

1 *Title:* Molecular dating of the emergence of anaerobic rumen fungi and the impact of laterally
2 acquired genes
3

4 *Short title:* Molecular dating and HGT of anaerobic gut fungi
5

6 *Authors:* Yan Wang^{*,†}, Noha Youssef[†], M.B. Couger[§], Radwa Hanafy[†], Mostafa Elshahed[†], Jason
7 E. Stajich^{*,†}
8

9 *Affiliations:*

10 ^{*} Department of Microbiology and Plant Pathology, University of California-Riverside,
11 Riverside, California, 92521 USA.

12 [†] Institute for Integrative Genome Biology, University of California-Riverside, Riverside,
13 California, 92521 USA.

14 [‡] Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater,
15 Oklahoma, 74074 USA.

16 [§] High Performance Computing Center, Oklahoma State University, Stillwater, Oklahoma, 74074
17 USA.

18

19 *To whom correspondence may be addressed:*

Yan Wang
University of California, Riverside
Department of Microbiology and Plant
Pathology
Riverside, CA 92521 USA
Phone: +1 951.386.5197
Email: yanxw.wang@gmail.com

Jason E. Stajich
University of California, Riverside
Department of Microbiology and Plant
Pathology
Riverside, CA 92521 USA
Phone: +1 951.827.2363
Email: jason.stajich@ucr.edu

20

21 *Keywords:* Comparative genomics, Divergence time estimation, HGT, Intron insertion,

22 Phylogenomics

23 **Abstract**

24 The anaerobic gut fungi (AGF) or Neocallimastigomycota inhabit the rumen and alimentary tract
25 of herbivorous mammals, where they play an important role in the degradation of plant fiber.
26 Comparative genomic and phylogenomic analysis of the AGF has long been hampered by their
27 fastidious growth pattern as well as their large and AT-biased genomes. We sequenced 21 AGF
28 transcriptomes and combined them with 5 available genome sequences of AGF taxa to explore
29 their evolutionary relationships, time their divergence, and characterize patterns of gene gain/loss
30 associated with their evolution. We estimate that the most recent common ancestor of the AGF
31 diverged 66 (± 10) million years ago, a timeframe that coincides with the evolution of grasses
32 (Poaceae), as well as the mammalian transition from insectivory to herbivory. The concordance
33 of these independently estimated ages of AGF evolution, grasses evolution, and mammalian
34 transition to herbivory suggest that AGF have been important in shaping the success of
35 mammalian herbivory transition by improving the efficiency of energy acquisition from
36 recalcitrant plant materials. Comparative genomics identified multiple lineage-specific genes and
37 protein domains in the AGF, two of which were acquired from an animal host (galectin) and
38 rumen gut bacteria (carbohydrate-binding domain) via horizontal gene transfer (HGT). Four of
39 the bacterial derived “Cthe_2159” genes in AGF genomes also encode eukaryotic Pfam domains
40 (“Atrophin-1”, “eIF-3_zeta”, “Nop14”, and “TPH”) indicating possible gene fusion events after
41 the acquisition of “Cthe_2159” domain. A third AGF domain, plant-like polysaccharide lyase N-
42 terminal domain (“Rhamnogal_lyase”), represents the first report from fungi that potentially aids
43 AGF to degrade pectin. Analysis of genomic and transcriptomic sequences confirmed the
44 presence and expression of these lineage-specific genes in nearly all AGF clades supporting the
45 hypothesis that these laterally acquired and novel genes in fungi are likely functional. These
46 genetic elements may contribute to the exceptional abilities of AGF to degrade plant biomass and
47 enable metabolism of the rumen microbes and animal hosts.

48 **Introduction**

49 Diverse microbes inhabit the digestive tract of ruminant mammals and contribute to degradation
50 of ingested plant fibers, a process that liberates nutrients for their hosts. Large scale genomic and
51 metagenomic sequencing of rumen microbes have produced hundreds of novel bacterial
52 genomes enabling discovery of plant biomass degrading enzymes and patterns of genomic

53 evolution (Seshadri et al., 2018; Stewart et al., 2018). However, eukaryotic members of the
54 rumen microbial community have been less intensely studied (Haitjema et al., 2017; Youssef et
55 al., 2013). Members of the phylum Neocallimastigomycota (anaerobic gut fungi or AGF) are
56 important members of the rumen and hindgut of a wide range of herbivorous mammals and
57 reptiles (Gruninger et al., 2014). To survive in this anoxic and prokaryotes-dominated
58 environment, extant AGF members have undergone multiple structures and metabolic
59 adaptations, including the loss of the mitochondria, a gain of a hydrogenosome, the loss of
60 respiratory capacities, and a substitution of ergosterol with tetrahymanol in the cell membrane
61 (Yarlett et al., 1986). Importantly, all known AGF taxa have a remarkably efficient plant biomass
62 degradation machinery, which may be critical for competing with other microbes for resources
63 and establishing growth in the herbivorous gut. Such capacity is reflected in the possession of an
64 impressive arsenal of plant biomass degradation enzymes and the production of the
65 cellulosomes—extracellular structures that harbor multiple enzymes bound to scaffoldins
66 (Haitjema et al., 2017). These metabolic and structural adaptations improve survivability, fitness,
67 and competitiveness of the AGF in the herbivorous gut, but the genetic and evolutionary origins
68 of these changes remain largely undescribed (Solomon et al., 2016; Youssef et al., 2013).
69 Previous genomic investigations of the AGF have identified a massive number of carbohydrate
70 active enzymes coded by genes with foreign origins presumably from multiple lineages of
71 bacteria through Horizontal Gene Transfer (HGT) events independently (Haitjema et al., 2017;
72 Solomon et al., 2016; Youssef et al., 2013). In fact, HGT examples from bacteria to fungi have
73 been documented extensively (Chaib De Mares et al., 2014; Dhillon et al., 2015; Gardiner et al.,
74 2012; Pombert et al., 2012). However, HGT elements in fungi that have been transferred from
75 other eukaryotes are still rare with only a few described cases from animals (Wang et al., 2016),
76 oomycetes (Sun et al., 2011), or plants (Richards et al., 2009). The rumen is an intriguing context
77 to explore patterns of HGT, where degradative enzymes break down sorts of cells liberating
78 DNA and RNA molecules. Competing organisms can find advantage by acquiring foreign genes
79 that operate efficiently in an anaerobic environment to obtain nutrients from recalcitrant plant
80 fibers or to recognize other microbes.

