Title

Toward a Fully Resolved Fungal Tree of Life

Journal

Annual Review of Microbiology, 74(1)

ISSN

0066-4227

Authors

James, Timothy Y
Stajich, Jason E
Hittinger, Chris Todd
et al.

Publication Date

2020-09-08

DOI

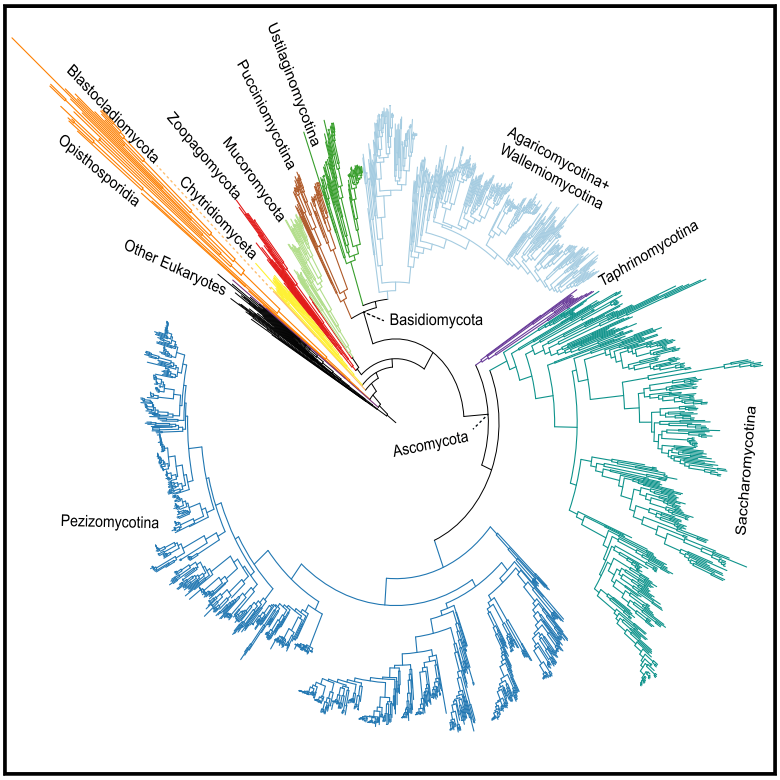
10.1146/annurev-micro-022020-051835

Article

Current Biology

A genome-scale phylogeny of the kingdom Fungi

Graphical Abstract



Highlights

- Genome-scale phylogeny of the fungal kingdom based on 290 genes and 1,644 species
- 85% of inferred phylogenetic relationships among fungi are robustly supported
- Certain unresolved relationships may be due to ancient diversification events
- Fungal higher rank taxonomy broadly reflects organisms' genome sequence divergence

Authors

Yuanning Li, Jacob L. Steenwyk, Ying Chang, ..., Chris Todd Hittinger, Xing-Xing Shen, Antonis Rokas

Correspondence

xingxingshen@zju.edu.cn (X.-X.S.), antonis.rokas@vanderbilt.edu (A.R.)

In Brief

Li et al. analyze 290 genes from 1,644 species to infer a genome-scale phylogeny of the fungal kingdom. Analyses using different approaches and data matrices show that 85% of inferred relationships among fungi are robustly supported. The results provide a robust phylogenomic framework to explore the tempo and mode of fungal evolution.

Li et al., 2021, Current Biology *31*, 1653–1665
April 26, 2021
https://doi.org/10.1016/j.cub.2021.01.074

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

A genome-scale phylogeny of the kingdom Fungi

Yuanning Li,[1] Jacob L. Steenwyk,[1] Ying Chang,[2] Yan Wang,[3,4] Timothy Y. James,[5] Jason E. Stajich,[3] Joseph W. Spatafora,[2] Marizeth Groenewald,[6] Casey W. Dunn,[7] Chris Todd Hittinger,[8] Xing-Xing Shen,[9,*] and Antonis Rokas[1,10,*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA
[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA
[3]Department of Microbiology and Plant Pathology, Institute for Integrative Genome Biology, University of California, Riverside, CA 92521, USA
[4]Department of Biological Sciences, University of Toronto Scarborough and Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada
[5]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA
[6]Westerdijk Fungal Biodiversity Institute, 3584 CT, Utrecht 85167, the Netherlands
[7]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA
[8]Laboratory of Genetics, Center for Genomic Science Innovation, J.F. Crow Institute for the Study of Evolution, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI 53706, USA
[9]State Key Laboratory of Rice Biology and Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, Hangzhou 310058, China
[10]Lead contact
*Correspondence: xingxingshen@zju.edu.cn (X.-X.S.), antonis.rokas@vanderbilt.edu (A.R.)
https://doi.org/10.1016/j.cub.2021.01.074

SUMMARY

Phylogenomic studies using genome-scale amounts of data have greatly improved understanding of the tree of life. Despite the diversity, ecological significance, and biomedical and industrial importance of fungi, evolutionary relationships among several major lineages remain poorly resolved, especially those near the base of the fungal phylogeny. To examine poorly resolved relationships and assess progress toward a genome-scale phylogeny of the fungal kingdom, we compiled a phylogenomic data matrix of 290 genes from the genomes of 1,644 species that includes representatives from most major fungal lineages. We also compiled 11 data matrices by subsampling genes or taxa from the full data matrix based on filtering criteria previously shown to improve phylogenomic inference. Analyses of these 12 data matrices using concatenation- and coalescent-based approaches yielded a robust phylogeny of the fungal kingdom, in which ~85% of internal branches were congruent across data matrices and approaches used. We found support for several historically poorly resolved relationships as well as evidence for polytomies likely stemming from episodes of ancient diversification. By examining the relative evolutionary divergence of taxonomic groups of equivalent rank, we found that fungal taxonomy is broadly aligned with both genome sequence divergence and divergence time but also identified lineages where current taxonomic circumscription does not reflect their levels of evolutionary divergence. Our results provide a robust phylogenomic framework to explore the tempo and mode of fungal evolution and offer directions for future fungal phylogenetic and taxonomic studies.

INTRODUCTION

Kingdom Fungi, one of the most diverse and ancient branches of the tree of life, includes an estimated 2–5 million species that play vital roles in terrestrial and aquatic ecosystems (Figure 1).[1–3] Fungi exhibit a wide variety of feeding lifestyles, morphologies, developmental patterns, and ecologies and are thought to have coevolved with plants.[1,4] A robustly resolved phylogeny of fungi is necessary for understanding how their genes, pathways, traits, and their biology in general evolved. However, the early history of diversification of major fungal lineages remains poorly resolved.[5]

There are more than 200 orders of fungi classified into 12 phyla (see an alternative scheme of classification[6]).[5] These 12 phyla are placed into six major groups: the subkingdoms Dikarya (which includes the phyla Ascomycota, Basidiomycota, and Entorrhizomycota) and Chytridiomyceta (which includes the phyla Chytridiomycota, Monoblepharidomycota, and Neocallimastigomycota); the phyla Mucoromycota, Zoopagomycota, and Blastocladiomycota; and the major group Opisthosporidia (which includes the phyla Aphelidiomycota, Cryptomycota/Rozellomycota, and Microsporidia and is possibly paraphyletic).[5]

Evolutionary relationships among several fungal higher taxonomic ranks are poorly resolved, with molecular studies providing support for conflicting hypotheses or being equivocal in their support (Figure S1).[5,7] For example, relationships among the three phyla within Opisthosporidia are ambiguous, especially the placement of Aphelidiomycota (Figure S1). This is likely due to the parasitic lifestyles, highly reduced morphologies, and very rapidly evolving genomes of many of the organisms involved (e.g., Microsporidia), which render their evolutionary placement challenging.[8,9] Ambiguity also exists with respect to

Figure 1. Diversity of major fungal lineages

Representative species for major fungal lineages.
(A) Crown coral *Artomyces pyxidata* (Agaricomycotina and Basidiomycota).
(B) Witch's butter *Tremella mesenterica* (Pucciniomycotina and Basidiomycota).
(C) Flowerpot parasol, *Leucocoprinus birnbaumii* (Agaricomycotina and Basidiomycota).
(D) Pearl oyster mushroom, *Pleurotus ostreatus* (Agaricomycotina and Basidiomycota).
(E) Snow fungus, *Tremella fuciformis* (Agaricomycotina and Basidiomycota).
(F) Turkey tail, *Trametes versicolor* (Agaricomycotina and Basidiomycota).
(G) Baker's yeast *Saccharomyces cerevisiae* (Saccharomycotina and Ascomycota).
(H) Fission yeast *Schizosaccharomyces pombe* (Taphrinomycotina and Ascomycota).
(I) *Mucor mucedo* (Mucoromycotina and Mucoromycota).
(J) Corn smut *Ustilago maydis* (Ustilaginomycotina and Basidiomycota).
(K) *Aspergillus oerlinghausenensis* (Pezizomycotina and Ascomycota).
(L) Fly agaric *Amanita muscaria* (Agaricomycotina and Basidiomycota).
(M) *Entomophthora muscae* (Entomophthoromycotina and Zoopagomycota).
(N) *Rozella allomycis* parasitizing the chytrid *Allomyces*.
(O) *Monoblepharis macrandra* (Monoblepharidomycetes and Chytridiomycota).
(P) *Coemansia braziliensis* (Kickxellomycotina and Zoopagomycota).
(Q) *Piptocephalis repens* (Zoopagomycotina and Zoopagomycota).
(R) *Mortierella elongata* (Mortierellomycotina and Mucoromycota).
(S) *Rhizopus* spp. (Mucoromycotina and Mucoromycota).
(T) *Penicillium digitatum* (Pezizomycotina and Ascomycota).
(A–C, E, and F) Photograph courtesy of Jacob L. Steenwyk. (D, G, J, L, M, S, and T) Images are available to the public domain through https://commons.wikimedia.org/wiki/Main_Page. (H) Photograph reproduced with permission of David O. Morgan. (K) Photograph courtesy of Jos Houbraken. (I) Photograph courtesy of Kerry O'Donnell. (N and O) Photographs courtesy of Kensuke Seto and Timothy James. (P–R) Photographs courtesy of Jason Stajich. See also Figure S1 and Data S1. Permission has been obtained to use all images.

the placement of Blastocladiomycota, a group of flagellated zoospore-producing fungi whose characteristics are similar to those of terrestrial fungi.[10] Previous analyses place Blastocladiomycota as diverging either before or after Chytridiomyceta (Figure S1), making their placement on the fungal phylogeny key for understanding the evolution of diverse fungal traits.[5,11] Mucoromycota and Zoopagomycota were previously classified as zygomycetes,[12] a now defunct taxonomic group, based on the production of coenocytic hyphae and sexual reproduction by zygospores. After arbuscular mycorrhizal fungi were segregated from zygomycetes into the new phylum Glomeromycota,[13] zygomycetes became paraphyletic[7,11] and the group was abandoned in favor of a classification of zygomycete taxa into two major lineages, Mucoromycota and Zoopagomycota.[7] The placement of Entorrhizomycota, a group of gall-forming root parasites of Poales flowering plants, with respect to Basidiomycota is also not clear.[5] Finally, evolutionary relationships among phyla within the chytrid clade Chytridiomyceta, among subphyla within Basidiomycota, and within phylum Ascomycota (e.g., between classes in Taphrinomycotina) are also elusive (Figure S1).[5,7]

In retrospect, previous molecular phylogenetic analyses have relied primarily on a few loci from many taxa that often provided little resolution of the deep internal branches (e.g., 6 genes/199 taxa)[10] or genomic data with scarce taxon sampling (e.g., 53

genes/121 taxa;[14] 192 genes/46 taxa;[15] 650 genes/104 taxa;[16] and 455 genes/72 taxa[17]). However, phylogenomic studies of specific fungal lineages that are well sampled, such as Saccharomycotina (e.g., 2,408 genes/332 taxa)[18] and Ascomycota (e.g., 815 genes/1,107 taxa),[19] suggest that denser gene and taxon sampling holds great potential for resolving relationships that previously seemed intractable.

