

## EVOLUTIONARY BIOLOGY

## Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota

Xing-Xing Shen<sup>1\*</sup>, Jacob L. Steenwyk<sup>2</sup>, Abigail L. LaBella<sup>2</sup>, Dana A. Opulente<sup>3</sup>, Xiaofan Zhou<sup>4</sup>, Jacek Kominek<sup>3</sup>, Yuaning Li<sup>2</sup>, Marizeth Groenewald<sup>5</sup>, Chris T. Hittinger<sup>3</sup>, Antonis Rokas<sup>2\*</sup>

Ascomycota, the largest and most well-studied phylum of fungi, contains three subphyla: Saccharomycotina (budding yeasts), Pezizomycotina (filamentous fungi), and Taphrinomycotina (fission yeasts). Despite its importance, we lack a comprehensive genome-scale phylogeny or understanding of the similarities and differences in the mode of genome evolution within this phylum. By examining 1107 genomes from Saccharomycotina (332), Pezizomycotina (761), and Taphrinomycotina (14) species, we inferred a robust genome-wide phylogeny that resolves several contentious relationships and estimated that the Ascomycota last common ancestor likely originated in the Ediacaran period. Comparisons of genomic properties revealed that Saccharomycotina and Pezizomycotina differ greatly in their genome properties and enabled inference of the direction of evolutionary change. The Saccharomycotina typically have smaller genomes, lower guanine-cytosine contents, lower numbers of genes, and higher rates of molecular sequence evolution compared with Pezizomycotina. These results provide a robust evolutionary framework for understanding the diversity and ecological lifestyles of the largest fungal phylum.

## INTRODUCTION

The fungal phylum Ascomycota is one of most diverse phyla of eukaryotes with at least 83,000 described species that represent more than half of all described species of fungi (1). The Ascomycota is divided into three subphyla. The Saccharomycotina subphylum is a lineage of more than 1000 known species and 12 major clades (2), commonly referred to as budding yeasts. Species in this lineage include the model organism *Saccharomyces cerevisiae* (3) and several notable pathogens, such as the opportunistic human pathogen *Candida albicans* (4) and the multidrug-resistant emerging pathogen *Candida auris* (5). The Pezizomycotina subphylum contains more than 82,000 described species in 16 classes (1), commonly referred to as filamentous fungi. This subphylum contains several major plant and animal pathogens belonging to diverse genera, such as *Fusarium* and *Aspergillus* (6, 7). Last, the Taphrinomycotina subphylum contains ~140 described species in five classes (1), commonly referred to as fission yeasts. This subphylum contains the model organism *Schizosaccharomyces pombe* (8).

To better understand the evolution of species diversity and ecological lifestyles in the Ascomycota, a robust framework of phylogenetic relationships and divergence time estimates is essential (9–11). In the last two decades, several studies have aimed to infer the phylogeny of the Ascomycota either using a handful of gene markers from hundreds of taxa (12–16) or using hundreds of gene markers from tens of taxa (17–20). To date, the most comprehensive “few-markers-from-many-taxa”

phylogeny used a 6-gene, 420-taxon (8 Taphrinomycotina, 16 Saccharomycotina, and 396 Pezizomycotina) data matrix (12), whereas the most comprehensive genome-scale phylogeny used a 238-gene, 496-taxon (12 Taphrinomycotina, 76 Saccharomycotina, and 408 Pezizomycotina) data matrix (21) but was inferred using FastTree, a program that is faster but typically yields phylogenies that have much lower likelihood scores than those obtained by IQ-TREE and RAxML/RAxML-NG (22). Key relationships supported by these studies include the monophyly of each subphylum and class and the sister group relationship of subphyla Saccharomycotina and Pezizomycotina. In contrast, relationships among classes are contentious between studies, particularly with respect to relationships between the 16 classes in Pezizomycotina (1). For example, there is disagreement whether the sister group to the rest of classes in the Pezizomycotina is class Pezizomycetes (13), class Orbiliomycetes (16), or a clade composed of both (18).

Previous molecular clock-based estimates of divergence times for Ascomycota have all been based on few-markers-from-many-taxa data matrices (13, 14, 23–25), resulting in age estimates for key events in Ascomycota evolution that have wide intervals. For example, analysis of a 6-gene, 121-taxon (1 Saccharomycotina, 118 Pezizomycotina, and 2 Taphrinomycotina) data matrix inferred that the origin of the phylum Ascomycota took place 531 million years (Ma) ago [95% credibility interval (CI), 671 to 410 Ma ago] (see their scenario 4) (14), while analysis of a 4-gene, 145-taxon (12 Saccharomycotina, 129 Pezizomycotina, and 4 Taphrinomycotina) data matrix inferred that the phylum originated 588 Ma ago (95% CI, 773 to 487 Ma ago) (13). The sparser taxon sampling of previous studies has prevented estimation of divergence times of several key divergence events of higher taxonomic ranks (23–25) and stymied our understanding of their evolutionary pace. While these studies have substantially advanced our understanding of Ascomycota evolution, a comprehensive, genome-scale phylogeny and timetree stemming from the sampling of hundreds of genes from thousands of taxa from the phylum are still lacking.

A robust phylogenomic framework would also facilitate comparisons of genome evolution across the subphyla of Ascomycota.

<sup>1</sup>Institute of Insect Sciences, Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China. <sup>2</sup>Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA. <sup>3</sup>Laboratory of Genetics, Wisconsin Energy Institute, Center for Genomic Science Innovation, DOE Great Lakes Bioenergy Research Center, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706, USA. <sup>4</sup>Guangdong Laboratory for Lingnan Modern Agriculture, Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou 510642, China. <sup>5</sup>Westerdijk Fungal Biodiversity Institute, 3584 CT, Netherlands.

\*Corresponding author. Email: xingxingshen@zju.edu.cn (X.-X.S.); antonis.rokas@vanderbilt.edu (A.R.)

For example, the three subphyla differ in their genome sizes, with the genomes of Pezizomycotina species being notably larger (~42 Mb) than those of Saccharomycotina (~13 Mb) and Taphrinomycotina (~14 Mb) (26). While several recent studies have analyzed major lineages within the two largest subphyla (in terms of number of described species), Saccharomycotina (2, 19, 27) and Pezizomycotina (28, 29), comparisons of genome evolution across the two subphyla are lacking. For example, a recent analysis of the tempo and mode of genome evolution in 332 Saccharomycotina found evidence of high evolutionary rates and reductive evolution across this subphylum (2), but whether budding yeasts are faster evolving than filamentous fungi remains unknown. However, a recent analysis of 71 Ascomycota genomes showed that Pezizomycotina has much higher levels of gene order divergence than Saccharomycotina (20). Similarly, genome-wide examinations of horizontal gene transfer events in dozens to more than a hundred Ascomycota genomes have revealed that Pezizomycotina acquired substantially higher numbers of genes from prokaryotic donors than did Saccharomycotina (30). Although these studies have contributed to our understanding of certain evolutionary processes in the phylum, we still know relatively little about the evolution of Ascomycota genomes and their properties.

There are currently more than 1000 publicly available genomes from Ascomycota species, which span the diversity of Saccharomycotina (332 genomes representing all 12 major clades), Pezizomycotina (761 genomes representing 9 of 16 classes), and Taphrinomycotina (14 genomes representing 4 of 5 classes) (1107 genomes as of 14 December 2018). These 1107 genomes represent a much larger and representative source of genomic data across the Ascomycota phylum than previously available, providing a unique opportunity to infer a genome-scale phylogeny and timetree for the phylum and compare the mode of genome evolution across its subphyla.

## RESULTS AND DISCUSSION

### A genome-scale phylogeny of the fungal phylum Ascomycota

To infer a genome-scale phylogeny of the Ascomycota, we used 1107 publicly available genomes from species belonging to Ascomycota (Saccharomycotina, 332; Pezizomycotina, 761; and Taphrinomycotina, 14) and six out-groups from the sister fungal phylum Basidiomycota. The numbers of species examined represent ~332 of 1000 (~30%) of all described species of Saccharomycotina, ~761 of 82,000 (~1%) of Pezizomycotina, and ~14 of 140 (~10%) of Taphrinomycotina. All genomes were retrieved from the National Center for Biotechnology Information (NCBI) GenBank database, ensuring that only one genome per species was included (tables S1 and S2). Analysis of genome assembly completeness reveals that 1021 of 1113 (~92%) genomes have more than 90% of the 1315 full-length Benchmarking Universal Single-Copy Orthologs (BUSCO) genes (fig. S1) (31).

