Supplementary Information for

Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants

Ying Chang, Sishuo Wang, Satoshi Sekimoto, Andrea L. Aerts, Cindy Choi, Alicia Clum, Kurt M. LaButti, Erika A. Lindquist, Chew Yee Ngan, Robin A. Ohm, Asaf A. Salamov, Igor V. Grigoriev, Joseph W. Spatafora, and Mary L. Berbee

This file includes:

Supplementary Online Material

Table of Contents

1. Material and Methods

1.1. Fungal strains and DNA/RNA extraction

Gonapodya prolifera JEL478 was inoculated into a 150 ml flask containing 50 ml PmTG broth (Longcore 2004) and incubated at ambient temperature with shaking at 150 rpm for 7 days, then transferred to new PmTG broth in 1 l flasks and incubated with shaking for 7 – 17 days. *Coemansia reversa* NRRL 1564 and *Conidiobolus coronatus* NRRL 28638 grown on potato dextrose agar were inoculated into 150 ml flasks containing 50 ml GYEP broth (ATCC Medium #1005; 20 g glucose, 10 g peptonized milk powder, and 5 g yeast extract in 1000 ml H_2O) and incubated at ambient temperature with shaking at 150 rpm for 2 -3 day, then transferred to new GYEP broth in multiple 0.5 l or 1.0 l flasks and incubated for 7 days.

Mycelia were harvested by filtering with Whatman #1 filters (GE Healthcare Bio-Sciences, NJ), frozen in liquid nitrogen and lyophilized, then ground with 0.1 mm Zirconia/Silica beads (BioSpec Products, Inc., Bartlesville, OK) in liquid nitrogen, and stored at -80 ºC until further processing. To prepare high molecular weight total DNA, lyophilized, ground mycelia were mixed with AP1 lysis buffer (DNeasy Plant Kit, Qiagen Inc. Toronto, ON) and incubated at 65 ºC for 30-60 minutes. The mixture was purified with phenol/chloroform then chloroform, precipitated with isopropanol and a high salt solution (0.8 M sodium citrate and 1.2 M sodium chloride) and resuspended in TE (10 mM Tris-HCl and 1 mM EDTA pH 8.0). After treatment with DNase-free RNase I (Lucigen, Middleton, WI), the extract was purified with chloroform, precipitated with isopropanol and the high salt solution, and resuspended in TE.

Total RNA was extracted from lyophilized ground mycelia with TRIzol Reagent (Life Technologies, Carlsbad, CA) using the manufacturer's instructions, treated with RQ1 RNase-free DNase (Promega, Madison, WI), purified with chloroform, precipitated with isopropanol and the

high salt solution, and resuspended in RNase-free water. After a preliminary check by electrophoresis on 1% TAE agarose gels, the purity and concentration of total RNA was assessed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) at CMMT/CFRI BioAnalyzer Core Facility, Faculty of Medicine, University of British Columbia.

1.2 Genome sequencing and assembly

The genome assemblies for fungi reported in this study were produced using several sequencing approaches and the corresponding assemblers. The *Gonapodya prolifera* JEL478 genome was sequenced using an Illumina (San Diego, CA) standard paired-end (PE) and long mate-pairs (LMP) libraries. For Illumina standard libraries, genomic DNA was sheared to 250-300 bp fragments using the Covaris E210 (Woburn MA), treated with end repair, A-tailing, ligated to adaptors, and sequenced in 2X150 bp read formats. The LMP library was generated (Peng, et al. 2012), where genomic DNA was sheared to 4 kb using the HydroShear® (Digilab, Inc., Marlborough, MA). The fragments were treated with end repair and ligation (NEB, New England BioLabs Inc., Ipswich, MA) of biotinylated adapters containing loxP (Integrated DNA Technologies Coralville, IA). The adapter ligated DNA fragments were circularized via recombination by a Cre excision reaction (NEB). The circularized DNA was digested using 4 base cutter restriction enzymes (NEB) followed by self-ligation and immobilization onto streptavidin beads (Invitrogen, Life Technologies, Grand Island, NY) followed by 12-15 cycles of inverse PCR (KAPA biosystems, Wilmington, MA). The final library was gel size selected for 300-600 bp and sequenced using 2X100 bp format on Illumina platform. Each FASTQ file was filtered for artifacts, process contamination, and subsequently assembled with ALLPATHS-LG (Gnerre, et al. 2011). The *G. prolifera* genome was assembled with ALLPATHS-LG v.R42328

using 100X PE and 50X LMP data and processed for submission using the ALLPATHS-LG SubmissionPrep module.

