Phylogenomics unravels the early diversification of fungi

Jürgen F. H. Strassert[1,2,*] and Michael T. Monaghan[1,3]

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

[2]Lead Contact

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

*Correspondence: juergen.strassert@igb-berlin.de (J.F.H.S.)

Summary

Phylogenomic analyses have boosted our understanding of the evolutionary trajectories of all living forms by providing continuous improvements to the tree of life[1–5]. Within this tree, fungi represent an ancient eukaryote group[6], having diverged from the animals ~1.35 billion years ago[7]. Estimates of the number of extant species range between 1.5 and 3.8 million[8,9]. Recent reclassifications and the discovery of the deep-branching Sanchytriomycota lineage[10] have brought the number of proposed phyla to 20[11]; 21 if the Microsporidia are included[12–14]. Uncovering how the diverse and globally distributed fungi are related to each other is fundamental for understanding how their lifestyles, morphologies, and metabolic capacities evolved. To date, many of the proposed relationships among the phyla remain controversial and no phylogenomic study has examined the entire fungal tree using a taxonomically comprehensive data set and suitable models of evolution. We assembled and curated a 299-protein data set with a taxon sampling broad enough to encompass all recognised fungal diversity with available data, but selective enough to run computationally intensive analyses using best-fitting models. Using a range of reconstruction methods, we were able to resolve many contested nodes, such as a sister-relationship of Chytridiomyceta to all other non-Opisthosporidia fungi (with Chytridiomycota being sister to Monoblepharomycota + Neocallimastigomycota), a branching of Blastocladiomycota + Sanchytriomycota after the Chytridiomyceta but before other non-Opisthosporidia fungi, and a branching of Glomeromycota as sister to the Dikarya. Our most up-to-date fungal tree of life will serve as a springboard for future investigations on the evolution of fungi.

Key words: fungal tree of life, fungi, evolution, phylogenomics, phylogeny

Results

35 **A well-curated and taxon-balanced phylogenomic data set.** To untangle the early
36 diversification in the fungal tree of life, we assembled a 299-protein data set with 661 taxa that
37 represent all extant recognised fungal lineages (Table S1). All proteins were carefully inspected
38 by reconstructing single-protein phylogenies to evaluate orthology and detect contamination.
39 Considering that the taxonomic distribution of publicly available fungal genomes and
40 transcriptomes is highly biased towards Ascomycota and Basidiomycota, we placed emphasis
41 on the inclusion of underrepresented phyla such as the Blastocladiomycota or the fast-evolving
42 early-branching Aphelidiomycota, Rozellomycota, and Microsporidia (together known as
43 Opisthosporidia). To uncover deep phylogenetic relationships, a taxon-reduced data set (101
44 taxa) was derived to allow for more computationally intensive analyses. This taxon-reduced
45 data set maintained a high diversity and coverage and selected slow-evolving over fast-evolving
46 lineages within phyla.
47
48 **Reconstructing the phylogeny of fungi.** Two trees were inferred from the full data set, one by
49 maximum likelihood (ML) under the site-homogeneous LG+G+F model, and the other under
50 the multi-species coalescent (MSC) model. These were largely congruent to each other (Figures
51 S1 and S2); however, as expected for the heterogeneous taxon sampling in our full data set,
52 some of the deeper nodes remained unsupported using these models. The taxon-reduced data
53 set, from which all final analyses were derived and to which is referred to hereafter, was
54 therefore analysed under two site-heterogeneous mixture models: CAT+GTR+Γ4 using
55 Bayesian inference (BI) (hereafter *CATGTR*; Figure 1A), and LG+C60+G+F with posterior
56 mean site frequency profiles using ML (hereafter *ML-C60*; Figures 1B and S3). Both models
57 have been shown to better account for rate variation across sites[15,16] and their implementation
58 led to the recovery of a similar overall topology with maximal statistical support for nearly all
59 nodes (Figure 1A, B). A further tree was inferred under the MSC model, which was highly
60 congruent to the *ML-C60* tree (Figures 1B, S3, and S4). All three analyses (*CATGTR*, *ML-C60*,

61 and MSC; Figure 1, Table 1) were congruent regarding the branching of Chytridiomyceta as
62 sister to all other non-Opisthosporidia fungi; the sister-relationship of Chytridiomycota to a
63 clade comprising Neocallimastigomycota + Monoblepharomycota; the branching of
64 Blastocladiomycota + Sanchytriomycota, which diverged after the Chytridiomyceta but before
65 other non-Opisthosporidia fungi; the sister-relationship of *Olpidium* to all other non-zoosporic
66 fungi; or the sister-relationship of Pucciniomycotina to Agaricomycotina/Wallemiomycotina +
67 Ustilaginomycotina. Aphelidiomycota was fully supported as sister taxon to the entire non-
68 Opisthosporidia clade and distinct from Rozellomycota and Microsporidia, although
69 Aphelidiomycota was represented by only one species in our analysis (Figure 1).
70 Ambiguity remained considering the diverging pattern of the Glomeromycota, either as
71 sister to Mucoromycota + Mortierellomycota as reconstructed in the MSC and *ML-C60* trees or
72 as sister to the Dikarya (Ascomycota + Basidiomycota, and possibly Entorrhizomycota that lack
73 genomic/transcriptomic data[12]) as reconstructed in the *CATGTR* tree (Figure 1). Another
74 discrepancy between the inferred tree topologies was the placement of the opisthosporid
75 *Mitosporidium* as either sister to only *Paramicrosporidium* (MSC and *ML-C60*; Figures S3 and
76 S4) or sister to all other Microsporidia including *Paramicrosporidium* (*CATGTR*; Figure 1A).
77
78 **Tree evaluation.** The conflicts in the branching of the Glomeromycota and *Mitosporidium*
79 were fully supported by bootstrap (*ML-C60*) and posterior probability (*CATGTR*) analyses.
80 Both alternative branching orders in the *CATGTR* topology were also rejected by ML in
81 Approximately Unbiased tests (Glomeromycota + Dikarya, p-AU = 0.022; *Mitosporidium* + all
82 other Microsporidia, p-AU = 0.011). To test whether the conflicts were due to the usage of
83 different tree reconstruction methods (ML or BI), a further tree was derived using BI under the
84 same C60 model that was used for the ML tree. This tree (hereafter *BI-C60*; Figure S5)
85 recapitulated the topology obtained with *ML-C60* indicating that differences were due to the
86 evolutionary model rather than the inference method. To test which model better minimised

