

A stable backbone for the fungi

Ingo Ebersberger¹, Matthias Gube², Sascha Strauss¹, Anne Kupczok¹, Martin Eckart^{2,3}, Kerstin Voigt^{2,3}, Erika Kothe², and Arndt von Haeseler¹

¹ Center for Integrative Bioinformatics Vienna (CIBIV), University of Vienna, Medical University of Vienna, University of Veterinary Medicine Vienna, Austria

² Friedrich Schiller University, Institute for Microbiology, Jena, Germany

³ current address: Fungal Reference Centre, Jena, Germany

Fungi are abundant in the biosphere. They have fascinated mankind as far as written history goes and have considerably influenced our culture. In biotechnology, cell biology, genetics, and life sciences in general fungi constitute relevant model organisms. Once the phylogenetic relationships of fungi are stably resolved individual results from fungal research can be combined into a holistic picture of biology. However, and despite recent progress¹⁻³, the backbone of the fungal phylogeny is not yet fully resolved. Especially the early evolutionary history of fungi⁴⁻⁶ and the order or below-order relationships within the ascomycetes remain uncertain. Here we present the first phylogenomic study for a eukaryotic kingdom that merges all publicly available fungal genomes and expressed sequence tags (EST) to build a data set comprising 128 genes and 146 taxa. The resulting tree provides a stable phylogenetic backbone for the fungi. Moreover, we present the first formal supertree based on 161 fungal taxa and 128 gene trees. The combined evidences from the trees support the deep-level stability of the fungal groups towards a comprehensive natural system of the fungi. They indicate that the classification of the fungi, especially their alliance with the Microsporidia, requires careful revision. Our analysis is also an inventory of present day sequence information for the fungi. It provides insights into which phylogenetic conclusions can and which cannot be drawn from the current data and may serve as a guide to direct further sequencing initiatives. Together with a comprehensive animal phylogeny⁷, we provide the second of three pillars to understand the evolution of the multicellular eukaryotic kingdoms, fungi, metazoa, and plants, in the past 1.6 billion years⁸.

Molecular data have proven useful to complement, and sometimes overrule, morphological evidences in attempts to re-classify the fungi³. However, the evolutionary backbone of the fungi could not yet be resolved with confidence. This is mainly due to a limited gene sampling² bearing the possibility of a biased view on evolutionary relationships⁹. Especially the widely used rDNA and its spacers are problematic¹⁰⁻¹³ since their complex mode of evolution is not sufficiently taken into account by the current models. Alternatively, a larger set of genes had been used, but the analyses then were confined to few taxa with sequenced genomes^{14, 15}. This substantially increased branch support values, however it was to the cost of bearing the risk of misleading conclusions on phylogenetic relationships due to insufficient taxon sampling¹⁶. Recently, EST data were proven useful for phylogenetic studies^{7, 17, 18}. This wealth of data has only been recently tapped for fungi¹⁹ and bears tremendous potential for the resolution of unstable fungal branches.

To arrive at a stable and refined phylogeny for the fungi, we maximized taxon and gene sampling by merging data from 63 completely sequenced fungi and 104 fungal EST projects. We screened these sequences for presence of orthologs to 1,035 evolutionary conserved protein coding nuclear genes with well-supported orthology from animals to fungi. 128 genes (Supplementary Table 2) and 146 taxa resulted in a data matrix with only 33% missing data. From the resulting concatenated multiple sequence alignment (supermatrix) a maximum likelihood (ML) tree and a Bayesian tree was inferred (Fig. 1, Fig. 2, Supplementary Fig. 1).

Most branches show a very high statistical support with a mean of 95% for the bootstrap probabilities (BP) and 0.98 for the Bayesian posterior probabilities (BPP). Only in four cases, at the base of Ophiostomatales-Sordariales-Diaporthales, the Dikarya-Mucoromycotina-Glomeromycota clade, within Hypocreaceae, and at the

base of the Agaricomycetes the branching pattern remained unresolved in the ML tree. Bayesian tree inference resolved all but the basal agaricomycete phylogeny. In a second approach, we computed a supertree based on 128 gene trees (Supplementary Fig. 2). With the exception of the Agaricomycetes and Hypocreaceae all polytomies of the ML tree were resolved. The combined evidence from three analyses allows sound conclusions on the evolutionary relationships of the major fungal clades (Fig. 1).