A robust phylogenetic framework for Fungi based on a broad sampling of genes and taxa is key for understanding the evolution of the kingdom and would greatly facilitate larger scale studies in fungal comparative biology, ecology, and genomics. In recent years, the 1000 Fungal Genomes Project (https://mycocosm.jgi.doe.gov/mycocosm/home/1000-fungal-genomes) has greatly expanded the availability of genomes from diverse understudied taxa.[20] Additionally, efforts focused on specific ecological or evolutionary groups, such as the Y1000+ Project (http://y1000plus.wei.wisc.edu/) that aims to sequence all known species of the subphylum Saccharomycotina,[21] the Dothideomycetes project that aims to study plant pathogenic fungi,[22] and the *Aspergillus* genome project that aims to examine the metabolic dexterity of this diverse genus of fungi,[23] have greatly increased the availability of genomes from specific lineages.

The availability of genomic data from a substantially expanded and more representative set of fungal species offers an opportunity to reconstruct a genome-scale fungal tree of life and examine its support for relationships that have heretofore remained poorly resolved (Figure S1). To this end, we analyzed data from 1,644 available fungal genomes that include representatives from most major lineages and provided a robust phylogenomic framework to explore the evolution of the fungal kingdom.

RESULTS

A pan-fungal phylogenomic matrix with high taxon sampling and occupancy

To assemble a phylogenomic data matrix, we sampled 1,707 publicly available genomes from NCBI (one representative genome per species; retrieved on January 30, 2020), representing every major lineage across fungi (1,679 taxa) and selected outgroups (28 taxa) based on the current understanding of the Opisthokonta phylogeny;[24,25] the sole exceptions were the Aphelidiomycota and Entorrhizomycota phyla, for which no genomes were available as of January 30, 2020 (Data S1).

To filter out low-quality genomes, we analyzed each genome using BUSCO[26] with the Fungi OrthoDB v9 database,[27] which contains 290 genes. To minimize missing data and remove potential low-quality genomes, we retained only those genomes that contained $\geq$100 single-copy BUSCO genes (Data S1). This analysis resulted in the removal of the genomes of 35 fungal species. The average genome assembly completeness for the remaining 1,672 taxa was 92.32% (average of 267.74/290 BUSCO genes). The full data matrix contains 124,700 amino acid sites from 290 BUSCO genes (90.6% taxon occupancy per BUSCO gene, an average length of 430 residues per gene after trimming, and 84.36% site occupancy) across 1,672 taxa (1,644 fungal taxa and 28 outgroups; Data S2). To conduct sensitivity analyses for potential systematic errors or biases that may influence the accuracy of phylogenetic inference, we generated 11 data matrices by subsampling genes (8 data matrices) or taxa (3 data matrices) from the full data matrix. The examined biases include the removal of genes (e.g., based on shorter alignment length and higher evolutionary rate) or taxa (e.g., by removing rogue taxa) according to filtering criteria previously shown to improve phylogenomic inference (Figure S2).[28,29]

A robust phylogenetic framework to explore fungal evolution

To infer the fungal phylogeny, we used concatenation-based single model (unpartitioned), concatenation-based data partitioning (one partition per gene), and coalescent-based approaches on the 12 data matrices (Figure S2). The gene occupancy for every taxon in each data matrix is shown in Data S2. These analyses produced 33 phylogenetic trees: 12 from concatenation-based single model analyses; nine from concatenation-based data-partitioning analyses (phylogenies were not inferred from three matrices for reasons of computational efficiency); and 12 from coalescent-based analyses; see STAR methods for more details. We found that ~85% (1,414/1,669 of bipartitions (or internodes/internal branches) were recovered consistently across these 33 phylogenies, suggesting that a large fraction of bipartitions in the fungal phylogeny were robustly supported (Figures S3 and S4).

Notable examples of relationships recovered in all 33 phylogenies included the placements of the cellular slime mold *Fonticula* as sister to fungi and of Opisthosporidia as sister to the rest of fungi (Figures 2, 3, and S3).[25,30] Our analyses also robustly placed Wallemiomycotina (previously placed sister to[31,32] or outside of, albeit with low support,[33] Agaricomycotina) as sister to Agaricomycotina with strong support (bootstrap [BS] = 100%; local posterior probability [LPP] = 100; Figures 2 and 3).

In general, robustly supported relationships were more commonly found in parts of the tree with higher taxon sampling. For Ascomycota, the phylum with the highest sampling of taxa in our data matrix, ~94% of bipartitions (1,036/1,101) were consistently recovered across the 33 phylogenies. For example, we found that all 33 phylogenies strongly supported Taphrinomycotina as the sister lineage to a clade of Saccharomycotina and Pezizomycotina (BS = 100%; LPP = 100; q1 = 0.62; Figures 3 and 4H). Similarly, all phylogenies strongly supported a clade consisting of Pezizomycetes and Orbiliomycetes as the sister group to the remaining Pezizomycotina (Figures 3 and S5). Both Saccharomycotina (332 taxa with representatives of all 12 major clades included) and Pezizomycotina (761 taxa with 9/17 known classes included) are the most well-sampled major lineages in our data matrix (Data S2), suggesting that genome sequencing of underrepresented taxa will improve the resolution of the fungal tree of life. Importantly, relationships among the 12 major clades of the subphylum Saccharomycotina and relationships among higher taxonomic ranks within Ascomycota recovered by our analyses are essentially the same as those of previous studies performed using different sets of genes and taxa.[18,19]

Finally, we note that a recent study used the alignment-free feature frequency profile (FFP) method to reconstruct a broad sketch of the fungal tree of life based on proteome data from over 400 fungal genomes.[34] However, it was recently shown that the performance of the FFP method is much worse than

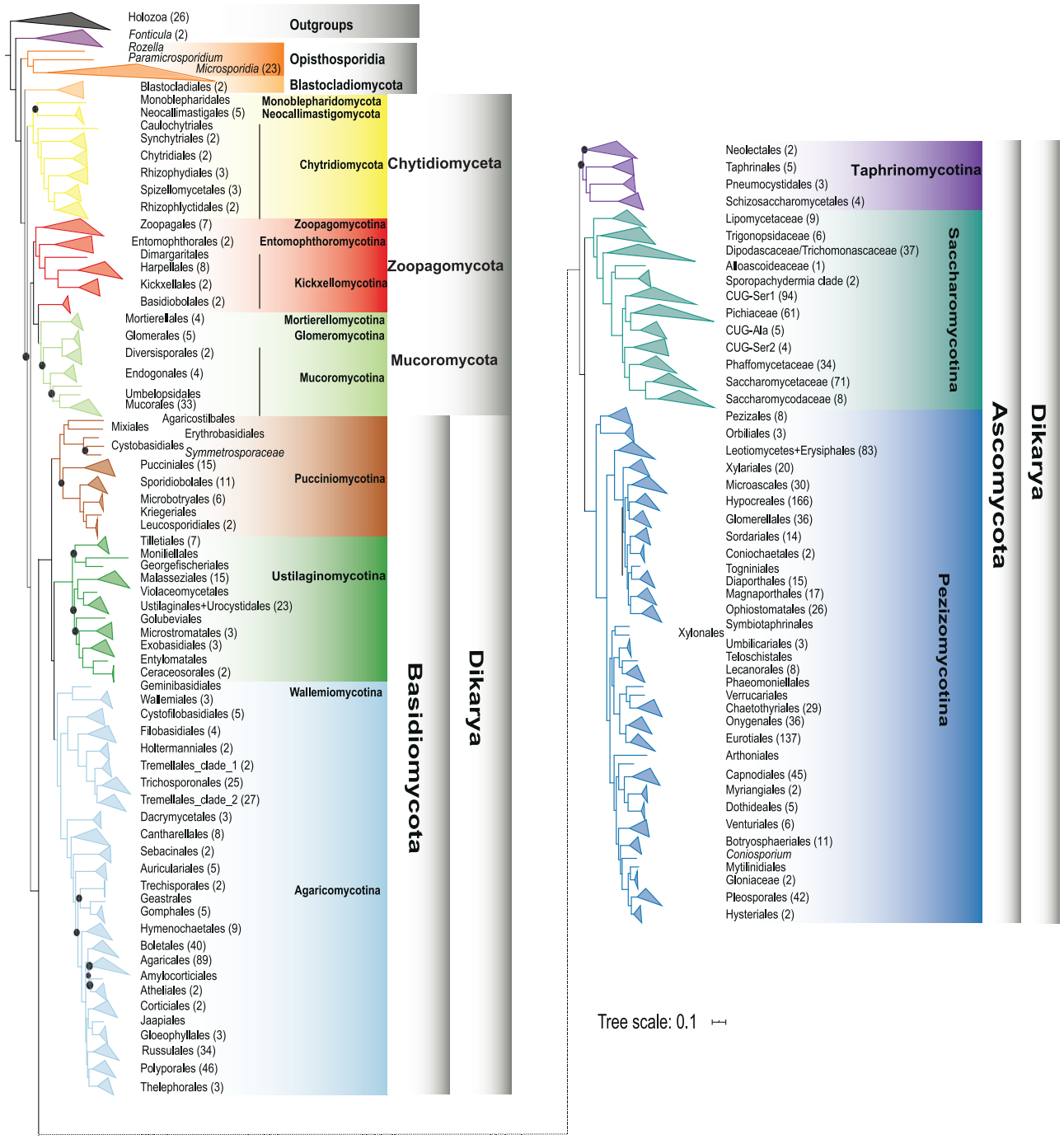


Figure 2. Genome-scale phylogeny of 1,644 species spanning the diversity of fungi

The topology shown is derived from maximum likelihood analysis using a concatenation single-model (LG+G4) approach on the full data matrix (1,672 taxa [1,644 fungi and 28 outgroups] and 290 genes; *lnL* = −78287339.984). Internal branches supported with 100% ultrafast bootstrap values are not shown; those with values lower than 100% are denoted by purple dots. Termini are labeled using order-level taxonomic names from NCBI, except for in Saccharomycotina, where informal and family-level names reflecting the 12 major clades comprising this group are used.[18] See also Figures S3 and S6 and Data S2.

concatenation and coalescence for reconstructing the phylogeny of major and ancient lineages,[35] such as fungi. The poor performance of the FFP method explains why many relationships reported by Choi and Kim[34] strongly contradict the current consensus view of the fungal tree of life.[5,19]

Most instances of incongruence stem from differences between concatenation- and coalescent-based phylogenies

By examining the distribution of incongruence across the 33 phylogenies, we found that the 21 phylogenies obtained from

