Phylogenomics unravels the early diversification of fungi

Jürgen F. H. Strassert[1,*] and Michael T. Monaghan[1,2]

[1]Department of Evolutionary and Integrative Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany

[2]Institut für Biologie, Freie Universität Berlin, Berlin, Germany

***Corresponding author:** E-mail: juergen.strassert@igb-berlin.de

Key words: fungal tree of life, evolution, phylogenomics

Abstract

Our understanding of the eukaryote tree of life is continually improving, although the branching events at some of the deepest nodes remain elusive. The fungi are an ancient group of eukaryotes with a wide range of morphologies, life-history strategies, and ecological roles. While several recent phylogenomic analyses have been shown to be a powerful tool for uncovering even earliest diversifications, no study has yet examined the entire tree of fungi using a taxonomically comprehensive data set and suitable models of evolution. Here, we assembled a data set of 299 proteins from all recognised fungal groups for which genomic/transcriptomic data were available and subjected this to a battery of analyses, including tree inferences using site-heterogeneous mixture models and the Multi-Species Coalescent model. Tree topology was highly congruent and well supported, and any incongruence was found to result from an inability of some frequently used evolutionary models to model fast-evolving and heterogeneous sites. Our results provide higher resolution among early-branching lineages than previous studies, and shed light on hitherto highly contested evolutionary origins of the major fungal groups.

Main

How organisms are related to each other is one of the oldest questions in natural sciences. In addition to morphological characteristics, molecular analyses of an ever-increasing number of phylogenetic markers have boosted our understanding of the evolutionary trajectories of all living forms, cumulating in a continuously improving tree of life[e.g. 1–5]. Within this tree, fungi represent one of the most ancient groups[6], having diverged from the animals about ~1.35 billion years ago[7]. Estimations of the number of extant species range between 1.5 and 3.8 million[8,9]. Recently, several reclassifications and the discovery of the deep-branching Sanchytriomycota lineage[10] have raised the number of proposed fungal phyla to 20[11]. Including the Microsporidia, although its status within the fungi is debated[12–14], would raise the number of

phyla to 21. The many proposed relationships among these phyla remain controversial. This is especially true for the unicellular, parasitic, and often fast-evolving early-branching lineages Aphelidiomycota and Rozellomycota, which together with the Microsporidia are known as Opisthosporidia[6,15–17]. Based on phylogenomic studies, the Aphelidiomycota are reported to branch separately as sister to free-living fungi[15], and also the monophyly of the three opisthosporidian phyla themselves is contested[e.g. 17,18].

The situation is comparable for the zoosporic (flagellated) subkingdom Chytridiomyceta, which live either as saprophytes, parasites, or as intermediate forms. Two recent phylogenomics studies proposed three different branching orders for the Chytridiomyceta depending on the method of tree inference: i) sister to Blastocladiomycota within an otherwise unresolved early-branching clade (coalescent-based approach)[18], ii) diverging after Blastocladiomycota but before all other non-Opisthosporidia fungi (supermatrix analysed by maximum likelihood)[18], or iii) sister to all other non-Opisthosporidia fungi (supermatrix analysed by Bayesian inference)[10]. Within the Chytridiomyceta, the branching order among the currently recognised phyla (Chytridiomycota, Neocallimastigomycota, Monoblepharomycota, and the proposed Caulochytriomycota[19]) is far from being resolved[6,12,20]. The unclear placement of Blastocladiomycota[12,18] raises an important evolutionary question, as members of this phylum produce zoospores, but with a distinct morphology from those produced by Chytridiomyceta. Another distinguishing feature is the sporic meiosis that results in a regular alternation of haploid and diploid generations in Blastocladiomycota[21].

Further uncertainties in the fungal tree of life concern the position of the Glomeromycota, either as sister to Mucoromycota + Mortierellomycota, or as sister to the Dikarya. The Dikarya includes the Ascomycota, Basidiomycota, and possibly the Entorrhizomycota, although the phylogenetic position is unknown for the latter and no genomic/transcriptomic data are available[6,10,12,18]. Finally, even within well-sampled and phylogenetically well-investigated lineages such as Basidiomycota, several internal

relationships remain unresolved, for example among Agaricomycotina/Wallemiomycotina, Pucciniomycotina, and Ustilaginomycotina[18,22].

Considering their great evolutionary diversity, it is not astonishing that fungi show very large differences in their morphologies, ecologies, and lifestyles. Globally distributed, they are the major decomposers of lignocellulosic plant litter in many ecosystems, and through the provision of nutrients in symbiotic interactions known as mycorrhiza, various lineages are essential for the survival of more than 80% of all plant species[23]. As parasites, fungi can influence their host population dynamics. Chytrid fungi are a major driver of amphibian declines[24] and species that infect phytoplankton are thought to cause a significant impact on the planet's carbon cycle by terminating algal blooms[25]. Uncovering how the highly diverse fungi are related to each other is fundamental for understanding how their lifestyles, morphological characteristics, and their metabolic capacities evolved. In this study, we aimed to resolve the deep phylogenetic relationships between major fungal groups. By first inferring single-protein phylogenies to evaluate orthology and potential contamination, we assembled a 299-proteins data set with a taxon sampling that is broad enough to encompass fungal diversity, and at the same time selective enough to run computationally intensive analyses using best-fitting models. A series of different phylogenetic reconstruction methods allowed us to provide what is currently the most robust backbone for the fungal tree of life.

Results

A well-curated and taxon-balanced phylogenomic data set. To untangle the early diversification in the fungal tree of life, we assembled a 299-proteins data set with broad taxon sampling. All proteins were carefully inspected by reconstructing single-protein phylogenies to evaluate orthology and detect contamination (see Methods). Considering that the taxonomic distribution of publicly available fungal genomes and transcriptomes is highly biased towards Ascomycota and Basidiomycota, we placed special emphasis on the inclusion of

underrepresented groups such as the Blastocladiomycota or Opisthosporidia. To uncover the deep phylogenetic relationships among fungal lineages, a taxon-reduced version of the full data set (661 taxa) was derived to allow a set of computationally intensive analyses. This taxon sampling (101 taxa) was aimed to keep a high diversity and coverage, and to select slow-evolving over fast-evolving lineages within phyla.