Conidiobolus coronatus and *Coemansia reversa* were sequenced using a 454 pyrosequencing (454 Life Sciences, a Roche company, Branford CT). For 454 standard libraries, genomic DNA was fragmented by nebulization to an average size of 500-800 bp. 454 LMP were produced using CRE-LoxP protocols as described above. Genomic DNA was sheared to its desired insert size (4 kb and 8 kb). Circularized DNA was randomly sheared using the Covaris E210, ligated with adapters, and amplified via 15-20 cycles of PCR. Sequence reads were produced using 454 Titanium chemistry. Each file was quality assessed, trimmed, screened for artifacts and process contamination, and subsequently and assembled with Newbler (v 2.5). For *C. reversa*, 454 data were also combined with shredded consensus from Velvet (Zerbino and Birney 2008) assembled Illumina data (2x75 bp reads, standard 300 bp insert size Illumina library). Improvement was performed with an in-house automated gap closure tool gapResolution (Trong, et al. 2009).

1.3 Genome annotation

The genome assemblies of *Gonapodya prolifera, Coemansia reversa* and *Conidiobolus coronatus* were annotated using the Pipeline JGI annotation (Grigoriev, et al. 2006), which combines several gene prediction and annotation methods, and integrates the annotated genomes into the web-based fungal resource MycoCosm (Grigoriev, et al. 2014), for comparative genomics. Before gene prediction, assembly scaffolds were masked using RepeatMasker (Smit, et al. 1996-2010) and RepBase library (Jurka, et al. 2005), with the most frequent (>150 times) repeats recognized by RepeatScout (Price, et al. 2005). The following combination of gene predictors was run on the masked assembly: *ab-initio* Fgenesh (Salamov and Solovyev 2000)

and GeneMark (Ter-Hovhannisyan, et al. 2008), trained for specific genomes; homology-based Fgenesh+ (Salamov and Solovyev 2000) and GeneWise (Birney, et al. 2004), seeded by BLASTx alignments (Altschul, et al. 1997) against the NCBI-NR protein database; and (for *G. prolifera* and *C. coronatus*) transcriptome-based CombEST (Zhou *et al*., personal communication). In addition to protein-coding genes, tRNAs were predicted using tRNAscan-SE (Lowe and Eddy 1997). All of the predicted proteins were functionally annotated using SignalP (Petersen, et al. 2011) for signal sequences, TMHMM (Krogh, et al. 2001) for transmembrane domains, InterProScan (Hunter, et al. 2009) for the integrated collection of functional and structured protein domains, and protein alignments to NCBI-NR, Swiss-Prot (http://www.expasy.org/sprot/), KEGG (Kanehisa, et al. 2008) for metabolic pathways, and KOG (Koonin, et al. 2004) for eukaryotic clusters of orthologs. InterPro and Swiss-Prot hits were also used to map the gene-ontology terms (Ashburner, et al. 2000). For each genomic locus, the best representative gene model was selected based on a combination of protein similarity and EST support.

1.4 Transcriptome sequencing and assembly

Illumina paired end sequencing was used for transcriptomes of *G. prolifera* and *C. coronatus*. First, total RNA from each species was used to generate stranded RNASeq libraries. The mRNA from each was purified from total RNA using an Absolutely mRNA™ purification kit (Stratagene, La Jolla, CA). The isolation procedure was repeated to ensure that the samples were free of rRNA. Subsequently, the mRNA samples were chemically fragmented to a size range of 200-250 bp using 1X fragmentation solution for 5 minutes at 70 ºC (RNA Fragmentation Reagents, AM8740 – Zn, Ambion® Foster City, CA). The first strand cDNA synthesis used Superscript II Reverse Transcriptase (Invitrogen) and random hexamers, followed

by purification with Ampure SPRI beads (Agencourt Bioscience, Beverly, MA). The second strand synthesis used a dNTP mix (with dUTP replacing dTTP) and *E. coli* RNase H, DNA ligase, and DNA polymerase I for nick translation. Two rounds of purification with Ampure SPRI beads selected for double stranded cDNA fragments in the range of 200-300 bp. The double stranded cDNA fragments were treated with end repair, A-tailing, and ligated with TruSeq adaptors using an Illumina DNA Sample Prep Kit, and then purified with Ampure SPRI beads. The second strand was removed by AmpErase UNG (Applied Biosystems®, Life Technologies, Grand Island, NY) (Parkhomchuk, et al. 2009). The digested cDNA was again purified with Ampure SPRI beads. Paired end sequencing reads were generated using an Illumina HiSeq System. For each species, *G. prolifera* and *C. coronatus, de novo* transcript contigs were assembled using Rnnotator (Martin, et al. 2010).

Total RNA from *Coemansia reversa* was used to generate a 454 RNA library. Poly A+ RNA was isolated from the total RNA using the Absolutely mRNA™ purification kit but then the mRNA was used to construct cDNA libraries using the cDNA rapid library preparation method outlined by the manufacturer (Roche 454 Life Sciences, Branford, CT, USA), followed by sequencing on the 454 GS-FLX platform. For the assembly of the *C. reversa* 454 sequence data, ribosomal RNA, low quality and low complexity reads were filtered out, and the remaining reads were assembled using Newbler (v2.3-PreRelease-6/30/2009) with default parameters.

