AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Host-Microbe Biology

Molecular Dating of the Emergence of Anaerobic Rumen Fungi and the Impact of Laterally Acquired Genes

Yan Wang,[a,b] Noha H. Youssef,[c] Matthew Brian Couger,[d] Radwa A. Hanafy,[c] Mostafa S. Elshahed,[c] Jason E. Stajich[a,b]

[a]Department of Microbiology and Plant Pathology, University of California—Riverside, Riverside, California, USA
[b]Institute for Integrative Genome Biology, University of California—Riverside, Riverside, California, USA
[c]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA
[d]High Performance Computing Center, Oklahoma State University, Stillwater, Oklahoma, USA

ABSTRACT The anaerobic gut fungi (AGF), or Neocallimastigomycota, inhabit the rumen and alimentary tract of herbivorous mammals, where they play important roles in the degradation of plant fiber. Comparative genomic and phylogenomic analyses of the AGF have long been hampered by their fastidious growth condition, as well as their large (up to 200 Mb) and AT-biased (78 to 84%) genomes. We sequenced 21 AGF transcriptomes and combined them with 5 available AGF genome sequences to explore their evolutionary relationships, time their divergence, and characterize gene gain/loss patterns associated with their evolution. We estimate that the most recent common ancestor of the AGF diverged 66 (±10) million years ago, a time frame that coincides with the evolution of grasses (Poaceae), as well as the mammalian transition from insectivory to herbivory. The concordance of independent estimations suggests that AGF have been important in shaping the success of mammalian herbivory transition by improving the efficiency of energy acquisition from recalcitrant plant materials. Comparative genomics identified multiple lineage-specific genes in the AGF, two of which were acquired from rumen gut bacteria and animal hosts via horizontal gene transfer (HGT). A third AGF domain, plant-like polysaccharide lyase, represents a novel gene in fungi that potentially aids AGF to degrade pectin. Analysis of genomic and transcriptomic sequences confirmed both the presence and expression of these lineage-specific genes in nearly all AGF clades. These genetic elements may contribute to the exceptional abilities of AGF to degrade plant biomass and enable metabolism of the rumen microbes and animal hosts.

IMPORTANCE Anaerobic fungi living in the rumen of herbivorous mammals possess an extraordinary ability to degrade plant biomass. We examined the origin and genomic composition of these poorly characterized anaerobic gut fungi using both transcriptome and genomic data. Phylogenomics and molecular dating analyses found remarkable concurrence of the divergence times of the rumen fungi, the forage grasses, and the dietary shift of ancestral mammals from primarily insectivory to herbivory. Comparative genomics identified unique machinery in these fungi to utilize plant polysaccharides. The rumen fungi were also identified with the ability to code for three protein domains with putative functions in plant pectin degradation and microbial defense, which were absent from all other fungal organisms (examined over 1,000 fungal genomes). Two of these domains were likely acquired from rumen gut bacteria and animal hosts separately via horizontal gene transfer. The third one is a plant-like polysaccharide lyase, representing a unique fungal enzyme with potential pectin breakdown abilities.

KEYWORDS comparative genomics, divergence time estimation, evolution, HGT, phylogenomics

Citation Wang Y, Youssef NH, Couger MB, Hanafy RA, Elshahed MS, Stajich JE. 2019. Molecular dating of the emergence of anaerobic rumen fungi and the impact of laterally acquired genes. mSystems 4:e00247-19. https://doi.org/10.1128/mSystems.00247-19.

Editor Olga Zhaxybayeva, Dartmouth College

Address correspondence to Yan Wang, yanxw.wang@gmail.com, or Jason E. Stajich, jason.stajich@ucr.edu.

Anaerobic gut fungi inhabit rumen of herbivorous mammals. Phylogenomics of 26 isolates, 7 genera shows divergence time == transitions to herbivory & evolution of grasses. Gene content comparison reveals horiz transfer from bacteria and animals.

Received 19 April 2019
Accepted 1 August 2019
Published 27 August 2019

Diverse microbes inhabit the digestive tract of ruminant mammals and contribute to the degradation of ingested plant fibers, a process that liberates nutrients for their hosts. Large-scale genomic and metagenomic sequencing of rumen microbes have produced hundreds of novel bacterial genomes, enabling the discovery of plant biomass-degrading enzymes and patterns of genomic evolution (1, 2). However, eukaryotic members of the rumen microbial community have been less intensely studied (3, 4). Members of the phylum Neocallimastigomycota (anaerobic gut fungi [AGF]) are important members of the rumen and hindgut of a wide range of herbivorous mammals and reptiles (5). To survive in this anoxic and prokaryote-dominated environment, extant AGF members have undergone multiple structures and metabolic adaptations, including the loss of the mitochondria, gain of a hydrogenosome, loss of respiratory capacities, and substitution of ergosterol with tetrahymanol in the cell membrane (6). Importantly, all known AGF taxa have a remarkably efficient plant biomass degradation machinery, which may be critical for competing with other microbes for resources and establishing growth in the herbivorous gut. Such capacity is reflected in the possession of an impressive arsenal of plant biomass degradation enzymes and the production of the cellulosomes, extracellular structures that harbor multiple enzymes bound to scaffoldins (4). These metabolic and structural adaptations improve the survivability, fitness, and competitiveness of the AGF in the herbivorous gut, but the genetic and evolutionary origins of these changes remain largely undescribed (3, 7). Previous genomic investigations of the AGF have identified a massive number of carbohydrate-active enzymes encoded by genes with foreign origins, presumably from multiple lineages of bacteria through independent horizontal gene transfer (HGT) events (3, 4, 7). Recently, hundreds of HGT elements were detected in AGF and are suggested to have enabled the fungi to expand their substrate utilization range, augment their biosynthetic capabilities, and shape a phylogenetically distinct fungal lineage (8). In fact, HGT examples from bacteria to fungi have been documented extensively (9–12). However, HGT elements in fungi that have been transferred from other eukaryotes are still rare, with only a few described cases from animals (13), oomycetes (14), or plants (15). The rumen is an intriguing context to explore patterns of HGT, where DNA and RNA are liberated when cells are disrupted by degradative enzymes. Competing organisms can find an advantage by acquiring foreign genes that operate efficiently in an anaerobic environment to obtain nutrients from recalcitrant plant fibers or to recognize other microbes. Our study took a conservative approach to identify candidate HGT by focusing on protein domains rather than entire genes and focusing on instances that appear to be unique gains in the AGF and that are missing in all other fungal lineages.

The Neocallimastigomycota are classified within the Chytridiomycota (chytrid) fungi, which share the trait of a flagellated zoospore stage (16–19). Efforts to resolve the phylogenetic relationship of AGF and their sister lineages using ribosomal markers have yielded conflicting topologies (20, 21). A multilocus phylogeny is needed to evaluate their evolutionary relationships and to estimate the divergence time of the AGF. Using genomes and transcriptomes from 26 different AGF taxa (Table 1) covering seven out of the 10 recognized genera, we reconstructed a robust phylogenomic tree of the AGF and estimated their divergence time. We compared the genomes or transcriptomes of AGF and their non-rumen-associated relatives in Chytridiomycota to identify unique and shared genome contents. This study examined the relatively recent divergence of the AGF clade and revealed a concordance of the divergence time of the Neocallimastigomycota fungi with both the mammalian host transition to herbivory and the diversification events of the forage grasses. As the AGF are well known for their exceptional efficiency at plant biomass degradation, we also explored the diverse genetic components of these fungi. We discovered two potential HGT elements that were found to be unique to the AGF, which are predicted to have originated from animals or bacteria. Examination of the family of bacterial transferred genes revealed multiple intron insertion events that occurred after the HGT acquisition process, which are present in all five AGF genomes. Comparative analyses of these genes suggest

TABLE 1 Information for the AGF strains included in this study

Organism	Strain	Accession no.	Type	Host	Reference or source
Anaeromyces contortous	Na	GGWN00000000	Transcriptome	Cow	This study
Anaeromyces contortous	C3J	GGWO00000000	Transcriptome	Cow	This study
Anaeromyces contortous	G3G	GGWR00000000	Transcriptome	Goat	This study
Anaeromyces contortous	O2	GGWQ00000000	Transcriptome	Cow	This study
Anaeromyces contortous	C3G	GGWR00000000	Transcriptome	Cow	This study
Anaeromyces robustus	S4	MCFG00000000	Genome	Sheep	4
Caecomyces sp.	Brit4	GGWS00000000	Transcriptome	Cow	This study
Caecomyces sp.	Iso3	GGXE00000000	Transcriptome	Cow	This study
Feramyces austinii	WSF3a	GGWU00000000	Transcriptome	Aoudad	This study
Feramyces austinii	WSF2c	GGWT00000000	Transcriptome	Aoudad	This study
Orpinomyces sp.	D3A	GGWV00000000	Transcriptome	Cow	This study
Orpinomyces sp.	D3B	GGWW00000000	Transcriptome	Cow	This study
Orpinomyces sp.	D4C	GGWX00000000	Transcriptome	Cow	This study
Pecoramyces ruminantium	C1A	ASRE00000000	Genome	Cow	3
Pecoramyces sp.	S4B	GGWY00000000	Transcriptome	Sheep	This study
Pecoramyces sp.	FX4B	GGWZ00000000	Transcriptome	Cow	This study
Pecoramyces sp.	FS3c	GGXF00000000	Transcriptome	Cow	This study
Pecoramyces sp.	YC3	GGXA00000000	Transcriptome	Cow	This study
Piromyces finnis	Pirfi3	MCFH00000000	Genome	Horse	4
Piromyces sp.	E2	MCNC00000000	Genome	Elephant	4
Piromyces sp.	A1	GGXB00000000	Transcriptome	Sheep	This study
Piromyces sp.	B4	GGXH00000000	Transcriptome	Cow	This study
Piromyces sp.	B5	GGXI00000000	Transcriptome	Cow	This study
Neocallimastix californiae	G1	MCOG00000000	Genome	Goat	4
Neocallimastix frontalis	Hef5	GGXJ00000000	Transcriptome	Cow	This study
Neocallimastix sp.	G3	GGXC00000000	Transcriptome	Sheep	This study

putative intron insert events involved the intragenic duplication of coding sequences. In addition, a novel plant polysaccharide lyase was revealed from both AGF genomes and transcriptomes that has never been reported from any known fungal genomes or genetic studies. The evolutionary genomic investigation of these rumen-inhabiting fungi provides perspective on the concordant timing of their divergence with the ecological niche they inhabit and the potential role of HGT in the accumulation of lineage-specific processes that may contribute to their unique biology.

RESULTS

Divergence time estimation and phylogenomic relationship of Neocallimastigomycota. Phylogenomic analysis placed the 26 AGF taxa into a single monophyletic clade with strong support of Bayesian posterior probability (1.0/1.0) and maximum likelihood bootstrap value (100%) (Fig. 1; see also Fig. S1 in the supplemental material). All AGF genera (*Anaeromyces*, *Caecomyces*, *Feramyces*, *Neocallimastix*, *Orpinomyces*, *Pecoramyces*, and *Piromyces*) included in this study formed individual monophyletic clades that were also supported by both Bayesian (Fig. 1) and maximum likelihood (Fig. S1) analyses. A conflict in the tree topology between the two phylogenetic reconstructions is the placement of the *Caecomyces* clade. This lineage is sister to the rest of the Neocallimastigomycota in the maximum likelihood tree (Fig. S1), while the *Caecomyces* position is swapped with *Piromyces* in the Bayesian phylogeny (Fig. 1). This is likely due to short internode distances, which suggests a rapid radiation of the ancestors of the two genera. The relative short bar of the highest-probability density (HPD) on the node of the AGF clade (Fig. 1) suggests the integrative natural history of this group of fungi and the outperforming resolving power of the genome-wide data in the molecular dating analyses.

