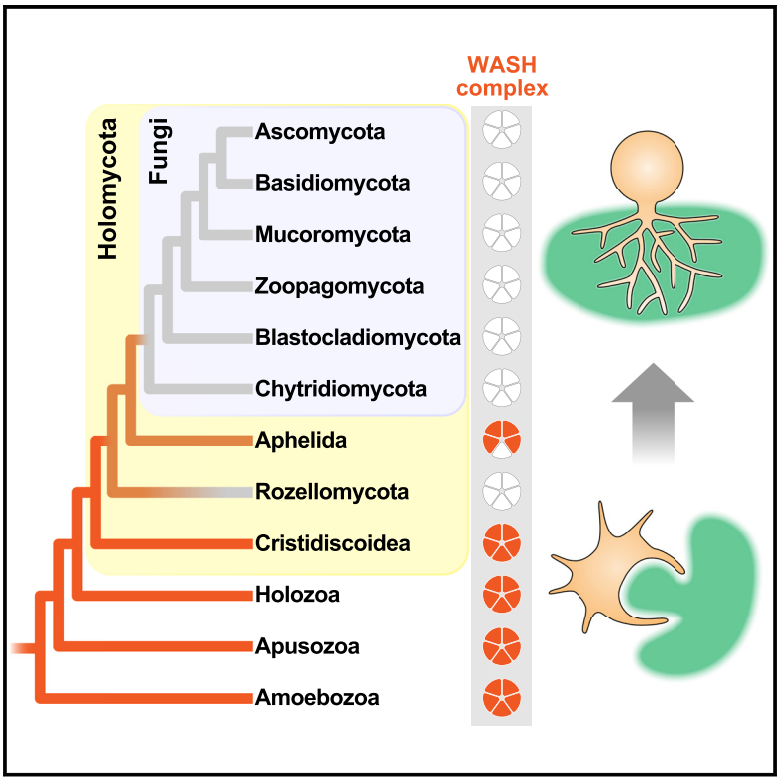
Article

Current Biology

Genomic analysis reveals cryptic diversity in aphelids and sheds light on the emergence of Fungi

Graphical abstract



Authors

Kirill V. Mikhailov, Sergey A. Karpov, Peter M. Letcher, ..., Dmitry Y. Sherbakov, Yuri V. Panchin, Vladimir V. Aleoshin

Correspondence

kv.mikhailov@belozersky.msu.ru

In brief

Mikhailov et al. report the genomes of *Amoeboaphelidium* species, revealing vast divergence within the phylum Aphelida and confirming their close phylogenetic relationship to Fungi. Genomic analyses show the conservation of phagocytic proteins, expansions of receptor-like kinases, and evidence for horizontally transferred genes in aphelids.

Highlights

- Strains of *A. protococcarum* have a complex evolutionary history involving hybridization
- Phylogenomic analysis confirms a sister relationship between Aphelida and Fungi
- Aphelids demonstrate independent reduction of flagella in two aphelid lineages
- Aphelids retain key genes involved in phagocytic activity that were lost in Fungi

Mikhailov et al., 2022, Current Biology *32*, 4607–4619
November 7, 2022
https://doi.org/10.1016/j.cub.2022.08.071

CellPress

Current Biology

CellPress

Article

Genomic analysis reveals cryptic diversity in aphelids and sheds light on the emergence of Fungi

Kirill V. Mikhailov,[1,2,12,*] Sergey A. Karpov,[3,4] Peter M. Letcher,[5] Philip A. Lee,[6] Maria D. Logacheva,[1,2,7] Aleksey A. Penin,[2] Maksim A. Nesterenko,[3,4] Igor R. Pozdnyakov,[3] Evgenii V. Potapenko,[8,9] Dmitry Y. Sherbakov,[10,11] Yuri V. Panchin,[1,2] and Vladimir V. Aleoshin[1,2]

[1]Belozersky Institute for Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119992, Russian Federation
[2]Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127051, Russian Federation
[3]Zoological Institute, Russian Academy of Sciences, St. Petersburg 199034, Russian Federation
[4]Biological Faculty, St. Petersburg State University, St. Petersburg 199034, Russian Federation
[5]Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL 35487-0344, USA
[6]Allegheny Science and Technology, Bridgeport, WV 26330, USA
[7]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow 121205, Russian Federation
[8]Institute of Evolution, University of Haifa, Haifa 3498838, Israel
[9]Department of Evolutionary and Environmental Biology, University of Haifa, Haifa 3498838, Israel
[10]Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk 664033, Russian Federation
[11]Novosibirsk State University, Novosibirsk 630090, Russian Federation
[12]Lead contact
*Correspondence: kv.mikhailov@belozersky.msu.ru
https://doi.org/10.1016/j.cub.2022.08.071

SUMMARY

Over the past decade, molecular phylogenetics has reshaped our understanding of the fungal tree of life by unraveling a hitherto elusive diversity of the protistan relatives of Fungi. Aphelida constitutes one of these novel deep branches that precede the emergence of osmotrophic fungal lifestyle and hold particular significance as the pathogens of algae. Here, we obtain and analyze the genomes of aphelid species *Amoeboaphelidium protococcarum* and *Amoeboaphelidium occidentale*. Genomic data unmask the vast divergence between these species, hidden behind their morphological similarity, and reveal hybrid genomes with a complex evolutionary history in two strains of *A. protococcarum*. We confirm the proposed sister relationship between Aphelida and Fungi using phylogenomic analysis and chart the reduction of characteristic proteins involved in phagocytic activity in the evolution of Holomycota. Annotation of aphelid genomes demonstrates the retention of actin nucleation-promoting complexes associated with phagocytosis and amoeboid motility and also reveals a conspicuous expansion of receptor-like protein kinases, uncharacteristic of fungal lineages. We find that aphelids possess multiple carbohydrate-processing enzymes that are involved in fungal cell wall synthesis but do not display rich complements of algal cell-wall-processing enzymes, suggesting an independent origin of fungal plant-degrading capabilities. Aphelid genomes show that the emergence of Fungi from phagotrophic ancestors relied on a common cell wall synthetic machinery but required a different set of proteins for digestion and interaction with the environment.

INTRODUCTION

Aphelids are a group of obligate intracellular parasitoids of algae that is closely related to the kingdom Fungi.[1,2] They are characterized by a life cycle similar to that of zoosporic fungi, but with a phagotrophic amoeba in the vegetative stage, as opposed to osmotrophic rhizoids.[3] The motile zoospores, which may be flagellated or amoeboid depending on the aphelid species, infect algal cells by encysting on their surface and puncturing the cell wall with a penetration tube. The parasitoid then migrates inside the host through the penetration tube and matures into a multinuclear plasmodium by phagocytizing the host cytoplasm until subsequent division and release of numerous zoospores.[4] The group comprises around a dozen described species in four genera, *Aphelidium*, *Amoeboaphelidium*, *Paraphelidium*, and *Pseudaphelidium*, while the environmental sequence data expose substantial unexplored biodiversity within the group.[3,5,6] The described aphelids are known predominantly from freshwater habitats where they infect various green, yellow-green, and diatom algae.[2,3] These parasitic organisms have a strong effect on the development of algae in water bodies, regulating their abundance, and can adversely affect aquacultures.[7]

Aphelids have been a subject of recent revisions in the high-level classification of Holomycota.[1,8,9] The initial molecular phylogenetic analyses hinted at the existence of a monophyletic group uniting aphelids with another prominent lineage of

Table 1. Genome assembly and annotation characteristics

	A. protococcarum X5	*A. protococcarum* FD95	*A. occidentale* FD01
Assembly size (Mb)	23.7	24.7	13.6
Number of scaffolds	348	258	946
Scaffold N50 (kb)	778	2,170	74
Largest scaffold (kb)	2,509	3,250	366
GC content (%)	40.5	40.5	40.0
Protein-coding genes	12,712	13,180	7,492
Coding sequence (%)	75.7	75.9	79.7
Gene density (genes/Mb)	537	533	553
Intron density (introns/gene)	0.47	0.46	1.17
Mean intron length (bp)	100.3	99.6	64.9
Mean exon length (bp)	962	977	666
Mean intergenic length (bp)	394	400	263
BUSCO completeness[a] (%)	92.1	92.7	91.4

[a]Evaluated using the eukaryota_odb9 dataset

parasitic early-diverging fungi-like protists, Rozellomycota or Cryptomycota including Microsporidia, leading to the proposition of a novel superphylum Opisthosporidia to house these organisms.[3] More recent analyses using ribosomal RNA genes and phylogenomic datasets featuring aphelid *Paraphelidium tribonemae* have opposed the monophyly of Opisthosporidia.[1,10,11] These analyses have recovered strong support for the sister relationship between Aphelida and true Fungi, highlighting the crucial significance of aphelids for the evolutionary studies of the emergence of Fungi.

Here, we present the genomic data for three strains of *Amoeboaphelidium* species: *Amoeboaphelidium protococcarum* (*A. protococcarum*) strains X5 and FD95 and *Amoeboaphelidium occidentale* (*A. occidentale*) strain FD01. Analysis of the genomic data accentuates the high level of divergence between the aphelid species and uncovers an intertwined evolutionary history of hybrid strains of *A. protococcarum*. We annotate the aphelid genomes and explore their kinomes and repertoires of carbohydrate-processing enzymes. Using novel data, we reconstruct the phylogeny of Holomycota and assess support for the placement of aphelids, consolidating the sister relationship between Aphelida and true Fungi. With this novel phylogenetic framework, we infer the evolutionary dynamics of gene family content in Holomycota and illustrate the reduction of key phagotrophy-related proteins in the evolution of Fungi.

RESULTS AND DISCUSSION

Genome assemblies of *Amoeboaphelidium* species

Sequencing data for three representatives of the aphelid genus *Amoeboaphelidium*—*A. protococcarum* strain X5, *A. protococcarum* strain FD95, and *A. occidentale* strain FD01—were assembled into draft genomes totaling 24 megabases (Mb) for the two strains of *A. protococcarum* and 13.6 Mb for *A. occidentale* strain. The genomes of *A. protococcarum* strains X5 and FD95 are predicted to encode around 13,000 protein-coding genes, and the genome of *A. occidentale* is predicted to encode 7,500 genes, which amount to similar coding sequence percentages and gene densities across all three genomes (Table 1). The genome of *A. occidentale* utilizes the standard genetic code, whereas both strains of *A. protococcarum* utilize non-canonical code with the traditional stop codons UAA and UAG encoding glutamine. The UAR codons constitute 54% of all glutamine codons in the genomes of *A. protococcarum*.

The genomes of both *A. protococcarum* strains show evidence of whole-genome duplication. The clustering of transcript sequences by similarity in each of the two strains of *A. protococcarum* revealed that the majority of predicted genes have a duplicate with an average nucleotide sequence identity of 92.6% for copies in X5 and 95.6% for copies in FD95 (Figure S1A). The estimated values of synonymous divergence (Ks) between the duplicated genes are distributed around a single peak in each of the strains of *A. protococcarum*, with an average Ks of 0.26 in the genome of X5 and 0.14 in the genome of FD95 (Figure S1B). The duplicated genes display a degree of intragenomic synteny: we found 115 blocks of colinear genes with an average span of 44 genes in the genome of X5 and 49 blocks with an average span of 111 genes in the genome of FD95 (Figures 1A and 1B). The distribution of colinear genes in the genomic sequences shows numerous rearrangements in a pairwise alignment of scaffolds and rarely extends to whole scaffolds, refuting conventional meiotic pairing between the duplicated sequences.

A comparison of transcripts that were clustered using both strains of *A. protococcarum* revealed that a dominant proportion (87% of clusters that contain a pair from each of the strains) supports a scenario where both strains emerge as products of independent hybridization events involving four closely related lineages (Figures 1C and S1C). The tree underlying the hybridization scenario allows us to differentiate the genes in most intragenomic pairs by assigning them to a particular ancestral genome. The mapping of inferred subgenome assignments to the genomic sequences shows regions of predominantly single ancestry spanning several hundred kilobases of sequence (Figures 1A and 1B). Genomic regions corresponding to individual subgenomes also cover multiple colinear blocks in both strains, indicating that a significant amount of genomic rearrangements have occurred in the ancestral lineages before the

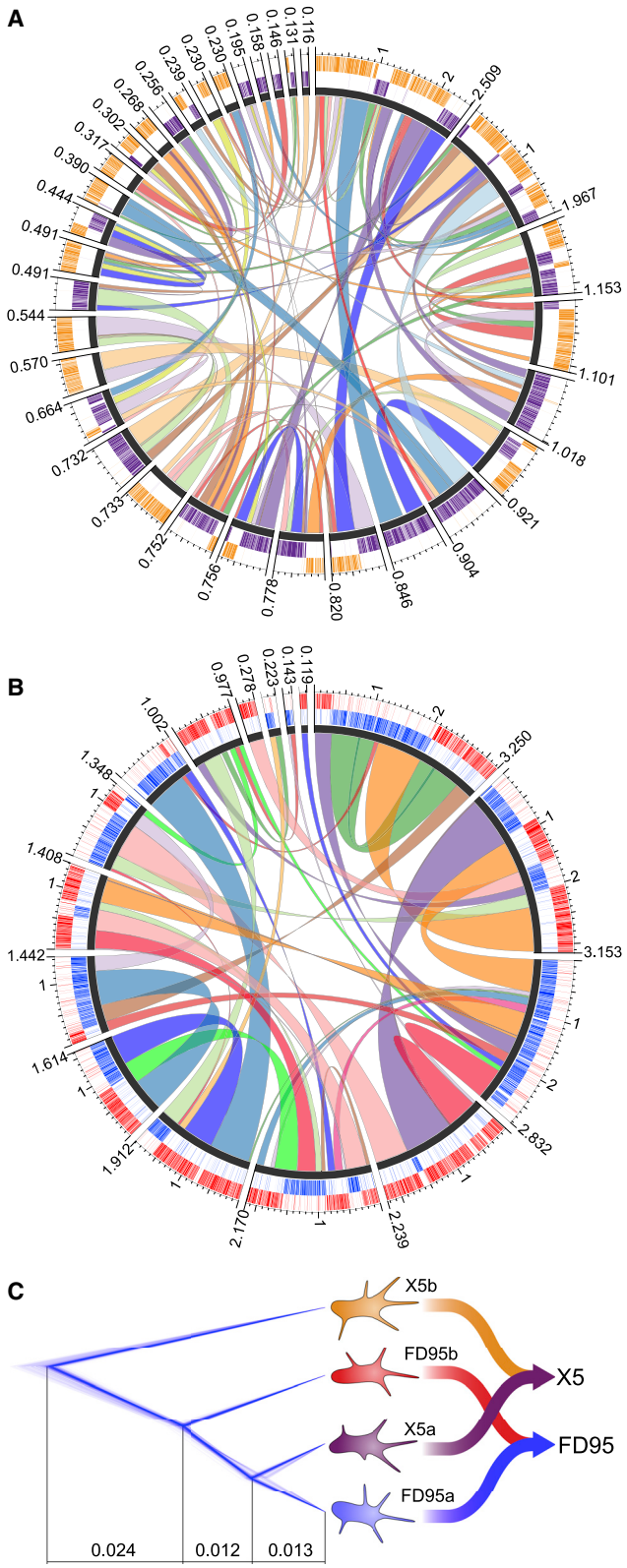


Figure 1. Intragenomic synteny and subgenome assignments in the genome assemblies of *A. protococcarum*

(A) Circular diagram of the genome assembly for strain X5.

(B) Circular diagram of the assembly for strain FD95. Links in circular diagrams connect the identified colinear regions within the genomes. The assignments of genes to the subgenomes are shown in the outer layer of the diagram: inner ring, subgenomes "a"; outer ring, subgenomes "b." Only scaffolds longer than 100 kb are shown in the diagrams; the ruler for genomic scaffolds is given in Mb.

(C) Phylogenetic analysis with a concatenated alignment of coding sequences attributed to the individual subgenomes in *A. protococcarum*, designated here as X5a, X5b, FD95a, and FD95b, performed by BEAST and visualized with DensiTree; the estimates of relative divergence times (substitutions per nucleotide site) for the ancestral lineages are shown below the tree.

See also Figures S1 and S2.

hybridization event. Consequently, the four lineages that have given rise to the hybrid strains X5 and FD95 could have diverged enough to be incompatible for meiotic recombination, meaning that the sequenced strains of *A. protococcarum* are likely interspecific hybrids or allopolyploids.

To assess the ploidy state of the sequenced strains, we examined the SNP frequency distributions within the read data. Not counting the variants introduced by hybridization, the sequencing reads from both strains of *A. protococcarum* are relatively homogeneous: the total number of potential polymorphisms is in the range of 100–600, and the allele frequency distributions do not reveal any peaks that would indicate ploidy higher than one (Figure S1E). By contrast, the number of potential variants found in *A. occidentale* is close to 21,000, which evaluates to per nucleotide SNP rate of 1.8E−3. The allele frequency distribution in *A. occidentale* has a single peak at 0.5, characteristic of a diploid state (Figure S1E).

Mitochondrial genomes of *A. protococcarum* strains X5 and FD95 were assembled into circular molecules of 30,437 and 32,381 bp, respectively. The assembly of *A. occidentale* contained three overlapping mitochondrial fragments, allowing a provisional contiguous circular molecule of 26,157 bp (Figure S2). Aphelids possess a common set of fungal mtDNA-encoded respiratory complex components, but lack an *atp8* gene or any ribosomal protein genes, and have a highly reduced set of tRNA genes. Several genes in *A. protococcarum* are punctuated by members of the LAGLIDADG family of homing endonucleases, commonly found in fungal mitochondrial genomes.[12] The intricate relationship between the sequenced strains of *A. protococcarum* is elaborated further by the mitochondrial genomes. The mtDNA sequences of X5 and FD95 have 99.7% identity, with only several indels accounting for the difference in the genome sizes. Divergence between the closest ancestral lineages that constitute hybrid nuclear genomes of X5 and FD95 is estimated to be nearly 10 times higher than the divergence observed for their mtDNAs (Figure 1C). This level of similarity refutes simple vertical inheritance of mitochondria in X5 and FD95, suggesting that the evolution of these strains or their progenitor lineages involved mitochondrial introgression.