The trees support the monophyly of fungi, metazoa and green plants as well as the monophyly of the opisthokonts. The Microsporidia, however, currently classified as fungi^{2,4}, are nested within the Mycetozoa and Amoebozoa. A number of biological features specific to fungi, *e.g.* chitinous cell walls, and hyphal or yeast growth forms, are not seen with the Microsporidia. Since these organisms are obligate parasites, fungal characteristics could have been secondarily lost during their adaptation to parasitism. However, a placement of the Microsporidia within fungi is not supported by any of our phylogenomic analyses, and long branch attraction artifacts most likely do not play a role. This supports the earlier view that Microsporidia are not derived fungi but protozoa²⁰ for which secondary loss of fungal characteristics has not to be postulated (see supplementary online information for further discussion).

Within the fungi, the monophyletic group of neocallimastigomycetes, blastocladiomycetes and chytridiomycetes (BP: 100, BPP: 0.98) split first from the backbone. This puts a new complexion of the early evolution of fungi. We conclude that the Neocallimastigomycota and the Blastocladiomycota^{2,21} have to be withdrawn as distinct phyla and subsumed as subphyla (suffixed with -mycotina) within the Chytridiomycota.

The Entomophthoromycotina³ are well separated from the earlier branching

Chytridiomycota and are placed outside of the remaining fungi in the ML and Bayesian trees (BP: 68, BPP: 0.98). The Glomeromycota split next from the fungal backbone in the Bayesian tree (BPP: 1). The extended taxon sampling in our complementary MRP supertree shows the Entomophthoromycotina and Glomeromycota each as monophyletic. However, the current data do not allow to confidently attach them to the phylogenetic backbone of the fungi. Whether or not the Glomeromycota have to be included into the zygomycetes or are a separate phylum, as suggested by the MRP-supertree and the Bayesian analysis, remains open. The ongoing *Glomus* genome sequencing initiative²² will help to elucidate this point. The monophyletic Mucoromycotina are the sister group of the Dikarya. Within the well-supported Dikarya consisting of Basidiomycota and Ascomycota, the latter subdivide into three sub-phyla: Taphrinomycotina, Saccharomycotina and Pezizomycotina. The Taphrinomycotina split off first as a monophyletic clade in the ML and Bayesian analysis. The monophyletic *Saccharomyces* complex has experienced difficulties in morphological classification in the past. Our data suggest a revision of that group (Fig. 2).

Within the Pezizomycotina, all classes are monophyletic and their phylogeny is well resolved. ML/Supertree and Bayesian tree disagree only in the phylogenetic position of the Dothideomycetes. The supertree analysis indicates that the majority of genes lend independent support for the placement of the Dothideomycetes as sister to the Sordariomycetes/Leothiomycetes. Finally, we note that four taxa with unclear systematic position were confidently placed in our trees: *Thermomyces* groups within the order Eurotiales, *Glomerella spec.* and *Verticillium dahliae* representing the Phyllacorales are placed as sister to the Hypocreales, and *Amorphotheca resinae* is associated with the Leotiomyces. A fifth taxon, *Geomyces pannorum*, that has been

originally described to belong to the Onygenales, is now placed within the Leotiomyces. Some of these placements have been seen individually in studies restricted to individual parts of the ascomycete tree²³⁻²⁵, but so far had not been generally recognized. Based on our analysis, a re-classification of the five species is proposed. The supertree provides further hints on the systematic position of individual taxa. For example, *Oidiodendron maius*, formerly described as *Ascomycota incerta sedis*, was placed within the Leotiomyces. This fungus forms ericoid mycorrhizae like the leotiomyces genus *Hymenoscyphus*. The similar lifestyle and morphology lends further biological support to our placement of *Oidiodendron* within the Leotiomyces.

