

Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes

Thomas A. Richards^{a,b,1}, Darren M. Soanes^a, Meredith D. M. Jones^{a,b}, Olga Vasieva^c, Guy Leonard^{a,b}, Konrad Paszkiewicz^a, Peter G. Foster^b, Neil Hall^c, and Nicholas J. Talbot^a

^aBiosciences, University of Exeter, Exeter EX4 4QD, United Kingdom; ^bDepartment of Zoology, Natural History Museum, London SW7 5BD, United Kingdom; and ^cSchool of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, United Kingdom

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved July 27, 2011 (received for review March 31, 2011)

Horizontal gene transfer (HGT) can radically alter the genomes of microorganisms, providing the capacity to adapt to new lifestyles, environments, and hosts. However, the extent of HGT between eukaryotes is unclear. Using whole-genome, gene-by-gene phylogenetic analysis we demonstrate an extensive pattern of cross-kingdom HGT between fungi and oomycetes. Comparative genomics, including the de novo genome sequence of *Hyphochytrium catenoides*, a free-living sister of the oomycetes, shows that these transfers largely converge within the radiation of oomycetes that colonize plant tissues. The repertoire of HGTs includes a large number of putatively secreted proteins; for example, 7.6% of the secreted proteome of the sudden oak death parasite *Phytophthora ramorum* has been acquired from fungi by HGT. Transfers include gene products with the capacity to break down plant cell walls and acquire sugars, nucleic acids, nitrogen, and phosphate sources from the environment. Predicted HGTs also include proteins implicated in resisting plant defense mechanisms and effector proteins for attacking plant cells. These data are consistent with the hypothesis that some oomycetes became successful plant parasites by multiple acquisitions of genes from fungi.

osmotrophy | pseudofungi | lateral gene transfer

Horizontal gene transfer (HGT) involves the transfer of genetic material between reproductively isolated lineages (1, 2) and has been an important factor in prokaryotic evolution (3–5). By contrast, the role of HGT between eukaryotic genomes is less clear (2, 6) and seems to have occurred at a lower frequency, with some reports suggesting that HGT between eukaryotic kingdoms is rare (7). Conversely, HGTs originating from prokaryotic genomes constitute a significant, but relatively small, proportion of the total inventory of genes in phagotrophic protists (8–12) and some osmotrophic fungi (13).

Fungi branch with animals on the tree of life (14) and encompass an extremely diverse group of organisms, including species adapted to forming intimate relationships with plants, including mutualistic and parasitic associations, as well as saprophytic growth (15–17). To explore the frequency of HGT between distinct eukaryotic groups that form parasitic associations with plants, we previously used the predicted protein-encoding sequences from the genome of a fungal plant parasite, *Magnaporthe oryzae* (18), as a BLASTp search seed and identified four HGTs from fungi to oomycetes with strong phylogenetic support (19). The oomycetes are distant relatives of fungi and branch within the stramenopile radiation, which includes a diversity of photosynthetic microbes possessing plastid organelles of secondary endosymbiotic ancestry (20). Oomycetes, however, are not photosynthetic and display filamentous growth, closely resembling fungi in many aspects of their biology. Both fungi and oomycetes, for instance, feed exclusively by osmotrophy, secreting depolymerizing enzymes to break down complex biological materials in the extracellular environment, followed by transport of broken-down metabolic units into the cell. Fungi and oomycetes also cause many of the world's most serious plant diseases. Sudden oak death is caused by the oomycete *Phytophthora*

ramorum, for example, whereas the Irish potato famine of the 19th century was caused by the late blight parasite *Phytophthora infestans*. Important crop diseases caused by fungi include the devastating rice blast disease caused by *M. oryzae* and the rusts, smuts, and mildews that affect wheat, barley, and maize. In this study we report that HGT between fungi and oomycetes has occurred to a far greater degree than hitherto recognized (19). Our previous analysis suggested four strongly supported cases of HGT, but by using a whole-genome, gene-by-gene phylogenetic analysis we now reveal a pattern of 34 transfers and propose that these transfers have been fundamental to the evolution of plant parasitic traits within the oomycetes.

Results and Discussion

Identifying and Testing the Pattern of HGT Between Fungi and Oomycetes. Among the best methods for identifying HGT is to identify a gene phylogeny that places a taxonomic group (recipient) within the branches of a distantly related group (donor) in direct contradiction to the known phylogenetic relationships of the respective taxa (2, 6). To identify all potential gene transfers between fungi and oomycetes, we selected the predicted proteomes of the oomycete species *Phytophthora ramorum*, *Phytophthora sojae*, *Phytophthora infestans*, and *Hyaloperonospora parasitica* (also named *Hyaloperonospora arabidopsidis*) (21–23). We processed each proteome separately, first excluding all putative transposable elements (*SI Materials and Methods*). Then we clustered the genes from each genome into closely related cluster groups (24) (*SI Materials and Methods*). Cluster groups that BLASTp demonstrated were exclusively found only in the oomycete genomes were then removed (*SI Materials and Methods*). This process left 3,014, 3,018, 3,233, and 2,169 cluster groups from *P. ramorum*, *P. sojae*, *P. infestans*, and *H. parasitica*, respectively, totaling 11,434 gene clusters (Table 1). We then used a gene-by-gene phylogeny pipeline (7) to generate fast maximum-likelihood trees for all 11,434 gene family groups, using a database of 795 (173 eukaryotic and 622 prokaryotic) genomes (Table S1). These data were searched manually for trees that demonstrated putative fungi/oomycete gene transfer. Gene families that produced unresolved tree topologies were reprocessed by running the tree-building pipeline again but using different gathering thresholds (*SI Materials and Methods*).

Author contributions: T.A.R., N.H., and N.J.T. designed research; T.A.R., D.M.S., M.D.M.J., O.V., and N.H. performed research; P.G.F. contributed new reagents/analytic tools; T.A.R., D.M.S., O.V., G.L., K.P., P.G.F., and N.H. analyzed data; and T.A.R., N.H., and N.J.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The genome sequence data reported in this paper (*Hyphochytrium*) have been submitted to Sequence Read Archive for the raw data (SRP004821), and the assembled genome has been archived at the European Bioinformatics Institute (Genome Project: 61035).

¹To whom correspondence should be addressed. E-mail: thomr@nhm.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1105100108/-DCSupplemental.

Table 1. Comparative gene-by-gene phylogenomic analysis and identified HGTs

Species	<i>P. ramorum</i>	<i>P. sojae</i>	<i>P. infestans</i>	<i>H. parasitica</i>
Total no. of predicted proteins	15,743	19,027	22,658	13,240
Total proteins used in phylogenetic analysis	6,871	6,079	7,093	3,365
No. of gene cluster groups used for phylogenetic analyses [after OrthoMCL (24) clustering]	3,014	3,018	3,233	2,169
No. of HGT gene families strongly supported by all methods	19	17	11	4
No. of "gray zone" HGT gene families (Table 2 and Table S3)	11	10	7	5
Total putative ORFs from fungal-derived HGTs	143	117	48	21
Total secreted proteins in genome identified by both WoLFPSORT and SignalP	1,326	1,586	1,123	411
Total no. of proteins analyzed with phylogeny predicted to be secreted	521	630	351	121
No. of fungal HGT-derived proteins with N-terminal secretion motif	101	89	30	13
% Total predicted secretome derived from fungi by HGT	7.61	5.61	2.67	3.16
% Proteins analyzed that are predicted to be secreted and are derived by HGT	19.38	14.13	8.55	10.74

All of the predicted fungi/oomycete HGTs were subject to reevaluation of taxon sampling using additional database searches and manual alignment improvement, followed by recalculation of the phylogenetic tree, combined with bootstrap analyses (see *SI Materials and Methods* and Table S2 for details of phylogenetic analysis). These analyses identified multiple cases in which oomycete genes (the recipient group) branched within a clade of fungal genes (the donor group) and a single case demonstrating the opposite relationship. To confirm the results of these phylogenetic analyses, we used alternative topology tests to test whether it was possible to reject monophyly of the donor group. Where taxon sampling allowed, we used a variety of different topology constraints, corresponding to different relationships within the fungi (15, 25). This approach enabled us to partially polarize the ancestry of the HGT event relative to the donor group (Table S3). Taken together, our analyses identified 20 gene families predicted to have been transferred from within the fungi to the oomycetes and one case in which the transfer occurred from the oomycetes to the fungi (Fig. 1 and Figs. S1.1–S1.21).

Our pipeline analysis also identified four gene families that were exclusively found in oomycete and fungal genomes and not present in any additional taxonomic groups analyzed. We checked to confirm that the taxon sampling was robust by comparison with the GenBank nr database, using both BLASTp and psi-BLAST with five iterations (26), and by interrogating the GenBank EST and the Taxonomically Broad EST database using tBLASTn (27). Because of the taxon distribution it was not possible to perform phylogenetic analysis to test an HGT hypothesis directly, because the taxon sampling included only oomycetes and fungi. However, punctate taxon distribution of

a gene family can be used to suggest HGT because alternative explanations involving ancient acquisition of a gene in the last common ancestor, coupled to gene loss, are less parsimonious (see ref. 6 for description of different HGT scenarios). On the basis of these criteria, we have included the additional four gene families among the putative fungi–oomycete HGTs (Table 2 and Table S3).

Our analysis also identified an additional nine gene families that demonstrated phylogenetic support for the oomycete gene branching within the fungi to the exclusion of all other taxa, but the approximately unbiased (AU) alternative topology tests proved inconclusive, suggesting that the case for HGT is not definitive given available data (Table S3). In four of these cases (Figs. S1.22–S1.24 and S1.30) our analyses demonstrated that the gene family was present in a mosaic scattering of prokaryote, fungi, and oomycete taxa, with the oomycetes branching within the fungal clade but with weak bootstrap support. These data sets implicate HGT (on the basis of taxonomic distribution of the gene family with phylogenetic analyses suggesting fungi-to-oomycete HGT), but alternative topology tests proved inconclusive. The remaining five gene families also suggested fungi-to-oomycete HGT, placing the oomycete gene within the fungi, separate from other eukaryotic taxa, with moderate to strong bootstrap support but with alternative topology tests again proving inconclusive (Figs. S1.25–S1.29 and Table S3). We therefore find strong evidence for 21 gene transfers and tentative evidence for a further 13 transfers, suggesting an important and pervasive pattern of HGT between these distantly related groups (Table S3).

Theoretically, the presence of conserved introns in both recipient and donor taxa would suggest that any HGT is the

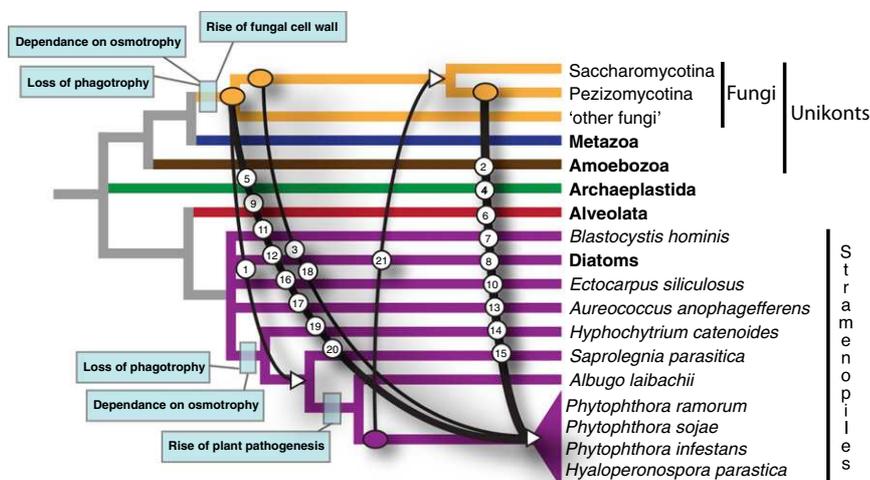


Fig. 1. Pattern of HGTs between fungi and oomycetes, demonstrating that the majority of the fungal-derived gene transfers are into/or retained by the plant parasitic oomycetes. Using the results of the phylogenetic analysis in combination with alternative topology tests, it was possible to estimate the earliest point of transfer for each of the 21 strongly supported HGTs (Table S3). By comparing the taxonomic distribution of the HGTs it was then possible to identify the putative primary point of acquisition. Incomplete phylogenetic resolution and incomplete taxon sampling may cause these estimates to misplace the HGT events. We also note that the figure is based on a hypothetical cladogram and does not identify the pattern of transfer relative to either phylogenetic distance or time. Additional genome sampling will enable improved resolution of these transfers. The number labels on each transfer event refer to phylogenetic data in Figs. S1.1–S1.21. Major events in cell evolution are marked to polarize HGTs in relation to evolutionary history of these microbes.

Table 2. Summary of support for each HGT and gene annotation in brief

HGT ID 1–34	GenBank accession no./Joint Genome Institute database protein ID	Support for HGT: phylogeny (P), bootstrap analysis (B), topology comparison tests (Ct), and/or taxon distribution (Td)	Conserved introns between donors and recipients (Table S4)	Annotation of putative function	Reported by Richards et al. (19)
1 (Fig. S1.1)	EEY57756	P, B, Ct, Td	1 shared intron	MFS transporters similar to saccharide monomer transporters	Yes (named AraJ)
2 (Fig. S1.2)	82760 (<i>P. ramorum</i>)	P, B, Ct		Extracellular esterase/lipase	Yes (named esterase/lipase (ref. 19, Fig. S3B))
3 (Fig. S1.3)	AAM48174.1	P, B, Ct		Aldose 1-epimerase	Yes (named GalM)
4 (Fig. S1.4)	71178 (<i>P. ramorum</i>)	P, B, Ct, Td		α -ketoglutarate dependent xanthine dioxygenase (XanA)	
5 (Fig. S1.5)	EEY56552	P, B, Ct		Dehydrogenase/reductase family protein	
6 (Fig. S1.6)	72257 (<i>P. ramorum</i>)	P, B, Ct, Td		Extracellular quercetin 2, 3-dioxygenase	
7 (Fig. S1.7)	EEY52979	P, B, AU test is borderline		Extracellular α -L-rhamnosidase (glycosyl hydrolase 78)	
8 (Fig. S1.8)	EEY63463	P, B, Ct		Lactonohydrolase/gluconolactonase	
9 (Fig. S1.9)	EEY64355	P, B, Ct		Extracellular glucooligosaccharide oxidase	
10 (Fig. S1.10)	EEY59160	P, B, Ct		Extracellular unsaturated rhamnogalacturonyl hydrolase (glycosyl hydrolase 88)	
11 (Fig. S1.11)	83543 (<i>P. ramorum</i>)	P, B, Ct		Transcription factor	
12 (Fig. S1.12)	85044 (<i>P. ramorum</i>)	P, B, Ct, Td	1 shared intron	3-octaprenyl-4-hydroxybenzoate carboxy-lyase/Phenylphosphate carboxylase family	
13 (Fig. S1.13)	EEY53137	P, B, Ct	1 shared intron	Phosphatidylinositol transfer protein	
14 (Fig. S1.14)	133521 (<i>P. sojae</i>)	P, B, Ct		Extracellular protein similar to prokaryote lipases	
15 (Fig. S1.15)	142730 (<i>P. sojae</i>)	P, B, Ct	1 shared intron	Esterase/lipase protein domain family	
16 (Fig. S1.16)	EEY68514	P, B, Ct		Xylitol dehydrogenase/sorbitol dehydrogenase	
17 (Fig. S1.17)	EEY55544	P, B, Ct, Td	2 shared introns	Extracellular arabinan-endo-1, 5- α -L-arabinosidase (glycosyl hydrolase 43)	
18 (Fig. S1.18)	EEY59913	P, B, Ct, Td		Transporter for purines and pyrimidines	Yes (named CodB)
19 (Fig. S1.19)	141189 (<i>P. sojae</i>)	P, B, Ct		Extracellular protein with an esterase/lipase domain	
20 (Fig. S1.20)	EEY58144	P, B, Ct, Td		Putative member of the NPP1 (necrosis and ethylene-inducing peptide 1 proteins)	
21 (Fig. S1.21)	EEY64154	P, B, Ct, Td	1 shared intron	Conserved hypothetical protein with similarity to a prokaryotic antibiotic biosynthesis monooxygenase	
22 (Fig. S1.22)	EEY62062	P, Td		Pectate lyase	
23 (Fig. S1.23)	EEY67135	P, Td		Extracellular intradiol dioxygenase	Yes (named PcaH)
24 (Fig. S1.24)	ABB22031	P, Td		Extracellular endoglucanase (glycosyl hydrolase 12)	
25 (Fig. S1.25)	EEY67552	P, B	1 shared intron	Extracellular histidine phosphatase domain protein	
26 (Fig. S1.26)	EEY65395	P, B	2 shared intron	Extracellular endo-1,4- β -xylanase (glycosyl hydrolase 10)	
27 (Fig. S1.27)	75147 (<i>P. ramorum</i>)	P, B		Member of Zinc metalloprotease proteins which include archaemetzincin	
28 (Fig. S1.28)	EEY56384	P, B		Extracellular arabinogalactan endo-1,4- β -galactosidase (glycosyl hydrolase 53)	

Table 2. Cont.