The divergence time of the Neocallimastigomycota clade is estimated at the Cretaceous/Paleogene (K/Pg) period boundary 66 (±10) million years ago (Mya) (Fig. 1). The chronogram (Fig. 1) displays a long branch leading to the emergence of the AGF clade, which extends from the end of Ediacaran period (~564 Mya) to the K/Pg period boundary (~66 Mya). This suggests that the extant members of AGF did not emerge until recently and then rapidly radiated into separate clades in the Paleogene. The

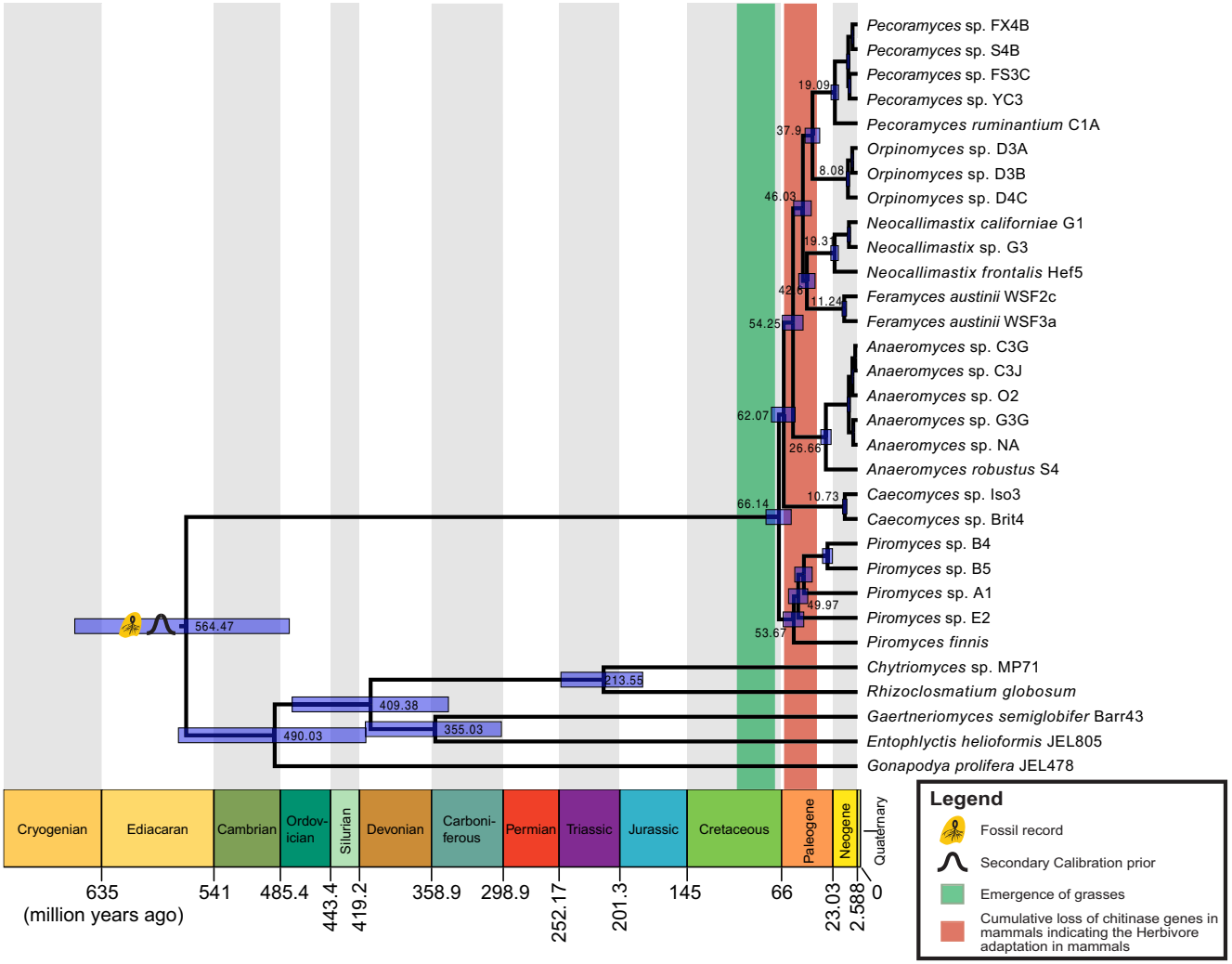

FIG 1 Bayesian phylogenomic maximum clade credibility tree of Neocallimastigomycota with divergence time estimation. All clades are fully supported by Bayesian posterior probabilities (BPP). For clarity, mean ages and 95% highest-probability density ranges (blue bars) are denoted on the nodes above the rank of genus.

estimated time frame for AGF divergence broadly coincides with the age of the grasses (70 to 95 Mya), previously estimated using molecular (nuclear and chloroplast) markers, and calibrated using fossils from pollen and dinosaur coprolite as well as the breakup time of the Gondwana (22–26). In addition, this inferred AGF divergence time also coincides with a major diet change of placental mammals, the transition from a primarily insectivorous to an herbivorous and omnivorous lifestyle. The loss of chitinase gene diversity, estimated to occurred from the K/Pg period boundary (66 Mya) to the mid-Paleogene period (34 Mya) (Fig. 1), is widely seen as a consequence of such a transition (27). Collectively, these overlapping estimates suggest that the evolution of the symbiotic association between herbivorous mammals and rumen fungi is tightly linked with the evolution of forage grasses and mammalian dietary transitions within a 66- to 95-Mya time frame. The exact chronology of these three divergence or transition events cannot be accurately determined partially due to the intervals of the estimates (Fig. 1). However, the dates inferred from phylogenetic analyses are consistent with the hypothesis that rumen fungi have played important roles in the dietary transition of some mammals to acquire nutrition from forage grasses.

Genome-wide comparison of protein domains and homologous genes. Comparative genomic analysis between AGF and their non-rumen-associated chytrid rela-

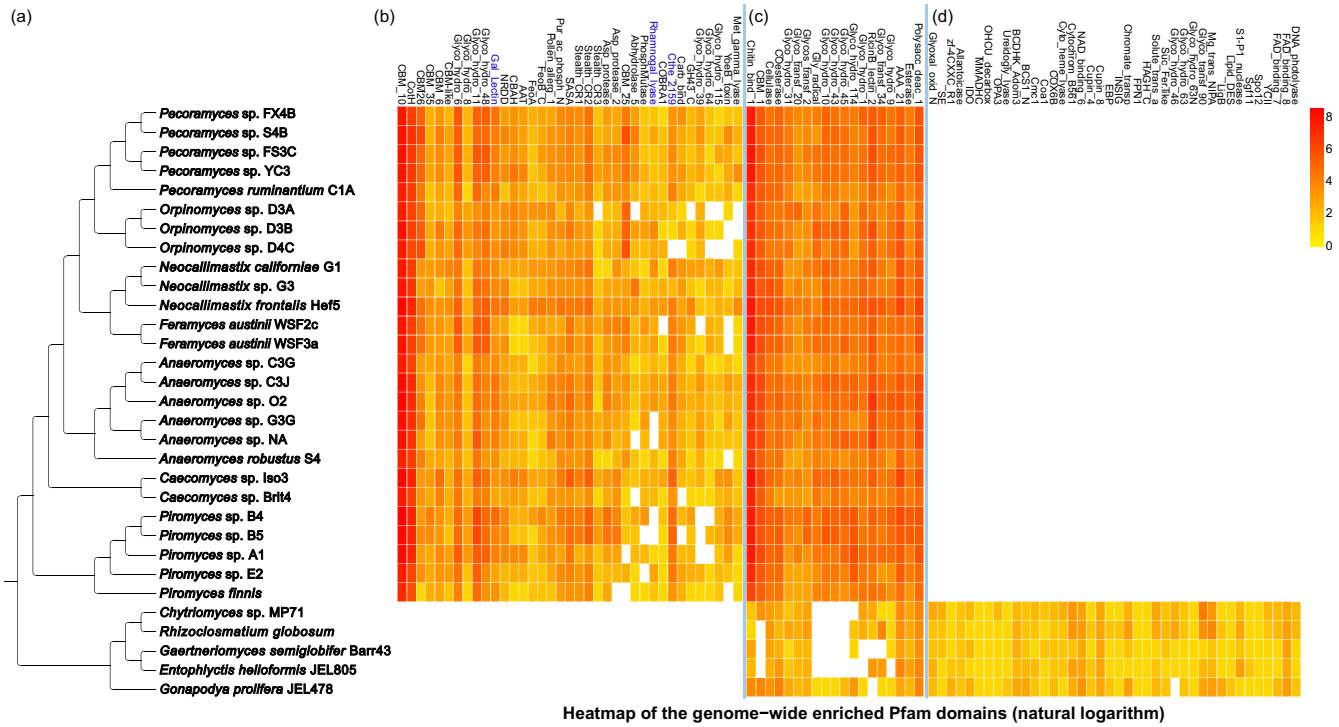

FIG 2 Cladogram and heatmap enrichment of the Pfam domains between Neocallimastigomycota and Chytridiomycota. (a) Cladogram showing the phylogenetic relationship of the compared taxa (Neocallimastigomycota genomes are in bold). (b) Heatmap plot of natural logarithm of the domain copy numbers showing the ones uniquely gained in Neocallimastigomycota (Pfam domains highlighted in this study are in blue). (c) Pfam domains highly enriched in Neocallimastigomycota. (d) Pfam domains absent in Neocallimastigomycota (presented domains are partial; see Table S1 for the full list).

tives (Fig. 2) identified 40 Pfam domains that are unique to the AGF, representing 0.67% of the total number of Pfams (5,980) in the AGF pangenome-transcriptome (Table S1 and Fig. 2b). The predicted functions of these domains include anaerobic ribonucleotide reductase (NRDD), metal transport and binding (FeoA and FeoB_C), carbohydrate binding (e.g., CBM_10, CBM-like, and Cthe_2159), atypical protein kinase (CotH), and glycoside hydrolase (e.g., Glyco_hydro_6 and Glyco_hydro_11) (Table S1 and Fig. 2b). In addition to these 40 unique AGF domains, many additional Pfams were also enriched in the AGF. Such domains mediate polysaccharide degradation and monosaccharide fermentations (Fig. 2c), including Chitin_binding_1, CBM_1, Cellulase, Glyco_hydro_10, Gly_radical, RicinB_lectin_2, Esterase, and Polysacc_deac_1 domains. Further, our analysis also identified 106 Pfam domains that are not present in AGF genomes and transcriptomes but found in sister Chytridiomycota. Most of these missing domains are related to oxidation reactions on cytochromes and mitochondria; instead, they possess specialized organelles called hydrogenosomes conducting metabolism under anaerobic conditions (6) (Table S1 and Fig. 2d). In addition, domains involved in the biosynthesis of nicotinic acid, uric acid, and photolyase, in purine catabolism, and in pathways of ureidoglycolate and kynurenine are also found to be absent in AGF species. Similar patterns were also identified in the comparison of homologous genes (Fig. S2).

A permissive criterion, allowing some missing copies, found a total of 2,728 gene families shared between AGF and chytrids. We discovered that 1,709 additional gene families are shared among AGF genomes (each gene presents in at least 21 out of the total 26 taxa) but absent in other chytrids, while another 367 families are missing in AGF members but present in the other chytrid lineages.