The Basidiomycota clade shows the rust fungi (Pucciniomycotina) and the Agaricomycotina each as a monophyletic group. The placement of the smut fungi (Ustilaginomycotina) could only be solved with confidence by the Bayesian analysis that indicates the smuts as sister taxon to the Agaricomycotina (BPP: 1). The Agaricomycotina comprise the well supported Tremellomyces and Agaricomycetes. At the base of the Agaricomycetes the branch lengths are short and the phylogenetic signal is not sufficient to allow a resolution of the branching pattern with the present data. A re-computation of the Basidiomycota subtree with an adapted data set gave no further improvement (Supplementary Fig. 2).

The difficulties in resolving the early splits of the basidiomycete phylogeny suggest that this part of the fungal tree is more bush-like, *i.e.*, the corresponding speciation events occurred in close succession. Exemplified for the Ustilaginomycotina (Supplementary discussion) we show that increasing the number of genes for the maximum likelihood phylogeny reconstruction to more than 1,200 resolves their position as sister taxon to the Agaricomycotina. However, a closer look

reveals a picture that resembles a problematic case in the animal phylogeny, the phylogenetic position of *Caenorhabditis elegans*. This particular taxon serves as a paragon of how limited taxon sampling paired with long external branches can result in statistically excellently supported – but presumably wrong – phylogenetic conclusions¹⁶. A definite phylogenetic placement of the Ustilaginomycotina must, therefore, await a better taxon sampling of this fungal sub-phylum (see supplementary information). Presumably the same applies to two more fungal phyla/sub-phyla that are currently represented only by a single taxon (Fig. 1).

Our taxon sampling is biased towards species with whole genome sequences or large EST sets available. This causes the scarce presence or even absence of some groups that are currently not considered commercial, medical, or scientific models. We encourage the fungal community to start EST sequencing projects for taxa that have been ignored so far, but are representative for missing and highly unique clades, e.g. the ascomycete *Neolecta*, the lichen *Lecanora* or the zygomycetes *Endogone* and *Kickxella*.

The biology of fungi is full of complexities. Classifications of the Mycota based on morphological characters have suffered, for instance, from the problem to assign sexual and asexual stages of a fungus to one species and from convergent evolution. Exemplified for the Ascomycota we analyze the evolution of the spore dispersal machinery as a phylogenetic informative morphological character complex. The presence of fruiting bodies is a derived character within the Ascomycota with Taphrinomycotina and Saccharomycotina lacking any ascomatal structures. Within the Pezizomycotina, fruiting body types are polymorphic (Fig. 3). Of the fruiting bodies, apothecial forms are found to be basal. Perithecia, cleistothecia, and also pseudothecia are therefore derived character states. The independent occurrence of

cleistothecia and pseudothecia is a result of convergent evolution. Similar difficulties arise with other characteristics of the spore release machinery limiting its usefulness in elucidating the evolutionary relationships among the individual classes of the Pezizomycotina (see supplementary online information for further discussion).

The stable phylogenetic backbone represents a major advance towards resolving the evolutionary history of fungi. It comprises the fundament to build the multiple, fascinating scenarios necessary to advance knowledge for applied purposes, e.g. to forecast fungal groups with high potential of natural compounds or to raise production levels in biotechnologically important fungi depending on similar regulatory mechanisms conserved in evolution. A well resolved phylogeny of the fungi will provide insight into the evolution of their peculiar features, e.g. fruiting body development, ecological impact, or even allow new insights into the evolution of multicellularity.