1.5 Additional alignments: matrices used to test the effects of removing outgroups and altering data density

To evaluate the effect of remotely-related outgroups on the phylogeny, we deleted all the non-opisthokont species from the original unaligned sequence data and then repeated the aligning and masking process with Aliscore and Alicut. The resulting concatenated

'Opisthokonts matrix' included 32 taxa and 29,126 sites from 136 genes. To optimize data density, we applied MARE v.0.1.2 (Meyer and Misof 2010), using the All-w6 matrix as input and creating the new alignment 'MARE matrix' as the output. We constrained the taxa to keep all of the opisthokont relatives and the non-terrestrial fungi (i.e., *Rozella, Allomyces, Catenaria,* the core chytrids, *Piromyces*) in 'MARE matrix'. To retain a larger subset of data with relatively less information content than the default setting would allow, we set the alpha value as one. 'MARE matrix' consisted of 44 out of 136 genes that were present most consistently across taxa and 7,818 sites.

MARE optimizes overall data density but because we aimed to resolve the earliest splits in fungi, we also created a matrix ' Opisthokonts-minimum-missing' with the most complete possible data density for the taxa relevant to early splits. Based on the Opisthokonts matrix, we selected genes present in at least in one member of each of the subgroups, i.e., opisthokont outgroups, *Rozella,* Blastocladiomycota, Chytridiomycota, and terrestrial fungi. The resulting matrix contained 32 taxa and 7,244 amino acid sites from 33 genes.

Although the Opisthokonts-minimum-missing matrix and the MARE matrix included comparable numbers of aligned sites, they supported different divergence orders for the first few splits in fungi (Table S4). Like the more inclusive All-w6 and Opisthokonts matrices, the MARE matrix showed the Blastocladiomycota and Chytridiomycota as paraphyletic. The Opisthokonts-minimum-missing matrix, which specifically excluded genes that were missing for the early diverging *Rozella*, *Allomyces*, and *Catenaria*, showed Blastocladiomycota as sister to Chytridiomycota (ML bootstrap support 71%). The MARE matrix, compared with the Opisthokonts-minimum-missing dataset, was missing 26 genes from *Rozella* and 20 from both

Allomyces and *Catenaria*. In other words, the placement of *Rozella,* and Blastocladiomycota (*Allomyces* and *Catenaria*) was partially determined by genes that were absent from one or both of the two groups.

1.6 FcLM, AU and RADICAL tests for signal and conflict

For FcLM analyses, we first divided the sampled species into four sub-groups according to the node of interest (Table S3, Fig. S1). We then extracted the genes that were present in at least one representative from each of the four groups. We sampled a quartet of species with one species from each of the four sub-groups defined above and calculated the likelihood of the three competing topologies using Tree-Puzzle v5.2 (Schmidt, et al. 2002). Repeating the sampling process and likelihood calculation, we tried all the possible combinations of quartets of species. We then counted the number of quartets showing unambiguous support (see Strimmer and von Haeseler 1997 for detailed explanation) for each of the three possible topologies for each node of interest. Conflicting signal was revealed when different genes showed majority support for different topologies. Similarly, in each AU test we selected alignments that contained at least one taxon from each of the clades involved in the competing arrangements (Table S3). For each gene, we used the substitution model from ProtTest, found the best constrained tree and calculated corresponding site-wise log-likelihood using RAxML, and then carried out AU tests in CONSEL (Shimodaira and Hasegawa 2001).

We used random addition concatenation analysis with RADICAL to explore the relative branch support from the data (Narechania, et al. 2012). We applied RADICAL analysis on Opisthokonts matrix to generate a library of concatenated matrices, randomly sampling subsets of genes from a given set of data with a step parameter of ten. With this setting, in the first step,

RADICAL randomly sampled one gene. In the second step, it sampled and concatenated 10 genes. It increased sample sizes by 10 genes per step until the concatenation size reached 130. The final concatenation point was 136. The whole concatenation process was then repeated 100 times. For each of the newly generated concatenated matrices we infer the best ML tree using RAxML and then compared this tree to the ML tree based on the whole Opisthokonts matrix.

1.7 Alternative approaches to reconcile the gene phylogenies with the species tree

We used two parsimony approaches to reconciliation. The third approach used the model-based program JPrIME-DLRS (Sjöstrand, et al. 2012; Mahmudi, et al. 2013). A model-based method has the potential to better estimate gene duplications/losses along the species tree as its makes more realistic assumptions about gene evolution (Doyon, et al. 2011; Wu, et al. 2013). However, due to the deep divergence times in our species tree and to the complex evolutionary histories of the enzyme families, JPrIME-DLRS was unable to conduct an effective search of the tree space for the larger enzyme families (e.g., GH28). Here, for any combination of gene genealogy and parameter values, the likelihood density for the DLR model (model of gene Duplication and Loss and Rate heterogeneity) was so close to zero that proposed new states in the tree search process were constantly rejected, resulting in failure of the exploration of tree space. For the smaller datasets of CE8, GH53, PL3 and PL4, JPrIME-DLRS analyses was feasible. As input, we used the ML species tree with the node age estimates from r8s. We ran each JPrIME-DLRS analysis for a minimum of 20,000,000 generations, until the estimated sample sizes for all parameters were over 200. We calculated the number of duplication/loss events on each node by averaging the estimations from individual sampled realizations after a burn-in of 25%.