Genomic interactions within the rumen of mammalian herbivores. We focused on three Pfam domains (Cthe_2159, Gal_Lectin, and Rhamnogal_lyase) that are unique to the Neocallimastigomycota and previously not observed in fungal genomes. Phylogenetic analyses support a horizontal transfer of Cthe_2159 from rumen bacteria into

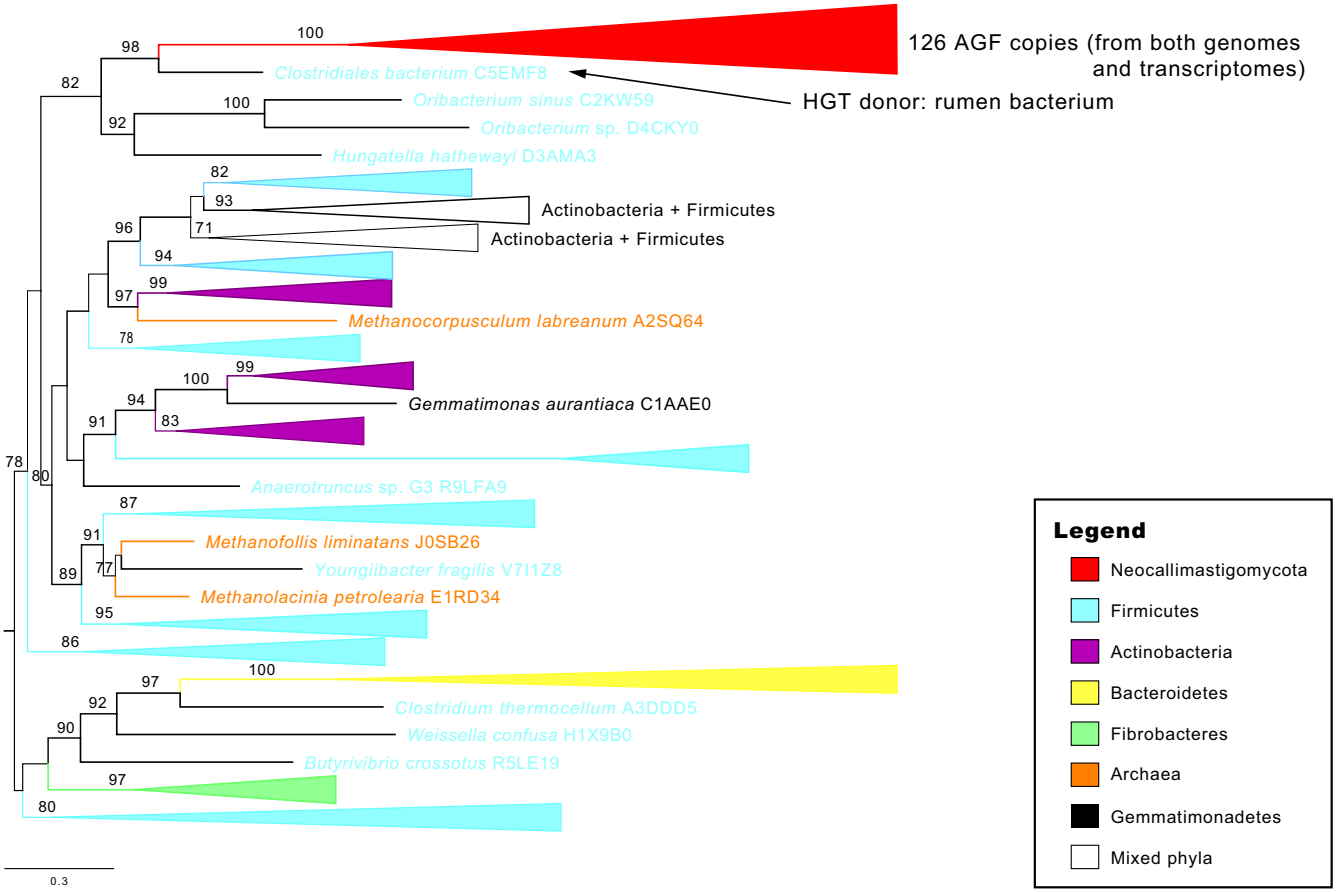

FIG 3 Midpoint-rooted phylogenetic tree of the Cthe_2159 domain. All 126 Neocallimastigomycota (AGF) copies (copies that have >90% identities have been removed) form a single clade (red) indicating the HGT donor, *Clostridiales bacterium* C5EMF8 (an obligate rumen bacterium), with strong support of maximum likelihood bootstrap (98/100). Included bacterial lineages were assigned different colors according to their phylogenetic classification (see legend for detailed information; the complete tree with all tip information is shown in Fig. S3).

AGF, followed by potential gene fusion to deliver eukaryotic specific functions. Similarly, analysis of Gal_Lectin domain copies in AGF suggests they were acquired from animal donor lineages. A similarity search of the AGF Rhamnogal_lyase domain finds most similar copies in plant genomes, and phylogenetic analysis indicates that the AGF polysaccharide lyase domain is distinct and not orthologous to related enzymes in other fungi.

A bacteria-like biomass-binding and putatively polysaccharide lyase domain, Cthe_2159. The Cthe_2159 domain was originally characterized as a polysaccharide lyase-like protein in the thermophilic and biomass-degrading bacterium *Clostridium thermocellum* (28). Proteins of the Cthe_2159 domain are beta-helix proteins with the ability to bind celluloses and acid sugars (polygalacturonic acid, a major component of the pectin), and homologs are primarily found in archaeal and bacterial genomes. Notably, a total 583 copies of the Cthe_2159 domain were identified in 5 genomes and 21 transcriptomes of AGF taxa, but this was reduced to a set of 126 clusters based on overall protein similarity (>90%) due to redundancy in transcriptome assemblies. This domain is absent in all other eukaryotic genomes examined in this study (Fig. 3 and Table 2). A phylogenetic tree of Cthe_2159 homologs identified from archaea, bacteria, and AGF suggests that the AGF Cthe_2159 domains were acquired from bacteria through HGT (Fig. 3). The likely donor was a Gram-positive firmicute (*Clostridiales*) (maximum likelihood bootstrap value, 98%), and the closest protein copies of Cthe_2159 domains are encoded in the *Oribacterium sinus*, *Oribacterium* sp., and *Hungatella hathewayi* genomes (Fig. 3). Members of the order *Clostridiales* are integral members of

TABLE 2 Distribution of the three studied domains in the fungal kingdom

Phylum	No. of examined genomes	No. of domains		
		Cthe_2159	Gal_Lectin	Rhamnogal_lyase
Ascomycota	652	0	0	0
Basidiomycota	324	0	0	0
Mucoromycota	76	0	0	0
Zoopagomycota	23	0	0	0
Chytridiomycota	14	0	0	0
Neocallimastigomycota	5	95	67	26
Blastocladiomycota	4	0	0	0
Cryptomycota	1	0	0	0
Microsporidia	22	0	0	0
Total	1,121			

the rumen microbiome. Four of these AGF Cthe_2159 domain-containing genes also encode eukaryotic Pfam protein domains (Atrophin-1, eIF-3_zeta, Nop14, and TPH) at the 3′ position of the Cthe_2159 domain. We hypothesize that these domains are the result of fusion after the acquisition of Cthe_2159 domain. The putative functions of these additional domains include initiation of the eukaryotic translation, maturation of 18S rRNA, production of 40S ribosome, and meiosis-specific activities (Fig. 4). Approximately 30% of these AGF Cthe_2159 gene models possess between 1 and 2 introns, but there is limited spliced transcript evidence to provide confidence in the gene structures, so the apparent intron gains could be artifacts of genome assembly or annotation (Text S1) (3, 4).

An animal-like galactose binding lectin domain, Gal_Lectin. Gal_Lectin domains were found in AGF genomes universally and absent in all other examined chytrid and fungal genomes (Table 2). Phylogenetic analysis recovered a monophyletic AGF Gal_Lectin clade which was not placed as a sister clade to the animals as expected for a fungal gene. Instead, it was embedded within the animal homologs in the tree and allies with one subgroup, polycystin-1 (PC-1) (Fig. 5a). The three separate animal subclades contain protein members that harbor the Gal_Lectin domain but with dissimilar functions based on sequence homology (Fig. 5). The genomes of ruminant hosts (e.g., horse and sheep) of the AGF also contain three gene families with the Gal_Lectin domain, which can be observed in each of the animal subclades (Fig. 5). The proteins in the animal subclade 1 were annotated as PC-1 based on similarity to the human polycystic kidney disease (*PKD1*) genes. The members of the animal subclade 2 were searched by BLAST against the NCBI nonredundant protein database and identified as homologs of the adhesion G protein-coupled receptor L1/3 (ADGRL1). The animal subclade 3 contains homologs of the EVA-1 protein, most of which contain two adjacent copies of the Gal_Lectin domain. The three subgroups of animal Gal_Lectin domains are also flanked by disparate Pfam domains (Fig. 5b). The gene phylogeny suggests an animal PC-1 protein as the likely donor lineage for the AGF Gal_Lectin gene (Fig. 5a), based on its closest sister relationship. In addition, the AGF proteins also contain a Pfam Glyco_transf_34 domain (Fig. 5b) which is absent in all animal homologs of the Gal_Lectin-containing genes, suggesting its involvement in fungus-specific activities in the rumen.

A novel fungal rhamnogalacturonate lyase domain, Rhamnogal_lyase, in AGF. In plants, the rhamnogalacturonate lyases are involved in the fruit ripening-related process, cell wall modification, and lateral root and root hair formation (29, 30). The Pfam database classifies two types of domains for rhamnogalactoside-degrading activity, Rhamnogal_lyase and RhgB_N. They are both N-terminal catalytic domains associated with the rhamnogalacturonan lyase protein (polysaccharide lyase family 4 [PL4]) and flanked persistently by the group of fn3_3 and carbohydrate-binding module (CBM)-like domains, with the particular function of degrading the rhamnogalacturonan I (RG-I) backbone of pectin. The Rhamnogal_lyase domain is found in the genomes of