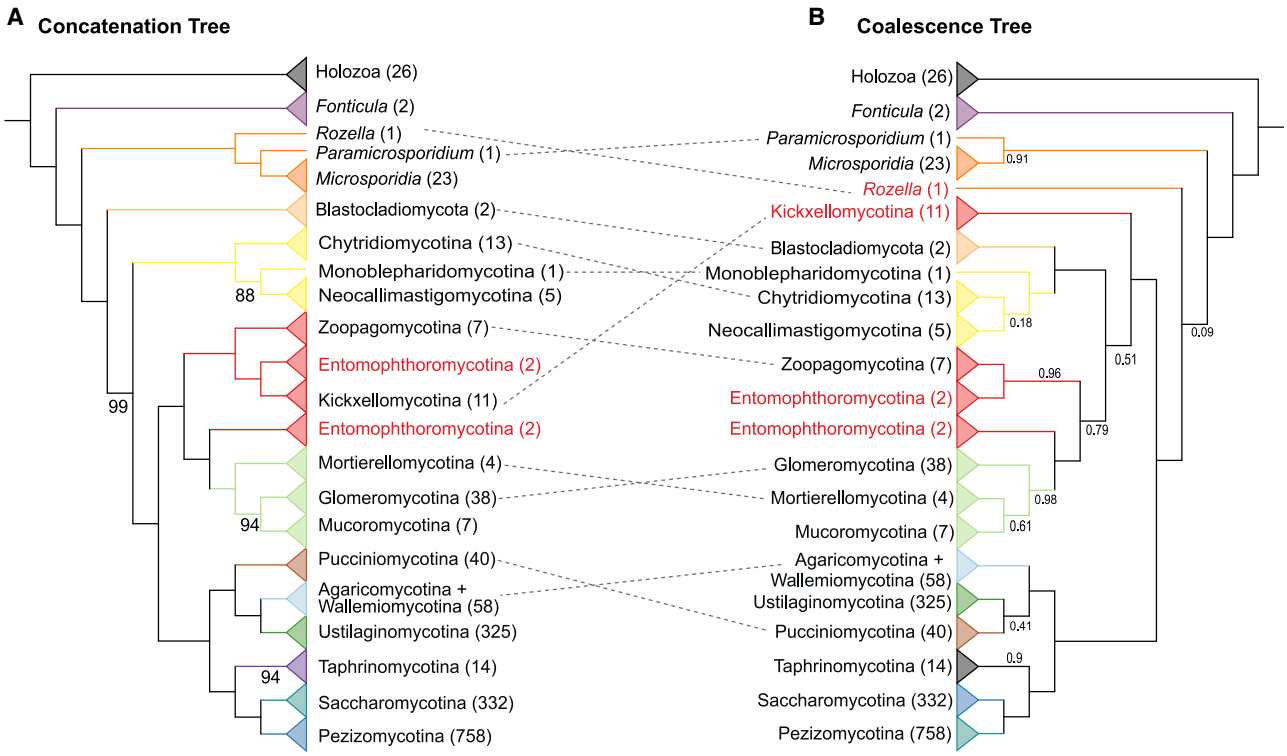


Figure 3. Incongruence between concatenation- and coalescent-based phylogenies of fungi

Topologies derived from maximum likelihood analysis using (A) a concatenation single-model (LG+G4) approach and (B) a coalescence approach. Numerical values below branches represent (A) ultrafast bootstrap (BS) values and (B) local posterior probabilities (LPP); unlabeled branches received 100% BS or 1.0 PP support. Termini are labeled using major lineages of fungi. Taxa in red correspond to groups inferred to be paraphyletic by the topology shown. The dashed line indicated the incongruent placements between topologies from concatenation and coalescence. See also Figure S5 and Data S3.

concatenation-based single model and data-partitioning analyses were largely congruent (Figure S4); an average of 98.6% (1,645/1,669) of bipartitions were recovered consistently.

In contrast, 145/255 (average = 58.9%) incongruent bipartitions found across the 33 phylogenies were mainly due to whether the data matrix was analyzed by concatenation or coalescence (Figure S4). Furthermore, these incongruent bipartitions were more concentrated in branches toward the base of the fungal phylogeny (Figures 3 and S5). By examining incongruence at the taxonomic levels of order, class, and phylum, we found four taxonomic groups that were recovered as non-monophyletic in concatenation-based analyses compared to six non-monophyletic groups in coalescent-based analyses (Figure S5; Data S3). Coalescent-based trees contradict well-established relationships supported by most previous phylogenetic studies, as well as by our concatenation-based analyses, such as the sister group relationship of Rozellomycota and Microsporidia[30,36] and the monophyly of Zoopagomycota (excluding *Basidiobolus*; Figures 3B and S5B).[15]

The observed differences between concatenation-based and coalescent-based analyses may stem from the fact that a substantial number of internodes in individual gene trees are not well supported. We found an average of 4.99%, 6.69%, 10.74%, and 19.18% of internodes in individual gene trees that received ultrafast bootstrap support values lower than 33%, 50%, 75%, and 95%, respectively. Given that values above 95% are considered as strong support,[37] these results suggest that nearly one in five internodes in individual gene trees lacks robust support. Because our coalescence-based analyses directly use these gene trees to infer the coalescent-based species trees, their accuracy may be disproportionally affected (compared to the concatenation-based species trees) by the poor resolution of individual gene trees.

Another possible explanation is that 290 genes are not sufficient to robustly resolve all internal branches of a tree with hundreds of taxa. The number of genes in a phylogenomic data matrix is known to impact the accuracy of both concatenation-based[38] and coalescent-based inference.[39] Moreover, the taxon occupancy values for non-Dikarya fungi (average of 207.02/290 BUSCO genes; 71.39%) are substantially lower than the ones of Dikarya (average of 279.59/290 BUSCO genes; 96.41%). Consequently, the placements of non-Dikarya taxa are based on many fewer genes and gene trees.[38]

Notwithstanding the debate on which of the two approaches is better or more appropriate for estimating species phylogenies,[40,41] these results suggest that concatenation-based phylogenies of this phylogenomic data matrix are likely more reliable than coalescent-based phylogenies due to the poor resolution of individual gene trees (see also Shen et al.[42]).

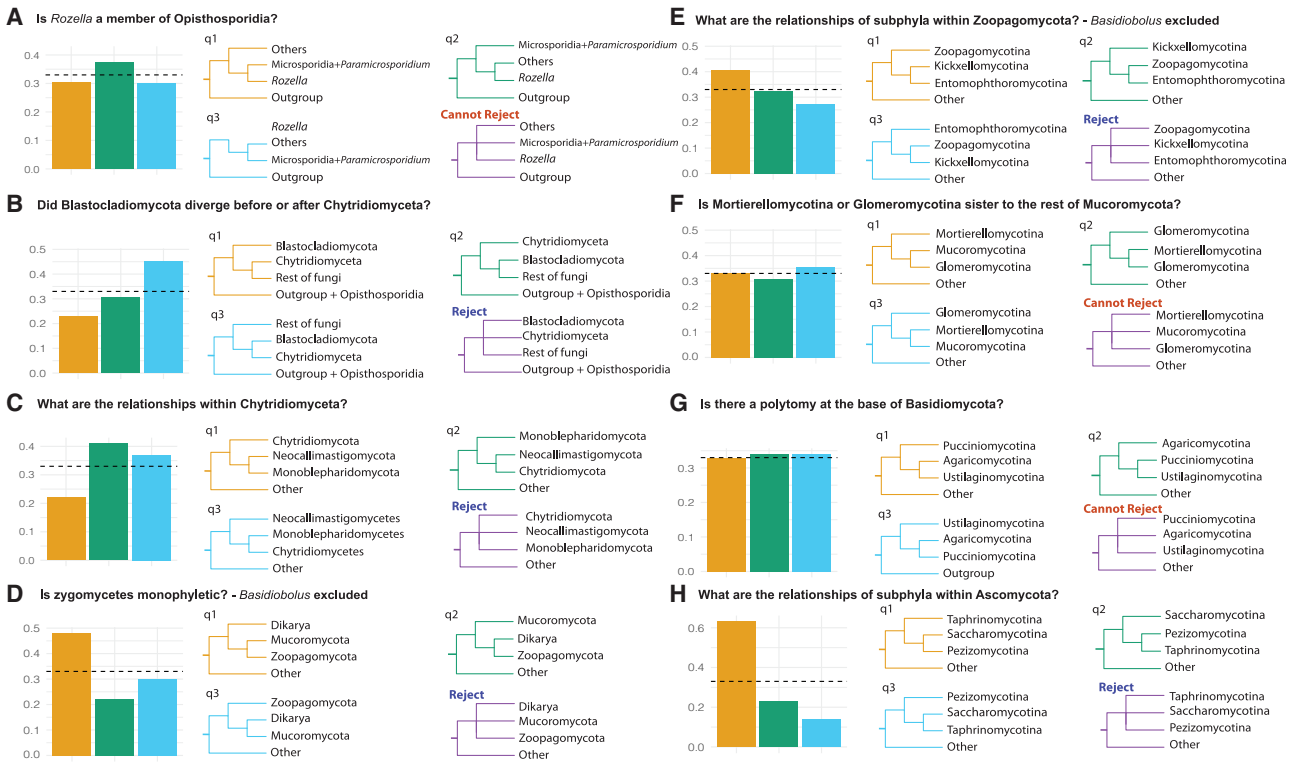


Figure 4. Examination of support among individual gene tees for alternative hypotheses for contentious relationships in the fungal phylogeny

The gene-tree quartet frequencies (bar graphs) for alternative branching orders for contentious relationships in the fungal phylogeny.

(A) Is *Rozella* a member of Opisthosporidia?

(B) Did Blastocladiomycota diverge before or after Chytridiomyceta?

(C) What are the relationships within Chytridiomyceta?

(D) Is zygomycetes monophyletic?

(E) What are the relationships of subphyla within Zoopagomycota?

(F) Is Mortierellomycotina or Glomeromycotina sister to the rest of Mucoromycota?

(G) Is there a polytomy at the base of Basidiomycota?

(H) What are the relationships of subphyla within Ascomycota?

Orange bars and topologies reflect the relationships inferred using a concatenation-based single-model approach on the full data matrix; blue and green bars and trees correspond to the two alternative hypotheses (supported by the two alternative resolutions of each quartet). The purple tree shows whether a polytomy scenario can be rejected by the quartet analysis or not. Dashed horizontal lines mark expectation for a hard polytomy. See also Figures S1 and S2.

Incongruence among major lineages and identification of ancient radiations

Although ~85% of internodes in our phylogeny of Fungi were robustly supported irrespective of approach and data matrix used, the remaining ~15% showed incongruence between analyses. Below, we discuss key incongruent relationships of interest. For each case, we present the results from our concatenation- and coalescent-based analyses and place our results in the context of the published literature. We also tested whether the data from the 290 gene trees rejected the hypothesis that the branch in question represents a polytomy (Figure 4). Briefly, the polytomy test evaluates whether the frequencies of quartet trees (obtained from all the gene trees) are significantly different for a branch of interest.[43] For every quartet tree, there are three possible topologies (i.e., three alternative hypotheses noted as q1, q2, and q3) of how the taxa are related. The test measures the frequencies of the quartet trees present in all gene trees; if there are no significant differences in their frequencies, then the hypothesis that the branch in question is a polytomy cannot be rejected. Given that the quartet frequencies are obtained from the individual gene trees, the analyses of Figure 4 generally reflect the results of the coalescent-based analyses.

Is Rozella a member of Opisthosporidia?

Opisthosporidia is a group of reduced, endoparasite taxa that includes Rozellomycota, Microsporidia (parasites of animals), and Aphelidiomycota (parasites of algae for which no genomes are currently available; Figure S1). Within Opisthosporidia, our concatenation-based analyses strongly supported a clade of Rozellomycota + Microsporidia (Figures 2 and 3A). To date, only two Rozellomycota genomes have been sequenced, *Paramicrosporidium saccamoebae*[30] and *Rozella allomycis*.[36] Both concatenation- and coalescent-based analyses placed *P. saccamoebae* sister to Microsporidia, suggesting that Rozellomycota is paraphyletic (Figures 2 and 3). These results are largely consistent with previous gene content and phylogenetic

analyses that *P. saccamoebae* is more closely related to Microsporidia than to other Rozellomycota (Figure S1).[30] In contrast, the two approaches differed in the placement of *R. allomycis* (Figures 3 and S5). Whereas concatenation-based analyses placed *R. allomycis* sister to the *P. saccamoebae* + Microsporidia clade (Figures 3A and S5A), coalescent-based analyses placed *R. allomycis* as sister to the remaining non-Opisthosporidia fungi with very low support (LPP = 0.07; Figures 3B and S5B). Finally, quartet tree support for the concatenation-based placement (q1 = 0.31) was lower than the coalescent-based placement (q2 = 0.38), but a polytomy scenario could not be rejected (Figure 4A).[36]

Given that only two genomes from Rozellomycota and none from Aphelidiomycota are available, the lack of resolution within Opisthosporidia may be due to scarce taxon sampling. Although previous phylogenomic analyses based on a single transcriptome from Aphelidiomycota placed this phylum as sister to free-living fungi,[8] which would render Opisthosporidia paraphyletic, further studies with more taxa will be necessary to confidently resolve relationships in this lineage.