Bacterial pectinase sequences were only included in final reconciliation analyses of GH53, PL1, PL3 and PL4, where adding the bacteria changed the interpretation of eukaryotic gene evolution with bootstrap support. We calculated the number of gene copies and gene duplication/loss events for all the ancestral nodes of the species tree using 'phylogenomics' option in Notung 2.8. For the species tree, we added bacteria in as sister taxa to the eukaryotes because lack of overlap in appropriately informative genes precluded formal phylogenetic analysis. We set the date of divergence of bacteria from the eukaryotes at 4.2 billion years before present in JPrIME analysis (Hedges, et al. 2001).

We also tested the effects of poorly supported nodes from JPrIME on reconciliation in Notung. We first collapsed nodes in the JPrIME trees with support values below 0.98. Secondly, we collapsed nodes with support values below 0.999. The enzyme phylogenies for GH53, PL3 and PL4, JPrIME trees showed little or no resolution when nodes with weak support were collapsed into polytomies and we analyzed these no further. We used the JPrIME trees that still showed resolution after collapsing branches as input for reconciliation in Notung, allowing rearrangement of the polytomies to minimize the number of duplication and loss events.

1.8 Dating the origin of pectins

Phylogenies show that streptophytes algae are paraphyletic (Timme, et al. 2012; Chang and Graham 2014; Wickett, et al. 2014). Pectin-specific polysaccharides, and genes for their synthesis, are only found in the land plants and the algal species showing close affinity to land plants, including those from Zygnematales, Coleochaetales, Charales and Klebsormidiales (Sørensen, et al. 2011; Mikkelsen, et al. 2014). Earlier streptophytes, i.e., Mesostimatales and Chlorokybales, presumably lacked pectins and the synthesizing enzymes. Pectin synthesis in

plants evolved after the divergence of land plants from these early-diverging streptophyte algae. However, most molecular clock based studies of divergence dates among plants have relied on a limited sampling of the streptophytes outside of the land plants. A study by Magallón, et al. (2013) suggested an older date for the divergence of streptophyte algae from land plants, but this may be the result of a technical error; the prior ages that they assigned to the node came from the divergence of chlorophyte algae from land plants, a much older event than the divergence of streptophyte algae from land plants (Becker 2013). We acknowledge that estimates of divergence dates are imperfect. Future studies may suggest an age older or younger than 750 Ma to constrain the divergence of Chytridiomycota and Dikarya. This would make all the ages in the chronogram (Fig. 3) proportionally older or younger and, we would hope, closer to their true geological age.

Table S1. Species sampled and repositories of genomic data.

	Higher Taxonomy	**Species Name**	**Data Source**
Outgroups	Plantae	*Arabidopsis thaliana*	*Arabidopsis* Genome Initiative, NCBI
	Chromalveolata	*Aureococcus anophagefferens*	JGI
		Phytophthora infestans	Broad
		Saprolegnia parasitica	Broad
		Thalassiosira pseudonana	JGI
	Amoebozoa	*Dictyostelium discoideum*	DictyBase
		Entamoeba histolytica	J. Craig Venter Institute
	Apusozoa	*Thecamonas trahens*	Broad, Origins of Multicellularity
	Opisthokonta	*Capsaspora owczarzaki*	Broad, Origins of Multicellularity
		Drosophila melanogaster	Flybase
		Monosiga brevicollis	JGI
		Salpingoeca rosetta	Broad, Origins of Multicellularity
		Sphaeroforma arctica	Broad, Origins of Multicellularity
Fungi	Cryptomycota	*Rozella allomycis*	University of Michigan, NCBI
	Blastocladiomycota	*Allomyces macrogynus*	Broad, Origins of Multicellularity
		Catenaria anguillulae	JGI
	Chytridiomycota	*Batrachochytrium dendrobatidis*	Broad
		Gonapodya prolifera	JGI

		Piromyces E2	JGI
		Spizellomyces punctatus	Broad, Origins of Multicellularity
	Zygomycota I	*Mortierella verticillata*	Broad, Origins of Multicellularity
		Mucor circinelloides	JGI
		Phycomyces blakesleeanus	JGI
		Rhizophagus irregularis	JGI
		Rhizopus oryzae	Broad
	Zygomycota II	*Coemansia reversa*	JGI
		Conidiobolus coronatus	JGI
	Ascomycota	*Aspergillus niger*	JGI
		Candida albicans	Broad
		Coccidioides immitis	Broad
		Metarhizium anisopliae	Laboratório Nacional de Computação Científica, NCBI
		Orbilia auricolor	Laboratory for Conservation and Utilization of Bio-resources, Yunnan University NCBI
		Saccharomyces cerevisiae	JGI
		Schizosaccharomyces pombe	Broad
		Tuber melanosporum	Genoscope

