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

**Authors:** 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,*]

**Affiliations:**
^Co-first authors
#Co-second authors
%Deceased
*Corresponding authors: antonis.rokas@vanderbilt.edu and cthittinger@wisc.edu

[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, 150 Research Campus Drive, Kannapolis, NC 28081, USA

[7] College of Agriculture and Biotechnology and Centre for Evolutionary & 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] Howards 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, U.S. Department of Agriculture, Peoria, IL 61604, USA

[23] Westerdijk Fungal Biodiversity Institute, 3584 CT Utrecht, The Netherlands

**Abstract:** 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: trade-offs between performance efficiency and breadth; and 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 (1,154 yeast strains from 1,051 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 limited evidence for trade-offs. These comprehensive data argue that intrinsic factors shape niche breadth variation in microbes.

**One-Sentence Summary:** A nearly complete genomic and phenotypic catalog of the yeast subphylum illuminates the evolution of metabolic breadth.

**Main Text:**

## Introduction

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, while 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*), begging 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 utilizing a specific food source will be coupled to selection against utilizing other food sources and vice versa. Over the long-term, such selection produces generalists that utilize more substrates less efficiently and specialists that utilize 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*). Here, 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 due to canopy cover (*19*); and the availability of nitrogen sources (*20*, *21*), carbon sources (*22–24*), and growth-inhibiting specialized metabolites can differ due to 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*, 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*). In addition, 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 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 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 1,154 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 novel associations between carbon generalism and specific metabolic pathways. In 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 here, 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 new genomes in this study and combined them with 140 genomes previously sequenced by the Y1000+ Project (*40*) and 61 publicly available genomes (data S1A). Our dataset contained 1,154 genomes from 1,051 species, including 1,037 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 new species. The genomic dataset spans 96 yeast genera, which is about 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 of 387.5 Kb, which was comparable to our previous smaller-scale dataset of 332 genomes (417.2 kb) (fig. S1A & data S1A) (*28*). All genomes were annotated to identify putative coding sequences. On average, 5,908 +/- 1069 (s.d.) protein-coding sequences were identified per genome with a range from 3,775 (*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 = 13.2Mb +/- 3.5Mb) ranged from 7.2Mb (*Starmerella lactis-condensi*) to 41.3Mb (*Magnusiomyces magnusii*) (fig. S1C-D & data S1A). Of the 1,154 yeast genomes, 1,000 (~87%) had ≥ 90% of the 2,137 predefined single-copy orthologs defined by OrthoDB v10 (data S1A & data S1B**)** (*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 1,403 orthologous groups (OGs) from 1,154 Saccharomycotina genomes and 21 outgroups. Nearly all internodes in both concatenation-based (1,136/1,153; 99%) and coalescent-based (1,123/1,153; 97%) phylogenies received strong (≥ 95%) support (Fig. 2 and fig. S2 and fig. S3). The two phylogenies were highly congruent, with only 60/1,153 (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, while 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 lab 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 & 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 four, eight, and twelve, 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, while 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 asking if 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-value < 2.2e-16); specialists were more often slow growers (38%: 146/381 growth rates) than fast growers (15%: 54/381), whereas generalists were more often fast growers (33%: 403/1,222 growth rates) than slow growers (20%: 238/1,222 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, 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, while 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-value = $7.8 \times 10^{-11}$). In this case, carbon niche breadth was significantly and positively correlated with growth rates on glucose (p-value = 0.0002, slope = 0.002), sucrose (p-value =0.0032, slope = 0.001), and fructose (p-value=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 to 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 than specialists, including under conditions preferred by specialists.

We next tested whether there was a trade-off 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-value = $3.26 \times 10^{-14}$) (Fig. 3C). Moreover, trait-trait co-evolutionary analysis found that carbon generalists tended to also be nitrogen generalists (Bayes factor >2). Furthermore, our analyses of co-evolution 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-value of 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 non-generalist 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, 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 (1,088 total). We then grouped strains by similar environments using a formal hierarchical ontology of isolation environments. This ontology contained 1,597 classes (specific environments) (fig. S8, 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 wood-boring 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, than 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 & fig. S10A-B). Strikingly, 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 due 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 and fig. S10C-F**,** 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. S10C-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 employed 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 (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.

Examination of the features on which the classifier relied using dropout analysis identified 2,050 KOs that significantly contributed to classification accuracy. Approximately 5,000 unique yeast KOs were used to train the algorithm, suggesting 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 fourth. 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 utilize 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). Indeed, in a portion of this pathway, 5 of the 8 reactions were among the 2,050 KOs (with two falling in the top 100 KOs) that contributed to the classification of carbon generalists and specialists (black boxes in Fig. 4D). Importantly, growth on xylose was included in our carbon classification, and the xylose metabolism genes *XYL1* (Step 2 in Fig. 4D), *XYL2* (Step 3), and *XYL3* (Step 8) were all identified as important features (with *XYL1* falling within the top 100), suggesting 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 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). Interestingly, 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 2,050 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, as 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 to 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 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), while 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 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*), while 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 to 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 than those that did not (z-test, $\chi^2$= 10.425, p-value = 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 to those of their non-pathogenic 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, suggesting that carbon niche breadth may play an important role in promoting or preventing fungal pathogenesis (*72*). Yeasts are subject to diverse micro-environments 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*). Interestingly, 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 non-pathogenic relatives (Fig. 5A-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 component 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 3-fold: *Candida parapsilosis* had a growth rate of 0.042, while *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 role 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 pathogens had an elevated rate of growth at 37°C (~64%) compared to all yeasts for which growth at this temperature was measured (~41%). This result likely reflects the necessity of growth at 37°C to evolve prior to pathogenicity. Heat shock proteins (HSPs) are also known to impact 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). The analyses shown here 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 non-pathogenic relatives supports the hypothesis that the traits and genetic elements contributing to pathogenicity are not broadly shared across pathogens but unique to each (*77*). The data and analyses presented here provide a model for the investigation of other complex traits across Saccharomycotina using our ensemble of genomic, metabolic, and environmental data.

## Conclusions

Here 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 over 10,000 growth rates measured across 853 yeasts. We found that generalists typically grew faster on carbon sources than 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.

In 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 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 unique traits that provide them with a selective advantage. However, we found that generalists, as compared to 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% 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 unique 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.

## Summary of Methods

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 (USA). 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 1,403 orthologous groups (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 co-evolution**

We generated quantitative growth data on 18 carbon and 6 nitrogen sources for 853 yeasts, measuring optical density every two 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 co-evolution of carbon traits and carbon generalism/specialism using BayesTraits (http://www.evolution.reading.ac.uk).

**Underlying factors driving generalist and specialist phenotypes**

We identified strain-specific isolation environments for 1,088 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 generalists 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/absence was most strongly associated with carbon generalism and specialism.

## References


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 (N Y)* **75**, 2494–2508 (2021).
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).
3. J. W. Wenger, J. Piotrowski, S. Nagarajan, K. Chiotti, G. Sherlock, F. Rosenzweig, Hunger Artists: Yeast Adapted to Carbon Limitation Show Trade-Offs under Carbon Sufficiency. *PLoS Genet* **7**, e1002202 (2011).

4. R. H. MacArthur, *Geographical Ecology: Patterns in the Distribution of Species* (Princeton University Press, 1984; https://books.google.com/books?id=3NAYEKc--WAC)*Biology / Princeton University Press*.
5. D. S. Wilson, J. Yoshimura, On the coexistence of specialists and generalists. *Am Nat* **144**, 692–707 (1994).
6. A. R. Burmeister, A. Fortier, C. Roush, A. J. Lessing, R. G. Bender, R. Barahman, R. Grant, B. K. Chan, P. E. Turner, Pleiotropy complicates a trade-off between phage resistance and antibiotic resistance. *Proceedings of the National Academy of Sciences* **117**, 11207–11216 (2020).
7. Z. Luo, J. Liu, P. Zhao, T. Jia, C. Li, B. Chai, Biogeographic patterns and assembly mechanisms of bacterial communities differ between habitat generalists and specialists across elevational gradients. *Front Microbiol* **10**, 169 (2019).
8. F. Seebacher, V. Ducret, A. G. Little, B. Adriaenssens, Generalist–specialist trade-off during thermal acclimation. *R Soc Open Sci* **2**, 140251 (2015).
9. J. D. Napier, P. P. Grabowski, J. T. Lovell, J. Bonnette, S. Mamidi, M. J. Gomez-Hughes, A. VanWallendael, X. Weng, L. H. Handley, M. K. Kim, A generalist–specialist trade-off between switchgrass cytotypes impacts climate adaptation and geographic range. *Proceedings of the National Academy of Sciences* **119**, e2118879119 (2022).
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. *Proceedings of the Royal Society B: Biological Sciences* **275**, 277–284 (2008).
11. Q. Xu, G. Luo, J. Guo, Y. Xiao, F. Zhang, S. Guo, N. Ling, Q. Shen, Microbial generalist or specialist: Intraspecific variation and dormancy potential matter. *Mol Ecol* **31** (2022).
12. R. Kassen, The experimental evolution of specialists, generalists, and the maintenance of diversity. *J Evol Biol* **15**, 173–190 (2002).
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).
14. L. Büchi, S. Vuilleumier, Coexistence of specialist and generalist species is shaped by dispersal and environmental factors. *Am Nat* **183**, 612–624 (2014).
15. C. Bleuven, C. R. Landry, Molecular and cellular bases of adaptation to a changing environment in microorganisms. *Proceedings of the Royal Society B: Biological Sciences* **283** (2016).
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).
17. D. A. Opulente, E. J. Rollinson, C. Bernick-Roehr, A. B. Hulfachor, A. Rokas, C. P. Kurtzman, C. T. Hittinger, Factors driving metabolic diversity in the budding yeast subphylum. *BMC Biol* **16**, 26 (2018).
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).
19. T. Vanneste, S. Govaert, F. Spicher, J. Brunet, S. A. O. Cousins, G. Decocq, M. Diekmann, B. J. Graae, P.-O. Hedwall, R. E. Kapás, Contrasting microclimates among hedgerows and woodlands across temperate Europe. *Agric For Meteorol* **281**, 107818 (2020).
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).
21. E.-U. Manuel, R. Juan-Luis, Expression of a Pseudomonas putidaAminotransferase Involved in Lysine Catabolism Is Induced in the Rhizosphere. *Appl Environ Microbiol* **67**, 5219–5224 (2001).
22. C. Schädel, A. Blöchl, A. Richter, G. Hoch, Quantification and monosaccharide composition of hemicelluloses from different plant functional types. *Plant physiology and Biochemistry* **48**, 1–8 (2010).
23. A. Ebringerová, Z. Hromádková, T. Heinze, Hemicellulose. *Polysaccharides I: Structure, characterization and use*, 1–67 (2005).
24. C. H. Jaeger Iii, S. E. Lindow, W. Miller, E. Clark, M. 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).
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).
26. A. L. Demain, From natural products discovery to commercialization: a success story. *J Ind Microbiol Biotechnol* **33**, 486–495 (2006).
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. *Amphibia-Reptilia* **43**, 299–311 (2022).
28. X.-X. Shen, D. A. Opulente, J. Kominek, X. Zhou, J. L. Steenwyk, K. V. Buh, M. A. B. B. Haase, J. H. Wisecaver, M. Wang, D. T. Doering, J. T. Boudouris, R. M. Schneider, Q. K. Langdon, M. Ohkuma, R. Endoh, M. Takashima, R. Manabe, N. Čadež, D. Libkind, C. A. Rosa, J. DeVirgilio, A. B. Hulfachor, M. Groenewald,