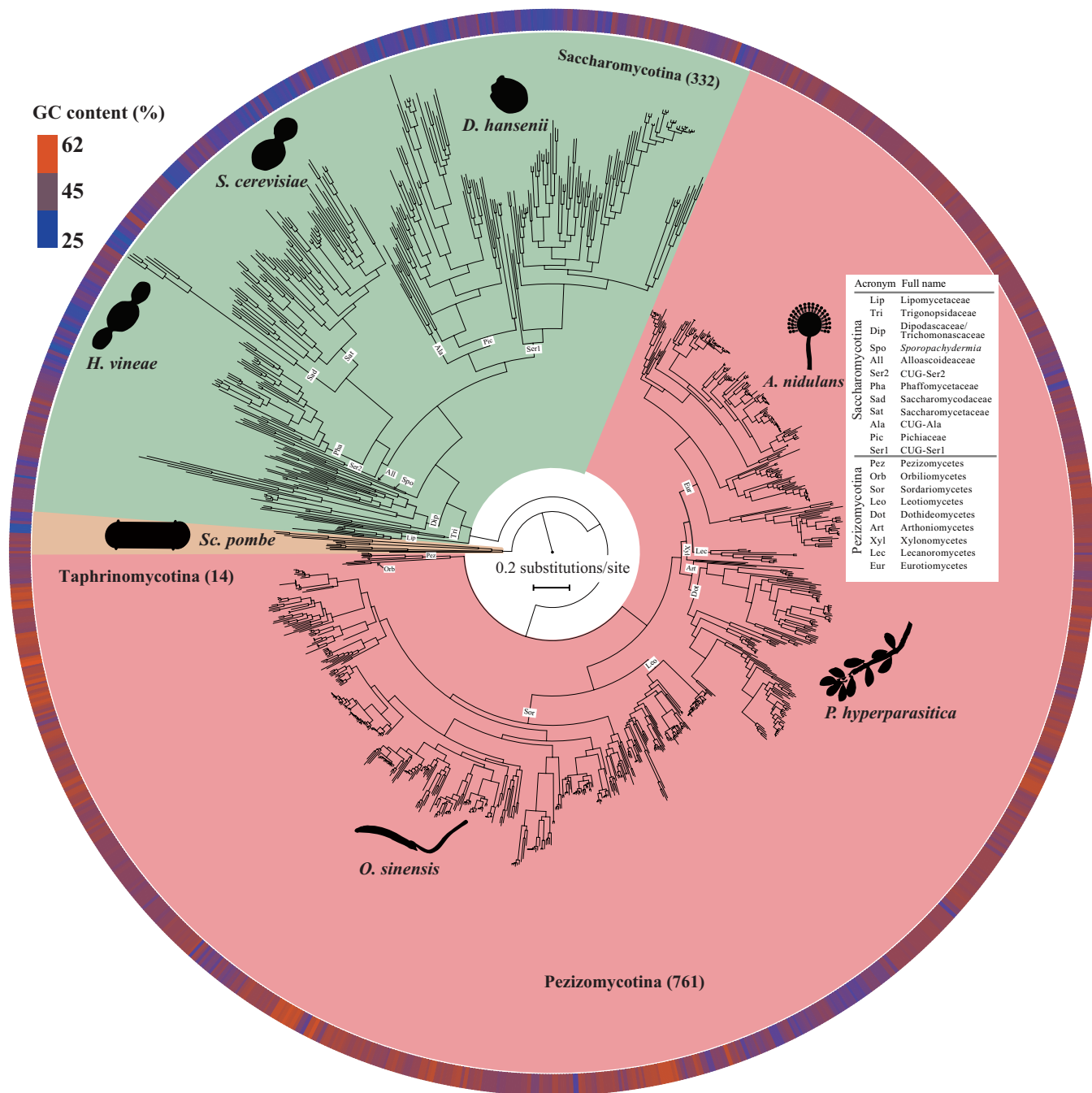
The 1315 BUSCO genes from 1107 Ascomycota species and six out-groups were used to construct a phylogenomic data matrix (see Materials and Methods). After constructing the multiple amino acid sequence alignment and trimming ambiguous regions for each of these 1315 BUSCO genes, we kept only the 815 BUSCO genes that had taxon occupancy of  $\geq 50\%$  for each subphylum (i.e.,  $\geq 7$  Taphrinomycotina,  $\geq 166$  Saccharomycotina, and  $\geq 381$  Pezizomycotina) and whose amino acid sequence alignments were  $\geq 300$  sites in length. In the final set of 815 BUSCO genes, alignment lengths range from

300 to 4585 amino acid sites (average, 690), and numbers of taxa range from 851 to 1098 (average, 1051) (table S3). The final data matrix contains 1107 taxa, 815 genes, and 562,376 amino acid sites.

Inference using concatenation- and coalescent-based approaches yielded a robust, comprehensive phylogeny of the Ascomycota phylum (Fig. 1). The vast majority of internodes in both the concatenation-based (1103 of 1110, 99%) and the coalescent-based phylogeny (1076 of 1110, 97%) received strong ( $\geq 95\%$ ) support and were congruent between the phylogenies inferred using the two approaches; only 46 of 1110 (4%) internodes were incongruent between the two phylogenies (figs. S2 and S3).

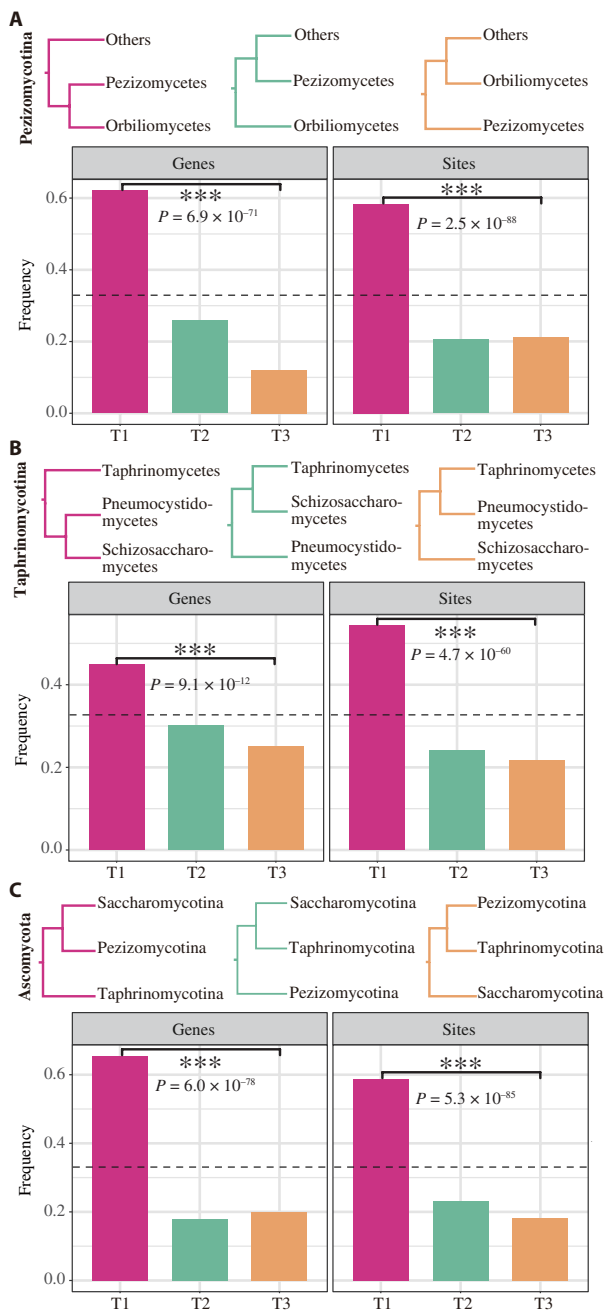
Our higher-level phylogeny of Ascomycota is generally more congruent with previous genome-scale phylogenies (2, 17, 18) than with few-genes-from-many-taxa phylogenies (12–15), particularly with respect to relationships among the nine classes in the subphylum Pezizomycotina. For example, genome-scale studies, including ours, consistently favor a clade consisting of Pezizomycetes and Orbiliomycetes as the sister group to the rest of the Pezizomycotina (17, 18), while studies based on a few genes recovered either Orbiliomycetes (14–16) or Pezizomycetes (13) as the sister class to the rest of the Pezizomycotina (Fig. 2A). Our phylogeny also strongly supported the placement of the class Schizosaccharomycetes, which includes the model organism *S. pombe*, as the sister group to the class Pneumocystidomycetes, which contains the human pathogen *Pneumocystis jirovecii* (Fig. 2B). A recent genome-scale study of 84 fungal genomes showed that our result is consistent with the phylogeny inferred using an alignment-free composition vector approach but not with the phylogeny inferred using maximum likelihood (ML), which instead recovered Schizosaccharomycetes as the sister group to Taphrinomycetes (17). Last, both concatenation- and coalescent-based approaches supported the placement of the subphylum Saccharomycotina as the sister group to the subphylum Pezizomycotina (Fig. 2C). This result is consistent with most previous studies that analyze multiple sequence alignment data (15, 17, 19, 32, 33), but not with a recent study that analyzed genomic data with an alignment-free method and placed the subphylum Saccharomycotina as the sister group to the subphylum Taphrinomycotina (34).

To evaluate whether our genome-scale data matrix robustly resolved the three historically contentious branches discussed in the previous paragraph, we quantified the distribution of phylogenetic signal for alternative topologies of these three phylogenetic hypotheses at the level of genes and sites using an ML framework presented by Shen *et al.* (35). First, we found that the phylogenetic support for each of the three branches stemmed from many genes, i.e., it was not dominated by a small number of genes with highly disproportionate influence (table S4). Second, we found that the topology recovered by both concatenation- and coalescent-based approaches in our study had significantly higher frequencies of supporting genes and supporting sites (*G*-test) than two alternative competing hypotheses, ranging from 0.45 to 0.65, in all three branches examined (Fig. 2, A to C, and tables S4 and S5). Neither of the two alternative conflicting phylogenetic hypotheses for each of the three branches received frequencies of supporting genes and supporting sites that were equal to or greater than one-third (0.33), the value expected if the relationships among the taxa were represented by a polytomy. In summary, we found that very few branches (<5%) were incongruent between concatenation- and coalescent-based phylogenies (figs. S2 and S3). We also found that individual genes and sites exhibited robust phylogenetic signal for specific, historically contentious



Downloaded from <http://advances.sciencemag.org/> on November 4, 2020

**Fig. 1. Maximum likelihood (ML) phylogeny of 1107 taxa in the fungal phylum Ascomycota.** The concatenation-based ML phylogeny (InL = -269,043,834.145) was inferred from a set of 815 BUSCO amino acid genes (total of 562,376 sites) under a single LG + G4 substitution model using IQ-TREE multicore version 1.5.1. The number of species sampled in each subphylum is given in parentheses. Internal branch labels are acronyms for 12 major clades in the subphylum Saccharomycotina and nine classes in the subphylum Pezizomycotina. *S. cerevisiae*, *Saccharomyces cerevisiae*; *H. vineae*, *Hanseniaspora vineae*; *D. hansenii*, *Debaryomyces hansenii*; *O. sinensis*, *Ophiocordyceps sinensis*; *P. hyperparasitica*, *Pseudovirgaria hyperparasitica*; *A. nidulans*, *Aspergillus nidulans*. Images representing taxa were drawn by hand, taken from PhyloPic (<http://phylopic.org>), or modified from Google Images. The bar next to each species indicates the guanine-cytosine (GC) content. On average, lineages in the subphylum Saccharomycotina have significantly lower GC content (49.6% versus 40.6%; Wilcoxon rank-sum test;  $P = 3.07 \times 10^{-103}$ ) but higher evolutionary rate (1.80 substitutions per site versus 1.12 substitutions per site; Wilcoxon rank-sum test;  $P = 6.57 \times 10^{-126}$ ) compared with lineages in the subphylum Pezizomycotina. The complete phylogenetic relationships of 1107 taxa are given in fig. S2 and in the Figshare repository. For easy determination of the relationships among any subset of taxa, the phylogeny is also available through Treehouse (63).