	Basidiomycota	*Cryptococcus neoformans*	Stanford Genome Technology Center
		Malassezia globosa	JGI
		Phanerochaete chrysosporium	JGI
		Puccinia graminis	Broad
		Ustilago maydis	Broad

Table S2. Assembly and Annotation Statistics

	Coemansia reversa	*Gonapodya prolifera*	*Conidiobolus coronatus*
Assembly statistics			
Assembly length (Mbp)	21.8	48.79	39.9
Contig length total (Mbp)	20.6	48.28	31.72
Number of contigs	1063	753	7809
Contig N50	81	78	1663
Contig L50 (bp)	64542	169758	5408
Number of scaffolds	346	352	1050
Scaffold N50	21	42	113
Scaffold L50 (bp)	347177	347324	102411
Number of scaffold gaps	717	401	6759
Scaffolds gaps length (bp)	1192887	511418	8185666
Percentage of scaffolds in gaps (%)	5.46	1.05	20.51
Number of repeat-covered regions	7115	6665	57594
Length of repeat-covered regions (bp)	3260529	1406083	4178337
Percentage assembly covered by repeats (%)	14.93	2.88	10.47
GC content (%)	45.18	52.41	27.65
Gene statistics			
Number of genes	7347	13902	10635

Protein length (amino acids, median)	324	334	287
Exon length (bp, median)	614	157	185
Gene length (bp, median)	1170	1541	1049
Transcript length (bp, median)	1074	1147	928
Intron length (bp, median)	135	68	52
Number of genes with intron	2388	12451	5979
Percentage of genes with an intron	32.5	89.56	56.22
Introns per gapped gene (median)	1	4	2
Functional annotations			
Genes with KEGG annotation [n, (%)]	5634 (76.68%)	9657 (69.46%)	7215 (67.84%)
Genes with KOG annotation [n, (%)]	5319 (72.40%)	8888 (63.93%)	6653 (62.56%)
Genes with Swissprot hit [n, (%)]	5393 (73.40%)	9056 (65.14%)	6813 (64.06%)
Genes with Pfam domain [n, (%)]	4440 (60.43%)	7419 (53.37%)	5652 (53.15%)
Genes with transmembrane domain [n, (%)]	1216 (16.55%)	2665 (19.17%)	2132 (20.05%)
Unique PFAM domains	2172	2348	2057

Table S3. Grouping schemes used in tests of support for alternative tree topologies.

Node of interest	Taxa tested in all three possible topologies using Four-cluster Likelihood Mapping analysis	Constraints used in Approximately Unbiased tests
Placement of *Rozella*	*Rozella*; *Catenaria* + *Allomyces*; Other fungi; Other opisthokonts.	*Rozella* sister to other fungi; *Rozella* sister to Chytridiomycota; *Rozella* sister to Blastocladiomycota; *Rozella* sister to (zygomycota I, zygomycota II, Ascomycota and Basidiomycota); Blastocladiomycota splits first from all other fungi.
Placement of *Catenaria* + *Allomyces*	*Rozella* + other opisthokonts; *Catenaria* + *Allomyces*; Chytridiomycota; zygomycota I, zygomycota II, Ascomycota and Basidiomycota.	Blastocladiomycota sister to (Chytridiomycota, zygomycota I, zygomycota II, Ascomycota and Basidiomycota); Blastocladiomycota sister to Chytridiomycota; Blastocladiomycota sister to (zygomycota I, zygomycota II, Ascomycota and Basidiomycota); (Blastocladiomycota sister to *Rozella*)

Table S3 cont.

Placement of zygomycota II	Other opisthokonts, Cryptomycota, Blastocladiomycota, Chytridiomycota; *Conidiobolus, Coemansia*; Mucorales, *Mortierella*, *Rhizophagus*; Ascomycota and Basidiomycota.	(*Conidiobolus*, *Coemansia*) sister to (Ascomycota, Basidiomycota, *Mortierella*, Mucorales, *Rhizophagus*); (*Conidiobolus*, *Coemansia*) sister to (Ascomycota and Basidiomycota); (*Conidiobolus*, *Coemansia*) sister to (*Mortierella*, Mucorales, *Rhizophagus*);
Placement of *Rhizophagus*	*Rhizophagus*; *Mortierella*; Mucorales; Other taxa.	*Rhizophagus* sister to *Mortierella*; *Mortierella* sister to other Mucorales; *Rhizophagus* sister to (Ascomycota and Basidiomycota).

Table S4. Support for branching order among Fungi was strong and consistent across three alternative analytical methods and four alternative matrices of concatenated loci.