Phylogenomics support sister-group relationship between aphelids and fungi

To reconstruct fungal phylogeny and investigate the position of aphelids, we assembled a 300-gene concatenated alignment. Bayesian inference (BI) and maximum likelihood (ML) analyses with the alignment, using site-heterogeneous models, successfully recover the majority of well-recognized fungal phyla[13]

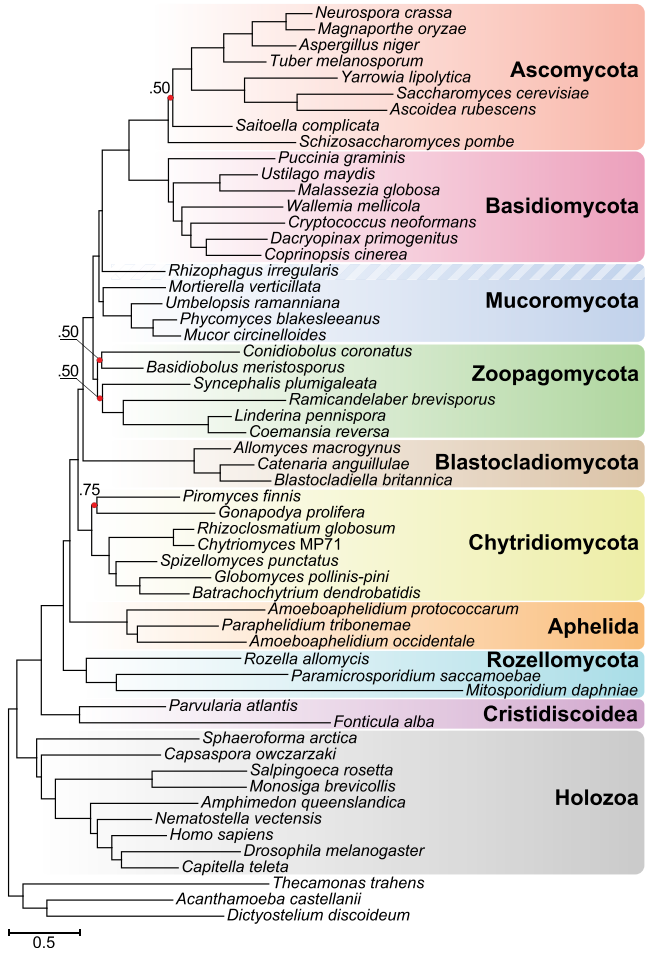


Figure 2. Bayesian inference with the concatenated 300-gene alignment

PhyloBayes consensus tree reconstructed with the alignment using four analysis chains under the CAT-GTR-G4 model; tree nodes that fail to reach maximal support in the analysis are labeled with red dots, and the corresponding posterior probability values are provided. See also Figure S3 and Table S1.

(Figures 2 and S3A). These phylogenies show no support for the earlier proposed superphylum Opisthosporidia,[3] a union of Aphelida and Rozellomycota (including Microsporidia), and instead place Aphelida in a sister relationship to Fungi with full support, in agreement with the previous analyses.[10,11] Several contentious deep nodes of fungal phylogeny[14] are also resolved: the trees fully support the basal placement of chytrids relative to the rest of fungi and back the paraphyly of zygomycete fungi,[15] Zoopagomycota and Mucoromycota. On the other hand, the analyses fail to converge in asserting the branching within the Zoopagomycota, or in resolving the relationship of early-diverging ascomycetes and chytrids, and show a striking inconsistency in the position of glomeromycete *Rhizophagus irregularis*, which groups with the Mucoromycota in the ML tree but shifts in favor of a classical grouping of Glomeromycota and Dikarya[16] in the BI tree.

We examined the stability of key tree nodes using gene subsampling and site elimination techniques. Phylogenetic signal supporting the sister position of aphelids relative to fungi and the basal placement of chytrids dominates in over 90% of all replicates obtained by random subsampling (Figure S3B). The relationship of zygomycete fungi or even the monophyly of Zoopagomycota is less stable: these nodes have a wide range of support values across all subsampling replicates, converging on 60%–90% for the early divergence of Zoopagomycota at 80% of the genes sampled. Approximately unbiased (AU) tests with alignments, generated by the removal of fast-evolving sites or compositionally heterogeneous partitions, firmly reject the monophyly of Opisthosporidia (Table S1). The rejection of Opisthosporidia is unaffected by the exclusion of highly divergent *Mitosporidium*. The tests also reject the alternative placements for chytrids until over 70% of the dataset is eliminated. Furthermore, alternative branching scenarios for zygomycete fungi are rejected following the removal of 10% of the fastest evolving sites, indicating that instability around this node is caused by systematic biases in the data.

Next, the phylogeny is further corroborated by an analysis with a recoded dataset, aiming to alleviate the impact of compositional biases and mutational saturation.[17] The analysis recovers the tree obtained with the non-recoded data (Figure 2), showing full support for all tree nodes, with the exception of nodes in Zoopagomycota, where both analyses fail to converge.

Within Aphelida we find that *A. occidentale* groups with *Paraphelidium tribonemae* rather than *A. protococcarum* (Figure 2). The AU tests confidently reject closer affinity of *A. occidentale* to *A. protococcarum*, rendering their attribution to a single genus incongruous with molecular data (Table S1). In the 300-gene dataset, the two *Amoeboaphelidium* species display only 60% identity at the amino acid level, which is comparable to the level of identity between members of different fungal phyla. In fact, using time-calibrated phylogeny, we obtained an estimate of 420–640 million years ago for the time of divergence of these aphelids (Figure S3C). This interval overlaps the estimated split of ascomycete and basidiomycete fungi, emphasizing just how underestimated the extent of aphelid diversity is currently.

Variation in gene family content and independent loss of flagellum in Aphelida

The predicted genes in *A. occidentale* and both strains of *A. protococcarum* are inferred to comprise 4,947 and 5,711 gene families, respectively. Gene content-wise the two aphelids are considerably diverged. Approximately 1,100 gene families are species specific in *A. occidentale*, and 2,000 families are specific to *A. protococcarum*. The number of families inherited from the last common holomycotan ancestor is estimated by the parsimony analysis to be in the range of 3,300–3,400 for both aphelid species. This puts them near the midpoint of ancestral family counts among holomycotan genomes, above the sequenced members of Rozellomycota and several Dikarya, but below Chytridiomycota, Blastocladiomycota, or Mucoromycota species (Figure 3A). A probabilistic model of gene content evolution suggests that major events of gene loss in the evolution of Holomycota were associated with the appearance of Holomycota, the transition from zoosporic to non-zoosporic fungi, and the emergence of Dikarya (Figure 3B). Extensive changes in gene content are also predicted in lineages of Aphelida,

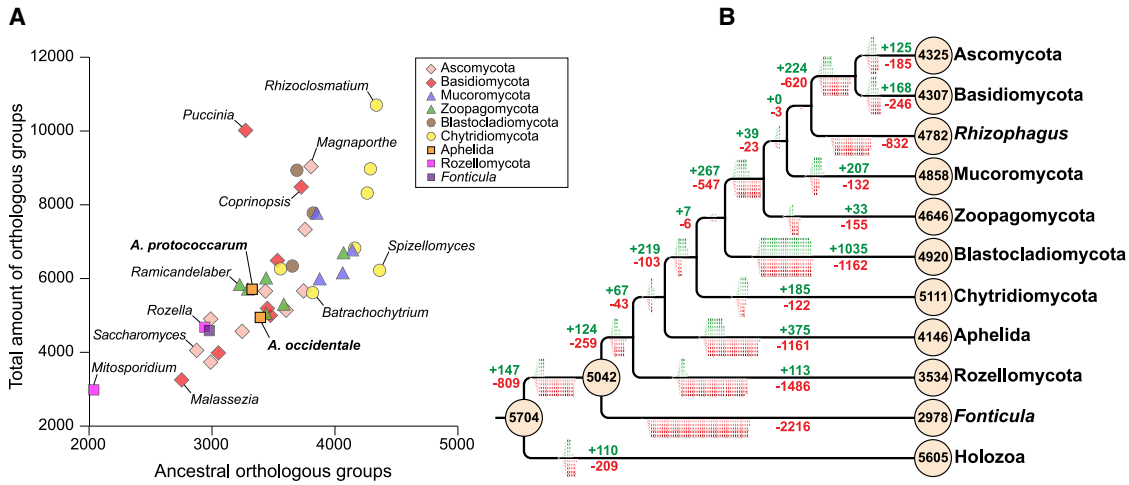


Figure 3. Conservation of inferred ancestral gene families in aphelids and fungi

(A) Scatterplot of the inferred ancestral family counts and the total gene family counts in the genomes of holomycotan species; the ancestral families are orthologous groups predicted to be inherited from the last common ancestor of Holomycota, according to the Dollo parsimony.

(B) Evolution of gene family content in Holomycota, according to the birth-and-death model. Numbers on the branches show the inferred gains (green) and losses (red) of gene families, illustrated schematically using waffle charts. The estimated total gene family counts at the ancestral nodes for Opisthokonta, Holomycota, and each phyla are shown for the corresponding nodes; only families shared by at least two genomes were considered in the analysis.

See also Figure S4 and Tables S2 and S3.

Rozellomycota, and Blastocladiomycota, although these estimates are likely biased by unequal representation of diversity across phyla.

Functional characterization of aphelid genomes, using KEGG orthology, reveals variation in the categories of amino acid metabolism, membrane trafficking, chromosome associated proteins, and cytoskeleton (Figures S4A and S4B). The sequenced aphelids display different capacities for amino acid biosynthesis. In *A. occidentale*, the pathways for biosynthesis of amino acids are largely intact, which resembles the situation seen in the majority of fungal species (Table S2). In *A. protococcarum*, entire pathways for histidine, tryptophan, and arginine biosynthesis are lost, suggesting stricter reliance on the amino acid uptake from the host. The lack of functional flagella in *Amoeboaphelidium* species[3] is reflected in the reduction of cytoskeletal proteins. Both species lack axonemal dyneins and experience loss of complexes involved in the assembly and maintenance of cilia (Table S3). The intraflagellar transport (IFT) complexes and the ciliary transition zone MKS complex, which are generally conserved in zoosporic fungi, undergo loss of components in both aphelids. Dynein assembly factors and the IFT-associated BBSome complex are completely lost in these aphelids. Reduction of the flagellar cytoskeleton extends further in *A. protococcarum* and involves crucial elements of the basal body biogenesis. *A. protococcarum* lacks centrosomal proteins SAS4/CENPJ, CEP135, POC1, POC5, and tubulins δ and ε, predicting centrioles with aberrant structure. By contrast, the transcriptome of *Paraphelidium tribonemae* shows conservation of axonemal motor and ciliogenesis proteins (Table S3), which agrees with the presence of flagellated zoospores in the species.[5] Phylogenetic analyses support closer affinity of *A. occidentale* to *P. tribonemae* (Figure 2). This indicates that the common ancestor of these aphelids was flagellated, and the reduction of flagellum occurred independently in *A. occidentale* and *A. protococcarum*.

Conserved cell wall synthesis and horizontal transfers shaped the repertoires of carbohydrate-active enzymes in aphelids

Aphelid genomes possess a unique composition of carbohydrate-active enzymes (CAZymes), which includes essential components of the fungal cell wall synthesis and remodeling machinery, and enzymes potentially capable of participating in algal cell wall degradation (Table S4). Clustering of the identified CAZy families by their phyletic profiles reveals a group of enzymes associated with the synthesis and processing of β-1,3-glucans that are shared by aphelids and fungi (Figure S5). The group includes 1,3-β-glucan synthases (GT48) and β-1,3-glucan-specific transglycosylases and endoglucanases of families GH72 and GH81, which are implicated in the remodeling of fungal cell walls.[18,19] Aphelid GT48 and GH72 family sequences are orthologous to the fungal enzymes and represent the earliest diverging members of these families in the holomycotan lineage. Aphelid GH81 family sequences, however, are not closely related to the fungal endo-β-1,3-glucanases and instead branch with prokaryotic representatives of the family, suggesting the enzyme was acquired independently through a horizontal gene transfer (HGT) event (Figure S6A). The largest CAZy family in aphelids, GT2, includes chitin synthases (CHSs)—central enzymes in the synthesis of the characteristic polymer of fungal cell walls. Aphelids mark an expansion of this family in Holomycota, with *A. occidentale* and *A. protococcarum* each containing 10 CHS genes. The domain architectures of aphelid sequences conform to the archetypes of the main fungal CHS classes and branch primarily with the homologs from lower fungi (Figure 4A). Staining of algal cultures infected with *A. protococcarum*, using wheat germ agglutinin, reveals chitinous walls of the invasive cysts (Figure 4B). The rigidity conferred by chitin is proposed to facilitate the mechanical penetration of host cell walls, while the release of mature parasitoid from the infected algal cells happens by

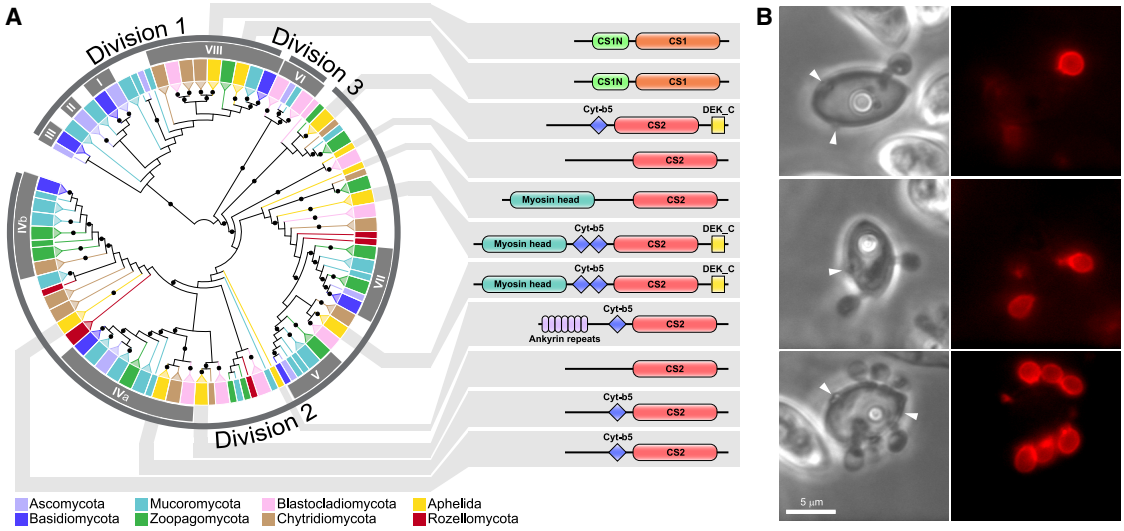


Figure 4. Evolution of chitin synthases in Holomycota and chitin staining of aphelid invasive cysts

(A) Maximum likelihood tree of chitin synthases (CHSs) in Holomycota reconstructed using an alignment of the core domain of 246 CHSs from 21 species. The tree only depicts the topology—branch lengths are not representative of true genetic distances. Branches receiving ≥95% UFBoot support are marked with a black dot. Clades formed by sequences belonging to major taxonomic lineages are collapsed in the tree. The CHS divisions (outer circle labels) and classes (inner circle labels) are named in accordance with Li et al.[22] Domain architectures of aphelid CHSs are shown to the right of the dendrogram: CS1N, chitin synthase N-terminal (PF08407); CS1, chitin synthase 1 (PF01644); CS2, chitin synthase 2 (PF03142); Cyt-b5, cytochrome b5-like heme/steroid binding domain (PF00173); myosin head, myosin motor domain (PF00063); and DEK_C, DEK C-terminal domain (PF08766).

(B) Fluorescent wheat germ agglutinin staining of aphelid invasive cysts. The cysts of *A. protococcarum* are attached to the cells of *Scenedesmus obliquus*, where most of the cytoplasm has been consumed by the parasitoid. The parasitoid escapes through the openings created by the penetration tubes and disintegrated invasive cysts (white arrows).

See also Figures S5 and S6 and Tables S4 and S5.

dehiscence of former invasive cysts,[7,20] which presumably involves lysis of chitinous walls. Notably, the most abundant glycoside hydrolase family in the sequenced aphelids, and also the family with the highest number of predicted secreted members, is GH18, known to contain the bulk of fungal chitinases.[21] *Amoeboaphelidium* species encode from 6 to 8 members of the GH18 family, and at least half of those sequences are predicted to possess a signal peptide (Table S4).

The repertoire of aphelid CAZymes potentially capable of degrading cellulose, hemicellulose, and pectin is very limited or lacking (Figure S5). Among the fungal CAZy families with cellulolytic activities, namely glycoside hydrolase families GH5, GH6, GH7, GH12, GH45, and the AA9/GH61 family of lytic polysaccharide monooxygenases,[23] aphelid sequences are found only in the GH5 family. Phylogenetic analysis of the diverse GH5 family enzymes (Figure S6B) places aphelid sequences in the subfamily GH5_12, with characterized β-glucosylceramidase and flavonoid β-glucosidase activities, and inside a cluster of subfamilies GH5_11, GH5_16, GH5_24, and GH5_51, harboring a single characterized enzyme with β-1,6-galactanase activity.[24] However, both the GH5_12 subfamily and the GH5_11-16-25-51 subfamily clusters also include homologs from mycoparasitic *Rozella allomycis*, challenging their association with algal cell wall degradation.

Exploring aphelid CAZymes we noticed several other candidates for bacterial and algal HGTs. A member of the GH1 family in aphelids groups with algal and plant homologs of the chloroplast membrane remodeling galactosyltransferase SFR2, although bacterial origin of the enzyme could not be ruled out (Figure S6C). Two more examples include GT2 family mannosyltransferase and GT34 family xylosyltransferase, where aphelids group with sequences from chlorophytes, streptophytes, haptophytes, and glaucophytes (Figures S6D and S6E). The sequences from both aphelid species branch together in the trees, which points to the antiquity of the presumable acquisition events, and might be indicative of a long-standing association between aphelids and their algal hosts. A systematic genome-wide search for HGT cases revealed around a 100 potential acquisition events in aphelids, with the majority of findings involving either Bacteria or eukaryotic groups Viridiplantae and stramenopiles (Table S5).