81 The Neocallimastigomycota are classified within the Chytridiomycota (chytrid) fungi,
82 which share the trait of a flagellated zoospore stage (James et al., 2006a, 2006b; Spatafora et al.,
83 2017; Stajich, 2017). Efforts to resolve the phylogenetic relationship of AGF and their sister

84 lineages using ribosomal markers have yielded conflicting topologies (Liggenstoffer et al., 2010;
85 Wang et al., 2017). Multilocus phylogeny or phylogenomics has not yet been applied to evaluate
86 their evolutionary relationships and to estimate the divergence time of the AGF. Using genomes
87 and transcriptomes from 26 different AGF taxa (Table 1) covering seven out of the ten
88 recognized genera, we reconstructed a robust phylogenomic tree of the AGF and estimated their
89 divergence time. We compared the genomes or transcriptomes of AGF and their non-rumen
90 associated relatives in Chytridiomycota to identify unique and shared genome contents. This
91 study examined the relatively recent divergence of the AGF clade and revealed a concordance of
92 the divergence time of the Neocallimastigomycota fungi with both the mammalian hosts
93 transition to herbivory and the diversification events of the forage grasses. As the AGF are well
94 known for their exceptional efficiency at plant biomass degradation, we also explored the diverse
95 genetic components of these fungi. We discovered two potential HGT elements that were found
96 unique to the AGF, which are predicted to have originated from animals or bacteria. Examination
97 of the family of bacterial transferred genes revealed multiple intron insertion events that occurred
98 after the HGT acquisition process, which are present in all five AGF genomes. Comparative
99 analyses of these genes suggest the intron insert events were related to intragenic duplication of
100 coding sequences. In addition, a novel plant polysaccharide lyase was revealed from both AGF
101 genomes and transcriptomes that has never been reported from any known fungal genomes or
102 genetic studies. The evolutionary genomic investigation of these rumen inhabiting fungi provides
103 perspective on the concordant timing of their divergence with the ecological niche they inhabit
104 and the potential role of HGT in accumulation of lineage-specific processes that may contribute
105 to their unique biology.

106 **Results**

107 **Divergence time estimation and phylogenomic relationship of Neocallimastigomycota**
108 Phylogenomic analysis placed the 26 AGF taxa into a single monophyletic clade with strong
109 support of Bayesian posterior probability (1.0/1.0) and maximum likelihood bootstrap value
110 (100%) (Figure 1 and Figure S1). All AGF genera (*Anaeromyces*, *Caecomyces*, *Feramyces*,
111 *Neocallimastix*, *Orpinomyces*, *Pecoramycetes*, and *Piromyces*) included in this study formed
112 individual monophyletic clades that were also supported by both Bayesian (Figure 1) and
113 maximum likelihood analyses (Figure S1). A conflict in the tree topology between the two

114 phylogenetic reconstructions is the placement of the *Caecomyces* clade. This lineage is sister to
115 the rest of the Neocallimastigomycota in the maximum likelihood tree (Figure S1), while the
116 *Caecomyces* position is swapped with *Piromyces* in the Bayesian phylogeny (Figure 1). This is
117 likely due to short internode distances, which suggest a rapid radiation of the ancestors of the
118 two genera. The relative short bar of the highest-probability density (HPD) on the node of the
119 AGF clade (Figure 1) suggests the integrative natural history of this group of fungi and the
120 outperforming resolving power of the genome-wide data in the molecular dating analyses.

121 The divergence time of the Neocallimastigomycota clade is estimated at the
122 Cretaceous/Paleogene (K/Pg) boundary 66 (± 10) Mya (Figure 1). The chronogram (Figure 1)
123 displays a long branch leading to the emergence of the AGF clade, which extends from the end
124 of Ediacaran (~564 Mya) to the K/Pg boundary (~66 Mya). This suggests that the extant
125 members of AGF did not emerge until recently and then rapidly radiated into separate clades in
126 the Paleogene. The estimated time frame for AGF divergence broadly coincides with the age of
127 the grasses (70-95 Mya), previously estimated using molecular (nuclear and chloroplast)
128 markers, and calibrated using fossils from pollen and dinosaur coprolite as well as the breakup
129 time of the Gondwana (Bremer, 2002; Christin et al., 2014; Gaut, 2002; Prasad et al., 2005;
130 Vicentini et al., 2008). In addition, this inferred AGF divergence time also coincides with a
131 major diet change of placental mammals: the transition from a primarily insectivorous to a
132 herbivorous and omnivorous lifestyles. The loss of chitinase genes diversity, estimated to
133 occurred from the Cretaceous/Paleogene (K/Pg) boundary (66 Mya) to the mid of Paleogene (34
134 Mya) (Figure 1), is widely seen as a consequence of such transition (Emerling et al., 2018).
135 Collectively, these overlapping estimates suggest that the evolution of the symbiotic association
136 between herbivorous mammals and rumen fungi is tightly linked with the evolution of forage
137 grasses and mammalian dietary transitions within a 66-95 Mya timeframe. The exact chronology
138 of these three divergence or transition events cannot be accurately determined partially due to the
139 intervals of the estimates (Figure 1). However, the dates inferred from phylogenetic analyses are
140 consistent with the hypothesis that rumen fungi have played important roles in the diet transition
141 of some mammals to acquire nutrition from forage grasses.