	Method	RAxML				MrBayes	P4[¶]		
							a) no polytomy	b) polytomy1	c) polytomy 2
	Matrix name	All-w6*	Opistho-konts*†	MARE†‡	Opistho-konts-minimum-missing†[§]	All-w6*	All-w6*, Recoded Dayhoff		
	Description of alignment	40 taxa 136 genes 28,807 sites	32 taxa 136 genes 29,126 sites	32 taxa 44 genes 7,818 sites	32 taxa 33 genes 7,243 sites	40 taxa 136 genes 28,807 sites	40 taxa 136 genes 28,807 sites	40 taxa 136 genes 28,807 sites	40 taxa 136 genes 28,807 sites
a	Monophyly of Fungi	100	81	88	78	1.0	1.0	1.0	1.0
b[#]	Fungi excl. *Rozella*	89	100	95	48	1.0	0.92	0.75	0.55
c[#]	Fungi excl. *Rozella* and Blastocladiomycota	83	83	100	-	1.0	0.99	0.99	0.99
d[#]	[Chytridiomycota + *Piromyces*]	93	91	97	100	1.0	1.0	1.0	1.0
e	(zygomycota I, zygomycota II,	100	100	98	90	1.0	1.0	1.0	1.0

	Ascomycota and Basidiomycota)								
f	Monophyly of zygomycota I	100	100	99	97	1.0	1.0	1.0	1.0
g[#]	[*Rhizophagus* + *Mortierella*]	96	89	-	58	1.0	1.0	1.0	1.0
g2	[*Mortierella* + Mucorales]	-	-	100	-	-	-	-	-
h	(zygomycota I, Ascomycota + Basidiomycota)	89	77	97	31	1.0	1.0	1.0	1.0
i	[Ascomycota + Basidiomycota]	100	100	100	100	1.0	1.0	1.0	1.0

*Matrices All-w6 and Opisthokonts include all available genes.

[†]Excludes non-Opisthokont outgroups.

[‡]For the opisthokonts matrix, MARE increases data density by excluding taxa and sites that exceed a missing data threshold.

[§]Opisthokonts-minimum-missing preferentially includes genes from the opisthokont outgroups, *Rozella,* Blastocladiomycota, Chytridiomycota, and terrestrial fungi.

[¶]The P4 Bayesian analyses show the effects of (a) allowing only dichotomous trees; (b) permitting polytomies, polytomy prior as 1 and LogC value as 1; and (c) permitting polytomies, polytomy prior as 1 and LogC value as 2.

[#]Asterisks indicates the nodes tested in Approximately Unbiased and Four-cluster Likelihood Mapping analyses.

Table S5. Summary of results of Four Taxon Likelihood Mapping tests on individual genes, regarding the placement of certain taxa of interest.

Arrangements tested*	# genes tested†	# of genes showing majority support for one of three candidate topologies
Placement of *Rozella*	34	0
Placement of Blastocladiales	62	2
Placement of zygomycetes II	126	1
Placement of *Rhizophagus*	107	107

*See **Table S3** for the specific arrangements tested.

†The number is smaller than the total number of genes (136) because some genes were absent from one or more taxa required for the tests.

Table S6. Few if any individual genes offered significant support to any possible candidate topology in Approximately Unbiased Tests of early branching order in Fungi.

Arrangements tested*	# genes tested†	# of genes showing significant support for any one candidate topologies ($p > 0.95$)
Placement of *Rozella*	34	0
Placement of Blastocladiomycota	62	2
Placement of zygomycetes II	126	4
Placement of *Rhizophagus*	105	2

*See **Table S3** for the specific topologies tested.

†The number is smaller than the total number of genes (136) because some genes were absent from one or more taxa required for the AU test.

Table S7. Growth of both the Chytridiomycota fungus *Gonapodya prolifera* and the Ascomycota species *Aspergillus niger* were significantly greater in medium that contained a pectin, either polygalacturonic acid (PGA) or rhamnogalacturonan I (RGI), compared with growth in the low sugar control medium.

Species		Medium			
		1/4mPmT Low sugar control	1/4mPmT + Glucose	1/4mPmT + PGA*	1/4mPmT + RGI†
Aspergillus niger	Mean*	6.0	8.7†	9.3†	10.7†
	SE	0.00	0.33	0.33	0.33
Gonapodya prolifera	Mean	6.0	6.7	8.0†	8.7†
	SE	0.00	0.33	0.00	0.33
Allomyces sp.	Mean	6.0	6.3	5.0	6.3
	SE	0.00	0.33	0.00	0.33

*Means are the mean mycelial ball diameters in mm from three replicates.

†These means differed significantly from those of the low sugar controls, based on two-tailed t-tests at $P < 0.05$.

Table S8. Gene numbers in pectinase families with names of pectin-specific families in bold. The pectin-specific families GH93 and PL11 were not included as they were not detected among the fungal species sampled in this study.