Expansions of receptor-like protein kinases in aphelids

Kinannote identifies 311 protein kinases in *A. occidentale* and 215 distinct protein kinases in *A. protococcarum*. Classification of the detected kinases shows an atypically large number of sequences among the unclassified kinase findings and in the group of tyrosine kinase-like (TKL) kinases (Table S6). Phylogenetic reconstruction with the catalytic domains of aphelid sequences clusters the majority of unclassified kinases with the TKL group, which together constitute 50% of the kinome in *A. occidentale* and 30% of the kinome in *A. protococcarum* (Figure 5A).

Evolution of fungal kinomes is marked by the loss of protein tyrosine kinases[25] and the narrowing of the receptor kinase complements to hybrid histidine kinases.[26] Aphelid genomes provide further evidence for the early loss of tyrosine kinases by the Holomycota lineage, as their kinases lack the characteristic residues for tyrosine recognition.[27,28] Additionally, we find no

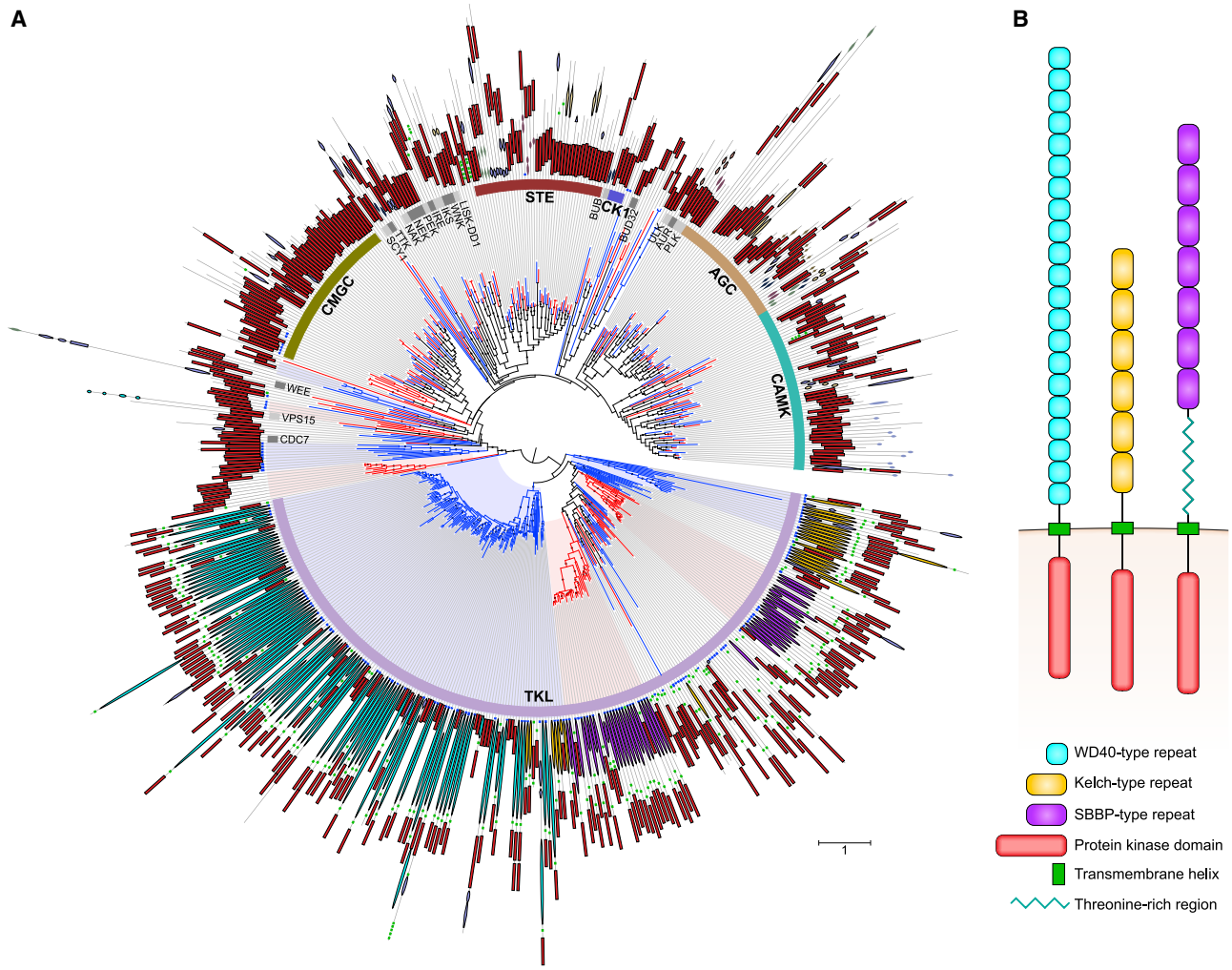


Figure 5. Protein kinases in *A. occidentale* and *A. protococcarum*

(A) Maximum likelihood tree reconstructed using an alignment of 569 catalytic domains of protein kinases detected in aphelids. Branches receiving ≥95% UFBoot support are marked with a black dot, the branches of *A. occidentale* are colored blue, and those of *A. protococcarum* are colored red. Blue and red shading in the tree additionally highlights species-specific expansions of kinases. Domain architectures for the kinases are shown, with the catalytic domain in red, transmembrane helices in green, signal peptide in blue, and the repeat arrays of TKL kinases in light blue (WD40-type), orange (Kelch-type), and purple (SBBP-type).

(B) Domain architecture archetypes of the receptor-like protein kinases in *A. occidentale* (WD40-type and Kelch-type) and *A. protococcarum* (SBBP-type).

See also Table S6 and Figure S7.

phosphotyrosine-interacting Src homology 2 (SH2) domains in their genomes, supporting the lack of tyrosine kinase activity. However, in contrast to the kinomes of fungal species, the TKL assemblage in aphelids comprises proteins that follow the common architecture of cell-surface receptor kinases, with a single transmembrane helix, a variable extracellular region, and a C-terminal kinase catalytic domain (Figure 5B). In the majority of cases, the extracellular ligand-binding region is composed of an array of repeats, which tend to be specific to the TKL clusters in the tree but are all recognized as members of the versatile β-propeller fold.[29] The receptor-like kinases in *A. occidentale* have WD40-type repeats or Kelch-type repeats, and the repeats in *A. protococcarum* show the highest similarity to the SBBP family (PF06739).

Sensing functions in many fungal species are mediated by the eukaryotic two-component system, which involves hybrid histidine kinases as receptor proteins.[30] We find five proteins with the domain architectures of hybrid histidine kinases in *A. occidentale* and nine proteins in *A. protococcarum*. Aphelid sequences are primarily species specific and fall into eight individual lineages in the tree with fungal histidine kinases (Figure S7). The only histidine kinase with orthology in both aphelids is closely related to receptors for the plant hormone ethylene and contains the ethylene binding domain. Phytohormone receptor homologs occur in the genomes of several early-diverging fungal species and are believed to be orchestrating interactions with the host organisms.[31] Participation of phytohormones in the regulation of cellular activities in algae[32] and the conservation

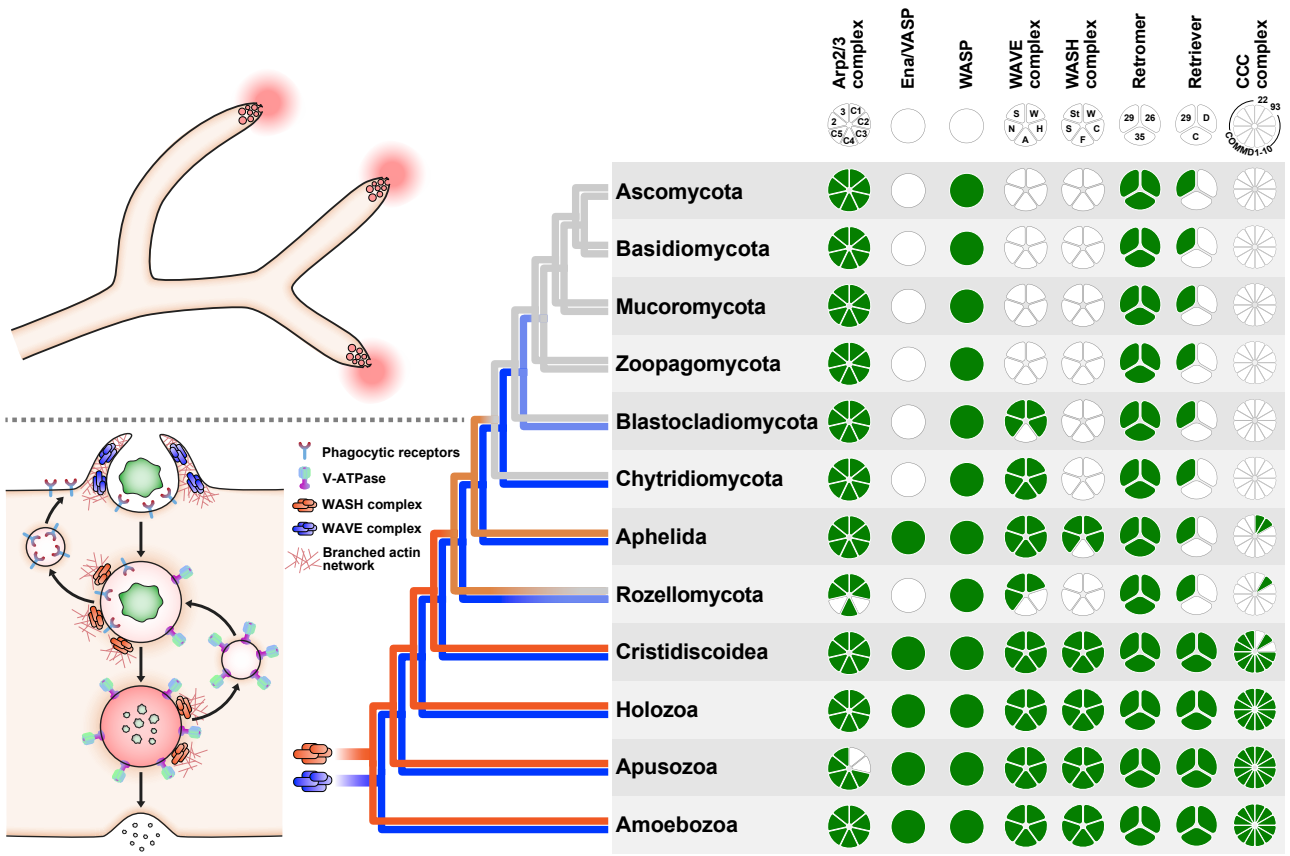


Figure 6. Conservation of the actin-remodeling and endosomal complexes in Holomycota

A parsimonious scenario for the reduction of WASH (orange) and WAVE (blue) complexes in the evolution of Holomycota. Components of the WAVE complex are WAVE/SCAR protein (W), HSPC300 (H), Abi (A), Nap1 (N), and Sra1 (S). Components of the WASH complex are WASH protein (W), CCDC53 (C), Fam21 (F), SWIP (S), and strumpellin/KIAA0196 (St). Components of the retromer cargo-selective complex are Vps26, Vps35, and Vps29. Components of the retriever complex are DSCR3 (D), C16orf62 (C), and Vps29. Components of the CCC complex are CCDC22 (22), CCDC93 (93), and 10 copper metabolism MURR1 domain-containing proteins (COMMD1-10). Components of the actin nucleation Arp2/3 complex are ARPC1-5 (C1-C5), ARP2 (2), and ARP3 (3). See also Table S7.

of the ethylene receptor in aphelids support the possibility of their involvement in interaction with their algal hosts.

Evolution of proteins associated with the transition to fungal lifestyle

Central to the evolutionary transition from ancestral Holomycota to Fungi is the loss of phagotrophy and the gradual restructuring of cellular machinery for osmotrophy and hyphal growth.[33] Previous comparative genomic analyses have linked the loss of the Wiskott-Aldrich syndrome protein (WASP) family members, WAVE and WASH, with the transition to hyphal organization in Holomycota.[26] With the inclusion of data from aphelids, a comparison of annotated protein domain families highlights the loss of WASH complex as one of the demarcating features, setting apart Fungi and the ancestral phagotrophic Holomycota.

The WASH complex includes five proteins that are conserved in the deeper-branching eukaryotes (Figure 6). In aphelids, we find orthologs of the WASH protein and three other components of the complex: CCDC53, SWIP, and strumpellin. Aphelid sequences are remarkably divergent—they display the highest evolutionary rates among eukaryotic orthologs. Sensitive profile searches could not detect the fifth component, Fam21, which is implicated in maintaining the stability of the complex and interacting with various other proteins.[34,35] Although the aphelids present exceedingly divergent orthologs, retention of the complex would be consistent with its crucial role in phagocytosis. WASH functions as a nucleation-promoting factor for branched actin filament networks in vesicular trafficking,[36] and the complex is necessary for organizing endosomal retrieval subdomains, which recycle surface receptors and acidifying V-ATPases in the endocytic pathway.[37] The reliance on WASH for maintaining phagocytic receptors and efficient proteolysis is conserved between mammalian cells and the amoebozoan *Dictyostelium discoideum*,[38,39] suggesting a similar function in phagotrophic holomycotans. The reported ability of the WASH complex to perform many of its functions without Fam21[40] supports the existence of this complex in aphelids, which might represent its minimal functionally active variant. The complete loss of WASH complex in the last common ancestor of Fungi marks the transition to extracellular digestion, superseding the requirement for phagocytosis-associated cargo recycling (Figure 6).

In the endosomal retrieval subdomains, the WASH complex interacts with other retrieval complexes, which include retromer,

retriever, and the CCC complex.[41] The three core subunits of the retromer are almost universally conserved in eukaryotes[42] and are retained in both Aphelida and Fungi. The other retrieval complexes with deep eukaryotic ancestry, retriever and CCC complexes, undergo reduction in the early evolution of the Holomycota and are completely lost in the fungal lineages (Figure 6; Table S7). Aphelids retain orthologs of only two CCC complex subunits, CCDC22 and CCDC93, and no proteins specific to the retriever complex.

In aphelids, we find orthologs of another major activator of branched actin filaments, protein WAVE, along with a more widely conserved and versatile WASP. All components of the pentameric WAVE complex, structurally analogous to the WASH complex,[43] are conserved in aphelids. WAVE is associated primarily with the actin-based plasma membrane protrusions and cell motility.[44,45] The loss of WAVE complex in the fungal evolution coincides with the appearance of non-zoosporic terrestrial fungi (Figure 6). However, the phylogenetic distribution of a wider set of actin-remodeling proteins traces an even earlier reduction of cytoskeletal complexity in the zoosporic fungi.[46] In contrast to the zoosporic fungi, aphelids retain members of the Ena/VASP family—mediators of actin elongation involved in the formation of filopodia.[47] In *A. protococcarum*, we also find a protein with the highest similarity to the regulators of actin assembly of the CARMIL family, which is lost by Fungi.[48] The searches also reveal aphelid orthologs of an animal actin filament stabilizer Lasp of the nebulin family,[49] moving the emergence of this family to the common ancestor of Holozoa and Holomycota.

Conclusions

Genomic data demonstrate that *A. protococcarum* and *A. occidentale* are not closely related aphelid species and challenge their attribution to a single genus, which was originally motivated by the apparent morphological likeness of their zoospores.[50] These species have different biosynthetic capabilities, use different genetic codes, and have independently experienced reduction of the flagellar apparatus. Their dissimilarity, first recognized from the comparison of rRNA sequences,[7] is also apparent in the time-calibrated analysis, dating back their divergence to over 400 million years ago, and phylogenetic reconstructions, asserting the paraphyly of the genus *Amoeboaphelidium*. The propensity to form hybrids, shown by the strains of *A. protococcarum*, adds further complexity to the scope of aphelid diversity. The sequenced aphelids serve as a good example of how unremarkable morphological features and similarities of life cycles can mask deep diversity within a group.

Our phylogenomic reconstructions firmly reject the monophyly of Opisthosporidia and support a sister-group relationship between Aphelida and Fungi, reinforcing the results of earlier analyses.[1,10,11] As the closest relatives of true Fungi, aphelids offer insight into the evolution of ancestral features of the fungal stem lineage. Common cell wall remodeling enzymes and the early diversification of CHSs seen in aphelids link their invasive cysts with the fungal cysts and sporangia, suggesting shared early origin of their cell walls. The presence of similar parasitic lifestyles among the early-diverging fungal lineages[51] hints at the possibility that the whole group has emerged from a parasitic aphelid-like ancestor in close association with algae. On the surface, however, the genomic data show that aphelids display a more derived state by retaining fewer ancestral gene families overall than the predominantly saprotrophic members of the zoosporic fungal lineages. Among the ancestral gene families that are retained in aphelids we find regulators of actin-based motility and endosomal cargo recycling, consistent with the presence of amoeboid and phagotrophic stages in aphelids. The losses of these proteins by the fungal lineages coincide with key transitions in their evolution: the loss of WASH complex marks the switch from phagotrophy to extracellular digestion, and the loss of WAVE complex along with the flagellum marks the transition from zoosporic to non-zoosporic fungi as an adaptation to dry terrestrial habitats. In essence, Aphelida appear to have emerged as a lineage capturing one of the niches explored by the fungal ancestors before they transitioned to exclusively osmotrophic nutrition, a niche tightly linked to predation on algae.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Genome sequencing and assembly
  - Genome annotation
  - Analysis of hybrid genomes
  - Variant calling
  - Dataset preparation for phylogenomic analysis
  - Phylogenomic analyses
  - Molecular dating
  - Inference of ancestral gene families
  - Functional annotation
  - Analysis of carbohydrate-active enzymes
  - Chitin staining
  - Global search for horizontally transferred genes
  - Analysis of protein kinases
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.08.071.