142 **Genome-wide comparison of protein domains and homologous genes**

143 Comparative genomic analysis between AGF and their non-rumen associated chytrid relatives
144 (Figure 2) identified 40 Pfam domains that are unique to the AGF, representing 0.67% of the

145 total number of Pfams (5,980) in the AGF pan-genome-transcriptome (Table S1 and Figure 2b).
146 The predicted functions of these domains include anaerobic ribonucleotide reductase (“NRDD”),
147 metal transport and binding (“FeoA”, “FeoB_C”), carbohydrate binding (e.g., “CBM_10”,
148 “CBM-like”, “Cthe_2159”), atypical protein kinase (“CotH”), and glycoside hydrolase (e.g.,
149 “Glyco_hydro_6”, “Glyco_hydro_11”) (Table S1 and Figure 2b). In addition to these 40 unique
150 AGF domains, many additional Pfams were also enriched in the AGF. Such domains mediate
151 polysaccharide degradation and monosaccharide fermentations (Figure 2c), including
152 “Chitin_binding_1”, “CBM_1”, “Cellulase”, “Glyco_hydro_10”, “Gly_radical”,
153 “RicinB_lectin_2”, “Esterase”, and “Polysacc_deac_1” domains. Further, our analysis also
154 identified 106 Pfam domains that are not present in AGF genomes and transcriptomes but found
155 in sister Chytridiomycota. Most of these missing domains are related to oxidation reactions on
156 cytochromes and mitochondria, instead, they possess specialized organelle called
157 hydrogenosome conducting metabolism in the anaerobic condition (Yarlett et al., 1986) (Table
158 S1 and Figure 2d). In addition, domains involved in the biosynthesis of nicotinic acid, uric acid,
159 purine catabolism, photolyase, and pathways of ureidoglycolate and kynurenine are also found to
160 be absent in AGF species. Similar patterns were also identified in the comparison of homologous
161 genes (Figure S2).

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

167 **Genomic interactions within the rumen of mammalian herbivores**

168 We focused on three Pfam domains (“Cthe_2159”, “Gal_Lectin”, and “Rhamnogal_lyase”) that
169 are unique to the Neocallimastigomycota and previously not observed in fungal genomes.
170 Phylogenetic analyses support a horizontal transfer of “Cthe_2159” from rumen bacteria into
171 AGF followed by potential gene fusion to deliver eukaryotic specific functions. Similarly,
172 analysis of “Gal_Lectin” domain copies in AGF suggests they were acquired from animal donor
173 lineages. Similarity search of AGF “Rhamnogal_lyase” domain finds most similar copies in
174 plant genomes and phylogenetic analysis indicates the AGF polysaccharide lyase domain is
175 distinct and not orthologous to related enzymes in other fungi.

176 A bacteria-like biomass-binding and putatively polysaccharide lyase domain (“Cthe_2159”)

177 The “Cthe_2159” domain was originally characterized as a polysaccharide lyase-like protein in
178 the thermophilic and biomass-degrading bacterium *Clostridium thermocellum* (Close et al.,
179 2014). “Cthe_2159” are beta-helix proteins with the ability to bind celluloses and acid-sugars
180 (polygalacturonic acid, a major component of the pectin) and homologs are primarily found in
181 archaeal and bacterial genomes. Notably, a total 583 copies of the “Cthe_2159” domain were
182 identified in 5 genomes and 21 transcriptomes of AGF taxa, but reduced to a set of 126 clusters
183 based on overall protein similarity (>90%) due to redundancy in transcriptome assemblies. This
184 domain is absent in all other eukaryotic genomes examined in this study (Figure 3 and Table 2).
185 A phylogenetic tree of “Cthe_2159” homologs identified from Archaea, Bacteria, and AGF
186 suggest that the AGF “Cthe_2159” domains were acquired from bacteria through HGT (Figure
187 3). The likely donor was a gram-positive Firmicute (*Clostridiales*) (Maximum Likelihood
188 bootstrap value 98%) and the closest protein copies of “Cthe_2159” domains are encoded in the
189 *Oribacterium sinus*, *Oribacterium* sp., and *Hungatella hathewayi* genomes (Figure 3). Members
190 of the order Clostridiales are integral members of the rumen microbiome. Four of these AGF
191 “Cthe_2159” domain containing genes also encode eukaryotic Pfam protein domains (“Atrophin-
192 1”, “eIF-3_zeta”, “Nop14”, and “TPH”) at the 3’ position of the “Cthe_2159” domain. We
193 hypothesize these domains are the result of fusion after the acquisition of “Cthe_2159” domain.
194 The functions of these additional domains include initiation of the eukaryotic translation,
195 maturation of 18S rRNA, production of 40S ribosome, and meiosis-specific activities (Figure
196 4a). Approximately 30% of these AGF “Cthe_2159” gene models possess between 1 and 2
197 introns but there is limited spliced transcript evidence to provide confidence in the gene
198 structures, so the apparent intron gains could be artifacts of genome assembly or annotation
199 (Supplementary results) (Haitjema et al., 2017; Youssef et al., 2013).

200 An animal-like galactose binding lectin domain (“Gal_Lectin”)

201 “Gal_Lectin” domains were found in AGF genomes universally and absent in all other
202 examined chytrid and fungal genomes (Table 2). Phylogenetic analysis recovered a
203 monophyletic AGF “Gal_Lectin” clade which was not placed as a sister clade to the animals as
204 expected for a fungal gene. Instead, it was embedded within the animal homologs in the tree and
205 allies with one subgroup, polycystin-1 (PC-1 protein) (Figure 5a). The three separate animal
206 subclades contain protein members that harbor the “Gal_Lectin” domain but with dissimilar

207 sequence information and annotated functions (Figure 5). The genomes of ruminant hosts (e.g.,
208 horse, sheep) of the AGF fungi also encode three gene families containing the “Gal_Lectin”
209 domain, which can be observed in each of the animal subclades (Figure 5). The proteins in the
210 animal subclade 1 were annotated as the PC-1, which are homologues to the human “polycystic
211 kidney disease” (*PKD1*) genes. The members of the animal subclade 2 were searched by BLAST
212 for the the “adhesion G-protein coupled receptor L1/3” (*ADGRL1*). The animal subclade 3
213 contains homologs of the “EVA-1 protein”, most of which contain two adjacent copies of the
214 “Gal_Lectin” domain. The three subgroups of animal “Gal_Lectin” domains are also flanked by
215 disparate Pfam domains (Figure 5b). The gene phylogeny suggests an animal PC-1 protein as the
216 likely donor lineage for the AGF “Gal_Lectin” gene (Figure 5a), based on its closest sister
217 relationship. In addition, the AGF proteins also contain a Pfam “Glyco_transf_34” domain
218 (Figure 5b) which is absent in all animal homologs of the “Gal_Lectin” containing genes
219 suggesting its involvement in fungus specific activities in the rumen.