	Non-terrestrial fungi							zygo-mycota II		zygomycota I					Basidiomycota					Ascomycota							
species (fungi)*	Ra	Am	Ca	Bd	Gp	Sp	Pi	Cr	Cc	Ri	Mv	Mc	Pb	Ro	Cn	Mg	Pc	Pg	Um	An	CA	Ci	Ma	Oa	Sc	SP	Tm
Total	0	8	1	1	33	1	87	2	1	2	3	3	14	27	8	1	22	14	8	74	0	1	13	56	1	0	10
GH28	0	4	0	0	14	0	1	0	0	0	1	2	8	18	1	0	4	1	1	22	0	0	1	7	1	0	2
GH53	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0
PL1	0	2	0	0	6	0	14	0	0	0	0	0	0	0	0	0	0	2	1	8	0	0	0	7	0	0	2
PL3	0	0	1	0	3	0	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
PL4	0	0	0	0	0	0	9	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	1	0	0	1
CE8	0	0	0	0	3	0	8	0	0	0	0	0	4	6	0	0	2	7	1	3	0	0	0	5	0	0	1
CE13	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH2	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	4	0	0	3	0	0	0	0
GH35	0	0	0	1	0	0	0	2	1	2	1	1	1	1	0	0	3	1	1	5	0	0	3	2	0	0	0

GH43	0	2	0	0	1	0	28	0	0	0	0	0	0	2	0	1	4	2	0	10	0	1	1	9	0	0	1
GH51	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	0	2	4	0	0	0	1	0	0	0
GH54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0
GH78	0	0	0	0	1	0	1	0	0	0	0	0	0	0	2	0	1	0	0	7	0	0	0	5	0	0	2
GH88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0
GH105	0	0	0	0	2	0	2	0	0	0	0	0	0	0	1	0	0	0	1	2	0	0	3	2	0	0	0
PL9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
CE1	0	0	0	0	0	0	8	0	0	0	0	0	0	0	1	0	4	0	1	2	0	0	1	8	0	0	0
CE12	0	0	0	0	1	0	7	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3	0	0	1

*Abbreviations for fungal species: Am=*Allomyces macrogynus*, An=*Aspergillus niger*, Bd=*Batrachochytrium dendrobatidis*, CA=*Candida albicans*, Ca=*Catenaria anguillulae*, Cc=*Conidiobolus coronatus*, Ci=*Coccidioides immitis*, Cn=*Cryptococcus neoformans*, Cr=*Coemansia reversa*, Gp=*Gonapodya prolifera*, Ma=*Metarhizium anisopliae*, Mc=*Mucor circinelloides*, Mg=*Malassezia globosa*, Mv=*Mortierella verticillata*, Oa=*Orbilia auricolor*, Pb=*Phycomyces blakesleeanus*, Pc=*Phanerochaete chrysosporium*, Pg=*Puccinia graminis*, Pi=*Piromyces* E2, Ra=*Rozella allomycis*, Ri=*Rhizophagus irregularis*, Ro=*Rhizopus oryzae*, Sc=*Saccharomyces cerevisiae*, SP=*Schizosaccharomyces pombe*, Sp=*Spizellomyces punctatus*, Tm=*Tuber melanosporum*, Um=*Ustilago maydis*.

Table S8 cont.

	Non-Opisthokonts								Opisthokonts				
species (non-fungal)‡	At	Aa	Pi	sp	Tp	Dd	Eh	TT	Dm	Mb	Sr	Co	Sa
Total	**230**	**27**	**104**	**7**	**6**	**2**	**0**	**1**	**11**	**27**	**12**	**3**	**7**
GH28	73	6	21	2	0	0	0	1	0	5	2	1	3
GH53	0	0	3	0	0	0	0	0	0	0	0	0	0
PL1	28	1	19	5	0	0	0	0	0	0	0	0	0
PL3	0	0	32	0	0	0	0	0	0	0	0	0	2
PL4	9	0	4	0	0	0	0	0	0	0	0	0	0
CE8	68	0	12	0	0	0	0	0	0	0	0	0	0
CE13	20	5	0	0	0	0	0	0	1	8	4	0	0
GH2	3	4	0	0	1	0	0	0	7	3	2	0	0
GH35	24	3	1	0	0	2	0	0	3	2	0	2	0
GH43	3	3	3	0	2	0	0	0	0	2	1	0	0

GH51	2	0	0	0	0	0	0	0	0	1	0	0	0
GH54	0	5	1	0	0	0	0	0	0	0	0	0	0
GH78	0	0	5	0	3	0	0	0	0	5	2	0	0
GH88	0	0	0	0	0	0	0	0	0	1	0	0	2
GH105	0	0	2	0	0	0	0	0	0	0	1	0	0
PL9	0	0	0	0	0	0	0	0	0	0	0	0	0
CE1	0	0	0	0	0	0	0	0	0	0	0	0	0
CE12	0	0	1	0	0	0	0	0	0	0	0	0	0

