers, N. Dippenaar, I. S. Pretorius, F. F. Bauer, Carnitine-dependent metabolic activities in *Saccharomyces cerevisiae*: Three carnitine acetyltransferases are essential in a carnitine-dependent strain. *Yeast* **18**, 585–595 (2001). doi: 10.1002/yea.712; pmid: 11329169
68. M.-C. Harrison *et al*., Machine learning illuminates how diet influences the evolution of yeast galactose metabolism. bioRxiv 2023.07.20.549758 [Preprint] (2023); https://doi.org/10.1101/2023.07.20.549758.
69. C. Gonçalves *et al*., Diverse signatures of convergent evolution in cacti-associated yeasts. bioRxiv 2023.09.14.557833 [Preprint] (2023); https://doi.org/10.1101/2023.09.14.557833.
70. K. T. David *et al*., Saccharomycotina yeasts defy long-standing macroecological patterns. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2316031121 (2024). doi: 10.1073/pnas.2316031121; pmid: 11329169
71. World Health Organization, "WHO fungal priority pathogen list to guide research, development and public health action" (2022); https://www.who.int/publications/i/item/9789240060241.
72. L. N. A. Ries, S. Beattie, R. A. Cramer, G. H. Goldman, Overview of carbon and nitrogen catabolite metabolism in the virulence of human pathogenic fungi. *Mol. Microbiol.* **107**, 277–297 (2018). doi: 10.1111/mmi.13887; pmid: 29197127
73. I. V. Ene *et al*., Host carbon sources modulate cell wall architecture, drug resistance and virulence in a fungal pathogen. *Cell. Microbiol.* **14**, 1319–1335 (2012). doi: 10.1111/j.1462-5822.2012.01813.x; pmid: 22587014

74. M. Takashima, T. Sugita, Taxonomy of Pathogenic Yeasts *Candida*, *Cryptococcus*, *Malassezia*, and *Trichosporon*. *Med. Mycol. J.* **63**, 119–132 (2022). doi: 10.3314/mmj.22.004; pmid: 36450564
75. P. T. Magee, Fungal pathogenicity and morphological switches. *Nat. Genet.* **42**, 560–561 (2010). doi: 10.1038/ng0710-560; pmid: 20581877
76. F. L. Mayer, D. Wilson, B. Hube, *Candida albicans* pathogenicity mechanisms. *Virulence* **4**, 119–128 (2013). doi: 10.4161/viru.22913; pmid: 23302789
77. A. Rokas, M. E. Mead, J. L. Steenwyk, N. H. Oberlies, G. H. Goldman, Evolving moldy murderers: *Aspergillus* section *Fumigati* as a model for studying the repeated evolution of fungal pathogenicity. *PLOS Pathog.* **16**, e1008315 (2020). doi: 10.1371/journal.ppat.1008315; pmid: 32106242
78. J. L. Steenwyk *et al*., Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLOS Biol.* **17**, e3000255 (2019). doi: 10.1371/journal.pbio.3000255; pmid: 31112549
79. K. C. Fugelsang, C. G. Edwards, *Wine Microbiology: Practical Applications and Procedures* (Springer, 2007).
80. See the supplementary materials available online.
81. X. Zhou *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). doi: 10.1534/g3.116.034249; pmid: 27638685
82. P. Jones *et al*., InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014). doi: 10.1093/bioinformatics/btu031; pmid: 24451626
83. M. Blum *et al*., The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021). doi: 10.1093/nar/gkaa977; pmid: 33156333
84. B. Q. Minh *et al*., IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020). doi: 10.1093/molbev/msaa015; pmid: 32011700
85. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018). doi: 10.1186/s12859-018-2129-y; pmid: 29745866
86. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016). doi: 10.1093/molbev/msw054; pmid: 27004904
87. M. Kahm, G. Hasenbrink, H. Lichtenberg-Fraté, J. Ludwig, M. Kschischo, grofit: Fitting biological growth curves with R. *J. Stat. Softw.* **33**, 1–21 (2010). doi: 10.18637/jss.v033.i07
88. D. Orme, The caper package: Comparative analysis of phylogenetics and evolution in R, R package version 5 (2013).
89. T. Wu *et al*., clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021). doi: 10.1016/j.xinn.2021.100141; pmid: 34557778
90. T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, 2016), pp. 785–794.
91. G. Zhang *et al*., Characterization of the sugar alcohol-producing yeast *Pichia anomala*. *J. Ind. Microbiol. Biotechnol.* **41**, 41–48 (2014). doi: 10.1007/s10295-013-1364-5; pmid: 24170383
92. D. Opulente, DanaO523/Y1000Project: Y1000+ Analyses (Y1000Analyses_all), Zenodo (2024); https://doi.org/10.5281/zenodo.10709452.
93. M.-C. Harrison, mcharrison95/RF_for_Yeast_GenSpec: Random forest code for y1000+ paper, version 1.0.0, Zenodo (2024); https://doi.org/10.5281/zenodo.10711059.