87 inadequacy in describing compositional heterogeneity, posterior predictive analyses were
88 conducted that resulted in a better fit of *CATGTR* compared to *BI-C60* (Table 2). We also
89 removed 25% and 50% of the most compositionally heterogeneous sites from the alignment in
90 order to reduce compositional heterogeneity, and additional trees were reconstructed from these
91 new alignments with *CATGTR* (Figure 2). The inferred trees were in agreement with the
92 *CATGTR* tree reconstructed from the full-length alignment (Figures 1A and 2).
93 In order to understand the extent to which fast-evolving sites, which are more likely to
94 be homoplasic, led to systematic bias in the *ML-C60* tree inference due to model
95 misspecification[17], the fastest-evolving sites were progressively removed from the alignment in
96 increments of 10,000 sites. Each sub-alignment was then subjected to further ML tree
97 inferences using the LG+C60+G+F model and node support was inferred by ultrafast bootstrap
98 approximation. Because of the higher computational demand of BI, this analysis was run for
99 ML inferences only. Interestingly, whereas the bootstrap support for Dikarya remained
100 maximal throughout all the analyses, the support for the sister-relationship of Glomeromycota
101 to Mucoromycota + Mortierellomycota decreased rapidly while the support for their closer
102 relationship to the Dikarya, as seen in the *CATGTR* trees reconstructed from complete and
103 pruned alignments, increased, reaching 92% after the removal of 30,000 sites (Figure 3).
104 Further removal of sites led either to an unsupported branching or recovered the
105 Glomeromycota + Dikarya topology again (100,000 sites removed; Figure 3). For the fast-
106 evolving microsporidians, the sister-relationship of *Mitosporidium* to all other Microsporidia, as
107 inferred by *CATGTR*, was only supported when the 120,000 fastest-evolving sites were
108 removed from the alignment (i.e. 8,450 sites remained; Figure 3).
109 Finally, to assess whether the here received topologies were skewed due to long branch
110 attraction (LBA) effects[18], further *ML-C60* and *CATGTR* trees were calculated but with the
111 Sanchytriomycota and the nine fastest-evolving Microsporidia species removed. In both trees,
112 the branching order of the remaining fungal groups was fully congruent to the branching order

113 that has been recovered before under the same evolutionary model, using the alignment that
114 includes the fast-evolving taxa (compare Figure 1 with Figure S6).
115

116 ## Discussion

117 Fungi show a great diversity in morphology and lifestyle. They are important decomposers and
118 through the provision of nutrients, various lineages are essential for the survival of >80% of all
119 plant species[19]. As parasites, chytrid infections can have a significant impact on amphibian
120 populations[21] and may even affect the planet's carbon cycle by terminating algal blooms[20].
121 Despite their ecological importance, worldwide distribution, and essential roles in human
122 history as food source, leavening and fermentation agents, and medicines, our knowledge of
123 how the huge diversity of fungi evolved is still limited. Phylogenomic studies encompassing the
124 entire fungal kingdom with nearly all its different phyla are scarce — either due to missing data
125 or to a focus on only a few certain groups of fungi[10,22–25]. In this study, we tested whether a
126 more balanced taxon-sampling, in combination with new filtering methods[26,27] (see
127 STAR*Methods) that improve the phylogenetic signal of each single-protein alignment and
128 best-fitting models of evolution, could untangle some of the hitherto debated early
129 diversifications in the fungal tree of life.

130 The branching order among fungal phyla remained consistent and well supported across
131 multiple analyses using site-heterogeneous models in BI (*CATGTR* and *BI-C60*) and ML (*ML-*
132 *C60*) tree reconstructions, and using an MSC approach (Figures 1, 2, and S3–S6). This applies
133 even to phyla that were represented by only a single species such as Olpidiomycota, whose
134 sister-relationship to non-flagellated fungi has recently been proposed based on
135 phylogenomics[28]. The sole exception (on the phylum level) concerned the placement of the
136 Glomeromycota, which was sister to the Dikarya in *CATGTR*-based analyses or was sister to
137 the Mucoromycota/Mortierellomycota clade in the other analyses. Glomeromycota were

138 suggested to form an independent phylum at the beginning of this century[29], but closer
139 phylogenetic affiliation has remained controversial ever since[12,25,30,31]. A sister-relationship to
140 the Dikarya has been proposed based on tree inferences using the elongation factor EF-1 alpha,
141 RNA polymerase II subunits and/or rRNA gene alignments[6,29,31–33]. This placement was
142 discussed to be artificial due to lineage-specific rates in these data sets[12] and was rejected in a
143 ML tree inference from a 192-protein matrix. However, we note that this matrix analysis used
144 an LG model with fixed base frequencies[34], and that our results support a more recent
145 *CATGTR*-based analysis of a 264-protein matrix that also recovered Glomeromycota as sister to
146 the Dikarya[10]. It is reasonable to assume that this finding depicts the real scenario in the
147 evolution of fungi and that the discovered sister-relationship to the
148 Mucoromycota/Mortierellomycota clade in the *ML-C60* tree is a consequence of the inadequacy
149 of the C60 model for capturing compositional heterogeneity. This is in line with the results
150 obtained by the posterior predictive test, in which the fit of both models, *CATGTR* and *BI-C60*,
151 was directly compared using the PhyloBayes software package (Table 2). Congruent topologies
152 reconstructed from the analyses of the same alignment using both *BI-C60* and *ML-C60* (Figures
153 S3 and S5) give further evidence that the found discrepancies in the branching order of the
154 Glomeromycota are rather caused by the less fitting evolutionary C60 model than by the use of
155 BI or ML. Interestingly, however, Glomeromycota + Dikarya was also supported by *ML-C60*
156 tree inferences after only 15% of the fastest-evolving sites were removed (Figure 3), indicating
157 that the conflicting placement of Glomeromycota in the C60 analysis of the full alignment may
158 be due to statistical problems this model encounters with homoplasic sites.
159 The relationships among the opisthosporidian phyla remained controversial for a long
160 time[6,22,35,36]. Here, the Aphelidiomycota representative was placed as sister to all non-
161 Opisthosporidia fungi in all analyses (Figures 1, 2, and S3–S6). Similar findings have been
162 reported before based on phylogenomic studies corroborating the paraphyletic nature of
163 Opisthosporidia[10,35]. The assignment of some of the earliest-branching fungal species to either

164 the Rozellomycota or the Microsporidia is more controversial, leading to debates as to whether
165 these phyla are monophyletic or not[6,11,22]. *Paramicrosporidium saccamoebae* was initially
166 assigned to the Rozellomycota based on a phylogenetic analysis of SSU rRNA gene sequences,
167 but an unsupported affiliation to the Microsporidia was recovered when 5.8S and partial LSU
168 rRNA gene sequences were added to the analyses[37]. We found the exact placement of *P.*
169 *saccamoebae* to be questionable. While trees inferred with the better fitting *CATGTR* model
170 recovered this species to be nested between the microsporidians *Mitosporidium daphniae* and
171 *Amphiamblys* sp. (Figures 1A and 2), a sister-relationship to *M. daphniae* was recovered in the
172 MSC tree and with the C60 model in BI and ML inferences (Figures S3–S5). For these fast-
173 evolving lineages, the latter topology was only challenged when at least 85% of the fastest-
174 evolving sites were removed (Figure 3). However, independent from the exact position of *P.*
175 *saccamoebae*, the only safely assigned Rozellomycota species with available genomic data,
176 *Rozella allomycis*, formed a sister to all Microsporidia and *P. saccamoebae* in all analyses
177 (Figures 1, 2, and S3–S6). Taking into account these results, the results obtained by rRNA
178 analyses[37] (see above) and the morphology of *P. saccamoebae*, which resembles that of
179 microsporidians (they lack a mitochondrion and have tiny spores with a polar filament[37]), its
180 assignment to the Microsporida seems to be reasonable and is proposed here. Thus,
181 Microsporidia and Rozellomycota form two distinct lineages for which a frequently discussed
182 paraphyly[6,22,25] cannot be confirmed here. Whether or not Microsporidia + Rozellomycota
183 should be classified as only one phylum is currently debated[6,31,38] and will require data from
184 additional representatives of Rozellomycota.