C. P. Kurtzman, C. T. Hittinger, A. Rokas, Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175** (2018).
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).
30. K. Voordeckers, C. A. Brown, K. Vanneste, E. van der Zande, A. Voet, S. Maere, K. J. Verstrepen, Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol* **10**, e1001446 (2012).
31. W. M. Patrick, E. M. Quandt, D. B. Swartzlander, I. Matsumura, Multicopy suppression underpins metabolic evolvability. *Mol Biol Evol* **24**, 2716–2722 (2007).
32. C. S. McBride, J. R. Arguello, B. C. O'Meara, Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**, 1395–416 (2007).
33. K. R. Christie, S. Weng, R. Balakrishnan, M. C. Costanzo, K. Dolinski, S. S. Dwight, S. R. Engel, B. Feierbach, D. G. Fisk, J. E. Hirschman, 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).
34. M. B. Arnaud, M. C. Costanzo, M. S. Skrzypek, G. Binkley, C. Lane, S. R. Miyasato, G. Sherlock, The Candida Genome Database (CGD), a community resource for Candida albicans gene and protein information. *Nucleic Acids Res* **33**, D358–D363 (2005).
35. C. Kurtzman, J. W. Fell, T. Boekhout, *The Yeasts: A Taxonomic Study* (Elsevier, 2011).
36. E. Rozpędowska, L. Hellborg, O. P. Ishchuk, F. Orhan, S. Galafassi, A. Merico, M. Woolfit, C. Compagno, J. Piškur, Parallel evolution of the make–accumulate–consume strategy in Saccharomyces and Dekkera yeasts. *Nat Commun* **2**, 302 (2011).
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).
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).
39. I. V Ene, S. Brunke, A. J. P. Brown, B. Hube, Metabolism in fungal pathogenesis. *Cold Spring Harb Perspect Med* **4**, a019695 (2014).
40. C. T. Hittinger, A. Rokas, F.-Y. Bai, T. Boekhout, P. Gonçalves, T. W. Jeffries, J. Kominek, M.-A. Lachance, D. Libkind, C. A. Rosa, J. P. Sampaio, C. P. Kurtzman, Genomics and the making of yeast biodiversity. *Curr Opin Genet Dev* **35**, 100–109 (2015).
41. M. Groenewald, C. T. Hittinger, K. Bensch, D. A. Opulente, X.-X. Shen, Y. Li, C. Liu, A. L. LaBella, X. Zhou, S. Limtong, A genome-informed higher rank classification of the biotechnologically important fungal subphylum Saccharomycotina. *Stud Mycol* (2023).
42. Data is available at: https://doi.org/10.25452/figshare.plus.c.6714042.
43. E. V Kriventseva, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F. A. Simão, E. M. Zdobnov, 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).
44. R. M. Waterhouse, M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V Kriventseva, E. M. Zdobnov, BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**, 543–548 (2018).
45. T. Krassowski, J. Kominek, X. X. Shen, D. A. Opulente, X. Zhou, A. Rokas, C. T. Hittinger, K. H. Wolfe, Multiple Reinventions of Mating-type Switching during Budding Yeast Evolution. *Current Biology* **29**, 2555-2562.e8 (2019).
46. S. Mühlhausen, H. D. Schmitt, K.-T. Pan, U. Plessmann, H. Urlaub, L. D. Hurst, M. Kollmar, Endogenous stochastic decoding of the CUG codon by competing Ser-and Leu-tRNAs in Ascoidea asiatica. *Current Biology* **28**, 2046–2057 (2018).
47. Y. Li, J. L. Steenwyk, Y. Chang, Y. Wang, T. Y. James, J. E. Stajich, J. W. Spatafora, M. Groenewald, C. W. Dunn, C. T. Hittinger, A genome-scale phylogeny of the kingdom Fungi. *Current Biology* **31**, 1653–1665 (2021).
48. X.-X. Shen, J. L. Steenwyk, A. L. LaBella, D. A. Opulente, X. Zhou, J. Kominek, Y. Li, M. Groenewald, C. T. Hittinger, A. Rokas, Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci Adv* **6**, eabd0079 (2020).
49. C. Gonçalves, J. H. Wisecaver, J. Kominek, M. S. Oom, M. J. Leandro, X.-X. Shen, D. A. Opulente, X. Zhou, D. Peris, C. P. Kurtzman, Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. *Elife* **7**, e33034 (2018).

50. F. M. Commichau, K. Forchhammer, J. Stülke, Regulatory links between carbon and nitrogen metabolism. **9**, 167–172 (2006).
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).
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).
53. A. Wagner, Metabolic networks and their evolution. *Adv Exp Med Biol* **751** (2012).
54. S. R. Proulx, D. E. L. Promislow, P. C. Phillips, Network thinking in ecology and evolution. *Trends Ecol Evol* **20**, 345–353 (2005).
55. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. [Preprint] (2000). https://doi.org/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).
57. R. L. Cordeiro, C. R. Santos, M. N. Domingues, T. B. Lima, R. A. S. Pirolla, M. A. B. Morais, F. M. Colombari, R. Y. Miyamoto, G. F. Persinoti, A. C. Borges, Mechanism of high-mannose N-glycan breakdown and metabolism by Bifidobacterium longum. *Nat Chem Biol* **19**, 218–229 (2023).
58. R. L. Nalabothu, K. J. Fisher, A. L. LaBella, T. A. Meyer, D. A. Opulente, J. F. Wolters, A. Rokas, C. T. Hittinger, Codon optimization improves the prediction of xylose metabolism from gene content in budding yeasts. *Mol Biol Evol*, msad111 (2023).
59. J. F. Wolters, A. L. LaBella, D. A. Opulente, A. Rokas, C. T. Hittinger, Mitochondrial Genome Diversity across the Subphylum Saccharomycotina. *bioRxiv* (2023).
60. M. A. H. Luttik, K. M. Overkamp, P. Kötter, S. De Vries, J. P. Van Dijken, J. T. Pronk, The Saccharomyces cerevisiae NDE1 and NDE2 genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. *Journal of Biological Chemistry* **273**, 24529–24534 (1998).
61. C. Wirth, U. Brandt, C. Hunte, V. Zickermann, Structure and function of mitochondrial complex I. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1857**, 902–914 (2016).
62. R. She, D. F. Jarosz, Mapping causal variants with single-nucleotide resolution reveals biochemical drivers of phenotypic change. *Cell* **172**, 478–490 (2018).
63. M. Bromley, A. Johns, E. Davies, M. Fraczek, J. Mabey Gilsenan, N. Kurbatova, M. Keays, M. Kapushesky, M. Gut, I. Gut, Mitochondrial complex I is a global regulator of secondary metabolism, virulence and azole sensitivity in fungi. *PLoS One* **11**, e0158724 (2016).
64. K. Strijbis, C. W. T. van Roermund, W. F. Visser, E. C. Mol, J. van den Burg, D. M. MacCallum, F. C. Odds, E. Paramonova, B. P. Krom, B. Distel, 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).
65. K. Strijbis, B. Distel, Intracellular acetyl unit transport in fungal carbon metabolism. [Preprint] (2010). https://doi.org/10.1128/EC.00172-10.
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).
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).
68. M.-C. Harrison, E. J. Ubbelohde, A. L. LaBella, D. A. Opulente, J. F. Wolters, X. Zhou, X.-X. Shen, M. Groenewald, C. T. Hittinger, A. Rokas, Machine learning illuminates how diet influences the evolution of yeast galactose metabolism. *bioRxiv*, 2023.07.20.549758 (2023).
69. C. Gonçalves, M.-C. Harrison, J. L. Steenwyk, D. A. Opulente, A. L. LaBella, J. F. Wolters, X. Zhou, X.-X. Shen, M. Groenewald, C. T. Hittinger, A. Rokas, Diverse signatures of convergent evolution in cacti-associated yeasts. *bioRxiv*, 2023.09.14.557833 (2023).
70. K. T. David, M.-C. Harrison, D. A. Opulente, A. L. LaBella, J. F. Wolters, X. Zhou, X.-X. Shen, M. Groenewald, M. Pennell, C. T. Hittinger, A. Rokas, Saccharomycotina yeasts defy longstanding macroecological patterns. *bioRxiv*, 2023.08.29.555417 (2023).
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).
73. I. V. Ene, A. K. Adya, S. Wehmeier, A. C. Brand, D. M. Maccallum, N. A. R. Gow, A. J. P. Brown, Host carbon sources modulate cell wall architecture, drug resistance and virulence in a fungal pathogen. *Cell Microbiol* **14** (2012).
74. T. Masako, S. Takashi, Taxonomy of Pathogenic Yeasts Candida, Cryptococcus, Malassezia, and Trichosporon. *Med Mycol J* **63**, 119–132 (2022).
75. P. T. Magee, Fungal pathogenicity and morphological switches. *Nat Genet* **42**, 560–561 (2010).
76. F. L. Mayer, D. Wilson, B. Hube, Candida albicans pathogenicity mechanisms. *Virulence* **4**, 119–128 (2013).
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).
78. J. L. Steenwyk, D. A. Opulente, J. Kominek, X. X. Shen, X. Zhou, A. L. Labella, N. P. Bradley, B. F. Eichman, N. Čadež, D. Libkind, J. De Virgilio, A. B. Hulfachor, C. P. Kurtzman, C. T. Hittinger, A. Rokas, Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLoS Biol* **17** (2019).
79. K. C. Fugelsang, C. G. Edwards, *Wine Microbiology: Practical Applications and Procedures* (2007).
80. Supplementary Materials: Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts. *Science (1979)*.
81. X. Zhou, D. Peris, J. Kominek, C. P. Kurtzman, C. T. Hittinger, A. Rokas, In silico Whole Genome Sequencer and Analyzer (iWGS): a computational pipeline to guide the design and analysis of de novo genome sequencing studies. *G3: Genes, Genomes, Genetics* **6**, 3655–3662 (2016).
82. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
83. M. Blum, H. Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G. A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D. H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, C. A. Orengo, A. P. Pandurangan, C. Rivoire, C. J. A. Sigrist, I. Sillitoe, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, A. Bateman, R. D. Finn, The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* **49**, D344–D354 (2021).
84. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, E. Teeling, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).
85. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19** (2018).
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).
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).
88. D. Orme, The caper package: Comparative analysis of phylogenetics and evolution in R. *R package version* **5**, 1–36 (2013).
89. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2** (2021).
90. T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system" in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016)vols. 13-17-August-2016.
91. Code available for growth and enrichment analyses at https://zenodo.org/records/10709452.
92. Code available for machine learning analysis at https://zenodo.org/doi/10.5281/zenodo.10711058.
93. G. Zhang, Y. Lin, P. He, L. Li, Q. Wang, Y. Ma, Characterization of the sugar alcohol-producing yeast Pichia anomala. *J Ind Microbiol Biotechnol* **41** (2014).
94. K. Sylvester, Q.-M. Wang, B. James, R. Mendez, A. B. Hulfachor, C. T. Hittinger, Temperature and host preferences drive the diversification of Saccharomyces and other yeasts: A survey and the discovery of eight new yeast species. *FEMS Yeast Res* **15**, fov002 (2015).
95. W. J. Spurley, K. J. Fisher, Q. K. Langdon, K. V Buh, M. Jarzyna, M. A. B. Haase, K. Sylvester, R. V Moriarty, D. Rodriguez, A. Sheddan, S. Wright, L. Sorlie, A. B. Hulfachor, D. A. Opulente, C. T. Hittinger, Substrate, temperature, and geographical patterns among nearly 2000 natural yeast isolates. *Yeast* **n/a**, 55–68 (2022).
96. J. Kominek, D. T. Doering, D. A. Opulente, X.-X. Shen, X. Zhou, J. DeVirgilio, A. B. Hulfachor, M. Groenewald, M. A. Mcgee, S. D. Karlen, Eukaryotic acquisition of a bacterial operon. *Cell* **176**, 1356–1366 (2019).

97. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
98. L. Song, L. Florea, B. Langmead, Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol* **15**, 1–13 (2014).
99. R. Chikhi, P. Medvedev, Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**, 31–37 (2014).
100. J. T. Simpson, K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, I. Birol, ABySS: A parallel assembler for short read sequence data. *Genome Res* **19** (2009).
101. A. V. Zimin, G. Marçais, D. Puiu, M. Roberts, S. L. Salzberg, J. A. Yorke, The MaSuRCA genome assembler. *Bioinformatics* **29** (2013).
102. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* **19**, 455–477 (2012).
103. D. R. Zerbino, E. Birney, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18** (2008).
104. C. L. Weib, M. Pais, L. M. Cano, S. Kamoun, H. A. Burbano, nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* **19** (2018).
105. S. Huang, M. Kang, A. Xu, HaploMerger2: Rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics* **33** (2017).
106. A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
107. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, E. M. Zdobnov, BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38** (2021).
108. E. V. Kriventseva, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F. A. Simão, E. M. Zdobnov, 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** (2019).
109. R. M. Waterhouse, M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V Kriventseva, E. M. Zdobnov, BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**, 543–548 (2018).
110. T. Brůna, K. J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3** (2021).
111. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24** (2008).
112. T. Brůna, A. Lomsadze, M. Borodovsky, GeneMark-EP+: Eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform* **2** (2020).
113. T. Aramaki, R. Blanc-Mathieu, H. Endo, K. Ohkubo, M. Kanehisa, S. Goto, H. Ogata, KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
114. Y. Shulgina, S. R. Eddy, A computational screen for alternative genetic codes in over 250,000 genomes. *Elife* **10** (2021).
115. J. L. Steenwyk, A. Rokas, orthofisher: A broadly applicable tool for automated gene identification and retrieval. *G3: Genes, Genomes, Genetics* **11** (2021).
116. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5** (2010).
117. S. Q. Le, O. Gascuel, An improved general amino acid replacement matrix. *Mol Biol Evol* **25** (2008).
118. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* **30** (2013).
119. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25** (2009).
120. I. Letunic, P. Bork, Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49** (2021).
121. T. Krassowski, A. Y. Coughlan, X. X. Shen, X. Zhou, J. Kominek, D. A. Opulente, R. Riley, I. V. Grigoriev, N. Maheshwari, D. C. Shields, C. P. Kurtzman, C. T. Hittinger, A. Rokas, K. H. Wolfe, Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. *Nat Commun* **9** (2018).

122. A. L. LaBella, D. A. Opulente, J. L. Steenwyk, C. T. Hittinger, A. Rokas, Signatures of optimal codon usage in metabolic genes inform budding yeast ecology. *PLoS Biol* **19**, e3001185 (2021).
123. S. W. Kembel, P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, C. O. Webb, Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
124. F. E. Harrell Jr, M. C. Dupont, The Hmisc package. *R package version* **3**, 0–12 (2006).
125. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* **13** (2003).
126. G. Csardi, T. Nepusz, G. Csárdi, T. Nepusz, G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal Complex Systems* **1695**, 1–9 (2006).
127. A. Ofir, K. Hofmann, E. Weindling, T. Gildor, K. S. Barker, P. D. Rogers, D. Kornitzer, Role of a Candida albicans Nrm1/Whi5 homologue in cell cycle gene expression and DNA replication stress response. *Mol Microbiol* **84**, 778–794 (2012).
128. R. M. Al-Reedy, R. Malireddy, C. B. Dillman, J. C. Kennell, Comparative analysis of Fusarium mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. Fungal Genetics and Biology 49 (2012).
129. C. R. Baker, B. B. Tuch, A. D. Johnson, Extensive DNA-binding specificity divergence of a conserved transcription regulator. Proc Natl Acad Sci U S A 108 (2011).
130. E. C. Baker, B. Wang, N. Bellora, D. Peris, A. B. Hulfachor, J. A. Koshalek, M. Adams, D. Libkind, C. T. Hittinger, The genome sequence of Saccharomyces eubayanus and the domestication of lager-brewing yeasts. Mol Biol Evol 32 (2015).
131. T. M. Batista, R. G. Moreira, H. O. Hilário, C. G. Morais, G. R. Franco, L. H. Rosa, C. A. Rosa, Draft genome sequence of Sugiyamaella xylanicola UFMG-CM-Y1884T, a xylan-degrading yeast species isolated from rotting wood samples in Brazil. Genom Data 11 (2017).
132. G. Borelli, J. José, P. J. P. L. Teixeir, L. V. dos Santos, G. A. G. Pereira, De novo assembly of Candida sojae and Candida boidinii genomes, unexplored xylose-consuming yeasts with potential for renewable biochemical production. Genome Announc 4 (2016).
133. B. Brejová, H. Lichancová, F. Brázdovič, A. Cillingová, M. Neboháčová, L. Tomáška, T. Vinař, J. Nosek, Draft genome sequence of an obligate psychrophilic yeast, Candida psychrophila NRRL Y-17665T. Genome Announc 5 (2017).
134. B. Brejová, H. Lichancová, V. Hodorová, M. Neboháčová, Ľ. Tomáška, T. Vinař, J. Nosek, Genome Sequence of an Arthroconidial Yeast, Saprochaete fungicola CBS 625.85. Microbiol Resour Announc 8 (2019).
135. B. Brejová, H. Lichancová, F. Brázdovič, E. Hegedűsová, M. Forgáčová Jakúbková, V. Hodorová, V. Džugasová, A. Baláž, L. Zeiselová, A. Cillingová, M. Neboháčová, V. Raclavský, Ľ. Tomáška, B. F. Lang, T. Vinař, J. Nosek, Genome sequence of the opportunistic human pathogen Magnusiomyces capitatus. Curr Genet 65 (2019).
136. G. Butler, M. D. Rasmussen, M. F. Lin, M. A. S. Santos, S. Sakthikumar, C. A. Munro, E. Rheinbay, M. Grabherr, A. Forche, J. L. Reedy, I. Agrafioti, M. B. Arnaud, S. Bates, A. J. P. Brown, S. Brunke, M. C. Costanzo, D. A. Fitzpatrick, P. W. J. De Groot, D. Harris, L. L. Hoyer, B. Hube, F. M. Klis, C. Kodira, N. Lennard, M. E. Logue, R. Martin, A. M. Neiman, E. Nikolaou, M. A. Quail, J. Quinn, M. C. Santos, F. F. Schmitzberger, G. Sherlock, P. Shah, K. A. T. Silverstein, M. S. Skrzypek, D. Soll, R. Staggs, I. Stansfield, M. P. H. Stumpf, P. E. Sudbery, T. Srikantha, Q. Zeng, J. Berman, M. Berriman, J. Heitman, N. A. R. Gow, M. C. Lorenz, B. W. Birren, M. Kellis, C. A. Cuomo, Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 459 (2009).
137. N. Čadež, N. Bellora, R. Ulloa, C. T. Hittinger, D. Libkind, Genomic content of a novel yeast species Hanseniaspora gamundiae sp. nov. from fungal stromata (Cyttaria) associated with a unique fermented beverage in Andean Patagonia, Argentina. PLoS One 14 (2019).
138. N. Čadež, N. Bellora, R. Ulloa, M. Tome, H. Petković, M. Groenewald, C. T. Hittinger, D. Libkind, Hanseniaspora smithiae sp. nov., a Novel Apiculate Yeast Species From Patagonian Forests That Lacks the Typical Genomic Domestication Signatures for Fermentative Environments. Front Microbiol 12, 679894 (2021).
139. S. Chatterjee, S. V. Alampalli, R. K. Nageshan, S. T. Chettiar, S. Joshi, U. S. Tatu, Draft genome of a commonly misdiagnosed multidrug resistant pathogen Candida auris. BMC Genomics 16 (2015).
140. F. S. Dietrich, S. Voegeli, S. Kuo, P. Philippsen, Genomes of Ashbya fungi isolated from insects reveal four mating-type loci, numerous translocations, lack of transposons, and distinct gene duplications. G3 (Bethesda) 3, 1225–1239 (2013).