‡Abbreviations for non-fungal species At=*Arabidopsis thaliana*, Aa=*Aureococcus anophagefferens*, Co=*Capsaspora owczarzaki*, Dd=*Dictyostelium discoideum*, Dm=*Drosophila melanogaster*, Eh=*Entamoeba histolytica*, Mb=*Monosiga brevicollis*, Pi=*Phytophthora infestans*, Sa=*Sphaeroforma arctica*, sp=*Saprolegnia parasitica*, Sr=*Salpingoeca rosetta*, Tp=*Thalassiosira pseudonana*, and TT=*Thecamonas trahens*

Supplementary References

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.
Ashburner M, et al. 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25: 25-29.
Becker B 2013. Snow ball earth and the split of Streptophyta and Chlorophyta. Trends Plant Sci 18: 180-183.
Birney E, Clamp M, Durbin R 2004. GeneWise and genomewise. Genome Res 14: 988-995.
Chang Y, Graham SW 2014. Patterns of clade support across the major lineages of moss phylogeny. Cladistics 30: 590-606.
Doyon J-P, Ranwez V, Daubin V, Berry V 2011. Models, algorithms and programs for phylogeny reconciliation. Briefings in Bioinformatics 12: 392-400.
Gnerre S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci USA 108: 1513-1518.
Grigoriev IV, Martinez DA, Salamov AA. 2006. Fungal Genomic Annotation. In: Arora DK, Berka RM, Singh GB, editors. Applied Mycology and Biotechnology, Vol 6: Bioinformatics. p. 123-142.
Grigoriev IV, et al. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42: D699-D704.
Hedges SB, et al. 2001. A genomic timescale for the origin of eukaryotes. BMC Evol Biol 1: 4.
Hunter S, et al. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res 37: D211-D215.
Jurka J, et al. 2005. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110: 462-467.
Kanehisa M, et al. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res 36: D480-D484.
Koonin EV, et al. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biology 5.
Krogh A, Larsson B, von Heijne G, Sonnhammer ELL 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 305: 567-580.
Longcore JE 2004. *Rhizophydium brooksianum* sp nov., a multipored chytrid from soil. Mycologia 96: 162-171.
Lowe TM, Eddy SR 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955-964.
Magallón S, Hilu KW, Quandt D 2013. Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Am J Bot 100: 556-573.
Mahmudi O, Sjostrand J, Sennblad B, Lagergren J 2013. Genome-wide probabilistic reconciliation analysis across vertebrates. BMC Bioinformatics 14: S10.
Martin J, et al. 2010. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics 11: 663.

Meyer B, Misof B. 2010. MARE: matrix reduction—a tool to select optimized data subsets from supermatrices for phylogenetic inference. Bonn (Germany): Zentrum für Molekulare Biodiversitätsforschung (zmb) am ZFMK.
Mikkelsen MD, et al. 2014. Evidence for land plant cell wall biosynthetic mechanisms in charophyte green algae. Ann Bot 114: 1217-1236.
Narechania A, et al. 2012. Random Addition Concatenation Analysis: A novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the Cyanobacteria. Genome Biol Evol 4: 30-43.
Parkhomchuk D, et al. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37.
Peng Z, et al. 2012. Generation of Long Insert Pairs Using a Cre-LoxP Inverse PCR Approach. PLoS ONE 7: e29437.
Petersen TN, Brunak S, von Heijne G, Nielsen H 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785-786.
Price AL, Jones NC, Pevzner PA 2005. De novo identification of repeat families in large genomes. Bioinformatics 21: I351-I358.
Salamov AA, Solovyev VV 2000. Ab initio gene finding in *Drosophila* genomic DNA. Genome Res 10: 516-522.
Schmidt HA, Strimmer K, Vingron M, von Haeseler A 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504.
Shimodaira H, Hasegawa M 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17: 1246-1247.
Sjöstrand J, Sennblad B, Arvestad L, Lagergren J 2012. DLRS: gene tree evolution in light of a species tree. Bioinformatics 28: 2994-2995.
RepeatMasker Open-3.0. [Internet]. 1996-2010. Available from: Http://www.repeatmasker.org
Sørensen I, et al. 2011. The charophycean green algae provide insights into the early origins of plant cell walls. Plant J 68: 201-211.
Strimmer K, von Haeseler A 1997. Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 94: 6815-6819.
Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18: 1979-1990.
Timme RE, Bachvaroff TR, Delwiche CF 2012. Broad Phylogenomic Sampling and the Sister Lineage of Land Plants. PLoS ONE 7: e29696.
Trong S, et al. 2009. GapResolution: A software package for improving Newbler genome assemblies. 4th Annual Meeting on Sequencing Finishing, Analysis in the Future; May 27-29 2009; Santa Fe.
Wickett NJ, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci USA 111: E4859-E4868.
Wu Y-C, Rasmussen MD, Bansal MS, Kellis M 2013. TreeFix: Statistically Informed Gene Tree Error Correction Using Species Trees. Syst Biol 62: 110-120.
Zerbino DR, Birney E 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821-829.