Did Blastocladiomycota split before or after Chytridiomyceta?

The relationships between flagellated zoosporic fungi in the Blastocladiomycota and Chytridiomyceta lineages and the rest of fungi (excluding Opisthosporidia) remain ambiguous.[5,44–46] Our concatenation analyses placed Blastocladiomycota as sister to a clade of Chytridiomyceta and the rest of fungi with strong support (BS = 99%; Figure 3A). In contrast, coalescent-based analyses strongly supported a sister taxon relationship between Blastocladiomycota and Chytridiomyceta (LPP = 1.00; Figure 3B). The quartet-based analyses showed low support for the concatenation-based placement (q1 = 0.24), intermediate support for Chytridiomyceta as sister to a clade of Blastocladiomycota and the rest of fungi (q2 = 0.31), and strong support for the coalescent-based placement (q3 = 0.45; Figure 4B). The low resolution of relationships between Blastocladiomycota and Chytridiomyceta in our coalescent-based analysis might be due to the lower taxon occupancy in these two clades (average of taxon occupancy: 73.68% in Chytridiomyceta; 42.59% in Blastocladiomycota; Data S2). Blastocladiomycota are zoospore-producing fungi that have been previously shown to be phylogenetically distinct from Chytridiomyceta and have characteristics that more resemble terrestrial fungi, such as well-developed hyphae, closed mitosis, cell walls with β-1-3-glucan, and a Spitzenkörper.[47,48] Thus, understanding the true branching order has important implications for the evolution of key traits and processes (e.g., life cycles and mitosis).[5] Interestingly, a recent study suggested that two atypical zoosporic fungi, *Amoeboradix gromovi* and *Sanchytrium tribonematis*, represent a new fungal phylum sister to Blastocladiomycota.[49]

Within the subkingdom Chytridiomyceta, phylogenetic relationships among Monoblepharidomycota, Chytridiomycota, and Neocallimastigomycota are also uncertain.[45,50] Our concatenation analyses recovered Chytridiomycota as the sister group to Monoblepharidomycota + Neocallimastigomycota (BS = 85%; Figures 3A and S5A), whereas coalescent analyses recovered Monoblepharidomycota as the sister to Chytridiomycota + Neocallimastigomycota (LPP = 0.18; Figures 3B and S5B). The quartet-based analyses showed lower support for the concatenation-based placement (q1 = 0.22) than for the coalescent-based placement (q2 = 0.41) or the third alternative hypothesis (q3 = 0.38; Figure 4C). Given that one genome was sampled from Monoblepharidomycota, 13 genomes were sampled from Chytridiomycota, and five genomes were sampled from Neocallimastigomycota, additional sampling of taxa, and perhaps genes as well, will be necessary for the confident resolution of relationships within Chytridiomyceta. Interestingly, a recent phylogenomic study placed the zoosporic obligate endoparasite *Olpidium bornovanus* as the closest zoosporic relative of the non-flagellated terrestrial fungi.[51]

Is zygomycetes monophyletic?

The monophyly of zygomycetes was not supported in recent phylogenetic studies, and relationships among these fungi are uncertain.[10,15,45,50] Consequently, several recent classifications have split zygomycetes into multiple subphyla and phyla, including Zoopagomycota and Mucoromycota.[11,15] Our concatenation analyses strongly supported the monophyly of Zoopagomycota and Mucoromycota (BS = 100%; Figures 3A and S5A).[11,15] Coalescent analyses recovered Mucoromycota as monophyletic, although as mentioned earlier, Chytridiomyceta and Blastocladiomycota are nested within Zoopagomycota in these coalescent-based phylogenies (Figures 3B and S5B). The quartet-based analysis shows that the quartets for the monophyly of Zoopagomycota and Mucoromycota received the highest support (q1 = 0.48; Figure 4D).

However, we found one subsampled data matrix (Top100_slow-evolving data matrix) that recovered the paraphyly of zygomycetes, albeit with very low support (BS = 28%; Figure S6B). This recovered topology is largely consistent with previous analyses, and Zoopagomycota is also recovered as monophyletic (BS = 28%).

To further explore the effect of gene sampling on the resolution of zygomycetes in different phylogenomic data matrices, we next quantified the support of phylogenetic signal over two alternative hypotheses (T1: zygomycetes-monophyly; T2: zygomycetes-paraphyly) using our Subset_Dikarya data matrix (see STAR methods) and a previously published 192-gene, 46-taxon data matrix (Spatafora2016_46taxa_192 genes data matrix; Figure 5; Data S4).[15] By calculating genewise log-likelihood scores between T1 and T2 (ΔlnL) for every gene in each matrix, we found that the proportions of genes supporting T1 versus T2 were similar in both data matrices (95 of 192: 49.5% versus 97 of 192: 50.5% in the Spatafora2016_46taxa_192 genes matrix; 161 of 290: 55.5% versus 129 of 290: 44.5% in the Subset_Dikarya data matrix; Figure 5), even though the results of our study support zygomycetes monophyly[52] and those of other studies support zygomycetes paraphyly.[10,15,45] Thus, phylogenomic analyses of zygomycetes should be interpreted with caution until further taxon and gene sampling of taxa from the lineages in question sheds more light onto this part of the fungal phylogeny.

Is Zoopagomycota paraphyletic?

Zoopagomycota, a group of pathogenic and saprophytic fungi,[53] are thought to be a monophyletic group based on previous phylogenomic analyses.[15,54] Surprisingly, we found that Zoopagomycota was paraphyletic because two *Basidiobolus* species

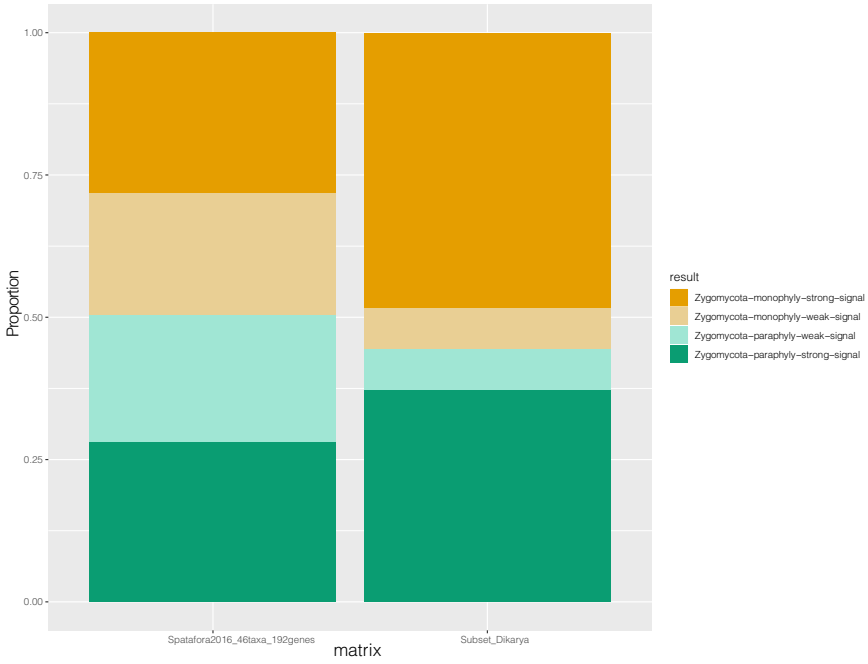


Figure 5. Distribution of phylogenetic signal for two alternative hypotheses on the zygomycetes lineage

The two alternative hypotheses are Mucoromycota is sister to Zoopagomycota (zygomycetes-monophyly; T1 orange) and Mucoromycota is sister to Dikarya (zygomycetes-paraphyly; T2 green). Proportions of genes supporting each of two alternative hypotheses in the Spatafora2016_46taxa_192 genes and Subset_Dikaya data matrices are shown. The difference in gene-wise log-likelihood scores between T1 and T2 (ΔGLS values) for each gene in each data matrix are provided in Data S4. We considered a gene with an absolute value of log-likelihood difference of two as a gene with strong (|ΔlnL| > 2) or weak (|ΔlnL| < 2) phylogenetic signal. See also Data S4.

were placed as the sister group to Mucoromycota (Figures 2, 3, and S3). The phylogenetic placement of *Basidiobolus* in previous phylogenetic analyses based on genomic[15] or multigene[55] studies was unstable, and a recent study has suggested that many genes in *Basidiobolus* genomes might have been acquired from Bacteria through horizontal gene transfers.[56] Notably, removal of the two *Basidiobolus* taxa in the removal-of-rogue-taxa data matrix did not alter the monophyly of zygomycetes (Figure S6A), suggesting that this result was not affected by the topological instability of *Basidiobolus*.

What are the relationships of subphyla within Zoopagomycota?

The evolutionary relationships of the three subphyla within Zoopagomycota are still uncertain, with either Entomophthoromycotina[15] or Zoopagomycotina[54] sister to the remaining Zoopagomycota. Our concatenation-based analyses recovered Zoopagomycotina as sister to Kickxellomycotina and Entomophthoromycotina with strong support (BS = 100%; Figure 2). This relationship is also supported in our quartet-based analysis (q1 = 0.41; q2 = 0.32; q3 = 0.27; Figure 4E).

Is Mortierellomycotina or Glomeromycotina sister to the rest of Mucoromycota?

Within Mucoromycota, the concatenation-based analysis moderately supported Mortierellomycotina as sister to Mucoromycotina and Glomeromycotina (BS = 98%), whereas the coalescent-based analysis placed Glomeromycotina sister to the remaining Mucoromycota with low support (LPP = 0.61; Figures 3 and S5). Quartet-tree support for the concatenation-based phylogeny was largely similar to the two alternative hypotheses (q1 = 0.33; q2 = 0.31; q3 = 0.36; Figure 4F), suggesting that a polytomy best explains relationships between subphyla of Mucoromycota based on current evidence. Nevertheless, the small number of genomes sampled suggests that these inferences may be subject to revision.

Is there a hard polytomy at the base of Basidiomycota?

Even though Basidiomycota have much denser taxon sampling than most other fungal lineages, reconstruction of the relationships among Pucciniomycotina, Ustilaginomycotina, and Agaricomycotina + Wallemiomycotina has proven challenging.[31,34,57,58] We too found discordant topologies between concatenation- and coalescent-based analyses (Figures 3 and S5) and nearly equal support for the three alternative hypotheses (Figure 4G). Concatenation analyses placed Ustilaginomycotina with Agaricomycotina + Wallemiomycotina (BS = 100%), whereas coalescence supported Pucciniomycotina + Ustilaginomycotina (LPP = 0.41). Notably, we found that gene-tree quartet support for the three alternative hypotheses was consistent with a polytomy (q1 = 0.33; q2 = 0.34; q3 = 0.34; Figure 4G). These results fail to reject the hypothesis that major relationships among Basidiomycota represent a hard polytomy (Figure 4G), consistent with a previous study[58] that used fewer taxa and genes (67 taxa/134 genes); however, Bayesian Markov chain Monte Carlo (MCMC) and likelihood mapping analyses led the study's authors to infer that the lack of resolution at the base of Basidiomycota does not represent a hard polytomy.[58] What is increasingly clear is that the origin of major lineages within Basidiomycota are likely the result of an ancient diversification. It should be noted that the Entorrhizomycota taxa were not sampled here.