141. F. S. Dietrich, S. Voegeli, S. Brachat, A. Lerch, K. Gates, S. Steiner, C. Mohr, R. Pöhlmann, P. Luedi, S. Choi, R. A. Wing, A. Flavier, T. D. Gaffney, P. Philippsen, The Ashbya gossypii Genome as a Tool for Mapping the Ancient Saccharomyces cerevisiae Genome. Science (1979) 304 (2004).
142. B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. De Montigny, C. Marck, C. Neuvéglise, E. Talla, Genome evolution in yeasts. Nature 430, 35–44 (2004).
143. S. Duplessis, C. A. Cuomo, Y. C. Lin, A. Aerts, E. Tisserant, C. Veneault-Fourrey, D. L. Joly, S. Hacquard, J. Amselem, B. L. Cantarel, R. Chiu, P. M. Coutinho, N. Feaue, M. Field, P. Frey, E. Gelhaye, J. Goldberg, M. G. Grabherr, C. D. Kodira, A. Kohler, U. Kües, E. A. Lindquist, S. M. Lucas, R. Mago, E. Mauceli, E. Morin, C. Murat, J. L. Pangilinan, R. Park, M. Pearson, H. Quesneville, N. Rouhier, S. Sakthikumar, A. A. Salamov, J. Schmutz, B. Selles, H. Shapiro, P. Tanguay, G. A. Tuskan, B. Henrissat, Y. Van De Peer, P. Rouzé, J. G. Ellis, P. N. Dodds, J. E. Schein, S. Zhong, R. C. Hamelin, I. V. Grigoriev, L. J. Szabo, F. Martin, Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci U S A 108 (2011).
144. K. C. Freel, A. Friedrich, V. Sarilar, H. Devillers, C. Neuvéglise, J. Schacherer, Whole-genome sequencing and intraspecific analysis of the yeast species *Lachancea quebecensis*. Genome Biol Evol 8 (2016).
145. T. Gabaldón, T. Martin, M. Marcet-Houben, P. Durrens, M. Bolotin-Fukuhara, O. Lespinet, S. Arnaise, S. Boisnard, G. Aguileta, R. Atanasova, C. Bouchier, A. Couloux, S. Creno, J. Almeida Cruz, H. Devillers, A. Enache-Angoulvant, J. Guitard, L. Jaouen, L. Ma, C. Marck, C. Neuvéglise, E. Pelletier, A. Pinard, J. Poulain, J. Recoquillay, E. Westhof, P. Wincker, B. Dujon, C. Hennequin, C. Fairhead, Comparative genomics of emerging pathogens in the Candida glabrata clade. BMC Genomics 14 (2013).
146. V. Galeote, F. Bigey, H. Devillers, C. Neuvéglise, S. Dequin, Genome sequence of the food spoilage yeast Zygosaccharomyces bailii CLIB 213T. Genome Announc 1 (2013).
147. R. Gazis, A. Kuo, R. Riley, K. LaButti, A. Lipzen, J. Lin, M. Amirebrahimi, C. N. Hesse, J. W. Spatafora, B. Henrissat, M. Hainaut, I. V. Grigoriev, D. S. Hibbett, The genome of Xylona heveae provides a window into fungal endophytism. Fungal Biol 120 (2016).
148. J. L. Souciet, B. Dujon, C. Gaillardin, M. Johnston, P. V. Baret, P. Cliften, D. J. Sherman, J. Weissenbach, E. Westhof, P. Wincker, C. Jubin, J. Poulain, V. Barbe, B. Ségurens, F. Artiguenave, V. Anthouard, B. Vacherie, M. E. Val, R. S. Fulton, P. Minx, R. Wilson, P. Durrens, G. Jean, C. Marck, T. Martin, M. Nikolski, T. Rolland, M. L. Seret, S. Casarégola, L. Despons, C. Fairhead, G. Fischer, I. Lafontaine, V. Leh, M. Lemaire, J. De Montigny, C. Neuvéglise, A. Thierry, I. Blanc-Lenfle, C. Bleykasten, J. Diffels, E. Fritsch, L. Frangeul, A. Goëffon, N. Jauniaux, R. Kachouri-Lafond, C. Payen, S. Potier, L. Pribylova, C. Ozanne, G. F. Richard, C. Sacerdot, M. L. Straub, E. Talla, Comparative genomics of protoploid Saccharomycetaceae. Genome Res 19 (2009).
149. A. Goffeau, G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S. G. Oliver, Life with 6000 genes. Science (1979) 274 (1996).
150. J. L. Gordon, D. Armisen, E. Proux-Weŕa, S. S. OhEígeartaigh, K. P. Byrne, K. H. Wolfe, Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. Proc Natl Acad Sci U S A 108 (2011).
151. M. A. B. Haase, J. Kominek, D. A. Opulente, X. X. Shen, A. L. LaBella, X. Zhou, J. DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, A. Rokas, C. T. Hittinger, Repeated horizontal gene transfer of GALactose metabolism genes violates Dollo's law of irreversible loss. Genetics 217 (2021).
152. J. K. Hane, R. G. T. Lowe, P. S. Solomon, K. C. Tan, C. L. Schoch, J. W. Spatafora, P. W. Crous, C. Kodira, B. W. Birren, J. E. Galagan, S. F. F. Torriani, B. A. McDonald, R. P. Oliver, Dothideomycete-plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen Stagonospora nodorum. Plant Cell 19 (2007).
153. F. E. Hartmann, R. C. Rodríguez De La Vega, P. Gladieux, W. J. Ma, M. E. Hood, T. Giraud, D. Falush, Higher Gene Flow in Sex-Related Chromosomes than in Autosomes during Fungal Divergence. Mol Biol Evol 37 (2020).
154. V. Hodorová, H. Lichancová, S. Zubenko, K. Sienkiewicz, S. M. U. Penir, P. Afanasyev, D. Boceck, S. Bonnin, S. Hakobyan, U. Smyczynska, E. Zhivkoplias, M. Zlatohurska, E. Tralle, A. Frolova, L. P. Pryszcz, B. Brejová, T. Vinař, J. Nosek, Genome Sequence of the Yeast Saprochaete ingens CBS 517.90 . Microbiol Resour Announc 8 (2019).
155. A. P. Jackson, J. A. Gamble, T. Yeomans, G. P. Moran, D. Saunders, D. Harris, M. Aslett, J. F. Barrell, G. Butler, F. Citiulo, D. C. Coleman, P. W. J. De Groot, T. J. Goodwin, M. A. Quail, J. McQuillan, C. A. Munro, A. Pain, R. T. Poulter, M. A. Rajandream, H. Renauld, M. J. Spiering, A. Tivey, N. A. R. Gow, B. Barrell, D. J. Sullivan, M. Berriman, Comparative genomics of the fungal pathogens Candida dubliniensis and Candida albicans. Genome Res 19 (2009).

156. T. W. Jeffries, I. V. Grigoriev, J. Grimwood, J. M. Laplaza, A. Aerts, A. Salamov, J. Schmutz, E. Lindquist, P. Dehal, H. Shapiro, Y. S. Jin, V. Passoth, P. M. Richardson, Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis. Nat Biotechnol 25 (2007).
157. T. Jones, N. A. Federspiel, H. Chibana, J. Dungan, S. Kalman, B. B. Magee, G. Newport, Y. R. Thorstenson, N. Agabian, P. T. Magee, R. W. Davis, S. Scherer, The diploid genome sequence of Candida albicans. Proc Natl Acad Sci U S A 101 (2004).
158. J. Kämper, R. Kahmann, M. Bölker, L. J. Ma, T. Brefort, B. J. Saville, F. Banuett, J. W. Kronstad, S. E. Gold, O. Müller, M. H. Perlin, H. A. B. Wösten, R. De Vries, J. Ruiz-Herrera, C. G. Reynaga-Peña, K. Snetselaar, M. McCann, J. Pérez-Martín, M. Feldbrügge, C. W. Basse, G. Steinberg, J. I. Ibeas, W. Holloman, P. Guzman, M. Farman, J. E. Stajich, R. Sentandreu, J. M. González-Prieto, J. C. Kennell, L. Molina, J. Schirawski, A. Mendoza-Mendoza, D. Greilinger, K. Münch, N. Rössel, M. Scherer, M. Vranĕs, O. Ladendorf, V. Vincon, U. Fuchs, B. Sandrock, S. Meng, E. C. H. Ho, M. J. Cahill, K. J. Boyce, J. Klose, S. J. Klosterman, H. J. Deelstra, L. Ortiz-Castellanos, W. Li, P. Sanchez-Alonso, P. H. Schreier, I. Häuser-Hahn, M. Vaupel, E. Koopmann, G. Friedrich, H. Voss, T. Schlüter, J. Margolis, D. Platt, C. Swimmer, A. Gnirke, F. Chen, V. Vysotskaia, G. Mannhaupt, U. Güldener, M. Münsterkötter, D. Haase, M. Oesterheld, H. W. Mewes, E. W. Mauceli, D. DeCaprio, C. M. Wade, J. Butler, S. Young, D. B. Jaffe, S. Calvo, C. Nusbaum, J. Galagan, B. W. Birren, Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature 444 (2006).
159. M. Kellis, B. W. Birren, E. S. Lander, Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature 428 (2004).
160. A. Kohler, A. Kuo, L. G. Nagy, E. Morin, K. W. Barry, F. Buscot, B. Canbäck, C. Choi, N. Cichocki, A. Clum, J. Colpaert, A. Copeland, M. D. Costa, J. Doré, D. Floudas, G. Gay, M. Girlanda, B. Henrissat, S. Herrmann, J. Hess, N. Högberg, T. Johansson, H. R. Khouja, K. Labutti, U. Lahrmann, A. Levasseur, E. A. Lindquist, A. Lipzen, R. Marmeisse, E. Martino, C. Murat, C. Y. Ngan, U. Nehls, J. M. Plett, A. Pringle, R. A. Ohm, S. Perotto, M. Peter, R. Riley, F. Rineau, J. Ruytinx, A. Salamov, F. Shah, H. Sun, M. Tarkka, A. Tritt, C. Veneault-Fourrey, A. Zuccaro, A. Tunlid, I. V. Grigoriev, D. S. Hibbett, F. Martin, Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet 47 (2015).
161. D. J. Krause, J. Kominek, D. A. Opulente, X. X. Shen, X. Zhou, Q. K. Langdon, J. DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, A. Rokas, C. T. Hittinger, Functional and evolutionary characterization of a secondary metabolite gene cluster in budding yeasts. Proc Natl Acad Sci U S A 115 (2018).
162. M. A. Lachance, E. Hurtado, T. Hsiang, A stable phylogeny of the large-spored Metschnikowia clade. Yeast 33 (2016).
163. H. Lichancová, V. Hodorová, K. Sienkiewicz, S. M. U. Penir, P. Afanasyev, D. Boceck, S. Bonnin, S. Hakobyan, P. S. Krawczyk, U. Smyczynska, E. Zhivkoplias, M. Zlatohurska, A. Odrzywolski, E. Tralle, A. Frolova, L. P. Pryszcz, B. Brejová, T. Vinař, J. Nosek, Genome Sequence of Flavor-Producing Yeast Saprochaete suaveolens NRRL Y-17571. Microbiol Resour Announc 8 (2019).
164. G. Liti, A. N. Nguyen Ba, M. Blythe, C. A. Müller, A. Bergström, F. A. Cubillos, F. Dafhnis-Calas, S. Khoshraftar, S. Malla, N. Mehta, C. C. Siow, J. Warringer, A. M. Moses, E. J. Louis, C. A. Nieduszynski, High quality de novo sequencing and assembly of the Saccharomyces arboricolus genome. BMC Genomics 14 (2013).
165. G. Liti, D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. A. James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. T. O'Kelly, A. Van Oudenaarden, D. B. H. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. A. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, E. J. Louis, Population genomics of domestic and wild yeasts. Nature 458 (2009).
166. F. P. Lobo, D. Gon{ç}alves, S. L. Alves, A. L. Gerber, de V. ATR, L. C. Basso, G. R. Franco, M. A. Soares, R. M. Cadete, C. A. Rosa, B. U. Stambuk, D. L. Gonçalves, S. L. Alves, A. L. Gerber, de V. ATR, L. C. Basso, G. R. Franco, M. A. Soares, R. M. Cadete, C. A. Rosa, B. U. Stambuk, Draft genome sequence of the D-xylose-fermenting yeast Spathaspora arborariae UFMG-HM19.1AT. Genome Announcments 2 (2014).
167. B. J. Loftus, E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, D. Bruno, J. Vamathevan, M. Miranda, I. J. Anderson, J. A. Fraser, J. E. Allen, I. E. Bosdet, M. R. Brent, R. Chiu, T. L. Doering, M. J. Donlin, C. A. D'Souza, D. S. Fox, V. Grinberg, J. Fu, M. Fukushima, B. J. Haas, J. C. Huang, G. Janbon, S. J. M. Jones, H. L. Koo, M. I. Krzywinski, J. K. Kwon-Chung, K. B. Lengeler, R. Maiti, M. A. Marra, R. E. Marra, C. A. Mathewson, T. G. Mitchell, M. Pertea, F. R. Riggs, S. L. Salzberg, J. E. Schein, A. Shvartsbeyn, H. Shin, M. Shumway, C. A. Specht, B. B. Suh, A. Tenney, T. R. Utterback, B. L. Wickes, J. R. Wortman, N. H. Wye, J. W. Kronstad, J. K. Lodge, J. Heitman, R. W. Davis, C. M. Fraser, R. W. Hyman, The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans. Science (1979) 307 (2005).