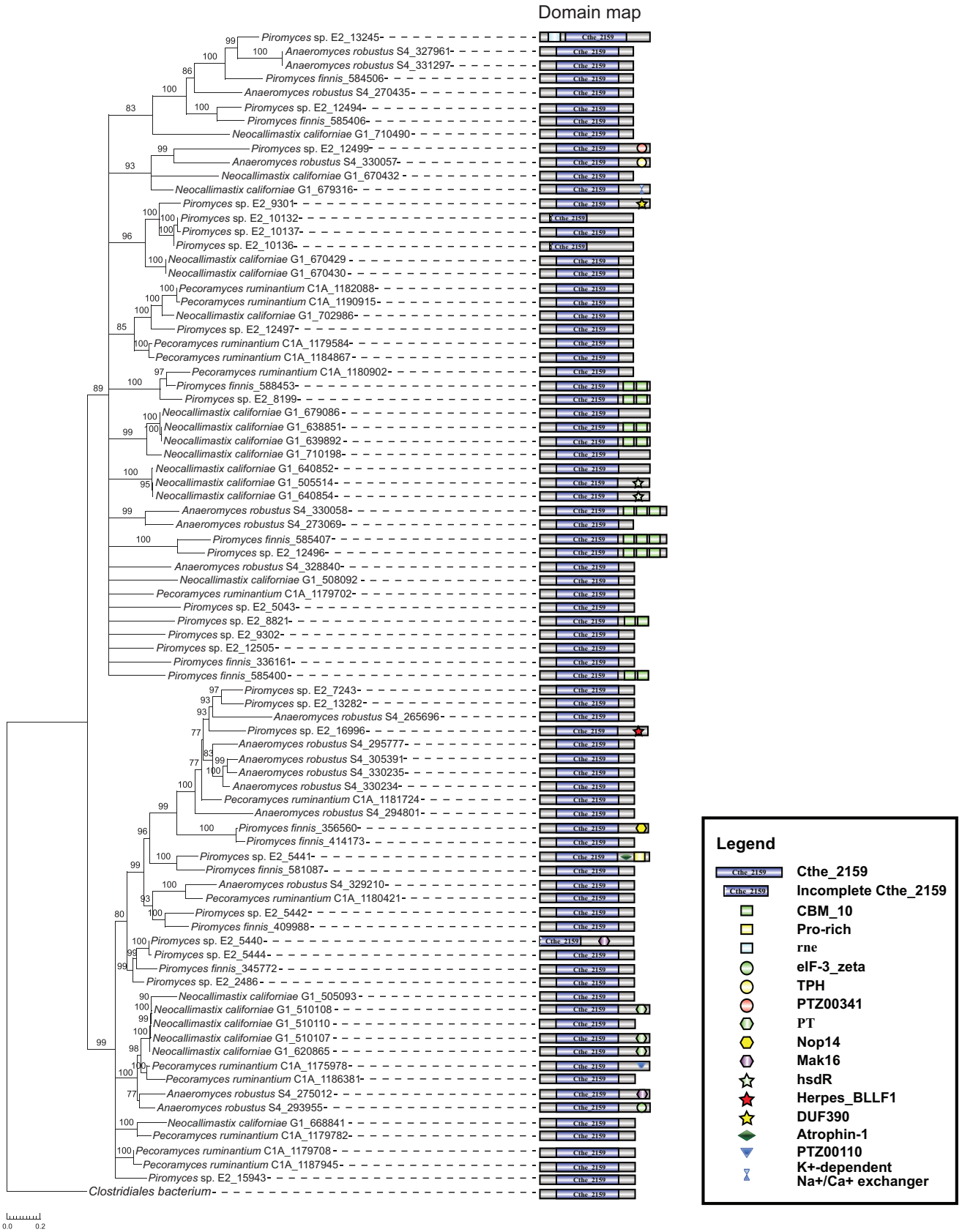

FIG 4 Phylogenetic tree of the 83 Cthe_2159 domains identified in five AGF genomes. Coding sequence identifiers used in the original annotation files are labeled with species names in tree tips. The tree is rooted with the closest related bacterial homolog found in *Clostridiales bacterium*. The domain map on the right shows the conserved domains produced by the Cthe_2159-containing genes.

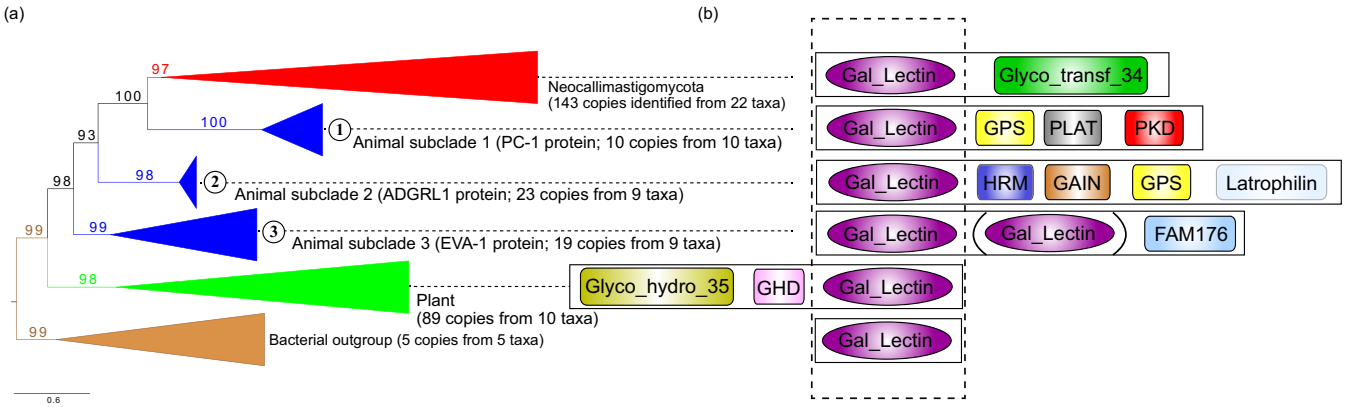

FIG 5 Phylogenetic tree of the animal-like Gal-Lectin domain identified in Neocallimastigomycota. (a) Collapsed phylogenetic tree based on protein sequences (rooted with the bacterial outgroup), including clades of Neocallimastigomycota (red), animals (blue; three clades are labeled 1 to 3), plants (green), and bacteria (brown) (a complete tree with all tip information is shown in Fig. S4). (b) Schematic diagrams showing the Gal_Lectin and other conserved domains on the same protein in each clade individually (dotted box highlights the aligned region used to produce the phylogenetic tree).

plants and plant-pathogenic bacteria (e.g., *Dickeya dadantii*, formerly *Erwinia chrysanthemi*), whereas the RhgB_N domain has a wider distribution and can be found in bacteria, fungi, and oomycetes (31). Sequence similarity searches using the Rhamnogal_lyase domain against various protein sequence databases (e.g., EnsEMBL, MycoCosm, and Pfam) returned no homolog in any other fungi (except the AGF members), which indicates that this domain is unique to AGF, plants, and bacteria. On the other hand, the RhgB_N domain is widely shared by Dikarya fungi, oomycetes, and bacteria. Although the RhgB_N and Rhamnogal_lyase domains are distantly related according to sequence similarity (24% between the copies of the *Aspergillus nidulans* and *Anaeromyces robustus*), they presumably share an origin due to the fact that they both physically located on the N-terminal region of the PL4 proteins and they have resembling functions to degrade the pectin RG-I region. The phylogenetic tree shows that although AGF Rhamnogal_lyase domains are more closely related to the plant homologs than to the clades of fungi and oomycetes, these AGF rhamnogalacturonate lyases likely have evolved a specific function in fungi (Fig. 6). The presence of the Rhamnogal_lyase domain in the rumen-associated fungi suggests that the AGF may support an ability to soften, modify, and degrade the plant pectin within the anaerobic rumen in a related but different way from plants.

DISCUSSION

Microbial diversity of ruminants is a research hot spot for development of bioenergy tools (2, 32, 33). The AGF are an important but understudied component of the ruminant microbiome, and their obligate anaerobic and relatively large (50 to 200 Mb) and AT rich (78 to 84%) genomes challenged the initial generation of genomic resources for the clade. In this study, we produced the most phylogenetically broad transcriptome sampling of the Neocallimastigomycota fungi to date to support phylogenomic and comparative analyses. Our results contribute new insights into the natural history and dynamic evolution of these cryptic ruminant gut fungi. The reconstructed phylogenomic species tree resolved previously unanswered questions about the evolutionary relationships of the members of the AGF. In addition, we provide the first estimation of the divergence time of AGF taxa, 66 (±10) Mya (Fig. 1), which is in remarkable concordance with the divergence of the forage Poaceae grasses (70 to 95 Mya) and dietary shifts in mammalian lineages (34 to 66 Mya) from insectivore to herbivore and omnivore. Grass evolution enabled the herbivory transition, and this diet adaptation drove an increase in the developmental and morphological complexity of the digestive tract, compartmentalization, and the development of dedicated anaerobic fermentation chambers (e.g., rumen and cecum) in the herbivorous alimentary tract to improve biomass degradation efficiency (34). This transition to plant-based (or

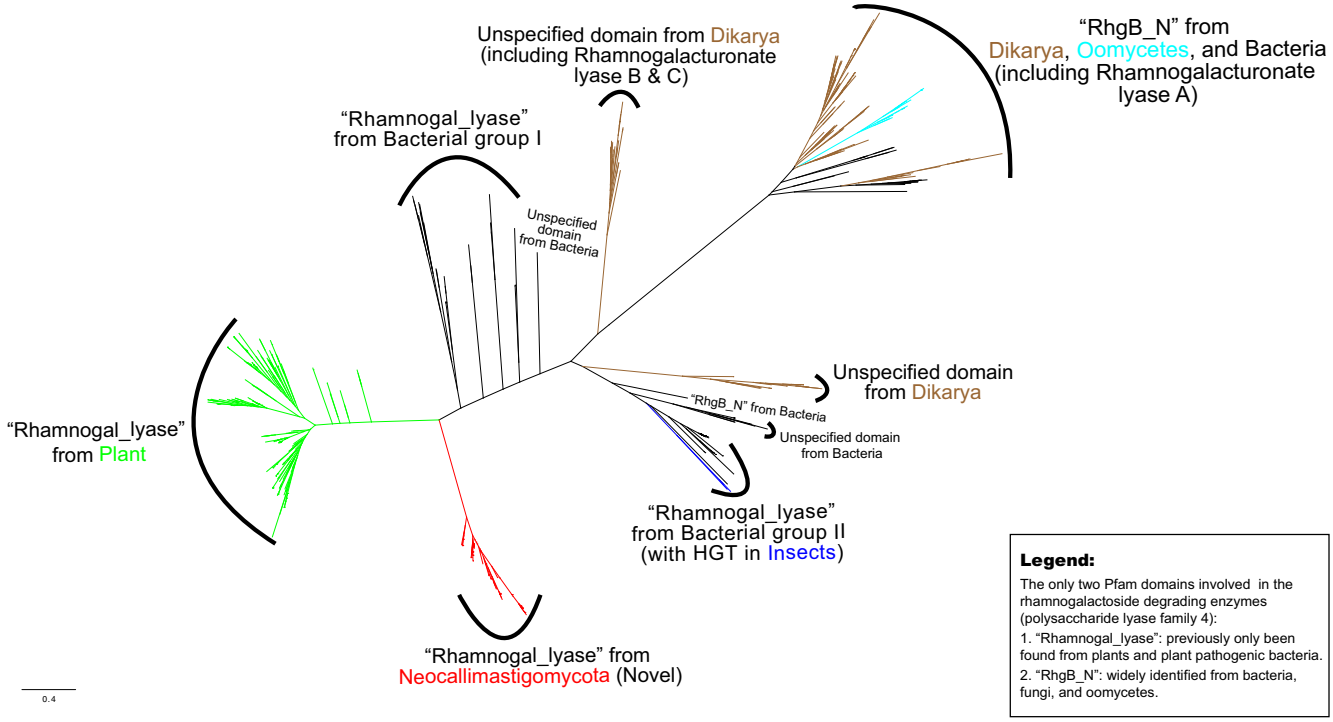

FIG 6 Radial phylogenetic tree of the Rhamnogal_lyase domain encoded by the Neocallimastigomycota (red). Plant copies are colored in green, and other homologous fungal genes are colored in brown. Oomycetes are in cyan, and animal copies only known in the mountain pine beetle *Dendroctonus ponderosae* are in blue. Bacterial branches are in black. The tree also included homologs of RhgB_N and Rhamnogalacturonan lyase A, B, and C. Domain names are suggested using NCBI's conserved domain search tool (cutoff $1E^{-5}$) with unaligned FASTA sequences (refer to Fig. S6 for a tree with detailed information).

plant-exclusive) diets required additional partnership with microbes since mammals lack cellulolytic and hemicellulolytic enzymes necessary to liberate sugars for absorption (5). In addition, the genome content comparisons help illustrate and predict new biological roles AGF play in the mammalian herbivore guts. The long branch that leads to the emergence of the Neocallimastigomycota clade indicates the distinctiveness of the extant group of obligate symbiotic fungi in the mammalian herbivores and implies the existence of undiscovered although possibly extinct relatives of the Neocallimastigomycota and Chytridiomycota (Fig. 1). Future environmental and metagenome sequence exploration of anaerobic environment testing for presence of these types of fungi may provide new observations that support their existence.