ACKNOWLEDGMENTS

S.A.K. thanks RSF for grant 21-74-20089 (study supervision, manuscript editing, and providing *Amoeboaphelidium protococcarum* strain X5) and the Ministry of Science and Higher Education of the Russian Federation for grant 075-15-2021-1069 for support in the cultivation of aphelids. We thank V.I. Shestopalov for assistance with delivering the biological material.

AUTHOR CONTRIBUTIONS

K.V.M., S.A.K., Y.V.P., and V.V.A. conceived and supervised the study; S.A.K., P.M.L., P.A.L., and V.V.A. provided and maintained the biological material; M.D.L. and A.A.P. obtained the sequencing data; K.V.M., M.A.N., I.R.P.,

E.V.P., and D.Y.S. performed the analyses; K.V.M. drafted the manuscript; S.A.K. and V.V.A. edited the manuscript; and all authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 8, 2022
Revised: June 26, 2022
Accepted: August 24, 2022
Published: September 19, 2022

REFERENCES

1. Tedersoo, L., Sánchez-Ramírez, S., Kõljalg, U., Bahram, M., Döring, M., Schigel, D., May, T., Ryberg, M., and Abarenkov, K. (2018). High-level classification of the Fungi and a tool for evolutionary ecological analyses. Fungal Divers. *90*, 135–159.
2. Letcher, P.M., and Powell, M.J. (2019). A taxonomic summary of *Aphelidiaceae*. IMA Fungus *10*, 4.
3. Karpov, S.A., Mamkaeva, M.A., Aleoshin, V.V., Nassonova, E., Lilje, O., and Gleason, F.H. (2014). Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. Front. Microbiol. *5*, 112.
4. Gromov, B.V. (2000). Algal parasites of the genera *Aphelidium*, *Amoeboaphelidium* and *Pseudoaphelidium* from the Cienkovski's "Monadea" group as representatives of new class. Zool. Z. *79*, 517–525.
5. Karpov, S.A., Tcvetkova, V.S., Mamkaeva, M.A., Torruella, G., Timpano, H., Moreira, D., Mamanazarova, K.S., and López-García, P. (2017). Morphological and genetic diversity of Opisthosporidia: new aphelid *Paraphelidium tribonemae* gen. et sp. nov. J. Eukaryot. Microbiol. *64*, 204–212.
6. Seto, K., Matsuzawa, T., Kuno, H., and Kagami, M. (2020). Morphology, ultrastructure, and molecular phylogeny of *Aphelidium collabens* sp. nov. (Aphelida), a parasitoid of a green alga *Coccomyxa* sp. Protist *171*, 125728.
7. Letcher, P.M., Lopez, S., Schmieder, R., Lee, P.A., Behnke, C., Powell, M.J., and McBride, R.C. (2013). Characterization of *Amoeboaphelidium protococcarum*, an algal parasite new to the cryptomycota isolated from an outdoor algal pond used for the production of biofuel. PLoS One *8*, e56232.
8. Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., et al. (2019). Revisions to the classification, nomenclature, and diversity of eukaryotes. J. Eukaryot. Microbiol. *66*, 4–119.
9. Wijayawardene, N.N. (2020). Outline of Fungi and fungus-like taxa. Mycosphere *11*, 1060–1456.
10. Torruella, G., Grau-Bové, X., Moreira, D., Karpov, S.A., Burns, J.A., Sebé-Pedrós, A., Völcker, E., and López-García, P. (2018). Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. Commun. Biol. *1*, 231.
11. Galindo, L.J., López-García, P., Torruella, G., Karpov, S., and Moreira, D. (2021). Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. Nat. Commun. *12*, 4973.
12. Megarioti, A.H., and Kouvelis, V.N. (2020). The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). Genome Biol. Evol. *12*, 1337–1354.
13. Spatafora, J.W., Aime, M.C., Grigoriev, I.V., Martin, F., Stajich, J.E., and Blackwell, M. (2017). The fungal tree of life: from molecular systematics to genome-scale phylogenies. Microbiol. Spectr. *5*, https://doi.org/10.1128/microbiolspec.FUNK-0053-2016.
14. Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., Spatafora, J.W., Groenewald, M., Dunn, C.W., Hittinger, C.T., et al. (2021). A genome-scale phylogeny of the kingdom Fungi. Curr. Biol. *31*, 1653–1665.e5.
15. Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia *108*, 1028–1046.
16. James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., et al. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818–822.
17. Susko, E., and Roger, A.J. (2007). On reduced amino acid alphabets for phylogenetic inference. Mol. Biol. Evol. *24*, 2139–2150.
18. Mouyna, I., Aimanianda, V., Hartl, L., Prevost, M.C., Sismeiro, O., Dillies, M.A., Jagla, B., Legendre, R., Coppee, J.Y., and Latgé, J.P. (2016). GH16 and GH81 family beta-(1,3)-glucanases in *Aspergillus fumigatus* are essential for conidial cell wall morphogenesis. Cell. Microbiol. *18*, 1285–1293.
19. Patel, P.K., and Free, S.J. (2019). The genetics and biochemistry of cell wall structure and synthesis in *Neurospora crassa*, a model filamentous fungus. Front. Microbiol. *10*, 2294.
20. Karpov, S.A., Mikhailov, K.V., Mirzaeva, G.S., Mirabdullaev, I.M., Mamkaeva, K.A., Titova, N.N., and Aleoshin, V.V. (2013). Obligately phagotrophic aphelids turned out to branch with the earliest-diverging fungi. Protist *164*, 195–205.
21. Oyeleye, A., and Normi, Y.M. (2018). Chitinase: diversity, limitations, and trends in engineering for suitable applications. Biosci. Rep. *38*. BSR2018032300.
22. Li, M., Jiang, C., Wang, Q., Zhao, Z., Jin, Q., Xu, J.R., and Liu, H. (2016). Evolution and functional insights of different ancestral orthologous clades of chitin synthase genes in the fungal tree of life. Front. Plant Sci. *7*, 37.
23. Kubicek, C.P., Starr, T.L., and Glass, N.L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. Annu. Rev. Phytopathol. *52*, 427–451.
24. Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., 3rd, and Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol. Biol. *12*, 186.
25. Suga, H., Dacre, M., de Mendoza, A., Shalchian-Tabrizi, K., Manning, G., and Ruiz-Trillo, I. (2012). Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. Sci. Signal. *5*, ra35.
26. Kiss, E., Hegedüs, B., Virágh, M., Varga, T., Merényi, Z., Kószó, T., Bálint, B., Prasanna, A.N., Krizsán, K., Kocsubé, S., et al. (2019). Comparative genomics reveals the origin of fungal hyphae and multicellularity. Nat. Commun. *10*, 4080.
27. Cowan-Jacob, S.W. (2006). Structural biology of protein tyrosine kinases. Cell. Mol. Life Sci. *63*, 2608–2625.
28. Zhao, Z., Jin, Q., Xu, J.R., and Liu, H. (2014). Identification of a fungi-specific lineage of protein kinases closely related to tyrosine kinases. PLoS One *9*, e89813.
29. Chaudhuri, I., Söding, J., and Lupas, A.N. (2008). Evolution of the beta-propeller fold. Proteins *71*, 795–803.
30. Hérivaux, A., So, Y.S., Gastebois, A., Latgé, J.P., Bouchara, J.P., Bahn, Y.S., and Papon, N. (2016). Major sensing proteins in pathogenic fungi: the hybrid histidine kinase family. PLoS Pathog. *12*, e1005683.
31. Hérivaux, A., Dugé de Bernonville, T., Roux, C., Clastre, M., Courdavault, V., Gastebois, A., Bouchara, J.P., James, T.Y., Latgé, J.P., Martin, F., and Papon, N. (2017). The identification of phytohormone receptor homologs in early diverging fungi suggests a role for plant sensing in land colonization by fungi. mBio *8*. e01739-16.
32. Han, X., Zeng, H., Bartocci, P., Fantozzi, F., and Yan, Y. (2018). Phytohormones and effects on growth and metabolites of microalgae: a review. Fermentation *4*, 25.

33. Richards, T.A., Leonard, G., and Wideman, J.G. (2017). What defines the "kingdom" fungi? Microbiol. Spectr. *5*, 57–77.

34. Lee, S., Chang, J., and Blackstone, C. (2016). FAM21 directs SNX27-retromer cargoes to the plasma membrane by preventing transport to the Golgi apparatus. Nat. Commun. *7*, 10939.

35. Wang, J., Fedoseienko, A., Chen, B., Burstein, E., Jia, D., and Billadeau, D.D. (2018). Endosomal receptor trafficking: retromer and beyond. Traffic *19*, 578–590.

36. Rotty, J.D., Wu, C., and Bear, J.E. (2013). New insights into the regulation and cellular functions of the ARP2/3 complex. Nat. Rev. Mol. Cell Biol. *14*, 7–12.

37. King, J.S., Gueho, A., Hagedorn, M., Gopaldass, N., Leuba, F., Soldati, T., and Insall, R.H. (2013). WASH is required for lysosomal recycling and efficient autophagic and phagocytic digestion. Mol. Biol. Cell *24*, 2714–2726.

38. Carnell, M., Zech, T., Calaminus, S.D., Ura, S., Hagedorn, M., Johnston, S.A., May, R.C., Soldati, T., Machesky, L.M., and Insall, R.H. (2011). Actin polymerization driven by WASH causes V-ATPase retrieval and vesicle neutralization before exocytosis. J. Cell Biol. *193*, 831–839.

39. Buckley, C.M., Gopaldass, N., Bosmani, C., Johnston, S.A., Soldati, T., Insall, R.H., and King, J.S. (2016). WASH drives early recycling from macropinosomes and phagosomes to maintain surface phagocytic receptors. Proc. Natl. Acad. Sci. USA *113*, E5906–E5915.

40. Park, L., Thomason, P.A., Zech, T., King, J.S., Veltman, D.M., Carnell, M., Ura, S., Machesky, L.M., and Insall, R.H. (2013). Cyclical action of the WASH complex: FAM21 and capping protein drive WASH recycling, not initial recruitment. Dev. Cell *24*, 169–181.

41. McNally, K.E., and Cullen, P.J. (2018). Endosomal retrieval of cargo: retromer is not alone. Trends Cell Biol. *28*, 807–822.

42. Koumandou, V.L., Klute, M.J., Herman, E.K., Nunez-Miguel, R., Dacks, J.B., and Field, M.C. (2011). Evolutionary reconstruction of the retromer complex and its function in *Trypanosoma brucei*. J. Cell Sci. *124*, 1496–1509.

43. Jia, D., Gomez, T.S., Metlagel, Z., Umetani, J., Otwinowski, Z., Rosen, M.K., and Billadeau, D.D. (2010). WASH and WAVE actin regulators of the Wiskott-Aldrich syndrome protein (WASP) family are controlled by analogous structurally related complexes. Proc. Natl. Acad. Sci. USA *107*, 10442–10447.

44. Campellone, K.G., and Welch, M.D. (2010). A nucleator arms race: cellular control of actin assembly. Nat. Rev. Mol. Cell Biol. *11*, 237–251.

45. Fritz-Laylin, L.K., Lord, S.J., and Mullins, R.D. (2017). WASP and SCAR are evolutionarily conserved in actin-filled pseudopod-based motility. J. Cell Biol. *216*, 1673–1688.

46. Prostak, S.M., Robinson, K.A., Titus, M.A., and Fritz-Laylin, L.K. (2021). The actin networks of chytrid fungi reveal evolutionary loss of cytoskeletal complexity in the fungal kingdom. Curr. Biol. *31*, 1192–1205.e6.

47. Mattila, P.K., and Lappalainen, P. (2008). Filopodia: molecular architecture and cellular functions. Nat. Rev. Mol. Cell Biol. *9*, 446–454.

48. Stark, B.C., Lanier, M.H., and Cooper, J.A. (2017). CARMIL family proteins as multidomain regulators of actin-based motility. Mol. Biol. Cell *28*, 1713–1723.

49. Pappas, C.T., Bliss, K.T., Zieseniss, A., and Gregorio, C.C. (2011). The nebulin family: an actin support group. Trends Cell Biol. *21*, 29–37.

50. Letcher, P.M., Powell, M.J., Lopez, S., Lee, P.A., and McBride, R.C. (2015). A new isolate of *Amoeboaphelidium protococcarum*, and *Amoeboaphelidium occidentale*, a new species in phylum Aphelida (Opisthosporidia). Mycologia *107*, 522–531.

51. Karpov, S.A., Kobseva, A.A., Mamkaeva, M.A., Mamkaeva, K.A., Mikhailov, K.V., Mirzaeva, G.S., et al. (2014). *Gromochytrium mamkaevae* gen. & sp. nov. and two new orders: Gromochytriales and Mesochytriales (Chytridiomycetes). Persoonia *32*, 115–126.

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

53. Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. *18*, 821–829.

54. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience *1*, 18.

55. Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. Genome Biol. *13*, R56.

56. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.

57. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

58. Noguchi, H., Park, J., and Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res. *34*, 5623–5630.

59. Pati, A., Heath, L.S., Kyrpides, N.C., and Ivanova, N. (2011). ClaMS: a classifier for metagenomic sequences. Stand. Genomic. Sci. *5*, 248–253.

60. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*, 637–644.

61. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-seq. Bioinformatics *25*, 1105–1111.

62. Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics *12*, 491.

63. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. *33*, 6494–6506.

64. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067.

65. Smit, A.F.A., and Hubley, R. (2008–2015). RepeatModeler Open-1.0. http://www.repeatmasker.org.

66. Smit, A.F.A., Hubley, R., and Green, P. (2013–2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.

67. Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques *28*, 1102. 1104.

68. Dutilh, B.E., Jurgelenaite, R., Szklarczyk, R., van Hijum, S.A., Harhangi, H.R., Schmid, M., de Wild, B., Françoijs, K.J., Stunnenberg, H.G., Strous, M., et al. (2011). FACIL: fast and accurate genetic code inference and logo. Bioinformatics *27*, 1929–1933.

69. Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: assessing genome assembly and annotation completeness. Methods Mol. Biol. *1962*, 227–245.

70. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

71. Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7–W13.

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

73. Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49.

74. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

75. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

76. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *10*, e1003537.

77. Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics *8*, 77–80.

78. Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA *108*, 1513–1518.

79. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

80. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. https://doi.org/10.48550/arXiv.1207.3907.

81. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296.

82. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.

83. Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. S. *41*, 95–98.

84. Roure, B., Rodriguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. *7* (*Suppl 1*), S2.

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

86. Kück, P., and Struck, T.H. (2014). BaCoCa – a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. *70*, 94–98.

87. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.

88. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. *62*, 611–615.

89. Tice, A.K., Žihala, D., Pánek, T., Jones, R.E., Salomaki, E.D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A.J., et al. (2021). PhyloFisher: a phylogenomic package for resolving eukaryotic relationships. PLoS Biol. *19*, e3001365.

90. Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D.C., Augspurger, T., Halchenko, Y., Cole, J.B., Warmenhoven, J., et al. (2017). mwaskom/seaborn, v0.8.1 (September 2017). (Zenodo). https://zenodo.org/record/883859#.Yw0R331BzDc.

91. Smith, S.A., Brown, J.W., and Walker, J.F. (2018). So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS One *13*, e0197433.

92. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. *67*, 901–904.

93. Csűrös, M. (2010). Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics *26*, 1910–1912.

94. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35*, W182–W185.

95. Kanehisa, M., Sato, Y., and Kawashima, M. (2022). KEGG mapping tools for uncovering hidden features in biological data. Protein Sci. *31*, 47–53.

96. Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., and Yin, Y. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. *46*, W95–W101.

97. Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. *7*, e1002195.

98. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

99. Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. *37*, 420–423.

100. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

101. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

102. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

103. Goldberg, J.M., Griggs, A.D., Smith, J.L., Haas, B.J., Wortman, J.R., and Zeng, Q. (2013). Kinannote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. Bioinformatics *29*, 2387–2394.

104. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

105. Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics *20*, 473.

106. Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. *49*, D458–D460.

107. Pinevich, A.V., Mamkaeva, K.A., Titova, N.N., Gavrilova, O.V., Ermilova, E.V., Kvitko, K.V., Pljusch, A.V., Voloshko, L.N., and Averina, S.G. (2004). St. Petersburg Culture Collection (CALU): four decades of storage and research with microscopic algae, cyanobacteria and other microorganisms. Nova Hedwigia *79*, 115–126.

108. Pinevich, A., Gromov, B., Mamkaeva, K., and Nasonova, E. (1997). Study of molecular karyotypes in *Amoeboaphelidium protococcarum*,the endotrophic parasite of Chlorophycean alga *Scenedesmus*. Curr. Microbiol. *34*, 122–126.

109. NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *44*, D7–D19.

110. The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. *45*, D158–D169.

111. Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. *110*, 462–467.

112. Bouckaert, R., and Heled, J. (2014). DensiTree 2: seeing trees through the forest. https://doi.org/10.1101/012401.

113. Torruella, G., de Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M.A., del Campo, J., Eme, L., Pérez-Cordón, G., Whipps, C.M., Nichols, K.M., et al. (2015). Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. Curr. Biol. *25*, 2404–2410.

114. Zhong, M., Hansen, B., Nesnidal, M., Golombek, A., Halanych, K.M., and Struck, T.H. (2011). Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. BMC Evol. Biol. *11*, 369.

115. Dayhoff, M., Schwartz, R., and Orcutt, B. (1978). A model of evolutionary change in proteins. In Atlas of Protein Sequence and Structure, M. Dayhoff, ed. (National Biomedical Research Foundation), pp. 345–352.

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

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

118. Berbee, M.L., Strullu-Derrien, C., Delaux, P.M., Strother, P.K., Kenrick, P., Selosse, M.A., and Taylor, J.W. (2020). Genomic and fossil windows into the secret lives of the most ancient fungi. Nat. Rev. Microbiol. *18*, 717–730.

119. Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. Proc. Natl. Acad. Sci. USA *115*, E2274–E2283.

120. Chang, Y., Wang, S., Sekimoto, S., Aerts, A.L., Choi, C., Clum, A., LaButti, K.M., Lindquist, E.A., Yee Ngan, C., Ohm, R.A., et al. (2015). Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. Genome Biol. Evol. *7*, 1590–1601.