Higher level taxonomic ranks generally reflect levels of evolutionary and molecular clock divergence across the fungal kingdom

The availability of a taxon-rich, genome-scale phylogeny for fungi provides an opportunity to evaluate the degree to which current fungal taxonomy reflects fungal evolutionary relationships and rates of fungal genome evolution. To test this, we normalized the fungal taxonomy ranks retrieved from the National Center for Biotechnology Information (NCBI) using the relative evolutionary divergence (RED) approach.[59] The RED approach normalizes the inferred phylogenetic distances

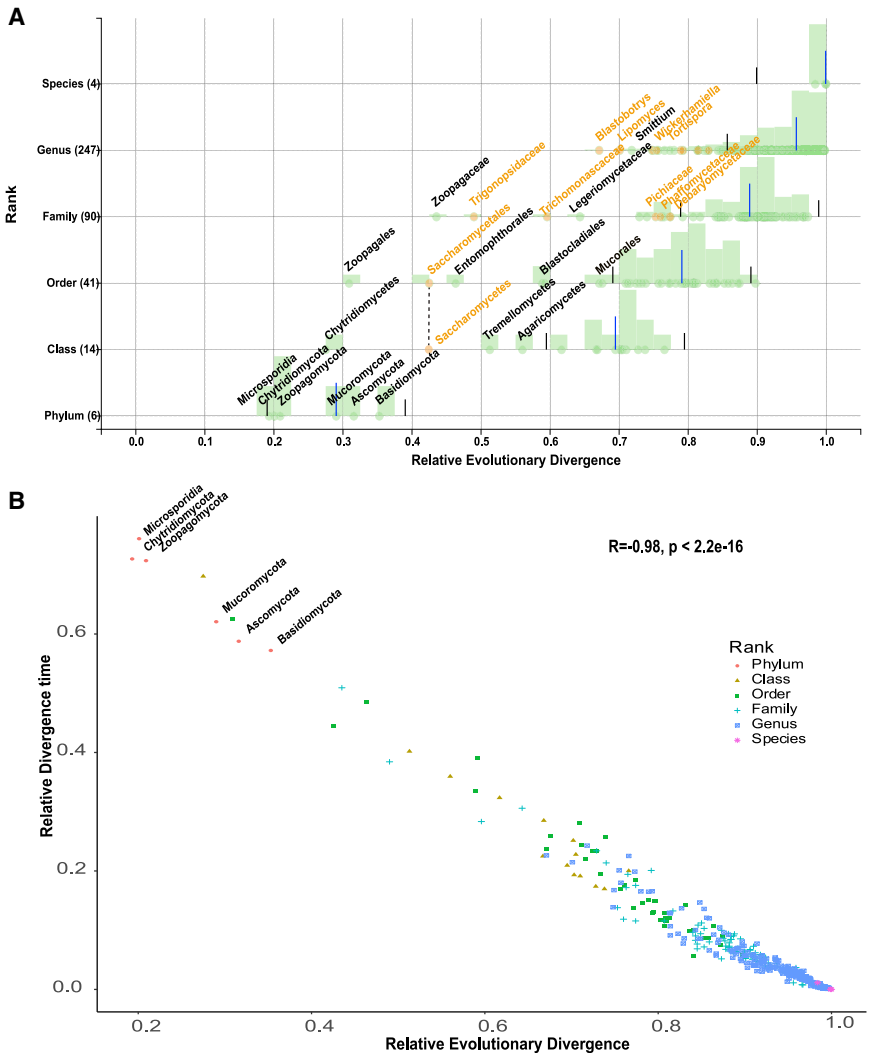


Figure 6. Higher level taxonomic ranks generally reflect levels of evolutionary divergence across the fungal kingdom

(A) Relative evolutionary divergence (RED) of taxa defined by the NCBI taxonomy based on the topology inferred from the concatenation-based single-model approach. Each data point (green or orange circle) represents a taxon distributed according to its RED value (x axis) and its taxonomic rank (y axis). Blue bars correspond to median RED values and black bars to the RED intervals (±0.1) for each rank. Orange circles represent taxa belonging to the subphylum Saccharomycotina (Ascomycota), which are the most notable instance of an underclassified lineage in the fungal kingdom. Note that RED values of ranks with a single subordinate rank will be identical to each other (e.g., class Saccharomycetes contains a single order, Saccharomycetales; thus, both ranks have the same RED value). Only a subset of taxon names is shown here; results for all taxa are reported in Data S5.

(B) The Pearson correlation coefficient (Pearson's r) between the RED values and relative divergence time estimated using relaxed molecular clock approaches for all internal nodes. The data points associated with six fungal phyla are shown for illustration purposes.

See also Data S5.

between the last common ancestor of fungi (RED = 0) to all extant fungal taxa (RED = 1) to provide an approximation of the relative amount of divergence (Figure 6A).

The RED approach was developed to revise taxonomy ranks in Bacteria and Archaea so that they reflect evolutionary divergence.[59,60] Although the RED approach has yet to be applied to fungi, several previous studies have suggested the use of divergence times as a ranking criterion.[61,62] Interestingly, we found that the RED values of fungal taxonomic ranks in our phylogeny are broadly consistent to their relative divergence times estimated using relaxed molecular clock approaches (Pearson's correlation coefficient $r = -0.98$; $p < 2.2e{-}16$; Figure 6B). Thus, our results suggest that RED and divergence time approaches capture similar aspects of evolutionary divergence and can be used to compare fungal taxonomy ranks in a phylogeny-informed way.

Of the 6 phyla, 14 classes, 41 orders, 90 families, and 247 genera examined (Figure 6A; Data S5), we found that ~85% of ranks fell within ±0.1 of the median RED value for taxa at that rank, suggesting they had comparable levels of evolutionary divergence. The only instance of a fungal rank that appears to be overclassified (i.e., has a much higher RED value than the rest) is the plant-associated order Diaporthales (RED = 0.897; average RED value for other fungal orders = 0.752). All other instances that were outside the ±0.1 RED interval concerned underclassification (i.e., ranks with a much lower RED value than the rest) and were concentrated on specific lineages. Remarkably, nearly 40% (22 of 49, including 1 order, 5 families, and 16 genera) of the underclassified ranks were within the Saccharomycotina subphylum of budding yeasts. Other underclassified taxa included classes Chytridiomycetes (2/49), Tremellomycetes (2/49), and Agaricomycetes (4/49).

The most underclassified lineage was order Zoopagales of Zoopagomycotina, whose RED value (0.309) was the lowest compared to other orders or classes included in our analysis. Because many Zoopagales are predacious or parasitic and non-culturable, all seven Zoopagales genomes have been sequenced using single-cell sequencing methods;[54] thus, it is possible the low RED value in this lineage stems from the typically higher nucleotide base calling errors of single-cell sequencing methods or from contamination. Moreover, it should be noted that the most serious instance of underclassification concerns the most well-sampled major lineage (Saccharomycotina). Thus, as the genomes of more species are sampled and added to the fungal phylogeny (especially from major lineages whose taxonomic diversity is not well represented in our phylogeny), it is possible that examination of RED values reveals further instances in the fungal tree of life, where classification is not on par with evolutionary divergence.

Taken together, these results suggest that the current fungal classification is largely concordant with our understanding of fungal phylogeny and evolutionary divergence. However, our results also identify lineages, such as Saccharomycotina, where taxonomic rank assignment appears to not truly reflect the observed levels of evolutionary divergence (compared to assignments in the rest of the fungal kingdom), reducing the utility of taxonomy for comparative fungal biology.

DISCUSSION

Fungi have undergone extensive diversification into numerous ecological roles, morphological forms, and genomic architectures over the last 1 Ga (Figure 1). Resolving relationships among major groups of the fungal tree has proven challenging due to the lack of data from organisms spanning fungal diversity and the relative paucity of phylogenomic studies for the entire kingdom. By synthesizing data from more than fifteen hundred publicly available genomes, we provide a robust phylogenetic framework to explore fungal evolution and examine sources of conflict and support for the backbone of the fungal phylogeny.

We find that most parts of the fungal phylogeny are robustly resolved with our 290-gene dataset, but a handful of challenging branches remain unresolved. We provide evidence that some of these relationships may actually reflect genuine instances of ancient evolutionary diversification events, or hard polytomies, such as those among subphyla in Basidiomycota. In contrast, other unresolved relationships likely stem from the relatively poor taxon and/or gene sampling of several fungal phyla, suggesting that improving the resolution of the fungal phylogeny will require continued efforts to sample genomes spanning the diversity of the fungal kingdom. This inference is further supported by the results of our examination of concatenation- and coalescent-based phylogenies from several different data matrices that vary in their gene and taxon occupancy, which also suggests that the elucidation of these unresolved relationships will likely require substantial additional data and analyses. In the case of the monophyly of the zygomycetes, we show that the distinction between a phylogenomic analysis recovering monophyly versus paraphyly rests on a handful of genes. As fungal phylogenomic analyses improve their gene and taxon sampling, it is important to be aware that, although the latest genome-scale phylogenies represent the currently best supported hypotheses, they are always potentially subject to revision and improvement. Given how often phylogenomic studies contradict each other on certain contentious relationships,[63,64] clearly identifying relationships that remain ambiguous, despite the many taxa, genes, and analyses, sets the stage for further exploration of contentious bipartitions by sampling additional taxa and genes. Furthermore, by quantifying the support for alternative hypotheses, our approach offers a way to illuminate controversial or ambiguous relationships and generate a more accurate fungal tree of life.

Finally, our study presents a novel examination of the relationship between the current state of taxonomic classification in fungi and genomic evolutionary divergence. Although fungal taxonomy broadly reflects evolutionary divergence, we identified instances of specific lineages, such as the subphylum Saccharomycotina, where the lack of correspondence hinders the utility of taxonomy as a yardstick for comparative biology. In conclusion, the generation and analyses of a phylogenomic data matrix from 1,644 species spanning the diversity of the kingdom establish an integrated and robust phylogenetic framework for studying the evolution of fungi.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Sequence data
- METHOD DETAILS
  - Quality assessment
  - Phylogenomic data matrix construction
  - Full – data matrix #1
  - Subset_Dikarya_taxa – data matrix #2
  - Top_100_DVMC – data matrix #3
  - Top_100_length – data matrix #4
  - Top100_low_LB – data matrix #5
  - Top100_low_RCFV – data matrix #6
  - Top100_low_saturation – data matrix #7
  - Top100_slow-evolving – data matrix #8
  - Top100_completeness – data matrix #9
  - Top100_high_ABS data matrix – data matrix #10
  - LB_taxa_removal – data matrix #11
  - Rogue_taxa_removal – data matrix #12
  - Phylogenomic analyses
  - Concatenation-based approach without and with data-partitioning
  - Coalescent-based approach
  - Quantification of incongruence
  - Polytomy test
  - Quantification of the distribution of phylogenetic signal
  - RED index
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2021.01.074.

ACKNOWLEDGMENTS

We thank members of the Rokas laboratory for discussions and comments, Phil Hugenholtz for encouraging us to use relative evolutionary distance to examine the correspondence between genome sequence divergence and taxonomic rank across fungi, and Donovan Parks for initial help with getting relative evolutionary distance analyses working. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University and Yale Center for Research Computing (Farnam HPC cluster) for use of the research computing infrastructure. Yuanning Li was partially supported by a scholarship from the China Scholarship Council (CSC) for studying and living abroad. This work was supported by the National Science Foundation (United States) (DEB-1442113 to

A.R.; DEB-1442148 to C.T.H.; DEB-1929738 and DEB-1441677 to T.Y.J.; DEB-1557110 and DEB-1441715 to J.E.S.; and DEB-1441604 to J.W.S.), in part by the DOE Great Lakes Bioenergy Research Center (United States) (DOE Office of Science BER DE-SC0018409 to C.T.H.), and the USDA National Institute of Food and Agriculture (Hatch project 1003258 to C.T.H.; Hatch project CA-R-PPA-5062-H to J.E.S.). C.T.H. is a Pew Scholar in the Biomedical Sciences and an H.I. Romnes Faculty Fellow, supported by the Pew Charitable Trusts and Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, respectively. X.-X.S. was supported by the National Natural Science Foundation of China (no. 32071665) and the Young Scholar 1000 Talents Plan. J.E.S. is a Fellow in the CIFAR program Fungal Kingdom: Threats and Opportunities. 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. received additional support from the Guggenheim Foundation, the Burroughs Wellcome Fund, and the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56 AI146096).