Reconstructing the phylogeny of fungi. Two trees were inferred from the full data set by (i) maximum likelihood (ML) under the site-homogeneous LG+G+F model, and (ii) under the Multi-Species Coalescent (MSC) model. These were largely congruent to each other (Supplementary Figs. 1 and 2); however, as expected for the heterogeneous taxon sampling in our full data set, some of the deeper nodes remained unsupported in these analyses. The taxon-reduced data set, from which all final analyses were derived and to which is referred to hereafter, was analysed under two site-heterogeneous mixture models: LG+C60+G+F with posterior mean site frequency profiles using ML (hereafter *ML-C60*; Fig. 1, Supplementary Fig. 3), and CAT+GTR+Γ4 using Bayesian inference (BI) (hereafter *CATGTR*; Fig. 1). Both models have been shown to account better for rate variation across sites[26,27] and their implementation led to the recovery of a similar overall topology with maximal statistical support for nearly all nodes. A further tree was inferred under the MSC model, which was highly congruent to the *ML-C60* tree (Fig. 1, Supplementary Figs. 3 and 4). This applies for example to the branching of Chytridiomyceta as sister to all other non-Opisthosporidia fungi; the sister-relationship of Chytridiomycota to a clade comprising Neocallimastigomycota + Monoblepharomycota; the branching of Blastocladiomycota + Sanchytriomycota, which diverged after the Chytridiomyceta but before other non-Opisthosporidia fungi; or the sister-relationship of Pucciniomycotina to Agaricomycotina/Wallemiomycotina + Ustilaginomycotina (Fig. 1). Aphelidiomycota was robustly recovered (full support) as a sister to the entire non-

Opisthosporidia clade and distinct from Rozellomycota and Microsporidia, although Aphelidiomycota was represented by only one species in our analysis (Fig. 1).

Ambiguity remained considering the diverging pattern of the Glomeromycota, either as sister to Mucoromycota + Mortierellomycota as reconstructed in the MSC and *ML-C60* trees or as sister to the Dikarya as reconstructed in the *CATGTR* tree (Fig. 1). Another discrepancy between the inferred tree topologies was the placement of the opisthosporid *Mitosporidium* as either sister to *Paramicrosporidium* (MSC and *ML-C60*; Supplementary Figs. 3 and 4) or sister to all other Microsporidia including *Paramicrosporidium* (*CATGTR*; Fig. 1).

Tree evaluation. The conflicts in the branching of the Glomeromycota and of *Mitosporidium* described above, shown by *ML-C60* and *CATGTR*, were fully supported by bootstrap and posterior probability analyses, respectively. Both alternative branching orders in the *CATGTR* topology were also rejected by ML in Approximately Unbiased tests (Glomeromycota + Dikarya, p-AU = 0.022; *Mitosporidium* + all other Microsporidia, p-AU = 0.011). To test whether the conflicts were due to the usage of different tree reconstruction methods (ML or BI), a further tree was derived using BI under the same model that was used for the ML tree. This tree (hereafter *BI-C60*; Supplementary Fig. 5) recapitulated the topology obtained with *ML-C60* indicating that differences were due to the evolutionary model rather than the inference method. To test which model better minimised inadequacy in describing compositional heterogeneity, posterior predictive analyses were conducted that resulted in a better fit of *CATGTR* compared to *BI-C60* (Table 1). However, both models failed to fully describe the data. Thus, 25% and 50% of the most compositionally heterogeneous sites were stripped from the alignment in order to reduce compositional heterogeneity, and additional trees were reconstructed from the pruned alignments with *CATGTR* (Fig. 2). The inferred trees were in agreement with the *CATGTR* tree reconstructed from the full-length alignment (Figs. 1, 2).

In order to understand the extent to which fast-evolving sites, which are more likely to be homoplasic, led to systematic bias in the *ML-C60* tree inference due to model misspecification[28], the fastest-evolving sites were progressively removed from the alignment in increments of 10,000 sites. Each sub-alignment was then subjected to further ML tree inferences using the LG+C60+G+F model and node support was inferred by ultrafast bootstrap approximation (please note that due to the higher computational demand of BI, this analysis was run for ML inferences only). Interestingly, whereas the bootstrap support for Dikarya remained maximal throughout all the analyses, the support for the sister-relationship of Glomeromycota to Mucoromycota + Mortierellomycota decreased rapidly while the support for their closer relationship to the Dikarya, as seen in the *CATGTR* trees reconstructed from complete and pruned alignments, increased, reaching 92% after the removal of 30,000 sites (Fig. 3). Further removal of sites led either to an unsupported branching or recovered the Glomeromycota + Dikarya topology again (100,000 sites removed; Fig. 3). For the fast-evolving microsporidians, the sister-relationship of *Mitosporidium* to all other Microsporidia, as inferred by *CATGTR*, was only supported when the 120,000 fastest-evolving sites were removed from the alignment (i.e. 8,450 sites remained; Fig. 3).

Discussion

The fungal kingdom shows a great diversity in morphology and lifestyle. Despite their worldwide distribution, their ecological importance, and the significant roles that fungi have played in human history (e.g. as a food source, leavening agent, medicine, or to ferment beer and other alcoholic beverages), our knowledge of how this diversity evolved is still limited. Phylogenomic studies encompassing the entire fungal kingdom with nearly all its different phyla are scarce — either due to missing data or to a focus on only a few certain groups of fungi[10,17,18,22,29]. In this study, we tested whether a more balanced taxon-sampling, in combination with new filtering methods[30,31] that improve the phylogenetic signal of each

single-protein alignment and best-fitting models of evolution, could untangle some of the hitherto debated early diversifications in the fungal tree of life.

The branching order among fungal phyla remained consistent and well supported across multiple analyses using site-heterogeneous models in BI (*CATGTR* and *BI-C60*) and ML (*ML-C60*) tree reconstructions, and using an MSC approach (Figs. 1, 2, Supplementary Figs. 3–5). The sole exception concerned the placement of the Glomeromycota, which was sister to the Dikarya in *CATGTR*-based analyses or was sister to the Mucoromycota/Mortierellomycota clade in the other analyses. Glomeromycota were suggested to form an independent phylum at the beginning of this century[32], but closer phylogenetic affiliation has remained controversial ever since[12,18,20,33]. A sister-relationship to the Dikarya has been proposed based on tree inferences using the elongation factor 1-a, RNA polymerase II subunits and/or rRNA gene alignments[6,32–35]. This placement was discussed to be artificial due to lineage-specific rates in these data sets[12] and was rejected in a ML tree inference from a 192-proteins supermatrix. However, we note that this supermatrix analysis used an LG model with fixed base frequencies[36], and that our results support a more recent *CATGTR*-based analysis of a 264-proteins supermatrix that recovered Glomeromycota as sister to the Dikarya[10]. It is reasonable to assume that this finding depicts the real scenario in the evolution of fungi and that the discovered sister-relationship to the Mucoromycota/Mortierellomycota clade in the *ML-C60* tree is a consequence of the higher inadequacy of the *C60* model in describing compositional heterogeneity. This is in line with the results obtained by the posterior predictive test, in which the fit of both models, *CATGTR* and *BI-C60*, was directly compared using the PhyloBayes software package (Table 1). Congruent topologies reconstructed from the analyses of the same alignment using both *BI-C60* and *ML-C60* (Supplementary Figs. 3, 5) give further evidence that the found discrepancies in the branching order of the Glomeromycota are rather caused by the less fitting evolutionary *C60* model than by the use of BI or ML. Interestingly, however, Glomeromycota + Dikarya was also supported by *ML-C60* tree inferences after only 15% of the