Our analyses identified multiple instances of Pfam domain gains ($n = 40$) and losses ($n = 106$) within the Neocallimastigomycota clade (Fig. 2 and Table S1). As the mRNA collected for most of the fungi was from isolates grown on a single substrate (cellobiose), these observations are limited to the genes expressed under this condition. More comprehensive sampling of growth conditions across developmental stages and substrates is important to fully categorize gene gain and loss. We have taken a conservative approach that considers genes with gain or loss patterns shared across multiple RNA sequencing (RNA-seq) data sets and additionally confirmed by gene content in the five available AGF genomes. This approach identified three AGF lineage-specific protein domains which are absent from all other examined fungal genomes (Table 2). Phylogenetic analyses support the hypothesis that they were acquired via HGT or other noncanonical events. Phylogenetic analyses of Cthe_2159 and Gal_Lectin domains indicate that they were separately transferred from the rumen bacteria and animal hosts horizontally (Fig. 3 and 5). Prior studies of multiple fungal lineages suggested that lineage-specific genes may have come from lateral acquisition (35–37). The absence of homologs in the entire fungal clades (except Neocallimastigomycota in this case) is a strong signal for their potential foreign origins. The absence of any fungal homologs of Cthe_2159 outside the AGF prevented us from testing the alternative hypotheses of a

fungal origin of the domain with an approximately unbiased (AU) topology constraint test (38). In analyzing Gal_Lectin, we compared the likelihoods of a constrained tree where all animal sequences were forced to be monophyletic to the topology of the unconstrained tree. The AU test did not find the constrained tree to be a statistically significantly worse fit, so the alternative hypothesis that the orphan Gal_Lectin in Neocallimastigomycota is a result of multiple independent loss events in all other fungal lineages could not be rejected by this analysis. Since the genetic distance between Neocallimastigomycota and sister chytrid relatives is quite large (Fig. 1), this multiple-loss scenario seems less likely than a single gain event. These identified domains highlight the diverse genomic tools utilized by AGF which we predict have improved their lignocellulolytic capacities (through Cthe_2159 and Rhamnogal_lyase) and recognition of molecules or other cells (through Gal_Lectin) within the rumen. The presence of four different eukaryotic Pfam domains fused with these bacterium-originated Cthe_2159 genes in AGF suggests that the genes are truly eukaryotic and present in the fungal genomes (Fig. 4) and not a contamination artifact. Studies of intron gains and losses in fungal lineages have suggested the ancestor was intron rich, an observation that is supported by intron-rich chytrid genomes (39–41). Although introns are present in several Cthe_2159 gene models in the Neocallimastigomycota genomes and flanked with duplicated coding sequences (*Piromyces* sp. strain E2), there is little support of spliced mRNA transcripts originating from these loci, preventing us from confidently declaring these as recent intron insertions (Text S1 and Fig. S7).

The Cthe_2159 protein family binds cellulosic and pectic substrates in the anaerobic and thermophilic bacterium *Clostridium thermocellum* (28). The crystal structure of the Cthe_2159 domain suggests that it is a polysaccharide lyase family with similarity to pectate lyases in the PL9 family. The Rhamnogal_lyase domains primarily function in the facilitation of cell wall modification in plants (29). The domain in phytopathogenic bacteria functions to disorganize plant tissues and support invasion (42). Although we cannot identify an unambiguous donor lineage of the AGF Rhamnogal_lyase domains (Fig. 6), their gain is a synapomorphy of the extant AGF taxa and may contribute to the ability of these fungi to access polysaccharides in plant cell walls. Both Cthe_2159 and Rhamnogal_lyase (PL4 family) domains have putative function in pectin binding or degradation activity, which we interpret as an indication of the importance of deconstruction of pectin in the lifestyle of AGF in the rumen (Table 2 and Fig. 2). Pectin is abundant in primary cell walls and the middle lamella in both dicotyledonous plants (making up 20 to 35% dry weight) and grasses (2 to 10%), serving as protection for plant cells from degrading enzymes produced by animals (43–46). The removal of pectin can increase the exposed surface area of a plant cell wall and improve the accessibility of degradation enzymes to other polysaccharides (cellulose and hemicellulose) masked by pectin (47). The Cthe_2159 and Rhamnogal_lyase proteins may contribute to the high efficiency of the AGF biomass degradability by uncoupling the pectin that glues cells together, increasing the exposed surface areas, and thus allowing diverse polysaccharide enzymes to work on plant cells simultaneously in the rumen. The fungi may benefit from these acquired domains in their capacity as primary degraders of ingested forage (48). Further investigation of their role in the multiple processes that AGF perform to weaken forage fibers and release polysaccharides is warranted (49, 50).

The Gal_Lectin domain bears the phylogenetic hallmark of being acquired from an animal donor. Animals use galactose-binding lectins to recognize foreign entities (51) and participate in antimicrobial defenses (52, 53). Our results suggest that the Gal_Lectin domains in AGF are homologous and closely related to animal PC-1 proteins (Fig. 5a), which are transmembrane proteins functioning in cell recognition (54, 55). *In vitro*, PC-1 shows binding ability to carbohydrate matrices and collagen types I, II, and IV (56). We postulate that the acquisition of the animal-like Gal_Lectin domain contributes to the AGF abilities of cell-cell recognition and interaction with other microbes in the rumen. The syntenic relationship of the coding genes shows that the AGF Gal_Lectin domains are flanked by the Glyco_transf_34 domain, which lacks homologs in any

other animals (Fig. 5b and S5). The AGF-equipped Glyco_transf_34 belongs to the galactosyltransferase GMA12/MNN10 family and may help catalyze the transfer of the sugar moieties in cooperating with the adjacent Gal_Lectin domain. Our investigation found that HGT has contributed to the AGF genome evolution, with donors from both prokaryotes and eukaryotes. HGT may have helped these fungi to acquire new functions and to thrive in the anaerobic gut as a key member of the microbial community degrading plant materials in animal hosts.

Other than the arsenal of diverse enzyme profiles, the AGF have also been known to use rhizoids and holdfasts to physically aid the fungal body to penetrate into the plant material deeply, which is superior to other rumen microorganisms in terms of efficiency (5, 57). Our study provides evidence that the rumen fungi are able to and have actively acquired functional domains from the animal hosts and coexisting anaerobic bacteria in the rumen. These exotic genetic elements encoded in Neocallimastigomycota genomes may contribute to the unique traits of these fungi which are distinct from their free-living relatives. The long branch leading to the recent radiation of Neocallimastigomycota (Fig. 1) also suggests an evolutionary trajectory distinct from those of the sister Chytridiomycota lineages. Living as gut dwellers in the strict anaerobic gut environment for over 66 million years, AGF have undergone reductive evolution on the mitochondria and eventually transformed it to a new organelle, the hydrosome (3, 32). Their ecological roles of AGF in such an extreme environment also endow their exceptional ability for plant degradation. The AGF use both physical (deconstruction of lignocelluloses) and biological (depolymerization) mechanisms before the fermentation of plant polysaccharides. These steps require diverse enzymes capable of breaking chemical bonds in carbohydrates, including cellulases, hemicellulases, ligninases, and pectinases (58). In turn, the acquisition of these enzymatic processes has driven the synapomorphic and autapomorphic characteristics described in the AGF. Currently, few close relatives have been found, and none have been cultured which subtend from the long branch. Environmental DNA investigations of extreme and anaerobic environment that may be a suitable niche of those Neocallimastigomycota-like microbes may reveal potential relatives (59). For example, a recent metagenomic survey from coastal marine sediments suggests that some operational taxonomic units (OTUs) could be assigned to Neocallimastigomycota using a 28S rRNA marker (60). Sampling of deep-sea habitats and marine mammalian herbivores could provide future discoveries of biodiversity and evolutionary importance for understanding the evolutionary trajectory of the Neocallimastigomycota.

MATERIALS AND METHODS

RNA extraction, sequencing, and data set preparation. In total, 21 strains of Neocallimastigomycota fungi were cultured from cow, sheep, horse, and goat feces and rumen fluid of fistulated cows in the Stillwater, OK, area (8) (Table 1). These strains were maintained under anaerobic conditions using the modified Hungate method, as described previously (61–64). Culture purity was ensured by serial dilution and incubation at 39°C for 24 to 48 h, followed by inoculation and a second round of isolation. Cellobiose was the sole carbon source of the fungal culture prepared for RNA extraction. The total volume of RNA was harvested at early stationary phase (48 to 60 h postinoculation) using the MasterPure yeast RNA purification kit (Epicentre, Madison, WI, USA) and processed for transcriptomics sequencing using the Illumina HiSeq 2500 platform and 2 × 150-bp paired-end library by Novogene (Beijing, China).

The RNA-seq data were assembled into *de novo* transcript assemblies using Trinity (v2.6.6) and used to predict ORFs using TransDecoder (v5.0.2) (65, 66). The generated proteomes and corresponding coding sequences were used as input to phylogenomic and comparative genomic analyses.

The five published Neocallimastigomycota genome sequences were obtained from the Joint Genome Institute (JGI) MycoCosm database (67, 68). These are the sequences for *Anaeromyces robustus* S4, *Neocallimastix californiae* G1, *Pecoramyces ruminantium* C1A (synonym *Orpinomyces* sp.), *Piromyces finnis* (v3.0), and *Piromyces* sp. E2 (3, 4). Five outgroup Chytridiomycota taxa with sequenced genomes were chosen. These are *Chytriomyces* sp. strain MP 71, *Entophlyctis helioformis* JEL805, *Gaertneriomyces semiglobifer* Barr 43, *Gonapodya prolifera* JEL478, and *Rhizoclosmatium globosum* JEL800 (69, 70).

Phylogenomics and divergence time estimation. A set of 434 highly conserved and generally single-copy protein-coding genes in fungi (https://doi.org/10.5281/zenodo.1251476) were developed through efforts of the 1000 Fungal Genomes Project and identified as single-copy genes in orthologous clusters provided in the Joint Genome Institute MycoCosm site (67, 71, 72). These markers were used for

phylogenomic analyses in the PHYling pipeline (https://doi.org/10.5281/zenodo.1257001). Profile-Hidden-Markov models of these markers were searched in the chytrid predicted protein sequences using HMMER3 (v3.1b2). A total of 426 (out of 434) conserved orthologous markers were identified with hmmsearch (cutoff = $1E^{-10}$) in the 26 Neocallimastigomycota and 5 Chytridiomycota. The identified protein sequence homologs in each species, for each phylogenetic marker, were aligned with hmmalign to the marker profile-HMM. The protein alignments were also back translated into codon alignments guided by the protein alignment using the tool bp_mrtrans.pl (73). The protein and coding sequences of the markers were concatenated into a superalignment with 426 partitions defined by each gene marker. The 426 gene partitions were further collapsed into 33 partitions by PartitionFinder v2.1.1 with a greedy search to find partitions with consistent phylogenetic signals (74). Phylogenetic trees were constructed from this superalignment and partition scheme with two methods, the maximum likelihood method, implemented in IQ-TREE (v1.5.5), and Bayesian inference method, implemented in BEAST (v1.8.4) (75, 76). Configuration files for divergence time estimation analysis were coded in BEAUti v1.8.4 using the 33 partitions and two calibration priors, (i) a direct fossil record of Chytridiomycota from the Rhynie Chert (407 Mya) (77, 78), and (ii) the emergence time of Chytridiomycota (573 to 770 Mya as 95% HPD) from earlier studies (69, 79, 80). The Birth-Death incomplete sampling tree model was employed for interspecies relationship analyses (81). Unlinked strict clock models were used for each partition. The archive of input files and analysis scripts used to perform the phylogenetic analyses are available at Zenodo (https://doi.org/10.5281/zenodo.1447225). Three independent runs were performed separately for 50 million generations each with random starting seeds. Sufficient effective sample size (ESS) (>200) values were obtained after the default burn-in (10%) for the final sampled trees. The maximum clade credibility (MCC) tree was compiled using TreeAnnotator v1.8.4.