AUTHOR CONTRIBUTIONS

Y.L., X.-X.S., T.Y.J., J.E.S., J.W.S., C.T.H., and A.R. designed this study. Y.L., X.-X.S., and J.L.S. conducted analyses and prepared figures. Y.L. and A.R. wrote the paper. All authors discussed the results and implications and commented on the manuscript at all stages.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 11, 2020
Revised: December 10, 2020
Accepted: January 21, 2021
Published: February 18, 2021

REFERENCES

1. Heitman, J., Howlett, B.J., Crous, P.W., Stukenbrock, E.H., James, T.Y., and Gow, N.A.R. (2017). The Fungal Kingdom (John Wiley & Sons).
2. Blackwell, M. (2011). The fungi: 1, 2, 3 ... 5.1 million species? Am. J. Bot. *98*, 426–438.
3. Hawksworth, D.L., and Lücking, R. (2017). Fungal diversity revisited: 2.2 to 3.8 million species. In The Fungal Kingdom, J. Heitman, B.J. Howlett, P.W. Crous, E.H. Stukenbrock, T.Y. James, and N.A.R. Gow, eds. (Wiley), pp. 79–95.
4. Lutzoni, F., Nowak, M.D., Alfaro, M.E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A.E., Lewis, L.A., Swofford, D.L., Hibbett, D., et al. (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. Nat. Commun. *9*, 5451.
5. James, T.Y., Stajich, J.E., Hittinger, C.T., and Rokas, A. (2020). Toward a fully resolved fungal tree of life. Annu. Rev. Microbiol. *74*, 291–313.
6. Wijayawardene, N.N., Hyde, K.D., Al-Ani, L.K.T., Tedersoo, L., Haelewaters, D., Rajeshkumar, K.C., Zhao, R.L., Aptroot, A., Leontyev, D.V., Saxena, R.K., et al. (2020). Outline of Fungi and fungus-like taxa. Mycosphere *11*, 1060–1456.
7. Spatafora, J.W., Aime, M.C., Grigoriev, I.V., Martin, F., Stajich, J.E., and Blackwell, M. (2017). The fungal tree of life: from molecular systematics to genome-scale phylogenies. In The Fungal Kingdom, J. Heitman, B.J. Howlett, P.W. Crous, E.H. Stukenbrock, T.Y. James, and N.A.R. Gow, eds. (Wiley), pp. 1–34.
8. Torruella, G., Grau-Bové, X., Moreira, D., Karpov, S.A., Burns, J.A., Sebé-Pedrós, A., Völcker, E., and López-García, P. (2018). Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. Commun. Biol. *1*, 231.
9. Bass, D., Czech, L., Williams, B.A.P., Berney, C., Dunthorn, M., Mahé, F., Torruella, G., Stentiford, G.D., and Williams, T.A. (2018). Clarifying the relationships between Microsporidia and Cryptomycota. J. Eukaryot. Microbiol. *65*, 773–782.
10. James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., et al. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818–822.
11. Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T., Kirk, P.M., Lücking, R., et al. (2007). A higher-level phylogenetic classification of the Fungi. Mycol. Res. *111*, 509–547.
12. Alexopoulos, C.J., Mims, C.W., and Blackwell, M. (1996). Introductory Mycology (John Wiley and Sons).
13. Schüßler, A., Schwarzott, D., and Walker, C. (2001). A new fungal phylum, the *Glomeromycota*: phylogeny and evolution. Mycol. Res. *105*, 1413–1421.
14. Capella-Gutiérrez, S., Marcet-Houben, M., and Gabaldón, T. (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol. *10*, 47.
15. Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia *108*, 1028–1046.
16. Varga, T., Krizsán, K., Földi, C., Dima, B., Sánchez-García, M., Sánchez-Ramírez, S., Szöllősi, G.J., Szarkándi, J.G., Papp, V., Albert, L., et al. (2019). Megaphylogeny resolves global patterns of mushroom evolution. Nat. Ecol. Evol. *3*, 668–678.
17. Kiss, E., Hegedüs, B., Virágh, M., Varga, T., Merényi, Z., Kószó, T., Bálint, B., Prasanna, A.N., Krizsán, K., Kocsubé, S., et al. (2019). Comparative genomics reveals the origin of fungal hyphae and multicellularity. Nat. Commun. *10*, 4080.
18. Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T., et al. (2018). Tempo and mode of genome evolution in the budding yeast subphylum. Cell *175*, 1533–1545.e20.
19. Shen, X.-X., Steenwyk, J.L., LaBella, A.L., Opulente, D.A., Zhou, X., Kominek, J., Li, Y., Groenewald, M., Hittinger, C.T., and Rokas, A. (2020). Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. Sci. Adv. *6*, 45.
20. Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. *42*, D699–D704.
21. Hittinger, C.T., Rokas, A., Bai, F.Y., Boekhout, T., Gonçalves, P., Jeffries, T.W., Kominek, J., Lachance, M.A., Libkind, D., Rosa, C.A., et al. (2015). Genomics and the making of yeast biodiversity. Curr. Opin. Genet. Dev. *35*, 100–109.
22. Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A., Andreopoulos, B., Baker, S.E., Barry, K., Bills, G., et al. (2020). 101 *Dothideomycetes* genomes: a test case for predicting lifestyles and emergence of pathogens. Stud. Mycol. *96*, 141–153.
23. Kjærbølling, I., Vesth, T.C., Frisvad, J.C., Nybo, J.L., Theobald, S., Kuo, A., Bowyer, P., Matsuda, Y., Mondo, S., Lyhne, E.K., et al. (2018). Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. Proc. Natl. Acad. Sci. USA *115*, E753–E761.
24. Torruella, G., Derelle, R., Paps, J., Lang, B.F., Roger, A.J., Shalchian-Tabrizi, K., and Ruiz-Trillo, I. (2012). Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Mol. Biol. Evol. *29*, 531–544.
25. Brown, M.W., Spiegel, F.W., and Silberman, J.D. (2009). Phylogeny of the "forgotten" cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. Mol. Biol. Evol. *26*, 2699–2709.
26. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO

applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. *35*, 543–548.

27. Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A., Ioannidis, P., Seppey, M., Loetscher, A., and Kriventseva, E.V. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. *45* (D1), D744–D749.
28. Struck, T.H. (2014). TreSpEx--detection of misleading signal in phylogenetic reconstructions based on tree information. Evol. Bioinform. Online *10*, 51–67.
29. Kocot, K.M., Struck, T.H., Merkel, J., Waits, D.S., Todt, C., Brannock, P.M., Weese, D.A., Cannon, J.T., Moroz, L.L., Lieb, B., and Halanych, K.M. (2017). Phylogenomics of Lophotrochozoa with consideration of systematic error. Syst. Biol. *66*, 256–282.
30. Quandt, C.A., Beaudet, D., Corsaro, D., Walochnik, J., Michel, R., Corradi, N., and James, T.Y. (2017). The genome of an intranuclear parasite, *Paramicrosporidium saccamoebae*, reveals alternative adaptations to obligate intracellular parasitism. eLife *6*, E29594.
31. Padamsee, M., Kumar, T.K.A., Riley, R., Binder, M., Boyd, A., Calvo, A.M., Furukawa, K., Hesse, C., Hohmann, S., James, T.Y., et al. (2012). The genome of the xerotolerant mold Wallemia sebi reveals adaptations to osmotic stress and suggests cryptic sexual reproduction. Fungal Genet. Biol. *49*, 217–226.
32. Zajc, J., Liu, Y., Dai, W., Yang, Z., Hu, J., Gostinčar, C., and Gunde-Cimerman, N. (2013). Genome and transcriptome sequencing of the halophilic fungus *Wallemia ichthyophaga*: haloadaptations present and absent. BMC Genomics *14*, 617.
33. Zhao, R.-L., Li, G.-J., Sánchez-Ramírez, S., Stata, M., Yang, Z.-L., Wu, G., Dai, Y.-C., He, S.-H., Cui, B.-K., Zhou, J.-L., et al. (2017). A six-gene phylogenetic overview of *Basidiomycota* and allied phyla with estimated divergence times of higher taxa and a phyloproteomics perspective. Fungal Divers. *84*, 43–74.
34. Choi, J., and Kim, S.-H. (2017). A genome Tree of Life for the Fungi kingdom. Proc. Natl. Acad. Sci. USA *114*, 9391–9396.
35. Li, Y., David, K.T., Shen, X.-X., Steenwyk, J.L., Halanych, K.M., and Rokas, A. (2020). Feature frequency profile-based phylogenies are inaccurate. Proc. Natl. Acad. Sci. USA *117*, 31580–31581.
36. James, T.Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J.E. (2013). Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. Curr. Biol. *23*, 1548–1553.
37. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.
38. Rokas, A., Williams, B.L., King, N., and Carroll, S.B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature *425*, 798–804.
39. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19* (*Suppl 6*), 153.
40. Gatesy, J., and Springer, M.S. (2014). Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. *80*, 231–266.
41. Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., et al. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. *94* (Pt A), 447–462.
42. Shen, X.-X., Li, Y., Hittinger, C.T., Chen, X.-X., and Rokas, A. (2020). An investigation of irreproducibility in maximum likelihood phylogenetic inference. Nat. Commun. *11*, 6096.
43. Sayyari, E., and Mirarab, S. (2018). Testing for polytomies in phylogenetic species trees using quartet frequencies. Genes (Basel) *9*, 132.
44. Ebersberger, I., de Matos Simoes, R., Kupczok, A., Gube, M., Kothe, E., Voigt, K., and von Haeseler, A. (2012). A consistent phylogenetic backbone for the fungi. Mol. Biol. Evol. *29*, 1319–1334.
45. Chang, Y., Wang, S., Sekimoto, S., Aerts, A.L., Choi, C., Clum, A., LaButti, K.M., Lindquist, E.A., Yee Ngan, C., Ohm, R.A., et al. (2015). Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. Genome Biol. Evol. *7*, 1590–1601.
46. Torruella, G., de Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M.A., del Campo, J., Eme, L., Pérez-Cordón, G., Whipps, C.M., Nichols, K.M., et al. (2015). Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. Curr. Biol. *25*, 2404–2410.
47. Ruiz-Herrera, J., and Ortiz-Castellanos, L. (2019). Cell wall glucans of fungi. A review. Cell Surf. *5*, 100022.
48. Dee, J.M., Mollicone, M., Longcore, J.E., Roberson, R.W., and Berbee, M.L. (2015). Cytology and molecular phylogenetics of Monoblepharidomycetes provide evidence for multiple independent origins of the hyphal habit in the Fungi. Mycologia *107*, 710–728.
49. Galindo, L.J., Lopez-Garcia, P., Torruella, G., Karpov, S.A., and Moreira, D. (2020). Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. bioRxiv, 2020.09.16.389700.
50. Liu, Y., Steenkamp, E.T., Brinkmann, H., Forget, L., Philippe, H., and Lang, B.F. (2009). Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. BMC Evol. Biol. *9*, 272.
51. Chang, Y., Rochon, D., Sekimoto, S., Wang, Y., Chovatia, M., Sandor, L., et al. (2021). Genome-scale phylogenetic analyses confirm *Olpidium* as the closest living zoosporic fungus to the non-flagellated, terrestrial fungi. Scientific Reports *11*, https://doi.org/10.1038/s41598-021-82607-4.
52. Sekimoto, S., Rochon, D., Long, J.E., Dee, J.M., and Berbee, M.L. (2011). A multigene phylogeny of Olpidium and its implications for early fungal evolution. BMC Evol. Biol. *11*, 331.
53. Berbee, M.L., James, T.Y., and Strullu-Derrien, C. (2017). Early diverging fungi: diversity and impact at the dawn of terrestrial life. Annu. Rev. Microbiol. *71*, 41–60.
54. Davis, W.J., Amses, K.R., Benny, G.L., Carter-House, D., Chang, Y., Grigoriev, I., Smith, M.E., Spatafora, J.W., Stajich, J.E., and James, T.Y. (2019). Genome-scale phylogenetics reveals a monophyletic Zoopagales (Zoopagomycota, Fungi). Mol. Phylogenet. Evol. *133*, 152–163.
55. Gryganskyi, A.P., Humber, R.A., Smith, M.E., Hodge, K., Huang, B., Voigt, K., and Vilgalys, R. (2013). Phylogenetic lineages in Entomophthoromycota. Persoonia *30*, 94–105.
56. Tabima, J.F., Trautman, I.A., Chang, Y., Wang, Y., Mondo, S., Kuo, A., Salamov, A., Grigoriev, I.V., Stajich, J.E., and Spatafora, J.W. (2020). Phylogenomic analyses of non-Dikarya fungi supports horizontal gene transfer driving diversification of secondary metabolism in the amphibian gastrointestinal symbiont, *Basidiobolus*. G3 (Bethesda) *10*, 3417–3433.
57. He, M.-Q., Zhao, R.-L., Hyde, K.D., Begerow, D., Kemler, M., Yurkov, A., McKenzie, E.H.C., Raspé, O., Kakishima, M., Sánchez-Ramírez, S., et al. (2019). Notes, outline and divergence times of Basidiomycota. Fungal Divers. *99*, 105–367.
58. Prasanna, A.N., Gerber, D., Kijpornyongpan, T., Aime, M.C., Doyle, V.P., and Nagy, L.G. (2020). Model choice, missing data, and taxon sampling impact phylogenomic inference of deep Basidiomycota relationships. Syst. Biol. *69*, 17–37.
59. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. *36*, 996–1004.
60. Rinke, C., Chuvochina, M., Mussig, A.J., Chaumeil, P.-A., Waite, D.W., Whitman, W.B., Parks, D.H., and Hugenholtz, P. (2020). A rank-normalized archaeal taxonomy based on genome phylogeny resolves widespread incomplete and uneven classifications. bioRxiv, 2020.03.01.972265.
61. Avise, J.C., and Johns, G.C. (1999). Proposal for a standardized temporal scheme of biological classification for extant species. Proc. Natl. Acad. Sci. USA *96*, 7358–7363.