121. Benton, M., Donoghue, P., Vinther, J., Asher, R., Friedman, M., and Near, T. (2015). Constraints on the timescale of animal evolutionary history. Palaeontol. Electron. *18*, 1–116.

122. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

123. Thorne, J.L., Kishino, H., and Painter, I.S. (1998). Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. *15*, 1647–1657.

124. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238.

125. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. *49*, D545–D551.

126. Busk, P.K., Pilgaard, B., Lezyk, M.J., Meyer, A.S., and Lange, L. (2017). Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. BMC Bioinformatics *18*, 214.

127. R Core Team (2021). R: a language and environment for statistical computing (R Foundation for Statistical Computing).

128. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

129. Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. *47*, D351–D360.

130. Chen, F., Mackey, A.J., Stoeckert, C.J., Jr., and Roos, D.S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. *34*, D363–D368.

131. Gladyshev, E.A., Meselson, M., and Arkhipova, I.R. (2008). Massive horizontal gene transfer in bdelloid rotifers. Science *320*, 1210–1213.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Wheat Germ Agglutinin, Tetramethylrhodamine Conjugate	Invitrogen	W849
Critical commercial assays		
Diatom DNA Prep 200 Kit	Isogen Laboratory	N/A
RNeasy Kit	QIAGEN	74104
TruSeq DNA Library Prep Kit	Illumina	FC-121-2001
TruSeq RNA Library Prep Kit	Illumina	RS-122-2001
Nextera Mate Pair Library Preparation Kit	Illumina	FC-132-1001
Deposited data		
Sequencing data and assembly of *Amoeboaphelidium protococcarum* strain X5	This paper	NCBI BioProject: PRJNA807603
Sequencing data and assembly of *Amoeboaphelidium protococcarum* strain FD95	This paper	NCBI BioProject: PRJNA817550
Sequencing data and assembly of *Amoeboaphelidium occidentale* strain FD01	This paper	NCBI BioProject: PRJNA817714
Phylogenetic datasets and analyses	This paper	Mendeley Data: https://doi.org/10.17632/7jrcwrxzb9.1
Experimental models: Organisms/strains		
Amoeboaphelidium protococcarum: strain X5	Stock Collection of ZIN RAS	N/A
Amoeboaphelidium protococcarum: strain FD95	Letcher et al.[50]	N/A
Amoeboaphelidium occidentale: strain FD01	Letcher et al.[7]	N/A
Software and algorithms		
Trimmomatic 0.36	Bolger et al.[52]	RRID: SCR_011848
Velvet 1.2.10	Zerbino and Birney[53]	RRID: SCR_010755
GapCloser 1.12	Luo et al.[54]	RRID: SCR_015026
GapFiller 1.10	Boetzer and Pirovano[55]	https://sourceforge.net/projects/gapfiller
QUAST 4.6.3	Gurevich et al.[56]	RRID: SCR_001228
NCBI BLAST 2.2.29+	Altschul et al.[57]	RRID: SCR_004870
MetaGene	Noguchi et al.[58]	http://metagene.nig.ac.jp/metagene/metagene.html
ClaMS	Pati et al.[59]	RRID: SCR_004929
Augustus 3.2.2	Stanke et al.[60]	RRID: SCR_008417
TopHat 2.0.13	Trapnell et al.[61]	RRID: SCR_013035
MAKER 2.31.6	Holt and Yandell[62]	RRID: SCR_005309
GeneMark-ES 4.32	Lomsadze et al.[63]	RRID: SCR_011930
CEGMA	Parra et al.[64]	RRID: SCR_015055
RepeatModeler open-1.0.8	Smit and Hubley[65]	RRID: SCR_015027
RepeatMasker open-4.0.5	Smit et al.[66]	RRID: SCR_012954
Sequence Manipulation Suite 2	Stothard[67]	https://www.bioinformatics.org/sms2/
FACIL 1.0	Dutilh et al.[68]	RRID: SCR_004375
BUSCO 3.0.0	Seppey et al.[69]	RRID: SCR_015008
CD-HIT 4.6.5	Li and Godzik[70]	RRID: SCR_007105
TranslatorX 1.1	Abascal et al.[71]	RRID: SCR_014733
MAFFT 7.215	Katoh and Standley[72]	RRID: SCR_011811
MCScanX	Wang et al.[73]	RRID: SCR_022067
Circos 0.69	Krzywinski et al.[74]	RRID: SCR_011798

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MEGA Software 7.0.21	Kumar et al.[75]	RRID: SCR_000667
BEAST 2.4.5	Bouckaert et al.[76]	RRID: SCR_017307
KaKs_Calculator 2.0	Wang et al.[77]	RRID: SCR_022068
ALLPATHS-LG 44837	Gnerre et al.[78]	RRID: SCR_010742
Bowtie 2.3.2	Langmead and Salzberg[79]	RRID: SCR_016368
FreeBayes 0.9.20	Garrison and Marth[80]	RRID: SCR_010761
iTOL	Letunic and Bork[81]	RRID: SCR_018174
OrthoFinder 2.5.4	Emms and Kelly[82]	RRID: SCR_017118
BioEdit 7.2.5	Hall[83]	RRID: SCR_007361
SCaFoS 1.25	Roure et al.[84]	https://megasun.bch.umontreal.ca/Software/scafos/scafos.html
IQ-TREE 1.6.12	Nguyen et al.[85]	RRID: SCR_017254
BaCoCa 1.105.r	Kück and Struck[86]	https://github.com/PatrickKueck/BaCoCa
PhyloBayes 4.1e	Lartillot et al.[87]	RRID: SCR_006402
PhyloBayes-MPI 1.8c	Lartillot et al.[88]	https://github.com/bayesiancook/pbmpi
PhyloFisher 1.0.11	Tice et al.[89]	https://github.com/TheBrownLab/PhyloFisher
Seaborn	Waskom et al.[90]	https://seaborn.pydata.org/
SortaDate	Smith et al.[91]	https://github.com/FePhyFoFum/SortaDate
Tracer 1.6	Rambaut et al.[92]	RRID: SCR_019121
Count 10.04	Csűrös[93]	http://www.iro.umontreal.ca/~csuros/gene_content/count.html
KEGG Automatic Annotation Server	Moriya et al.[94]	https://www.genome.jp/kegg/kaas/
KEGG Mapper	Kanehisa et al.[95]	https://www.genome.jp/kegg/mapper/
dbCAN2	Zhang et al.[96]	http://cys.bios.niu.edu/dbCAN2
HMMER 3.1b2	Eddy[97]	RRID: SCR_005305
trimAl 1.4.1	Capella-Gutiérrez et al.[98]	RRID: SCR_017334
SignalP 4.1	Almagro Armenteros et al.[99]	RRID: SCR_015644
ImageJ 1.48	Schneider et al.[100]	RRID: SCR_003070
DIAMOND 2.0.11.149	Buchfink et al.[101]	RRID: SCR_016071
InterProScan 5.52	Jones et al.[102]	RRID: SCR_005829
Kinannote 1.0	Goldberg et al.[103]	RRID: SCR_000352
TMHMM 2.0c	Krogh et al.[104]	RRID: SCR_014935
HH-suite 3.1.0	Steinegger et al.[105]	RRID: SCR_016133
SMART	Letunic et al.[106]	RRID: SCR_005026

RESOURCE AVAILABILITY

Lead contact

Further information and requests for data generated in the study should be directed to and will be fulfilled by the lead contact, Kirill V. Mikhailov (kv.mikhailov@belozersky.msu.ru)

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Assemblies, annotations, and raw sequencing reads for *Amoeboaphelidium protococcarum* strains X5, FD95, and *Amoeboaphelidium occidentale* strain FD01 have been deposited at the NCBI database and are publicly available. Corresponding NCBI BioProject accession numbers are listed in the key resources table. Datasets used in the phylogenetic analyses have been deposited at Mendeley Data and are publicly available as of the date of publication. DOI is listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The strain X5 of *Amoeboaphelidium protococcarum* was part of the Collection of Algae of Leningrad University (CALU),[107] currently maintained at the stock collection of ZIN RAS. The strain was originally isolated from an algal enrichment culture of aqueous samples collected from the hydrothermal springs near the Paratunka township in southern Kamchatka in August of 1966. The culture is maintained using the algal host *Scenedesmus obliquus* grown on a solid mineral medium[108] at 25°C in the presence of white light. For genomic sequencing the cultures of *A. protococcarum* X5 and its algal host were treated with antibiotics to remove bacterial contamination. An aliquot of *A. protococcarum* X5 cell suspension was incubated for 16 hours in a solution with kanamycin (40 μg/ml) and streptomycin (100 μg/ml), collected by centrifugation, mixed with an aliquot of the axenic culture of *S. obliquus* and plated on a solid culture medium containing ampicillin (100 μg/ml) and cefotaxime (100 μg/ml). The inoculated algal culture was incubated under normal growth conditions for 2 weeks, producing plaques in the algal lawn as a result of the spread of the parasitoid. A single plaque was then used for inoculating another aliquot of the algal culture and incubated to obtain sufficient material for sequencing.

Amoeboaphelidium protococcarum strain FD95 and *Amoeboaphelidium occidentale* strain FD01 were isolated from outdoor algal ponds of *Scenedesmus dimorphus* grown for biofuel production in New Mexico, USA. The parasitoids were detected in samples collected from the ponds using microscopy and isolated by plaque-plating as described in Letcher et al.[7] Ten-fold serial dilutions of the infected culture were placed in 96-well plates. One-tenth mL of each dilution was added to 1 mL of a saturated axenic culture of *S. dimorphus* grown in modified artificial seawater media MASM(D) and 4 mL of 0.75% soft agar in 15 mL culture tubes. Culture tubes were mixed thoroughly and poured onto solid agar plates. Plates were placed in an acrylic box maintained at 33°C with continuous light (Utilitech Lighting 4100 K T8 light bulbs, 200 microEinsteins) and a CO_2 flow rate of 0.3 L/min. Plaques were generated in approximately 5 to 7 days.

METHOD DETAILS

Genome sequencing and assembly

The DNA was extracted from plaques in algal cultures using a Diatom DNA Prep kit (Isogen Laboratory) after lysing the cells in the RLT buffer (RNeasy kit, Qiagen) and incubating the lysate with silica-gel granules in solution with 1/3 volume of ethanol. The RNA was extracted using an RNeasy kit (Qiagen) following the manufacturer's protocol with on column DNase digestion. The genome sequencing of *Amoeboaphelidium* species was done with an Illumina HiSeq 2000 system. The genomes of *A. protococcarum* strains X5 and FD95 were sequenced using paired-end and mate-pair libraries, and the genome of *A. occidentale* strain FD01 was sequenced using a single paired-end library. The paired-end and mate-pair libraries were prepared following the TruSeq and Nextera library preparation protocols (Illumina). Two paired-end and two mate-pair libraries were prepared for *A. protococcarum* X5 with the estimated mean insert lengths of 170, 280, 3550, and 6950 bp. One paired-end and two mate-pair libraries were prepared for *A. protococcarum* FD95 with the insert lengths of 320, 4700, and 7700 bp. The paired-end library for *A. occidentale* FD01 was prepared with an insert length of 320 bp. The 2×100 bp read pairs were generated for each library, with the total number of read pairs in the range of 17-20 million for the paired-end libraries and 24-29 million for the mate-pair libraries, which ultimately amounts to the average sequencing depth of approximately 300× for *A. protococcarum* X5 and 150× for *A. protococcarum* FD95 and *A. occidentale* FD01.

Genomic assemblies were performed with Velvet[53] using k-mer length of 77. Prior to assembly the reads were trimmed to remove sequencing and junction adapters with Trimmomatic.[52] For the assemblies of hybrid genomes of *A. protococcarum* X5 and FD95 the maximum divergence rate and maximum gap count parameters of the Velvet assembler were set to 0 to avoid undue merging of homoeologous genomic regions. The gaps in scaffolds were closed successively by the GapCloser of the SOAPdenovo package[54] and the GapFiller[55] programs. The genome assembly statistics were evaluated by QUAST.[56]

The assemblies were screened for prokaryotic contamination by performing BLAST searches[57] against the NCBI non-redundant database[109] with translated ORF sequences predicted by MetaGene.[58] Additional filtering was done for the assembly of *A. occidentale* FD01 using iterative ClaMS[59] to categorize several sources of contamination. Contigs with identified bacterial or eukaryotic contaminants or contigs shorter than 500 bp were discarded from the assemblies.

Genome annotation

Gene prediction in the genomes of *A. protococcarum* X5 and FD95 was done with Augustus[60] using a non-standard genetic code: TAA and TAG coding for glutamine. To assist the gene prediction we constructed and sequenced an RNA-Seq library for *A. protococcarum* X5. The paired-end RNA-Seq library was prepared using the TruSeq library preparation protocol and sequenced with a HiSeq 2000 system (Illumina) generating 11 million 100-bp read pairs. The reads were mapped to the assembly of *A. protococcarum* X5 with TopHat2[61] and the mapping was provided to Augustus as extrinsic evidence for gene prediction. The model parameters of Augustus were trained and optimized iteratively using predictions supported by the RNA-Seq data. The constructed Augustus model was used for gene prediction in both strains of *A. protococcarum*. Gene prediction in the genome of *A. occidentale* FD01 was performed with the MAKER2 pipeline[62] using Augustus and GeneMark-ES[63] gene predictors. The Augustus model for *A. occidentale* was constructed using gene predictions reported by the CEGMA pipeline[64] and the best scoring predictions derived from the preliminary run of the MAKER2 pipeline. The UniProtKB/Swiss-Prot database[110] and Augustus gene predictions from *A. protococcarum* were provided as homology based evidence in the annotation pipeline for *A. occidentale*. Repeats in the

assemblies were identified by the RepeatModeler program[65] and classified with the RepeatMasker[66] using the 2014-01-31 repeatmasker version of the Repbase.[111] Codon usage tables for predicted transcript sequences were generated with the Sequence Manipulation Suite.[67] Prediction of the genetic code in mitochondrial genomes was done with FACIL.[68] Completeness of predicted gene sets was evaluated using BUSCO.[69]

Analysis of hybrid genomes

Similarity clustering of predicted transcript sequences in the genomes of *A. protococcarum* X5 and FD95 was performed with the cd-hit-est program of the CD-HIT package.[70] The following clustering algorithm parameters were used: accurate clustering mode (-g 1), only +/+ strand alignment (-r 0) with an alignment bandwidth of 1000 (-b 1000); the final set of clusters was generated with the sequence identity and length difference thresholds of 0.8 (-c 0.8 -s 0.8). The intra and intergenomic alignments of transcript sequences in the resulting clusters were done with the TranslatorX[71] and MAFFT[72] alignment programs using the "ciliate nuclear" genetic code for amino acid-based nucleotide alignments. Analysis of intragenomic synteny was performed by identifying the colinear blocks of homoeologous genes with MCScanX[73] using the default parameters; all-vs-all search for predicted protein sequences was done with BLAST.[57] Genomic diagrams depicting the identified colinear regions were drawn with Circos.[74]

For the alignments of "2+2" clusters, containing a homologous gene pair from each of the strains of *A. protococcarum*, we performed phylogeny reconstructions using the UPGMA algorithm implemented in MEGA.[75] An UPGMA tree was constructed for each cluster on the basis of nucleotide p-distance. The resulting tree topologies were classified into six categories according to the inferred root position and the branching order of homologous genes from the two strains. The dominant tree topology (((X5,FD95),FD95),X5), which was recovered in 3,968 out of 4,585 clusters, was then used to differentiate the homoeologous genes in intragenomic pairs: the closest related genes of X5 and FD95 were classified as subgenomes "a" of the respective strains, and their deeper-branching homoeologs were assigned to subgenomes "b". For the estimates of genetic distances between the four subgenomes we selected a total of 2,132 clusters where subgenome assignments were additionally supported by the adjacent genes in the genomic sequences. The aligned transcript sequences representing each of the four subgenomes (X5a, X5b, FD95a, and FD95b) were concatenated into a 3.4 Mb alignment and analyzed as an unpartitioned dataset with BEAST2.[76] The BEAST analysis was performed using the lognormal relaxed clock, Yule model, and with GTR+G4+I substitution model parameters estimated from the data in 100 million MCMC samples. The generated trees were summarized with a 50% burn-in using the TreeAnnotator program and visualized with DensiTree.[112] The values of synonymous and nonsynonymous divergence for the concatenate and for individual gene alignments were calculated with the KaKs_Calculator2.0[77] using the "ciliate nuclear" genetic code and the model averaging method of parameter estimation.

Variant calling

The variants in the three sequenced genomes were detected using read mapping of the respective paired-end libraries. Prior to mapping the sequencing reads were processed using the error correction module of the ALLPATHS-LG pipeline.[78] The mapping of reads to the genome assemblies was performed with bowtie2.[79] The variants were called from the generated read alignments with FreeBayes[80] using naive variant calling and filtered to remove calls with phred quality below 20. Potentially artefactual variants associated with repetitive or poorly covered genomic regions were excluded by filtering out calls from the alignment regions with over twice or below half the mean depth of coverage.

Dataset preparation for phylogenomic analysis

The dataset for phylogenetic reconstructions was assembled on the basis of orthogroup inference performed with OrthoFinder.[82] Orthogroups were generated from the genomic data of 52 opisthokont species, including the newly obtained *Amoeboaphelidium* species, and 3 non-opisthokont species. The protein sequences were collected from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) and the JGI Genome Portal (https://genome.jgi.doe.gov/portal/). For orthogroup inference and phylogenetic analysis the protein sequences from both strains of *A. protococcarum* were clustered, and the clustered sequences were treated as a single proteome of the *A. protococcarum* species. Data from the transcriptomes of *Paraphelidium tribonemae*[10] and *Parvularia atlantis*[113] were added to the dataset using bi-directional BLAST approach with the orthogroup sequences. Algal contamination was filtered from the *Paraphelidium* data using BLAST searches against the available stramenopile genomic (*Nannochloropsis gaditana*, *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*) and transcriptomic (*Spumella elongata*) data.