168. D. D. Lopes, S. P. Cibulski, F. Q. Mayer, F. M. Siqueira, C. A. Rosa, R. E. Hector, M. A. Z. Ayub, Draft genome sequence of the D-xylose-fermenting yeast Spathaspora xylofermentans UFMG-HMD23.3. Genome Announc 5 (2017).
169. M. R. Lopes, C. G. Morais, J. Kominek, R. M. Cadete, M. A. Soares, A. P. T. Uetanabaro, C. Fonseca, M.-A. Lachance, C. T. Hittinger, C. A. Rosa, Genomic analysis and D-xylose fermentation of three novel Spathaspora species: Spathaspora girioi sp. nov., Spathaspora hagerdaliae f. a., sp. nov. and Spathaspora gorwiae f. a., sp. nov. FEMS Yeast Res 16 (2016).
170. M. R. Lopes, T. M. Batista, G. R. Franco, L. R. Ribeiro, A. R. O. Santos, C. Furtado, R. G. Moreira, A. Goes-Neto, M. J. S. Vital, L. H. Rosa, M. A. Lachance, C. A. Rosa, Scheffersomyces stambukii f.A., sp. Nov., a D-xylose-fermenting species isolated from rotting wood. Int J Syst Evol Microbiol 68 (2018).
171. K. R. Love, K. A. Shah, C. A. Whittaker, J. Wu, M. C. Bartlett, D. Ma, R. L. Leeson, M. Priest, J. Borowsky, S. K. Young, J. C. Love, Comparative genomics and transcriptomics of Pichia pastoris. BMC Genomics 17 (2016).
172. I. MacCallum, D. Przybylski, S. Gnerre, J. Burton, I. Shlyakhter, A. Gnirke, J. Malek, K. McKernan, S. Ranade, T. P. Shea, L. Williams, S. Young, C. Nusbaum, D. B. Jaffe, ALLPATHS 2: Small genomes assembled accurately and with high continuity from short paired reads. Genome Biol 10 (2009).
173. D. Martinez, J. Challacombe, I. Morgenstern, D. Hibbett, M. Schmoll, C. P. Kubicek, P. Ferreira, F. J. Ruiz-Duenas, A. T. Martinez, P. Kersten, K. E. Hammel, A. Vanden Wymelenberg, J. Gaskell, E. Lindquist, G. Sabat, S. S. BonDurant, L. F. Larrondo, P. Canessa, R. Vicuna, J. Yadav, H. Doddapaneni, V. Subramanian, A. G. Pisabarro, J. L. Lavín, J. A. Oguiza, E. Master, B. Henrissat, P. M. Coutinho, P. Harris, J. K. Magnuson, S. E. Baker, K. Bruno, W. Kenealy, P. J. Hoegger, U. Kües, P. Ramaiya, S. Lucas, A. Salamov, H. Shapiro, H. Tu, C. L. Chee, M. Misra, G. Xie, S. Teter, D. Yaver, T. James, M. Mokrejs, M. Pospisek, I. V. Grigoriev, T. Brettin, D. Rokhsar, R. Berka, D. Cullen, Genome, transcriptome, and secretome analysis of wood decay fungus Postia placenta supports unique mechanisms of lignocellulose conversion. Proc Natl Acad Sci U S A 106 (2009).
174. D. Martinez, L. F. Larrondo, N. Putnam, M. D. Sollewijn Gelpke, K. Huang, J. Chapman, K. G. Helfenbein, P. Ramaiya, J. C. Detter, F. Larimer, P. M. Coutinho, B. Henrissat, R. Berka, D. Cullen, D. Rokhsar, Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78. Nat Biotechnol 22 (2004).
175. C. G. Morais, T. M. Batista, J. Kominek, B. M. Borelli, C. Furtado, R. G. Moreira, G. R. Franco, L. H. Rosa, C. Fonseca, C. T. Hittinger, M. A. Lachance, C. A. Rosa, Spathaspora boniae sp. nov., a D-xylose-fermenting species in the Candida albicans/lodderomyces clade. Int J Syst Evol Microbiol 67 (2017).
176. L. Morales, B. Noel, B. Porcel, M. Marcet-Houben, M. F. Hullo, C. Sacerdot, F. Tekaia, V. Leh-Louis, L. Despons, V. Khanna, J. M. Aury, V. Barbe, A. Couloux, K. Labadie, E. Pelletier, J. L. Souciet, T. Boekhout, T. Gabaldon, P. Wincker, B. Dujon, Complete DNA sequence of Kuraishia capsulata illustrates novel genomic features among budding yeasts (saccharomycotina). Genome Biol Evol 5 (2013).
177. G. Morel, L. Sterck, D. Swennen, M. Marcet-Houben, D. Onesime, A. Levasseur, N. Jacques, S. Mallet, A. Couloux, K. Labadie, J. Amselem, J. M. Beckerich, B. Henrissat, Y. Van De Peer, P. Wincker, J. L. Souciet, T. Gabaldón, C. R. Tinsley, S. Casaregola, Differential gene retention as an evolutionary mechanism to generate biodiversity and adaptation in yeasts. Sci Rep 5 (2015).
178. S. Naseeb, S. A. James, H. Alsammar, C. J. Michaels, B. Gini, C. Nueno-Palop, C. J. Bond, H. McGhie, I. N. Roberts, D. Delneri, Saccharomyces jurei sp. Nov., isolation and genetic identification of a novel yeast species from Quercus robur. Int J Syst Evol Microbiol 67 (2017).
179. D. E. Neafsey, B. M. Barker, T. J. Sharpton, J. E. Stajich, D. J. Park, E. Whiston, C. Y. Hung, C. McMahan, J. White, S. Sykes, D. Heiman, S. Young, Q. Zeng, A. Abouelleil, L. Aftuck, D. Bessette, A. Brown, M. FitzGerald, A. Lui, J. P. Macdonald, M. Priest, M. J. Orbach, J. N. Galgiani, T. N. Kirkland, G. T. Cole, B. W. Birren, M. R. Henn, J. W. Taylor, S. D. Rounsley, Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control. Genome Res 20 (2010).
180. A. Riccombeni, G. Vidanes, E. Proux-Wéra, K. H. Wolfe, G. Butler, Sequence and analysis of the genome of the pathogenic yeast Candida orthopsilosis. PLoS One 7 (2012).
181. R. Riley, S. Haridas, K. H. Wolfe, M. R. Lopes, C. T. Hittinger, M. Göker, A. A. Salamov, J. H. Wisecaver, T. M. Long, C. H. Calvey, A. L. Aerts, K. W. Barry, C. Choi, A. Clum, A. Y. Coughlan, S. Deshpande, A. P. Douglass, S. J. Hanson, H.-P. Klenk, K. M. LaButti, A. Lapidus, E. A. Lindquist, A. M. Lipzen, J. P. Meier-Kolthoff, R. A. Ohm, R. P. Otillar, J. L. Pangilinan, Y. Peng, A. Rokas, C. A. Rosa, C. Scheuner, A. A. Sibirny, J. C. Slot, J. B. Stielow, H. Sun, C. P. Kurtzman, M. Blackwell, I. V Grigoriev, T. W. Jeffries, Comparative genomics of biotechnologically important yeasts. Proceedings of the National Academy of Sciences 113, 9882–9887 (2016).

182. V. Sarilar, H. Devillers, K. C. Freel, J. Schacherer, C. Neuvéglise, Draft genome sequence of Lachancea lanzarotensis CBS 12615T, an ascomycetous yeast isolated from grapes. Genome Announc 3 (2016).
183. D. R. Scannell, A. C. Frank, G. C. Conant, K. P. Byrne, M. Woolfit, K. H. Wolfe, Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci U S A 104 (2007).
184. D. R. Scannell, O. A. Zill, A. Rokas, C. Payen, M. J. Dunham, M. B. Eisen, J. Rine, M. Johnston, C. T. Hittinger, The awesome power of yeast evolutionary genetics: New genome sequences and strain resources for the Saccharomyces sensu stricto genus. G3: Genes, Genomes, Genetics 1 (2011).
185. B. Schäfer, M. Hansen, B. F. Lang, Transcription and RNA-processing in fission yeast mitochondria. RNA 11 (2005).
186. A. S. M. Sonnenberg, W. Gao, B. Lavrijssen, P. Hendrickx, N. Sedaghat-Tellgerd, M. Foulongne-Oriol, W. S. Kong, E. G. W. M. Schijlen, J. J. P. Baars, R. G. F. Visser, A detailed analysis of the recombination landscape of the button mushroom Agaricus bisporus var. bisporus. Fungal Genetics and Biology 93 (2016).
187. T. Suzuki, T. Hoshino, A. Matsushika, Draft genome sequence of Kluyveromyces marxianus strain DMB1, isolated from sugarcane bagasse hydrolysate. Genome Announc 2 (2014).
188. M. Toome, R. A. Ohm, R. W. Riley, T. Y. James, K. L. Lazarus, B. Henrissat, S. Albu, A. Boyd, J. Chow, A. Clum, G. Heller, A. Lipzen, M. Nolan, L. Sandor, N. Zvenigorodsky, I. V. Grigoriev, J. W. Spatafora, M. C. Aime, Genome sequencing provides insight into the reproductive biology, nutritional mode and ploidy of the fern pathogen Mixia osmundae. New Phytologist 202 (2014).
189. N. Vakirlis, V. Sarilar, G. Drillon, A. Fleiss, N. Agier, J. P. Meyniel, L. Blanpain, A. Carbone, H. Devillers, K. Dubois, A. Gillet-Markowska, S. Graziani, N. Huu-Vang, M. Poirel, C. Reisser, J. Schott, J. Schacherer, I. Lafontaine, B. Llorente, C. Neuvéglise, G. Fischer, Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. Genome Res 26 (2016).
190. J. A. L. Van Kan, J. H. M. Stassen, A. Mosbach, T. A. J. Van Der Lee, L. Faino, A. D. Farmer, D. G. Papasotiriou, S. Zhou, M. F. Seidl, E. Cottam, D. Edel, M. Hahn, D. C. Schwartz, R. A. Dietrich, S. Widdison, G. Scalliet, A gapless genome sequence of the fungus Botrytis cinerea. Mol Plant Pathol 18 (2017).
191. S. Vaux, A. Criscuolo, M. Desnos-Ollivier, L. Diancourt, C. Tarnaud, M. Vandenbogaert, S. Brisse, B. Coignard, F. Dromer, Multicenter outbreak of infections by Saprochaete clavata, an unrecognized opportunistic fungal pathogen. mBio 5 (2014).
192. L. Vega-Alvarado, J. Gómez-Angulo, Z. Escalante-García, R. Grande, A. Gschaedler-Mathis, L. Amaya-Delgado, A. Sanchez-Flores, J. Arrizon, High-quality draft genome sequence of Candida apicola NRRL Y-50540. Genome Announc 3 (2015).
193. Y. Vervoort, B. Herrera-Malaver, S. Mertens, V. Guadalupe Medina, J. Duitama, L. Michiels, G. Derdelinckx, K. Voordeckers, K. J. Verstrepen, Characterization of the recombinant Brettanomyces anomalus β-glucosidase and its potential for bioflavouring. J Appl Microbiol 121 (2016).
194. J. Wendland, A. Walther, Chromosome number reduction in eremothecium coryli by two telomere-to-telomere fusions. Genome Biol Evol 6 (2014).
195. D. J. Wohlbach, A. Kuo, T. K. Sato, K. M. Potts, A. A. Salamov, K. M. LaButti, H. Sun, A. Clum, J. L. Pangilinan, E. A. Lindquist, S. Lucas, A. Lapidus, M. Jin, C. Gunawan, V. Balan, B. E. Dale, T. W. Jeffries, R. Zinkel, K. W. Barry, I. V. Grigoriev, A. P. Gasch, Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. Proc Natl Acad Sci U S A 108 (2011).
196. J. R. Wortman, J. M. Gilsenan, V. Joardar, J. Deegan, J. Clutterbuck, M. R. Andersen, D. Archer, M. Bencina, G. Braus, P. Coutinho, H. von Döhren, J. Doonan, A. J. M. Driessen, P. Durek, E. Espeso, E. Fekete, M. Flipphi, C. G. Estrada, S. Geysens, G. Goldman, P. W. J. de Groot, K. Hansen, S. D. Harris, T. Heinekamp, K. Helmstaedt, B. Henrissat, G. Hofmann, T. Homan, T. Horio, H. Horiuchi, S. James, M. Jones, L. Karaffa, Z. Karányi, M. Kato, N. Keller, D. E. Kelly, J. A. K. W. Kiel, J. M. Kim, I. J. van der Klei, F. M. Klis, A. Kovalchuk, N. Krasevec, C. P. Kubicek, B. Liu, A. Maccabe, V. Meyer, P. Mirabito, M. Miskei, M. Mos, J. Mullins, D. R. Nelson, J. Nielsen, B. R. Oakley, S. A. Osmani, T. Pakula, A. Paszewski, I. Paulsen, S. Pilsyk, I. Pócsi, P. J. Punt, A. F. J. Ram, Q. Ren, X. Robellet, G. Robson, B. Seiboth, P. van Solingen, T. Specht, J. Sun, N. Taheri-Talesh, N. Takeshita, D. Ussery, P. A. vanKuyk, H. Visser, P. J. I. van de Vondervoort, R. P. de Vries, J. Walton, X. Xiang, Y. Xiong, A. P. Zeng, B. W. Brandt, M. J. Cornell, C. A. M. J. J. van den Hondel, J. Visser, S. G. Oliver, G. Turner, The 2008 update of the Aspergillus nidulans genome annotation: a community effort. Fungal Genet Biol 46 Suppl 1 (2009).
197. H. Xiao, Z. Shao, Y. Jiang, S. Dole, H. Zhao, Exploiting Issatchenkia orientalis SD108 for succinic acid production. [Preprint] (2014). https://doi.org/10.1186/s12934-014-0121-4.
198. J. Xu, C. W. Saunders, P. Hu, R. A. Grant, T. Boekhout, E. E. Kuramae, J. W. Kronstad, Y. M. DeAngelis, N. L. Reeder, K. R. Johnstone, M. Leland, A. M. Fieno, W. M. Begley, Y. Sun, M. P. Lacey, T. Chaudhary, T.