185 Throughout all analyses, our study revealed not only consistency regarding the
186 previously contested phylogenetic position of the zoosporic subkingdom Chytridiomyceta[e.g.
187 10,25], which live either as saprophytes, parasites, or as intermediate forms, but also regarding the
188 discussed branching order among its phyla[6,12,30]. With that, however, the justification for the
189 separation of the chytridiomycete phylum "Caulochytriomycota" from the Chytridiomycota

190 remains enigmatic (the erection of this phylum was proposed by Doweld in 2014[39] and has been
191 adapted by other researchers since). Only two species of the genus *Caulochytrium* are known,
192 and DNA sequences are available only for *Caulochytrium protostelioides*. In our analysis, *C.*
193 *protostelioides* was either closely related to, or even nested within, the Chytridiomycota, and
194 the clade was well-supported in all analyses (Figures 1, 2, and S3–S6). We therefore suggest
195 *Caulochytrium* be assigned to the Chytridiomycota based on current data. The here
196 unambiguously documented sister-relationship of Chytridiomycota to Neocallimastigomycota +
197 Monoblepharomycota makes them the earliest diverging group within the Chytridiomyceta — a
198 finding that is rather surprising taking into account that the zoospores of Chytridiomycota show
199 more morphological similarities to those of Monoblepharomycota than to those of
200 Neocallimastigomycota (in addition, members of the latter are the only opisthokonts, whose
201 zoospores can bear multiple flagella)[6,12]. Our study is the first to position the chytrid
202 *Hyaloraphidium curvatum* based on phylogenomic analyses and clearly shows its affiliation to
203 the Monoblepharomycota, which was represented by *Gonapodya* in our data set.
204 *Hyaloraphidium* was initially miss-classified as a green alga and later proposed to cluster with
205 Chytridiomyceta based on SSU rRNA gene sequence analysis[40]. However, its closer affiliation
206 has remained enigmatic ever since.
207 The previously unclear[12,25] but here consistent placement of Blastocladiomycota, which
208 together with the Sanchytriomycota diverged after the Chytridiomyceta but before other non-
209 Opisthosporidia fungi, seems in agreement with the finding of a distinct morphology of their
210 zoospores from those produced by Chytridiomyceta. Another distinguishing feature is the
211 sporic meiosis that results in a regular alternation of haploid and diploid generations in
212 Blastocladiomycota[41].
213
214 **Final remarks.** The here presented tree (Figure 1A) provides a robust backbone topology for
215 the diversification of major fungal groups. For the recovery of ancient nodes, not only site rate

216 variation but also variation of the sequence composition across branches and/or alignment sites
217 can hamper the reconstruction of topology and branch length. Thus, 25% and 50% of the sites
218 that contribute the most to branch heterogeneity were removed from the alignment and new
219 *CATGTR* trees were inferred (Figure 2). Both trees reflected the overall topology of the tree
220 inferred from the complete alignment (Figure 1A) indicating that compositional biases had
221 either no or a negligible negative impact that could result in model misspecification despite
222 using the best-fitting CAT+GTR+Γ4 model. Considering that also LBA effects did not seem to
223 have a significant impact on the received *ML-C60* and *CATGTR* topologies (Figures 1 and S6),
224 it is likely that the found incongruence between these trees resulted from an inability of the *ML-*
225 *C60* model to model fast-evolving and heterogeneous sites. This would be in line with the
226 findings of the fast-evolving sites removal analysis (Figure 3). However, it is noteworthy that
227 the *CATGTR* backbone topology provided here — like all phylogenetic trees — represents a
228 hypothesis and it is safe to assume that with an increasing diversity of fungi for which
229 genomic/transcriptomic data will be available, several clades will be reshuffled. It should also
230 be noted that the taxonomic classification in fungi is changing rapidly. The lineage information
231 in this study was assigned according to the NCBI taxonomy (ncbi.nlm.nih.gov/taxonomy) with
232 some modifications especially regarding the phylum assignments, which, if not otherwise
233 noted, were adopted in order to match those recently proposed by Wijayawardene *et al*.[11] A
234 comprehensive revision of the taxonomic classification in fungi, based on this and other recent
235 studies[e.g. 10,25], is desirable.
236

237 **Supplemental information**

238 Table S1: Lineage and source information for all taxa used in this study.
239 Document S1: Figures S1–S6.

240 Data S1: Sequence data, single-protein trees, and BI consensus trees from individual MCMC
241 chains.
242

243 Acknowledgments

244 J.F.H.S. acknowledges support from the German Research Foundation (DFG; Grant
245 STR1349/2-1 Project No. 432453260) and research was partially funded by the German Federal
246 Ministry of Education and Research (BMBF, Förderkennzeichen 033W034A). The authors
247 thank the High-Performance Computing Service of ZEDAT, Freie Universität Berlin, for
248 computing time. Finally, we thank David Moreira and Luis Javier Galindo for sharing sequence
249 data for *Amoeboradix gromovi* and *Sanchytrium tribonematis*.
250

251 Author contributions

252 J.F.H.S. conceived the study, assembled and curated the data, performed all analyses, and
253 drafted the manuscript. M.T.M. contributed to the final manuscript. Both authors read and
254 approved the final version.
255

256 Declaration of interests

257 The authors declare no competing interests.
258

259 Figure legends

260 **Figure 1. Phylogenetic trees showing the early diversification of fungi**
261 (A) Bayesian consensus tree inferred from a concatenated alignment of 299 proteins and 101
262 taxa under the CAT+GTR+Γ4 (*CATGTR* in the text) model. Node support is given by posterior
263 probabilities (full support is indicated by black circles). (B) Two phylogenetic trees showing
264 perfect congruence. Left: Tree inferred by ML under the LG+C60+G+F-PMSF (*ML-C60* in the

265 text) model using the same alignment as used for the Bayesian tree. Nonparametric bootstrap
266 support is not indicated but was maximal for all nodes shown (for details, see Figure S3). Right:
267 Tree based on summary-coalescent analyses of 299 single-protein ML phylogenies.
268 Quadripartition support values are given for nodes that are not fully supported (for details, see
269 Figure S4). The conflicting branching of the Glomeromycota in the CAT+GTR+Γ4 analysis
270 and the LG+C60+G+F-PMSF and summary-coalescent analyses is highlighted.
271 *_Caulochytrium_ and the subphyla of the Zoopagomycota were recently proposed to be
272 independent phyla[11] (see Discussion).
273
274 **Figure 2. Heterogeneous sites removal analysis**
275 Simplified Bayesian consensus tree inferred under the CAT+GTR+Γ4 (*CATGTR* in the text)
276 model from a concatenated alignment of 299 proteins and 101 taxa with 50% of the most
277 heterogeneous sites removed. All nodes shown were fully supported. A further tree, which was
278 inferred from the same alignment but with 25% of the most heterogeneous sites removed, had
279 an identical topology (not shown). Microsporidians are only partly collapsed in order to depict
280 the position of *Paramicrosporidium saccamoebae*. *Caulochytrium protostelioides* was nested
281 within the Chytridiomycota (same branching pattern as in Figure 1A; see Discussion).
282
283 **Figure 3. Fast-evolving sites removal analysis**
284 ML ultrafast bootstrap approximation for selected nodes using the full matrix and subsets from
285 which fast-evolving sites were removed in 10,000 increments.
286
287 **Tables**
288 **Table 1.** Nodes of major interest.