Prospective orthogroups for phylogenetic reconstructions were identified by enumerating species with a single ortholog in each group. Orthogroups with the highest number of single gene representatives were inspected manually at the level of sequence alignments to ensure lack of artefactual or unexpectedly divergent sequences or spurious orthologies. Additionally, we considered several orthogroups with multiple representatives per species, when they included easily distinguishable components of conserved macromolecular complexes, such as subunits of RNA polymerases, mini-chromosome maintenance complex components, structural maintenance of chromosomes proteins, and proteasome subunits. In all such cases where the OrthoFinder algorithm failed to properly divvy up the group, orthologs were identified using phylogeny reconstructions. The orthogroups were selected from the candidate set through manual review, generally favoring longer and likely less compositionally-biased genes. No requirements were applied to single gene phylogenies to avoid artificially influencing the analysis. Three hundred orthogroups were selected for phylogenetic reconstructions, aiming for a total of approximately 100K sites – a tradeoff between the alignment length and the available computational resources. Orthologous sequences were aligned using MAFFT[72] with the L-INS-i algorithm. The alignments were inspected using BioEdit[83] and ambiguously aligned regions were excluded using a custom-made mask. The trimmed alignments were

concatenated using SCaFoS[84] into a matrix with 113,270 aligned amino acid sites. Constant sites were eliminated from the full concatenated alignment to reduce computation time, resulting in a 100,256-site alignment for phylogenetic reconstructions.

Site-specific rates in the concatenated alignment were estimated with IQ-TREE[85] using the LG+C60+F+G4 evolutionary model. Alignments for the data removal analysis were generated by iteratively discarding 10% of the fastest-evolving sites. Compositional heterogeneity in the alignment partitions was evaluated using the relative composition frequency variability (RCFV) metric.[114] The RCFV values were calculated for the 300-gene alignment using BaCoCa.[86] The partitions were rated by the respective RCFV values (ranging from 0.06 to 0.29) and discarded in batches comprising approximately 10% of the alignment sites starting from the most heterogeneous partitions. Recoding of the concatenated alignment was done using the Dayhoff recoding scheme with 6 amino acid groups: AGPST, DENQ, HKR, MIVL, WFY, C,[115] via the recode option of the PhyloBayes program.[88]

The gene subsampling procedure was carried out using the random resampler tool of the PhyloFisher package.[89] Datasets were generated by randomly sampling 20%, 40%, 60%, or 80% of the genes in the 300-gene dataset. Forty replicates were generated by sampling 20% of the dataset, twenty replicates by sampling 40%, ten replicates by sampling 60%, and five replicates by sampling 80%.

Phylogenomic analyses

Bayesian inference with the concatenated 300-gene alignment was performed with PhyloBayes MPI[88] using the CAT-GTR+G4 model – an evolutionary model with site-specific profiles, global substitution rates inferred from the data, and across-site rate variation with 4 discrete Gamma-distributed categories. Analysis of the Dayhoff-recoded 300-gene alignment was performed similarly with the CAT-GTR+G4 model and utilizing the special alphabet. The inference for each dataset used four independent chains that were run for 10,000 cycles each. The consensus trees were constructed on the basis of all four chains, sampled with a 50% burn-in and a frequency of 0.02.

Maximum likelihood analyses with the concatenated 300-gene alignment and the alignments generated by the subsampling procedure were performed with IQ-TREE.[85] Tree inference with IQ-TREE was done using the LG+C60+F+G4 evolutionary model: the LG substitution matrix combined with a 60-profile mixture model, empirical AA frequencies, and 4 categories for Gamma-distributed rates. Node support for the IQ-TREE analysis was evaluated using the ultrafast bootstrap approximation[116] with 1000 replicates and the nearest neighbor interchange tree optimization. Box plots summarizing the support values for the bipartitions of interest across the subsampling replicates were drawn using the Python data visualization library Seaborn.[90]

Approximately unbiased (AU) tests[117] were performed by IQ-TREE with the site-wise likelihoods estimated using the LG+C60+F+G4 evolutionary model. The tree obtained by IQ-TREE with the complete alignment was used as the starting tree for evaluating the alternative hypotheses. The AU tests were conducted for the complete 300-gene alignment and the alignment variants generated by the site or partition removal procedures. Alternative tree topologies for testing were constructed using the MEGA software.[75]

Molecular dating

Estimation of divergence ages was performed using PhyloBayes.[87] The constrained phylogeny for the analysis was obtained by reconciling the conflicting nodes of the PhyloBayes trees with the ML tree obtained for the 300-gene dataset. Three calibration points were applied, based on the proposed links between the evolution of Streptophyta and Fungi[118] and the available fossil record for animals. We limited the maximal age of true fungi to 890 Ma, linking it to the previously estimated upper bound for the emergence of Streptophyta[119] – a connection motivated by the reported shared appearance of pectin-degrading capabilities in the fungal lineage.[120] The minimal age for the divergence of mycorrhizal fungi was set to 470 Ma – a lower bound linked to the estimated emergence of land plants.[119] The third constraint (550-636 Ma) was applied to the radiation of bilaterian animals, confining it to the Ediacaran period, based on the fossil data.[121] A gamma of mean 2000 and standard deviation 2000 million years were used for the prior on the age of the root.

To make computations with a complex model more feasible we subsampled the original 300-gene dataset, selecting only 30 genes with the most clock-like behavior: the genes were selected using SortaDate[91] favoring the lowest root-to-tip variance in individual gene trees. Individual gene phylogenies were reconstructed using IQ-TREE[85] with automatic model selection by ModelFinder.[122] Prior to molecular dating analysis we confirmed that the 30-gene alignment (14,318 sites) reproduced the phylogeny fixed for the analysis, by reconstructing the ML tree with IQ-TREE. The PhyloBayes analysis was run under the lognormal relaxed clock model[123] with the substitution process defined by the CAT-GTR+G4 model. Two analysis chains were run for 60,000 cycles, monitoring the behavior of optimized parameters using the Tracer tool.[92] The age estimates were obtained by summarizing one of the analysis chains, using a 25% burn-in and sampling 100 data points.

Inference of ancestral gene families

For the analysis of gene family content evolution we treated the orthologous groups reconstructed using OrthoFinder[124] as synonyms of gene families. Genomic data for the following species were used for the analysis: *Neurospora crassa*, *Magnaporthe oryzae*, *Aspergillus niger*, *Tuber melanosporum*, *Yarrowia lipolytica*, *Saccharomyces cerevisiae*, *Ascoidea rubescens*, *Saitoella complicata*, *Schizosaccharomyces pombe*, *Puccinia graminis*, *Ustilago maydis*, *Malassezia globosa*, *Wallemia mellicola*, *Cryptococcus neoformans*, *Dacryopinax primogenitus*, *Coprinopsis cinerea*, *Rhizophagus irregularis*, *Mortierella verticillata*, *Umbelopsis ramanniana*, *Phycomyces blakesleeanus*, *Mucor circinelloides*, *Syncephalis plumigaleata*, *Basidiobolus meristosporus*, *Conidiobolus coronatus*, *Ramicandelaber brevisporus*, *Linderina pennispora*, *Coemansia reversa*, *Allomyces macrogynus*, *Catenaria anguillulae*, *Blastocladiella*

britannica, *Piromyces finnis*, *Gonapodya prolifera*, *Rhizoclosmatium globosum*, *Chytriomyces* sp. MP71, *Spizellomyces punctatus*, *Globomyces pollinis-pini*, *Batrachochytrium dendrobatidis*, *Amoeboaphelidium occidentale*, *Amoeboaphelidium protococcarum*; *Rozella allomycis*, *Mitosporidium daphniae*, *Fonticula alba*, *Sphaeroforma arctica*, *Capsaspora owczarzaki*, *Salpingoeca rosetta*, *Monosiga brevicollis*, *Amphimedon queenslandica*, *Nematostella vectensis*, *Homo sapiens*, *Capitella teleta*, *Drosophila melanogaster*, *Thecamonas trahens*, *Dictyostelium discoideum*, *Acanthamoeba castellanii*. The predicted genes in *A. protococcarum* strains X5 and FD95 were clustered into a non-redundant gene set, representing the *A. protococcarum* species, as detailed in the hybrid genome analysis section. The searches for the OrthoFinder workflow were performed using the BLAST algorithm.[57] For computational tractability only families found in at least two species were used in the analysis (18,062 families). Identification of ancestral families and inference of gains and losses were performed with the software package Count[93] using two approaches: the Dollo parsimony principle and the probabilistic birth-and-death model. The family-specific rates of loss, gain, and duplication along with edge lengths were optimized for the birth-and-death model using 3 Gamma-distributed categories. The optimization was performed iteratively in 5 steps, progressively increasing the model complexity, per recommendation in the manual. The maximum number of rounds for optimization and the convergence delta used default values: 100 and 0.001, respectively. For 87 families where the posterior probability calculations returned a "NaN" error, the birth-and-death modeling results for family presence, gains, and losses were substituted with simple Dollo parsimony calculations.

Functional annotation

The predicted proteomes of *Amoeboaphelidium* species were annotated using the KEGG database.[125] KEGG orthology (KO) assignments were generated by the KEGG Automatic Annotation Server[94] using the bi-directional best hit method and the default score threshold. Pathway mappings were done using the KEGG Mapper annotation tool.[95] Comparative analyses of KEGG annotations in aphelids and other holomycotan species were performed on the basis of KO assignments generated for each of their proteomes: the results of orthology assignments were summarized with a comparative table, incorporating the KEGG BRITE classification system for orthologs. The KO assignments for each species were reduced to the KO presence/absence data, and the ancestral holomycotan KO entries were determined using the Count software[93] and the Dollo parsimony principle. The violin plot and the heatmap of KO entry counts for each genome in a selection of functional categories defined by BRITE were created using the Python data visualization library Seaborn.[90]

Analysis of carbohydrate-active enzymes

Carbohydrate-active enzymes (CAZymes) were searched and classified with the dbCAN2 meta server[96] using a union of HMMER[97] and Hotpep[126] annotations. To obtain accurate counts of orthologous CAZymes in aphelids, we reconstructed phylogenies for all OrthoFinder-inferred orthogroups containing aphelid CAZymes, and inspected the trees for monophyletic groupings of aphelid sequences. The orthogroup alignments were prepared with MAFFT[72] using the L-INS-i algorithm, and trimmed with trimAl[98] using the -gappyout option. The phylogenies were reconstructed by IQ-TREE[85] with the LG+C20+F+G4 evolutionary model. The heatmap featuring CAZyme counts in aphelids and other holomycotan species was created with the Python data visualization library Seaborn,[90] and the hierarchical clustering of CAZyme families was done using the Ward's method and binary distance measure, implemented in R.[127] Prediction of signal peptides in aphelid CAZymes was performed with SignalP using default score thresholds.[99]

Domain architectures of chitin synthases (CHSs) were examined using Pfam[128] and InterPro[129] database searches. The dataset of holomycotan CHSs was assembled from a sample of 21 holomycotan species, where the CHS sequences were identified by profile searches with the Pfam domain families "chitin synthase 1" (PF01644) and "chitin synthase 2" (PF03142). The dataset of GH5 family sequences was assembled on the basis of the sequence set examined by Aspeborg et al.,[24] and expanded here with Pfam domain PF00150 searches in the genomes of fungal species and early-branching holomycotans. The datasets for CAZyme families with aphelid HGT candidates (GH1, GH81, GT2, GT34) were assembled using OrthoFinder-inferred orthogroups and expanded further with BLAST[57] searches against the NCBI's non-redundant database[109] and with sequences found in the corresponding OrthoMCL-DB[130] ortholog groups. Sequence alignments for all datasets were generated with MAFFT[72] using the L-INS-I algorithm. For the alignment of CHS sequences we used a custom mask to eliminate columns outside of the core chitin synthase domain; the alignments of GH5 family sequences and families with HGT candidates were processed with trimAl[98] using the -gappyout option or a gap threshold of 0.2, adjusted to exclude unreliable alignment regions while preventing overtrimming. The trees were reconstructed with IQ-TREE[85] using the LG+C20+F+G4 evolutionary model for the CHSs and GH5 family alignments or automatic best-fit model selection by ModelFinder[122] for HGT candidate alignments, and the node support was estimated with ultrafast bootstrap[116] and 1000 replicates. Approximately unbiased (AU) tests[117] for aphelid CAZyme HGT candidates were performed with IQ-TREE using best-fit models for each dataset.

Chitin staining

Chitin was detected using the wheat germ agglutinin tetramethylrhodamine conjugate (Life Technologies). Samples of *Amoeboaphelidium protococcarum* strain X5 were collected from plaques in the inoculated algal cultures of *Scenedesmus obliquus* grown on a solid mineral medium. Collected samples were suspended in distilled water, incubated for 10 min with the dye at 5 μg/ml concentration, and imaged using an inverted microscope Zeiss Axiovert 200M. Photographs were taken with the ORCAII-ERG2 CCD-camera and processed using ImageJ software.[100]

Global search for horizontally transferred genes

A genome-wide search for HGT candidates in aphelid genomes was performed using an Alien Index (AI) analysis[131] with a taxonomically broad selection of 67 opisthokont species and 291 non-opisthokont species acquired from the UniProt reference proteomes database.[110] We used an AI threshold of 23, which roughly corresponds to an E-value difference of ten orders of magnitude between the best hit in non-opisthokont and opisthokont species, to identify an initial set of HGT candidates. This initial set was scrutinized further using phylogenetic analyses and BLAST searches[57] against a wider set of species. To obtain orthologous groups for phylogenetic reconstructions we used OrthoFinder clustering[124] with the selected set of 358 reference proteomes plus two aphelid species. Similarity searches in OrthoFinder were carried out using DIAMOND.[101] The OrthoFinder-inferred orthogroups containing aphelid HGT candidates were aligned using MAFFT[72] with the L-INS-i algorithm and processed with trimAl[98] using a gap threshold of 0.1. Phylogenetic reconstructions were performed by IQ-TREE[85] using ModelFinder[122] for automatic best-fit model selection, and evaluating support using ultrafast bootstrap[116] with 1000 replicates. By relying on the obtained phylogenies and alignment inspection we filtered the 442 initial findings in 203 orthogroups, discarding dubious candidates that fell into clusters with UFBoot support below 95% or generally had poorly resolved phylogenetic affinities or grouped with opisthokont species. Additionally, we identified and excluded multiple candidates picked up through spurious matches over architectures with expansions of repeat domains, such as Ankyrin, Kelch, and WD40. The filtered set of aphelid HGT candidates, which included 193 findings in 98 orthogroups, was queried against the entire reference proteomes database and the NCBI non-redundant database[109] to check for agreement between the phylogeny-derived taxonomic affiliations and similarity search results, and determine the likely source of horizontal transfer. Annotations for the HGT candidates were acquired using InterProScan searches.[102]

Analysis of protein kinases

Identification and classification of protein kinases was performed with Kinannote[103] using non-metazoan classification setting and built-in cutoffs. The sequences of catalytic domains of identified protein kinases in aphelids were aligned with MAFFT[72] using the L-INS-i algorithm and trimmed with trimAl[98] using a gap threshold of 0.5. The tree of aphelid protein kinases was reconstructed from the trimmed alignment of catalytic domains with IQ-TREE[85] using the LG+C20+F+G4 evolutionary model, with node support estimated by ultrafast bootstrap[116] with 1000 replicates. The domains in protein kinases were explored using HMMER searches[97] with the PfamScan tool against the Pfam database.[128] Transmembrane regions were predicted with the TMHMM program[104] and secretory signal peptides were predicted with SignalP[99] using default score thresholds. The illustration of aphelid protein kinase phylogeny with the domain architectures was drawn using the iTOL service.[81] The predicted extracellular regions of kinases in the aphelid TKL assemblage were explored by constructing alignments of full kinase sequences and identifying archetypes for the domain architectures in each sequence cluster. The repeat units found in these extracellular regions were aligned by MAFFT[72] and classified using HMM searches with the HH-suite[105] against the Pfam database.[128]

Hybrid histidine kinases were detected in aphelids by BLAST searches.[57] The identified hybrid histidine kinase sequences were added to the dataset of Hérivaux et al.,[31] and aligned with MAFFT[72] using the L-INS-i algorithm. The alignment was trimmed by trimAl[98] using a gap threshold of 0.1, and the phylogeny was reconstructed by IQ-TREE[85] using the best-fitting LG+R6 evolutionary model, with node support estimated by ultrafast bootstrap[116] with 1000 replicates. Domain architectures of aphelid sequences were explored using SMART analysis service.[106]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical support for reconstructed phylogenies was evaluated using parameters specific to each of the employed programs. Bayesian inference with PhyloBayes[88] used four independent chains for each analysis. The chains were run for 10,000 cycles each and sampled with a frequency of 0.02. The consensus tree and posterior probability values were obtained by sampling all four analysis chains with a 50% burn-in. Chain convergence details were checked using the bpcomp utility of the PhyloBayes software, and the effective sample sizes were inspected using the Tracer tool.[92] We note that the analyses fail to achieve convergence across chains: maximal difference between bipartitions is 1.0. Support values for the ML phylogenies reconstructed with IQ-TREE[85] were calculated using ultrafast bootstrap approximation UFBoot.[116] Bootstrap support calculations utilized 1,000 replicates and the nearest neighbor interchange tree refinement. Hypothesis testing for tree topologies was carried out with IQ-TREE using 10,000 replicates for multiscale bootstrap. The alternative tree topologies were evaluated by the approximately unbiased (AU) test[117] p-values, with the criterion for rejection defined by the 0.05 significance level. Tree topologies tested with the phylogenomic dataset and the corresponding AU test p-values are summarized in Table S1. The subsampling procedure for the phylogenomic dataset was carried out using the random resampler tool of the PhyloFisher package.[89] Confidence level for sampling all genes in the dataset was set in excess of 99.9%. Forty replicates were generated by sampling 20% of the dataset, twenty replicates by sampling 40%, ten replicates by sampling 60%, and five replicates by sampling 80%. Ultrafast bootstrap support values for bipartitions across subsampling replicates are summarized in Figure S3B using box plots with a 1.5 interquartile range threshold to specify outliers.