Keough, L. Chu, R. Sears, B. Yuan, T. L. Dawson, Dandruff-associated Malassezia genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. Proc Natl Acad Sci U S A 104 (2007).
199. J. Yang, L. Wang, X. Ji, Y. Feng, X. Li, C. Zou, J. Xu, Y. Ren, Q. Mi, J. Wu, S. Liu, Y. Liu, X. Huang, H. Wang, X. Niu, J. Li, L. Liang, Y. Luo, K. Ji, W. Zhou, Z. Yu, G. Li, Y. Liu, L. Li, M. Qiao, L. Feng, K. Q. Zhang, Genomic and proteomic analyses of the fungus arthrobotrys oligospora provide insights into nematode-trap formation. PLoS Pathog 7 (2011).
200. X. Zhang, X. Cheng, L. Liu, S. Liu, Genome Sequence Resource for the Plant Pathogen Sclerotinia sclerotiorum WH6 Isolated in China. Plant Dis 105 (2021).
201. G. Kunze, C. Gaillardin, M. Czernicka, P. Durrens, T. Martin, E. Böer, T. Gabaldón, J. A. Cruz, E. Talla, C. Marck, A. Goffeau, V. Barbe, P. Baret, K. Baronian, S. Beier, C. Bleykasten, R. Bode, S. Casaregola, L. Despons, C. Fairhead, M. Giersberg, P. P. Gierski, U. Hähnel, A. Hartmann, D. Jankowska, C. Jubin, P. Jung, I. Lafontaine, V. Leh-Louis, M. Lemaire, M. Marcet-Houben, M. Mascher, G. Morel, G. F. Richard, J. Riechen, C. Sacerdot, A. Sarkar, G. Savel, J. Schacherer, D. J. Sherman, N. Stein, M. L. Straub, A. Thierry, A. Trautwein-Schult, B. Vacherie, E. Westhof, S. Worch, B. Dujon, J. L. Souciet, P. Wincker, U. Scholz, C. Neuvéglise, The complete genome of Blastobotrys (Arxula) adeninivorans LS3 - A yeast of biotechnological interest. Biotechnol Biofuels 7 (2014).
202. Opulente DA, Langdon QK, Jarzyna M, Buh KV, Haase MAB, Groenewald M, Hittinger CT. Taxogenomic analysis of a novel yeast species isolated from soil, Pichia galeolata sp. nov. Yeast 40 (2023).

**Acknowledgments:**

We thank Linda C. Horianopoulos, Kaitlin J. Fisher, Liang Sun, David J. Krause, Kyle T. David, Christina M. Chavez, David C. Rinker, Trey K. Sato, and Hittinger Lab and Rokas Lab members for helpful discussions; Barbara Robbertse and Conrad 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://doi.org/10.21231/GNT1-HW21). 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 is an equal opportunity provider and employer.

**Funding:**

National Science Foundation Grant DEB-1442148 (CTH)

National Science Foundation Grant DEB-2110403 (CTH)

National Science Foundation Grant DEB-1442113 (AR)

National Science Foundation Grant DEB-2110404 (AR)

DOE Great Lakes Bioenergy Research Center, funded by BER Office of Science Grant DE-SC0018409 (CTH)

USDA National Institute of Food and Agriculture Hatch Project 1020204 (CTH)

USDA National Institute of Food and Agriculture Hatch Project 7005101 (CTH)

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 (CTH)

National Institutes of Health/National Institute of Allergy and Infectious Diseases Grant R56 AI146096 (AR)

National Institutes of Health/National Institute of Allergy and Infectious Diseases Grant R01 AI153356 (AR)

Burroughs Wellcome Fund (AR)

Research supported by the National Key R&D Program of China Grant 2022YFD1401600 (XXS)

National Science Foundation for Distinguished Young Scholars of Zhejiang Province Grant LR23C140001 (XXS)

Fundamental Research Funds for the Central Universities Grant 226-2023-00021 (XXS)

National Institutes of Health Grant T32 HG002760-16 (JFW)

National Science Foundation Grant Postdoctoral Research Fellowship in Biology 1907278 (JFW)

Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program (JLS, AR)

National Science Foundation Graduate Research Fellowship Grant DGE-1256259 (QLK)

Predoctoral Training Program in Genetics, funded by the National Institutes of Health Grant 5T32GM007133 (QLK)

Slovenian Research Agency Grant P4–0116 (NC)

Slovenian Research Agency Grant MRIC-UL ZIM, IP-0510 (NC)

Fundação para a Ciência e a Tecnologia Grant UIDB/04378/2020 (CG, PG, JPS)

Fundação para a Ciência e a Tecnologia Grant LA/P/0140/2020 (CG, PG, JPS)

Fundação para a Ciência e a Tecnologia Grant PTDC/BIA-EVL/0604/2021 (CG)

Fundação para a Ciência e a Tecnologia Grant PTDC/BIA-EVL/1100/2020 (PG)

Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil CNPq Grant 408733/2021 (CAR)

Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil CNPq Grant 406564/2022-1 - "INCT Yeasts: Biodiversity, preservation and biotechnological innovation" (CAR)

MINCyT Grant PICT-2020-SERIE A-00226 (DL)

CONICET Grant PIP 11220200102948CO (DL)

UNComahue Grant 04/B247 (DL)

JLS is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation

**Author contributions:**

DAO designed and implemented research, led phenotypic data collection, led genome sequencing, performed computational analyses and statistical analyses, managed data, and prepared figures.

ALL designed and implemented computational analyses, managed data, and prepared figures.

MCH designed and implemented machine learning analyses.

JFW designed and implemented genome filtering and data curation pipelines.

JK and JLS conducted data curation and filtering.

XXS led the phylogenomic analyses with CL and Yuanning Li.

XZ led the annotation of genomes with Yonglin Li.

DAO, HRS, JV, CRM, QKL, EJU, and ABH phenotyped and sequenced strains.

MS and CG performed fructophilic phenotyping experiments.

DAO, KVB, MJ, MABH, QKL, CAR, NC, DL, CPK, MG, and CTH provided yeast strains.

JHD, ABH, CPK, and MG curated and organized strains and metadata.

JPS contributed resources to fructophilic phenotyping experiments.

PG supervised fructophilic phenotyping experiments.

CPK and MG led the taxonomy.

DAO and ALL co-wrote the manuscript with contributions from MCH, JFW, JK, JLS, CG, PG, XZ, XXS, MG, AR, and CTH.

AR and CTH edited the manuscript.

CPK, AR, and CTH designed the research, obtained funding, and supervised the project.

All authors provided comments, input, and approved the manuscript.

**Competing interests:** JLS was a scientific adviser for WittGen Biotechnologies and is an adviser for ForensisGroup, Inc. AR is a scientific consultant for LifeMine Therapeutics, Inc. The other 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 at https://doi.org/10.25452/figshare.plus.c.6714042 *(42)*. All code has been deposited in GitHub at https://zenodo.org/records/10709452 *(91)* and https://zenodo.org/doi/10.5281/zenodo.10711058 *(92)* and is available in Figshare at https://doi.org/10.25452/figshare.plus.c.6714042 (*42)*. Nearly all strains came from globally recognized yeast culture collections and may be ordered from the United States Department of Agriculture (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 Lab that represent candidates for novel species that have not yet been formally described or deposited at CBS or NRRL may be obtained from cthittinger@wisc.edu under the Uniform Biological MTA or other mutually acceptable MTA.