fastest-evolving sites were removed (Fig. 3), indicating that the conflicting placement of Glomeromycota in the *C60* analysis of the full alignment may be due to statistical problems this model encounters with homoplasic sites.

The Aphelidiomycota representative was placed as sister to all non-Opisthosporidia fungi in all analyses (Figs. 1, 2, Supplementary Figs. 3–5). Similar findings have been reported before[10,15] corroborating the paraphyletic nature of Opisthosporidia. The assignment of some of the earliest-branching fungal species to either the Rozellomycota or the Microsporidia, on the other hand, is more controversial, leading to debates as to whether these phyla are monophyletic or not[6,11,17]. Whereas *Paramicrosporidium saccamoebae* was initially assigned to the Rozellomycota based on a phylogenetic analysis of the SSU rRNA gene sequence, its (unsupported) affiliation to the Microsporidia was recovered when 5.8S and partial LSU rRNA gene sequences were added to such analyses[37]. In the present study, the exact placement of *P. saccamoebae* remains questionable. While trees inferred with the better fitting *CATGTR* model recovered this species to be nested between the microsporidians *Mitosporidium daphnia* and *Amphiamblys* sp. (Figs. 1, 2), a sister-relationship to *M. daphnia* was recovered in all analyses using the MSC approach or the *C60* model in BI and ML inferences (Supplementary Figs. 3–5). For these fast-evolving lineages, the latter topology was only challenged when at least 85% of the fastest-evolving sites were removed (Fig. 3). However, independent from the exact position of *P. saccamoebae*, the only safely assigned Rozellomycota species with available genomic data, *Rozella allomycis*, formed a sister to all Microsporidia and *P. saccamoebae* throughout all analyses (Figs. 1, 2, Supplementary Figs. 3–5). Taking into account these results, the results obtained by rRNA analyses[37] (see above) and the morphology of *P. saccamoebae*, which resembles that of microsporidians (they lack a mitochondrion and have tiny spores with a polar filament[37]), its assignment to the Microsporida seems to be reasonable and is proposed here. Thus, both Microsporidia and Rozellomycota form two distinct lineages for which a frequently discussed paraphyly[6,17,18] cannot be confirmed here. Whether or not Microsporidia +

Rozellomycota should be classified as only one phylum is currently debated[6,33,38] and will require data from additional representatives of Rozellomycota.

The erection of the phylum Caulochytriomycota was proposed by Doweld in 2014[19] and has been adapted by other researchers since. Yet, the justification for the separation of this phylum from the Chytridiomycota remains enigmatic. Only two species of the genus *Caulochytrium* are known, and DNA sequences are available only for *Caulochytrium protostelioides*. In our analysis, *C. protostelioides* was either closely related to, or even nested within, the Chytridiomycota, and the clade was well-supported (Figs. 1, 2, Supplementary Figs. 3–5). An assignment of *Caulochytrium* to the Chytridiomycota is therefore suggested.

The here presented tree (Fig. 1a) provides a robust backbone topology for the diversification of major fungal groups. For the recovery of ancient nodes, not only site rate variation but also variation of the sequence composition across branches and/or alignment sites can hamper the reconstruction of topology and branch length. Thus, 25% and 50% of the sites that contribute the most to branch heterogeneity were removed from the alignment and new *CATGTR* trees were inferred (Fig. 2). Both trees reflected the overall topology of the tree inferred from the complete alignment (Fig. 1a) indicating that compositional biases had either no or a negligible negative impact that could result in model misspecification despite using the best-fitting CAT+GTR+Γ4 model. However, it is noteworthy that the backbone topology provided here — like all phylogenetic trees — represents a hypothesis and it is safe to assume that with an increasing diversity of fungi for which genomic/transcriptomic data will be available, several clades will be reshuffled. It should also be noted that the taxonomic classification in fungi is changing rapidly. The lineage information in this study was assigned according to the NCBI taxonomy (ncbi.nlm.nih.gov/taxonomy) with some modifications especially regarding the phylum assignments, which, if not otherwise noted, were adopted in order to match those recently proposed by Wijayawardene *et al.*[11] A comprehensive revision of the taxonomic classification in fungi, based on this and other recent studies [e.g. 10,18], is desirable.

Methods

Analyses were carried out using the Zedat high-performance computing infrastructure at Freie Universität Berlin[39].

Phylogenomic data set construction. Genome and transcriptomes of a broad range of fungi and selected non-fungal lineages were downloaded from EukProt[40], fungi.ensembl.org, metazoa.ensembl.org, protists.ensembl.org, imicrobe. us/#/projects/104, mycocosm.jgi.doe.gov, and ncbi.nlm.nih.gov. For unassembled sequence data, paired end reads were trimmed using Trimmomatic v. 0.39[41] with the following flags employed: LEADING: 3, TRAILING: 3, SLIDINGWINDOW: 4:15, and MINLEN: 36. Transcriptomic data was assembled and translated into amino acids by using the default settings in Trinity v. 2.10.0 and TransDecoder v. 5.5.0, respectively[42]. Reads of genomic data were merged with PEAR v. 0.9.11[43], quality filtered with Sickle v. 1.33[44] (only the paired but unmerged reads), and assembled with SPAdes v. 3.14.1[45] using the --isolate option. Genes were predicted using BUSCO v. 5.1.2 with the database fungi_odb10[46].