Node	**Model(s) supporting node**
Pucciniomycotina + (Agaricomycotina/Wallemiomycotina + Ustilaginomycotina)	ALL

Glomeromycota + Dikarya	*CATGTR*
Glomeromycota + (Mucoromycota + Mortierellomycota)	MSC, *ML-C60*[§]
Olpidiomycota + all other non-zoosporic fungi	ALL
(Blastocladiomycota + Sanchytriomycota) + non-Opisthosporidia excl. Chytridiomyceta*	ALL
Chytridiomyceta* + non-Opisthosporidia	ALL
Chytridiomycota + (Neocallimastigomycota + Monoblepharomycota)	ALL
Aphelidiomycota + non-Opisthosporidia	ALL
Mitosporidium + *Paramicrosporidium*	MSC, *ML-C60*[§]
Mitosporidium + (*Paramicrosporidium* + all other Microsporidia)	*CATGTR*

*Chytridiomycota, Neocallimastigomycota, and Monoblepharomycota.
[§]Not supported after the removal of fast-evolving sites (see Figure 3).

Table 2. 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

STAR*Methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Sequence data, single-protein trees, and BI consensus trees from individual MCMC chains	This paper	Data S1
Software and algorithms		
ASTRAL-III v. 5.7.7	[42]	https://github.com/smirarab/ASTRAL
BLAST v. 2.7.1	[43]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
BUSCO v. 5.1.2	[44]	https://busco.ezlab.org/
CD-HIT v. 4.8.1	[45]	http://weizhong-lab.ucsd.edu/cd-hit/
Divvier v. 1.01	[27]	https://github.com/simonwhelan/Divvier
FastTreeMP v. 2.1.11	[46]	http://www.microbesonline.org/fasttree/
IQ-Tree v. 1.6.12	[47]	http://www.iqtree.org/
MAFFT v. 7.475	[48]	https://mafft.cbrc.jp/alignment/software/
PEAR v. 0.9.11	[49]	https://cme.h-its.org/exelixis/web/software/pear/

PhyloBayes-MPI v. 1.8	[50]	https://github.com/bayesiancook/pbmpi
Prequal v. 1.02	[49]	https://github.com/simonwhelan/prequal
SCaFoS v. 1.25	[51]	https://megasun.bch.umontreal.ca/Software/scafos/scafos.html
Sickle v. 1.33	[52]	https://github.com/najoshi/sickle
SPAdes v. 3.14.1	[53]	https://github.com/ablab/spades
TransDecoder v. 5.5.0	[54]	https://github.com/sghignone/TransDecoder
trimAL v. 1.4	[55]	http://trimal.cgenomics.org/
Trimmomatic v. 0.39	[56]	https://github.com/usadellab/Trimmomatic
Trinity v. 2.10.0	[54]	https://github.com/trinityrnaseq/trinityrnaseq/releases

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jürgen F. H. Strassert (juergen.strassert@igb-berlin.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data needed to evaluate the conclusions of this study are presented in the paper and the Supplemental information. 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. The script alignment_pruner.pl is available at github.com/novigit/davinciCode/blob/master/perl. Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

Experimental model and subject details

312 Genomes and transcriptomes were downloaded from EukProt, Ensembl, Imicrobe, Mycocosm,
313 and NCBI. Sequences of outgroup taxa were added from Strassert et al.[7] For detailed source
314 information, see Table S1. Taxa were selected such that all major fungal groups were
315 represented, neglecting taxa with more missing data. Analyses were carried out using the Zedat
316 high-performance computing infrastructure at Freie Universität Berlin[57].
317
318 **Method details**
319 **Phylogenomic data set construction.** Genomes and transcriptomes of a broad range of fungi
320 and selected non-fungal lineages were downloaded from EukProt[58], fungi.ensembl.org,
321 metazoa.ensembl.org, protists.ensembl.org, imicrobe. us/#/projects/104, mycocosm.jgi.doe.gov,
322 and ncbi.nlm.nih.gov (Table S1). For unassembled sequence data, paired end reads were
323 trimmed using Trimmomatic v. 0.39[56] with the following flags employed: LEADING: 3,
324 TRAILING: 3, SLIDINGWINDOW: 4:15, and MINLEN: 36. Transcriptomic data was
325 assembled and translated into amino acids by using the default settings in Trinity v. 2.10.0 and
326 TransDecoder v. 5.5.0, respectively[54]. Reads of genomic data were merged with PEAR v.
327 0.9.11[49], quality filtered with Sickle v. 1.33[52] (only the paired but unmerged reads), and
328 assembled with SPAdes v. 3.14.1[53] using the --isolate option. Genes were predicted using
329 BUSCO v. 5.1.2 with the database fungi_odb10[44].
330 A recently published data set comprising 320 proteins for diverse eukaryotic lineages
331 (733) was used as starting point for data set construction[7]. With a few exceptions that were kept
332 as outgroup, non-fungal linages were removed. The data set was then expanded by adding a
333 broad range of fungal lineages as follows: 1) Protein sequences for each taxon were clustered
334 with CD-HIT v. 4.8.1[45] using an identity threshold of 85%. 2) Candidates for homologous
335 copies were retrieved by BLASTP searches[43] using the 320 proteins as queries (e-value: 1e-20;
336 coverage cutoff: 0.5). 3) In three rounds, phylogenetic trees were inferred and carefully
337 inspected in order to detect and remove paralogs and contaminants and select orthologs. In the

338 first round, sequences were aligned with MAFFT v. 7.475[48] using the --auto option, filtered
339 with trimAL v. 1.4[55] using a gap threshold of 0.8, and ML trees were inferred with FastTreeMP
340 v. 2.1.11[46] using -lg -gamma and options for more accurate performances. In the second and
341 third round, non-homologues sequence stretches were filtered out prior to aligning using
342 Prequal v. 1.02[26] with a threshold of 0.95. Protein sets were then aligned using MAFFT G-INS-
343 i with the following options: --allowshift, --unalignlevel 0.6, --maxiterate 0. Further non-
344 homologues residues were removed with Divvier v. 1.01[27] employing the -partial flag and
345 alignments were trimmed with trimAL using a soft gap threshold of 0.05. Finally, partial
346 sequences belonging to the same taxon that did not show evidence for paralogy or
347 contamination were merged. Single-protein ML phylogenies (Data S1) were reconstructed with
348 IQ-Tree v. 1.6.12[47] under BIC-selected models including site-homogeneous models (such as
349 LG) and empirical profile models (C10–C60) and node support was inferred by ultrafast
350 bootstrap approximation[59] (1,000 replicates).
351 Of the initial 320 proteins, 21 were removed due to a high proportion of missing taxa or
352 incongruous clustering of several clades. The final data set comprised 299 proteins and 637
353 fungal taxa plus 24 non-fungal taxa (Data S1). The protein sequences were concatenated using
354 SCaFoS v. 1.25[51]. The obtained matrix (129,814 amino acid positions) was then used to
355 calculate an initial ML tree with IQ-Tree employing the site-homogeneous model LG+G+F and
356 ultrafast bootstrap approximation (1,000 replicates; Figure S1). To allow computationally more
357 demanding analyses, a taxon-reduced data set was built representing all major fungal groups
358 and discarding taxa with more missing data. Corresponding sequences were newly aligned,
359 filtered, and concatenated (same settings as described above; 101 taxa, 128,450 amino acid
360 positions). New ML trees were inferred for each of the taxon-reduced protein alignments as
361 described above (Data S1).
362