## Supplementary Materials

Materials and Methods
Figures S1-S9
Tables S1-S4
Data files S1-S7
References (94-202)

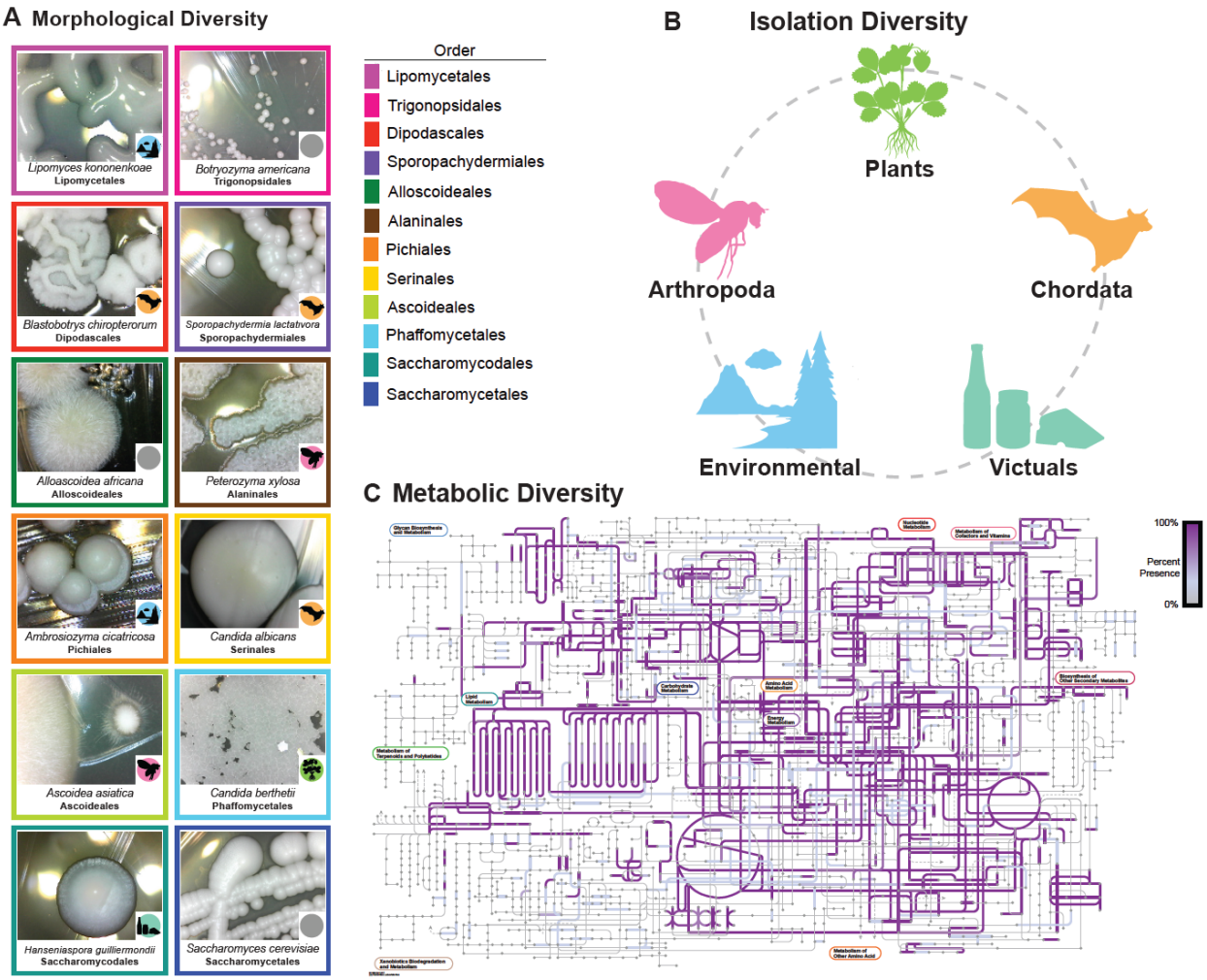



**Figure 1: Yeasts are morphologically, ecologically, and metabolically diverse.**

**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, animals, in soil, and many other environments. Strain-level isolation data were placed into an ecological ontology to allow for 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.

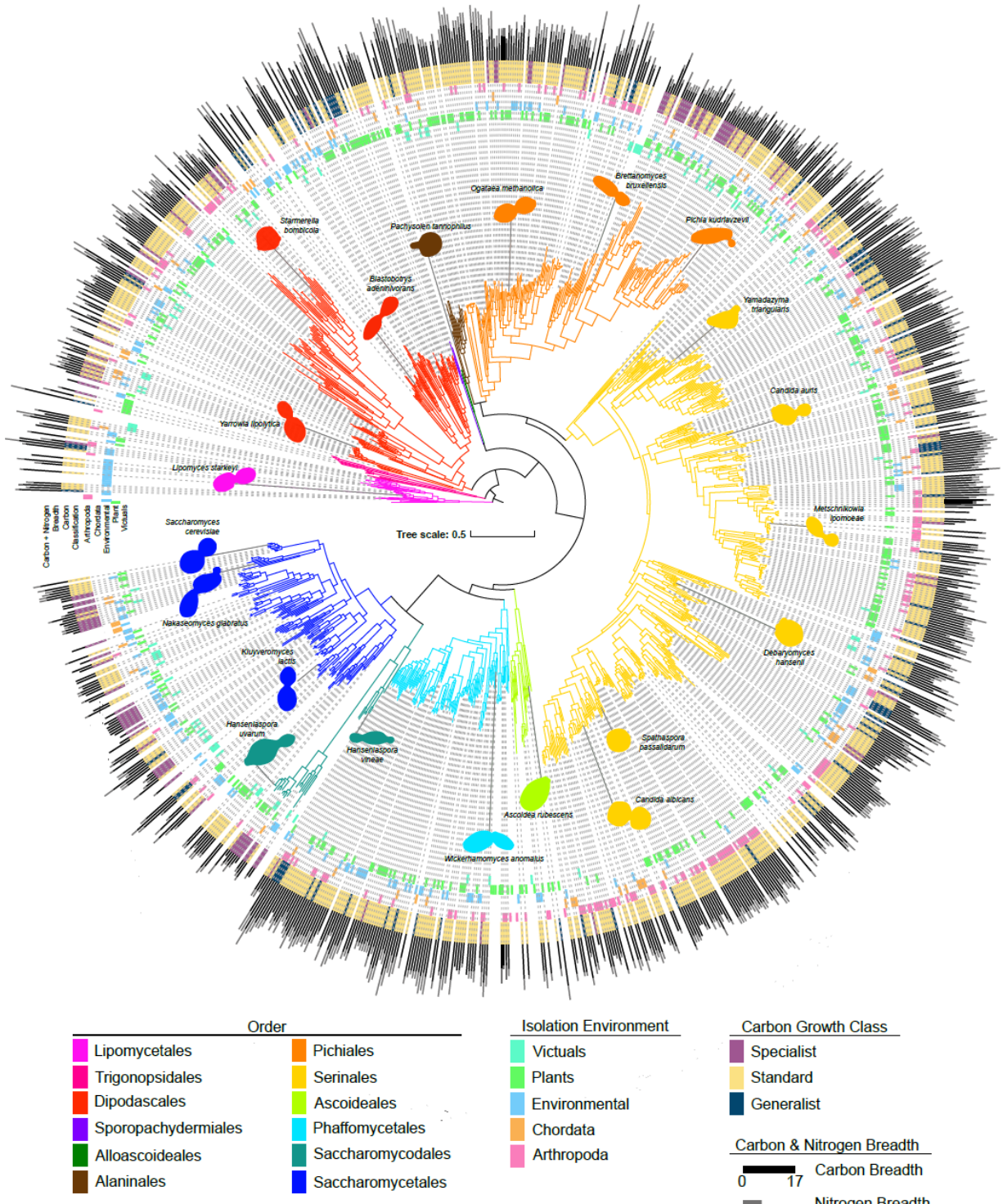



**Figure 2: Yeast traits are widely distributed across the phylogeny.**
The phylogeny of 1,154 yeasts and fungal outgroups built from 1,403 orthologous groups of genes. Branches are colored according to their taxonomic assignment to an order of Saccharomycotina (*41*). The innermost rings are colored by the top-level type of isolation environment in which each specific strain was isolated. The purple, yellow, and blue ring identifies 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.