220 A novel fungal rhamnogalacturonate lyase (“Rhamnogal_lyase”) in AGF

221 In plants, the rhamnogalacturonate lyases are involved in the fruit ripening-related process, cell-
222 wall modification, and lateral-root and root-hair formation (Molina-Hidalgo et al., 2013; Ponniah
223 et al., 2017). The Pfam database classifies two types of domains for rhamnogalactoside
224 degrading activity: “Rhamnogal_lyase” and “RhgB_N”. They are both N-terminal catalytic
225 domains associated with the rhamnogalacturonan lyase protein (polysaccharide lyase family 4,
226 PL4) and flanked persistently by the group of “fn3_3” and “CBM-like” domains, with the
227 particular function to degrade the rhamnogalacturonan I (RG-I) backbone of pectin. The
228 “Rhamnogal_lyase” domain exists in members of plants and plant-pathogenic bacteria (e.g.
229 *Erwinia chrysanthemi*), whereas the “RhgB_N” domain has a wider distribution and can be
230 found in bacteria, fungi, and oomycetes (Finn et al., 2016). Sequence similarity searches using
231 the “Rhamnogal_lyase” domain against various protein sequence databases (e.g., EnsEMBL,
232 Mycocosm, Pfam) returned no homolog in any other fungi (except the AGF members), which
233 indicates that this domain is unique to AGF, plants, and bacteria. On the other hand, the
234 “RhgB_N” domain is widely shared by Dikarya fungi, Oomycetes, and bacteria. Although
235 “RhgB_N” and “Rhamnogal_lyase” domains are distantly related according to the sequence
236 similarity (24% between the copies of the *Aspergillus nidulans* and *An. robustus*), they
237 presumably share a common origin due to that they both physically located on the N-terminal

238 region of the PL4 proteins and they have resembling functions to degrade the pectin RG-I region.
239 The phylogenetic tree shows that although AGF “Rhamnogal_lyase” domains are more closely
240 related to the plant homologs than to the clades of fungi and oomycetes, these AGF
241 rhamnogalacturonate lyases likely have evolved a specific function in fungi (Figure 3). The
242 presence of the “Rhamnogal_lyase” domain in the rumen-associated fungi suggests that the AGF
243 may support an ability to soften, modify, and degrade the plant pectin within the anaerobic
244 rumen in a related but different way from plants.

245

246 **Discussion**

247 Microbial diversity of ruminants is a research hotspot for development of bioenergy tools
248 (Bryant, 1959; Marvin-Sikkema et al., 1994; Seshadri et al., 2018). The AGF fungi are an
249 important but understudied component of the ruminant microbiome and their obligate anaerobic
250 and relatively A-T rich genomes have limited the initial genomic resources for the group. In this
251 study, we produced the most phylogenetically broad transcriptome sampling of the
252 Neocallimastigomycota fungi to date to support phylogenomic and comparative analyses. Our
253 results contribute new insights into the natural history and dynamic evolution of these cryptic
254 ruminant gut fungi. The reconstructed phylogenomic species tree resolved previously
255 unanswered questions about the evolutionary relationships of the members of the AGF. In
256 addition, we provide the first estimation of the divergence time of AGF taxa, 66 (± 10) Mya
257 (Figure 1), which is in remarkable concordance with the divergence of the forage Poaceae
258 grasses (70-95 Mya) and dietary shifts in mammalian lineages (34-66 Mya) from insectivore to
259 herbivore and omnivore. Grass evolution enabled the herbivory transition, and this diet
260 adaptation drove an increase in developmental and morphological complexity of the digestive
261 tract, compartmentalization, and the development of dedicated anaerobic fermentation chambers
262 (e.g., rumen and caecum) in the herbivorous alimentary tract to improve biomass degradation
263 efficiency (Hackmann and Spain, 2010). This transition to plant-based (or plant-exclusive) diets
264 required additional partnership with microbes since mammals lack cellulolytic and hemi-
265 cellulolytic enzymes necessary to liberate sugars for absorption (Gruninger et al., 2014). In
266 addition, the genome content comparisons help illustrate and predict new biological roles AGF
267 play in the mammalian herbivore guts. The long branch that leads to the emergence of the

268 Neocallimastigomycota clade indicates the distinctiveness of the extant group of obligate
269 symbiotic fungi in the mammalian herbivores and implies the existence of undiscovered although
270 possibly extinct relatives of the Neocallimastigomycota and Chytridiomycota (Figure 1). Future
271 environmental and metagenome sequence exploration of anaerobic environments testing for
272 presence of these types of fungi may provide new observations that support their existence.