METHODS SUMMARY

All available (as of July 2008) Expressed Sequence Tags (ESTs) from fungi and annotated gene sets from all fungal genome sequences were downloaded from the public databases (Supplementary Table 1). Overlapping ESTs from the same taxon were clustered into contigs. Two sets of evolutionary conserved genes (core-orthologs) were identified from two selections of completely sequenced fungal and metazoan genomes. The sequences in each core-ortholog cluster were aligned and converted into a Hidden Markov Model. The core-ortholog cluster were then extended with sequences from further taxa using a combination of a Hidden Markov Model based search followed by a reciprocal BLAST search (HaMStR). Ortholog cluster were then individually aligned with MAFFT²⁶. Phylogenetic trees were computed from the concatenated alignments (supermatrix) with RaXML²⁷ and with PhyloBayes²⁸. Alternatively, gene trees were computed from the individual alignments with RaXML and were used for Matrix Representation with Parsimony supertree reconstruction^{29, 30}.

Figure 1 The phylogenetic backbone of the fungi

The backbone of the phylogeny as inferred from two supermatrix approaches (maximum likelihood, Bayesian) and a supertree approach. Triangles denote clades represented by at least two taxa in the supermatrix analyses (size not drawn to scale). Branch support is given as bootstrap probability (supermatrix)/Bayesian posterior probability (supermatrix)/bootstrap probability (supertree). * denotes 100% support, - denotes 'not resolved', ! denotes different branching pattern, *n.a.* denotes clades that are represented by a single taxon in the supermatrix approaches. Branch lengths are taken from the Bayesian tree, dashed branches are not drawn to scale.

Figure 2 The phylogeny of the fungal kingdom

The deep-level maximum likelihood phylogeny of the fungi. Branch support values represent bootstrap probabilities where * denotes 100% support. Branches with a higher line weight have at least 95% Bayesian posterior probability. Names of taxa *incerta sedis* are written in purple, names of taxa positioned different to their systematic description are written in orange.

Figure 3 Distribution of fruiting body types in the Pezizomycotina.

Subtree of the maximum likelihood tree shown in Figure 2 with clades collapsed on the order level. *Amorphotheca resiniae* is not associated to an order. *Geomyces pannorum* is described as *Onygenales* but requires re-classification. Fruiting body morphology is given next to the taxon names. A = Apothecia, C = Cleistothecia, Ps = Pseudothecia, P = Perithecia, *unkn** = unknown

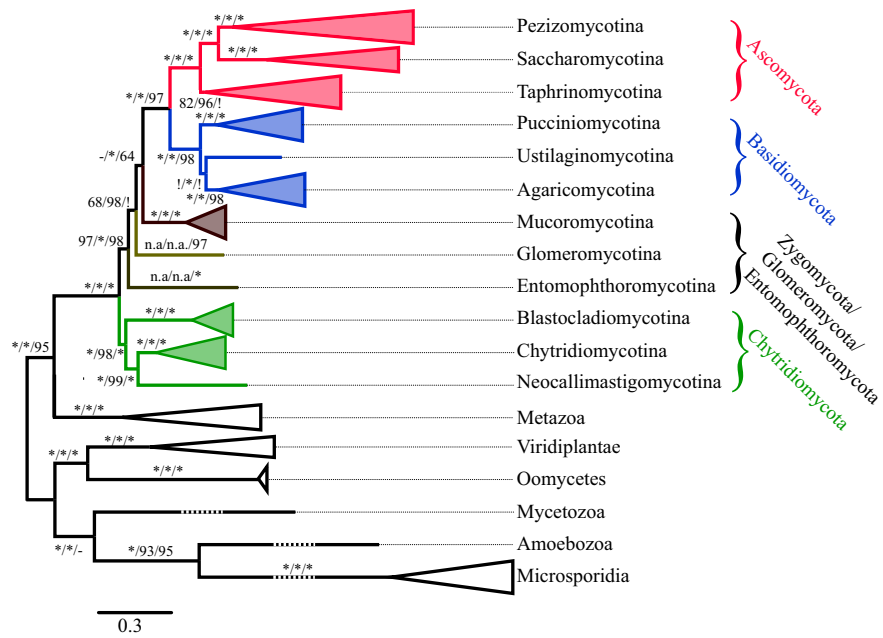
1. Lutzoni, F. et al. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *American Journal of Botany* **91**, 1446-1480 (2004).
2. James, T. Y. et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818-22 (2006).
3. Hibbett, D. S. et al. A higher-level phylogenetic classification of the Fungi. *Mycol Res* **111**, 509-47 (2007).
4. 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).
5. Schüßler, A., Schwarzott, D. & Walker, C. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol Res* **105**, 1413-1421 (2001).
6. Idnurm, A., Walton, F. J., Floyd, A. & Heitman, J. Identification of the sex genes in an early diverged fungus. *Nature* **451**, 193-6 (2008).
7. Dunn, C. W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745-9 (2008).
8. Hedges, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol* **4**, 2 (2004).
9. Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798-804 (2003).
10. Alvarez, I. & Wendel, J. F. Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* **29**, 417-34 (2003).
11. Mayol, M. & Rossello, J. A. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Mol Phylogenet Evol* **19**, 167-76 (2001).
12. Ko, K. S. & Jung, H. S. Three nonorthologous ITS1 types are present in a polypore fungus *Trichaptum abietinum*. *Mol Phylogenet Evol* **23**, 112-22 (2002).
13. Spatafora, J. W. et al. A five-gene phylogeny of Pezizomycotina. *Mycologia* **98**, 1018-28 (2006).
14. Robbertse, B., Reeves, J. B., Schoch, C. L. & Spatafora, J. W. A phylogenomic analysis of the Ascomycota. *Fungal Genet Biol* **43**, 715-25 (2006).
15. Cornell, M. J. et al. Comparative genome analysis across a kingdom of