HGT ID 1–34	GenBank accession no./Joint Genome Institute database protein ID	Support for HGT: phylogeny (P), bootstrap analysis (B), topology comparison tests (Ct), and/or taxon distribution (Td)	Conserved introns between donors and recipients (Table S4)	Annotation of putative function	Reported by Richards et al. (19)
29 (Fig. S1.29)	77558 (<i>P. ramorum</i>)	P, B		Aliphatic nitrilase (CN hydrolase family)	
30 (Fig. S1.30)	EEY68947	P, Td		Extracellular α -L-rhamnosidase (glycosyl hydrolase 78)	
31	76863 (<i>P. ramorum</i>)	Td		LysM domain protein	
32	EEY55495	Td		Extracellular conserved hypothetical protein	
33	134308 (<i>P. sojae</i>)	Td		NmrA, a negative transcriptional regulator	
34	EEY58177	Td		Conserved hypothetical protein	

Sequences for *P. ramorum* and *P. sojae* are available from the Department of Energy, Joint Genome Institute website. Protein sequences can be obtained using the protein ID and searching at http://genome.jgi-psf.org/pages/search-for-genes.jsf?organism=Phyra1_1 for *P. ramorum* and http://genome.jgi-psf.org/pages/search-for-genes.jsf?organism=Physo1_1 for *P. sojae*.

product of a direct eukaryotic-to-eukaryotic transfer. To test for evidence of such a pattern, we searched all 30 HGT gene families supported by phylogenetic analysis (Figs. S1.1–S1.30) for evidence of a conserved intron present in the donor and recipient taxa. We found eight such cases (Table 2 and Table S4), providing additional data to support the hypothesis that in these cases the transfers are eukaryotic-to-eukaryotic HGT events.

Distribution of HGTs Relative to Evolution of Parasitic Traits in the Oomycetes. To provide additional data for polarizing the ancestry of these HGT events, we sequenced the genome of *Hyphochytrium catenoides* using a combination of DNA and RNA (transcribed as cDNA) sequencing using both the Roche 454 GS FLX titanium and Illumina GAIx paired-end methods (SI Materials and Methods). *H. catenoides* forms a sister branch to the oomycetes and, like the oomycetes, was originally classified as a fungus because of its fungal-like morphology and osmotrophic feeding habit (20, 28). However, molecular phylogenetics have since demonstrated that oomycetes and hyphochytriomycetes are sister groups and branch separately from fungi within the stramenopile radiation (20, 28). Analysis of this draft genome assembly revealed that all of the 34 oomycete/fungi HGTs are absent from the *Hyphochytrium* genome (SI Materials and Methods). Of the 34 putative HGTs identified, we noted that only 2 were present in the genome assembly of the oomycete fish parasite *Saprolegnia parasitica* (Figs. S1.1 and S1.21), one of which is an HGT from the oomycete lineage to the fungi (Fig. 1). To further test this pattern we compared the 34 HGTs with 13,807 proteins from the oomycete *Albugo laibachii* genome available in GenBank (29). *Albugo* forms a deeper branch sister to the *Phytophthora/Hyaloperonospora* clade and possesses a distinct repertoire of effector proteins compared with other oomycete plant pathogens (29). Comparative analysis suggests that only one of the HGTs was present in the *Albugo* assembly (HGT 34; Table 2 and Table S3). However, all 34 HGTs were present in one or more of the *Phytophthora/Hyaloperonospora* plant parasitic oomycete species. The majority of the HGTs therefore seem to have been acquired specifically within the oomycete plant parasite clade, although this result will need retesting as more oomycete genomes become available.

In 2006 Richards et al. (19) provided preliminary evidence of a pattern of HGT between fungi and oomycetes. These data were based on a BLAST survey of a single fungal genome and provided strong evidence for four gene transfers from fungi to oomycetes, with an additional four candidates for which the tree topologies could not be resolved but could potentially indicate HGT. We used these data to propose that HGT led to transfer of osmo-

trophic characteristics from fungi to oomycetes and was an important step in the evolution of the fungal-like biology of oomycetes and possibly their sisters (e.g., hyphochytridiomycetes), which together form the “Pseudofungi” (20). The data reported here were based on a comprehensive gene-by-gene phylogeny approach and confirmed five of these HGTs [including the four strongly supported HGTs reported previously (19); Table 2 and Table S5]. The remaining three HGTs, which were listed as tentative in the 2006 article, are not supported by this analysis (Table S5). Using additional genome sampling including a de novo assembly of *H. catenoides*, we were then able to test the hypothesis that fungal-to-oomycete HGTs were important for the transition from a phagotrophic algal form to an osmotrophic filamentous form deep within the oomycete/hyphochytrium clade. Contrary to our previous suggestion (19), the results of this study support an alternative hypothesis, with the pattern of HGT seeming to have occurred much later, with the majority of the HGTs specifically retained by oomycete plant parasites, a pattern consistent with the putative function of these proteins, discussed below (Fig. 2, Table 2, and Table S5). Furthermore, the pattern demonstrates that the HGTs occurred after the last common ancestor of the oomycetes and the hyphochytriomycetes had lost the capacity for phagotrophy and had instead evolved an osmotrophic/filamentous lifestyle (Fig. 1), demonstrating that the absence of phagotrophy, at least for the oomycetes, is not a barrier to HGT.

The pattern of fungal-derived HGTs implicates the gene acquisitions as being important in the evolution of plant parasitism. To test this idea, we examined the HGTs in detail. First, we found that of the 34 HGTs, 21 demonstrated evidence of gene duplication after transfer. In some cases the number of gene duplication events was extensive, such that the 33 fungi-to-oomycete HGTs contributed a total of 329 predicted genes, including 143 genes to *P. ramorum*, 117 to *P. sojae*, 48 to *P. infestans*, and 21 to *H. parasitica* (Table 1).

Putative Function of HGTs in the Oomycetes. We set out to determine the proportion of the HGT candidates that were predicted to encode secreted proteins (SI Materials and Methods). Both fungi and oomycetes secrete large numbers of proteins, including plant cell wall-degrading enzymes, attachment factors, and effector proteins to subvert host defenses. We used a combination of SignalP and WolfPSORT to identify HGT-derived genes that putatively encoded proteins with N-terminal secretion signals (Table S5). This analysis showed that between 62% and 76% of all the HGT-derived genes in each of the four oomycete genomes investigated possess putative N-terminal secretion sig-

HGTs also encompasses a putative histidine phosphatase domain protein, which putatively functions in breakdown of phytic acid and other organophosphate substrates (33). Phytic acid is the primary storage form of phosphate and inositol in plants (34). Consequently, this acquisition might provide an adaptation for uptake of phosphates from plant tissue.

Finally, the list of HGTs includes two proteins that in fungi have been directly implicated in plant parasitism. The LysM domain-containing gene family acquired by HGT from fungi has, for instance, been shown in fungal plant parasites to be linked with suppression of plant defenses. In *Cladosporium fulvum*, the causal agent of leaf mold of tomato, implicated in suppression of chitin-triggered plant immunity (35). Second, there is an example of a necrosis-inducing protein or (Nep1)-like protein (NLP) that is a candidate fungi-to-oomycete HGT (36). Infiltration of NPP1 protein into leaves of *Arabidopsis thaliana* results in transcript accumulation of plant defense-related genes, production of reactive oxygen species and ethylene, callose apposition, and localized cell death (36).

Conclusion

When considered together, our analysis demonstrates a pattern of at least 21 HGTs, and probably more than 34, between fungi and oomycetes, with the vast majority of predicted HGTs (33 in total) transferring gene functions from fungi to oomycetes. This equates to a small proportion of the total genome for which phylogenetic analysis was possible (0.5–1% of the genome analyzed; Table 1), but in many cases these gene transfers have subsequently undergone numerous gene duplications. A large fraction of the HGT-derived proteins are predicted to be secreted (Table 1), strongly suggesting that HGT from fungi has played a significant role in the evolution of

the oomycete secretome. The observed pattern of transfer may have facilitated, or aided, the spread of oomycetes to plant hosts and their evolution into successful plant parasites. Comparative genomics demonstrates that these transfers do not seem to date back to the initial transition to an osmotrophic lifestyle (19) because they are, in the majority, absent in the animal parasitic oomycete *Saprolegnia* and the free-living filamentous osmotrophic *Hyphochytrium*. This suggests that osmotrophy, and the capacity to grow as strand-like hyphal cells, probably arose before diversification of these lineages and before the acquisition of fungal genes by HGT. Interestingly, fungal-derived HGTs may, however, represent specific acquisitions to life as a plant parasite. Our conclusion is consistent with the predicted functions of the HGT candidates aiding entry to plant cells, allowing efficient nutrient acquisition and leading to microbial proliferation in plant tissue.

Materials and Methods

Detailed descriptions of whole-genome gene-by-gene phylogenetic analysis, alternative topology tests, *H. catenoides* DNA and RNA preparation and sequencing, gene annotation, and identification of secreted proteins are provided in *SI Materials and Methods*. All potential HGTs found in a single oomycete genome were treated as possible cases of contamination. To test these, we conducted phylogeny of linked genes on DNA contigs. In all cases, we could confirm that the HGT was present within the oomycete genome (Table S6). An example of each HGT oomycete gene family is included as a combined FASTA file in [Dataset S1](#).

ACKNOWLEDGMENTS. T.A.R. received fellowship support from the Leverhulme Trust and funding from the Natural Environment Research Council (NERC) and the Biotechnology and Biological Sciences Research Council (BBSRC). M.D.M.J. is supported by NERC Grant NE/F011709/1. G.L. is supported by BBSRC Grant BB/G00885X/1.

- Doolittle WF (1999) Lateral genomics. *Trends Cell Biol* 9:M5–M8.
- Andersson JO (2009) Gene transfer and diversification of microbial eukaryotes. *Annu Rev Microbiol* 63:177–193.
- Doolittle WF, et al. (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos Trans R Soc Lond B Biol Sci* 358:39–57, discussion 57–58.
- Boucher Y, et al. (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 37:283–328.
- Jain R, Rivera MC, Moore JE, Lake JA (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol* 20:1598–1602.
- Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9:605–618.
- Richards TA, et al. (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21:1897–1911.
- Doolittle WF (1998) You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* 14:307–311.
- Berriman M, et al. (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309:416–422.
- Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc Natl Acad Sci USA* 100:7678–7683.
- Loftus B, et al. (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* 433:865–868.
- Andersson JO, Sjögren AM, Davis LAM, Embley TM, Roger AJ (2003) Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr Biol* 13:94–104.
- Marcet-Houben M, Gabaldón T (2010) Acquisition of prokaryotic genes by fungal genomes. *Trends Genet* 26:5–8.
- Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. *Biol Lett* 4:366–369.
- James TY, et al. (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443:818–822.
- Pirozynski KA, Malloch DW (1975) The origin of land plants: A matter of mycotrophism. *Biosystems* 6:153–164.
- Wang B, Qiu YL (2006) Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* 16:299–363.
- Dean RA, et al. (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434:980–986.
- Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ (2006) Evolution of filamentous plant pathogens: Gene exchange across eukaryotic kingdoms. *Curr Biol* 16:1857–1864.
- Cavalier-Smith T, Chao EE (2006) Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). *J Mol Evol* 62:388–420.
- Tyler BM, et al. (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261–1266.
- Haas BJ, et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:393–398.
- Baxter L, et al. (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549–1551.
- Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
- Fitzpatrick DA, Logue ME, Stajich JE, Butler G (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol* 6:99.
- Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- O'Brien EA, et al. (2007) TBESTDB: A taxonomically broad database of expressed sequence tags (ESTs). *Nucleic Acids Res* 35(Database issue):D445–D451.
- Van der Auwera G, et al. (1995) The phylogeny of the Hyphochytriomycota as deduced from ribosomal RNA sequences of *Hyphochytrium catenoides*. *Mol Biol Evol* 12:671–678.
- Kemen E, et al. (2011) Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol* 9:e1001094.
- Cultrone A, et al. (2005) Convergent evolution of hydroxylation mechanisms in the fungal kingdom: Molybdenum cofactor-independent hydroxylation of xanthine via alpha-ketoglutarate-dependent dioxygenases. *Mol Microbiol* 57:276–290.
- Sealy-Lewis HM, Scazzocchio C, Lee S (1978) A mutation defective in the xanthine alternative pathway of *Aspergillus nidulans*: Its use to investigate the specificity of uaY mediated induction. *Mol Gen Genet* 164:303–308.
- Sexton AC, Howlett BJ (2000) Characterisation of a cyanide hydratase gene in the phytopathogenic fungus *Leptosphaeria maculans*. *Mol Gen Genet* 263:463–470.
- Ullah AH, Gibson DM (1987) Extracellular phytase (E.C. 3.1.3.8) from *Aspergillus ficuum* NRRL 3135: purification and characterization. *Prep Biochem* 17:63–91.
- Reddy NR, Sathe SK, Salunkhe DK (1982) Phytates in legumes and cereals. *Adv Food Res* 28:1–92.
- de Jonge R, et al. (2010) Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science* 329:953–955.
- Qutob D, et al. (2006) Phytotoxicity and innate immune responses induced by Nep1-like proteins. *Plant Cell* 18:3721–3744.

Supporting Information

Richards et al. 10.1073/pnas.1105100108

SI Materials and Methods

Gene-by-Gene Phylogenetic Analysis of Four Oomycete Genomes to Identify Cases of Horizontal Gene Transfer (HGT). The predicted proteome of four oomycetes genomes, *Phytophthora ramorum*, *Phytophthora sojae*, *Phytophthora infestans*, and *Hyaloperonospora parasitica* (also named *Hyaloperonospora arabidopsidis*) (1–3), were used for gene-by-gene phylogenetic analysis using a bioinformatic protocol described previously (4). The protocol followed a multistep process. Each oomycete genome was treated separately. First, for each genome we identified and removed all candidate transposable elements by comparison with Repbase (5), a database of eukaryotic repetitive elements using tBLASTn with e-value cutoff 10^{-20} . Second, we selected the remaining protein sequences and clustered them into genes of closely related groups across the genome of origin (i.e., identifying recent gene duplications) using OrthoMCL (6) with an e-value cutoff 10^{-20} and an inflation value 1.5. Next, we identified and removed all oomycete cluster groups that were only found in oomycete genomes by performing BLASTp (7) searches of a representative sequence from each cluster group against the genomes in our local database (Table S1); proteins that had no hits other than oomycetes were removed. This left 11,434 cluster groups ready for phylogenetic analysis identifying 3,014, 3,018, 3,233, and 2,169 cluster groups from *P. ramorum*, *P. sojae*, *P. infestans*, and *H. parasitica*, respectively. Because the 11,434 gene sets were identified using a parallel process for all four genomes, this generated some four-way redundancy within the 11,434 gene sets. We did not attempt to remove this redundancy and instead analyzed all 11,434 gene sets. We took this approach to use multiple starting seeds for each gene dataset (where possible) and to control for human error in the manual tree selection stages (described below).

We next generated a fast-ML phylogeny (8) for all 11,434 cluster groups using a bespoke gene-by-gene phylogeny pipeline (4). Briefly, this process consists of a series of PERL scripts, which automatically constructed phylogenetic trees for each sequence cluster group identified. The phylogenies were calculated from taxon sampling using a custom-built MySQL database (www.mysql.com) containing the complete genome project-derived, predicted proteome sequence from 795 species, representing a wide diversity of eukaryotes and prokaryote taxa (Table S1). Each candidate sequence was compared against sequences in the database using BLASTp (7) and the best-similarity hits from each species extracted (using the e-value 10^{-20} gathering threshold). These sequences were aligned using Muscle (9), conserved regions from this alignment were sampled using GBLOCKS (10), and phylogenetic trees constructed using PhyML (8) with a WAG (11) + Γ + I substitution model (Γ + I parameters estimated by PhyML).