A recently published data set comprising 320 proteins for diverse eukaryotic lineages (733) was used as starting point for data set construction[7]. With a few exceptions that were kept as outgroup, non-fungal linages were removed. The data set was then expanded by adding a broad range of fungal lineages as follows: 1) Protein sequences for each taxon were clustered with CD-HIT v. 4.8.1[47] using an identity threshold of 85%. 2) Candidates for homologous copies were retrieved by BLASTP searches[48] using the 320 proteins as queries (e-value: 1e-20; coverage cutoff: 0.5). 3) In three rounds, phylogenetic trees were inferred and carefully inspected in order to detect and remove paralogs and contaminants and select orthologs. In the first round, sequences were aligned with MAFFT v. 7.475[49] using the --auto option, filtered with trimAL v. 1.4[50] using a gap threshold of 0.8, and ML trees were inferred with FastTreeMP

v. 2.1.11[51] using -lg -gamma and options for more accurate performances. In the second and third round, non-homologues sequence stretches were filtered out prior to aligning using Prequal v. 1.02[30] with a threshold of 0.95. Protein sets were then aligned using MAFFT G-INS-i with the following options: --allowshift, --unalignlevel 0.6, --maxiterate 0. Further non-homologues residues were removed with Divvier v. 1.01[31] employing the -partial flag and alignments were trimmed with trimAL using a soft gap threshold of 0.05. Finally, partial sequences belonging to the same taxon that did not show evidence for paralogy or contamination were merged. Single-protein ML phylogenies (Supplementary Data 1) were reconstructed with IQ-Tree v. 1.6.12[52] under BIC-selected models including site-homogeneous models (such as LG) and empirical profile models (C10–C60) and node support was inferred by ultrafast bootstrap approximation[53] (1,000 replicates).

Of the initial 320 proteins, 21 were removed due to a high proportion of missing taxa or incongruous clustering of several clades. The final data set comprised 299 proteins and 637 fungal taxa plus 24 non-fungal taxa (Supplementary Data 1). The protein sequences were concatenated using SCaFoS v. 1.25[54]. The obtained supermatix (129,814 amino acid positions; Supplementary Data 1) was then used to calculate an initial ML tree with IQ-Tree employing the site-homogeneous model LG+G+F and ultrafast bootstrap approximation (1,000 replicates; Supplementary Fig. 1). To allow computationally more demanding analyses, a taxon-reduced data set was built representing all major fungal groups and discarding taxa with more missing data. Corresponding sequences were newly aligned, filtered, and concatenated (same settings as described above; 101 taxa, 128,450 amino acid positions; Supplementary Data 1). New ML trees were inferred for each of the taxon-reduced protein alignments as described above (Supplementary Data 1).

Phylogenomic analyses. The taxon-reduced supermatrix was used to reconstruct a ML tree with IQ-Tree using the best-fitting site-heterogeneous model LG+C60+G+F with the PMSF

approach[27] to obtain nonparametric bootstrap support (100 replicates; Fig. 1, Supplementary Fig. 3). Copies of this tree were manually edited to test alternative topologies with the approximately unbiased test (AU test[55]). The taxon-reduced supermatrix was also subjected to Bayesian analyses using PhyloBayes-MPI v. 1.8[56] with the site-heterogeneous CAT+GTR+Γ4 model and the -dc option to remove constant sites. Three independent Markov Chain Monte Carlo (MCMC) chains were run for 1,600 generations (all sampled). For each chain, the burnin period was estimated by monitoring evolution of the log-likelihood (Lnl) at each sampled point, and 750 generations were removed from all chains. A consensus tree (Fig. 1) was built and global convergence between chains was assessed with PhyloBayes' built-in tools. As frequently recognised in PhyloBayes, global convergence was not received in all combinations of the chains with maxdiff = 1 and meandiff = 0.0167504. To discriminate whether potential discrepancies between this tree and the tree calculated by ML are due to the usage of different software packages or due to different evolutionary models, a further Bayesian tree (Supplementary Fig. 5) was inferred under the LG+C60 model (-dc -catfix C60 -lg -dgam 4; three chains, each with 2,000 generations; convergence was not received; maxdiff = 1, meandiff = 0.0268007). Posterior predictive analyses were then run for both tree inferences in order to informatively select the best topology.

The single-protein ML trees of both the full and the taxon-reduced data sets (Supplementary Data 1) were also analysed under the Multi-Species Coalescent (MSC) model in ASTRAL-III v. 5.7.7[57] to incorporate protein tree uncertainties (Supplementary Figs. 2 and 4). Prior to that, branches with bootstrap support <10 were collapsed as recommended by the developers. Quadripartition supports were calculated from quartet frequencies among the set of input trees[58].

Sites removal analyses. 25% and 50% of the most heterogeneous sites were removed from the taxon-reduced supermatrix using the script alignment_pruner.pl

(https://github.com/novigit/davinciCode/blob/master/perl). Bayesian trees (Fig. 2) were inferred form the pruned alignments with the CAT+GTR+Γ4 model: 25% pruned: three chains, each with 1,600 generations (burnin 1,000); convergence was not received; maxdiff = 1, meandiff = 0.0187389. 50% pruned: three chains, each with 2,450 generations (burnin 1,250); convergence was not received; maxdiff = 1, meandiff = 0.0128211. Also for the taxon-reduced supermatrix, fast-evolving sites were estimated with IQ-Tree under the LG+C60+G+F model, employing the -wsr flag and giving the PMSF tree as fixed tree topology. Sites were then removed in 10,000 increments and ML trees were inferred from the site-reduced alignments using IQ-Tree (LG+C60+G+F -bb 1,000 -bnni -wbtl; Fig. 3).

Data availability

All data needed to evaluate the conclusions of this study are presented in the paper, the Supplementary Information, and the Supplementary Data. Raw sequence data are available under the following web-links: doi.org/10.6084/m9.figshare.12417881.v2, fungi.ensembl.org, metazoa.ensembl.org, protists.ensembl.org, imicrobe.us/#/projects/104, mycocosm.jgi.doe.gov, and ncbi.nlm.nih.gov

Code availability

The script alignment_pruner.pl is available at github.com/novigit/davinciCode/blob/master/perl

References

1. Brown, M. W. *et al.* Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biol Evol* **10**, 427–433 (2018).
2. Strassert, J. F. H., Jamy, M., Mylnikov, A. P., Tikhonenkov, D. V. & Burki, F. New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. *Mol Biol Evol* **36**, 757–765 (2019).

3. Irisarri, I., Strassert, J. F. H. & Burki, F. Phylogenomic insights into the origin of primary plastids. *bioRxiv* 2020.08.03.231043 (2020) doi:10.1101/2020.08.03.231043.