363 **Phylogenomic analyses.** The taxon-reduced matrix was used to reconstruct a ML tree with IQ-
364 Tree using the best-fitting site-heterogeneous model LG+C60+G+F with the PMSF approach[16]
365 to obtain nonparametric bootstrap support (100 replicates; Figures 1B and S3). Copies of this
366 tree were manually edited to test alternative topologies with the approximately unbiased test
367 (AU test[60]). The taxon-reduced matrix was also subjected to Bayesian analyses using
368 PhyloBayes-MPI v. 1.8[50] with the site-heterogeneous CAT+GTR+Γ4 model and the -dc option
369 to remove constant sites. Three independent Markov Chain Monte Carlo (MCMC) chains were
370 run for 1,600 generations (all sampled). For each chain, the burnin period was estimated by
371 monitoring evolution of the log-likelihood (Lnl) at each sampled point, and 750 generations
372 were removed from all chains. A consensus tree (Figure 1A) was built and global convergence
373 between chains was assessed with PhyloBayes' built-in tools. As frequently recognised in
374 PhyloBayes, global convergence was not received in all combinations of the chains with
375 maxdiff = 1 and meandiff = 0.0167504. To discriminate whether potential discrepancies
376 between this tree and the tree calculated by ML are due to the usage of different software
377 packages or due to different evolutionary models, a further Bayesian tree (Figure S5) was
378 inferred under the LG+C60 model (-dc -catfix C60 -lg -dgam 4; three chains, each with 2,000
379 generations; convergence was not received; maxdiff = 1, meandiff = 0.0268007). Posterior
380 predictive analyses were then run for both tree inferences in order to informatively select the
381 best topology.
382 The single-protein ML trees of both the full and the taxon-reduced data sets (Data S1)
383 were also analysed under the multi-species Coalescent (MSC) model in ASTRAL-III v. 5.7.7[42]
384 to incorporate protein tree uncertainties (Figures 1, S2, and S4). Prior to that, branches with
385 bootstrap support <10 were collapsed as recommended by the developers. Quadripartition
386 supports were calculated from quartet frequencies among the set of input trees[61].
387

388 **Sites removal analyses.** 25% and 50% of the most heterogeneous sites were removed from the
389 taxon-reduced matrix using the script alignment_pruner.pl
390 (https://github.com/novigit/davinciCode/blob/master/perl). Bayesian trees (Figure 2) were
391 inferred form the pruned alignments with the CAT+GTR+Γ4 model: 25% pruned: three chains,
392 each with 1,600 generations (burnin 1,000); convergence was not received; maxdiff = 1,
393 meandiff = 0.0187389. 50% pruned: three chains, each with 2,450 generations (burnin 1,250);
394 convergence was not received; maxdiff = 1, meandiff = 0.0128211. Also for the taxon-reduced
395 matrix, fast-evolving sites were estimated with IQ-Tree under the LG+C60+G+F model,
396 employing the -wsr flag and giving the PMSF tree as fixed tree topology. Sites were then
397 removed in 10,000 increments and ML trees were inferred from the site-reduced alignments
398 using IQ-Tree (LG+C60+G+F -bb 1,000 -bnni -wbtl; Figure 3).
399
400 **Long branch attraction test.** From the taxon-reduced data set, the eleven fastest-evolving
401 lineages (two Sanchytriomycota species and nine Microsporidia species) were removed and the
402 corresponding sequences were newly aligned, filtered, and concatenated as described above (90
403 taxa, 129,049 amino acid positions). The matrix was then used to infer a ML tree with IQ-Tree
404 using the best-fitting LG+C60+G+F model and the PMSF approach to obtain nonparametric
405 bootstrap support (100 replicates; Figure S6). Also, one further Bayesian tree was reconstructed
406 using the CAT+GTR+Γ4 model (Figure S6): three chains, each with 900 generations (burnin
407 500); convergence was not received; maxdiff = 1, meandiff = 0.0188019.
408
409 **Quantification and statistical analysis**
410 Phylogenetic model selection was based on IQ-Tree's Bayesian information criterion[47].
411 Statistical support for phylogenies was obtained using non-parametric bootstraps, ultrafast
412 bootstraps[59], Bayesian posterior probabilities in BI[50], and quadripartition support values in
413 ASTRAL[42,61]. Alternative topologies were tested using the approximately unbiased test

414 implemented in IQ-Tree[47], and the fit of evolutionary models (*CATGTR* and *BI-C60*) was tested
415 using posterior predictive analyses implemented in the PhyloBayes software package[50].
416

417 **References**

418 1. Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K.,
419 Shiratori, T., Ishida, K.I., Hashimoto, T., Simpson, A.G.B., et al. (2018). Phylogenomics
420 places orphan protistan lineages in a novel eukaryotic super-group. Genome Biol Evol
421 *10*, 427–433.
422 2. Strassert, J.F.H., Jamy, M., Mylnikov, A.P., Tikhonenkov, D. V., and Burki, F. (2019).
423 New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the
424 eukaryote tree of life. Mol Biol Evol *36*, 757–765.
425 3. Irisarri, I., Strassert, J.F.H., and Burki, F. (2022). Phylogenomic insights into the origin
426 of primary plastids. Syst Biol *71*, 105–120.
427 4. Gawryluk, R.M.R., Tikhonenkov, D. V., Hehenberger, E., Husnik, F., Mylnikov, A.P.,
428 and Keeling, P.J. (2019). Non-photosynthetic predators are sister to red algae. Nature
429 *572*, 240–243.
430 5. Schön, M.E., Zlatogursky, V. V, Singh, R.P., Poirier, C., Wilken, S., Mathur, V.,
431 Strassert, J.F.H., Pinhassi, J., Worden, A.Z., Keeling, P.J., et al. (2021). Single cell
432 genomics reveals plastid-lacking Picozoa are close relatives of red algae. Nat Commun
433 *12*, 6651.
434 6. Voigt, K., James, T.Y., Kirk, P.M., Santiago, A.L.C.M. de A., Waldman, B., Griffith,
435 G.W., Fu, M., Radek, R., Strassert, J.F.H., Wurzbacher, C., et al. (2021). Early-diverging
436 fungal phyla: taxonomy, species concept, ecology, distribution, anthropogenic impact,
437 and novel phylogenetic proposals. Fungal Divers *109*, 59–98.
438 7. Strassert, J.F.H., Irisarri, I., Williams, T.A., and Burki, F. (2021). A molecular timescale

439 for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat
440 Commun *12*, 1879.
441 8. Hawksworth, D.L. (2001). The magnitude of fungal diversity: the 1.5 million species
442 estimate revisited. Mycol Res *105*, 1422–1432.
443 9. Hawksworth, D.L., Robert, L., Joseph, H., and James, T.Y. (2017). Fungal diversity
444 revisited: 2.2 to 3.8 million species. Microbiol Spectr *5*, 5.4.10.
445 10. Galindo, L.J., López-García, P., Torruella, G., Karpov, S., and Moreira, D. (2021).
446 Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution
447 across Holomycota. Nat Commun *12*, 4973.
448 11. Wijayawardene, N.N. (2020). Outline of fungi and fungus-like taxa. Mycosphere *11*,
449 1060–1456.
450 12. James, T.Y., Stajich, J.E., Hittinger, C.T., and Rokas, A. (2020). Toward a fully resolved
451 fungal tree of life. Annu Rev Microbiol *74*, 291–313.
452 13. James, T.Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J.E.
453 (2013). Shared signatures of parasitism and phylogenomics unite cryptomycota and
454 microsporidia. Curr Biol *23*, 1548–1553.
455 14. Keeling, P.J., Luker, M.A., and Palmer, J.D. (2000). Evidence from beta-tubulin
456 phylogeny that microsporidia evolved from within the fungi. Mol Biol Evol *17*, 23–31.
457 15. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site
458 heterogeneities in the amino-acid replacement process. Mol Biol Evol *21*, 1095–1109.
459 16. Wang, H.C., Minh, B.Q., Susko, E., and Roger, A.J. (2018). Modeling site heterogeneity
460 with posterior mean site frequency profiles accelerates accurate phylogenomic
461 estimation. Syst Biol *67*, 216–235.
462 17. Rodríguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., and
463 Philippe, H. (2007). Detecting and overcoming systematic errors in genome-scale
464 phylogenies. Syst Biol *56*, 389–399.