This process generated many phylogenies that were unresolved, either because taxon/sequence sampling was too narrow or because highly divergent sequences and paralogues were sampled, limiting resolution of the tree topology. In cases in which highly divergent sequences or too many paralogues limited tree resolution, we adjusted the sampling threshold using 10^{-30} and 10^{-40} to exclude divergent branches or to minimize paralogue sampling. In cases in which the taxon/sequence sampling was too narrow, we increased the threshold to 10^{-10} and 10^{-5} to sample additional members of the gene family. We then repeated the phylogenetic analyses pipeline for these data sets.

All 11,434 phylogenies were manually inspected for tree topologies that suggested fungi–oomycete gene transfer in either

direction. An HGT topology is defined as an oomycete sequence branching within a clade of fungal sequences, or vice versa.

Phylogenetic Analysis of Putative HGT Gene Families. Candidate HGT phylogenies totalling 51 separate cluster groups were recovered demonstrating either oomycetes branching within the fungi or vice versa. For each candidate HGT we checked, the genome sampling encompassed all available data by comparison with the GenBank nr database, GenBank EST database, and the Taxonomically Broad EST database (12). Specific attention was made to check for additional sequence data from the *Ectocarpus* and *Blastocystis* sequence data available in GenBank, the genome project of the Diatom *Fragilariopsis cylindrus* available at the Department of Energy genome portal, the genome of the oomycete fish parasite *Saprolegnia parasitica* available at the Broad Institute's genome portal, and the de novo-generated genome sequence of the sister group to the oomycetes, the free-living osmotrophic protist *Hyphochytrium catenoides* (description of sequencing protocol is outlined below). Additional sequences were added to the alignments as required. This process was facilitated using the sequence management for phylogeny programs Refgen and Treenamer (13). In many cases there were several representatives of the gene family in a single fungal or oomycete genome, with some of these genes often having large sections of the amino acid sequence missing relative to the gene family alignment. This is most likely the product of incomplete assembly and poor gene prediction, specifically intron/exon boundaries during automatic annotation of the genomes—meaning that some predicted protein sequences from the genomes are incomplete. Where the presence of these putatively incomplete sequences did not significantly alter the taxonomic representation of the gene, incomplete sequences were excluded from the alignment. Where the taxon sampling was vital for the phylogenetic analysis, the genome sequence data were manually edited and the ORF repredicted.

Each alignment was manually edited and masked to remove gaps and ambiguous alignment positions using the alignment program Seaview (14). All gene alignments are available at http://cogeme.ex.ac.uk/hgt/Oomycete_fungi_HGT_alignment_files.zip.

For each candidate HGT alignment, we identified the optimal model for phylogenetic analysis using Modelgenerator (15). RAXML v 7.2.6 (16) analysis was then used to assess topology and bootstrap support via our easyRAX script (<http://projects.exeter.ac.uk/ceem/easyRAX.html>). RAXML- and Modelgenerator-predicted models were generally the same, but where they were not, Modelgenerator analyses were used (Table S2). The best-scoring RAXML tree was determined with the PROTMIX method, starting with 10 randomized maximum parsimony trees. Statistical support was evaluated with 100 bootstrap replicates.

The manual alignment checks combined with the second round of phylogenetic analysis demonstrated that 10 of the alignments did not provide strong evidence for HGT, either because the amended taxon sampling demonstrated an alternative tree topology that did not support the HGT hypothesis or because the bootstrap support and tree resolution was too weak to infer HGT.

Four of the HGT gene families were found only in fungi and oomycetes, and therefore HGT is inferred on the basis of taxon sampling only. This process left 37 datasets with phylogenetic support for HGT ready for alternative topology comparison tests (see below).

Testing Phylogenetic Support for Each HGT Using Alternative Tree Topology Tests. For the 37 gene families with phylogenetic trees suggesting HGT we identified and labeled the major fungal and oomycetes taxonomic groups relative to the HGT with a taxon code [i.e., Pezizomycotina 1, Saccharomycotina 2, Taphrinomycotina 3, Basidiomycota 4, “other fungi (including paralogs with or without resolution)” 5, and Oomycetes 6] (Figs. S1.1–S1.30). We then systematically calculated phylogenetic trees with constrained monophyly of these groups and collective groupings (i.e., 1, 2, 3, 4, 1–2, 1–3, 1–4, 1–5, and 6). This process identified 88 constraint groups for the 37 alignments. Searches for the maximum likelihood (ML) trees, constrained or unconstrained, were performed using RAxML v 7.2.6. It was found that a bootstrap analysis gave better trees in most cases, and so 100 bootstraps were used. Searches were carried out using the PROT-CAT-WAG-F model, as a two-step process (default in RAxML 7.2.6.). All of the trees, unconstrained and constrained, from a given alignment were grouped into a single file and the site log likelihoods calculated using RAxML. The output from this analysis was used as input to Consel version 0.1k (17) and the trees from each alignment compared using the approximately unbiased (AU) test (18).

The process identified 21 gene families for which the phylogenetic data suggest horizontal gene transfer between fungi and the oomycetes (Table S3; 20 from fungi to oomycetes and 1 from oomycetes to fungi) and where topology comparison test could reject the monophyly of the donor group [AU test: 1 phylogeny at <0.1 (borderline), 5 at <0.05, 3 at <0.01, 12 at <0.001; Table S3], demonstrating a complex pattern of transfer between these distantly related eukaryotic microbes.

In a further nine cases the AU test could not reject monophyly of the fungi. However, these putative HGTs are included here because in each case the HGT hypothesis was supported by moderate to strong bootstrap support, and/or the taxon sampling of the gene family was restricted to fungi, oomycetes, and a few prokaryote groups, suggesting HGT on the basis of taxon distribution. The evidence for HGT is noted on a case-by-case basis in Table S3.

Analysis to Determine Whether Any Putative HGT Candidates Are the Product of DNA Contamination. Four of the identified HGTs were only present in a single oomycete genome. It is therefore possible that these genes may be annotated as oomycete genes but may be the product of DNA contamination, potentially from fungi, during a genome-sequencing project. To investigate the four single species HGTs, we identified genes adjacent to the putative HGT on genome contigs and generated phylogenies using the pipeline described above. In all four cases we could provide evidence that the gene of putative HGT ancestry was surrounded by oomycete genes of vertical inheritance, suggesting that the HGT derived gene was located on the genome of the oomycete and physically linked to native oomycete genes (Table S6).

Genome Sequencing of *Hyphochytrium*. An *H. catenoides* isolate (ATCC 18719) was inoculated onto Emerson YpSs agar (4 g yeast extract, 15 g soluble starch, 1 g dipotassium phosphate, 0.5 g magnesium sulfate, and 20 g agar dissolved in 1 L deionized water; medium was boiled for 1 min with agitation to dissolve the powder and sterilized by autoclaving at 121 °C for 15 min) and incubated at 25 °C for 2 wk. Colonies were removed, transferred to YpSs medium (made to the YpSs agar recipe but with the agar omitted) and incubated at 25 °C, with agitation, for 3 wk. Biomass was harvested by filtration through microcloth. It was then washed with ultrapure water, frozen in liquid nitrogen, and ground to a powder using a pestle and mortar previously sterilized with 0.1 M NaOH to render them DNA and RNase free. The resulting powder was split into two aliquots, half for DNA and half for RNA extraction. RNA was extracted using a LiCl

RNA extraction protocol. Ground mycelium was added to equal volumes of extraction buffer [0.1 M LiCl, 0.1 M Tris (pH 8) with HCl, 10 mM EDTA, and 1% SDS, made up to volume with ddH₂O] and phenol and mixed by inverting for 1 min. To this, 0.5 volumes of CIA (24:1 chloroform/iso-amyl alcohol) was added and mixed by inversion for 30 s. The sample was centrifuged at 4 °C, 9,500 × g for 30 min and the upper phase transferred to a sterile tube. To this, 1 volume of 4 M LiCl was added and the tube left on ice overnight. Centrifugation was then carried out at 4 °C, 9,500 × g for 20 min, the supernatant removed, and the pellet washed in 70% ethanol and resuspended in Diethylpyrocarbonate water. An equal volume of phenol: CIA was added and the tube vortexed and centrifuged at 4 °C, 16,000 × g for 10 min. The aqueous phase was carefully recovered and 2 volumes of 100% ethanol and 0.1 volumes of 3M sodium acetate added. The RNA was left to precipitate overnight at –20 °C. RNA was recovered by centrifugation at 4 °C, 10,000 × g for 20 min, the supernatant removed, and the pellet washed with 70% ethanol. Finally, RNA was resuspended in 300 µL ultrapure water. Total RNA was quantified and the purity checked using an ND-1000 spectrophotometer (Thermo Fisher Scientific).

DNA was extracted using the Cambio UltraCleansoil DNA kit according to the manufacturer’s alternative protocol for maximum yields and quantified in an ND-1000 spectrophotometer. DNA was checked for both eukaryotic and prokaryotic contamination using SSU rDNA PCR in an MJ mini personal thermal cycler. Each 50-µL reaction contained 2 µL of each primer (10 pMµL⁻¹), 25 µL of Master Mix (Promega, containing 3 mM MgCl₂, 400 µM of each dNTP, and 50 U/mL of Taq DNA polymerase), 19 µL of PCR water, and 2 µL of a 1/1,000 dilution of template DNA. For universal eukaryotic SSU rRNA amplification primers 1F (5'-CTGGTTGATCCTGCCAG-3') and 1520R (5'-CTGCAGG-TTCACCTA-3') and the following cycling conditions were used: initial denaturation at 95 °C for 5 min, followed by 30 identical cycles of denaturation at 95 °C for 1 min, annealing at 57 °C for 1 min, and extension at 72 °C for 1.5 min, with a final extension at 72 °C for 10 min (19). For universal prokaryotic SSU rRNA amplification primers PA (5'-AGAGTTTGATCCTGGCTCAG-3') and PH (5'-AAGGAGGTCATCCAGCCGCA-3') (20) and the following cycling conditions were used, with *Escherichia coli* DNA acting as a positive control: initial denaturation of 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 2 min, with a final extension step of 72 °C for 10 min. Successful amplification was checked by agarose gel electrophoresis on a 0.8% agarose gel, run at 110 V for 45 min. The prokaryotic PCR was negative, whereas the eukaryotic SSU PCR resulted in a clean band of appropriate size. The 1F-1520R PCR product was purified using the Wizard SV gel and PCR clean-up system (Promega) and sequenced externally on both strands by Cogenic (Essex). Chromatograms were checked by eye for inconsistencies, which could be the result of a multitemplate amplification before the derived sequence was used as the seed in a BLASTn search on the National Center for Biotechnology Information BLAST server, suggesting that the DNA was derived from a pure culture of the target microbe.

We sequenced the *Hyphochytrium* genome using two approaches: (i) 454 FLX Titanium (Roche) and (ii) Illumina GA2 paired-end 76-bp sequencing. The DNA was prepared for both sequencing platforms using the standard protocols. Five micrograms of DNA was fragmented by nebulization. Fragmented DNA was analyzed using a Bioanalyzer (Agilent Technologies) to ensure that the majority of the fragments were between 350 and 1,000 bp. The purified fragmented DNA was processed according to the 454 FLX Titanium Library construction kit and protocol (Roche Applied Science) to ligate adaptors specific to the Titanium sequencing chemistry. The resulting single-stranded DNA library was assessed for size distribution using a Bioanalyzer (Agilent Technologies). Library fragments were added to emul-

sion PCR beads at a ratio of 1:1 to emPCR at the optimal ratio of 1.5 DNA copy per beads and amplified according to the manufacturer's instructions (Roche Applied Science), and a full picotitre plate was sequenced. This generated 1,219,849 sequence reads resulting in 416,601,381 bp of sequence.

For Illumina GA2 sequencing we used a Bioruptor shearing device to obtain 500-bp fragments (this figure is inclusive of the adaptor sequences). V2 Cluster generation kits were used along with v4 SBS sequencing kits. Two lanes of Illumina GA2 paired-end 76-bp sequencing was performed, yielding a total of 63,082,628 reads. Reads with adaptor present were removed. Additionally, any reads without Q > 20 across >90% of bases were removed. After this filtering process, 45,453,312 reads were left in total.

The *Hyphochytrium* genome sequence was assembled using the Velvet assembler (0.7.63) (reference <http://genome.cshlp.org/content/18/5/821>) and resulted in 311,983 contigs spanning 85,813,724 bp at a mean coverage of 22 \times . Kmer size of 41 was used along with expected kmer coverage of 8 and coverage cutoff of 3 as determined by the Velvet Optimizer 2.1.7 script bundled with Velvet. At this sampling level the Lander-Waterman model (21) predicts the chance of a given base being absent from the sequence reads as less than 1 in 1 billion. However, this is in the ideal case of even coverage, which is never achieved in practice. In addition, the N50 contig length of the assembly was 611 bp. It was therefore possible that genes would be missed in homology searches owing to the short contig sizes.

To check genome coverage further we sequenced *Hyphochytrium* cDNA using 454 FLX Titanium (Roche) chemistry. The resulting sequence reads were assembled using the Newbler assembler (Roche) on default settings and yielded 31,368 contigs. These contigs were compared with the *Hyphochytrium* assembly using BLASTn. In this simple analysis more than 99% of all cDNAs matched the assembly with >99% identity and e-value <1 \times 10⁻¹⁰. The remaining contigs were either rich in simple sequence repeats or were very short (<100 bp) or of relatively low raw sequence quality. This strongly indicates that there are very few genic sequences missing from the draft assembly. We then used the CEGMA (22) (Core Eukaryotic Genes Mapping Approach) core conserved gene dataset to investigate recovery of core eukaryotic genes from the draft *Hyphochytrium* assembly. Using CEGMA 248 and the 458 genes only 17 and 28 genes, respectively, appear absent from the *Hyphochytrium* assembly using an e-value cutoff of 1 \times 10⁻⁴ and a sequence identity threshold of 90%. A more stringent cutoff of 1 \times 10⁻¹⁰ yielded 51 and 88 missing genes, respectively. The CEGMA core dataset is based on comparison of a range of different eukaryotic organisms (e.g., animals, plants, alveolates) but does not include any close relatives of *Hyphochytrium*. Nevertheless, depending upon the stringency used the gene recovery rate for the CEGMA dataset was between 80% and 94%.

Hyphochytrium genome sequence data have been submitted to Sequence Read Archive for the raw data (SRP004821), and the assembled genome has been archived at the European Bioinformatics Institute (Genome Project: 61035).

We then searched the assembly for candidate homologs of the 34 oomycete HGTs in the *Hyphochytrium* genome using a tBLASTn approach. Of the 34 oomycete HGTs, in seven cases candidate homolog sequences were identified in *Hyphochytrium*. In five cases the multiple alignment analyses demonstrated that the *Hyphochytrium* sequences were distant relatives, or too divergent to be true homologs, and were therefore not relevant to the putative HGT and excluded from further analysis. In the remaining two cases we added the *Hyphochytrium* genes to the alignments and recalculated the phylogenetic analysis. In both cases the *Hyphochytrium* genes were distantly related paralogues to the oomycetes HGTs branching separately from the fungal/

oomycetes HGT branches (Figs. S1.5 and S1.16), demonstrating that the *Hyphochytrium* genes have a different evolutionary derivation and that all 34 fungal-derived HGTs are absent from *Hyphochytrium* genome assembly.

Putative Functional Annotations of the HGT Gene Families and Comparison with *Phytophthora* Metabolic Network Analysis. To annotate putative functions for all 34 HGT gene families, a representative of each HGT gene was used for BLAST (7) and PFAM HMM homology searches (23). This process was used to assign a putative function to each HGT gene family. Each putative annotation was checked to exclude false annotations present in GenBank. To investigate putative cellular locations of each HGT-encoded protein, the complete HGT gene families, including all duplicate forms from all four oomycete genomes, were recovered and subject to WoLFPSORT (24), SignalP (25), and TMHMM (26) analysis for evidence of N-terminal secretion motifs and/or transmembrane domains. The results of these annotations are listed in Table S5. Evidence for an N-terminal secretion motif is only listed in Table S5 when both WoLFPSORT (24) and SignalP (25) methods suggest that the protein character is present.