Current Biology, Volume *32*

Supplemental Information

Genomic analysis reveals cryptic diversity in aphelids and sheds light on the emergence of Fungi

Kirill V. Mikhailov, Sergey A. Karpov, Peter M. Letcher, Philip A. Lee, Maria D. Logacheva, Aleksey A. Penin, Maksim A. Nesterenko, Igor R. Pozdnyakov, Evgenii V. Potapenko, Dmitry Y. Sherbakov, Yuri V. Panchin, and Vladimir V. Aleoshin

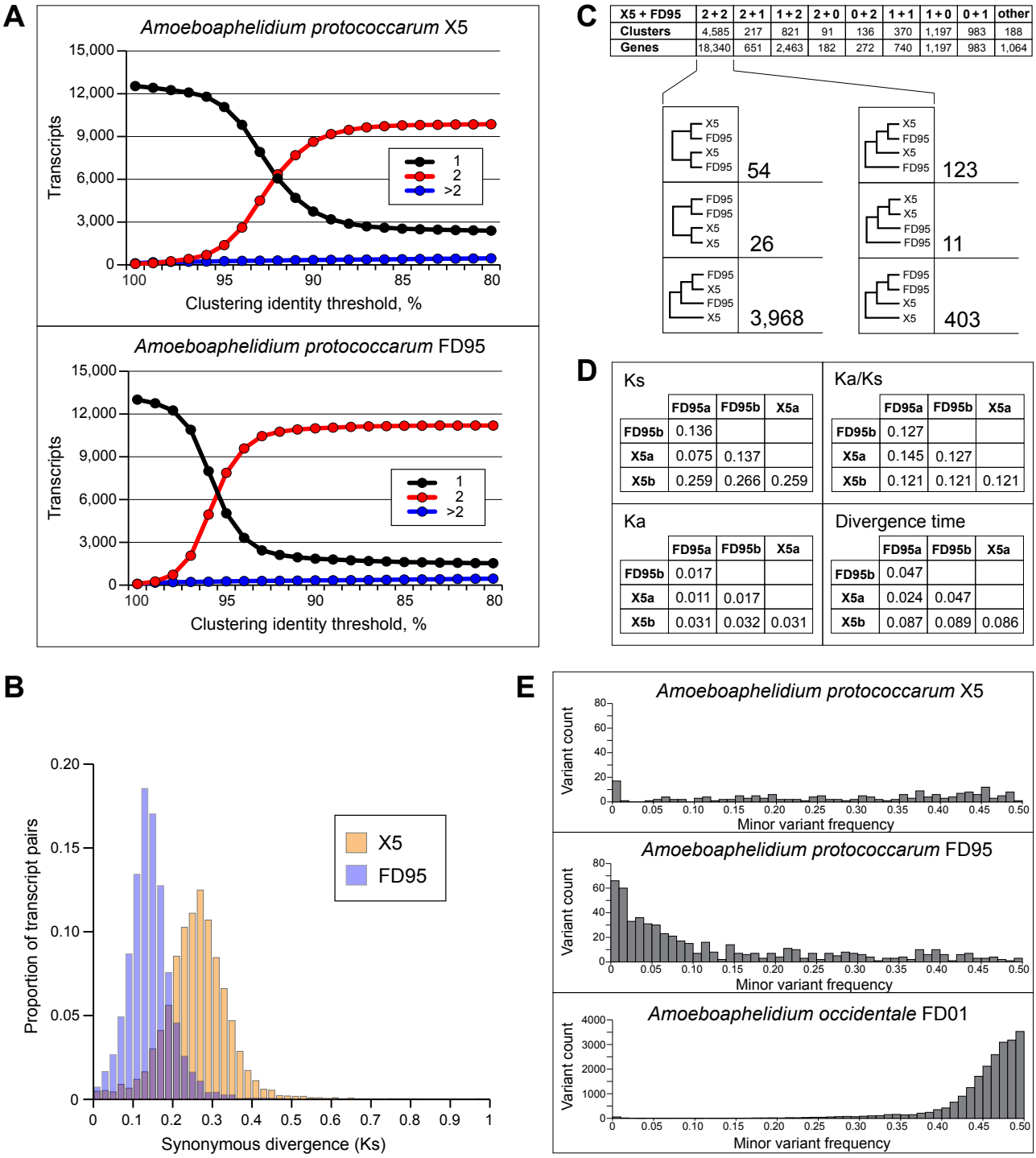


Figure S1. Characteristics of hybrid genomes of *A. protococcarum*. Related to Figure 1. (**A**) Clustering of transcripts in the genomes of *A. protococcarum* strains X5 and FD95. The cd-hit clustering was performed with a varying identity threshold (from 100% to 80% with a step of 1%), for each value of the threshold the graphs show the number of genes that fall in clusters of size 2 (red), clusters with over 2 members (blue) or remain singular (black). (**B**) Distributions of per-site synonymous divergence (Ks) values between the gene pairs in the genomes of *A. protococcarum* strains X5 and FD95. (**C**) Similarity clustering of pooled transcripts from the genomes of *A. protococcarum* strains X5 and FD95, and UPGMA tree inference for clusters containing a gene pair from each of the strains. The cd-hit clusters of all predicted transcripts in the genomes of X5 and FD95 were classified into categories according to the number genes from each strain that formed the cluster; sequences in clusters with a pair of genes from each strain ("2+2" clusters) were aligned and their phylogenetic relationship was inferred using the UPGMA method – the resulting trees were classified into the six topologies depicted in the diagram. (**D**) Estimates of synonymous divergence (Ks), nonsynonymous divergence (Ka), and divergence time in substitutions per site in the concatenated alignments of "2+2" cluster sequences following subgenome assignments of homoeologous genes. (**E**) Histograms of variant frequencies (minor variant to read depth ratios) in the mappings of paired-end reads to genome assemblies of *A. protococcarum* strains X5 and FD95 and *A. occidentale* strain FD01.

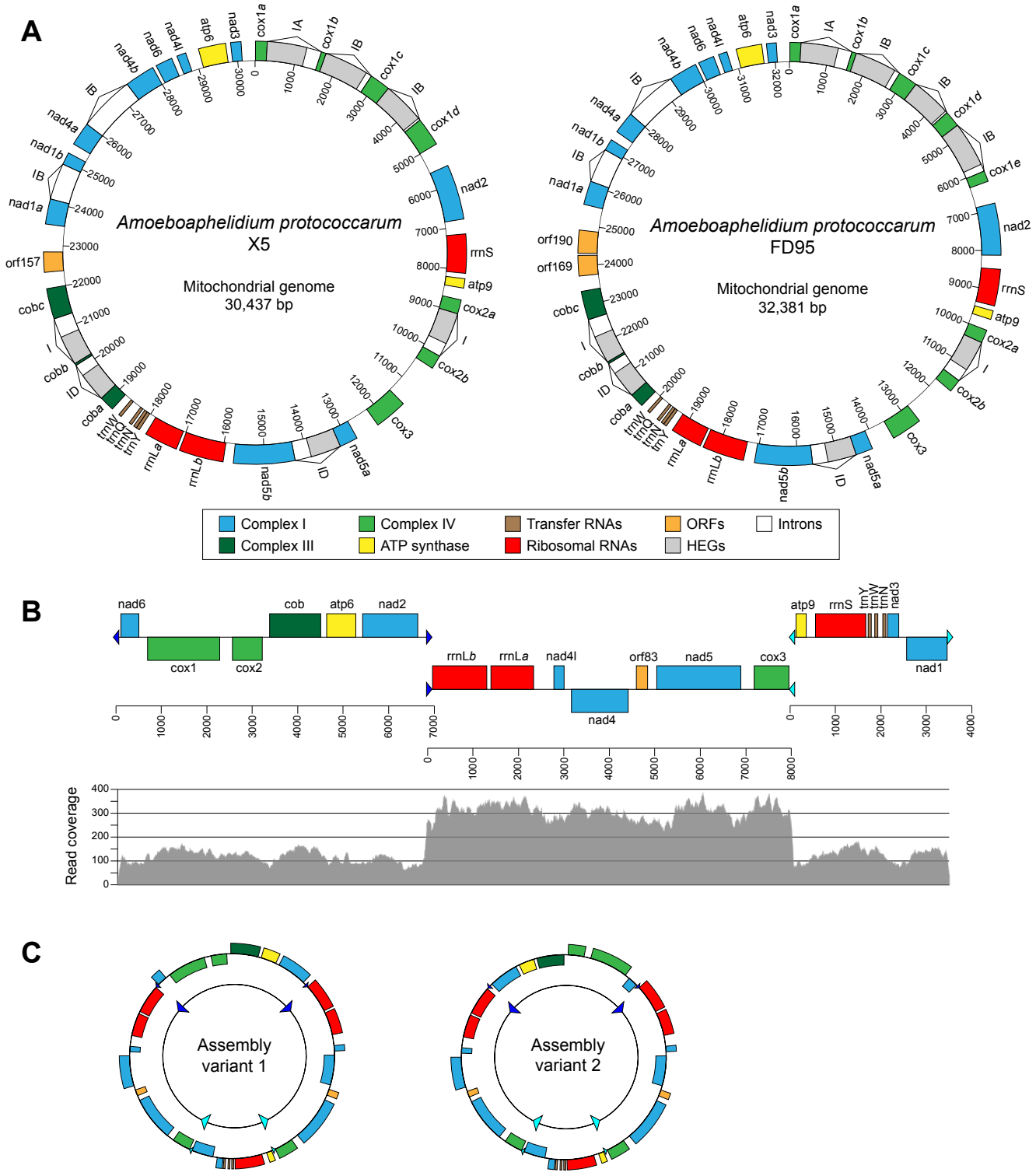


Figure S2. Mitochondrial genome assemblies of *A. protococcarum* and *A. occidentale*. Related to Figure 1. (**A**) Mitochondrial genome maps of *A. protococcarum* strains X5 and FD95. (**B**) Mitochondrial contigs of *A. occidentale* strain FD01. All genes in the mitochondrial genomes of *A. protococcarum* are encoded on a single strand; they include 13 respiratory complex components, small and large subunits of ribosomal RNA, with the latter one split into 2 parts (*rrnLa* and *rrnLb*), 4 transfer RNAs, and homing endonuclease genes (HEGs) – 7 in X5 and 8 in FD95. The mitochondrial genome in *A. occidentale* is found in 3 contigs and contains no introns or HEGs; the average read coverage of the 8 Kb contig is approximately 2.5 times higher than the coverage of the shorter contigs. The overlaps between contigs of *A. occidentale* are represented with light blue and dark blue arrowheads. (**C**) Two possible models for the assembly of a 26,157 bp circular mitochondrial genome for *A. occidentale* with the 8 Kb contig incorporated as an inverted repeat.

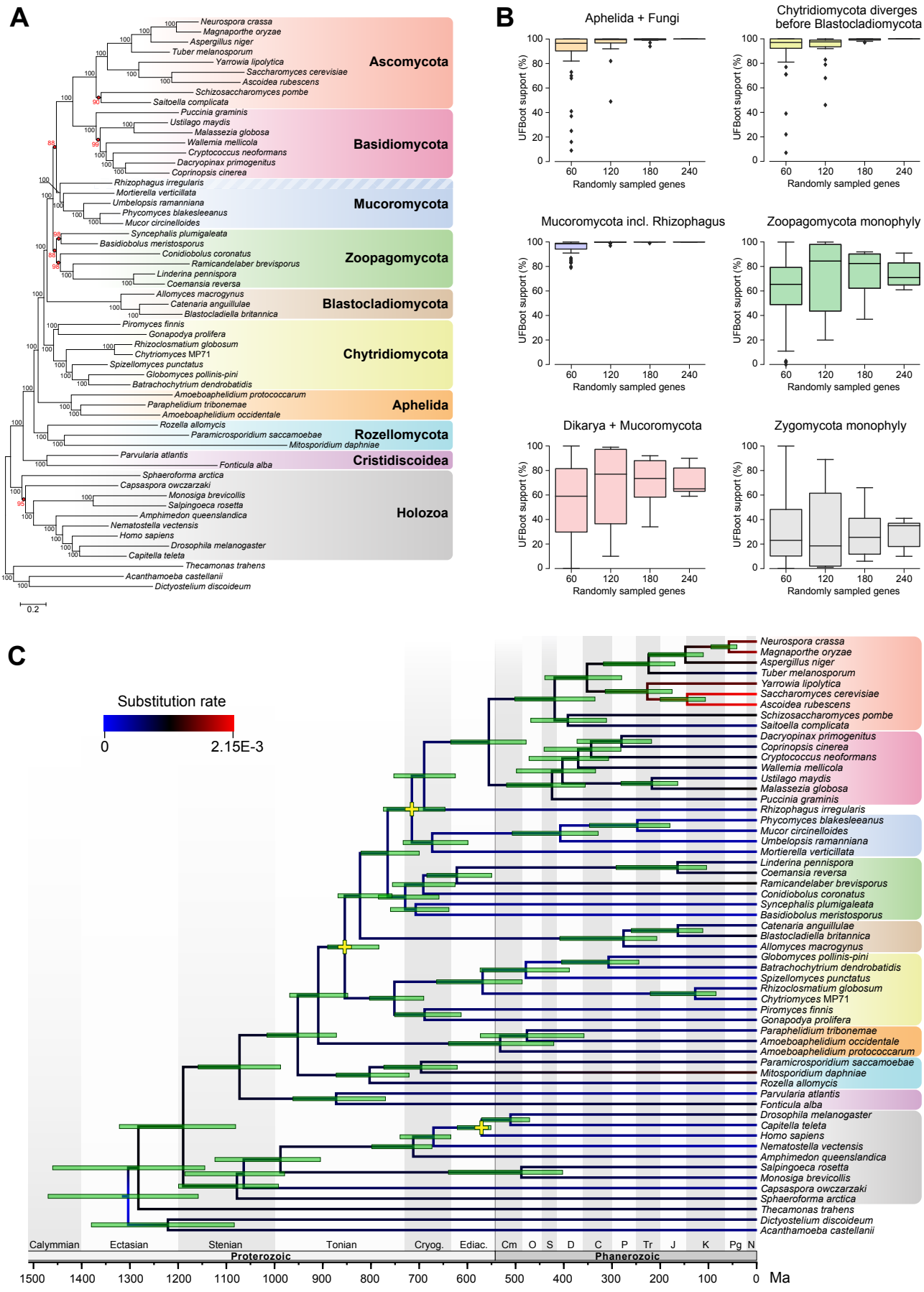


Figure S3. Assessment of the influence of gene subsampling on the stability of phylogeny and divergence date estimates using time-calibrated analysis. Related to Figure 2. (**A**) IQ-TREE maximum likelihood reconstruction with the 300-gene alignment using the LG+C60+F+G4 evolutionary model; node support was calculated using the ultrafast bootstrap approximation with 1000 replicates; nodes with support values below 100% are marked in red. (**B**) Impact of gene subsampling on the bipartitions of interest; the 300-gene dataset was used to randomly sample 20%, 40%, 60%, and 80% of

genes, which were concatenated and analyzed with IQ-TREE; 40 replicates were generated for the 60-gene dataset, 20 replicates for the 120-gene dataset, 10 replicates for the 180-gene dataset, and 5 replicates for the 240-gene dataset; ultrafast bootstrap support values for bipartitions across replicates are presented using box plots with a 1.5 interquartile range threshold to specify outliers. (**C**) Time-calibrated phylogeny inferred by PhyloBayes under the CAT-GTR model using a 30-gene dataset – a subset of the 300-gene dataset, selected for most clock-like behavior; green bars at the tree nodes represent 95% confidence intervals for posterior probability estimates of divergence times; the analysis was performed under a lognormal autocorrelated relaxed clock model with three calibration points (nodes marked with yellow crosses): setting the maximal age of true fungi at 890 Ma, the minimal age for the divergence of mycorrhizal fungi at 470 Ma, and confining the divergence of the bilaterian lineage within the 550-636 Ma time interval; the maximal and minimal ages for fungal divergences are motivated by the proposed links between the evolution of streptophytes and fungi[S1], and the previously estimated bounds on the emergence of streptophytes and land plants[S2]; the estimated evolutionary rates of branches (in substitutions per site per million years) are presented using a color gradient.