62. Tedersoo, L., Sánchez-Ramírez, S., Kõljalg, U., Bahram, M., Döring, M., Schigel, D., May, T., Ryberg, M., and Abarenkov, K. (2018). High-level classification of the Fungi and a tool for evolutionary ecological analyses. Fungal Divers. *90*, 135–159.

63. Shen, X.-X., Hittinger, C.T., and Rokas, A. (2017). Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. *1*, 126.

64. Li, Y., Shen, X.-X., Evans, B., Dunn, C.W., and Rokas, A. (2020). Rooting the animal tree of life. bioRxiv. https://doi.org/10.1101/2020.10.27.357798.

65. Zhang, Z., and Wood, W.I. (2003). A profile hidden Markov model for signal peptides generated by HMMER. Bioinformatics *19*, 307–308.

66. Gertz, E.M., Yu, Y.-K., Agarwala, R., Schäffer, A.A., and Altschul, S.F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol. *4*, 41.

67. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. *34*, W435–W439.

68. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

69. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

70. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics *30*, i541–i548.

71. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534.

72. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

73. Team, R.C., Team, M.R.C., Suggests, M., and Matrix, S. (2018). Package stats (The R Stats Package).

74. Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *47* (W1), W256–W259.

75. Shen, X.-X., Zhou, X., Kominek, J., Kurtzman, C.P., Hittinger, C.T., and Rokas, A. (2016). Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. G3 (Bethesda) *6*, 3927–3939.

76. Steenwyk, J.L., Shen, X.-X., Lind, A.L., Goldman, G.H., and Rokas, A. (2019). A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. MBio *10*, e00925-19.

77. Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. *27*, 401–410.

78. Struck, T.H., Nesnidal, M.P., Purschke, G., and Halanych, K.M. (2008). Detecting possibly saturated positions in 18S and 28S sequences and their influence on phylogenetic reconstruction of Annelida (Lophotrochozoa). Mol. Phylogenet. Evol. *48*, 628–645.

79. Whelan, N.V., Kocot, K.M., Moroz, L.L., and Halanych, K.M. (2015). Error, signal, and the placement of Ctenophora sister to all other animals. Proc. Natl. Acad. Sci. USA *112*, 5773–5778.

80. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.

81. Salichos, L., and Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. Nature *497*, 327–331.

82. Aberer, A.J., Krompass, D., and Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Syst. Biol. *62*, 162–166.

83. Zhou, X., Shen, X.-X., Hittinger, C.T., and Rokas, A. (2018). Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. Mol. Biol. Evol. *35*, 486–503.

84. Le, S.Q., Lartillot, N., and Gascuel, O. (2008). Phylogenetic mixture models for proteins. Philos. Trans. R. Soc. Lond. B Biol. Sci. *363*, 3965–3976.

85. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

86. Salichos, L., Stamatakis, A., and Rokas, A. (2014). Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol. Biol. Evol. *31*, 1261–1271.

87. Smith, S.A., Walker-Hale, N., Walker, J.F., and Brown, J.W. (2020). Phylogenetic conflicts, combinability, and deep phylogenomics in plants. Syst. Biol. *69*, 579–592.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genome assemblies	NCBI	See Table S1; Zenodo repository: ; http://10.6084/m9.figshare.12751736
Phylogenetic data matrices	This paper	Figshare repository: https://figshare.com/articles/dataset/Scripts_and_analyses_used_for_the_fungal_phylogeny/12751736; http://10.6084/m9.figshare.12751736
Phylogenetic trees	This paper	Figshare repository: https://figshare.com/articles/dataset/Scripts_and_analyses_used_for_the_fungal_phylogeny/12751736; http://10.6084/m9.figshare.12751736
Software and algorithms		
BUSCO v2.02.1	Waterhouse et al.[26]	https://busco.ezlab.org/
HMMER v3.1b2	Zhang and Wood[65]	http://hmmer.org
OrthoDB v9	Zdobnov et al.[27]	https://busco.ezlab.org/
tBLASTn	Gertz et al.[66]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
AUGUSTUS v2.5.5	Stanke et al.[67]	http://bioinf.uni-greifswald.de/augustus/downloads/
MAFFT v7.299	Katoh and Standley[68]	https://mafft.cbrc.jp/alignment/software/
trimAl v1.4	Capella-Gutiérrez et al.[69]	http://trimal.cgenomics.org/
Gotree v1.13.6	https://github.com/evolbioinfo/gotree	https://github.com/evolbioinfo/gotree
ASTRAL-III v5.1.1	Mirarab et al.[70]	https://github.com/smirarab/ASTRAL
IQ-TREE v1.6.8	Minh et al.[71]	http://www.iqtree.org/
PhyloRank v0.0.37	Parks et al.[59]	https://github.com/dparks1134/PhyloRank/
MEGA7	Kumar et al.[72]	https://mafft.cbrc.jp/alignment/software/
R package stats v3.6.2	R.C. Team et al.[73]	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html
ITOL v3	Letunic and Bork[74]	https://itol.embl.de/help.cgi#batch

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Antonis Rokas (antonis.rokas@vanderbilt.edu).

Materials availability

There are no materials to report.

Data and code availability

All genome assemblies were downloaded from NCBI and are publicly available in the Zenodo repository: https://zenodo.org/record/3970286. All scripts, data matrices, and phylogenetic trees are deposited at Figshare repository: https://figshare.com/articles/dataset/Scripts_and_analyses_used_for_the_fungal_phylogeny/12751736. Original data have been deposited to Zenodo repository: 10.5281/zenodo.3970286 and Figshare repository: 10.6084/m9.figshare.12751736.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sequence data

All 1,679 fungal genomes were downloaded from NCBI and only one representative genome from every species was included (last accession date: January 30, 2020). Moreover, the genomes of 28 outgroup taxa (11 representative taxa from Holozoa and 17 representative taxa from Metazoa) were downloaded from Ensembl or NCBI (Last accession date: January 1, 2020). The outgroups were

selected based on the current understanding of Opisthokonta phylogeny[24,25]. NCBI taxonomy, strain ID, and source information in this study are also provided in Data S1.

METHOD DETAILS

Quality assessment

To assess the qualities of the genome assemblies of the 1,679 fungal genomes we used the Benchmarking Universal Single-Copy Orthologs (BUSCO), version 2.02.1[26] and the Fungi odb9 database (Last accession date: January 15, 2020). Briefly, BUSCO uses a consensus sequence built from a hidden Markov model-based alignment of orthologous sequences derived from 85 different fungal species using HMMER, version 3.1b2[65], as a query in tBLASTn[66] to search an individual genome. A total of 290 predefined orthologs (referred to as fungal BUSCO genes) were used. To examine the presence of each BUSCO gene in a genome, gene structure was predicted using AUGUSTUS, version 2.5.5[67], with default parameters, from the nucleotide coordinates of putative genes identified using BLAST and then aligned to the HMM alignment of the same BUSCO gene. Genes were considered ''single-copy'' if there was only one complete predicted gene present in the genome, ''duplicated'' if there were two or more complete predicted genes for one BUSCO gene, ''fragmented'' if the predicted gene was shorter than 95% of the aligned sequence lengths from the 85 different fungal species, and ''missing'' if there was no predicted gene. For each genome, the fraction of single-copy BUSCO genes present corresponded to the completeness of each genome. To minimize missing data and remove potential low-quality genomes, we retained only those genomes that contained 100 or more single-copy BUSCO genes. The final dataset contained 1,644 fungi and 28 outgroup taxa (Data S1).

Phylogenomic data matrix construction

In addition to their use as a measure of genome completeness, BUSCO genes have also been widely used as markers for phylogenomic inference in diverse lineages[26], especially in exploring fungi relationships[18,19,75,76]. Therefore, we used the BUSCO genes to generate the full data matrix (1,672 taxa / 290 genes), as well as 11 additional data matrices by subsampling subsets of taxa or BUSCO genes. We used these 12 data matrices to assess the stability of phylogenetic relationships and identify putative sources of error in our analyses (Figure S2).

Full – data matrix #1

To construct the full data matrix, we only included single-copy BUSCO genes for each species. For each BUSCO gene, we extracted individual nucleotide sequences that have the BUSCO gene present and translated to amino acid sequences with their corresponding codon usage for each taxon (CUG-Ser1, CUG-Ser2 clades in yeasts: NCBI genetic code 12; CUG-Ala clades in yeasts: NCBI genetic code 26; all others: NCBI standard genetic code 1). Each gene was aligned with MAFFT version 7.299[68] with options ''—auto –maxiterate 1000.'' Ambiguously aligned regions were removed using trimAl version 1.4[69] with the ''gappyout'' option. The AA alignments of these 290 BUSCO genes, each of which has more than 50% of taxon occupancy, were then concatenated into the full data matrix, which contains 124,700 amino acid sites.

Subset_Dikarya_taxa – data matrix #2

Our taxon sampling is biased toward Ascomycota and Basidiomycota (Dikarya), especially in Saccharomycotina (332 taxa; 20.1% total), Pezizomycotina (758 taxa; 46% total), and Agaricomycotina (321 taxa; 19.5% total). To discern the potential effects of biased taxon sampling (i.e., effects associated with the tree search algorithm spending most time in those parts of the tree that contain the largest numbers of taxa than in the other, less well sampled, parts of the tree), we subsampled one representative of each genus in Saccharomycotina (reducing their sampling from 332 taxa to 79; 14.6% total), and one representative of each family in Pezizomycotina (758 - > 108 taxa; 20.0% total) and in Agaricomycotina (321 - > 92 taxa; 17.0%). This sampling resulted in a data matrix with 540 taxa and 124,700 amino acid sites.