To investigate where the oomycete HGT proteins putatively function within the metabolic network of the *Phytophthora* species, we sought to reconstruct the oomycete metabolic network. Complete proteome sequence files of *P. ramorum*, *P. sojae*, and *P. infestans* (proteins. fasta files) were recovered from the Broad Institute *Phytophthora* database. We used Kaas (Kegg automated annotation server) to project the downloaded proteome sequences onto Kegg's collection of metabolic pathways (27), using BDH (bidirectional hit) orthology search (similarity index threshold: 60) to explore the Kaas reference genome set (enriched with available fungi and protozoa genomes). Initially this analysis was performed to check and correct the annotations for the 34 HGT gene families (discussed above). Then we used this process to investigate where the HGT candidates fitted into the oomycete metabolic network map. This network analysis often demonstrated partial or missing pathways. In these cases we exploited a minimal threshold (28) in a combination with an SDH (single-direction hit) method to recheck specific components of pathways putatively connected to the HGT functions. The BDH and SDH methods were completed for all three *Phytophthora* predicted proteomes. We also used BLASTn and tBLASTn additionally to query the Broad Institute *Phytophthora* databases for the missing functions by known protein and gene sequences to reconfirm incomplete pathways. Finally, we used the SEED database (29) "Compare regions" tool to investigate synteny of each HGT gene family across distantly related taxa to help define evidence of potential functional linkage of putative homologs of the HGT genes.

Using microarray data for the *P. infestans* transcriptome, we identified evidence of altered transcript abundance both before and after infection. To do this we used publicly available data sets (2) recording the expression of genes in *P. infestans* over a 5-d time course of a potato infection to identify genes previously classified as "genes induced or repressed during infection." These candidate genes were identified by comparing expression intensities derived from samples of infected potato tissue to baseline expression intensities provided by samples derived from mycelium growing in axenic culture (Pea, V8, or RS agar). The data were published as part of the *P. infestans* genome project (2). Normalized transcript abundance values between the two samples were compared using the Student *t* test. Of the 18 HGT gene families present in the *P. infestans* genome, 6 HGT gene families showed evidence of up-regulation *in planta* [2 gene families at <0.1 (borderline), 2 at <0.05, and 2 at <0.01; Table S5].

1. Tyler BM, et al. (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261–1266.
2. Haas BJ, et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:393–398.
3. Baxter L, et al. (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549–1551.
4. Richards TA, et al. (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21:1897–1911.
5. Jurka J, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462–467.
6. Li L, Stoeckert CJ Jr., Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
7. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
8. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
9. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
10. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552.
11. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–699.
12. O'Brien EA, et al. (2007) TBestDB: A taxonomically broad database of expressed sequence tags (ESTs). *Nucleic Acids Res* 35(Database issue):D445–D451.
13. Leonard G, Stevens JR, Richards TA (2009) REFGEN and TREENAMER: Automated sequence data handling for phylogenetic analysis in the genomic era. *Evol Bioinform Online* 5:1–4.
14. Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 12:543–548.
15. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29.
16. Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
17. Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
18. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508.
19. Lefèvre E, et al. (2007) Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environ Microbiol* 9:61–71.
20. Edwards U, Rogall T, Blöcker H, Emde M, Böttger EC (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res* 17:7843–7853.
21. Lander ES, Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 2:231–239.
22. Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067.
23. Marchler-Bauer A, et al. (2005) CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Res* 33(Database issue):D192–D196.
24. Horton P, et al. (2007) WoLF PSORT: Protein localization predictor. *Nucleic Acids Res* 35(Web Server issue):W585–W587.
25. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
26. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6: 175–182.
27. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35(Web Server issue):W182–W185.
28. Bateman A, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32 (Database issue):D138–D141.
29. Overbeek R, et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691–5702.

Figs. S1.1–S1.30. Phylogenetic evidence for fungi–oomycete HGTs. Phylogenies are calculated as described in *SI Materials and Methods*. Each figure contains a key describing the figure notations. Note that Figs. S1.5 and S1.16 are the datasets with representation from *Hyphochytrium*

[Figs. S1.1–S1.30](#)

Table S1. List of genomes included in the phylogenetic pipeline database used in this study

[Table S1](#)

Table S2. Model parameters used for phylogenetic analysis estimated by Modelgenerator (RaxML has a preset α parameter for the Γ distribution, so no α value is given, etc.)

[Table S2](#)

Table S3. Phylogeny topology comparison tests to confirm branching position of HGT within donor group

[Table S3](#)

HGT in red demonstrate transfer in opposite direction. Transfers in gray are not supported by alternative topology comparison test, or alternatively the HGT hypothesis is based on taxon distribution data only, suggesting a weaker standard of support for the HGT. Sequences for *P. ramorum* and *P. sojae* are available from the Department of Energy, Joint Genome Institute website. Protein sequences can be obtained using the protein ID and searching at http://genome.jgi-psf.org/pages/search-for-genes.jsf?organism=Phyra1_1 for *P. ramorum* and http://genome.jgi-psf.org/pages/search-for-genes.jsf?organism=Physo1_1 for *P. sojae*.

Table S4. Evidence of conserved introns across donor and recipient groups in eight HGT gene families[Table S4](#)

Of the 30 HGTs supported by phylogenetic data, we searched for conserved introns in the HGT gene families of the recipient taxa. We found 15 cases in which the transferred gene family had one or more conserved intron. Of these 15 HGT families, 8 conserved introns were also present in the donor groups, suggesting that the pattern of HGT is from eukaryote-to-eukaryote and not as a eukaryotic-to-prokaryotic-to-eukaryotic transfer.

Table S5. Annotation of putative HGT candidates from fungi to oomycetes[Table S5](#)

The HGT shaded in red demonstrates a putative transfer in opposite direction (oomycete to fungi). Transfers in gray are not supported by AU alternative topology comparison test (1), or alternatively the HGT hypothesis is based on taxon distribution data only, suggesting a weaker standard of support for the HGT. Evidence for secretion was identified using both the SignalP and WolfP-Sort methods, as described in *SI Materials and Methods*. Presence of a putative N-terminal secretion motif is only suggested when both methods agree. Sequences for *P. ramorum* and *P. sojae* are available from the Department of Energy, Joint Genome Institute website. Protein sequences can be obtained using the protein ID and searching at the Department of Energy Joint Genome Institute website: http://genome.jgi-psf.org/pages/search-for-genes.jsf?organism=Phyra1_1 for *P. ramorum* and http://genome.jgi-psf.org/pages/search-for-genes.jsf?organism=Physo1_1 for *P. sojae*. Overlap between this analysis and the results reported by Richards et al. (2) is also listed in column 3. In this previous study we identified an additional three tentative HGTs [one Aconitase genes family and two divergent paralogues of the Esterase/Lipase gene family; Fig. S3 of the 2006 paper (2)]. In the 2006 study we could not provide strong evidence either to support or refute the HGT hypothesis. These three HGTs were not recovered in this analysis.

1. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492e508.
2. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ (2006) Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr Biol* 16:1857–1864.

Table S6. Check for contaminant sequences by phylogeny of genes adjacent to HGT[Table S6](#)**Other Supporting Information Files**[Dataset S1](#)

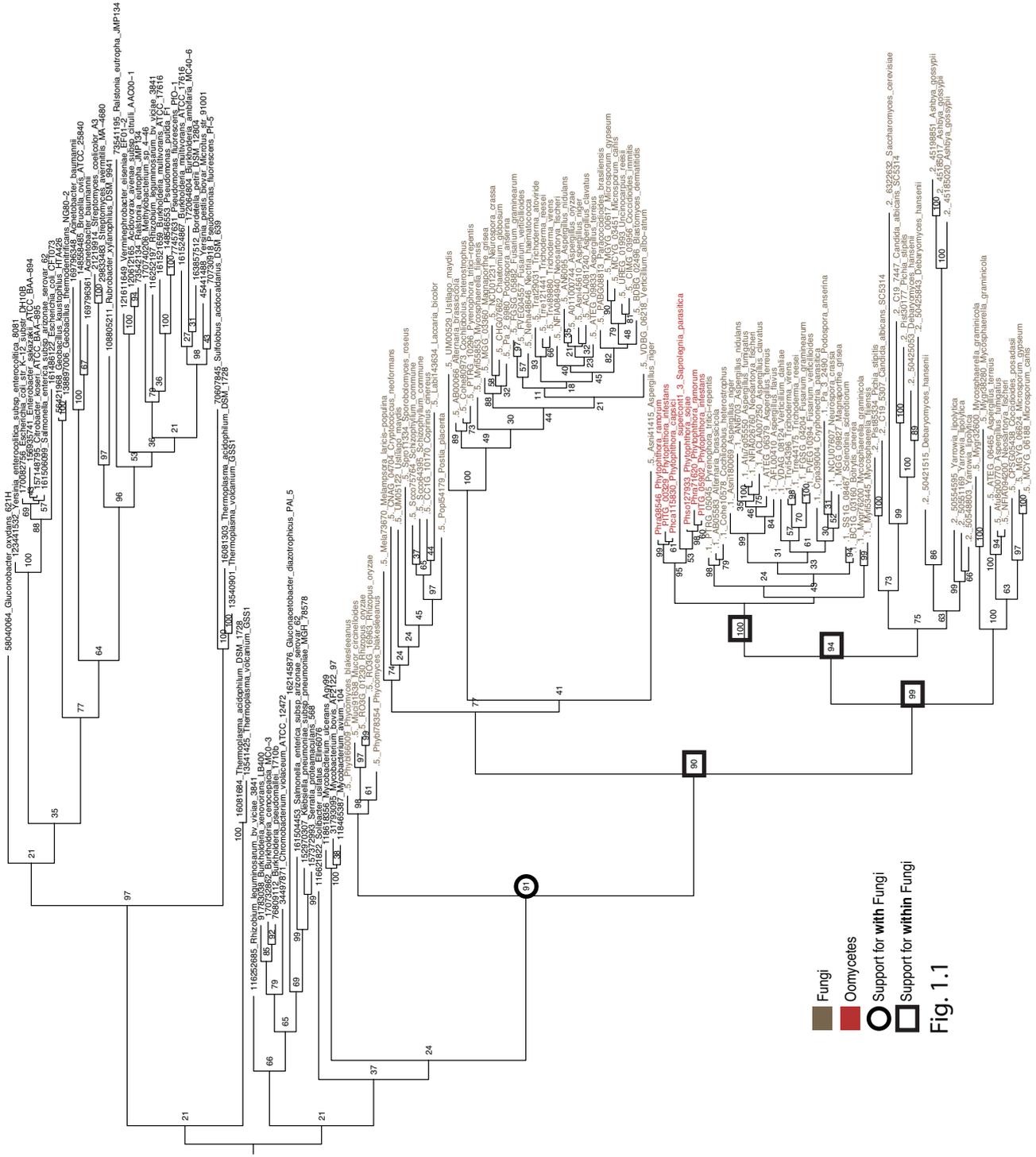
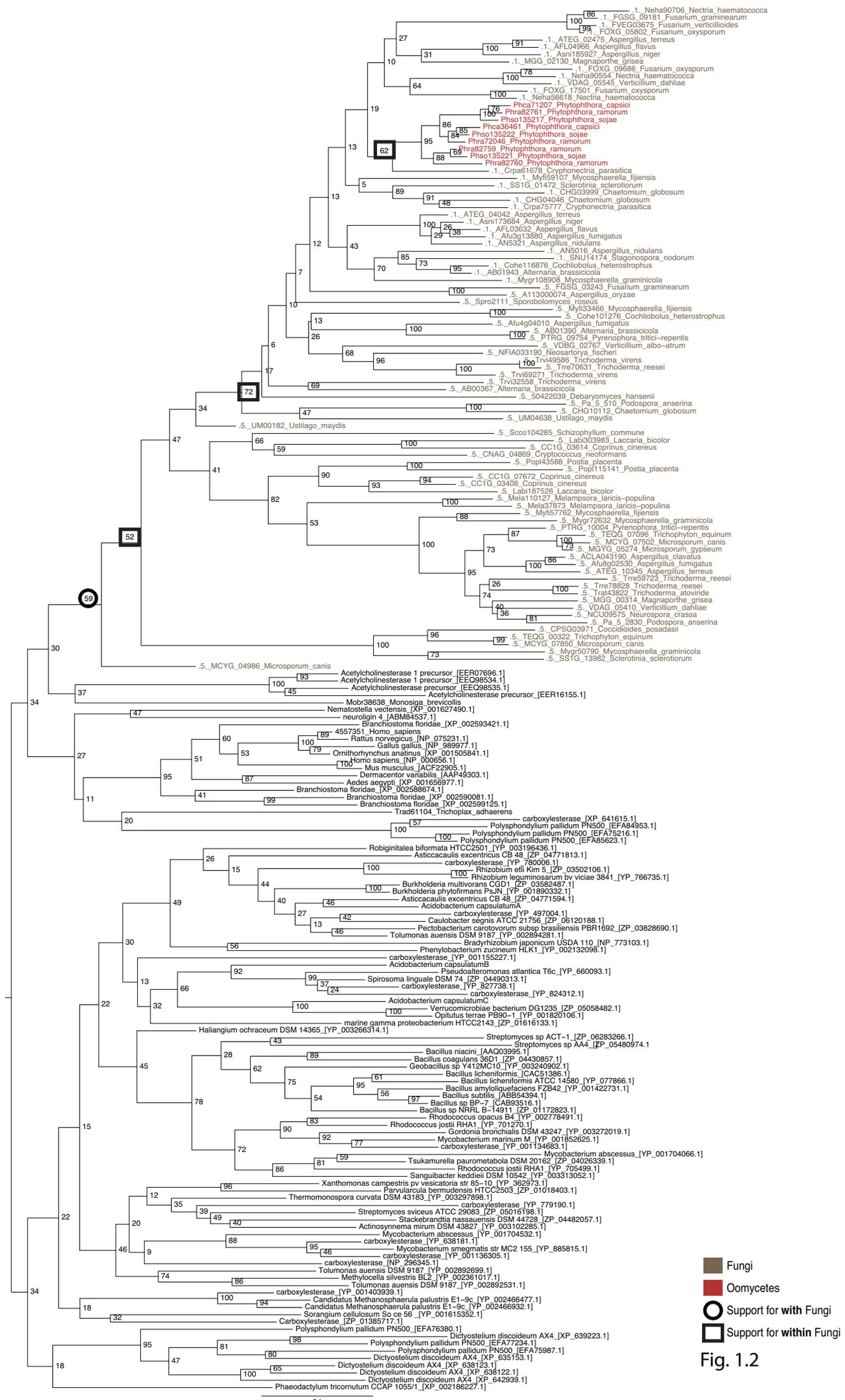
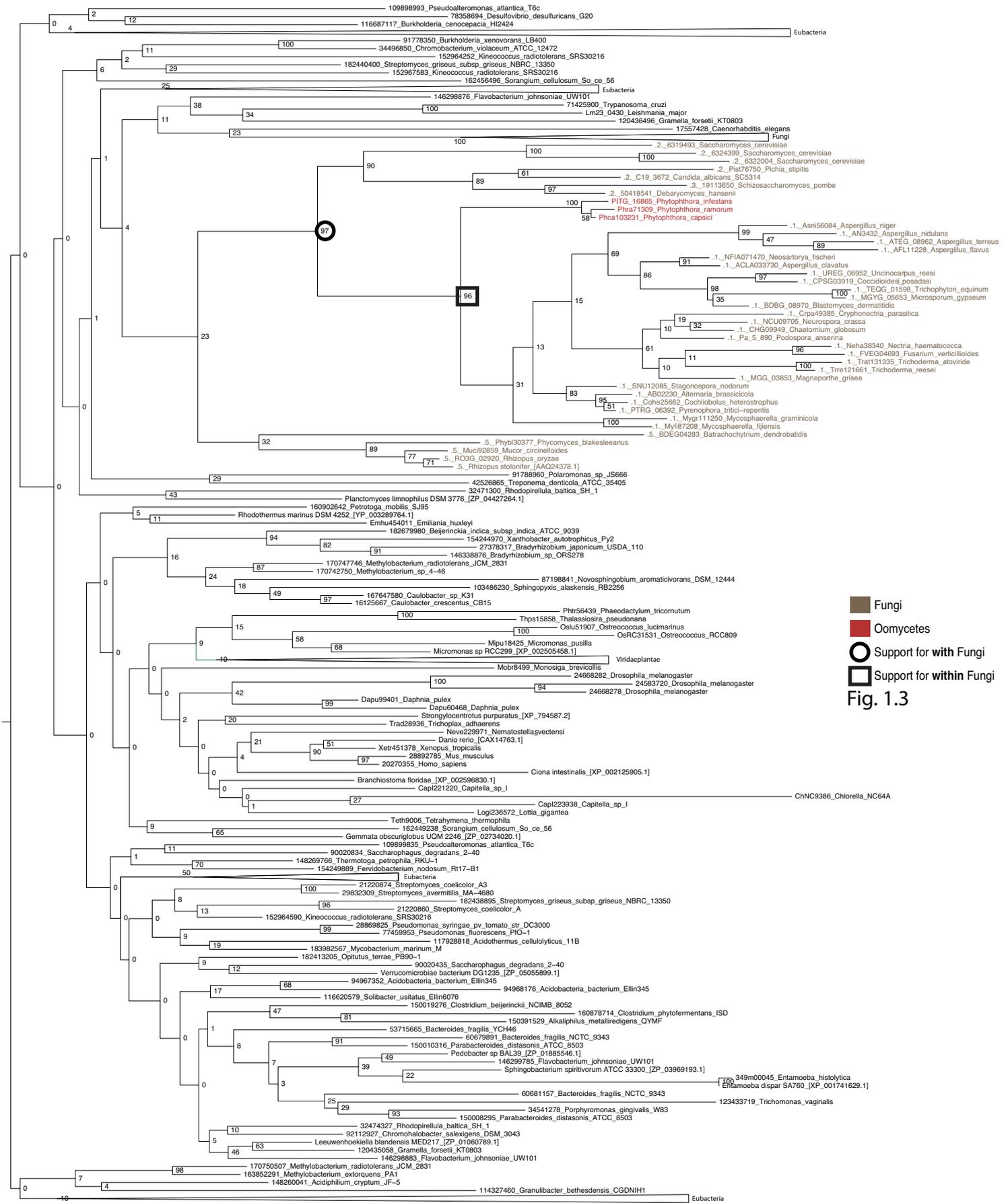