**Fig. 2. Distribution of phylogenetic signal for three historically contentious relationships within Ascomycota.** For each relationship/internal branch [(A) Which class(es) is the sister group to the rest of the Pezizomycotina? (B) What is the relationship among three classes Schizosaccharomycetes, Pneumocystidomycetes, and Taphrinomycetes in the subphylum Taphrinomycotina? (C) What is the relationship among three subphyla Pezizomycotina, Saccharomycotina, and Taphrinomycotina in the phylum Ascomycota?], we applied the framework presented by Shen *et al.* (35) to examine proportions of genes (left) and sites (right) supporting each of three competing hypotheses (topology 1 or T1 in red, topology 2 or T2 in green, and topology 3 or T3 in yellow). Note that both concatenation- and coalescent-based approaches supported T1 in our study. Dashed horizontal lines on one-third y axis value denote expectation of proportion of genes/sites under a polytomy scenario. The G-test was used to test whether the sets of three values are significantly different ( $***P \leq 0.001$ ). All values are given in tables S4 and S5. Input and output files associated with phylogenetic signal estimation are also deposited in the Figshare repository.

branches (Fig. 2, A to C). These results suggest that genome-scale amounts of data from a comprehensive sampling of taxa will robustly resolve major lineages of the tree of life.

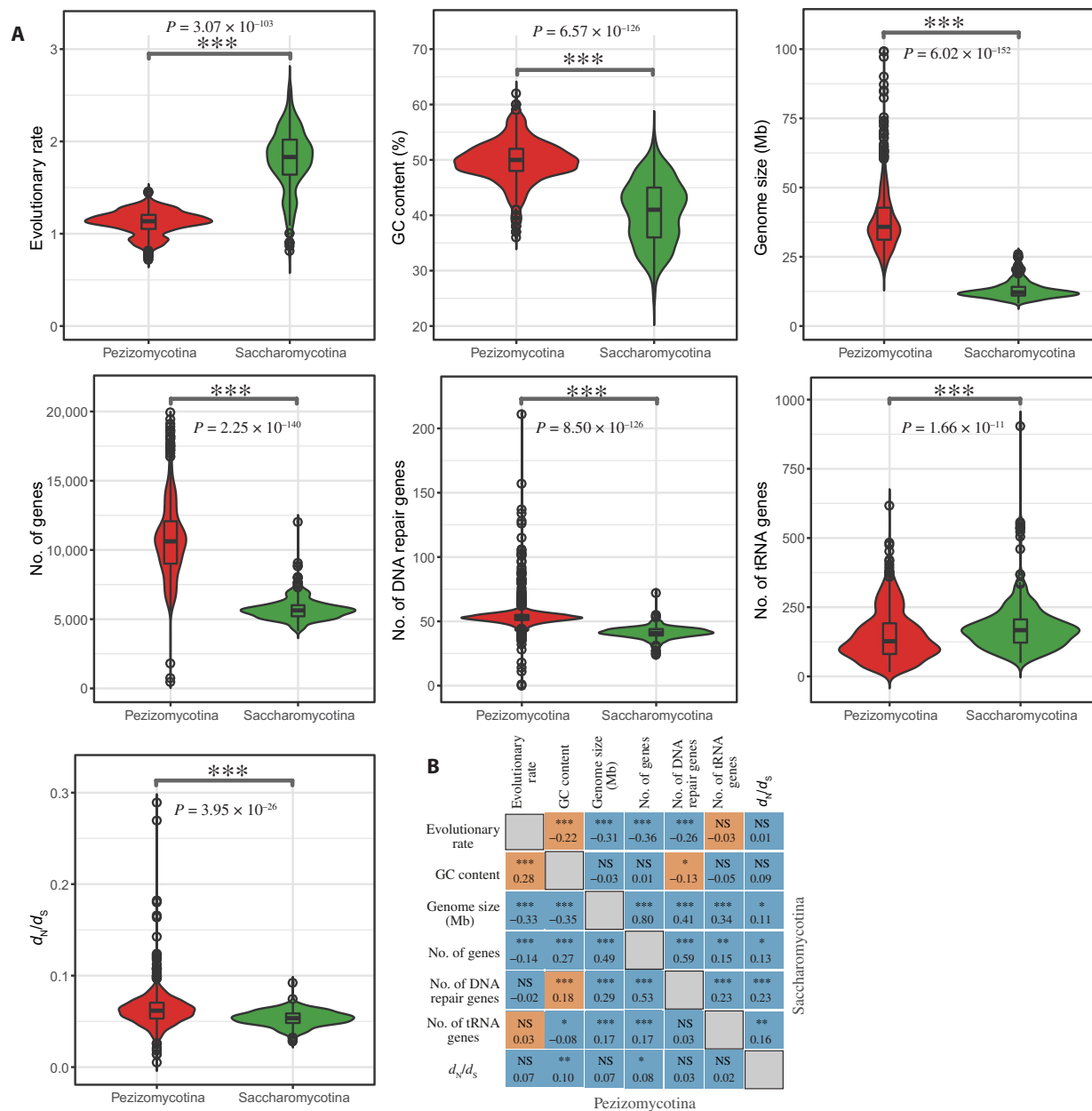
### A genome-scale evolutionary timetree of the fungal phylum Ascomycota

We next used the robust phylogeny, a relaxed molecular clock approach, and six widely accepted time calibration nodes (see Materials and Methods) to infer the time scale of evolution of Ascomycota. We inferred the origin of the phylum to have taken place 563 Ma ago (95% CI, 631 to 495 Ma ago), the origin of the subphylum Saccharomycotina 438.4 Ma ago (CI, 590 to 304 Ma ago), the origin of the subphylum Pezizomycotina 407.7 Ma ago (CI, 631 to 405 Ma ago), and the origin of Taphrinomycotina crown group 530.5 Ma ago (CI, 620 to 417 Ma ago). Notably, the taxonomic placement of all budding yeasts into a single class, Saccharomycetes, whose origin coincides with the origin of the subphylum Saccharomycotina, means that the last common ancestor of this sole class of budding yeasts is much more ancient than those of any of the nine classes (based on current taxon sampling) in the subphylum Pezizomycotina (fig. S4 and table S6). For example, the most ancient class in Pezizomycotina is Pezizomycetes, whose origin is dated 247.7 Ma ago (CI, 475 to 193 Ma ago) (fig. S4 and table S6). The other outlier, albeit with much larger confidence intervals, is class Neolectomycetes in Taphrinomycotina, which we estimate to have originated 480.4 Ma ago (CI, 607 to 191 Ma ago) (fig. S4 and table S6).

Comparison of our inferred dates of divergence to recent studies using a few genes (11, 13, 14) shows that our estimates are generally younger. For example, our estimate for the phylum Ascomycota is around 563 Ma ago, whereas the origin of the Ascomycota was dated at least 533 Ma ago by Hyde *et al.* (9). However, recent studies that used 18S and 28S ribosomal RNA markers from 111 fungal taxa (11) and four gene markers from 145 Ascomycota taxa (13) inferred the origin of the Ascomycota to have taken place around 606 and 588 Ma ago, respectively. This result is consistent with findings of previous studies [e.g., (36)], where inclusion of large numbers of genes was found to also result in younger estimates of divergence times, perhaps because of the influence of larger amounts of data in decreasing the stochastic error involved in date estimation. In summary, generation of a genome-scale timetree for more than 1000 ascomycete species spanning the diversity of the phylum provides a robust temporal framework for understanding and exploring the origin and diversity of Ascomycota lifestyles.