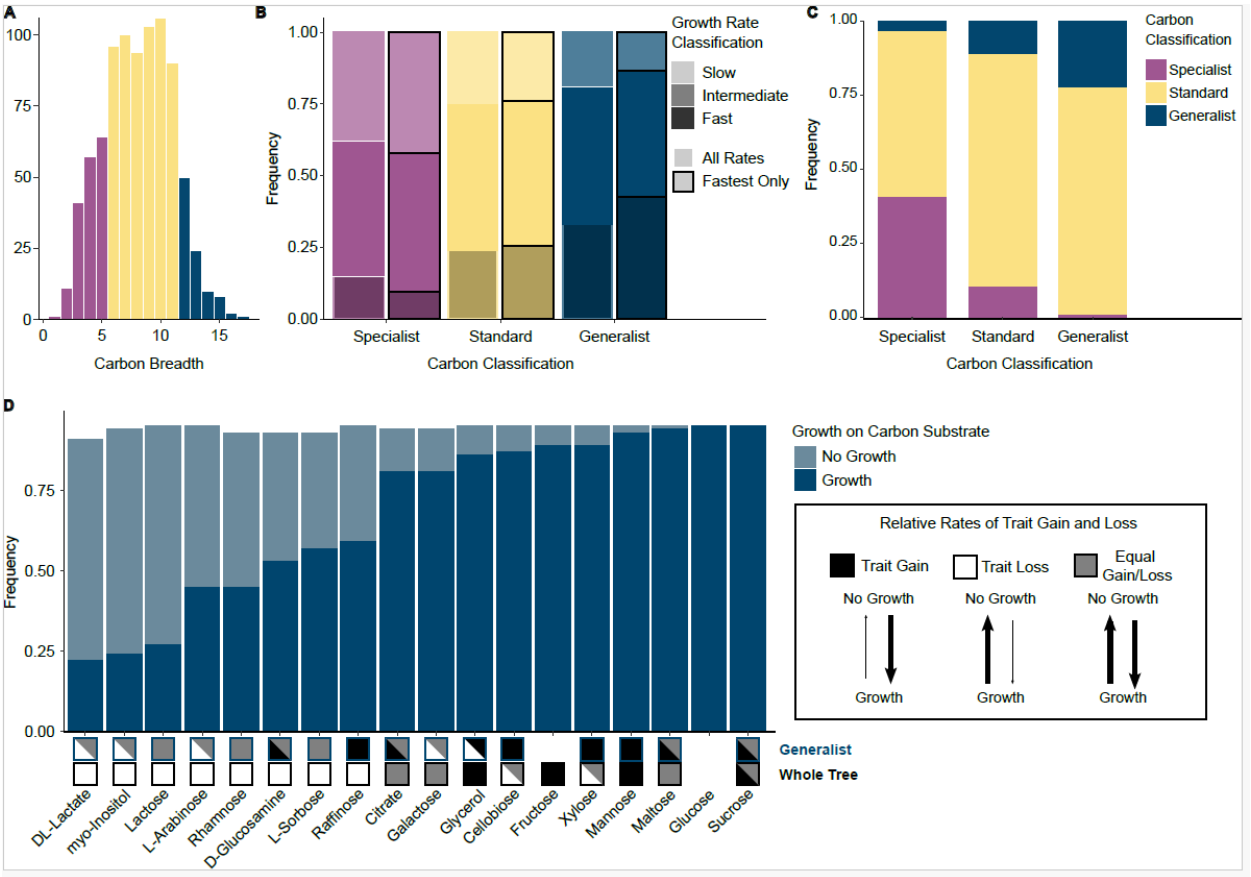


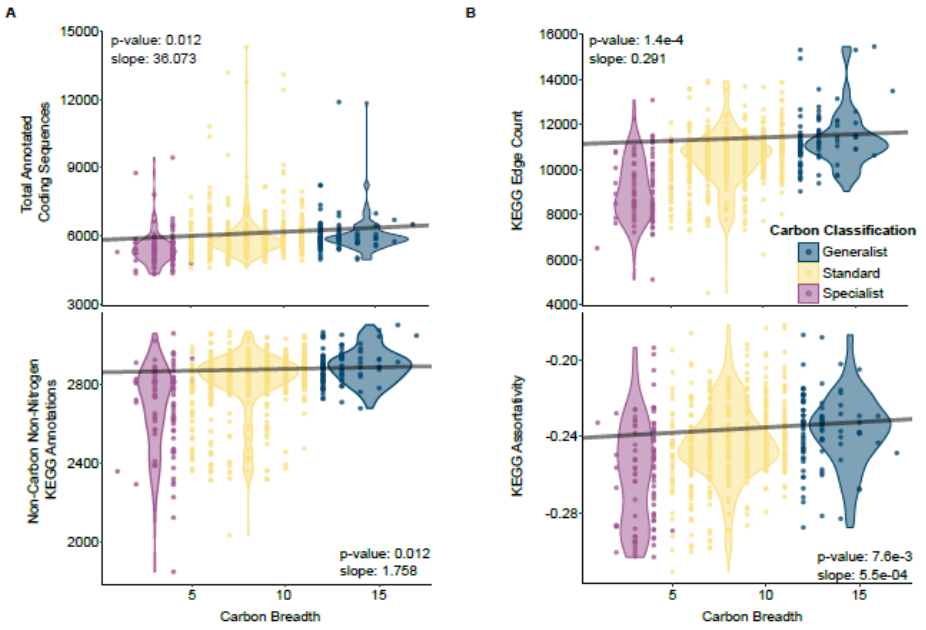

**Figure 3: Carbon specialists and generalists differ in nitrogen breadth, growth rate, and evolutionary history.**

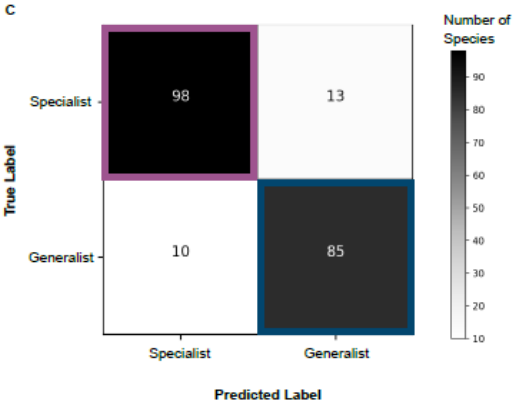
**A.** Histogram of carbon niche breadth across yeasts (n = 853). The colors of the bars represent the ranges for the different carbon classifications. Metabolic classifications were determined by permuting the binary carbon growth matrix (n = 1000 permutations). To determine the metabolic strategy of 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.

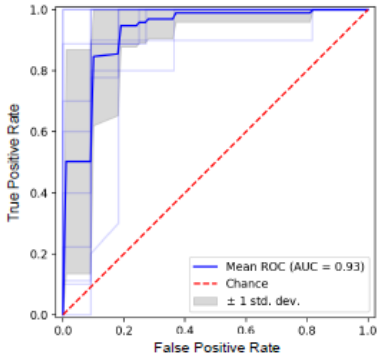
**B.** The growth rates for each yeast on each of the 18 carbon sources were categorized as slow (bottom 25%), intermediate (median 50%), or fast (top 25%) using either all the rates per yeast (white outline) or only the highest rate per yeast (black outline). Carbon generalists had the highest proportion of fast growth rates (33% all rates, 43% fastest rates), while specialists had the smallest proportion (15% all rates, 9% fastest rates) The inverse was also true, with carbon generalists having the smallest proportion of slow growth rates (19% all rates, 14% fastest rates) and carbon specialists having the highest proportion of slow growth rates (38% all rates, 42% fastest rates).

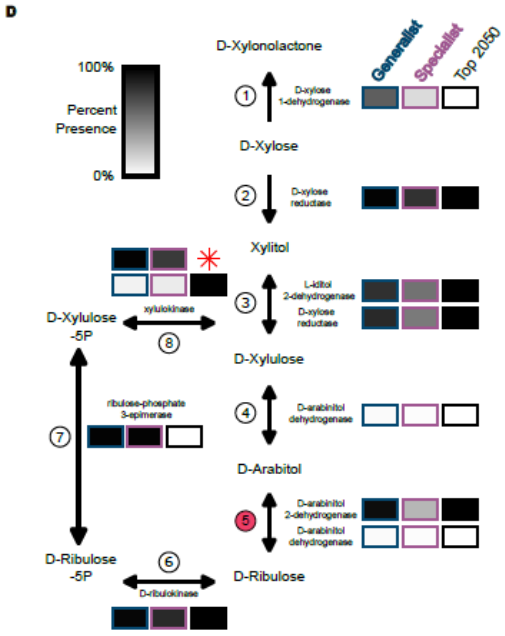
**C.** Stacked bar graph of carbon metabolic strategies within each nitrogen metabolic strategy.

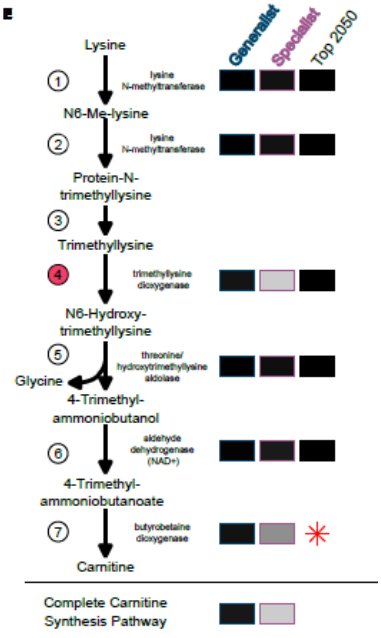
**D.** Carbon generalists shared many of the same growth traits: 10 out of 18 growth traits were found in more than 75% of generalists. Many of the carbon sources had different evolutionary trends in a generalist background as compared to across the whole tree. Three different evolutionary models are shown: trait gain (black), trait loss (white), and equal rates of trait gain and loss (gray). No box indicates that the trait was not co-evolving 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.


A
B
p-value: 0.012
slope: 36.073
Total Annotated Coding Sequences
Non-Carbon Non-Nitrogen KEGG Annotations
p-value: 0.012
slope: 1.758
Carbon Breadth
p-value: 1.4e-4
slope: 0.291
KEGG Edge Count
KEGG Assortativity
p-value: 7.6e-3
slope: 5.5e-04
Carbon Classification
Generalist
Standard
Specialist
Carbon Breadth


C
Specialist
Generalist
98
13
10
85
True Label
Predicted Label
Number of Species


True Positive Rate
False Positive Rate
Mean ROC (AUC = 0.93)
Chance
± 1 std. dev.


D
100%
Percent Presence
0%
D-Xylonolactone
D-Xylose
Xylitol
D-Xylulose
D-Arabitol
D-Ribulose
D-Xylulose -5P
D-Ribulose -5P
Generalist
Specialist
Top 2050


E
Lysine
N6-Me-lysine
Protein-N-trimethyllysine
Trimethyllysine
N6-Hydroxy-trimethyllysine
Glycine
4-Trimethyl-ammoniobutanol
4-Trimethyl-ammoniobutanoate
Carnitine
Complete Carnitine Synthesis Pathway
Generalist
Specialist
Top 2050

1357

**Figure 4: Generalist and specialist metabolism differs in expected and unexpected ways.**
**A.** Total annotated coding sequences (top) and total number of annotated KEGG ortholog groups (KOs; bottom) are 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 (*Magnusiomyces 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 ROC analysis suggests that both the sensitivity and specificity of our model is 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 2,050 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. Interestingly, experimental studies suggest that yeast D-arabinitol 2-dehydrogenase is also capable of completing the reaction in Step 4 (*93*). 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* is well studied (*58*). Therefore, we re-annotated 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 *Candida albicans* as being facilitated by the trimethyllysine dioxygenase enzyme encoded by *BBH2* (*64*). We re-annotated *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.

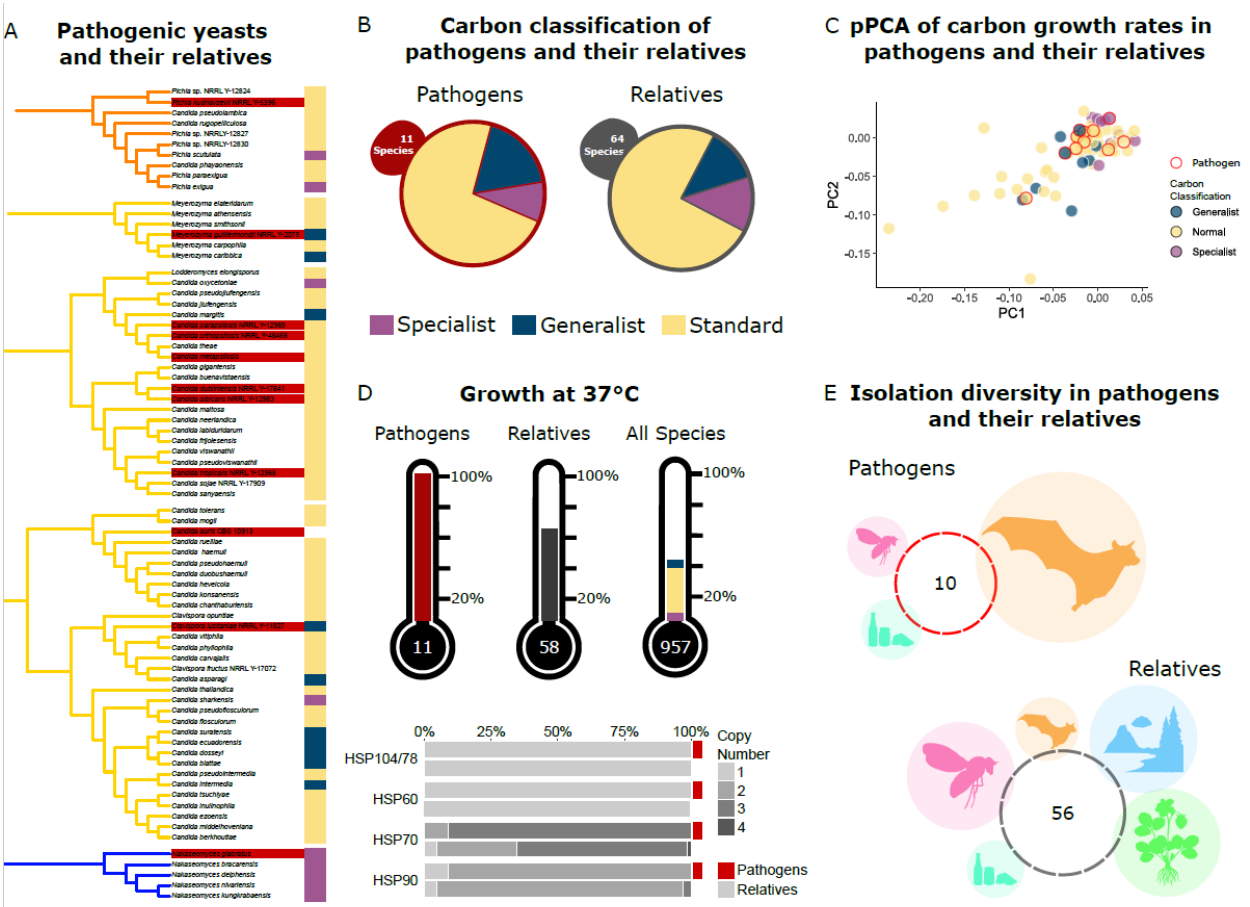



**Figure 5: Carbon generalism and specialism are not associated with yeast pathogenicity.**
**A.** The phylogenetic clades containing human fungal pathogens. Clades reflect all species within a specific phylogenetic distance from the identified pathogen. Pathogens are found in three different orders, 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 component 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, while 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 than their non-pathogenic relatives ($\chi^2$, $p = 0.042$). 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).