Fig. 1.1



■ Fungi
■ Oomycetes
○ Support for with Fungi
□ Support for within Fungi

Fig. 1.2



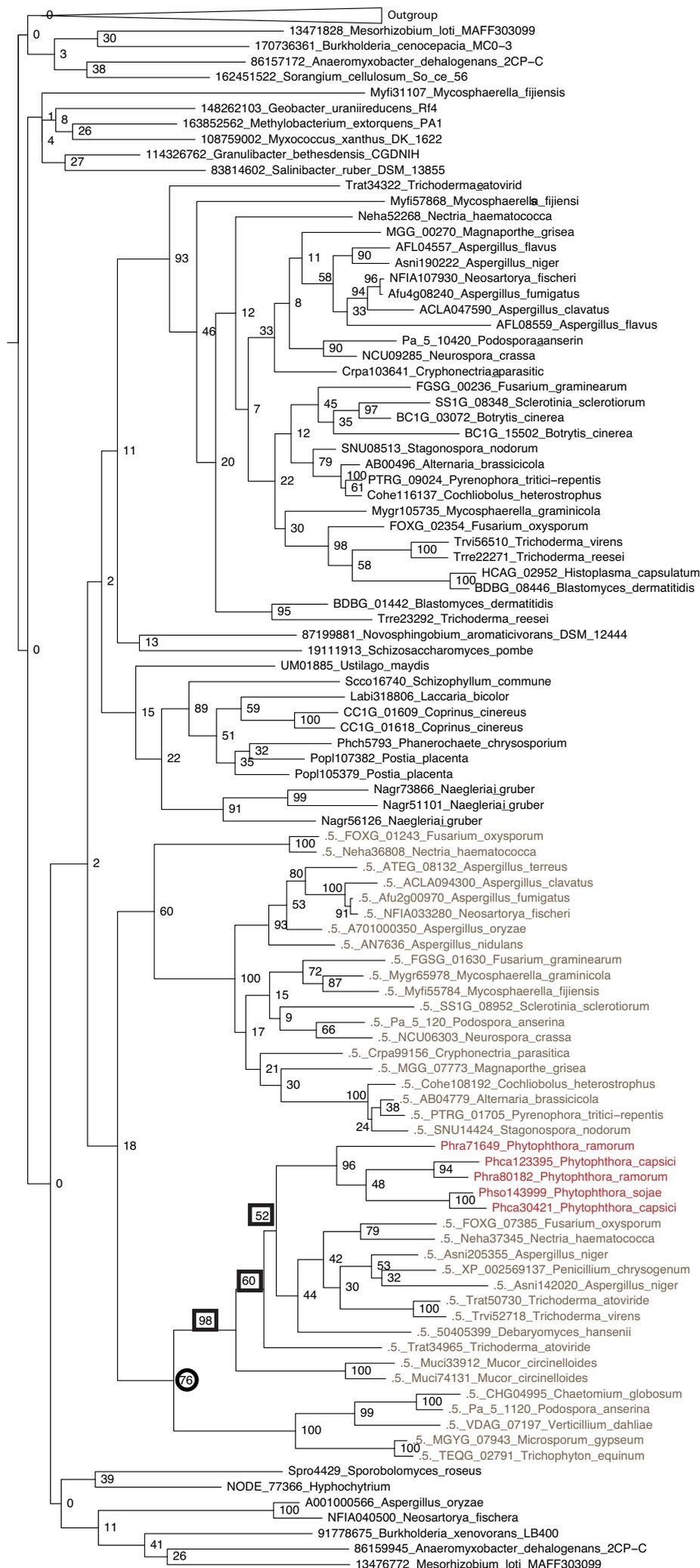


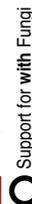
Fig. 1.5



Fungi



Oomycetes

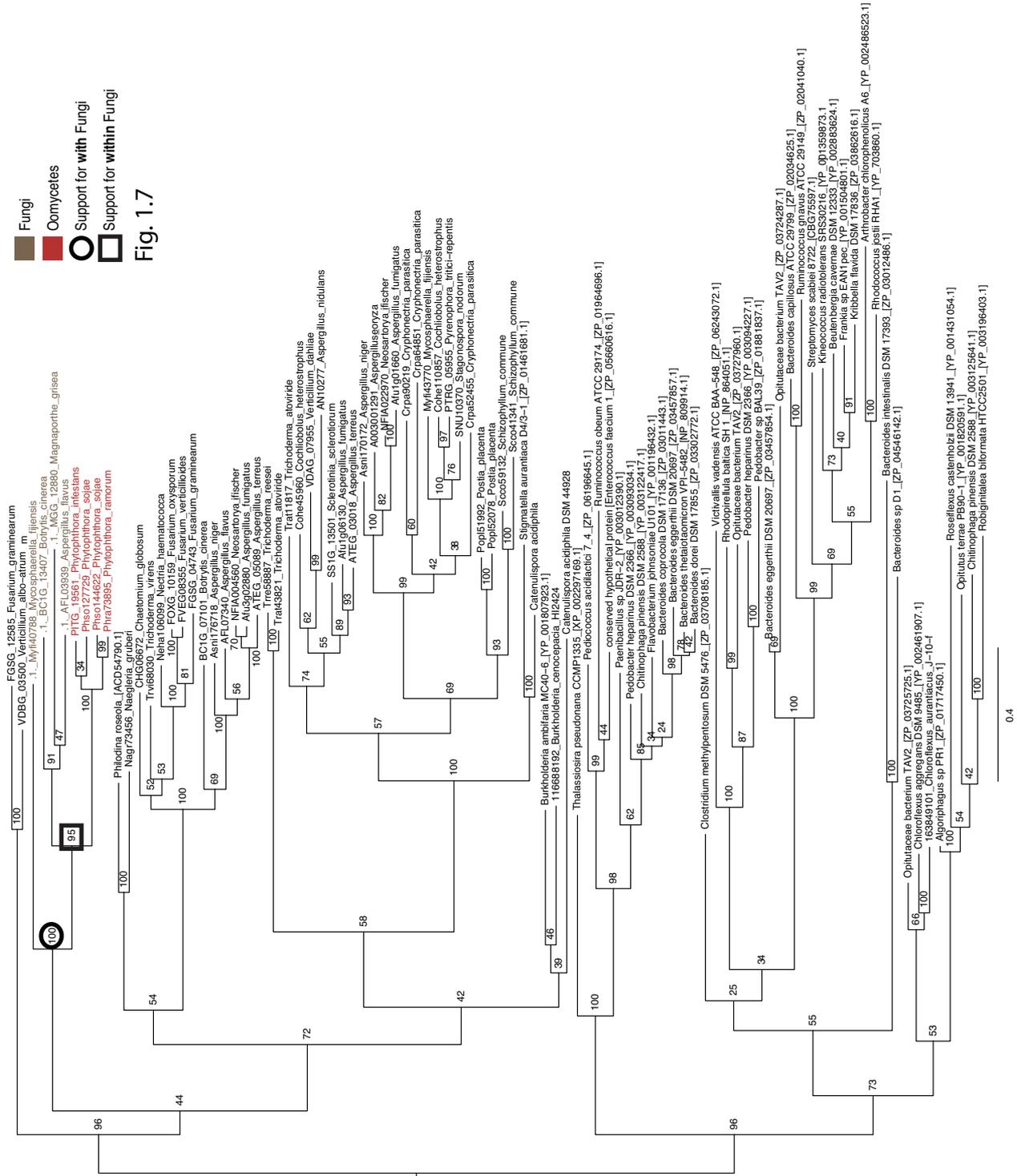


Support for with Fungi

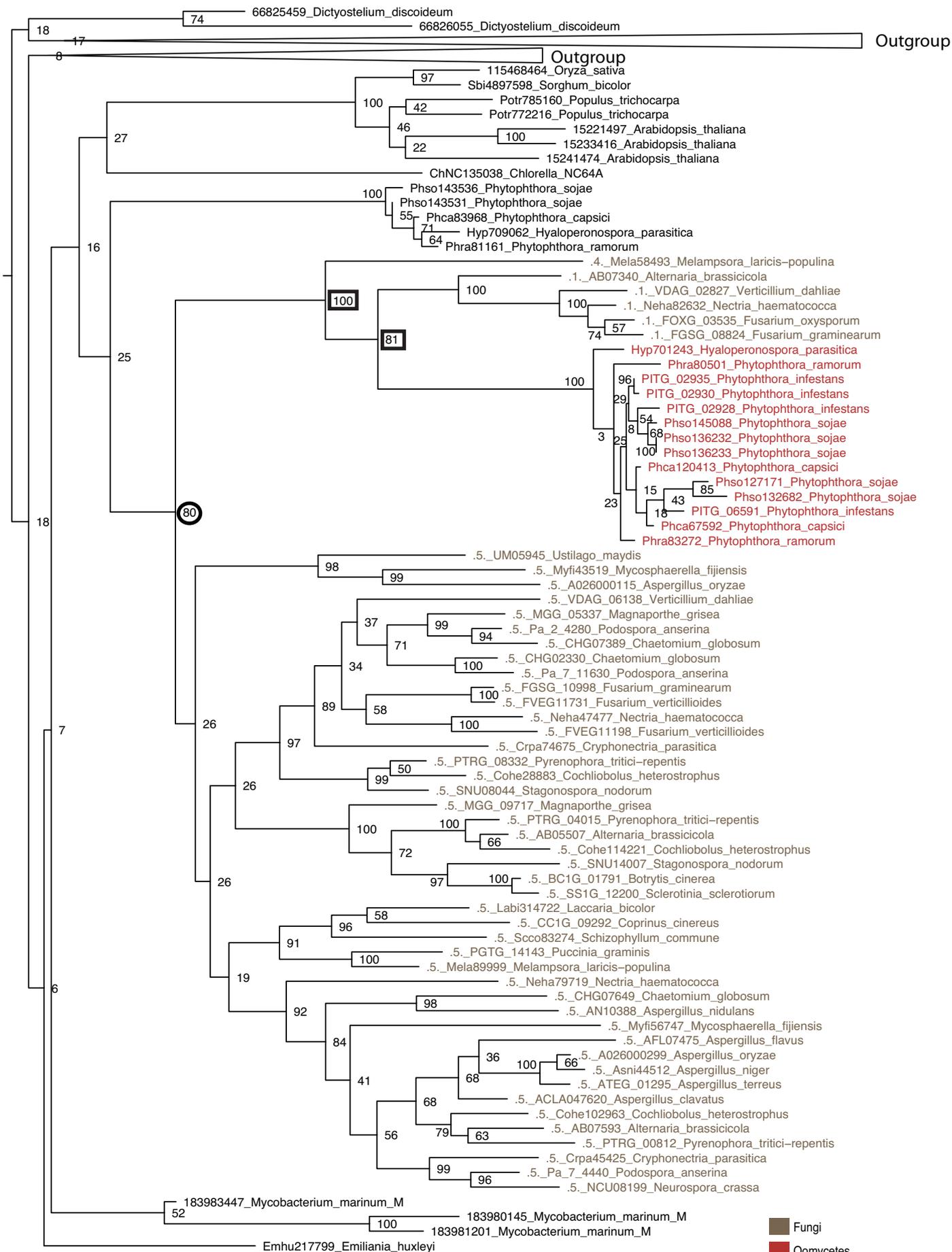