Identification of AGF-specific genes and Pfam domains. Orthologous genes across the 31 genomes or transcriptomes were identified using a comparative genomic pipeline that utilized all-versus-all BLASTp (cutoff = $1E^{-5}$) to obtain the similarity pairs, orthAgogue to identify putative orthologous relationships, and the Markov clustering algorithm (MCL, using the inflation value of 1.5) to generate disjoint clusters and deployed in an analysis pipeline (https://doi.org/10.5281/zenodo.1447225) (82–84). Comparisons of the shared gene content of the orthologous clusters were performed among the Chytridiomycota lineages using a permissive strategy of counting a gene family as shared if it is missing in up to 5 of the 26 Neocallimastigomycota taxa and 1 of the 5 chytrid genomes. In this scenario, genes absent in all chytrid genomes and maintained by more than 21 out of the 26 Neocallimastigomycota genomes/transcriptomes are defined as AGF unique genes; on the other hand, genes missing from all Neocallimastigomycota and present in at least 4 out of the 5 chytrid genomes are treated as AGF lost genes.

Protein domains were identified by searching the predicted proteomes from each genome assembly or transcriptome assembly against the Protein Family (Pfam) database (v31.0). The enrichment heatmap of the Pfam domains across the included taxa was produced using the aheatmap function in the R package NMF based on the total copy number (based on hmmscan results of searches against the Pfam database using a cutoff E-value of 1e−2) count in each assembly (85). Genes only present in the AGF genomes and missing from all of the included free-living chytrids relatives were identified.

To identify genes in AGF that are likely important for interactions with mammalian hosts and plant material breakdown, we further compared the five available AGF genomes to the genomes of their animal hosts (e.g., sheep, horse, elephant, and yak) (https://www.broadinstitute.org/elephant/elephant-genome-project and references 86–88), the diet plant (e.g., moss, rice, palm, maize, and sorghum) (89–98) (Table S2), and the 1,165 available fungal genomes from the ongoing 1000 Fungal Genomes Project (http://1000.fungalgenomes.org; https://mycocosm.jgi.doe.gov) (18, 19, 67, 68). Comparison of AGF genes to host or plant genomes was intended to test if any copies were likely donated from these lineages by searching for high-identity nucleotide matches. To prioritize AGF genes that may have been laterally acquired from these hosts, a Python script (13) and similarity search tool BLAT (99) were applied to filter out DNA elements in AGF with higher similarity to animal or plant homologs than any fungal ones, excluding the AGF themselves. Candidate genes for lateral transfer were ranked by the combination of the two strategies. The candidate genes with an assigned functional or biological process annotation were analyzed with priority using phylogenetic reconstruction to infer their potential origin.

Identification of homologous sequences and potential origin of HGT candidate loci. Three Pfam domains, Cthe_2159, Gal_Lectin, and Rhamnogal_lyase, were identified to be unique to the AGF genomes compared to the Chytridiomycota fungi or all other fungal members. To predict the donor lineages for these putative HGT events, we searched more broadly for homologues in genome databases of plant, metazoa, fungi, bacteria, and protists in EnsEMBL (v37) (100) via the Web-implemented HMMER tool (https://www.ebi.ac.uk/Tools/hmmer/) (cutoff = $1E^{-3}$). Additional fungal homologues were found by searching the Department of Energy (DOE) JGI's MycoCosm database (67, 68). The profile Hidden Markov Model tool phmmer in the HMMer package (101) was used to search for similar sequences in the 1,165 fungal genomes using the query of edge-trimmed domain sequences from *A. robustus* (cutoff = $1E^{-3}$).

Members of the RhgB_N sequences were obtained from the Pfam database classified in the RhgB_N (PF09284) family (31), along with the N-terminal sequences of the rhamnogalacturonate lyase families A, B, and C from GenBank (102–104). A single data set of RhgB_N and Rhamnogal_lyase family members from animals, fungi, plants, and bacteria was constructed from these searches. Domain names were confirmed using NCBI's conserved domain search tool (cutoff = $1E^{-5}$) with unaligned FASTA sequences (105). Similarly, homologs of the Gal_Lectin and Cthe_2159 domains were obtained by searching for similar sequences in the previously described genome databases and

the categorized Pfam database (families of Gal_Lectin [PF02140] and Cthe_2159 [PF14262]). Homologous sequences containing the Cthe_2159 domain were only identified in archaea and bacteria, while the AGF copies are the first eukaryotic representatives identified with this domain. Homologs of the flanking domain Glyco_transf_34 were obtained similarly from EnsEMBL genome databases described above using the edge-trimmed domain sequence from *A. robustus* (cutoff = $1E^{-5}$). Highly similar sequences (>90%) were filtered using CD-HIT v4.6.4, followed by multiple-sequence alignment with MUSCLE v3.8.31 (106, 107).

Phylogenetic analyses of the HGT candidates. In total, 747 sequences of the rhamnogalacturanate degradation proteins (including both Rhamnogal_lyase and RhgB_N) were included in the alignment. For the other two domains, Gal_Lectin and Cthe_2159, the alignments include 297 and 234 unique variants, respectively. The Cthe_2159 domain-containing genes in the 5 AGF genomes were aligned separately using MUSCLE v3.8.31 in the Mesquite software (107, 108). Both the upstream and downstream flanking regions of the studied Pfam domain sequences were trimmed using the Mesquite software (108). Selection of the appropriate substitutional model, the maximum likelihood phylogenetic tree reconstruction, and the ultrafast bootstrapping (1,000 replicates) were conducted using the IQ-TREE v1.5.5 package (75, 109, 110).

Data availability. Assembled transcriptomes, raw Illumina read sequences, and isolate metadata are deposited in GenBank with the BioProject number PRJNA489922. All accession numbers are listed in Table 1. All generated RNA-seq reads were deposited in the Sequence Read Archive, and assembled transcriptomes were deposited in the Transcriptome Shotgun Assembly archive.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00247-19.

TEXT S1, DOCX file, 0.1 MB.
FIG S1, PDF file, 0.2 MB.
FIG S2, PDF file, 2.5 MB.
FIG S3, PDF file, 0.1 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.1 MB.
FIG S6, PDF file, 1.1 MB.
FIG S7, PDF file, 1.6 MB.
TABLE S1, DOCX file, 0.1 MB.
TABLE S2, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (grants DEB-1557110 to J.E.S. and DEB-1557102 to N.Y. and M.E.). Y.W. acknowledges the Mycological Society of America for the Translational Mycology Postdoctoral Award. Data analyses were performed at the University of California Riverside High-Performance Computational Cluster supported by NSF grant DBI-1429826 and NIH grant S10-OD016290. J.E.S. is supported by the U.S. Department of Agriculture, National Institute of Food and Agriculture Hatch project CA-R-PPA-5062-H, and is a CIFAR Fellow in the Fungal Kingdom: Threats & Opportunities Program.

REFERENCES

1. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat Commun 9:870. https://doi.org/10.1038/s41467-018-03317-6.
2. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Eloe-Fadrosh EA, Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-Espino D, Perry R, Henderson G, Creevey CJ, Terrapon N, Lapebie P, Drula E, Lombard V, Rubin E, Kyrpides NC, Henrissat B, Woyke T, Ivanova NN, Kelly WJ. 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat Biotechnol 36:359–367. https://doi.org/10.1038/nbt.4110.
3. Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ, Atiyeh HK, Wilkins MR, Elshahed MS. 2013. The genome of the anaerobic fungus orpinomyces sp. strain c1a reveals the unique evolutionary history of a remarkable plant biomass degrader. Appl Environ Microbiol 79:4620–4634. https://doi.org/10.1128/AEM.00821-13.
4. Haitjema CH, Gilmore SP, Henske JK, Solomon KV, de Groot R, Kuo A, Mondo SJ, Salamov AA, LaButti K, Zhao Z, Chiniquy J, Barry K, Brewer HM, Purvine SO, Wright AT, Hainaut M, Boxma B, van Alen T, Hackstein JHP, Henrissat B, Baker SE, Grigoriev IV, O'Malley MA. 2017. A parts list for fungal cellulosomes revealed by comparative genomics. Nat Microbiol 2:17087. https://doi.org/10.1038/nmicrobiol.2017.87.
5. Gruninger RJ, Puniya AK, Callaghan TM, Edwards JE, Youssef N, Dagar SS, Fliegerova K, Griffith GW, Forster R, Tsang A, Mcallister T, Elshahed MS. 2014. Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. FEMS Microbiol Ecol 90:1–17. https://doi.org/10.1111/1574-6941.12383.
6. Yarlett N, Orpin CG, Munn EA, Yarlett NC, Greenwood CA. 1986. Hydrogenosomes in the rumen fungus *Neocallimastix patriciarum*. Biochem J 236:729–739. https://doi.org/10.1042/bj2360729.
7. Solomon KV, Haitjema CH, Henske JK, Gilmore SP, Borges-Rivera D, Lipzen A, Brewer HM, Purvine SO, Wright AT, Theodorou MK, Grigoriev IV, Regev A, Thompson DA, O'Malley MA. 2016. Early-branching gut fungi possess a