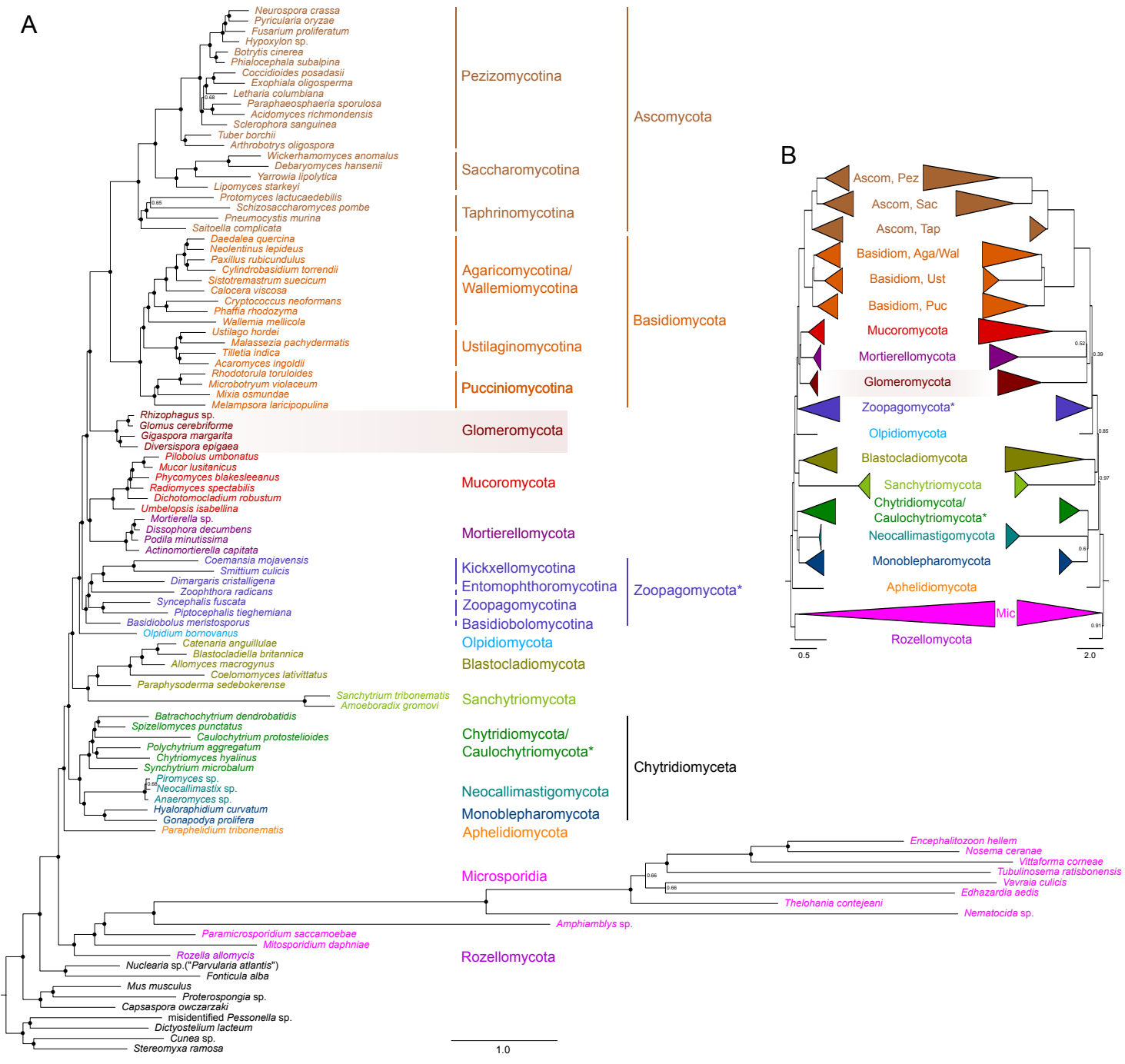
465 18. Bergsten, J. (2005). A review of long-branch attraction. Cladistics *21*, 163–193.
466 19. Wang, B., and Qiu, Y.-L. (2006). Phylogenetic distribution and evolution of mycorrhizas
467 in land plants. Mycorrhiza *16*, 299–363.
468 20. Frenken, T., Alacid, E., Berger, S.A., Bourne, E.C., Gerphagnon, M., Grossart, H.-P.,
469 Gsell, A.S., Ibelings, B.W., Kagami, M., Kupper, F.C., et al. (2017). Integrating chytrid
470 fungal parasites into plankton ecology: research gaps and needs. Environ Microbiol *19*,
471 3802–3822.
472 21. Fisher, M.C., and Garner, T.W.J. (2020). Chytrid fungi and global amphibian declines.
473 Nat Rev Microbiol *18*, 332–343.
474 22. Quandt, C.A., Beaudet, D., Corsaro, D., Walochnik, J., Michel, R., Corradi, N., and
475 James, T.Y. (2017). The genome of an intranuclear parasite, *Paramicrosporidium*
476 *saccamoebae,* reveals alternative adaptations to obligate intracellular parasitism. Elife *6*,
477 e29594.
478 23. Choi, J., and Kim, S.-H. (2017). A genome tree of life for the Fungi kingdom. Proc Natl
479 Acad Sci U S A *114*, 9391–9396.
480 24. Ahrendt, S.R., Quandt, C.A., Ciobanu, D., Clum, A., Salamov, A., Andreopoulos, B.,
481 Cheng, J.-F., Woyke, T., Pelin, A., Henrissat, B., et al. (2018). Leveraging single-cell
482 genomics to expand the fungal tree of life. Nat Microbiol *3*, 1417–1428.
483 25. Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., Spatafora, J.W.,
484 Groenewald, M., Dunn, C.W., Hittinger, C.T., et al. (2021). A genome-scale phylogeny
485 of the kingdom Fungi. Curr Biol *31*, 1653–1665.
486 26. Whelan, S., Irisarri, I., and Burki, F. (2018). PREQUAL: detecting non-homologous
487 characters in sets of unaligned homologous sequences. Bioinformatics *34*, 3929–3930.
488 27. Ali, R.H., Bogusz, M., and Whelan, S. (2019). Identifying clusters of high confidence
489 homologies in multiple sequence alignments. Mol Biol Evol *36*, 2340–2351.
490 28. Chang, Y., Rochon, D., Sekimoto, S., Wang, Y., Chovatia, M., Sandor, L., Salamov, A.,