4. Gawryluk, R. M. R. *et al.* Non-photosynthetic predators are sister to red algae. *Nature* **572**, 240–243 (2019).

5. Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The new tree of eukaryotes. *Trends Ecol Evol* **35**, 43–55 (2020).

6. Voigt, K. *et al.* Early-diverging fungal phyla: taxonomy, species concept, ecology, distribution, anthropogenic impact, and novel phylogenetic proposals. *Fungal Divers* **109**, 59–98 (2021).

7. Strassert, J. F. H., Irisarri, I., Williams, T. A. & Burki, F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat Commun* **12**, 1879 (2021).

8. Hawksworth, D. L. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol Res* **105**, 1422–1432 (2001).

9. Hawksworth, D. L., Robert, L., Joseph, H. & James, T. Y. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol Spectr* **5**, 5.4.10 (2017).

10. Galindo, L. J., López-García, P., Torruella, G., Karpov, S. & Moreira, D. Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. *Nat Commun* **12**, 4973 (2021).

11. Wijayawardene, N. N. Outline of fungi and fungus-like taxa. *Mycosphere* **11**, 1060–1456 (2020).

12. James, T. Y., Stajich, J. E., Hittinger, C. T. & Rokas, A. Toward a fully resolved fungal tree of life. *Annu Rev Microbiol* **74**, 291–313 (2020).

13. James, T. Y. *et al.* Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Curr Biol* **23**, 1548–1553 (2013).

14. Keeling, P. J., Luker, M. A. & Palmer, J. D. Evidence from beta-tubulin phylogeny that

microsporidia evolved from within the fungi. *Mol Biol Evol* **17**, 23–31 (2000).

15. Torruella, G. *et al.* Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. *Commun Biol* **1**, 231 (2018).
16. Bass, D. *et al.* Clarifying the relationships between Microsporidia and Cryptomycota. *J Eukaryot Microbiol* **65**, 773–782 (2018).
17. Quandt, C. A. *et al.* The genome of an intranuclear parasite, *Paramicrosporidium saccamoebae,* reveals alternative adaptations to obligate intracellular parasitism. *Elife* **6**, e29594 (2017).
18. Li, Y. *et al.* A genome-scale phylogeny of the kingdom Fungi. *Curr Biol* **31**, 1653–1665 (2021).
19. Doweld, A. B. Nomenclatural novelties. *Index Fungorum* **49**, (2014).
20. Spatafora, J. W. *et al.* The fungal tree of life: from molecular systematics to genome-scale phylogenies. in *Microbiology Spectrum* (eds. Heitmann, J. et al.) 3–34 (2017). doi:10.1128/microbiolspec.FUNK-0053-2016.
21. Powell, M. J. Blastocladiomycota. in *Handbook of the Protists* (eds. Archibald, J. M., Simpson, A. G. B. & Slamovits, C. H.) 1497–1521 (Springer International Publishing, 2017). doi:10.1007/978-3-319-28149-0_17.
22. Choi, J. & Kim, S.-H. A genome tree of life for the Fungi kingdom. *Proc Natl Acad Sci U S A* **114**, 9391–9396 (2017).
23. Wang, B. & Qiu, Y.-L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**, 299–363 (2006).
24. Fisher, M. C. & Garner, T. W. J. Chytrid fungi and global amphibian declines. *Nat Rev Microbiol* **18**, 332–343 (2020).
25. Frenken, T. *et al.* Integrating chytrid fungal parasites into plankton ecology: research gaps and needs. *Environ Microbiol* **19**, 3802–3822 (2017).
26. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in

the amino-acid replacement process. *Mol Biol Evol* **21**, 1095–1109 (2004).

27. Wang, H. C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol* **67**, 216–235 (2018).

28. Rodríguez-Ezpeleta, N. *et al.* Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol* **56**, 389–399 (2007).

29. Ahrendt, S. R. *et al.* Leveraging single-cell genomics to expand the fungal tree of life. *Nat Microbiol* **3**, 1417–1428 (2018).

30. Whelan, S., Irisarri, I. & Burki, F. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**, 3929–3930 (2018).

31. Ali, R. H., Bogusz, M. & Whelan, S. Identifying clusters of high confidence homologies in multiple sequence alignments. *Mol Biol Evol* **36**, 2340–2351 (2019).

32. Schüβler, A., Schwarzott, D. & Walker, C. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol Res* **105**, 1413–1421 (2001).

33. Tedersoo, L. *et al.* High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Divers* **90**, 135–159 (2018).

34. James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818–822 (2006).

35. Strassert, J. F. H. *et al.* Long rDNA amplicon sequencing of insect-infecting nephridiophagids reveals their affiliation to the Chytridiomycota and a potential to switch between hosts. *Sci Rep* **11**, 396 (2021).

36. Spatafora, J. W. *et al.* A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028–1046 (2016).

37. Corsaro, D. *et al.* Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. *Parasitol Res* **113**, 1909–1918 (2014).

38. Wijayawardene, N. N. *et al.* Notes for genera: basal clades of Fungi (including

Aphelidiomycota, Basidiobolomycota, Blastocladiomycota, Calcarisporiellomycota, Caulochytriomycota, Chytridiomycota, Entomophthoromycota, Glomeromycota, Kickxellomycota, Monoblepharomycota, Mortierellomyc. *Fungal Divers* **92**, 43–129 (2018).

39. Bennett, L., Melchers, B. & Proppe, B. Curta: a general-purpose high-performance computer at Zedat, Freie Universität Berlin. *http://dx.doi.org/10.17169/refubium-26754* (2020).

40. Richter, D. J., Berney, C., Strassert, J. F. H., Burki, F. & de Vargas, C. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotic life. *bioRxiv* 2020.06.30.180687 (2020) doi:10.1101/2020.06.30.180687.

41. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

42. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).

43. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).

44. Joshi, N. A. & Fass, J. N. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. *https://github.com/najoshi/sickle* (2011).

45. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477 (2012).

46. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

47. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local

alignment search tool. *J Mol Biol* **215**, 403–410 (1990).

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

50. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

51. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641–1650 (2009).

52. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).

53. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* **35**, 518–522 (2018).

54. Roure, B., Rodriguez-Ezpeleta, N. & Philippe, H. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol Biol* **7**, (2007).

55. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol* **51**, 492–508 (2002).

56. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. Phylobayes mpi: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol* **62**, 611–615 (2013).

57. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).