### Contrasting modes of genome evolution in fungal phylum Ascomycota

To begin to understand the similarities and differences in the modes of genome evolution between subphyla, we assessed seven different genomic properties for each species. Because the number of genomes from Taphrinomycotina taxa is substantially smaller (14) than Saccharomycotina (332 genomes) and Pezizomycotina (761) (Fig. 3), we focused on examining the evolution of different genomic properties between Saccharomycotina and Pezizomycotina. Specifically, we found that Saccharomycotina exhibited a 1.6-fold higher evolutionary rate (on average, 1.80 substitutions per site in Saccharomycotina versus 1.12 substitutions per site in Pezizomycotina), 1.24-fold lower guanine-cytosine (GC) content (40% versus 50%), 3-fold smaller genome size (13 Mb versus 39 Mb), 1.9-fold lower number of protein-coding genes (5734 versus 10,847), 1.3-fold lower



**Fig. 3. Contrasting patterns for seven genomic properties between Pezizomycotina and Saccharomycotina.** As the subphylum Taphrinomycotina (no. of species = 14) has a much smaller number of species than the subphylum Saccharomycotina (no. of species = 332) and the subphylum Pezizomycotina (no. of species = 761) in our dataset, we focused our analyses on the comparisons of seven genome properties between Saccharomycotina and Pezizomycotina. **(A)** For each species in Pezizomycotina (colored in red,  $n = 761$ ) and Saccharomycotina (colored in green,  $n = 332$ ), we calculated evolutionary rate, GC content, genome size, number of protein-coding genes, number of DNA repair genes, number of tRNA genes, and  $d_N/d_S$  (see Materials and Methods for details). The Wilcoxon rank-sum test was used to test whether the sets of values in two subphyla are significantly different. **(B)** Pairwise standard Pearson's correlation coefficient among pairs of the seven genomic properties was conducted using R 3.4.2 for Pezizomycotina (lower diagonal) and Saccharomycotina (upper diagonal), respectively. For each cell, the top value corresponds to the  $P$  value (NS,  $P > 0.05$ ; \* $P \leq 0.05$ ; \*\* $P \leq 0.01$ ; \*\*\* $P \leq 0.001$ ), whereas the bottom value corresponds to Pearson's coefficient value. Orange cells denote instances where correlation trends in Pezizomycotina and Saccharomycotina are in opposite directions, whereas blue cells denote instances where the trends are in the same direction. The detailed values of all seven properties in Pezizomycotina and Saccharomycotina are given in table S7. The correlations among these seven properties are largely consistent before (i.e., standard Pearson's correlations) and after (i.e., phylogenetically independent contrasts) accounting for correlations due to phylogeny (see table S8).

number of DNA repair genes (41 versus 54), 1.2-fold higher number of tRNA genes (179 versus 146), and 1.3-fold smaller estimates of nonsynonymous to synonymous substitution rate ratio ( $d_N/d_S$ ) (0.053 versus 0.063), compared with Pezizomycotina (Fig. 3A, Table 1, and table S7). In addition, we found that gene density (that is, the

number of genes divided by genome size) in Saccharomycotina is much higher than that in Pezizomycotina (on average, 451 genes per megabase in Saccharomycotina versus 290 genes per megabase in Pezizomycotina), which is in agreement with previous work showing that gene density is inversely correlated with genome size (37).

**Table 1. Summary of values for seven genomic properties in extant Saccharomycotina and Pezizomycotina and in the last common ancestors of Saccharomycotina and Pezizomycotina.**

Property	Extant Saccharomycotina* (n = 332)	Extant Pezizomycotina* (n = 761)	Saccharomycotina ancestor	Pezizomycotina ancestor	Mean difference between two extant lineages	Difference between two ancestors
Evolutionary rate (amino acid substitutions per site)	1.80 ± 0.33	1.12 ± 0.13	1.1	0.9	0.68	0.2
GC content (%)	40 ± 5	50 ± 4	43	47	10	4
Genome size (Mb)	13 ± 3	39 ± 15	23	42	26	19
No. of genes	5734 ± 834	10,847 ± 2,919	7000	9400	5113	2400
No. of DNA repair genes	41 ± 5	54 ± 13	44	52	13	8
No. of tRNA genes	179 ± 92	146 ± 89	160	170	33	10
$d_N/d_S$	0.053 ± 0.008	0.063 ± 0.021	0.052	0.058	0.01	0.006

\*Average value ± SD. The detailed values of all seven properties for each taxon are given in table S7.

Analysis of standard Pearson's correlations among the seven genomic properties revealed that two pairs exhibited statistically significant contrasting patterns between Saccharomycotina and Pezizomycotina. Specifically, evolutionary rate shows negative correlation with GC content in Saccharomycotina but positive correlation in Pezizomycotina, and GC content shows negative correlation with number of DNA repair genes in Saccharomycotina but positive correlation in Pezizomycotina (Fig. 3B). These correlations are largely consistent before (i.e., standard Pearson's correlations) and after (i.e., phylogenetically independent contrasts) accounting for correlations due to phylogeny (table S8).

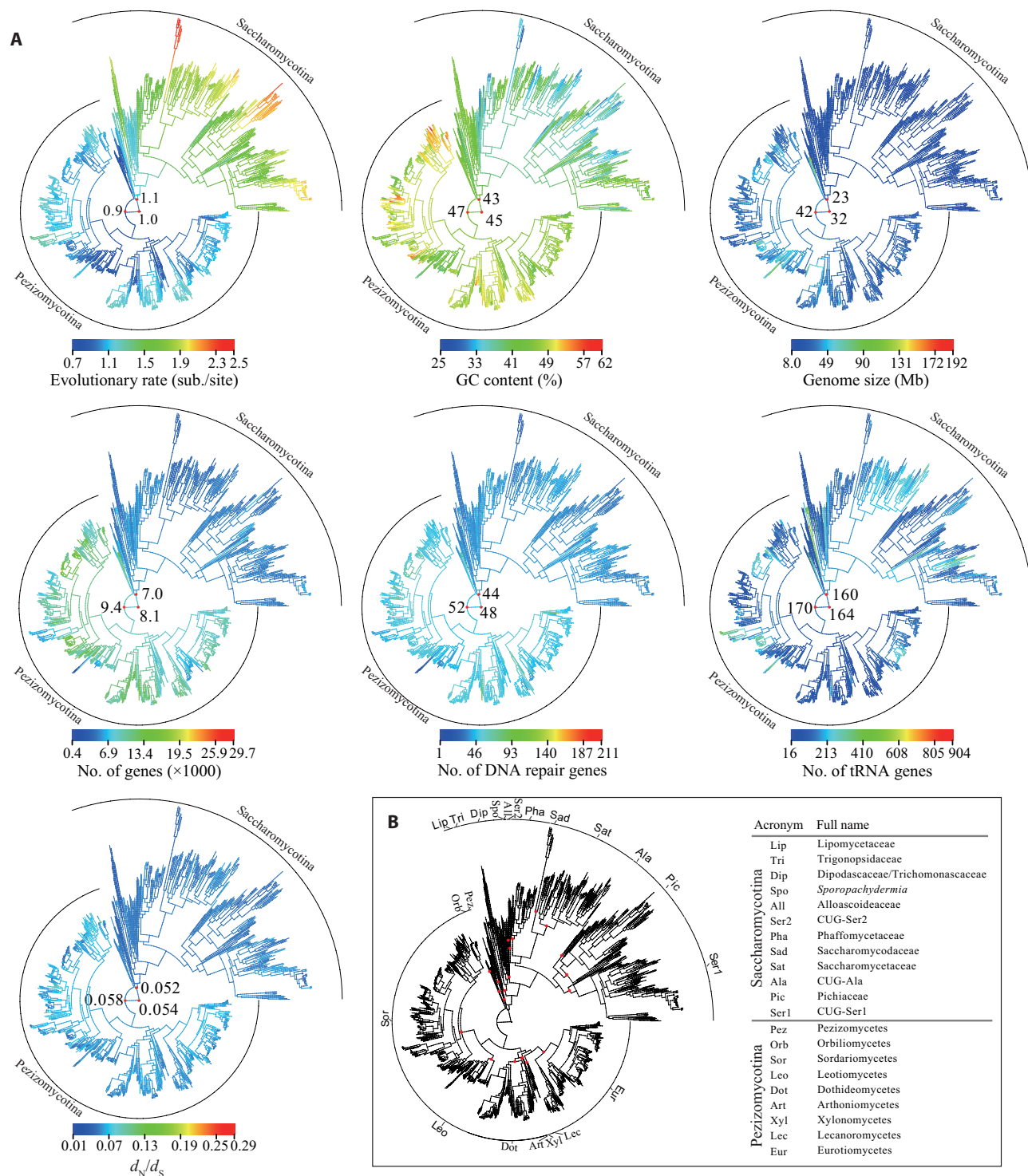
For each of the seven properties, we used our genome-scale phylogeny (Fig. 1) to infer the ancestral character states and reconstruct their evolution in the Saccharomycotina ancestor and the Pezizomycotina ancestor. Comparison of ancestral states along branches on the Saccharomycotina part of the phylogeny to those on the Pezizomycotina part of the phylogeny showed that all genomic properties, except the number of tRNA genes, exhibited different modes of evolution (Fig. 4 and Table 1). For example, most Saccharomycotina branches exhibit evolutionary rates of at least 1.0 amino acid substitutions per site, whereas those of Pezizomycotina exhibit evolutionary rates between 0.7 and 1.4 substitutions per site (Fig. 4A). However, the inferred values for these properties of the last common ancestors of Saccharomycotina and Pezizomycotina are quite similar. For example, the inferred state values for the Saccharomycotina last common ancestor and the Pezizomycotina last common ancestor are 1.1 and 0.9 substitutions per site for evolutionary rate and 43 and 47% for GC content (Table 1), respectively. The same trends are also observed across lineages, such as Lipomycetaceae, which is the sister group to the rest of the Saccharomycotina, and the clade consisting of Pezizomycetes and Orbiliomycetes, which is the sister group to the rest of the Pezizomycotina (Fig. 4, A and B).