Top_100_DVMC – data matrix #3

This data matrix was constructed by retaining the top 100 BUSCO genes whose evolutionary rates were most ''clock-like'' (inferred by examining the degree of violation of a molecular clock (DVMC) values among single-gene trees[76]) and contains 51,494 amino acid sites (from all 1,672 taxa). DVMC is the standard deviation of root to tip distances in a phylogeny.

Top_100_length – data matrix #4

This data matrix was constructed by retaining the top 100 BUSCO genes with the longest alignment lengths after trimming and contains 75,529 amino acid sites (from all 1,672 taxa).

Top100_low_LB – data matrix #5

Long-Branch (LB) scores are widely used as a measurement for identifying genes that might be subject to long branch attraction[77]. LB score is the average of the upper quartile of the tip-to-root distances in a phylogeny and was calculated for each BUSCO gene

using a customized python script (available at https://github.com/JLSteenwyk/Phylogenetic_scripts/blob/master/LB_score.py). This data matrix was constructed by retaining the top 100 BUSCO genes with the lowest average LB scores and contains 39,347 amino acid sites (from all 1,672 taxa).

Top100_low_RCFV – data matrix #6

This data matrix was constructed by retaining the 100 BUSCO genes with the lowest relative composition frequency variability (RCFV)[28]. Base composition heterogeneity can potentially influence phylogenetic analysis; one way to assess it is using the RCFV value measured from the frequencies of the amino acid or nucleotide data in each BUSCO gene alignment[28]. The RCFV value for each gene was calculated following the protocols outlined by a previous study[18]. This data matrix contains 60,647 amino acid sites (from all 1,672 taxa).

Top100_low_saturation – data matrix #7

This data matrix was constructed by retaining the 100 BUSCO genes with the highest values of the slope of patristic distance – i.e., sum of the lengths of the branches that link two nodes in a tree – versus uncorrected p-distance (larger slope values denote lower levels of saturation than smaller values), which are thought to improve phylogenetic inference[29,78]. Slope values were measured by TreSpEx[28]. This data matrix contains 32,947 amino acid sites (from all 1,672 taxa).

Top100_slow-evolving – data matrix #8

This data matrix was constructed by retaining the 100 BUSCO genes with the lowest values of average pairwise patristic distance, which has previously been used to evaluate if fast-evolving genes bias phylogenetic inference[29,79]. The average patristic distance of each gene was measured by TreSpEx[28]. This data matrix contains 33,111 amino acid sites (from all 1,672 taxa).

Top100_completeness – data matrix #9

This data matrix was constructed by retaining the 100 BUSCO genes with the highest taxon occupancy. This data matrix contains 42,731 amino acid sites (from all 1,672 taxa).

Top100_high_ABS data matrix – data matrix #10

This data matrix was constructed by retaining the top 100 genes with the highest average bootstrap support (ABS) value of all internal branches on the gene tree in R package ape[80], which has previously been shown to improve inference[81]. This data matrix contains 71,225 amino acid sites (from all 1,672 taxa).

LB_taxa_removal – data matrix #11

Long-Branch (LB) scores can also be used to identify taxa that might be subject to long branch attraction[77]. By examining the distribution of LB scores among sampled taxa, we identified one large break (LB score > 79.0) between taxa (Data S2). Thus, we constructed this data matrix by removing the 23 taxa with an LB score > 79.0; the LB score was measured by a customized python script (https://github.com/JLSteenwyk/Phylogenetic_scripts/blob/master/LB_score.py). All 23 removed taxa were from the Microsporidia lineage. This removal resulted in a data matrix with 1,649 taxa and 124,700 amino acid sites.

Rogue_taxa_removal – data matrix #12

This data matrix was constructed by pruning 33 taxa that varied in their placement between analyses of the full data matrix by concatenation-based single model and coalescence using RogueNaRok[82]. A given taxon is considered a rogue taxon when its removal from the dataset results in an increase in the overall support values or in a better resolved consensus tree[82]. This removal resulted in a data matrix with 1,639 taxa and 124,700 amino acid sites.

Phylogenomic analyses

For the full data matrix as well as for each of these 11 data matrices constructed above, we used three different approaches to infer the fungal phylogeny: (1) the concatenation (i.e., supermatrix) approach with a single model or partition, (2) the concatenation approach with data-partitioning by gene, and (3) the multi-species coalescent-based approach that used the individual gene trees to construct the species phylogeny. All phylogenetic analyses were performed using IQ-TREE, version 1.6.8[71], which has previously been shown to consistently perform well in analyses of phylogenomic data in a maximum likelihood (ML) framework[83].

Concatenation-based approach without and with data-partitioning

For concatenation-based analyses using a single model, we used the LG+G4 model[84] because it was the best-fitting model for 89% of 290 gene trees. For analyses with data-partitioning by gene we used the best-fitting model for each gene (see coalescent-based approach section). Two independent runs were employed in all data matrices and the topological robustness of each gene tree was evaluated by 1,000 ultrafast bootstrap replicates[37]. A single tree search for the full data matrix (290 genes / 1,672 taxa) with a single model required ~4,620 CPU hours.

Coalescent-based approach

Individual gene trees were inferred using IQ-TREE, version 1.6.8 with an automatic detection for the best-fitting model with "-MFP" option using ModelFinder[85] under the Bayesian information criterion (BIC). For each gene tree, we conducted 5 independent tree searches to obtain the best-scoring ML tree with "-runs 5" option. The topological robustness of each gene tree was evaluated by 1000 ultrafast bootstrap replicates.

To account for gene tree heterogeneity by taking incomplete lineage sorting (ILS) into account, we used the individual ML gene trees to infer the coalescent-based species tree using ASTRAL-III version 5.1.1[70] for each data matrix. We applied contraction filters (BS < 33) such that poorly supported bipartitions within each gene tree were collapsed to polytomies, an approach recently suggested to improve the accuracy of ASTRAL[43]. The topological robustness was evaluated using the local posterior probability (LPP).

Quantification of incongruence

From the set of 12 data matrices (the full one and 11 subsampled ones) and 3 analyses (concatenation with single model, concatenation with data-partitioning, and coalescence), we expect a total of 36 phylogenies. Data matrices 2, 11, and 12 have different sets of taxa that have been removed, so they cannot be straightforwardly compared to the rest of the data matrices, which contain the full set of taxa. To reduce the burden of computation (each tree search required thousands of CPU hours), we did not perform concatenation-based data-partitioning analyses for data matrices 1, 11 and 12. Thus, a total of 33 phylogenetic trees were compared. Lastly, we rooted each concatenation and coalescence tree based on outgroups using the ape and phangorn R packages and visualized it using ITOL v4[74].

For the 33 species phylogenies inferred from the 12 data matrices (12 from concatenation-based single model analyses, 9 from concatenation-based data-partitioning analyses, and 12 from coalescent-based analyses), we quantified the degree of incongruence for every internode by considering all prevalent conflicting bipartitions among individual ML gene trees[81,86] using the "compare" function in Gotree version 1.13.6 (https://github.com/evolbioinfo/gotree).

It should be noted that all our trees suggested Agaricales is paraphyletic due to *Pleurotus eryngii* being placed within Russulales. In contrast to other three *Pleurotus* species, the *P. eryngii* genome contains a significantly higher amount of duplicated BUSCO genes (166 / 290 genes) (Data S1A). Moreover, we blasted several single-copy BUSCO genes from the *P. eryngii* genome to GenBank and found the top BLAST hits were from Russulales instead of Agaricales species. Thus, these results suggested that the paraphyletic of Agaricales might be a result of misidentification or contamination of the *P. eryngii* genome.

Polytomy test

To examine the support in individual gene trees for contentious bipartitions (and the alternative, conflicting bipartitions) and potentially identify evidence for hard polytomies of major fungal lineages, we used the polytomy test in ASTRAL, version 1.6.8[43]. The test evaluates whether a polytomy can be rejected by examining the frequencies of the three alternative quartet tree topologies in a set of trees. In our case, we used all gene trees as input for the calculation of the frequencies of the three alternative quartet trees for bipartitions of interest. In all cases, we used a P value cutoff of < 0.05 to reject the null hypothesis of a polytomy (see Figure 4 for eight tested hypotheses). We used scripts available at https://github.com/smirarab/1kp/tree/master/scripts/hypo-test. We used pos-for-hyp-4-11-2.sh (-t 4 option) and quart-for-hyp-4-11-2.sh (-t 8 option) to compute the posterior probabilities for all three alternative topologies of a given quartet. To evaluate the discordance of gene trees in our single-copy gene dataset, we used the *Q* value in ASTRAL to display the percentages of quartets in gene trees in support of the topology inferred by concatenation (q1) as well as the other two possible alternative topologies (q2 and q3); We used poly-for-hyp-4-11-02.sh to compute the p value for a hard polytomy under the null hypothesis using ASTRAL (-t 10 option).

Quantification of the distribution of phylogenetic signal

To investigate the distribution of phylogenetic signal of whether zygomycetes are monophyletic or paraphyletic, we considered two data matrices that had different topologies between ML analyses. To save computation time, we used the subset Dikarya data matrix (#2) since it has essentially the same topology as the full data matrix but has many fewer taxa. We also analyzed the Spatafora2016_46taxa_192 genes data matrix from a previous study that recovered the paraphyly of zygomycetes[15]. We examined two hypotheses: zygomycetes-monophyly (T1) and zygomycetes-paraphyly (T2: Zoopagomycota sister to Dikarya + Mucoromycota). For ML analysis in each data matrix, site-wise likelihood scores were inferred for both hypotheses using IQ-TREE, version 1.6.8 (option -g) with the LG+G4 model. The two different phylogenetic trees passed to IQ-TREE (via -z) were the tree where zygomycetes is monophyletic and a tree modified to have Zoopagomycota placed as the sister to Dikarya + Mucoromycota. The numbers of genes and sites supporting each hypothesis were calculated from IQ-TREE output and Perl scripts from a previous study[63]. By calculating gene-wise log-likelihood scores between T1 and T2 for every gene, we considered a gene with an absolute value of log-likelihood difference of two as a gene with strong ($|\Delta \ln L| > 2$) or weak ($|\Delta \ln L| < 2$) phylogenetic signal as done in a previous study[87].

RED index

To evaluate whether fungal taxonomy is consistent with evolutionary genomic divergence, we calculated relative evolutionary divergence (RED) values from the annotated tree inferred from the full data matrix using concatenation with a single model by PhyloRank (v0.0.37; https://github.com/dparks1134/PhyloRank/), as described previously[59]. Briefly, the NCBI taxonomy associated with every fungal genome was obtained from the NCBI Taxonomy FTP site on January 17, 2020. PhyloRank linearly interpolates the RED values

of every internal node according to lineage-specific rates of evolution under the constraints of the root being defined as zero and the RED of all present taxa being defined as one[59,60]. The RED intervals for each rank were defined as the median RED value ± 0.1 to serve as a guide for the normalization of taxonomic ranks from genus to phylum.

We also compared RED values to relative time divergence under a relaxed-molecular clock model for every taxonomic rank from genus to phylum, since both methods are based on inferring lineage-specific rates of evolution. We used the RelTime algorithm employed in the command line version of MEGA7[72] since it is computationally much less demanding than Bayesian tree-dating methods. We conducted divergence time estimation using the full data matrix with the same ML tree that we used for the RED analysis (see above) without fossil calibrations. Correlation between the RED values and relative divergence time estimated by RelTime was calculated using Pearson's correlation coefficient using the cor.test function in R package stats v.3.6.2[73].

QUANTIFICATION AND STATISTICAL ANALYSIS

Best-fitting phylogenetic models were selected according to the Bayesian Information Criterion implemented in IQ-Tree[71]. Branch supports were estimated using UFBoot2[37] bootstrapping in IQ-Tree. The topological robustness was evaluated using the local posterior probability (LPP) in ASTRAL[39]. For polytomy test, we used a P value cutoff of < 0.05 to reject the null hypothesis of a polytomy[43].