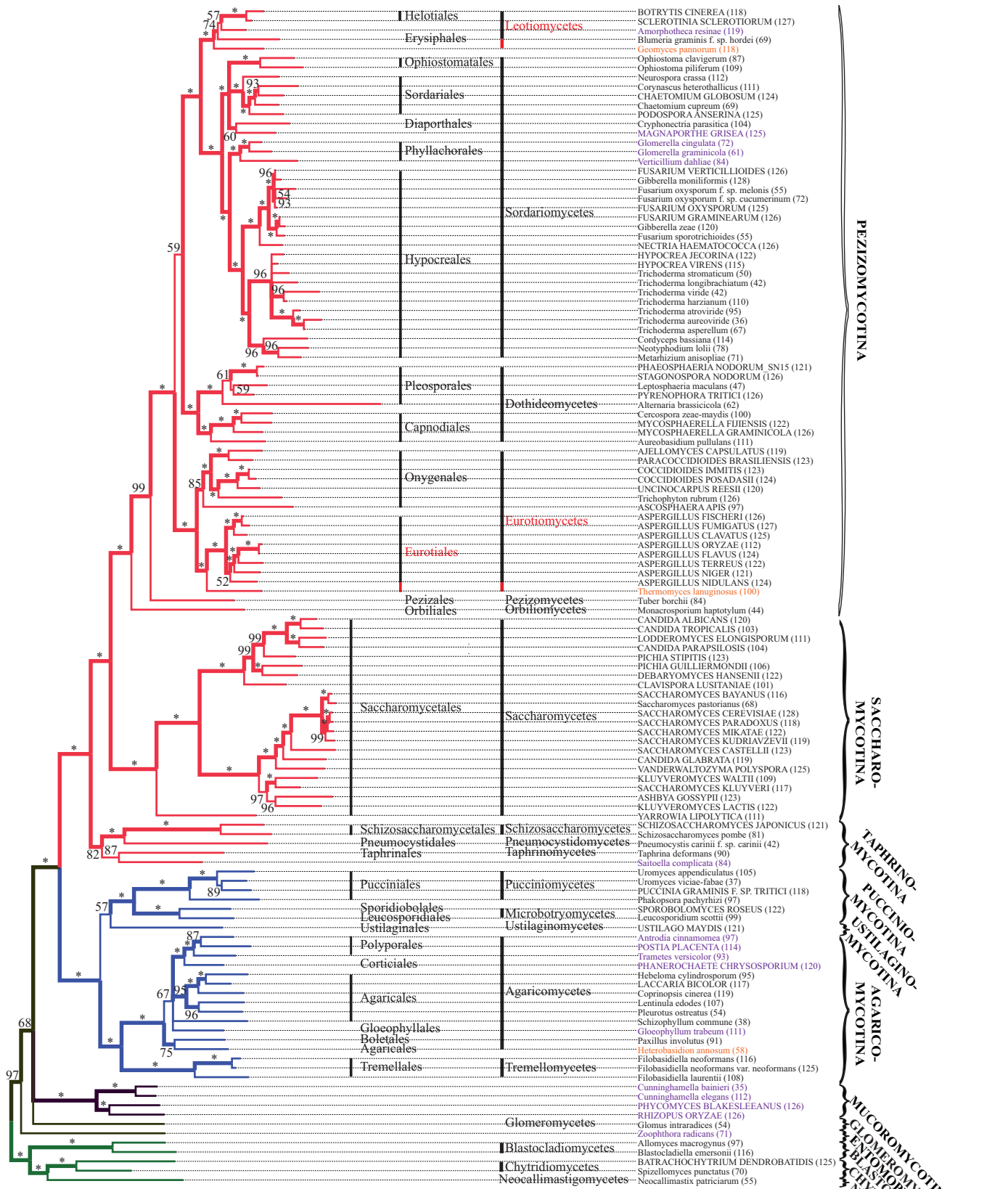
- eukaryotic organisms: specialization and diversification in the fungi. *Genome Res* **17**, 1809-22 (2007).
16. Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* **22**, 1246-53 (2005).
 17. Roeding, F. et al. EST sequencing of Onychophora and phylogenomic analysis of Metazoa. *Mol Phylogenet Evol* **45**, 942-51 (2007).
 18. Hughes, J. et al. Dense taxonomic EST sampling and its applications for molecular systematics of the Coleoptera (beetles). *Mol Biol Evol* **23**, 268-78 (2006).
 19. Liu, Y. et al. Phylogenomic analyses support the monophyly of Taphrinomycotina, including Schizosaccharomyces fission yeasts. *Mol Biol Evol* **26**, 27-34 (2009).
 20. Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326**, 411-4 (1987).
 21. James, T. Y. et al. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98**, 860-71 (2006).
 22. <http://www.jgi.doe.gov/genome-projects>.
 23. Wang, Z., Johnston, P. R., Takamatsu, S., Spatafora, J. W. & Hibbett, D. S. Toward a phylogenetic classification of the leotiomycetes based on rDNA data. *Mycologia* **98**, 1065-75 (2006).
 24. Hambleton, S., Nickerson, N. L. & Seifert, K. A. *Leohumicola*, a new genus of heat-resistant hyphomycetes. *Studies in Mycology* **53**, 29-52 (2005).
 25. Zhang, N. et al. An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. *Mycologia* **98**, 1076-87 (2006).
 26. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**, 511-8 (2005).
 27. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-90 (2006).
 28. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**, 1095-109 (2004).
 29. Baum, B. R. Combining trees as a way of combining data sets phylogenetic inference, and the desirability of combining gene trees. *Taxon* **41**, 3-10 (1992).

30. Ragan, M. A. Phylogenetic inference based on matrix representation of trees. *Mol Phylogenet Evol* **1**, 53-8 (1992).

Acknowledgements We thank M. Sogin and H. Morrison from the Josephine Bay Paul Center for generously providing us with data from *A. locustae*. I.E. and A.v.H acknowledge financial support by the Wiener Wissenschafts-, Forschungs- und Technologie Fonds (WWTF), and from the DFG priority program SPP 1174 Deep Metazoan Phylogeny. E.K and K.V. acknowledge financial support by the DFG.

Author information All sequence data as well as the alignments are available at <http://www.deep-phylogeny.org/fungi>. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. Correspondence should be addressed to I.E. (ingo.ebersberger@univie.ac.at).





0.1