Support for within Fungi

Fig. 1.7

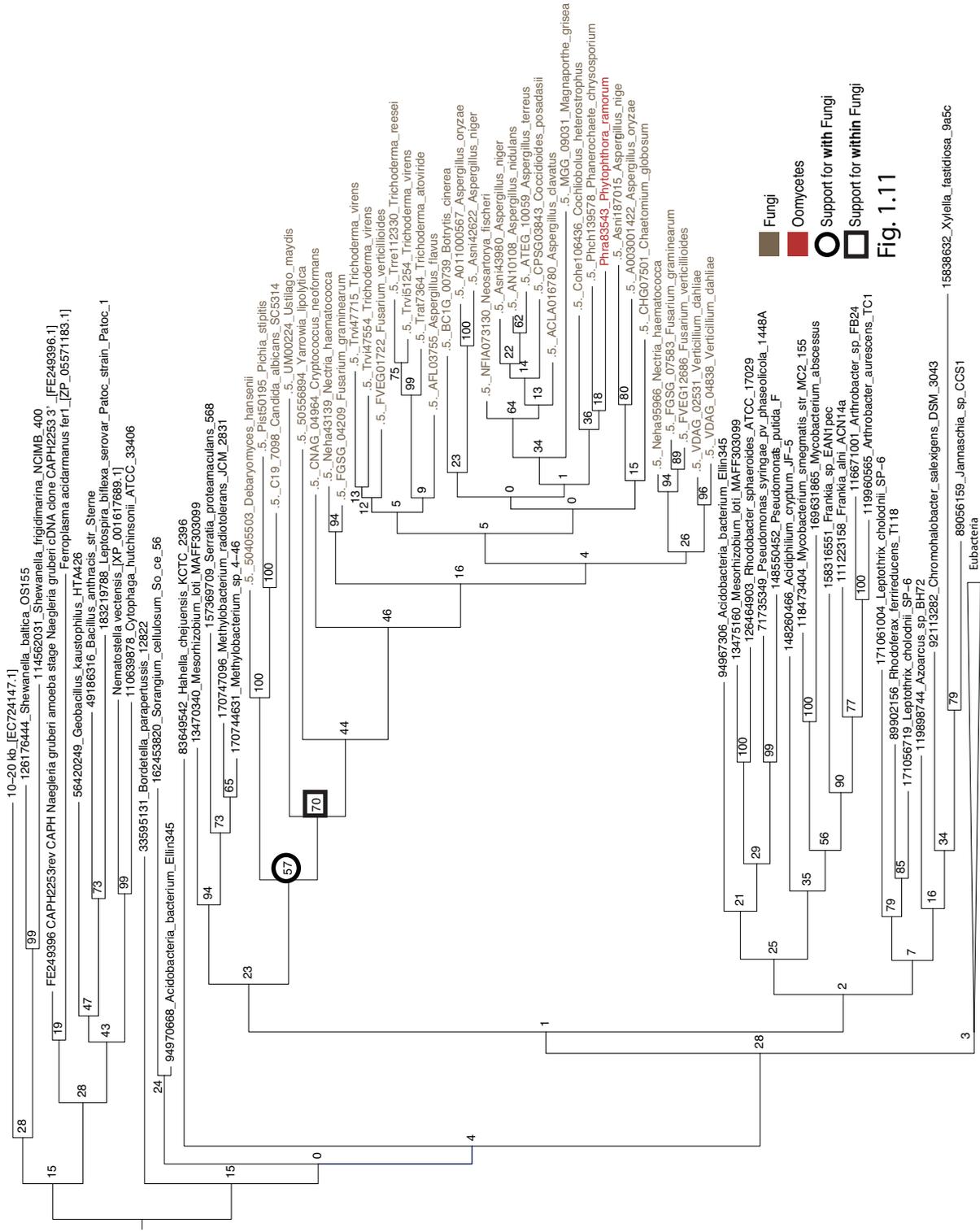


0.4



Fungi
 Oomycetes
 Support for **with** Fungi
 Support for **within** Fungi

Fig. 1.9



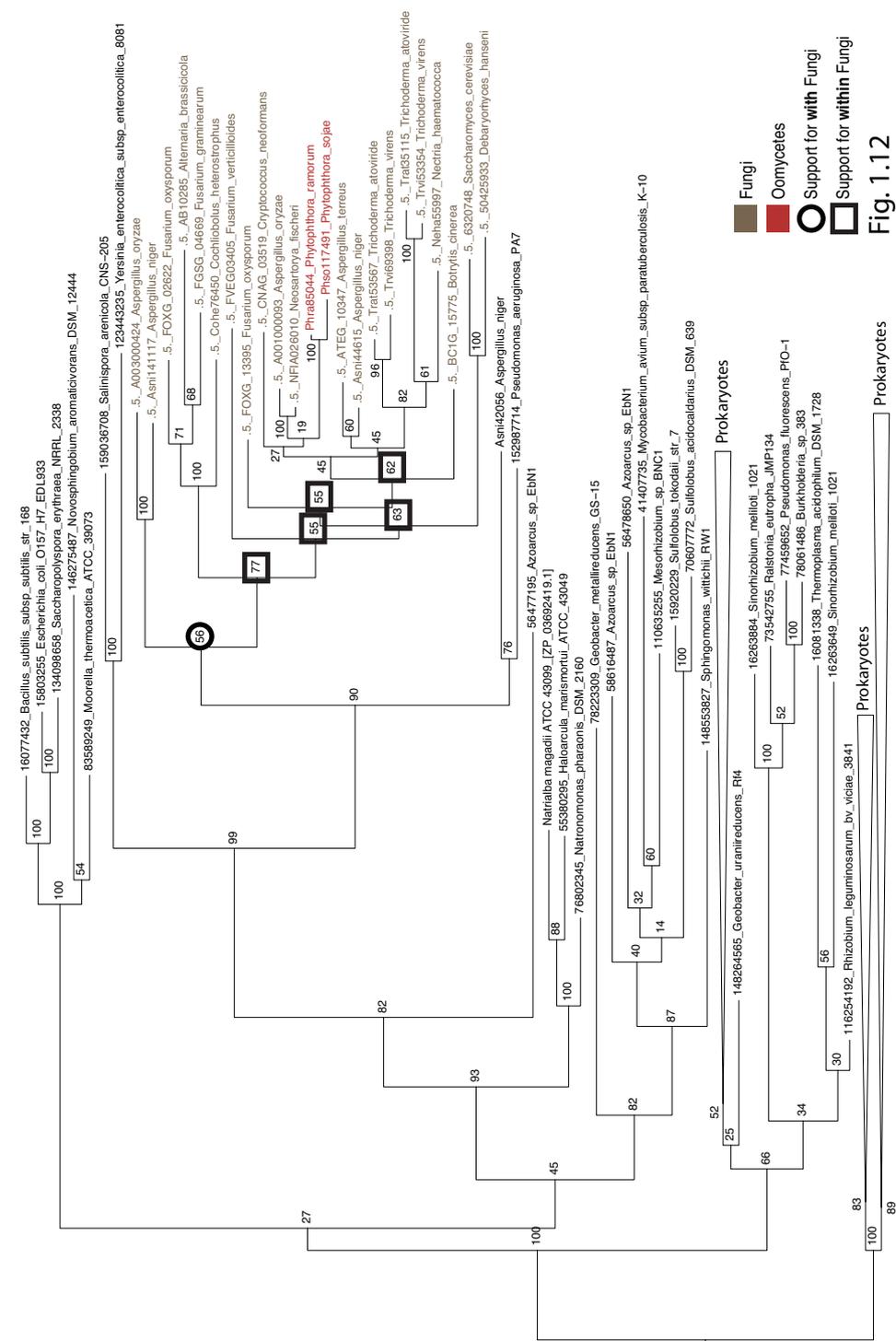
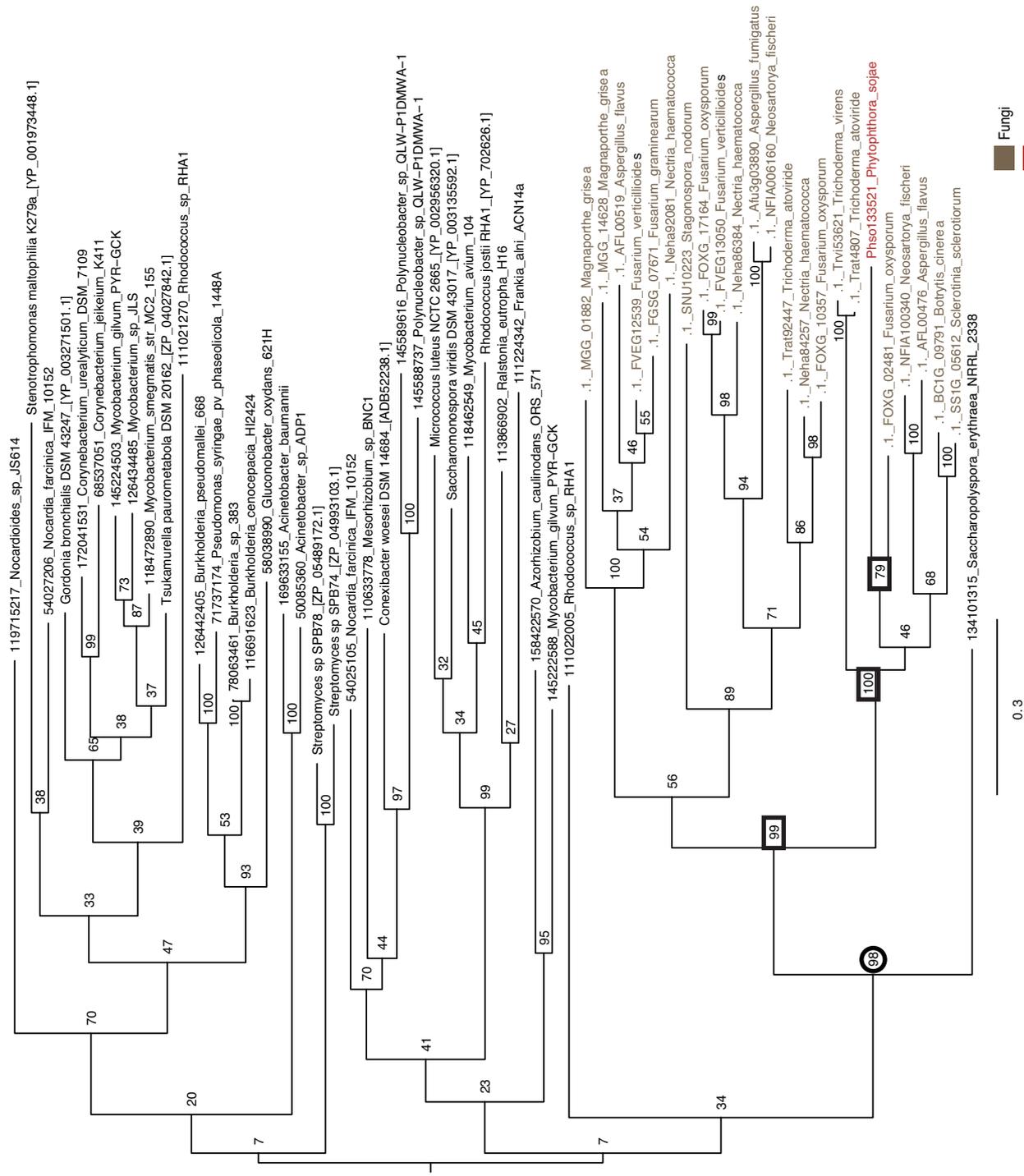
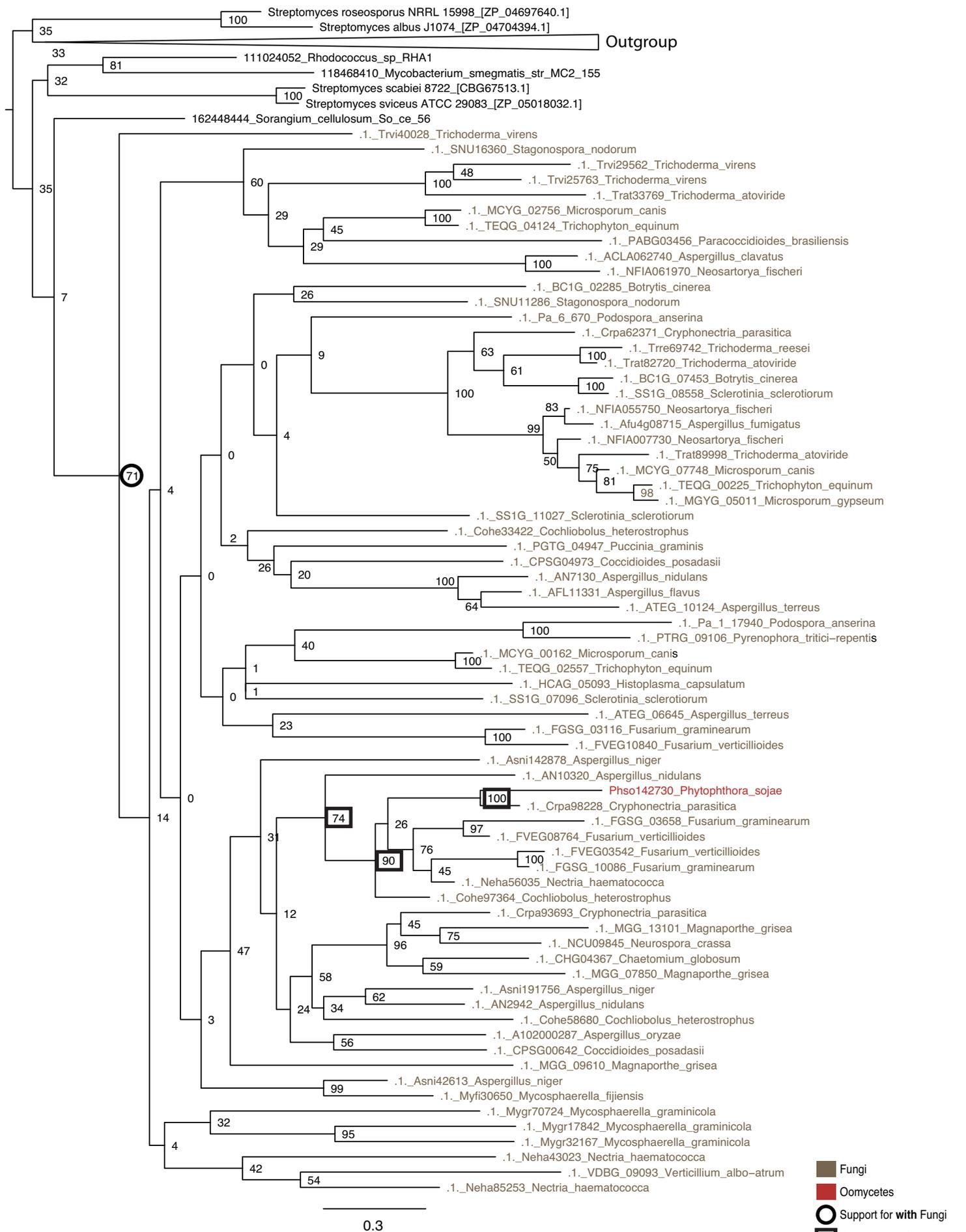