large, comprehensive array of biomass-degrading enzymes. Science 351:1192–1195. https://doi.org/10.1126/science.aad1431.
8. Murphy CL, Youssef NH, Hanafy RA, Couger MB, Stajich JE, Wang Y, Baker K, Dagar SS, Griffith GW, Farag IF, Callaghan TM, Elshahed MS. 2019. Horizontal gene transfer as an indispensable driver for evolution of Neocallimastigomycota into a distinct gut-dwelling fungal lineage. Appl Environ Microbiol 85:e00988-19. https://doi.org/10.1128/AEM.00988-19.
9. Dhillon B, Feau N, Aerts AL, Beauseigle S, Bernier L, Copeland A, Foster A, Gill N, Henrissat B, Herath P, LaButti KM, Levasseur A, Lindquist EA, Majoor E, Ohm RA, Pangilinan JL, Pribowo A, Saddler JN, Sakalidis ML, de Vries RP, Grigoriev IV, Goodwin SB, Tanguay P, Hamelin RC. 2015. Horizontal gene transfer and gene dosage drives adaptation to wood colonization in a tree pathogen. Proc Natl Acad Sci U S A 112:3451–3456. https://doi.org/10.1073/pnas.1424293112.
10. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, Kazan K, Chakraborty S, McDonald BA, Manners JM. 2012. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. PLoS Pathog 8:e1002952. https://doi.org/10.1371/journal.ppat.1002952.
11. Pombert J-F, Selman M, Burki F, Bardell FT, Farinelli L, Solter LF, Whitman DW, Weiss LM, Corradi N, Keeling PJ. 2012. Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. Proc Natl Acad Sci U S A 109:12638–12643. https://doi.org/10.1073/pnas.1205020109.
12. Chaib De Mares M, Hess J, Floudas D, Lipzen A, Choi C, Kennedy M, Grigoriev IV, Pringle A. 2015. Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal *Amanita*. New Phytol 205:1552–1564. https://doi.org/10.1111/nph.13140.
13. Wang Y, White MM, Kvist S, Moncalvo J-M. 2016. Genome-wide survey of gut fungi (Harpellales) reveals the first horizontally transferred ubiquitin gene from a mosquito host. Mol Biol Evol 33:2544–2554. https://doi.org/10.1093/molbev/msw126.
14. Sun G, Yang Z, Kosch T, Summers K, Huang J. 2011. Evidence for acquisition of virulence effectors in pathogenic chytrids. BMC Evol Biol 11:195. https://doi.org/10.1186/1471-2148-11-195.
15. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. 2009. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell 21:1897–1911. https://doi.org/10.1105/tpc.109.065805.
16. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung G-H, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schüßler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts RA, et al. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. Nature 443:818–822. https://doi.org/10.1038/nature05110.
17. James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D, Powell MJ, Griffith GW, Vilgalys R. 2006. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). Mycologia 98:860–871. https://doi.org/10.1080/15572536.2006.11832616.
18. Spatafora JW, Aime MC, Grigoriev IV, Martin F, Stajich JE, Blackwell M. 2017. The fungal tree of life: from molecular systematics to genome-scale phylogenies. Microbiol Spectr 5:FUNK-0053-2016. https://doi.org/10.1128/microbiolspec.FUNK-0053-2016.
19. Stajich JE. 2017. Fungal genomes and insights into the evolution of the kingdom. Microbiol Spectr 5: FUNK-0055-2016. https://doi.org/10.1128/microbiolspec.FUNK-0055-2016.
20. Wang X, Liu X, Groenewald JZ. 2017. Phylogeny of anaerobic fungi (phylum Neocallimastigomycota), with contributions from yak in China. Antonie Van Leeuwenhoek 110:87–103. https://doi.org/10.1007/s10482-016-0779-1.
21. Liggenstoffer AS, Youssef NH, Couger MB, Elshahed MS. 2010. Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. ISME J 4:1225–1235. https://doi.org/10.1038/ismej.2010.49.
22. Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA. 2008. The age of the grasses and clusters of origins of C_4 photosynthesis. Glob Chang Biol 14:2963–2977. https://doi.org/10.1111/j.1365-2486.2008.01688.x.
23. Prasad V, Strömberg CAE, Alimohammadian H, Sahni A. 2005. Dinosaur coprolites and the early evolution of grasses and grazers. Science 310:1177–1180. https://doi.org/10.1126/science.1118806.
24. Gaut BS. 2002. Evolutionary dynamics of grass genomes. New Phytol 154:15–28. https://doi.org/10.1046/j.1469-8137.2002.00352.x.
25. Bremer K. 2002. Gondwanan evolution of the grass alliance of families (Poales). Evolution 56:1374–1387. https://doi.org/10.1111/j.0014-3820.2002.tb01451.x.
26. Christin PA, Spriggs E, Osborne CP, Strömberg CAE, Salamin N, Edwards EJ. 2014. Molecular dating, evolutionary rates, and the age of the grasses. Syst Biol 63:153–165. https://doi.org/10.1093/sysbio/syt072.
27. Emerling CA, Delsuc F, Nachman MW. 2018. Chitinase genes (CHIAs) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. Sci Adv 4:eaar6478. https://doi.org/10.1126/sciadv.aar6478.
28. Close DW, D'Angelo S, Bradbury ARM. 2014. A new family of β-helix proteins with similarities to the polysaccharide lyases. Acta Crystallogr D Biol Crystallogr 70:2583–2592. https://doi.org/10.1107/S1399004714015934.
29. Molina-Hidalgo FJ, Franco AR, Villatoro C, Medina-Puche L, Mercado JA, Hidalgo MA, Monfort A, Caballero JL, Muñoz-Blanco J, Blanco-Portales R. 2013. The strawberry (*Fragaria*×*ananassa*) fruit-specific *rhamnogalacturonate lyase 1* (*FaRGLyase1*) gene encodes an enzyme involved in the degradation of cell-wall middle lamellae. J Exp Bot 64:1471–1483. https://doi.org/10.1093/jxb/ers386.
30. Ponniah SK, Thimmapuram J, Bhide K, Kalavacharla VK, Manoharan M. 2017. Comparative analysis of the root transcriptomes of cultivated sweetpotato (*Ipomoea batatas* [L.] Lam) and its wild ancestor (*Ipomoea trifida* [Kunth] G. Don). BMC Plant Biol 17:1–14. https://doi.org/10.1186/s12870-016-0950-x.
31. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344.
32. Marvin-Sikkema FD, Driessen AJM, Gottschal JC, Prins RA. 1994. Metabolic energy generation in hydrogenosomes of the anaerobic fungus *Neocallimastix*: evidence for a functional relationship with mitochondria. Mycol Res 98:205–212. https://doi.org/10.1016/S0953-7562(09)80187-1.
33. Bryant MP. 1959. Bacterial specis of the rumen. Bacteriol Rev 23:125–153.
34. Hackmann TJ, Spain JN. 2010. Ruminant ecology and evolution: perspectives useful to ruminant livestock research and production. J Dairy Sci 93:1320–1334. https://doi.org/10.3168/jds.2009-2071.
35. Rolland T, Neuvéglise C, Sacerdot C, Dujon B. 2009. Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. PLoS One 4:e6515. https://doi.org/10.1371/journal.pone.0006515.
36. Wisecaver JH, Slot JC, Rokas A. 2014. The evolution of fungal metabolic pathways. PLoS Genet 10:e1004816. https://doi.org/10.1371/journal.pgen.1004816.
37. Kominek J, Doering DT, Opulente DA, Shen X-X, Zhou X, DeVirgilio J, Hulfachor AB, Groenewald M, Mcgee MA, Karlen SD, Kurtzman CP, Rokas A, Hittinger CT. 2019. Eukaryotic acquisition of a bacterial operon. Cell 176:1356–1311. https://doi.org/10.1016/j.cell.2019.01.034.
38. Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492–508. https://doi.org/10.1080/10635150290069913.
39. Stajich JE, Dietrich FS, Roy SW. 2007. Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. Genome Biol 8:R223. https://doi.org/10.1186/gb-2007-8-10-r223.
40. Csuros M, Rogozin IB, Koonin EV. 2011. A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. PLoS Comput Biol 7:e1002150–9. https://doi.org/10.1371/journal.pcbi.1002150.
41. Nielsen CB, Friedman B, Birren B, Burge CB, Galagan JE. 2004. Patterns of intron gain and loss in fungi. PLoS Biol 2:e422. https://doi.org/10.1371/journal.pbio.0020422.
42. Laatu M, Condemine G. 2003. Rhamnogalacturonate lyase rhiE is secreted by the out system in *Erwinia chrysanthemi*. J Bacteriol 185:1642–1649. https://doi.org/10.1128/jb.185.5.1642-1649.2003.

43. Vogel J. 2008. Unique aspects of the grass cell wall. Curr Opin Plant Biol 11:301–307. https://doi.org/10.1016/j.pbi.2008.03.002.
44. Voragen AGJ, Coenen GJ, Verhoef RP, Schols HA. 2009. Pectin, a versatile polysaccharide present in plant cell walls. Struct Chem 20:263–275. https://doi.org/10.1007/s11224-009-9442-z.
45. Xiao C, Anderson CT. 2013. Roles of pectin in biomass yield and processing for biofuels. Front Plant Sci 4:67. https://doi.org/10.3389/fpls.2013.00067.
46. Salem H, Bauer E, Kirsch R, Vogel H, Fukatsu T, Kaltenpoth M, Salem H, Bauer E, Kirsch R, Berasategui A, Cripps M, Weiss B, Koga R, Fukumori K, Vogel H, Fukatsu T, Kaltenpoth M. 2017. Drastic genome reduction in an herbivore's pectinolytic symbiont. Cell 171:1520–1525. https://doi.org/10.1016/j.cell.2017.10.029.
47. Pakarinen A, Zhang J, Brock T, Maijala P, Viikari L. 2012. Enzymatic accessibility of fiber hemp is enhanced by enzymatic or chemical removal of pectin. Bioresour Technol 107:275–281. https://doi.org/10.1016/j.biortech.2011.12.101.
48. Haitjema CH, Solomon KV, Henske JK, Theodorou MK, O'Malley MA. 2014. Anaerobic gut fungi: advances in isolation, culture, and cellulolytic enzyme discovery for biofuel production. Biotechnol Bioeng 111:1471–1482. https://doi.org/10.1002/bit.25264.
49. Borneman WS, Akin DE, Ljungdahl LG. 1989. Fermentation products and plant cell wall-degrading enzymes produced by monocentric and polycentric anaerobic ruminal fungi. Appl Environ Microbiol 55:1066–1073.
50. Nagpal R, Puniya AK, Griffith GW, Goel G, Puniya M, Sehgal JP, Singh K. 2009. Anaerobic rumen fungi: potential and applications, p 375–393. *In* Khachatourians GG, Arora DK, Rajendran TP, Srivastava AK (ed), Agriculturally important microorganisms, 1st ed. Springer, Singapore.
51. García-Maldonado E, Cano-Sánchez P, Hernández-Santoyo A. 2017. Molecular and functional characterization of a glycosylated galactose-binding lectin from *Mytilus californianus*. Fish Shellfish Immunol 66:564–574. https://doi.org/10.1016/j.fsi.2017.05.057.
52. Low DHP, Frecer V, Le Saux A, Srinivasan GA, Ho B, Chen J, Ding JL. 2010. Molecular interfaces of the galactose-binding protein tectonin domains in host-pathogen interaction. J Biol Chem 285:9898–9907. https://doi.org/10.1074/jbc.M109.059774.
53. Uhlenbruck G, Steinhausen G. 1977. Tridacnins: symbiosis-profit or defense-purpose? Dev Comp Immunol 1:183–192. https://doi.org/10.1016/S0145-305X(77)80028-1.
54. Hughes J, Ward C, Peral B, Aspinwall R, Clark K, San Millán J, Gamble V, Harris P. 1995. The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. Nat Genet 10:151–160. https://doi.org/10.1038/ng0695-151.
55. Weston BS, Malhas AN, Price RG. 2003. Structure-function relationships of the extracellular domain of the autosomal dominant polycystic kidney disease-associated protein, polycystin-1. FEBS Lett 538:8–13. https://doi.org/10.1016/s0014-5793(03)00130-3.
56. Weston BS, Bagnéris C, Price RG, Stirling JL. 2001. The polycystin-1 C-type lectin domain binds carbohydrate in a calcium-dependent manner, and interacts with extracellular matrix proteins in vitro. Biochim Biophys Acta Mol Basis Dis 1536:161–176. https://doi.org/10.1016/S0925-4439(01)00046-1.
57. Berlemont R. 2017. Distribution and diversity of enzymes for polysaccharide degradation in fungi. Sci Rep 7:1–11. https://doi.org/10.1038/s41598-017-00258-w.
58. Brandt A, Gräsvik J, Hallett JP, Welton T. 2013. Deconstruction of lignocellulosic biomass with ionic liquids. Green Chem 15:550–583. https://doi.org/10.1039/c2gc36364j.
59. James TY, Porter D, Leander CA, Vilgalys R, Longcore JE. 2000. Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. Can J Bot 78:336–350. https://doi.org/10.1139/cjb-78-3-336.
60. Picard KT. 2017. Coastal marine habitats harbor novel early-diverging fungal diversity. Fungal Ecol 25:1–13. https://doi.org/10.1016/j.funeco.2016.10.006.
61. Hanafy RA, Elshahed MS, Liggenstoffer AS, Griffith GW, Youssef NH. 2017. *Pecoramyces ruminantium*, gen. nov., sp. nov., an anaerobic gut fungus from the feces of cattle and sheep. Mycologia 5514:37–41. https://doi.org/10.1080/00275514.2017.1317190.
62. Bryant MP. 1972. Commentary on the Hungate technique for culture of anaerobic bacteria. Am J Clin Nutr 25:1324–1328. https://doi.org/10.1093/ajcn/25.12.1324.
63. Balch WE, Wolfe RS. 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of Methanobacterium ruminantium in a pressurized atmosphere. Appl Environ Microbiol 32:781–791.
64. Hungate RE, Macy J. 1973. The roll-tube method for cultivation of strict anaerobes. Bull Ecol Res Comm 17:123–126.
65. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084.
66. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527. https://doi.org/10.1038/nbt.3519.
67. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42:D699–D704. https://doi.org/10.1093/nar/gkt1183.
68. Spatafora J, Bakum J, Redhead S, Norvell L, Seifert K, Miller A. 2011. 1000 fungal genomes to be sequenced. IMA Fungus 2:41. https://doi.org/10.1007/BF03449505.
69. Chang Y, Wang S, Sekimoto S, Aerts A, Choi C, Clum A, LaButti K, Lindquist E, Ngan CY, Ohm RA, Salamov A, Grigoriev IV, Spatafora JW, Berbee M. 2015. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. Genome Biol Evol 7:1590–1601. https://doi.org/10.1093/gbe/evv090.
70. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, Haridas S, Kuo A, Salamov A, Ahrendt SR, Lau R, Bowen BP, Lipzen A, Sullivan W, Andreopoulos BB, Clum A, Lindquist E, Daum C, Northen TR, Kunde-Ramamoorthy G, Schmitz RJ, Gryganskyi A, Culley D, Magnuson J, James TY, O'Malley MA, Stajich JE, Spatafora JW, Visel A, Grigoriev IV. 2017. Widespread adenine N6-methylation of active genes in fungi. Nat Genet 49:964–968. https://doi.org/10.1038/ng.3859.
71. Beaudet D, Chen ECH, Mathieu S, Yildirir G, Ndikumana S, Dalpé Y, Séguin S, Farinelli L, Stajich JE, Corradi N. 2018. Ultra-low input transcriptomics reveal the spore functional content and phylogenetic affiliations of poorly studied arbuscular mycorrhizal fungi. DNA Res 25:217–227. https://doi.org/10.1093/dnares/dsx051.
72. Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, James TY, O'Donnell K, Roberson RW, Taylor TN, Uehling J, Vilgalys R, White MM, Stajich JE. 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia 108:1028–1046. https://doi.org/10.3852/16-042.
73. Stajich JE, Block D, Boulez K, Steven E, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JGR, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res 12:1611–1618. https://doi.org/10.1101/gr.361602.
74. Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695–1701. https://doi.org/10.1093/molbev/mss020.
75. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
76. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. https://doi.org/10.1186/1471-2148-7-214.
77. Krings M, Taylor TN, Martin H. 2016. An enigmatic fossil fungus from the 410 Ma Rhynie chert that resembles *Macrochytrium* (Chytridiomycota) and *Blastocladiella* (Blastocladiomycota). Mycologia 108:303–312. https://doi.org/10.3852/15-224.
78. Strullu-Derrien C, Goral T, Longcore JE, Olesen J, Kenrick P, Edgecombe GD. 2016. A new chytridiomycete fungus intermixed with crustacean resting eggs in a 407-million-year-old continental freshwater environment. PLoS One 11:e0167301. https://doi.org/10.1371/journal.pone.0167301.
79. Lutzoni F, Nowak MD, Alfaro ME, Reeb V, Miadlikowska J, Krug M, Arnold AE, Lewis LA, Swofford D, Hibbett D, Hilu K, James TY, Quandt