273 Our analyses identified multiple instances of Pfam domain gains (n=40) and losses
274 (n=106) within the Neocallimastigomycota clade (Figure 2 and Table S1). We identified three
275 AGF lineage specific protein domains which are absent from all other examined fungal genomes
276 (Table 2). Phylogenetic analyses support the hypothesis that they were acquired via HGT or
277 other noncanonical events. Phylogenetic analyses of “Cthe_2159” and “Gal_Lectin” indicate the
278 domains were separately transferred from the rumen bacteria and animal hosts horizontally
279 (Figures 3 and 5). The gains of these domains highlight how HGT has contributed to broaden the
280 lignocellulolytic capacities (through “Cthe_2159”) of the AGF and potentially increase their
281 abilities for cell recognition (through “Gal_Lectin”) within the rumen. The presence of four
282 eukaryotic Pfam domains fused with these bacteria-originated “Cthe_2159” genes in AGF
283 suggests they are truly eukaryotic and encoded in the fungal genomes (Figure 4) and not a
284 contamination artifact. Studies of intron gains and losses in fungal lineages have suggested the
285 ancestor was intron rich, an observation that is supported by intron rich chytrid genomes (Csuros
286 et al., 2011; Nielsen et al., 2004; Stajich et al., 2007). Although introns are present in gene
287 models of several “Cthe_2159” copies found in all available AGF genomes, and many are
288 flanked with duplicated coding sequences (in *Piromyces* sp. E2), we are not able to confidently
289 conclude these models experienced recent intron-insertion events as there is little support of
290 spliced mRNA transcripts originating from these loci (Supplementary results).

291 The “Cthe_2159” is a newly described protein family that bind cellulosic and pectic
292 substrates in the anaerobic and thermophilic bacterium *Clostridium thermocellum* (Close et al.,
293 2014). The crystal structure of the “Cthe_2159” suggests that it is a polysaccharide lyase family
294 with similarity with pectate lyases in the PL9 family. Similarly, “Rhamnogal_lyase” domains are
295 primarily function in the facilitation of cell wall modification in plants (Molina-Hidalgo et al.,
296 2013). Phytopathogenic bacteria can utilize their prokaryotic versions of the domain to
297 disorganize plant tissues to support the invasion (Laatu and Condemeine, 2003). Although we
298 cannot locate the original donor lineages of the AGF “Rhamnogal_lyase” domains (Figure 6),

299 their gain is a key synapomorphy of the extant AGF taxa and may contribute to the ability of
300 these fungi to access polysaccharides in plant cell walls. Both “Cthe_2159” and
301 “Rhamnogal_lyase” (PL4 family) domains function in pectin binding or degradation activities,
302 and the possession both suggests that AGF may have evolved abilities to deconstruct pectin with
303 an exceptional efficiency that distinguish themselves from other fungi (Table 2 and Figure 2).
304 Pectin is abundant in primary cell walls and the middle lamella in both dicotyledonous plants
305 (making up 20-35% dry weight) and grasses (2-10%) serving as a protection of plant cells from
306 degrading enzymes produced by animals (Salem et al., 2017; Vogel, 2008; Voragen et al., 2009;
307 Xiao and Anderson, 2013). Removal of pectin can effectively increase the surface of exposed
308 plant cell wall, and thus improve the accessibility of other polysaccharides (cellulose and
309 hemicellulose) masked by pectin (Pakarinen et al., 2012). Both of the “Cthe_2159” and
310 “Rhamnogal_lyase” proteins may have contributed to the high efficiency of the AGF biomass
311 degradability by uncoupling the pectin that glues cells together, increasing the exposed surface
312 areas, and thus allowing diverse polysaccharide enzymes to work on plant cells simultaneously
313 in the rumen. These protein domains could account for the superior performance of AGF to
314 weaken forage fibers and release polysaccharides (Borneman et al., 1989; Nagpal et al., 2009).
315 The AGF may benefit or depend on these acquired domains in their capacity as primary
316 degraders of ingested forage (Haitjema et al., 2014).

317 The “Gal_Lectin” domain bears the phylogenetic hallmark of being acquired from an
318 animal donor. Animals use galactose-binding lectins to recognize foreign entities (García-
319 Maldonado et al., 2017) and participate in anti-microbial defenses (Low et al., 2010; Uhlenbruck
320 and Steinhausen, 1977). Our results suggest that the “Gal_Lectin” domains in AGF are
321 homologous and closely related to animal PC-1 proteins (Figure 5a), which are transmembrane
322 proteins functioning in cell recognition (Hughes et al., 1995; Weston et al., 2003). *In vitro*, PC-1
323 shows binding ability to carbohydrate matrices and collagens type I, II, and IV (Weston et al.,
324 2001). As such, we postulate that the acquisition of the animal-like “Gal_Lectin” domain
325 contributes to the AGF abilities of cell-cell recognition and interaction with other microbes in the
326 rumen. Syntenic relationship of the coding genes shows that the AGF “Gal_Lectin” domains are
327 flanked by the “Glyco_transf_34” domain, which lacks of homologs in any other animals (Figure
328 5b and Figure S5). The AGF-equipped “Glyco_transf_34” belongs to the galactosyl transferase
329 GMA12/MNN10 family and may help catalyze the transfer of the sugar moieties in cooperating

330 with the adjacent “Gal_Lectin” domain. Our investigation found that HGT has contributed to the
331 AGF genome evolution with donors from both prokaryotes and eukaryotes. HGT may have
332 helped these fungi to acquire new functions and to thrive in the anaerobic gut as a key member of
333 the microbial community degrading plant materials in animal hosts.