58. Sayyari, E. & Mirarab, S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol* **33**, 1654–1668 (2016).

Acknowledgments

JFHS acknowledges support from the German Research Foundation (DFG; Grant STR1349/2-1 Project No. 432453260) and research was partially funded by the German Federal Ministry of Education and Research (BMBF, Förderkennzeichen 033W034A). The authors thank the High-Performance Computing Service of ZEDAT, Freie Universität Berlin, for computing time. Finally, we thank David Moreira and Luis Javier Galindo for sharing sequence data for *Amoeboradix gromovi* and *Sanchytrium tribonematis*.

Author contributions

JFHS conceived the study, assembled and curated the data, performed all analyses, and drafted the manuscript. MTM contributed to the final manuscript. Both authors read and approved the final version.

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. Phylogenetic trees showing the early diversification of fungi. The conflicting branching of the Glomeromycota in the CAT+GTR+Γ4 analysis and the LG+C60+G+F-PMSF and summary-coalescent analyses is highlighted. **Caulochytrium* and the subphyla of the Zoopagomycota were recently proposed to build independent phyla[11] (see Discussion). (a) Bayesian consensus tree inferred from a concatenated alignment of 299 proteins under the CAT+GTR+Γ4 model. Node support is given by posterior probabilities (full support is indicated by black circles). (b) Two phylogenetic trees showing perfect congruence. Left: Tree inferred by ML under the LG+C60+G+F-PMSF model using the same alignment as used for the Bayesian tree. Nonparametric bootstrap support is not indicated but was maximal for all

nodes shown (for details, see Supplementary Fig. 3). Right: Tree based on summary-coalescent analyses of 299 single-protein ML phylogenies. Quadripartition support values are given for nodes that are not fully supported (for details, see Supplementary Fig. 4).

Figure 2. Simplified Bayesian consensus tree inferred under the CAT+GTR+Γ4 model from a concatenated alignment of 299 proteins and 101 taxa with 50% of the most heterogeneous sites removed. All nodes shown were fully supported. A further tree, which was inferred from the same alignment but with 25% of the most heterogeneous sites removed, recapitulated an identical topology (not shown). Microsporidians are only partly collapsed in order to depict the position of *Paramicrosporidium saccamoebae*. *Caulochytrium protostelioides* was nested within the Chytridiomycota (same branching pattern as in Fig. 1a; see Discussion).

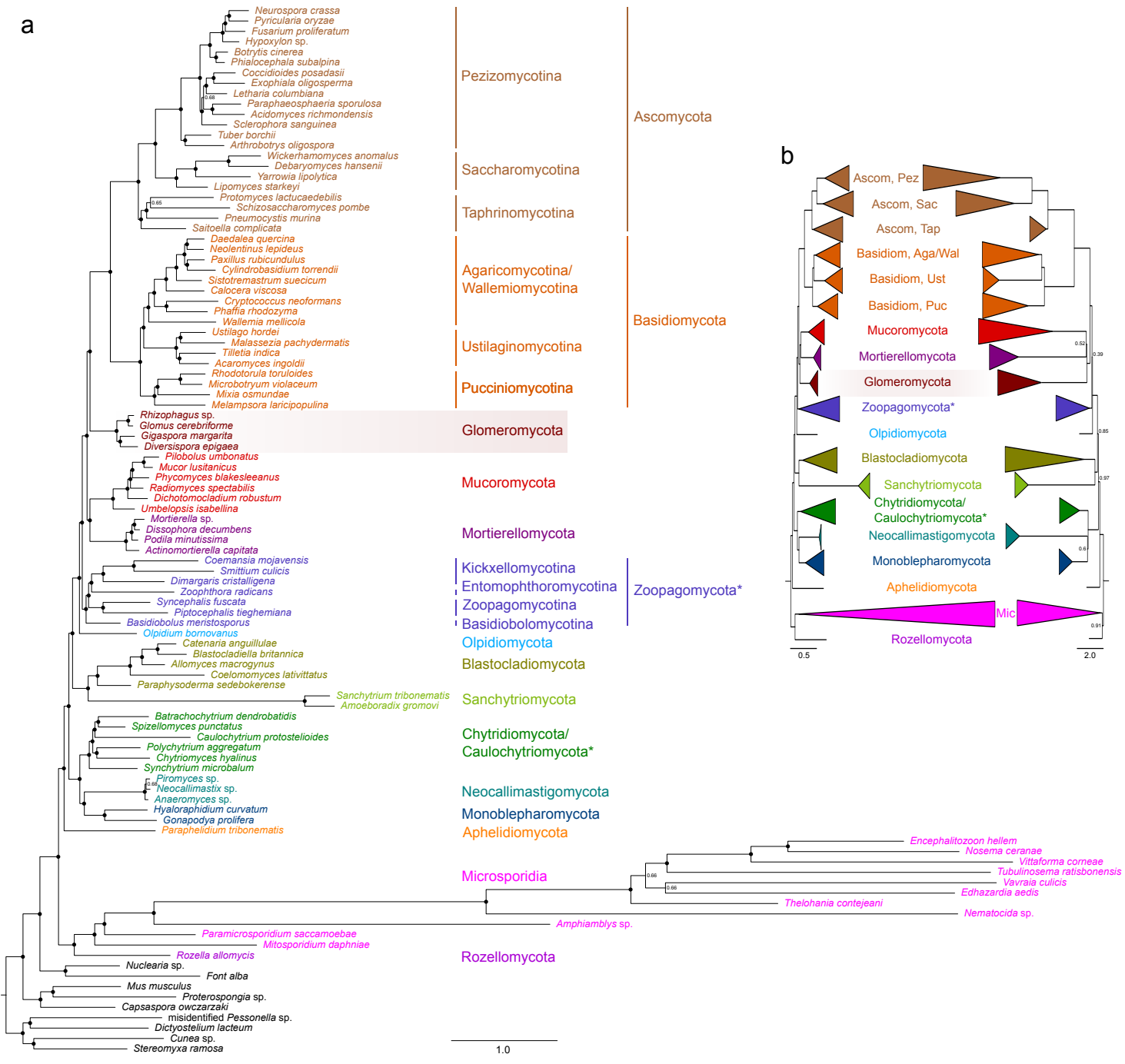
Figure 3. Fast-evolving sites removal analyses. ML ultrafast bootstrap approximation for selected nodes using the full supermatrix and subsets from which fast-evolving sites were removed in 10,000 increments (see Methods).