Comparison of the trait values for the seven genome properties between extant Saccharomycotina and Pezizomycotina branches to those of the Saccharomycotina and Pezizomycotina last common ancestors showed that evolutionary rate, GC content, genome size, and number of protein-coding genes were the properties with the highest amounts of evolutionary change (Figs. 3 and 4 and Table 1). Ancestral state reconstruction also enabled inference of the direction

of evolutionary change for each of the evolutionary properties. For example, the Saccharomycotina and Pezizomycotina last common ancestors, as well as branches in Lipomycetaceae and branches across Pezizomycotina, exhibit similar evolutionary rates, whereas the rest of the nodes and branches in the Saccharomycotina part of the phylogeny exhibit much higher evolutionary rates. This pattern suggests that the higher levels of genomic diversity in Saccharomycotina stem from an acceleration of evolutionary rate that occurred within the subphylum, after the divergence of Lipomycetaceae from the rest of the Saccharomycotina (Fig. 4, A and B).

Why do Saccharomycotina exhibit higher evolutionary rates compared with Pezizomycotina? Studies in other lineages, such as vertebrate animals (38), have previously shown that evolutionary rate is positively associated with generation time. Assuming that mutation rates are equal, species with shorter generation times will replicate their genomes more frequently, accruing more mutations per unit time. While the generation times of most fungi in our phylogeny are unknown, the generation times of model organisms in Saccharomycotina are thought to be shorter than those in Pezizomycotina. For example, the doubling time of the budding yeasts *S. cerevisiae* and *C. albicans* under optimal conditions is 90 min (39), while that of the filamentous fungi *Aspergillus nidulans* and *Neurospora crassa* is between 2 and 3 hours (40, 41). It is also well established that absence or loss of DNA repair genes can lead to an increase in genomes' mutation rates (42). We found that Saccharomycotina have, on average, fewer DNA repair genes (41) than Pezizomycotina (54) (Fig. 3 and Table 1). Thus, an alternative but not mutually exclusive explanation for the higher evolutionary rates in Saccharomycotina may stem at least partially from the presence of a smaller set of genes involved in DNA repair in their genomes than in the genomes of Pezizomycotina. Last, other life history traits (e.g., smaller cell size, faster metabolism, and larger population size) that have been associated with variation in the rate of molecular evolution (43) might also contribute to higher evolutionary rates of the Saccharomycotina.

Variation in genomic GC content has historically been of broad interest in biology. Average GC content values of different genomic regions (e.g., intergenic regions and protein-coding regions) in Saccharomycotina are consistently lower than those in Pezizomycotina (figs. S5 and S6A). Similarly, gene-wise average estimates of GC



**Fig. 4. Contrasting modes of genome evolution in Pezizomycotina and Saccharomycotina.** (A) For each of the seven genomic properties examined (see Materials and Methods for details), we reconstructed them as continuous traits on the species phylogeny (Fig. 1) and visualized their ancestral states with the R package phytools v0.6.44 (62). Heatmap bars denote ancestral state values from small (blue) to large (red). Three ancestral state values next to three red dots are shown for the ancestor of the subphyla Pezizomycotina and Saccharomycotina, the ancestor of the subphylum Pezizomycotina, and the ancestor of the subphylum Saccharomycotina, respectively. Sub., substitution. (B) Phylogeny key showing the placement of the 21 nodes representing the last common ancestors of the 12 major clades in the subphylum Saccharomycotina and of the nine classes in the subphylum Pezizomycotina; the 21 nodes are indicated by the red dots. The orders of branches in (A) are identical to those in (B).

content showed that all 815 BUSCO genes in Saccharomycotina have lower GC content values than those in Pezizomycotina (fig. S6B). Moreover, we found that the frequencies of amino acids encoded by GC-rich codons in Saccharomycotina are much lower than those of amino acids encoded by GC-rich codons in Pezizomycotina (fig. S7). Ancestral state reconstruction of genomic GC content along branches on the phylogeny shows that the last common ancestors of the Saccharomycotina and Pezizomycotina, as well as branches in Lipomycetaceae and branches in classes Pezizomycetes and Orbiliomycetes, exhibit intermediate GC content of around 45%. In contrast, GC content of most branches within the rest of Saccharomycotina (i.e., all major clades of Saccharomycotina, including extant taxa, except Lipomycetaceae) evolved toward 40%, while GC content within the rest of Pezizomycotina (i.e., all classes, including extant taxa, except Pezizomycetes and Orbiliomycetes) evolved toward 50%. This pattern suggests that the evolution of lower overall GC content in Saccharomycotina occurred after the divergence of Lipomycetaceae from the rest of Saccharomycotina and that the evolution of higher overall GC content in Pezizomycotina occurred after the divergence of the clade consisting of Pezizomycetes and Orbiliomycetes from the rest of Pezizomycotina (Fig. 4, A and B).

Why are Pezizomycotina genomes more GC rich compared with Saccharomycotina genomes? There are two possible explanations. The first one is that mutational biases have skewed the composition of Saccharomycotina genomes toward AT (adenine-thymine) content (44). For example, Steenwyk *et al.* showed that *Hanseniaspora* budding yeasts with higher AT content lost a greater number of DNA repair genes than those with lower AT content (42), suggesting that the loss of DNA repair genes is associated with AT richness. Consistent with these results, we found that Pezizomycotina genomes contain a higher number of DNA repair genes than Saccharomycotina (Fig. 3 and Table 1). The second potential, not necessarily mutually exclusive, explanation is that mutational biases have skewed Pezizomycotina genomes toward GC richness. It was recently shown that increasing GC-biased gene conversion (gBGC), a process associated with recombination that favors the transmission of GC alleles over AT alleles, can result in a systematic underestimate of  $d_N/d_S$  in birds (45). If this is true for Ascomycota, because of the higher GC content of Pezizomycotina genomes, we would expect that their  $d_N/d_S$  would be underestimated due to the higher levels of gBGC compared with Saccharomycotina. Consistent with this expectation, by calculating differences in  $d_N/d_S$  before and after accounting for gBGC across 815 codon-based BUSCO genes, we found that the underestimate of  $d_N/d_S$  in Pezizomycotina is twofold higher than that in Saccharomycotina (Pezizomycotina: average of differences in  $d_N/d_S = 0.004$ ; Saccharomycotina: average of differences in  $d_N/d_S = 0.002$ ) (fig. S8).

## CONCLUSION AND FUTURE WORK

In this study, we took advantage of the recent availability of the genome sequences of 1107 Ascomycota species from Saccharomycotina (332), Pezizomycotina (761), and Taphrinomycotina (14) to infer a genome-scale phylogeny and timetree for the phylum and compare the mode of genome evolution across its subphyla. Although we collected publicly available genomes for the greatest possible taxonomic sampling as of 14 December 2018, we note that the genome sequences of the 1107 taxa sampled are unevenly distributed across the three subphyla of Ascomycota. Furthermore, our sample of taxa did not contain novel species found in environmental samples

(46). Similarly, while our data matrix includes 815 genes, thousands of additional homologous genes could potentially be added. Therefore, it is possible that both the inference of the species phylogeny and the estimation of divergence times of Ascomycota may change as more taxa and genes are added.

These caveats notwithstanding, leveraging genome-scale amounts of data from the most comprehensive taxon set to date enabled us to test the robustness of our inference for several contentious branches, potentially resolving controversies surrounding key higher-level relationships within the Ascomycota phylum. For example, our study robustly supported Saccharomycotina as the sister group to Pezizomycotina, and a clade composed of classes Pezizomycetes and Orbiliomycetes as the sister group to the rest of the Pezizomycotina. Our first genome-scale timetree suggests the last common ancestor of Ascomycota likely originated in the Ediacaran period. Examination of mode of genome evolution revealed that Saccharomycotina, which contains the single currently described class Saccharomycetes, and Pezizomycotina, which contains 16 classes, exhibited greatly contrasting evolutionary processes for seven genomic properties, in particular for evolutionary rate, GC content, and genome size. Our results provide a robust evolutionary framework for understanding the fundamental genetic and ecological processes that have generated the biodiversity of the largest fungal phylum.

## MATERIALS AND METHODS

### Data collection

To collect the greatest possible set of genome representatives of the phylum Ascomycota as of 14 December 2018, we first retrieved the 332 publicly available Saccharomycotina yeast genomes (<https://doi.org/10.6084/m9.figshare.5854692>) from a recent comprehensive genomic study of the Saccharomycotina yeasts (2). We then used “Pezizomycotina” and “Taphrinomycotina” as search terms in NCBI’s Genome Browser ([www.ncbi.nlm.nih.gov/genome/browse#!/eukaryotes/Ascomycota](http://www.ncbi.nlm.nih.gov/genome/browse#!/eukaryotes/Ascomycota)) to obtain the basic information of strain name, assembly accession number, assembly release date, assembly level (e.g., contig and scaffold), and GenBank File Transfer Protocol (FTP) access number for draft genomes from the subphyla Pezizomycotina and Taphrinomycotina, respectively. For species with multiple isolates sequenced, we only included the genome of the isolate with the highest assembly level and the latest release date. We next downloaded genome assemblies from GenBank data via FTP access number (<https://ftp.ncbi.nlm.nih.gov/genomes/>). Collectively, we included 332 species representing 77 genera and all 12 major clades of the subphylum Saccharomycotina, 761 species representing 257 genera and nine classes of the subphylum Pezizomycotina, and 14 species representing 6 genera and four classes of the subphylum Taphrinomycotina. Last, we used the genomes of six representatives of the phylum Basidiomycota as out-groups. To check whether our sample of taxa were biased toward particular ecological lifestyles, we obtained the trophic categories for the 1107 Ascomycota species used in this study [symbiotrophs (e.g., ectomycorrhizae and lichens), pathotrophs (e.g., biotrophs, parasites, and pathogens), and saprotrophs (e.g., wood rotters and litter rotters)] from the FUNGuild database (47). We found that Saccharomycotina and Pezizomycotina, the two subphyla with the largest numbers of taxa sampled, contained taxa from all three trophic categories in similar proportions. Specifically, the numbers of Saccharomycotina species classified as symbiotrophs, pathotrophs, saprotrophs, and at least two trophic categories are



2 of 154 (1%), 45 of 154 (29%), 47 of 154 (30%), and 60 of 154 (40%), respectively; the numbers of Pezizomycotina species classified as symbiotrophs, pathotrophs, saprotrophs, and at least two trophic categories are 43 of 702 (6%), 183 of 702 (26%), 176 of 702 (25%), and 300 of 702 (43%), respectively. Detailed information on taxonomy and source of the 1107 Ascomycota and six out-group genomes in our study are provided in tables S1, S2, and S7.