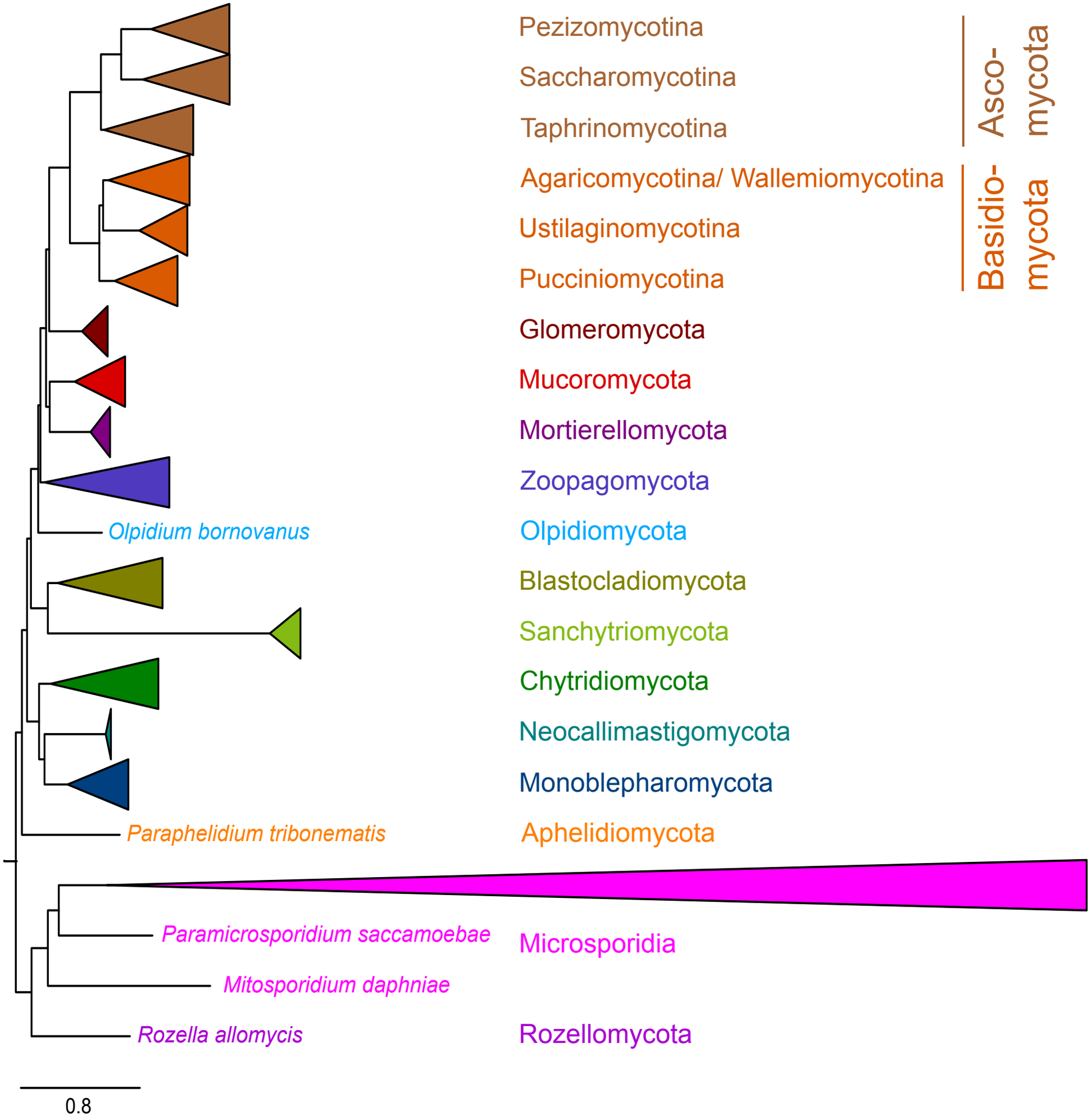
491 Grigoriev, I. V, Stajich, J.E., and Spatafora, J.W. (2021). Genome-scale phylogenetic
492 analyses confirm Olpidium as the closest living zoosporic fungus to the non-flagellated,
493 terrestrial fungi. Sci Rep *11*, 3217.
494 29. Schüβler, A., Schwarzott, D., and Walker, C. (2001). A new fungal phylum, the
495 Glomeromycota: phylogeny and evolution. Mycol Res *105*, 1413–1421.
496 30. Spatafora, J.W., Aime, M.C., Grigoriev, I. V, Martin, F., Stajich, J.E., and Blackwell, M.
497 (2017). The fungal tree of life: from molecular systematics to genome-scale phylogenies.
498 In Microbiology Spectrum, J. Heitmann, B. Howlett, P. Crous, E. Stuckenbrock, T.
499 James, and N. Gow, eds., pp. 3–34.
500 31. Tedersoo, L., Sánchez-Ramírez, S., Kõljalg, U., Bahram, M., Döring, M., Schigel, D.,
501 May, T., Ryberg, M., and Abarenkov, K. (2018). High-level classification of the Fungi
502 and a tool for evolutionary ecological analyses. Fungal Divers *90*, 135–159.
503 32. James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G.,
504 Gueidan, C., Fraker, E., Miadlikowska, J., et al. (2006). Reconstructing the early
505 evolution of Fungi using a six-gene phylogeny. Nature *443*, 818–822.
506 33. Strassert, J.F.H., Wurzbacher, C., Hervé, V., Antany, T., Brune, A., and Radek, R.
507 (2021). Long rDNA amplicon sequencing of insect-infecting nephridiophagids reveals
508 their affiliation to the Chytridiomycota and a potential to switch between hosts. Sci Rep
509 *11*, 396.
510 34. Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L.,
511 Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., et al. (2016). A phylum-level
512 phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia
513 *108*, 1028–1046.
514 35. Torruella, G., Grau-Bové, X., Moreira, D., Karpov, S.A., Burns, J.A., Sebé-Pedrós, A.,
515 Völcker, E., and López-García, P. (2018). Global transcriptome analysis of the aphelid
516 *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. Commun Biol *1*,