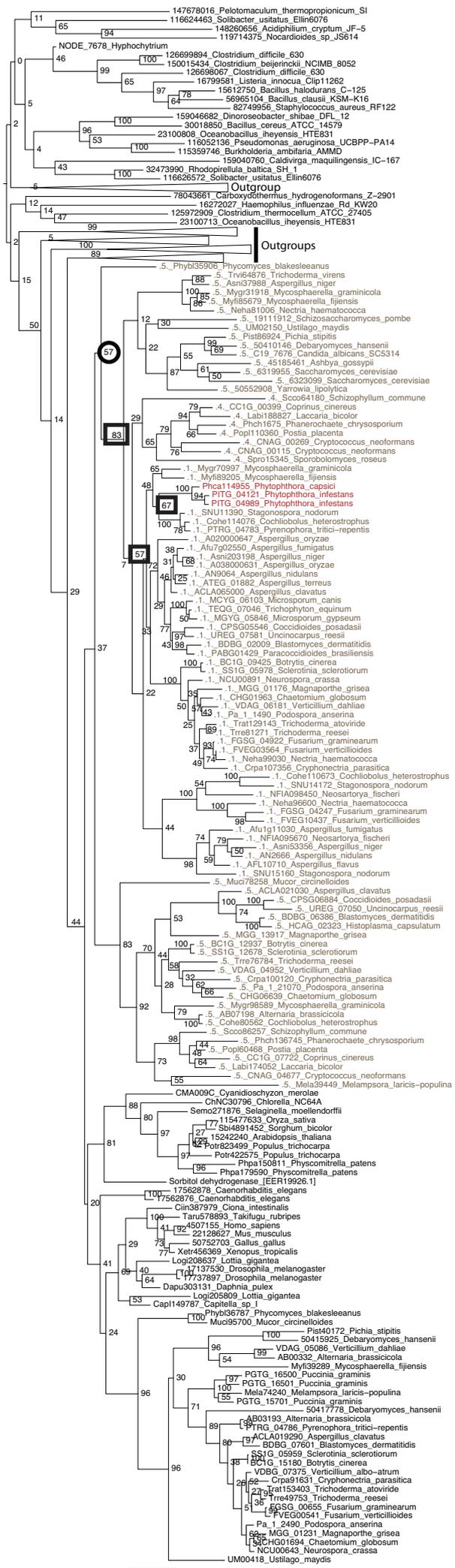
Fig. 1.12



Fungi
 Comycetes
 Support for **with** Fungi
 Support for **within** Fungi

Fig. 1.14





Fungi
 Oomycetes
 Support for with Fungi
 Support for within Fungi
Fig. 1.16

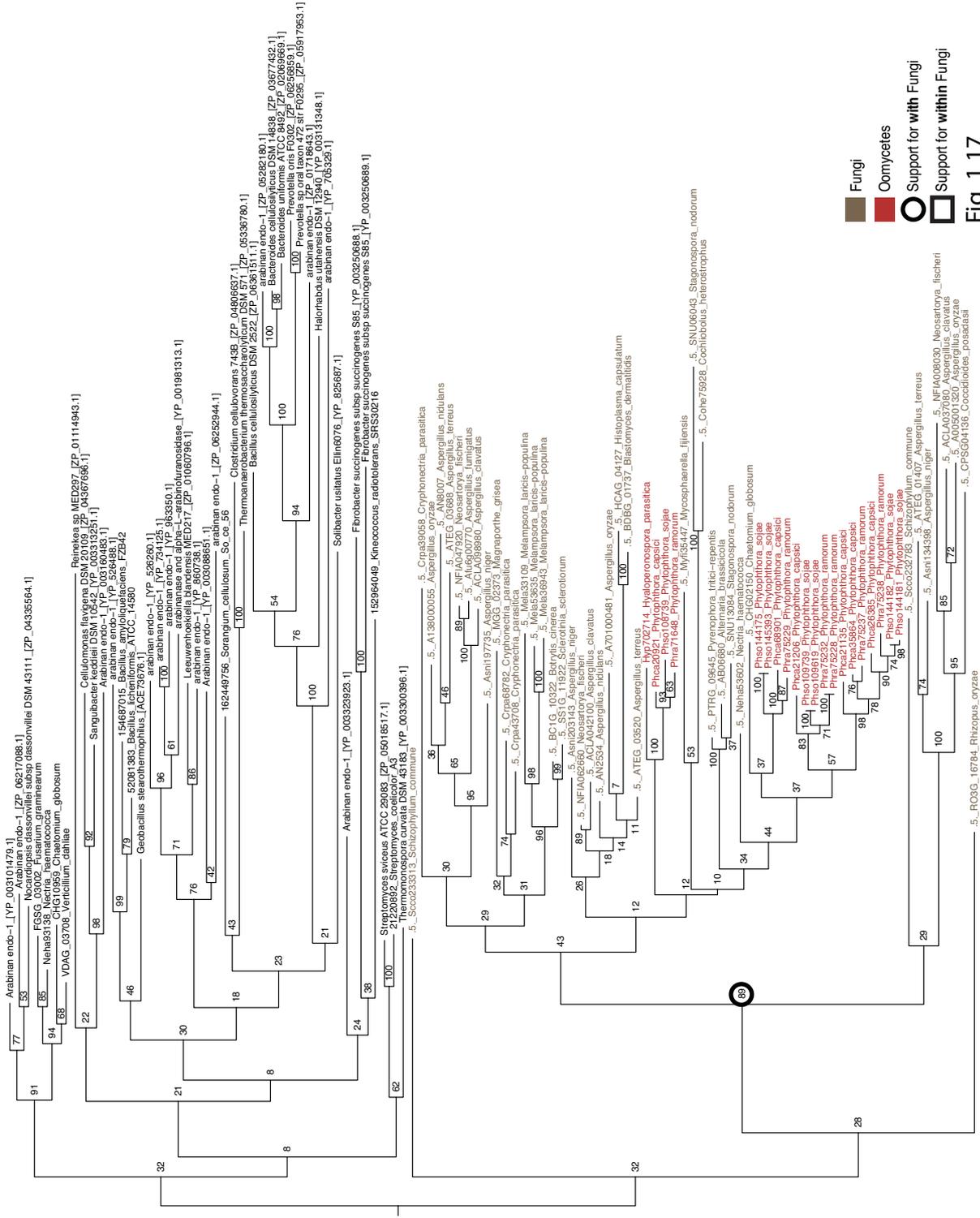
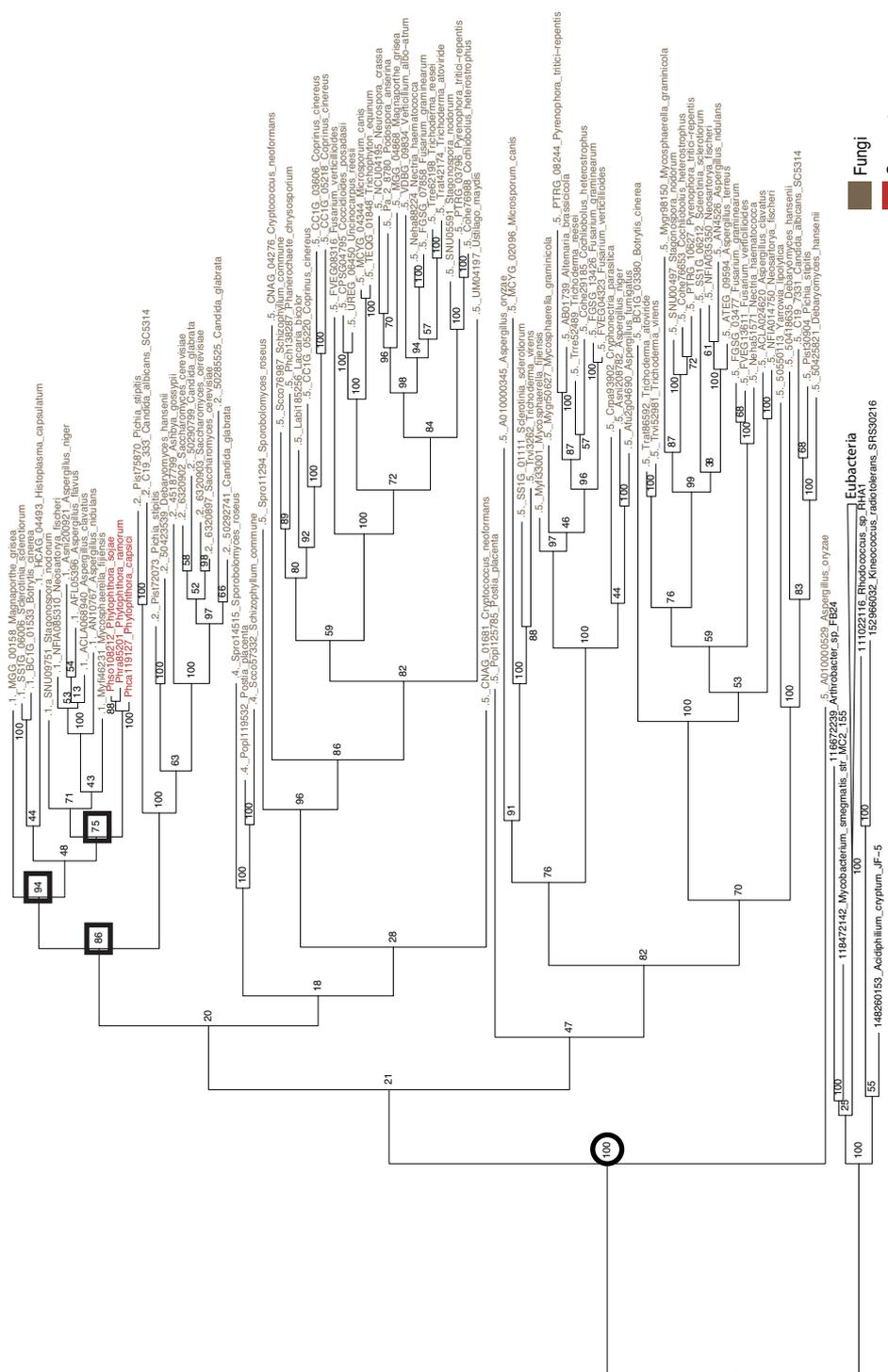
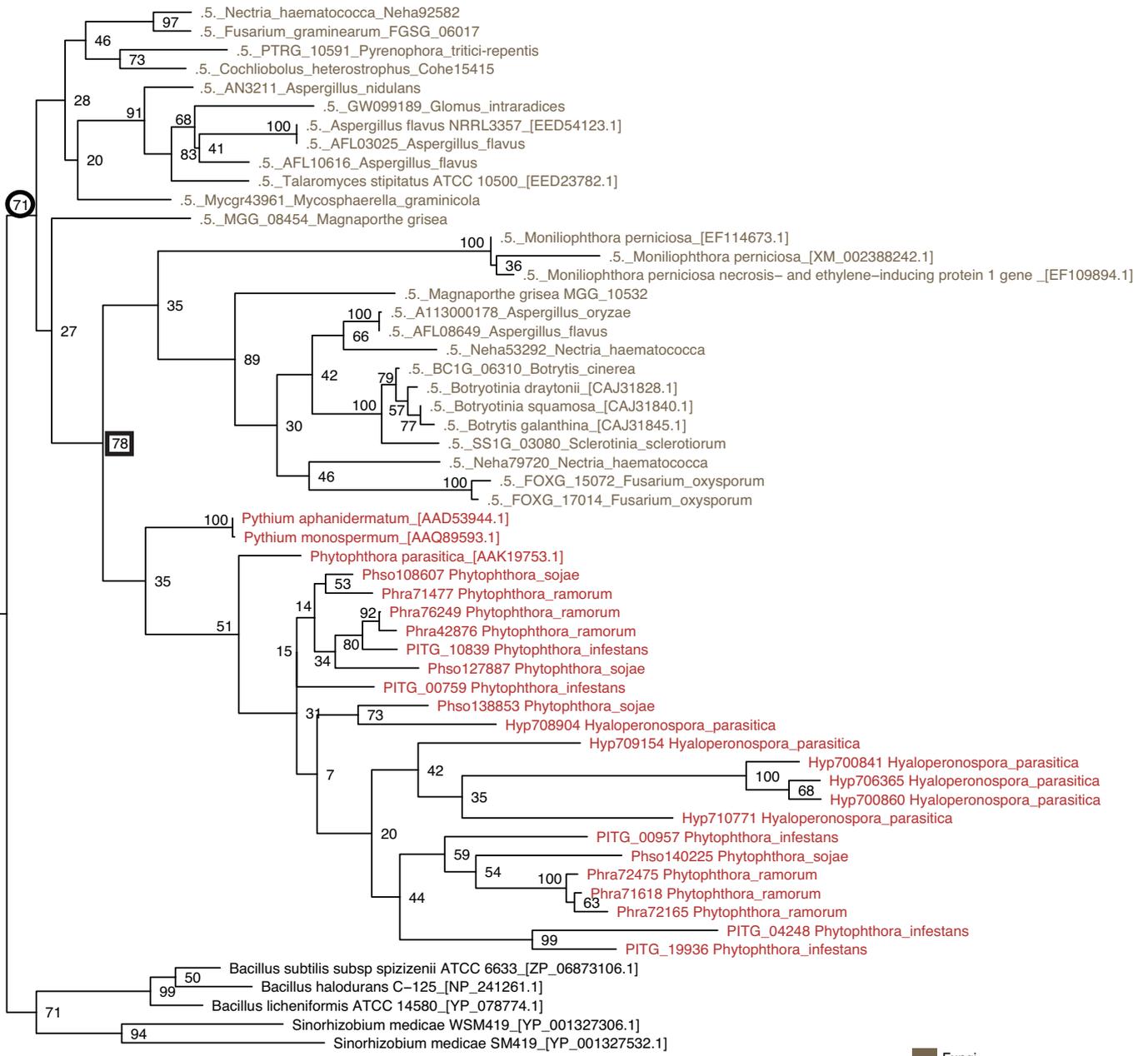


Fig. 1.17



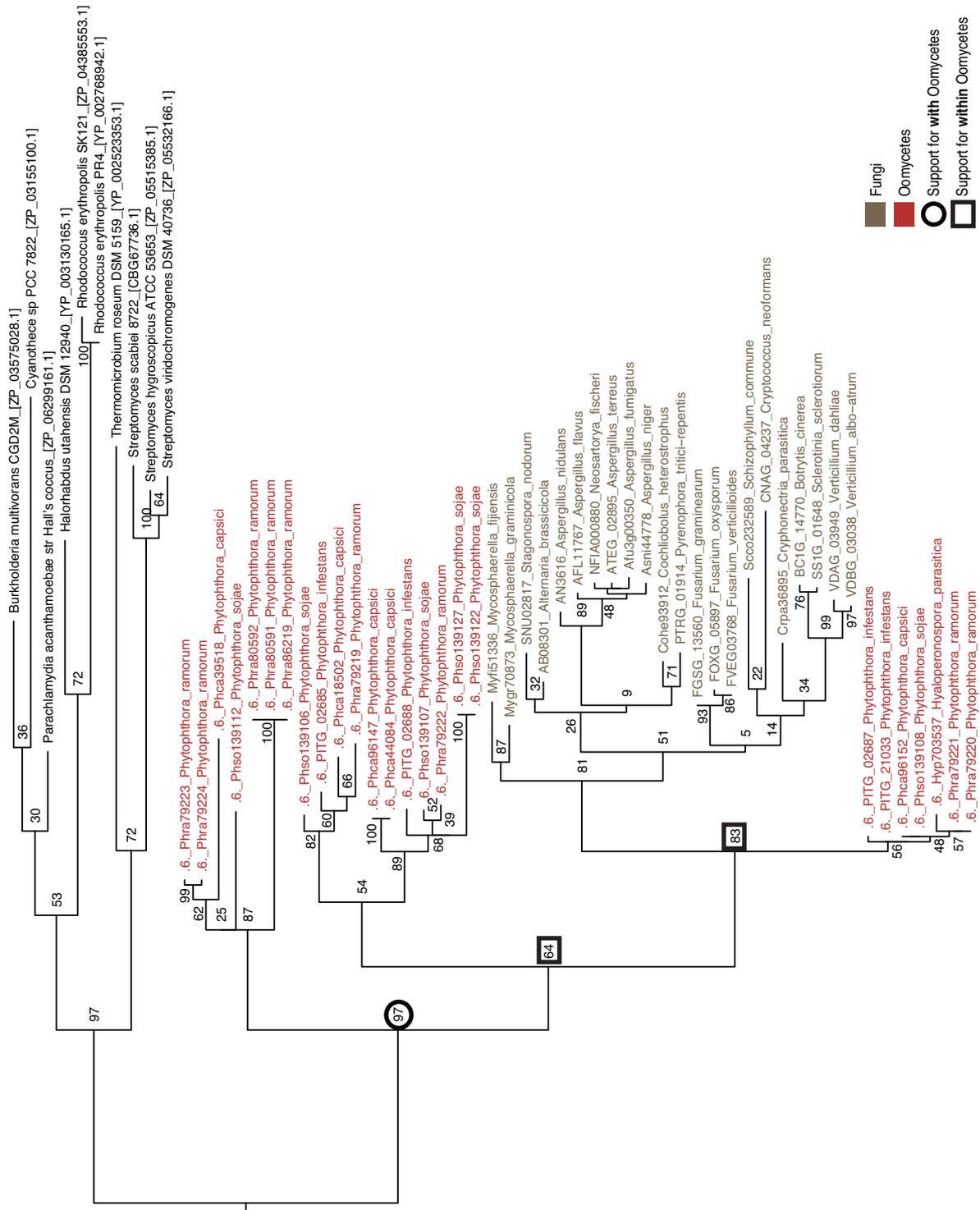
 Fungi
 Oomyetes
 Support for with Fungi
 Support for within Fungi