ACKNOWLEDGMENTS

We thank L. C. Horianopoulos, K. J. Fisher, L. Sun, D. J. Krause, K. T. David, C. M. Chavez, D. C. Rinker, T. K. Sato, and Hittinger laboratory and Rokas laboratory members for helpful discussions; B. Robbertse and C. Schoch for coordinating GenBank taxonomy; the yeast community for publicly depositing taxonomic type strains; the University of Wisconsin Biotechnology Center DNA Sequencing Facility (Research Resource Identifier, RRID:SCR_017759) for providing DNA sequencing facilities and services; Wisconsin Energy Institute staff for computational support; and the Center for High-Throughput Computing at the University of Wisconsin–Madison (https://chtc.cs.wisc.edu/). This work was performed in part using resources contained within the Advanced Computing Center for Research and Education at Vanderbilt University in Nashville, TN. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture (USDA). USDA is an equal opportunity provider and employer. **Funding:** This study was supported by National Science Foundation (NSF) grant DEB-1442148 (C.T.H.); NSF grant DEB-2110403 (C.T.H.); NSF grant DEB-1442113 (A.R.); NSF grant DEB-2110404 (A.R.); in part by DOE Great Lakes Bioenergy Research Center, funded by BER Office of Science grant DE-SC0018409 (C.T.H.); USDA National Institute of Food and Agriculture Hatch projects 1020204 and 7005101 (C.T.H.); an H. I. Romnes Faculty Fellow, supported by the Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (C.T.H.); National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID) grant R56 AI146096 (A.R.); NIH, NIAID grant R01 AI153356 (A.R.); the Burroughs Wellcome Fund (A.R.); National Key R&D Program of China grant 2022YFD1401600 (X.-X.S.); National Science Foundation for Distinguished Young Scholars of Zhejiang Province grant LR23C140001 (X.-X.S.); Fundamental Research Funds for the Central Universities grant 226-2023-00021 (X.-X.S.); NIH grant T32 HG002760-16 (J.F.W.); NSF grant Postdoctoral Research Fellowship in Biology 1907278 (J.F.W.); the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program (J.L.S. and A.R.); NSF Graduate Research Fellowship grant DGE-1256259 (Q.K.L.); Predoctoral Training Program in Genetics, funded by the NIH grant 5T32GM007133 (Q.K.L.); Slovenian Research Agency grant P4–0116 (N.Č.); Slovenian Research Agency grant MRIC-UL ZIM, IP-0510 (N.Č.); Fundação para a Ciência e a Tecnologia grant UIDB/04378/2020 (C.G., P.G., and J.P.S.); Fundação para a Ciência e a Tecnologia grant LA/P/0140/2020 (C.G., P.G., and J.P.S.); Fundação para a Ciência e a Tecnologia grant PTDC/BIA-EVL/0604/2021 (C.G.); Fundação para a Ciência e a Tecnologia grant PTDC/BIA-EVL/1100/2020 (P.G.); Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil CNPq grant 408733/2021 (C.A.R.); Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil CNPq grant 406564/2022-1, "INCT Yeasts: Biodiversity, preservation and biotechnological innovation" (C.A.R.); MINCyT grant PICT-2020-SERIE A-00226 (D.L.); CONICET grant PIP 11220200102948CO (D.L.); and UNComahue grant 04/B247 (D.L.). J.L.S. is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation. **Author contributions:** D.A.O. designed and implemented research, led phenotypic data collection, led genome sequencing, performed computational analyses and statistical analyses, managed data, and prepared figures. A.L.L. designed and implemented computational analyses, managed data, and prepared figures. M.C.H. designed and implemented machine learning analyses. J.F.W. designed and implemented genome filtering and data curation pipelines. J.K. and J.L.S. conducted data curation and filtering. X.-X.S. led the phylogenomic analyses with C.L. and Yu.L. X.Z. led the annotation of genomes with Yo.L. D.A.O., H.R.S., J.V., C.R.M., Q.K.L., E.J.U., and A.B.H. phenotyped and sequenced strains. M.S. and C.G. performed fructophilic phenotyping experiments. D.A.O., K.V.B., M.J., M.A.B.H., Q.K.L., C.A.R., N.Č., D.L., C.P.K., M.G., and C.T.H. provided yeast strains. J.H.D., A.B.H., C.P.K., and M.G. curated and organized strains and metadata. J.P.S. contributed resources to fructophilic phenotyping experiments. P.G. supervised fructophilic phenotyping experiments. C.P.K. and M.G. led the taxonomy. D.A.O. and A.L.L. cowrote the manuscript with contributions from M.-C.H., J.F.W., J.K., J.L.S., C.G., P.G., X.Z., X.-X.S., M.G., A.R., and C.T.H. A.R. and C.T.H. edited the manuscript. C.P.K., A.R., and C.T.H. designed the research, obtained funding, and supervised the project. All authors provided comments and input and approved the manuscript. **Competing interests:** J.L.S. was a scientific adviser for WittGen Biotechnologies and is an adviser for ForensisGroup, Inc. A.R. is a scientific consultant for LifeMine Therapeutics, Inc. The authors declare no other competing interests. **Data and materials availability:** All genome sequence assemblies and raw sequencing data have been deposited in GenBank under the accessions noted in data S1. All other data, including data on growth on different carbon and nitrogen sources and isolation environment data, have been deposited in Figshare (*42*). All code has been deposited in GitHub, available through Zenodo (*92*, *93*), and is available in Figshare (*42*). Nearly all strains came from globally recognized yeast culture collections and may be ordered from the USDA (https://nrrl.ncaur.usda.gov for NRRL strains) or Westerdijk Fungal Biodiversity Institute (https://wi.knaw.nl for CBS strains) under their respective material transfer agreements (MTAs) for publicly deposited strains; currently, NRRL only requires an MTA for strains requiring BSL-2 precautions. Strains from the Hittinger laboratory that represent candidates for novel species that have not yet been formally described or deposited at CBS or NRRL may be obtained from C.T.H. under the Uniform Biological MTA or another mutually acceptable MTA.

SUPPLEMENTARY MATERIALS

Submitted 28 June 2023; resubmitted 28 July 2023
Accepted 22 March 2024
10.1126/science.adj4503