D, Magallón S. 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. Nat Commun 9:1–11. https://doi.org/10.1038/s41467-018-07849-9.
80. Wang Y, White MM, Moncalvo J-M. 2019. Diversification of the gut fungi *Smittium* and allies (Harpellales) co-occurred with the origin of complete metamorphosis of their symbiotic insect hosts (lower Diptera). Mol Phylogenet Evol 139:106550. https://doi.org/10.1016/j.ympev.2019.106550.
81. Stadler T. 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. J Theor Biol 261:58–66. https://doi.org/10.1016/j.jtbi.2009.07.018.
82. Ekseth OK, Kuiper M, Mironov V. 2014. orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics 30:734–736. https://doi.org/10.1093/bioinformatics/btt582.
83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
84. Van Dongen S. 2000. Graph clustering by flow simulation. University of Utrecht, Utrecht, The Netherlands.
85. Gaujoux R, Seoighe C. 2010. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11:367. https://doi.org/10.1186/1471-2105-11-367.
86. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blöcker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MCT, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, Della Valle G, Fryc S, Guérin G, Hasegawa T, Hill EW, Jurka J, Kiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Røed KH, Ryder OA, Searle S, Skow L, Swinburne JE, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326:865–867. https://doi.org/10.1126/science.1178158.
87. The International Sheep Genomics Consortium, Archibald AL, Cockett NE, Dalrymple BP, Faraut T, Kijas JW, Maddox JF, McEwan JC, Hutton Oddy V, Raadsma HW, Wade C, Wang J, Wang W, Xun X. 2010. The sheep genome reference sequence: a work in progress. Anim Genet 41:449–453. https://doi.org/10.1111/j.1365-2052.2010.02100.x.
88. Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, Auvil L, Capitanu B, Ma J, Lewin HA, Qian X, Lang Y, Zhou R, Wang L, Wang K, Xia J, Liao S, Pan S, Lu X, Hou H, Wang Y, Zang X, Yin Y, Ma H, Zhang J, Wang Z, Zhang Y, Zhang D, Yonezawa T, Hasegawa M, Zhong Y, Liu W, Zhang Y, Huang Z, Zhang S, Long R, Yang H, Wang J, Lenstra JA, Cooper DN, Wu Y, Wang J, Shi P, Wang J, Liu J. 2012. The yak genome and adaptation to life at high altitude. Nat Genet 44:946–949. https://doi.org/10.1038/ng.2343.
89. Martin G, Baurens FC, Droc G, Rouard M, Cenci A, Kilian A, Hastie A, Doležel J, Aury JM, Alberti A, Carreel F, D'Hont A. 2016. Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. BMC Genomics 17:12. https://doi.org/10.1186/s12864-016-2579-4.
90. Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, Ooi LCL, Ooi SE, Chan KL, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan M, Budiman MA, Lee EK, Desalle R, Kudrna D, Goicoechea JL, Wing RA, Wilson RK, Fulton RS, Ordway JM, Martienssen RA, Sambanthamurthi R. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds. Nature 500:335–339. https://doi.org/10.1038/nature12309.
91. Peng Z, Lu Y, Li L, Zhao Q, Feng Q, Gao Z, Lu H, Hu T, Yao N, Liu K, Li Y, Fan D, Guo Y, Li W, Lu Y, Weng Q, Zhou C, Zhang L, Huang T, Zhao Y, Zhu C, Liu X, Yang X, Wang T, Miao K, Zhuang C, Cao X, Tang W, Liu G, Liu Y, Chen J, Liu Z, Yuan L, Liu Z, Huang X, Lu T, Fei B, Ning Z, Han B, Jiang Z. 2013. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). Nat Genet 45:456–461. https://doi.org/10.1038/ng.2569.
92. Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL. 2017. Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res 27:787–792. https://doi.org/10.1101/gr.213405.116.
93. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, Ware D. 2017. Improved maize reference genome with single-molecule technologies. Nature 546:524–527. https://doi.org/10.1038/nature22971.
94. The Rice Annotation Project, Itoh T, Tanaka T, Barrero RA, Yamasaki C, Fujii Y, Hilton PB, Antonio BA, Aono H, Apweiler R, Bruskiewich R, Bureau T, Burr F, Costa de Oliveira A, Fuks G, Habara T, Haberer G, Han B, Harada E, Hiaki AT, Hirochika H, Hoen D, Hokari H, Hosokawa S, Hsin YI, Ikawa H, Ikeo K, Imanishi T, Ito Y, Jaiswal P, Kanno M, Kawahara Y, Kawamura T, Kawashima H, Khurana JP, Kikuchi S, Komatsu S, Koyanagi KO, Kubooka H, Lieberherr D, Lin YC, Lonsdale D, Matsumoto T, Matsuya A, McCombie WR, Messing J, Miyao A, Mulder N, Nagamura N, Nam J, et al. 2007. Curated genome annotation of *Oryza sativa* ssp. japonica and comparative genome analysis with *Arabidopsis thaliana*. Genome Res 17:175–183. https://doi.org/10.1101/gr.5509507.
95. The International Brachypodium Initiative, Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, Tice H, Grimwood J, McKenzie N, Huo N, Gu YQ, Lazo GR, Anderson OD, You FM, Luo MC, Dvorak J, Wright J, Febrer M, Idziak D, Hasterok R, Lindquist E, Wang M, Fox SE, Priest HD, Filichkin SA, Givan SA, Bryant DW, Chang JH, Wu H, Wu W, Hsia AP, Schnable PS, Kalyanaraman A, Barbazuk B, Michael TP, Hazen SP, Bragg JN, Laudencia-Chingcuanco D, Weng Y, Haberer G, Spannagl M, Mayer K, Rattei T, Mitros T, Lee SJ, et al. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–768. https://doi.org/10.1038/nature08747.
96. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-Ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556. https://doi.org/10.1038/nature07723.
97. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E. 2007. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36:D1009–D1014. https://doi.org/10.1093/nar/gkm965.
98. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, et al. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69. https://doi.org/10.1126/science.1150646.
99. Kent WJ. 2002. BLAT—the BLAST-like alignment tool. Genome Res 12:656–664. https://doi.org/10.1101/gr.229202.
100. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, et al. 2018. Ensembl 2018. Nucleic Acids Res 46:D754–D761. https://doi.org/10.1093/nar/gkx1098.
101. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
102. Yoshino-Yasuda S, Karita S, Kato M, Kitamoto N. 2012. Sequence analysis and heterologous expression of rhamnogalacturonan lyase A gene (*AsrglA*) from *Shoyu Koji* mold, *Aspergillus sojae* KBN1340. Food Sci Technol Res 18:901–909. https://doi.org/10.3136/fstr.18.901.
103. Gómez-Cortecero A, Harrison RJ, Armitage AD. 2015. Draft genome sequence of a European isolate of the apple canker pathogen Neonectria ditissima. Genome Announc 3:e01243-15. https://doi.org/10.1128/genomeA.01243-15.

104. Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery JF, Hainaut M, Henrissat B, Lespinet O, Sacristán S, Ver Loren Van Themaat E, Kemen E, McHardy AC, Schulze-Lefert P, O'Connell RJ. 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. Nat Commun 7:11362. https://doi.org/10.1038/ncomms11362.
105. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203. https://doi.org/10.1093/nar/gkw1129.
106. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.
107. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
108. Maddison W, Maddison D. 2007. Mesquite: a modular system for evolutionary analysis, version 2.75. http://mesquiteproject.org.
109. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
110. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285.