Fig. 1.18



Fungi
 Oomycetes
 Support for with Fungi
 Support for within Fungi

Fig. 1.20



0.4

Fig. 1.21

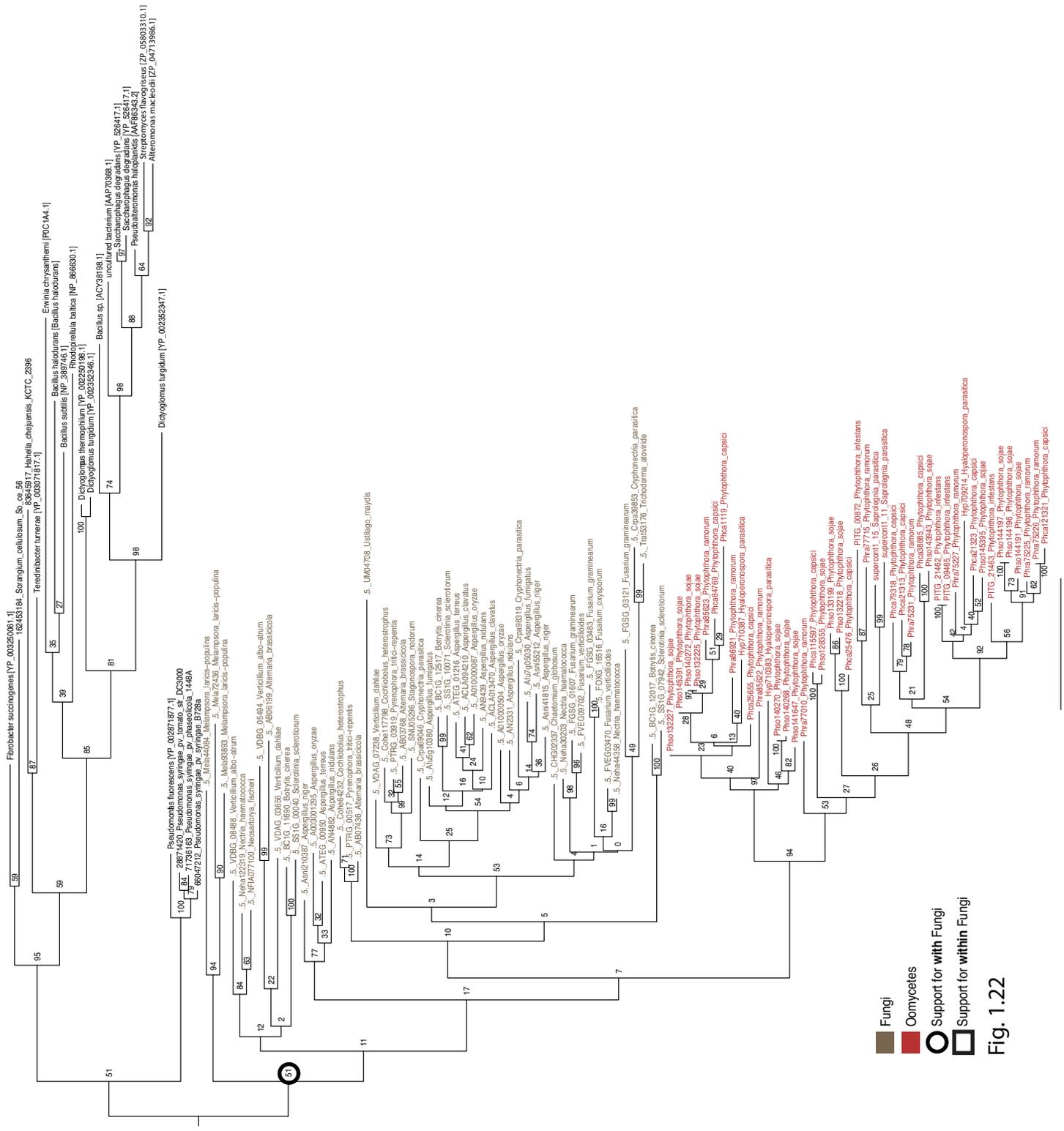
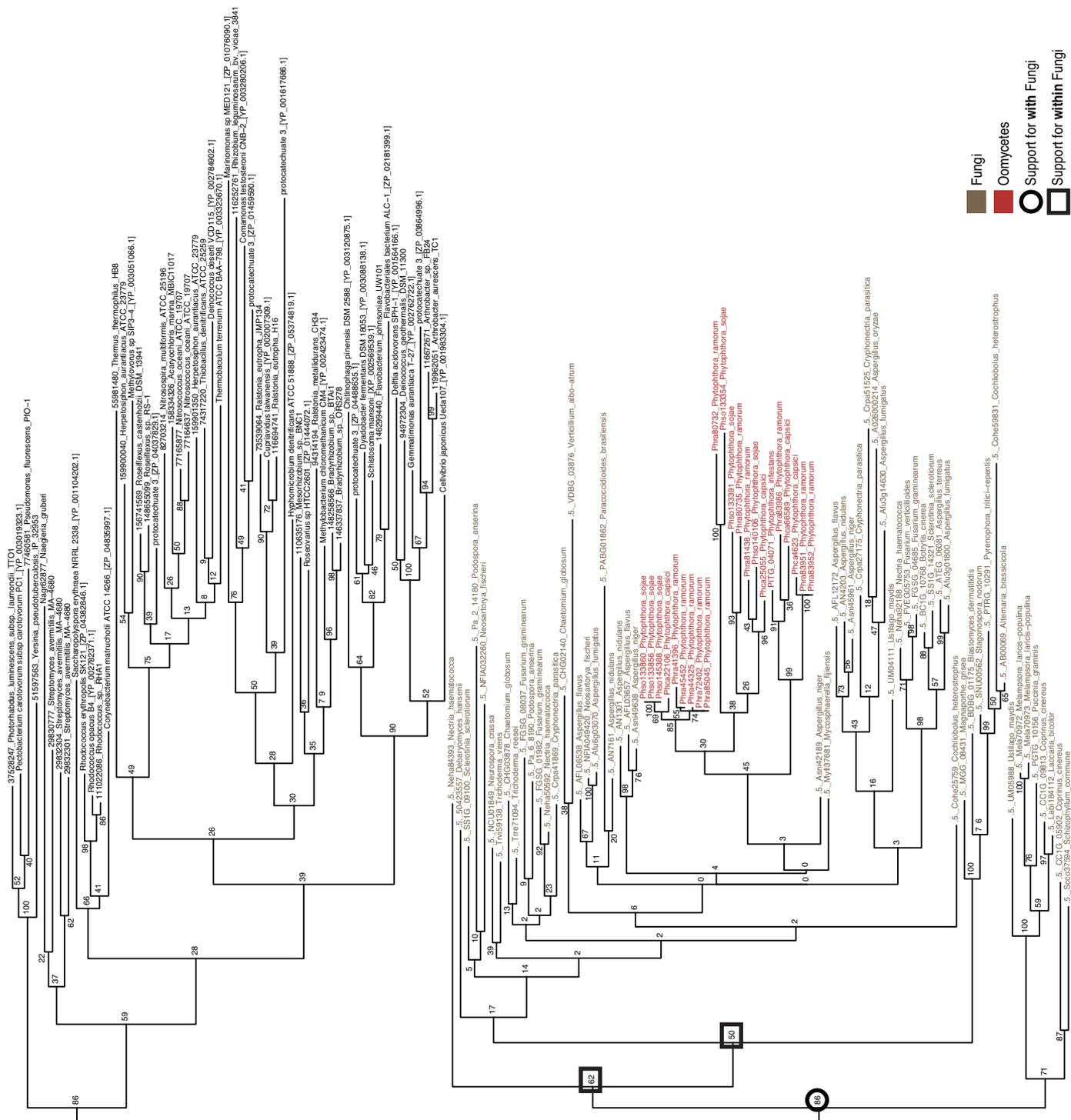


Fig. 1.22



Fungi
 Oomycetes
 Support for with Fungi
 Support for within Fungi

Fig. 1.23

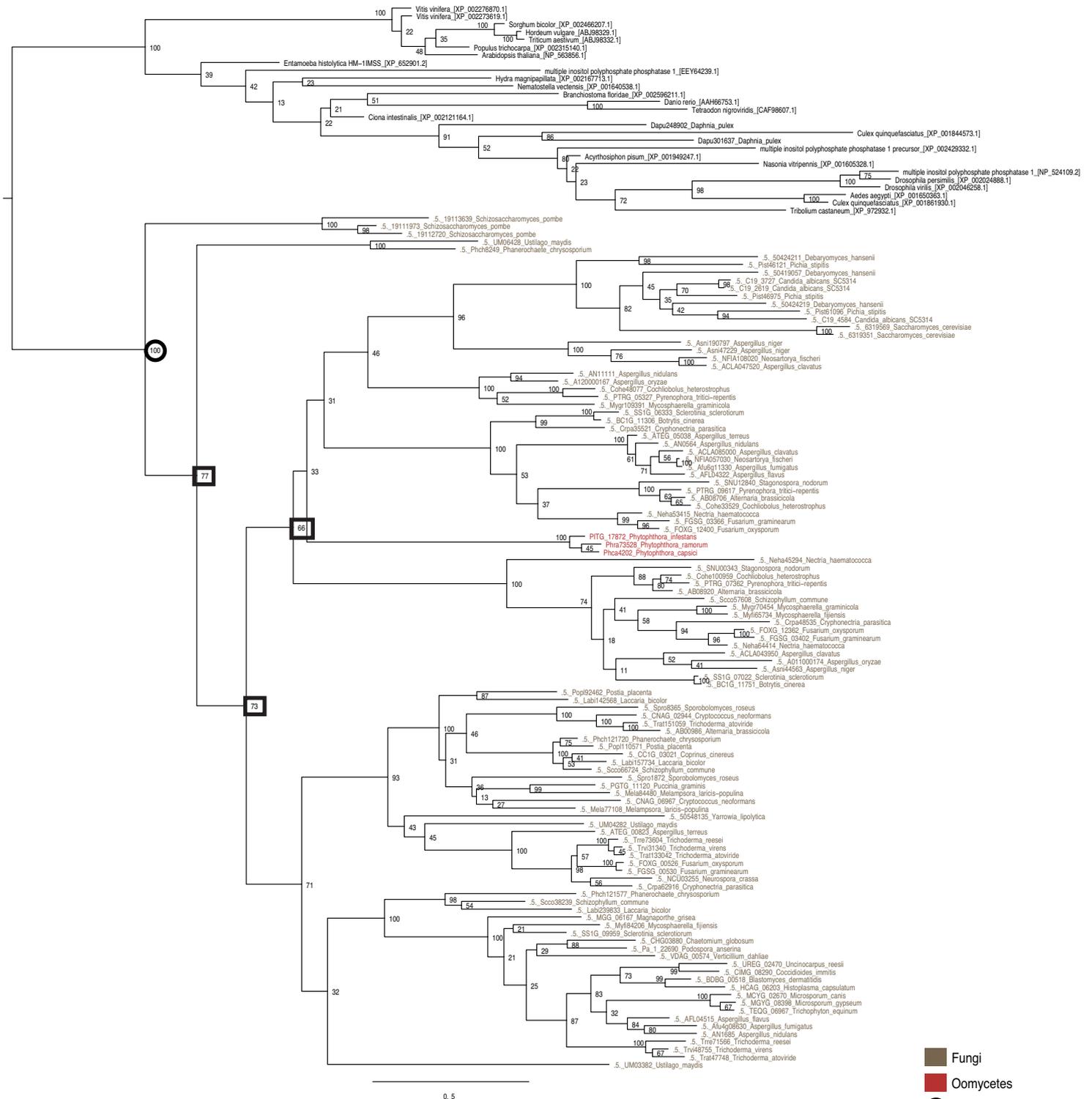


Fig. 1.25

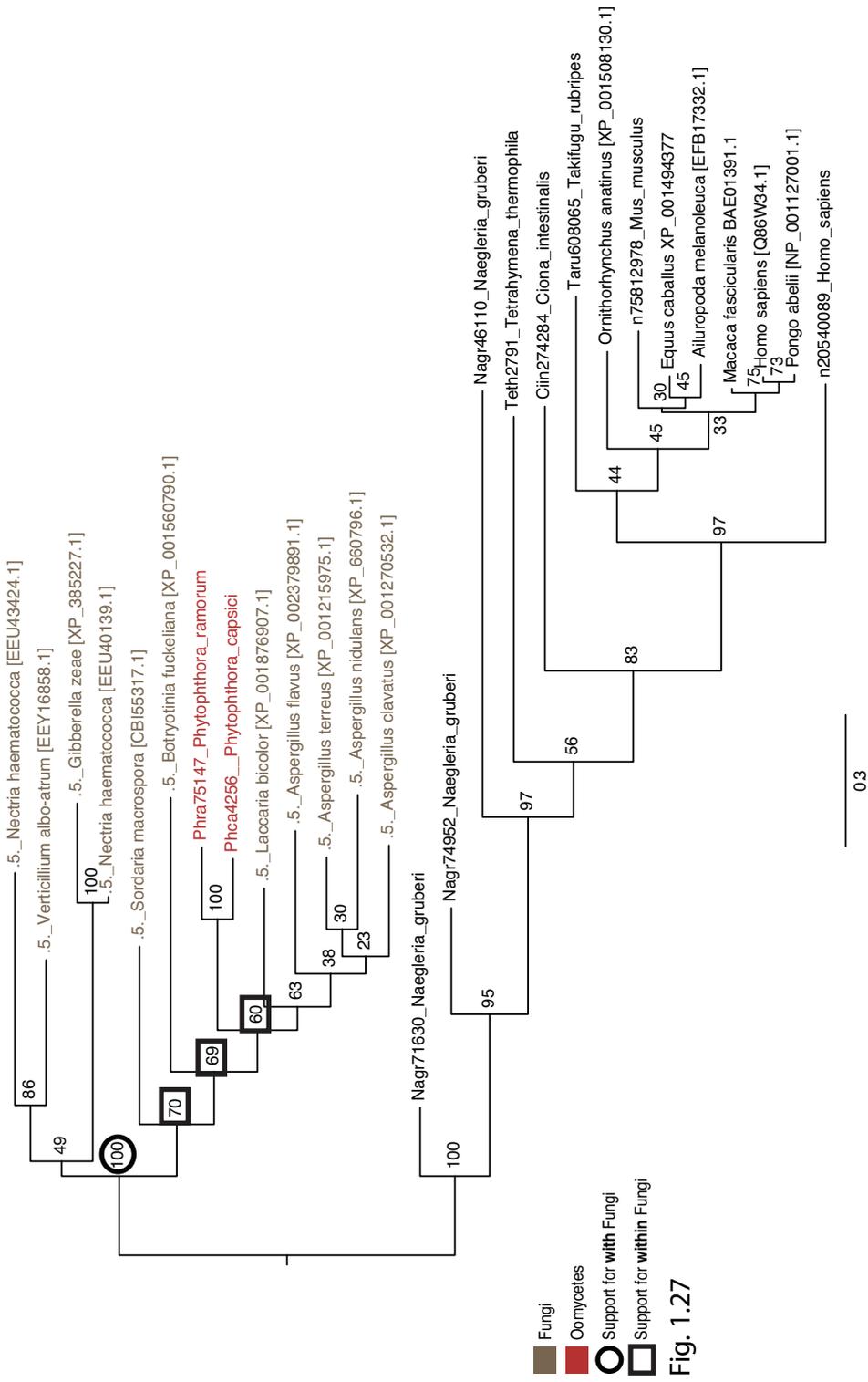


Fig. 1.27

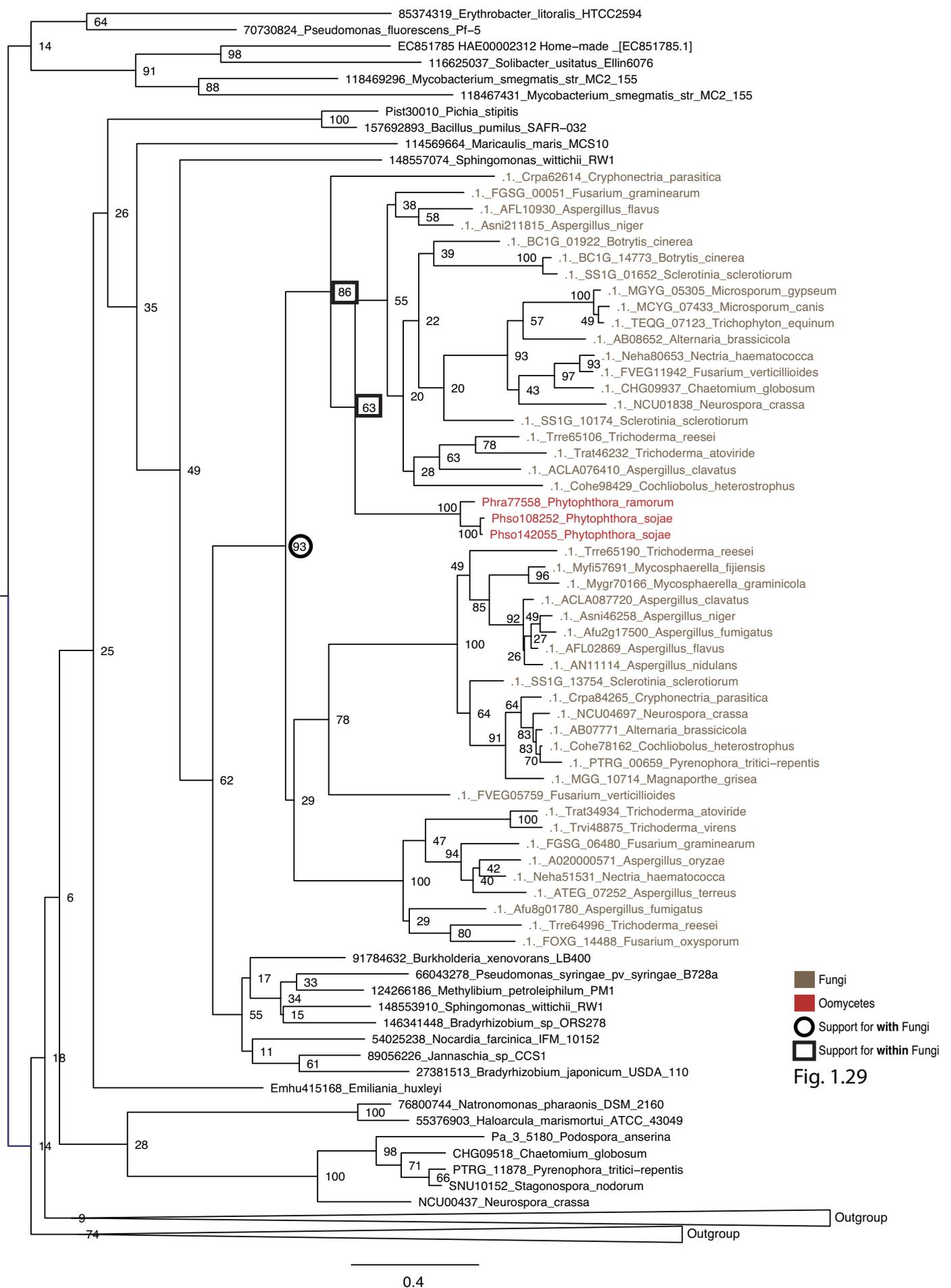


Fig. 1.29



Fig. 1.30