334 Other than the arsenal of diverse enzyme profiles, the AGF have also been known to use
335 rhizoids and holdfasts to physically aid the fungal body to penetrate into the plant material
336 deeply, which is superior than other rumen microorganisms in terms of efficiency (Berlemont,
337 2017; Gruninger et al., 2014). Our study provides evidence that the rumen fungi are able to and
338 have actively acquire functional domains from the animal hosts and co-existing anaerobic
339 bacteria in the rumen. These exotic genetic elements encoded in Neocallimastigomycota
340 genomes may contribute to their distinctive function comprised of unique genomic assets
341 comparing to their free-living relatives. The long branch leading to the recent radiation of
342 Neocallimastigomycota (Figure 1) also suggests distinct evolutionary trajectory from the sister
343 Chytridiomycota lineages. Living as gut-dwellers in the strict anaerobic gut environment for over
344 66 million years, AGF have undergone reductive evolution on the mitochondria and eventually
345 transformed it to a new organelle—hydrosome (Marvin-Sikkema et al., 1994; Youssef et al.,
346 2013). Their ecological roles of AGF in such an extreme environment also endow their
347 exceptional ability for plant degradation. The AGF use both physical (deconstruction of
348 lignocelluloses) and biological (depolymerization) mechanisms before the fermentation of plant
349 polysaccharides. These steps require diverse enzymes capable of breaking chemical bonds in
350 carbohydrates including cellulases, hemicellulases, ligninases, and pectinases (Brandt et al.,
351 2013). These all drive the synapomorphic and autapomorphic characters discovered in AGF.
352 Currently few close relatives have been found and none cultured which subtend the long branch.
353 Environmental DNA investigations of extreme environment that may be a suitable niche of those
354 Neocallimastigomycota-like microbes may reveal potential relatives (James et al., 2000). For
355 example, a recent metagenomic survey from costal marine sediments suggests that some
356 operational taxonomic units (OTUs) could be assigned to Neocallimastigomycota using 28S
357 rRNA marker (Picard, 2017). Sampling of deep sea habitats and marine mammalian herbivores
358 could provide future discoveries of biodiversity and evolutionary importance for understanding
359 the evolutionary trajectory of the Neocallimastigomycota.
360

361 **Materials and Methods**

362 **Transcriptome and genome datasets**

363 We generated the transcriptomes of 21 strains of Neocallimastigomycota fungi from cow, sheep,
364 horse, and goat feces, and rumen fluid of fistulated cows in the Stillwater, OK area (Murphy et
365 al., n.d.) (Table 1). These strains were maintained under anaerobic conditions using the modified
366 Hungate method as described previously (Balch and Wolfe, 1976; Bryant, 1972; Hanafy et al.,
367 2017; Hungate and Macy, 1973) . Total volume of RNA was harvested from the growing fungal
368 strains and processed for transcriptomics sequencing, which was performed using an external
369 commercial service provided by Novogene (Beijing, China). The RNAseq data were assembled
370 into *de novo* transcript assemblies using Trinity (v2.6.6), followed by TransDecoder (v5.0.2) to
371 predict ORFs (Haas et al., 2013). The generated proteomes and corresponding coding sequences
372 were used as input to phylogenomic and comparative genomic analyses. The five published
373 Neocallimastigomycota genome sequences were obtained from JGI MycoCosm database
374 (Grigoriev et al., 2014; Spatafora, 2011). These are *Anaeromyces robustus* S4, *Neocallimastix*
375 *californiae* G1, *Pecoramyces ruminantium* C1A (synonym *Orpinomyces* sp.), *Piromyces finnis*
376 (v3.0), and *Piromyces* sp. E2 (Haitjema et al., 2017; Youssef et al., 2013). Five outgroup
377 Chytridiomycota taxa with sequenced genomes were chosen. These are *Chytromyces* sp. MP 71,
378 *Entophysycis helioformis* JEL805, *Gaertneriomyces semiglobifer* Barr 43, *Gonapodya prolifera*
379 JEL478, and *Rhizoclostridium globosum* JEL800 (Chang et al., 2015; Mondo et al., 2017).
380 Assembled transcriptomes, raw Illumina read sequences, and isolates metadata are deposited in
381 the GenBank with the BioProject ID PRJNA489922. All accession numbers are listed in the
382 Table 1.

383 **Phylogenomics and divergence time estimation**

384 A set of 434 highly conserved and generally single-copy protein coding genes in fungi and
385 animal and plant outgroups (DOI: 10.5281/zenodo.1413687) were used for phylogenomic
386 analyses in the PHYling pipeline (DOI: 10.5281/zenodo.1257002). Profile-Hidden-Markov-
387 Models of these markers were searched in the chytrid predicted protein sequences using
388 HMMER3 (v3.1b2). A total of 426 (out of 434) conserved orthologous markers were identified
389 with hmmsearch (cutoff=1E⁻¹⁰) in the 26 Neocallimastigomycota and 5 Chytridiomycota. The
390 identified protein sequence homologs in each species, for each phylogenetic marker, were

391 aligned with hmalign to the marker profile-HMM. The protein alignments were also back
392 translation into codon alignments guided by the protein alignment using the tool bp_mrtrans.pl
393 (Stajich et al., 2002). The protein and coding sequences of the markers were concatenated into a
394 super-alignment with 426 partitions defined by each gene marker. The 426 gene partitions were
395 further collapsed into 33 partitions by PartitionFinder v.2.1.1 with a greedy search to find
396 partitions with consistent phylogenetic signals (Lanfear et al., 2012). Phylogenetic trees were
397 constructed from this super-alignment and partition scheme with two methods—maximum
398 likelihood implemented in IQ-TREE (v.1.5.5) and Bayesian inference implemented in BEAST
399 (v.1.8.4) (Drummond and Rambaut, 2007; Nguyen et al., 2015). Configuration files for
400 divergence time estimation analysis were coded in BEAUTi v.1.8.4 using the 33 partitions and
401 two calibration priors: 1) a direct fossil record of Chytridiomycota from the Rhynie Chert (407
402 Mya) (Krings et al., 2016; Strullu-Derrien et al., 2016), and 2) the emergence time of
403 Chytridiomycota (573-770 Mya as 95% HPD) from earlier studies (Chang et al., 2015; Lutzoni
404 et al., 2018; Wang et al., n.d.). The Birth-Death incomplete sampling tree model was employed
405 for inter-species relationships analyses (Stadler, 2009). Unlinked strict clock models were used
406 for each partition. Archive of input files and analysis scripts used to perform the phylogenetic
407 analyses are available at Zenodo (DOI: 10.5281/zenodo.1447226). Three independent runs were
408 performed separately for 50 million generations each with random starting seeds. Sufficient ESS
409 (>200) values were obtained after the default burn-in (10%) for the final sampled trees. The
410 Maximum Clade Credibility (MCC) tree was compiled using TreeAnnotator v.1.8.4.