### Assessment of genome assemblies and phylogenomic data matrix construction

To assess the quality of each of the 1113 genome assemblies, we used the BUSCO version 3.0.2 (31). Each assembly's completeness was assessed on the basis of the presence/absence of a set of 1315 predefined orthologs (referred to as BUSCO genes) from 75 genomes in the OrthoDB version 9 database from the Ascomycota database, as described previously (27). In brief, for each BUSCO gene, its consensus orthologous protein sequence among the 75 reference genomes was used as query in a tBLASTn search against each genome to identify up to three putative genomic regions, and the gene structure of each putative genomic region was predicted by AUGUSTUS v3.2.2 (48). Next, the sequences of these predicted genes were aligned to the hidden Markov model profile of the BUSCO gene. BUSCO genes in a given genome assembly were considered as single copy, "full-length" if there was only one complete predicted gene present in the genome; duplicated, "full-length" if there were two or more complete predicted genes present in the genome; "fragmented" if the predicted gene was shorter than 95% of the aligned sequence lengths from the 75 reference species; and "missing" if there was no predicted gene present in the genome.

To construct the phylogenomic data matrix, we started with the set of 1315 single-copy, full-length BUSCO genes from 1107 representatives of the phylum Ascomycota and six out-groups. For each BUSCO gene, we first translated nucleotide sequences into amino acid sequences, taking into account the different usage of the CUG (cytosine-uracil-guanine) codon in Saccharomycotina (2). Next, we aligned the amino acid sequences using Multiple Alignment using Fast Fourier Transform (MAFFT) v7.299b (49) with the options "--thread 4 --auto --maxiterate 1000" and trimmed amino acid alignments using the trimAl v1.4.rev15 (50) with the options "--gappyout --col-numbering." We mapped the nucleotide sequences on the trimmed amino acid alignment based on the column numbers in the original alignment and to generate the trimmed codon-based nucleotide alignment. Last, we removed BUSCO gene alignments whose taxon occupancy (i.e., percentage of taxa whose sequences were present in the trimmed amino acid alignment) was <50% for each subphylum (i.e., <7 Taphrinomycotina, <166 Saccharomycotina, and <381 Pezizomycotina) or whose trimmed alignment length was <300 amino acid sites. These filters resulted in the retention of 815 BUSCO gene alignments, each of which had  $\geq 50\%$  taxon occupancy for each subphylum and alignment length  $\geq 300$  amino acid sites.

### Phylogenetic analysis

For each of the 815 BUSCO genes, we first inferred its best-fitting amino acid substitution model using IQ-TREE multithread version 1.6.8 (51) with options "-m TEST -mrate G4" with the Bayesian information criterion. We then inferred best-scoring ML gene tree under 10 independent tree searches using IQ-TREE. The detailed parameters for running each gene were kept in log files (see the Figshare repository). We inferred the concatenation-based ML tree using IQ-

TREE on a single node with 32 logical cores under a single "LG + G4" model with the options "-seed 668688 -nt 32 -mem 220G -m LG+G4 -bb 1000", as 404 of 815 genes favored "LG + G4" as best-fitting model (table S3). We also inferred the coalescent-based species phylogeny with ASTRAL-III version 4.10.2 (52) using the set of 815 individual ML gene trees. The reliability of each internal branch was evaluated using 1000 ultrafast bootstrap replicates and local posterior probability in the concatenation- and coalescence-based species trees, respectively. We visualized phylogenetic trees using the R package ggtree v1.10.5 (53).

We used the RelTime method, as implemented in the command line version of MEGA7 (54), to estimate divergence times. The concatenation-based ML tree with branch lengths was used as the input tree. Six time calibration nodes, which were retrieved from the TimeTree database (55), were used for molecular dating analyses: the *S. cerevisiae*-*Saccharomyces uvarum* split (14.3 to 17.94 Ma ago), the *S. cerevisiae*-*Kluyveromyces lactis* split (103 to 126 Ma ago), the *S. cerevisiae*-*C. albicans* split (161 to 447 Ma ago), the origin of the subphylum Saccharomycotina (304 to 590 Ma ago), the *S. cerevisiae*-*Saitoella complicata* split (444 to 631 Ma ago), and the origin of the subphylum Pezizomycotina (at least 400 Ma ago) based on the *Paleopyrenomycites devonicus* fossil (10, 56).

To explore whether divergence time estimates were robust when using different methods, we also inferred divergence times using a Bayesian MCMCTree analysis (57). As the MCMCTree approach is computationally intractable for very large phylogenomic data matrices like the one we assembled for this study, we pruned the full data matrix to retain 48 representatives of 25 major groups (four classes in Taphrinomycotina, nine classes in Pezizomycotina, and 12 major clades in Saccharomycotina) so that we can estimate the divergence times of 47 nodes on the backbone of the Ascomycota phylogeny. We first estimated branch lengths under a single LG + G4 model with codeml in the PAML 4.9e package (57) and obtained a rough mean of the overall mutation rate. Next, we applied the approximate likelihood method to estimate the gradient vector and Hessian matrix with Taylor expansion (option usedata = 3). Last, we assigned (i) the Gamma-Dirichlet prior for the overall substitution rate (option rgene\_gamma) as G(1, 8.36), with a mean of 0.11 (meaning  $11 \times 10^{-10}$  amino acid substitutions per site per year), (ii) the Gamma-Dirichlet prior for the rate-drift parameter (option sigma2\_gamma) as G(1, 4.5), and (iii) the parameters for the birth-death sampling process with birth and death rates  $\lambda = \mu = 1$  and sampling fraction  $\rho = 0$ . We used the autocorrelated-rate model (option clock = 3) to account for the rate variation across different lineages and used soft bounds (left and right tail probabilities equal 0.025) to set minimum and maximum values for the four calibration splits mentioned above. The Markov chain Monte Carlo (MCMC) run was first run for 100,000 iterations as burn-in, then sampled every 500 iterations until a total of 3000 samples were collected. Our results showed that divergence time estimates for the 47 backbone nodes between the RelTime and MCMCTree methods are very similar, and their correlation is highly significant (Pearson's correlation coefficient  $r = 0.9907$ ;  $P < 0.00001$ ) (fig. S9). Furthermore, the average percentage of deviation (that is the average percentage of the absolute deviation from a mean point) between the two methods is small ( $\sim 3\%$ ) (table S6).

### Examination of seven genome properties

As the subphylum Taphrinomycotina (no. of species = 14) has a much smaller number of species than the subphylum Saccharomycotina

(no. of species = 332) and the subphylum Pezizomycotina (no. of species = 761) in our dataset, we focused our analyses on the comparisons of seven genome properties (evolutionary rate, GC content, genome size, number of genes, number of DNA repair genes, number of tRNA genes, and  $d_N/d_S$ ) between Saccharomycotina and Pezizomycotina. Specifically, for a given taxon, (i) evolutionary rate is a sum of path distances from the most common ancestor of the subphyla Saccharomycotina and Pezizomycotina to its tip on the concatenation-based ML tree (Fig. 1); (ii) GC content is the percentage of GC nucleotides in genome; (iii) genome size is the total number of base pairs in genome in megabases; (iv) number of genes is the number of protein-coding genes in genome [the gene structure was predicted with AUGUSTUS v3.3.1 (48) on *Aspergillus fumigatus* and *S. cerevisiae* S288C trained models for Pezizomycotina and Saccharomycotina, respectively]; (v) number of DNA repair genes was estimated by counting the number of unique protein-coding genes with Gene Ontology terms related to DNA repair using InterProScan version 5 (58); (vi) the number of tRNA genes is the number of tRNA genes inferred to be present using the tRNAscan-SE 2.0 program (59); and (vii)  $d_N/d_S$  was estimated by calculating the average of the ratio of the expected numbers of nonsynonymous ( $d_N$ ) and synonymous substitutions ( $d_S$ ) along the external branch (tip) of a given taxon across 815 trimmed codon-based BUSCO gene alignments and the concatenation-based ML tree. Note that the concatenation-based ML tree was pruned to match the taxon set in the codon-based alignment if the concatenation-based ML tree contained taxa that were absent from the codon-based alignment. We used the YN98 (F3X4) codon model and the free ratio model using bppml and MapNH in the bio++ libraries, following the study by Bolívar *et al.* (45).

### Statistical analyses

All statistical analyses were performed in R v. 3.4.2 (R core team 2017). Pearson's correlation coefficient was used to test for correlations among seven variables. To account for phylogenetic relationships of species in correlation analysis, we used the R package ape v5.1 (60) to compute phylogenetically independent contrasts following the method described by Felsenstein (61).