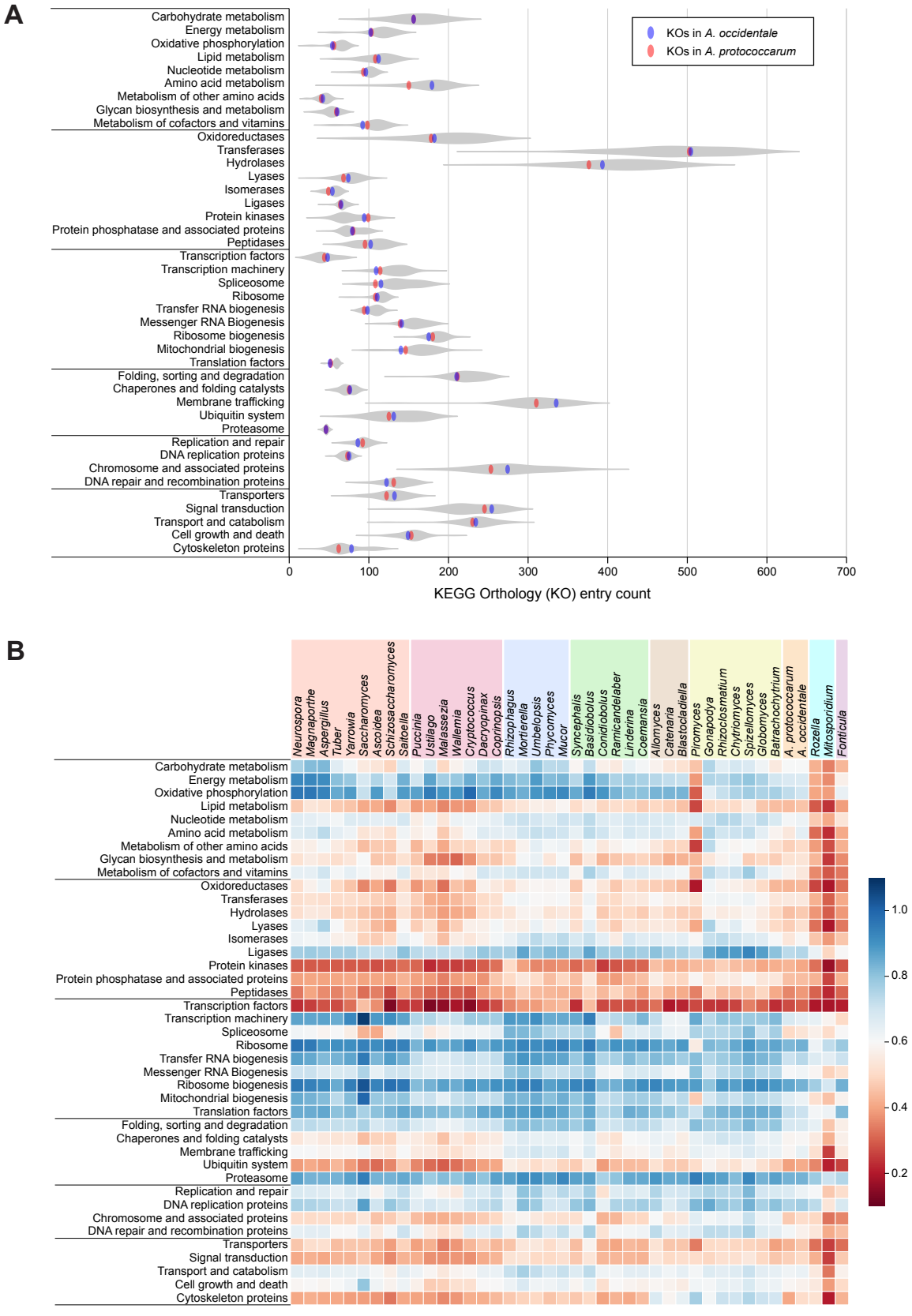


Figure S4. Comparative analysis of functional annotations for the genomes in Holomycota using KEGG. Related to Figure 3. (**A**) Distributions of annotated KEGG orthology (KO) counts in holomycotan genomes by functional categories, according to the KEGG BRITE classification; the distributions, shown as grey violin plots, were constructed from the genomic data of 40 holomycotan species, annotated using KAAS; KO annotations for each genome were reduced to KO presence/absence data; KO counts for the genomes of *A. occidentale* and *A. protococcarum* are shown as blue and red data points, respectively. (**B**) Heatmap of KO entries (presence/absence data) for BRITE functional categories in the genomes of holomycotan species, normalized to the inferred counts of unique KOs in the last common ancestor of Holomycota.

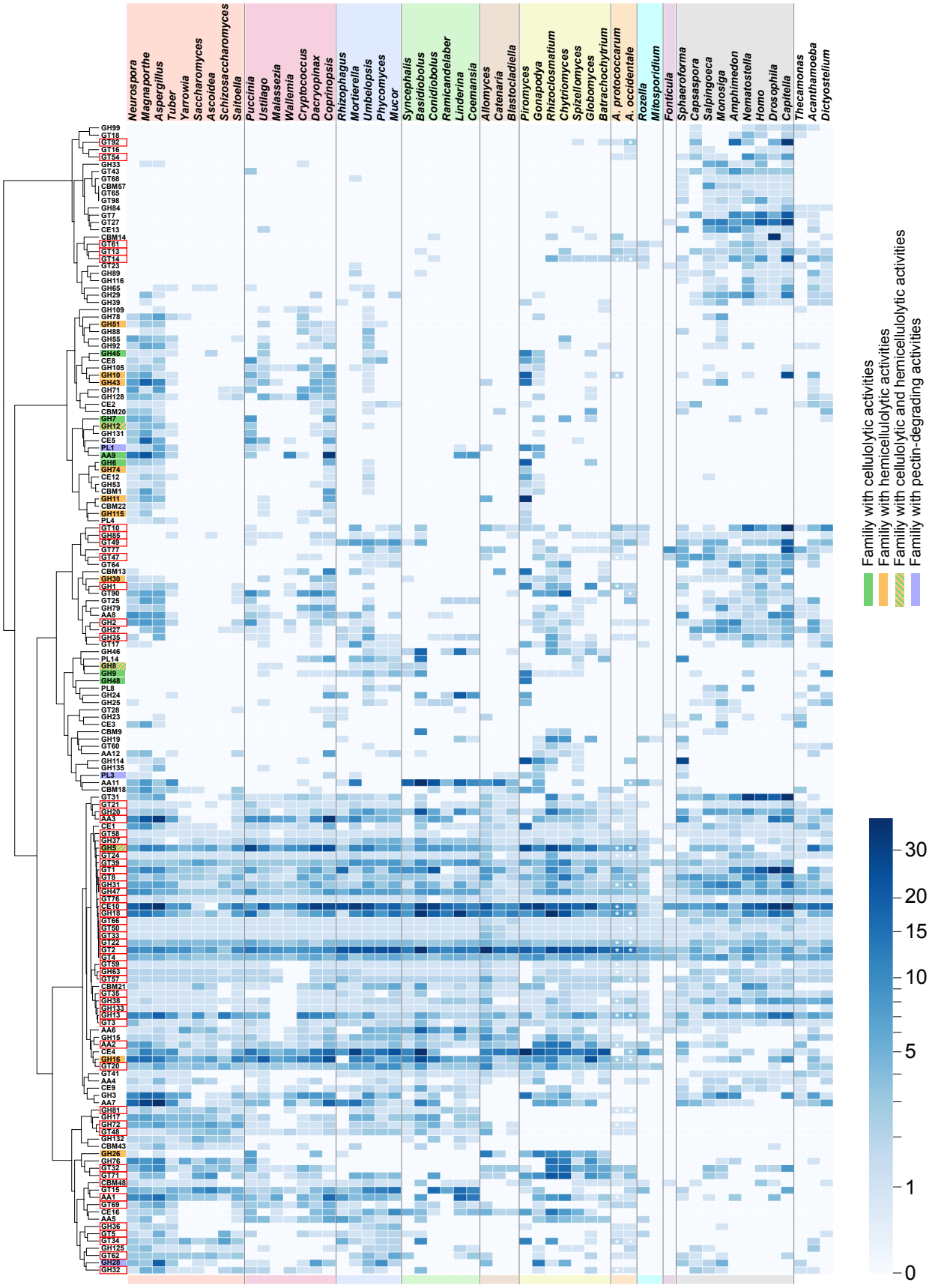


Figure S5. Heatmap of CAZy family member counts in the genomes of 54 eukaryotic species. Related to Figure 4. The coloring intensity is scaled logarithmically (scale bar to the right); CAZy families are clustered on the basis of binary presence/absence profiles in the inspected genomes using the Ward's method; families containing characterized enzymes with plant cell wall degrading activities are marked with colors: cellulolytic – green, hemicellulolytic – orange, pectin-degrading – blue; families were the two *Amoeboaphelidium* species have shared orthologous groups are marked with a red frame; white asterisks for aphelid families denote presence of sequences with a predicted secretory signal peptide.

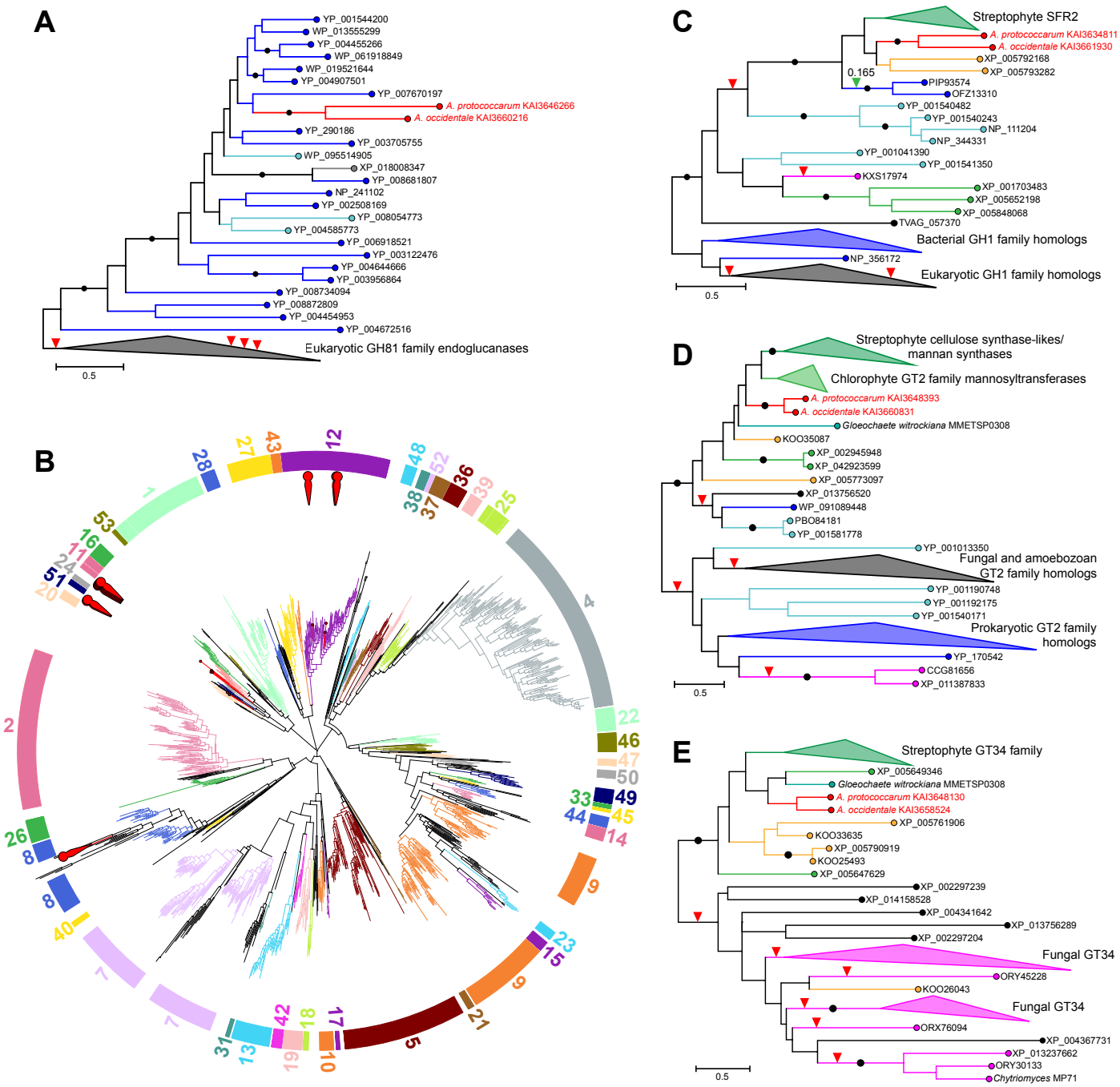


Figure S6. Maximum likelihood phylogenetic trees for GH5 family sequences and horizontal gene transfer candidates among the aphelid CAZymes. Related to Figure 4. (**A**) Phylogenetic tree with GH81 family endoglucanases; aphelid sequences are marked in red, bacterial sequences are in dark blue, and archaeal in light blue; eukaryotic cluster includes fungal, streptophyte and algal sequences; black circles on tree branches correspond to 100% UFboot support in the analysis; alternative placements for aphelid sequences, which were examined using the AU test, are labeled with triangles on tree branches: the placements rejected by the AU test at the 5% level are labeled with red triangles; for GH81 family we tested three alternative positions for aphelids within the cluster of eukaryotic sequences, by placing them at the bases of three fungal clusters – all alternatives were rejected by the test. (**B**) Phylogenetic tree with GH5 family sequences; the tree was reconstructed using the LG+C20+F+G4 model with an alignment of 1,431 Cellulase GH5 domain (PF00150) sequences; GH5 subfamilies and the corresponding subtrees are indicated in color and labeled on the outer rim of the diagram; subfamily classification and the initial sequence set are based on Aspeborg *et al*., 2012[S3] – the sequence set from the 2012 study was expanded with PF00150 domain hits from aphelid, rozellid, and fungal GH5 family enzymes; aphelid sequences in the tree are shown in red and highlighted on the rim of the diagram with red pins. (**C**) Phylogenetic tree with GH1 family SFR2 homologs; the tree colors and labels are as in (A),

additionally, green marks chlorophyte and streptophyte sequences/clusters, hapthophytes are marked in orange, and fungal sequences with magenta; a green triangle on the bacterial branch marks an alternative position for aphelid sequences that was not rejected by the AU test (p-value 0.165). (**D**) Phylogenetic tree with GT2 family putative mannosyltransferases; using the same color scheme and labels as in (A) and (C). (**E**) Phylogenetic tree with GT34 family putative xylosyltransferases; using the same color scheme and labels as in (A) and (C). The putative enzyme activities for (C-E) are based on the characterized enzymes of *Arabidopsis thaliana* most closely related to the aphelid sequences.

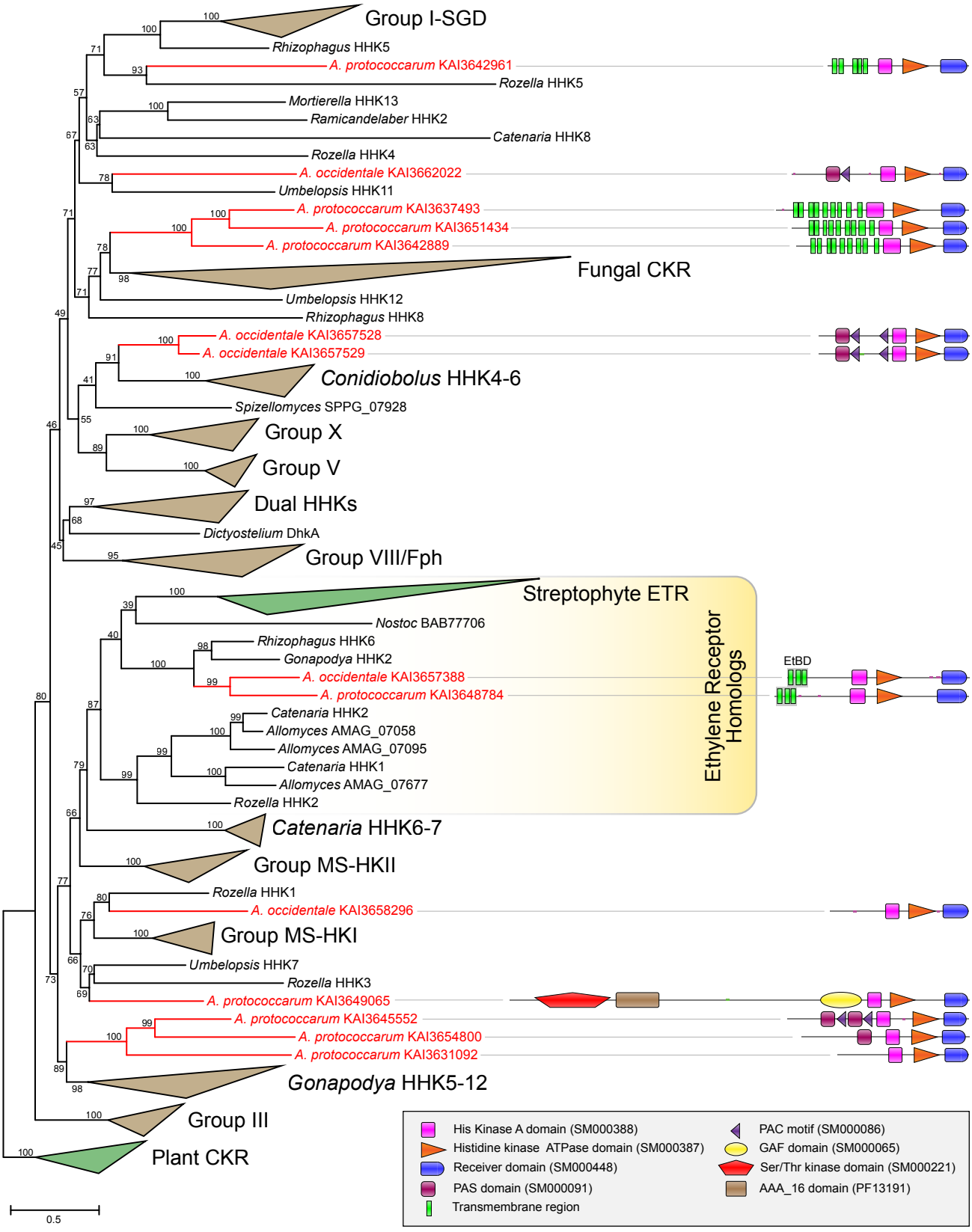


Figure S7. Phylogeny of hybrid histidine kinases with aphelid sequences. Related to Figure 5. Maximum likelihood phylogenetic tree was reconstructed by IQ-TREE with the LG+R6 model using an alignment of hybrid histidine kinase regions spanning the conserved Histidine Kinase A, ATPase, and the Receiver domains; the dataset of hybrid histidine kinases along with the sequence names and group designations are based on the Herivaux *et al*., 2017[S4]; branch support was evaluated using UF bootstrap with 1,000 replicates; highly-supported groups are collapsed in the tree; aphelid sequences are marked with red color and the corresponding protein domain architectures are depicted on the right; EtBD – ethylene binding domain.

Supplemental References

S1. Berbee, M.L., Strullu-Derrien, C., Delaux, P.M., Strother, P.K., Kenrick, P., Selosse, M.A., and Taylor, J.W. (2020). Genomic and fossil windows into the secret lives of the most ancient fungi. Nature reviews. Microbiology *18*, 717-730.

S2. Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. Proc Natl Acad Sci U S A *115*, E2274-E2283.

S3. Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., 3rd, and Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol *12*, 186.

S4. Herivaux, A., Duge de Bernonville, T., Roux, C., Clastre, M., Courdavault, V., Gastebois, A., Bouchara, J.P., James, T.Y., Latge, J.P., Martin, F., and Papon, N. (2017). The Identification of Phytohormone Receptor Homologs in Early Diverging Fungi Suggests a Role for Plant Sensing in Land Colonization by Fungi. mBio *8*.