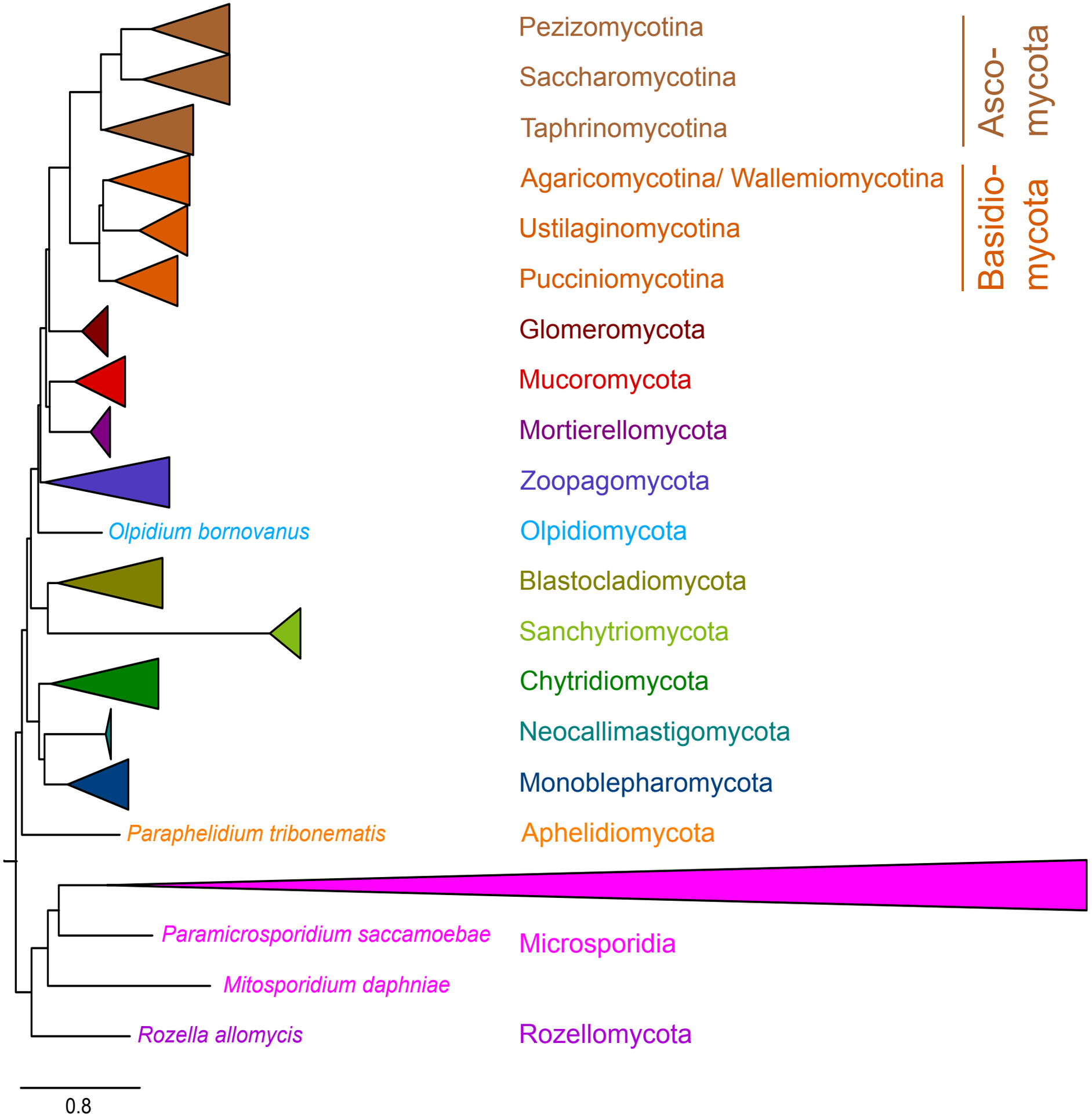
Tables

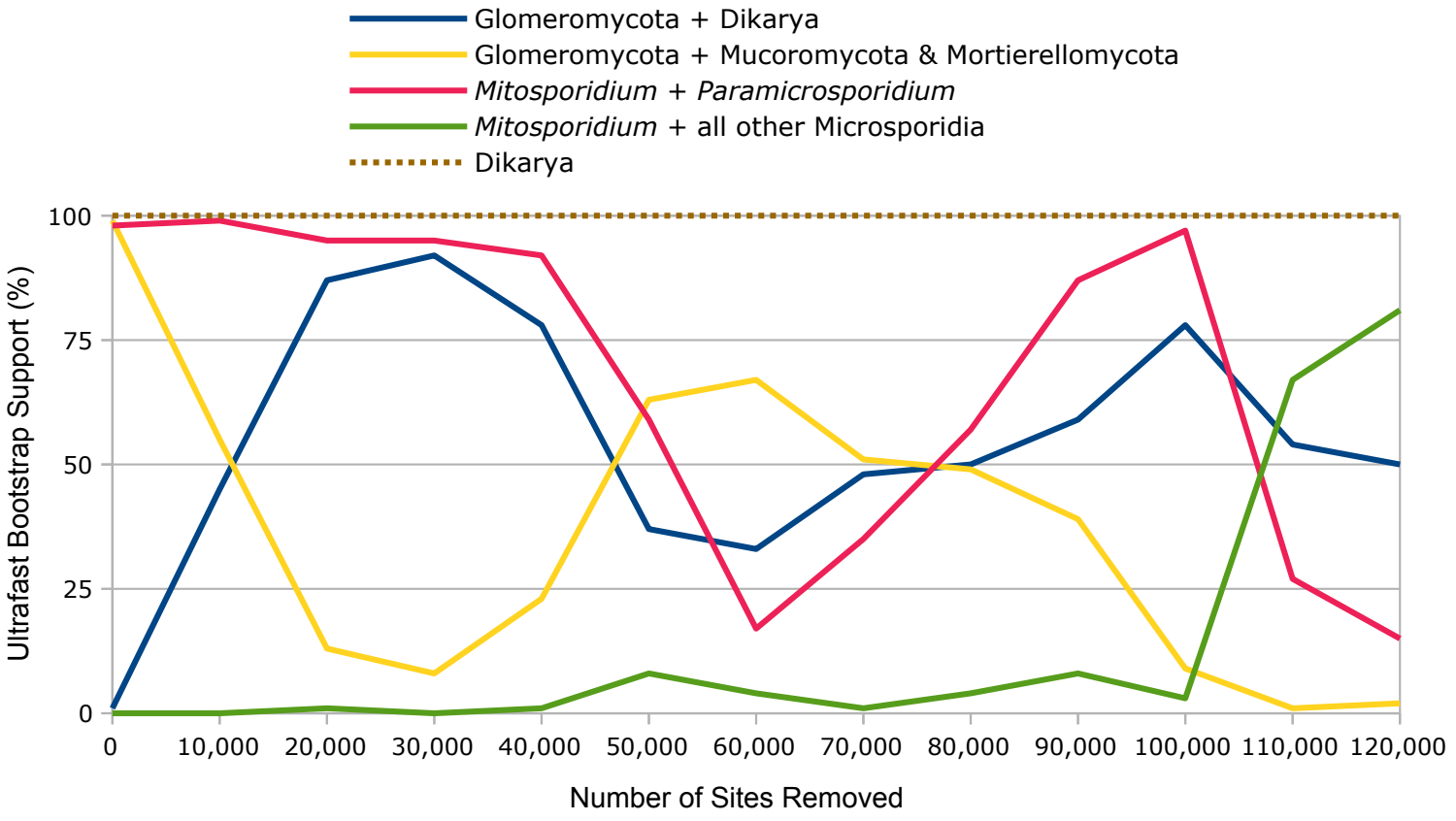
Table 1. Results of the posterior predictive analyses (PPA).