### Ancestral state reconstruction

To reconstruct ancestral character states for each of seven continuous properties, we used the R package phytools v0.6.44 function contMap (62) to infer ancestral character states across internal nodes using the ML method with the function fastAnc and to interpolate the states along each edge using Eq. 2 of Felsenstein (61). The input tree was derived from the concatenation-based ML with branch lengths, which was then pruned to keep the 1093 taxa from the subphyla Pezizomycotina and Saccharomycotina.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/45/eabd0079/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

1. T. Y. James, J. E. Stajich, C. T. Hittinger, A. Rokas, Toward a fully resolved fungal tree of life. *Annu. Rev. Microbiol.* **74**, 291–313 (2020).
2. X.-X. Shen, D. A. Ofulente, J. Kominek, X. Zhou, J. L. Steenwyk, K. V. Buh, M. A. 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.-i. 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**, 1533–1545.e20 (2018).
3. J. Peter, M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre, K. Freel, A. Llored, C. Cruaud, K. Labadie, J.-M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemaître, P. Wincker, G. Liti, J. Schacherer, Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).
4. F. C. Odds, A. J. P. Brown, N. A. R. Gow, *Candida albicans* genome sequence: A platform for genomics in the absence of genetics. *Genome Biol.* **5**, 230 (2004).
5. C. Sharma, N. Kumar, R. Pandey, J. F. Meis, A. Chowdhary, Whole genome sequencing of emerging multidrug resistant *Candida auris* isolates in India demonstrates low genetic variation. *New Microbes New Infect.* **13**, 77–82 (2016).
6. L.-J. Ma, H. C. van der Does, K. A. Borkovich, J. J. Coleman, M.-J. Daboussi, A. Di Pietro, M. Dufresne, M. Freitag, M. Grabherr, B. Henrissat, P. M. Houterman, S. Kang, W.-B. Shim, C. Woloshuk, X. Xie, J.-R. Xu, J. Antoniw, S. E. Baker, B. H. Bluhm, A. Breakpear, D. W. Brown, R. A. E. Butchko, S. Chapman, R. Coulson, P. M. Coutinho, E. G. J. Danchin, A. Diener, L. R. Gale, D. M. Gardiner, S. Goff, K. E. Hammond-Kosack, K. Hilburn, A. Hua-Van, W. Jonkers, K. Kazan, C. D. Kodira, M. Koehrsen, L. Kumar, Y.-H. Lee, L. Li, J. M. Manners, D. Miranda-Saavedra, M. Mukherjee, G. Park, J. Park, S.-Y. Park, R. H. Proctor, A. Regev, M. C. Ruiz-Roldan, D. Sain, S. Sakhthikumar, S. Sykes, D. C. Schwartz, B. G. Turgeon, I. Wapinski, O. Yoder, S. Young, Q. Zeng, S. Zhou, J. Galagan, C. A. Cuomo, H. C. Kistler, M. Rep, Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373 (2010).
7. 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).
8. V. Wood, R. Gwilliam, M.-A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Felwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerrutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, P. Nurse, The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).
9. K. D. Hyde, S. S. N. Maharachchikumbura, S. Hongsanan, M. C. Samarakoon, R. Lücking, D. Pem, D. Harishchandra, R. Jeewon, R.-L. Zhao, J.-C. Xu, J.-K. Liu, A. M. Al-Sadi, A. H. Bahkali, A. M. Elgorban, The ranking of fungi: A tribute to David L. Hawksworth on his 70th birthday. *Fungal Divers.* **84**, 1–23 (2017).
10. M. C. Samarakoon, K. D. Hyde, S. Hongsanan, E. H. C. McKenzie, H. A. Ariyawansa, I. Promputtha, X.-Y. Zeng, Q. Tian, J.-K. J. Liu, Divergence time calibrations for ancient lineages of Ascomycota classification based on a modern review of estimations. *Fungal Divers.* **96**, 285–346 (2019).
11. L. Tedersoo, S. Sánchez-Ramírez, U. Kõljalg, M. Bahram, M. Döring, D. Schigel, T. May, M. Ryberg, K. Abarenkov, High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Divers.* **90**, 135–159 (2018).
12. C. L. Schoch, G.-H. Sung, F. López-Giráldez, J. P. Townsend, J. Miodlikowska, V. Hofstetter, B. Robbertse, P. B. Matheny, F. Kauff, Z. Wang, C. Gueidan, R. M. Andrie, K. Trippel, L. M. Ciuffetti, A. Wynns, E. Fraker, B. P. Hodkinson, G. Bonito, J. Z. Groenewald, M. Arzanlou, G. S. de Hoog, P. W. Crous, D. Hewitt, D. H. Pfister, K. Peterson, M. Gryzenhout, M. J. Wingfield, A. Aptroot, S.-O. Suh, M. Blackwell, D. M. Hillis, G. W. Griffith, L. A. Castlebury, A. Y. Rossman, H. T. Lumbsch, R. Lücking, B. Bédel, A. Rauhut, P. Diederich, D. Ertz, D. M. Geiser, K. Hosaka, P. Inderbitzin, J. Köhlmeier, B. Volkmann-Köhlmeier, L. Mostert, K. O'Donnell, H. Sipman, J. D. Rogers, R. A. Shoemaker, J. Sugiyama, R. C. Summerbell, W. Untereiner, P. R. Johnston, S. Stenroos, A. Zuccaro, P. S. Dyer, P. D. Crittenden, M. S. Cole, K. Hansen, J. M. Trappe, R. Yahr, F. Lutzoni, J. W. Spatafora, The Ascomycota tree of life: A phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Syst. Biol.* **58**, 224–239 (2009).
13. C. Beimforde, K. Feldberg, S. Nylander, J. Rikkinen, H. Tuovila, H. Dörfelt, M. Gube, D. J. Jackson, J. Reitner, L. J. Seyfullah, A. R. Schmidt, Estimating the Phanerozoic history of the Ascomycota lineages: Combining fossil and molecular data. *Mol. Phylogenet. Evol.* **78**, 386–398 (2014).