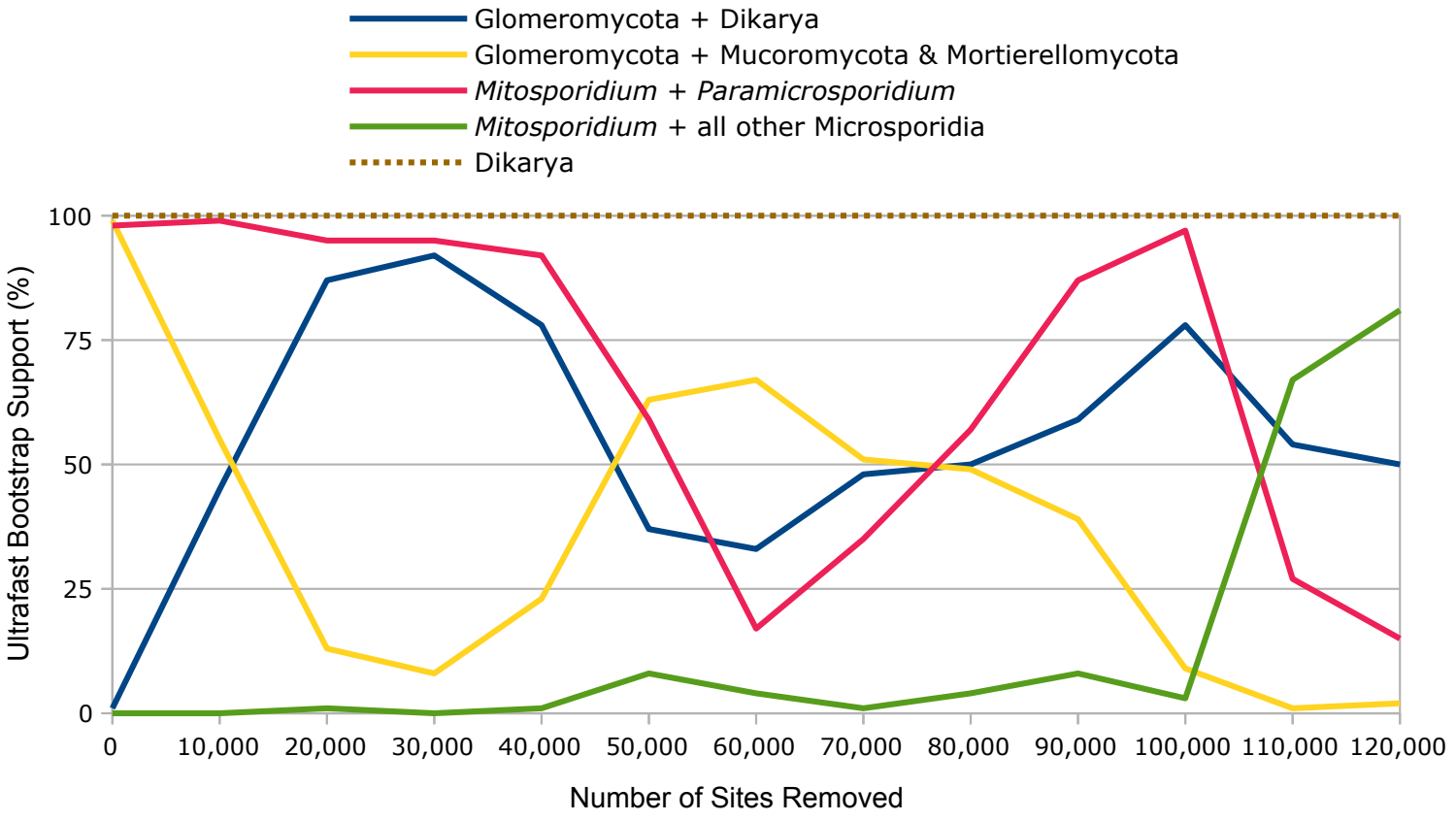
517 231.
518 36. Bass, D., Czech, L., Williams, B.A.P., Berney, C., Dunthorn, M., Mahé, F., Torruella,
519 G., Stentiford, G.D., and Williams, T.A. (2018). Clarifying the relationships between
520 Microsporidia and Cryptomycota. J Eukaryot Microbiol *65*, 773–782.
521 37. Corsaro, D., Walochnik, J., Venditti, D., Steinmann, J., Müller, K.-D., and Michel, R.
522 (2014). Microsporidia-like parasites of amoebae belong to the early fungal lineage
523 Rozellomycota. Parasitol Res *113*, 1909–1918.
524 38. Wijayawardene, N.N., Pawłowska, J., Letcher, P.M., Kirk, P.M., Humber, R.A.,
525 Schüßler, A., Wrzosek, M., Muszewska, A., Okrasińska, A., Istel, Ł., et al. (2018). Notes
526 for genera: basal clades of Fungi (including Aphelidiomycota, Basidiobolomycota,
527 Blastocladiomycota, Calcarisporiellomycota, Caulochytriomycota, Chytridiomycota,
528 Entomophthoromycota, Glomeromycota, Kickxellomycota, Monoblepharomycota,
529 Mortierellomyc. Fungal Divers *92*, 43–129.
530 39. Doweld, A.B. (2014). Nomenclatural novelties. Index Fungorum *49*.
531 40. Ustinova, I., Krienitz, L., and Huss, V.A.R. (2000). Hyaloraphidium curvatum is not a
532 green alga, but a lower fungus; Amoebidium parasiticum is not a fungus, but a member
533 of the DRIPs. Protist *151*, 253–262.
534 41. Powell, M.J. (2017). Blastocladiomycota. In Handbook of the Protists, J. M. Archibald,
535 A. G. B. Simpson, and C. H. Slamovits, eds. (Springer International Publishing), pp.
536 1497–1521.
537 42. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial
538 time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics
539 *19*, 153.
540 43. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local
541 alignment search tool. J Mol Biol *215*, 403–410.
542 44. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E.M.

543 (2015). BUSCO: assessing genome assembly and annotation completeness with single-
544 copy orthologs. Bioinformatics *31*, 3210–3212.
545 45. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering
546 the next-generation sequencing data. Bioinformatics *28*, 3150–3152.
547 46. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum
548 evolution trees with profiles instead of a distance matrix. Mol Biol Evol *26*, 1641–1650.
549 47. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a
550 fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.
551 Mol Biol Evol *32*, 268–274.
552 48. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software
553 version 7: improvements in performance and usability. Mol Biol Evol *30*, 772–780.
554 49. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate
555 Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.
556 50. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). Phylobayes mpi:
557 phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment.
558 Syst Biol *62*, 611–615.
559 51. Roure, B., Rodriguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: a tool for selection,
560 concatenation and fusion of sequences for phylogenomics. BMC Evol Biol *7*.
561 52. Joshi, N.A., and Fass, J.N. (2011). Sickle: a sliding-window, adaptive, quality-based
562 trimming tool for FastQ files (Version 1.33) [Software].
563 https://github.com/najoshi/sickle.
564 53. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S.,
565 Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new
566 genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol
567 *19*, 455–477.
568 54. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J.,

569 Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence
570 reconstruction from RNA-seq using the Trinity platform for reference generation and
571 analysis. Nat Protoc *8*, 1494–1512.
572 55. Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for
573 automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*,
574 1972–1973.
575 56. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for
576 Illumina sequence data. Bioinformatics *30*, 2114–2120.
577 57. Bennett, L., Melchers, B., and Proppe, B. (2020). Curta: a general-purpose high-
578 performance computer at Zedat, Freie Universität Berlin.
579 http://dx.doi.org/10.17169/refubium-26754.
580 58. Richter, D.J., Berney, C., Strassert, J.F.H., Poh, Y.-P., Herman, E.K., Muñoz-Gómez,
581 S.A., Wideman, J.G., Burki, F., and de Vargas, C. (2020). EukProt: a database of
582 genome-scale predicted proteins across the diversity of eukaryotes. bioRxiv
583 2020.06.30.180687; https://doi.org/10.1101/2020.06.30.180687.
584 59. Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018).
585 UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol *35*, 518–522.
586 60. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection.
587 Syst Biol *51*, 492–508.
588 61. Sayyari, E., and Mirarab, S. (2016). Fast coalescent-based computation of local branch
589 support from quartet frequencies. Mol Biol Evol *33*, 1654–1668.
590


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.("Parvularia atlantis")
Fonticula alba
Mus musculus
Proterospongia sp.
Capsaspora owczarzaki
misidentified Pessonella sp.
Dictyostelium lacteum
Cunea sp.
Stereomyxa ramosa
0.68
0.65
1.0
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
Chytridiomyceta
Microsporidia
Rozellomycota
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
0.52
0.39
0.85
0.97
0.6
0.91
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