	Site-specific AA heterogeneity*			Lineage-specific AA heterogeneity*	
Model	**div**	**conv**	**var**	**max**	**mean**
CAT+GTR+G	7.93110	13.16270	13.58923	179.34	340.850
LG+C60+G	75.71623	46.49947	44.97260	270.14	472.514

*z-scores


a
Neurospora crassa
Pyricularia oryzae
Fusarium proliferatum
Hypoxylon sp.
Botrytis cinerea
Phialocephala subalpina
Coccidioides posadasii
Exophiala oligosperma
Letharia columbiana
Paraphaeosphaeria sporulosa
Acidomyces richmondensis
Sclerophora sanguinea
Tuber borchii
Arthrobotrys oligospora
Wickerhamomyces anomalus
Debaryomyces hansenii
Yarrowia lipolytica
Lipomyces starkeyi
Protomyces lactucaedebilis
Schizosaccharomyces pombe
Pneumocystis murina
Saitoella complicata
Daedalea quercina
Neolentinus lepideus
Paxillus rubicundulus
Cylindrobasidium torrendii
Sistotremastrum suecicum
Calocera viscosa
Cryptococcus neoformans
Phaffia rhodozyma
Wallemia mellicola
Ustilago hordei
Malassezia pachydermatis
Tilletia indica
Acaromyces ingoldii
Rhodotorula toruloides
Microbotryum violaceum
Mixia osmundae
Melampsora laricipopulina
Rhizophagus sp.
Glomus cerebriforme
Gigaspora margarita
Diversispora epigaea
Pilobolus umbonatus
Mucor lusitanicus
Phycomyces blakesleeanus
Radiomyces spectabilis
Dichotomocladium robustum
Umbelopsis isabellina
Mortierella sp.
Dissophora decumbens
Podila minutissima
Actinomortierella capitata
Coemansia mojavensis
Smittium culicis
Dimargaris cristalligena
Zoophthora radicans
Syncephalis fuscata
Piptocephalis tieghemiana
Basidiobolus meristosporus
Olpidium bornovanus
Catenaria anguillulae
Blastocladiella britannica
Allomyces macrogynus
Coelomomyces lativittatus
Paraphysoderma sedebokerense
Sanchytrium tribonematis
Amoeboradix gromovi
Batrachochytrium dendrobatidis
Spizellomyces punctatus
Caulochytrium protostelioides
Polychytrium aggregatum
Chytriomyces hyalinus
Synchytrium microbalum
Piromyces sp.
Neocallimastix sp.
Anaeromyces sp.
Hyaloraphidium curvatum
Gonapodya prolifera
Paraphelidium tribonematis
Encephalitozoon hellem
Nosema ceranae
Vittaforma corneae
Tubulinosema ratisbonensis
Vavraia culicis
Edhazardia aedis
Thelohania contejeani
Nematocida sp.
Amphiamblys sp.
Paramicrosporidium saccamoebae
Mitosporidium daphniae
Rozella allomycis
Nuclearia sp.
Font alba
Mus musculus
Proterospongia sp.
Capsaspora owczarzaki
misidentified Pessonella sp.
Dictyostelium lacteum
Cunea sp.
Stereomyxa ramosa
Pezizomycotina
Saccharomycotina
Taphrinomycotina
Ascomycota
Agaricomycotina/
Wallemiomycotina
Ustilaginomycotina
Pucciniomycotina
Basidiomycota
Glomeromycota
Mucoromycota
Mortierellomycota
Kickxellomycotina
Entomophthoromycotina
Zoopagomycotina
Basidiobolomycotina
Zoopagomycota*
Olpidiomycota
Blastocladiomycota
Sanchytriomycota
Chytridiomycota/
Caulochytriomycota*
Neocallimastigomycota
Monoblepharomycota
Aphelidiomycota
Microsporidia
Rozellomycota
1.0
b
Ascom, Pez
Ascom, Sac
Ascom, Tap
Basidiom, Aga/Wal
Basidiom, Ust
Basidiom, Puc
Mucoromycota
Mortierellomycota
Glomeromycota
Zoopagomycota*
Olpidiomycota
Blastocladiomycota
Sanchytriomycota
Chytridiomycota/
Caulochytriomycota*
Neocallimastigomycota
Monoblepharomycota
Aphelidiomycota
Mic
Rozellomycota
0.5
2.0


Pezizomycotina
Saccharomycotina
Taphrinomycotina
Asco-mycota
Agaricomycotina/ Wallemiomycotina
Ustilaginomycotina
Pucciniomycotina
Basidio-mycota
Glomeromycota
Mucoromycota
Mortierellomycota
Zoopagomycota
Olpidium bornovanus
Olpidiomycota
Blastocladiomycota
Sanchytriomycota
Chytridiomycota
Neocallimastigomycota
Monoblepharomycota
Paraphelidium tribonematis
Aphelidiomycota
Paramicrosporidium saccamoebae
Microsporidia
Mitosporidium daphniae
Rozella allomycis
Rozellomycota
0.8


Glomeromycota + Dikarya
Glomeromycota + Mucoromycota & Mortierellomycota
Mitosporidium + Paramicrosporidium
Mitosporidium + all other Microsporidia
Dikarya
Ultrafast Bootstrap Support (%)
100
75
50
25
0
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
90,000
100,000
110,000
120,000
Number of Sites Removed