14. M. Prieto, M. Wedin, Dating the diversification of the major lineages of Ascomycota (Fungi). *PLOS ONE* **8**, e65576 (2013).
15. T. Y. James, F. Kauff, C. L. Schoch, P. B. Matheny, V. Hofstetter, C. J. Cox, G. Celio, C. Gueidan, E. Fraker, J. Miadlikowska, H. T. Lumbsch, A. Rauhut, V. Reeb, A. E. Arnold, A. Amtoft, J. E. Stajich, K. Hosaka, G.-H. Sung, D. Johnson, B. O'Rourke, M. Crockett, M. Binder, J. M. Curtis, J. C. Slot, Z. Wang, A. W. Wilson, A. Schüßler, J. E. Longcore, K. O'Donnell, S. Mozley-Standridge, D. Porter, P. M. Letcher, M. J. Powell, J. W. Taylor, M. M. White, G. W. Griffith, D. R. Davies, R. A. Humber, J. B. Morton, J. Sugiyama, A. Y. Rossman, J. D. Rogers, D. H. Pfister, D. Hewitt, K. Hansen, S. Hambleton, R. A. Shoemaker, J. Kohnmeyer, B. Volkman-Kohnmeyer, R. A. Spotts, M. Serdani, P. W. Crous, K. W. Hughes, K. Matsuura, E. Langer, G. Langer, W. A. Untereiner, R. Lücking, B. Büdel, D. M. Geiser, A. Aptroot, P. Diederich, I. Schmitt, M. Schultz, R. Yahr, D. S. Hibbett, F. Lutzoni, D. J. McLaughlin, J. W. Spatafora, R. Vilgalys, Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818–822 (2006).
16. J. W. Spatafora, G.-H. Sung, D. Johnson, C. Hesse, B. O'Rourke, M. Serdani, R. Spotts, F. Lutzoni, V. Hofstetter, J. Miadlikowska, V. Reeb, C. Gueidan, E. Fraker, T. Lumbsch, R. Lücking, I. Schmitt, K. Hosaka, A. Aptroot, C. Roux, A. N. Miller, D. M. Geiser, J. Hafellner, G. Hestmark, A. E. Arnold, B. Büdel, A. Rauhut, D. Hewitt, W. A. Untereiner, M. S. Cole, C. Scheidegger, M. Schultz, H. Sipman, C. L. Schoch, A five-gene phylogeny of Pezizomycotina. *Mycologia* **98**, 1018–1028 (2006).
17. C. G. P. McCarthy, D. A. Fitzpatrick, Multiple approaches to phylogenomic reconstruction of the fungal kingdom. *Adv. Genet.* **100**, 211–266 (2017).
18. L. G. Nagy, G. Szöllösi, Fungal phylogeny in the age of genomics: Insights into phylogenetic inference from genome-scale datasets. *Adv. Genet.* **100**, 49–72 (2017).
19. 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. Dougllass, 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. Otilar, 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. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9882–9887 (2016).
20. A. Rajeh, J. Lv, Z. Lin, Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in Ascomycota. *BMC Genomics* **19**, 282 (2018).
21. S. A. Higgins, C. W. Schadt, P. B. Matheny, F. E. Löffler, Phylogenomics reveal the dynamic evolution of fungal nitric oxide reductases and their relationship to secondary metabolism. *Genome Biol. Evol.* **10**, 2474–2489 (2018).
22. X. Zhou, X.-X. Shen, C. T. Hittinger, A. Rokas, Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. *Mol. Biol. Evol.* **35**, 486–503 (2018).
23. J. W. Taylor, M. L. Berbee, Dating divergences in the Fungal Tree of Life: Review and new analyses. *Mycologia* **98**, 838–849 (2006).
24. A. C. B. Padovan, G. F. O. Sanson, A. Brunstein, M. R. S. Briones, Fungi evolution revisited: Application of the penalized likelihood method to a Bayesian fungal phylogeny provides a new perspective on phylogenetic relationships and divergence dates of Ascomycota groups. *J. Mol. Evol.* **60**, 726–735 (2005).
25. C. Gueidan, C. Ruibal, G. S. de Hoog, H. Schneider, Rock-inhabiting fungi originated during periods of dry climate in the late Devonian and middle Triassic. *Fungal Biol.* **115**, 987–996 (2011).
26. J. E. Stajich, Fungal genomes and insights into the evolution of the kingdom. *Microbiol. Spectr.* **5**, 619–633 (2017).
27. X.-X. Shen, X. Zhou, J. Kominik, C. P. Kurtzman, C. T. Hittinger, A. Rokas, Reconstructing the backbone of the *Saccharomycotina* yeast phylogeny using genome-scale data. *G3 (Bethesda)* **6**, 3927–3939 (2016).
28. I. Kjørboelling, T. Vesth, J. C. Frisvad, J. L. Nybo, S. Theobald, S. Kildgaard, T. I. Petersen, A. Kuo, A. Sato, E. K. Lyhne, M. E. Kogle, A. Wiebenga, R. S. Kun, R. J. M. Lubbers, M. R. Mäkelä, K. Barry, M. Chovatia, A. Clum, C. Daum, S. Haridas, G. He, K. LaButti, A. Lipzen, S. Mondo, J. Pangilinan, R. Riley, A. Salamov, B. A. Simmons, J. K. Magnuson, B. Henrissat, U. H. Mortensen, T. O. Larsen, R. P. de Vries, I. V. Grigoriev, M. Machida, S. E. Baker, M. R. Andersen, A comparative genomic study of 23 *Aspergillus* species from section *Flavi*. *Nat. Commun.* **11**, 1106 (2020).
29. S. Haridas, R. Albert, M. Binder, J. Bloem, K. LaButti, A. Salamov, B. Andreopoulos, S. E. Baker, K. Barry, G. Bills, B. H. Bluhm, C. Cannon, R. Castanera, D. E. Culley, C. Daum, D. Ezra, J. B. González, B. Henrissat, A. Kuo, C. Liang, A. Lipzen, F. Lutzoni, J. Magnuson, S. J. Mondo, M. Nolan, R. A. Ohm, J. Pangilinan, H.-J. Park, L. Ramírez, M. Alfaro, H. Sun, A. Tritt, Y. Yoshinaga, L.-H. Zwiars, B. G. Turgeon, S. B. Goodwin, J. W. Spatafora, P. W. Crous, I. V. Grigoriev, 101 *Dothideomycetes* genomes: A test case for predicting lifestyles and emergence of pathogens. *Stud. Mycol.* **96**, 141–153 (2020).
30. M. Marcet-Houben, T. Gabaldón, Acquisition of prokaryotic genes by fungal genomes. *Trends Genet.* **26**, 5–8 (2010).
31. 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).
32. M. A. Naranjo-Ortiz, T. Gabaldón, Fungal evolution: Diversity, taxonomy and phylogeny of the Fungi. *Biol. Rev. Camb. Philos. Soc.* **94**, 2101–2137 (2019).
33. B. Robbertse, J. B. Reeves, C. L. Schoch, J. W. Spatafora, A phylogenomic analysis of the Ascomycota. *Fungal Genet. Biol.* **43**, 715–725 (2006).
34. J. Choi, S.-H. Kim, A genome Tree of Life for the Fungi kingdom. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9391–9396 (2017).
35. X.-X. Shen, C. T. Hittinger, A. Rokas, Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 0126 (2017).
36. X.-X. Shen, D. Liang, M.-Y. Chen, R.-L. Mao, D. B. Wake, P. Zhang, Enlarged multilocus data set provides surprisingly younger time of origin for the Plethodontidae, the largest family of salamanders. *Syst. Biol.* **65**, 66–81 (2016).
37. J. E. Galagan, M. R. Henn, L.-J. Ma, C. A. Cuomo, B. Birren, Genomics of the fungal kingdom: Insights into eukaryotic biology. *Genome Res.* **15**, 1620–1631 (2005).
38. J. J. Welch, O. R. P. Bininda-Emonds, L. Bromham, Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol. Biol.* **8**, 53 (2008).
39. R. Salari, R. Salari, Investigation of the best *Saccharomyces cerevisiae* growth condition. *Electron Physician* **9**, 3592–3597 (2017).
40. A. P. J. Trinci, A kinetic study of the growth of *Aspergillus nidulans* and other fungi. *J. Gen. Microbiol.* **57**, 11–24 (1969).
41. O. J. Gillie, Growth of *Neurospora crassa* in unstirred liquid cultures. *J. Gen. Microbiol.* **51**, 179–184 (1968).
42. J. L. Steenwyk, D. A. Opulente, J. Kominik, X.-X. Shen, X. Zhou, A. L. Labella, N. P. Bradley, B. F. Eichman, N. Čadež, D. Libkind, J. DeVirgilio, 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**, e3000255 (2019).
43. L. Bromham, The genome as a life-history character: Why rate of molecular evolution varies between mammal species. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2503–2513 (2011).
44. Y. O. Zhu, M. L. Siegal, D. W. Hall, D. A. Petrov, Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2310–E2318 (2014).
45. P. Bolívar, L. Guéguen, L. Duret, H. Ellegren, C. F. Mugal, GC-biased gene conversion conceals the prediction of the nearly neutral theory in avian genomes. *Genome Biol.* **20**, 5 (2019).
46. L. Tedersoo, M. Bahram, R. Puusepp, R. H. Nilsson, T. Y. James, Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome* **5**, 42 (2017).
47. N. H. Nguyen, Z. Song, S. T. Bates, S. Branco, L. Tedersoo, J. Menke, J. S. Schilling, P. G. Kennedy, FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
48. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** Suppl 2, ii215–ii225 (2003).
49. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
50. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
51. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
52. M. L. Z. Mendoza, Z. Xiong, M. Escalera-Zamudio, A. K. Runge, J. Thézé, D. Streicker, H. K. Frank, E. Loza-Rubio, S. Liu, O. A. Ryder, J. A. S. Castruita, A. Katzurakis, G. Pacheco, B. Taboada, U. Löber, O. G. Pybus, Y. Li, E. Rojas-Anaya, K. Bohmann, A. C. Baez, C. F. Arias, S. Liu, A. D. Greenwood, M. F. Bertelsen, N. E. White, M. Bunce, G. Zhang, T. Sicheritz-Pontén, M. P. T. Gilbert, Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat. Ecol. Evol.* **2**, 659–668 (2018).
53. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
54. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
55. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
56. T. N. Taylor, H. Hass, H. Kerp, M. Krings, R. T. Hanlin, Perithecial ascomycetes from the 400 million year old Rhynie chert: An example of ancestral polymorphism. *Mycologia* **97**, 269–285 (2005).
57. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
58. 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).
59. T. M. Lowe, P. P. Chan, tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57 (2016).

60. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
61. J. Felsenstein, Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
62. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
63. J. L. Steenwyk, A. Rokas, Treehouse: A user-friendly application to obtain subtrees from large phylogenies. *BMC Res. Notes* **12**, 541 (2019).

**Acknowledgments:** We thank members of the Rokas and Hittinger laboratories, especially members of the Y1000+ Project for constructive feedback. **Funding:** This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University. X.-X.S. was supported by a grant from the National Natural Science Foundation of China (No. 32071665) and the Fundamental Research Funds for the Central Universities (no. 2020QNA6019). X.Z. was supported by the National Key Project for Basic Research of China (973 Program, no. 2015CB150600) and the open fund from Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, College of Crop Science, Fujian Agriculture and Forestry University (GBMUC-2018-005). M.G. was supported by the Royal Netherlands Academy of Arts and Sciences. C.T.H. was supported by the National Science Foundation (DEB-1442148), the USDA National Institute of Food and Agriculture (Hatch Project No. 1020204), in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science No. DE-SC0018409), the Pew Charitable Trusts (Pew Scholar in the Biomedical Sciences), and the Office of the Vice Chancellor for Research and Graduate

Education with funding from the Wisconsin Alumni Research Foundation (H. I. Romnes Faculty Fellow). J.L.S. and A.R. were supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. A.R. was supported by the National Science Foundation (DEB-1442113), the Guggenheim Foundation, and the Burroughs Wellcome Fund. **Author contributions:** Study conception and design: X.-X.S., C.T.H., and A.R. Acquisition of data: X.-X.S. Analysis and interpretation of data: X.-X.S., J.L.S., A.L.L., D.A.O., X.Z., J.K., Y.L., M.G., C.T.H., and A.R. Drafting of manuscript: X.-X.S. and A.R. Critical revision: all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All genome assemblies and proteomes are publicly available in the Zenodo repository: <https://doi.org/10.5281/zenodo.3783970>. Multiple sequence alignments, phylogenetic trees, trait ancestral character state reconstructions, log files, R codes, and custom Perl scripts are available on the Figshare repository: <https://doi.org/10.6084/m9.figshare.12196149>.

Submitted 26 May 2020

Accepted 21 September 2020

Published 4 November 2020

10.1126/sciadv.abd0079

**Citation:** 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).

## Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota

Xing-Xing Shen, Jacob L. Steenwyk, Abigail L. LaBella, Dana A. Opulente, Xiaofan Zhou, Jacek Kominek, Yuanning Li, Marizeth Groenewald, Chris T. Hittinger and Antonis Rokas

*Sci Adv* 6 (45), eabd0079.  
DOI: 10.1126/sciadv.abd0079

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/6/45/eabd0079>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2020/11/02/6.45.eabd0079.DC1>

### REFERENCES

This article cites 63 articles, 4 of which you can access for free  
<http://advances.sciencemag.org/content/6/45/eabd0079#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).