411 **Identification of AGF-specific genes and Pfam domains**

412 Orthologous genes across the 31 genomes or transcriptomes were identified using a comparative
413 genomic pipeline that utilized all-vs-all BLASTp (cutoff=1E⁻⁵) to obtain the similarity pairs,
414 Orthologue to identify putative orthologous relationships, and the Markov-Clustering Algorithm
415 (MCL using the inflation value of 1.5) to generate disjoint clusters and deployed in an analysis
416 pipeline (DOI: 10.5281/zenodo.1447224) (Altschul et al., 1990; Ekseth et al., 2014; Van
417 Dongen, 2000). Comparisons of shared gene content of the Orthologous clusters was performed
418 among the Chytridiomycota lineages using a permissive strategy of counting a gene family as
419 shared if it is missing in up to 5 of the 26 Neocallimastigomycota taxa and 1 of the 5 chytrids
420 genomes. In this scenario, genes absent in all chytrids genomes and maintained by more than 21
421 out of the 26 Neocallimastigomycota genomes/transcriptomes are defined as AGF unique genes;

422 on the other hand, genes missing from all Neocallimastigomycota and present in at least 4 out of
423 the 5 chytrids genomes are treated as AGF lost genes.

424 Protein domains were identified by searching the predicted proteomes from each genome
425 assembly or transcriptome assembly against the Protein Family (Pfam) database (v31.0, last
426 accessed at March 20th, 2018). The enrichment heatmap of the Pfam domains across the included
427 taxa was produced using the “aheatmap” function in the R package “NMF” based on the total
428 copy number count in each assembly (Gaujoux and Seoighe, 2010). Genes only present in the
429 AGF genomes and missing from all of the included free-living chytrids relatives were identified.

430 To identify genes in AGF that are likely important for interactions with mammalian hosts
431 and plant material breakdown, we further compared the five available AGF genomes to the
432 genomes of their animal hosts (e.g., sheep, horse, elephant, yak) (Broad Institute, 2018; Qiu et
433 al., 2012; The International Sheep Genomics Consortium et al., 2010; Wade et al., 2009), the diet
434 plant (e.g., moss, rice, palm, maize, sorghum) (Jiao et al., 2017; Martin et al., 2016; Paterson et
435 al., 2009; Peng et al., 2013; Rensing et al., 2008; Singh et al., 2013; Swarbreck et al., 2008; The
436 International Brachypodium Initiative et al., 2010; The Rice Annotation Project, 2007; Zimin et
437 al., 2017) (Table S2), and the 1,165 available fungal genomes from the ongoing 1KFG project
438 (Grigoriev et al., 2014; Spatafora, 2011; Spatafora et al., 2017; Stajich, 2017). To prioritize AGF
439 genes that may have been laterally acquired from these hosts, a Python script (Wang et al., 2016)
440 and similarity search tool BLAT (Kent, 2002) was applied to filter out genetic elements in AGF
441 with higher similarity to animal or plant homologs than any fungal ones, excluding the AGF
442 themselves. Candidate genes for lateral transfer were ranked by the combination of the two
443 strategies. The candidate genes with an assigned functional or biological process annotation were
444 analyzed with priority using phylogenetic reconstruction to infer their potential origin.

445 **Identification of homologous sequences and potential origin of HGT candidate loci**

446 Three Pfam domains “Cthe_2159”, “Gal_Lectin”, and “Rhamnogal_lyase” were identified to be
447 unique to the AGF genomes as compared to the Chytridiomycota fungi or all other fungal
448 members. To predict the donor lineages for these putative HGT events, we searched more
449 broadly for homologues in genome databases of Plant, Metazoa, Fungi, Bacteria, and Protists in
450 Ensembl (v37) (Zerbino et al., 2018) via the web-implemented HMMER tool
451 (<https://www.ebi.ac.uk/Tools/hmmer/>) (cutoff=1E⁻³). Additional fungal homologues were found
452 by searching the DOE JGI’s MycoCosm database (Grigoriev et al., 2014; Spatafora, 2011). The

453 profile Hidden Markov Model tool phmmer in the HMMer package (Eddy, 2011) was used to
454 search for similar sequences in the 1,165 fungal genomes using the query of edge-trimmed
455 domain sequences from *An. robustus* (cutoff=1E⁻³).

456 Members of the “Rhgb_N” sequences were obtained from the Pfam database classified in
457 the “Rhgb_N” (PF09284) family (Finn et al., 2016) along with the N-terminal sequences of the
458 rhamnogalacturonate lyase families A, B, and C from GenBank (Gomez-Cortecero et al., 2015;
459 Hacquard et al., 2016; Yoshino-Yasuda et al., 2012). A single dataset of “Rhgb_N” and
460 “Rhamnogal_lyase” family members from animals, fungi, plants, and bacteria was constructed
461 from these searches. Domain names were confirmed using NCBI’s conserved domain search tool
462 (cutoff=1E⁻⁵) with unaligned FASTA sequences (Marchler-Bauer et al., 2017). Similarly,
463 homologs of the “Gal_Lectin” and “Cthe_2159” were obtained by searching for similar
464 sequences in the previously described genome databases and the categorized Pfam database
465 (families of “Gal_Lectin (PF02140)” and “Cthe_2159 (PF14262)”). Homologous sequences
466 containing the “Cthe_2159” domain were only identified in Archaea and Bacteria, while the
467 AGF copies are the first eukaryotic representatives identified with this domain. Homologs of the
468 flanking domain “Glyco_transf_34” was obtained similarly from EnsEMBL genome databases
469 described above using the edge-trimmed domain sequence from *An. robustus* (cutoff=1E⁻⁵).
470 Highly similar sequences (>90%) were filtered using CD-HIT v4.6.4 followed by multiple
471 sequence alignment with MUSCLE v3.8.31 (Edgar, 2004; Fu et al., 2012).

472 Phylogenetic trees of the HGT candidates

473 In total, 747 sequences of the rhamnogalacturanate degradation proteins (including both
474 “Rhamnogal_lyase” and “Rhgb_N”) were included in the alignment. For the other two domains,
475 “Gal_Lectin” and “Cthe_2159”, the alignments include 297 and 234 unique variants
476 respectively. The “Cthe_2159” domain containing genes in the 5 AGF genomes were aligned
477 separately using MUSCLE v3.8.31 in Mesquite software (Edgar, 2004; Maddison and Maddison,
478 2007). Both the upstream and downstream flanking regions of the studied Pfam domain
479 sequences were trimmed using the Mesquite software (Maddison and Maddison, 2007).
480 Selection of the appropriate substitutional model, the maximum-likelihood phylogenetic tree
481 reconstruction, and the ultrafast bootstrapping (1000 replicates) were conducted using the IQ-
482 TREE v1.5.5 package (Hoang et al., 2017; Kalyaanamoorthy et al., 2017; Nguyen et al., 2015).

483 **Acknowledgements**

484 This work was supported by National Science Foundation Grants (DEB-1557110 to J.E.S. and
485 DEB-1557102 to N.Y. and M.E.). Y.W. acknowledges the Mycological Society of America for
486 the Translational Mycology Postdoctoral Award. Data analyses were performed on the
487 University of California Riverside High-Performance Computational Cluster supported by NSF
488 DBI-1429826 and NIH S10-OD016290.

489 **References**

490 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool.
491 *J Mol Biol* **215**:403–410. doi:10.1016/S0022-2836(05)80360-2

492 Balch WE, Wolfe RS. 1976. New approach to the cultivation of methanogenic bacteria: 2-
493 mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium*
494 *ruminantium* in a pressurized atmosphere. *Appl Environ Microbiol* **32**:781–791.

495 Berlemont R. 2017. Distribution and diversity of enzymes for polysaccharide degradation in
496 fungi. *Sci Rep* **7**:1–11. doi:10.1038/s41598-017-00258-w

497 Borneman WS, Akin DE, Ljungdahl LG. 1989. Fermentation products and plant cell wall-
498 degrading enzymes produced by monocentric and polycentric anaerobic ruminal fungi. *Appl*
499 *Environ Microbiol* **55**:1066–1073.

500 Brandt A, Gräsvik J, Hallett JP, Welton T. 2013. Deconstruction of lignocellulosic biomass with
501 ionic liquids. *Green Chem* **15**:550–583. doi:10.1039/c2gc36364j

502 Bremer K. 2002. Gondwanan evolution of the grass alliance of families (Poales). *Evolution (N Y)*
503 **56**:1374–1387. doi:10.1111/j.0014-3820.2002.tb01451.x

504 Broad Institute. 2018. Elephant Genome Project. [http://www.broadinstitute.org/scientific-
505 community/science/projects/mammals-models/elephant/elephant-genome-project](http://www.broadinstitute.org/scientific-community/science/projects/mammals-models/elephant/elephant-genome-project)

506 Bryant MP. 1972. Commentary of anaerobic on the Hungate technique for culture. *Am J Clin
507 Nutr.*

508 Bryant MP. 1959. Bacterial species of the rumen. *Bact Rev* **23**:125–153.

509 Chaib De Mares M, Hess J, Floudas D, Lipzen A, Choi C, Kennedy M, Grigoriev I V., Pringle
510 A. 2014. Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal
511 *Amanita*. *New Phytol* **205**:1552–1564. doi:10.1111/nph.13140

512 Chang Y, Wang S, Sekimoto S, Aerts A, Choi C, Clum A, LaButti K, Lindquist E, Ngan CY,
513 Ohm RA, Salamov A, Grigoriev I V., Spatafora JW, Berbee M. 2015. Phylogenomic
514 analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land
515 plants. *Genome Biol Evol* **7**:1590–1601. doi:10.1093/gbe/evv090

516 Christin PA, Spriggs E, Osborne CP, Strömberg CAE, Salamin N, Edwards EJ. 2014. Molecular
517 dating, evolutionary rates, and the age of the grasses. *Syst Biol* **63**:153–165.
518 doi:10.1093/sysbio/syt072

519 Close DW, D'angelo S, Bradbury ARM. 2014. A new family of β -helix proteins with similarities
520 to the polysaccharide lyases. *Acta Crystallogr Sect D Biol Crystallogr* **70**:2583–2592.
521 doi:10.1107/S1399004714015934

522 Csuros M, Rogozin IB, Koonin E V. 2011. A detailed history of intron-rich eukaryotic ancestors
523 inferred from a global survey of 100 complete genomes. *PLoS Comput Biol* **7**:1–9.
524 doi:10.1371/journal.pcbi.1002150

525 Dhillon B, Feau N, Aerts AL, Beauseigle S, Bernier L, Copeland A, Foster A, Gill N, Henrissat
526 B, Herath P, LaButti KM, Levasseur A, Lindquist EA, Majoor E, Ohm RA, Pangilinan JL,
527 Pribowo A, Saddler JN, Sakalidis ML, de Vries RP, Grigoriev I V., Goodwin SB, Tanguay
528 P, Hamelin RC. 2015. Horizontal gene transfer and gene dosage drives adaptation to wood
529 colonization in a tree pathogen. *Proc Natl Acad Sci* **112**:3451–3456.
530 doi:10.1073/pnas.1424293112

531 Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees.
532 *BMC Evol Biol* **7**:214. doi:10.1186/1471-2148-7-214

533 Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* **7**:e1002195.
534 doi:10.1371/journal.pcbi.1002195

535 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
536 throughput. *Nucleic Acids Res* **32**:1792–1797. doi:10.1093/nar/gkh340

537 Ekseth OK, Kuiper M, Mironov V. 2014. OrthAgogue: an agile tool for the rapid prediction of
538 orthology relations. *Bioinformatics* **30**:734–736. doi:10.1093/bioinformatics/btt582

539 Emerling CA, Delsuc F, Nachman MW. 2018. Chitinase genes (CHIAs) provide genomic
540 footprints of a post-Cretaceous dietary radiation in placental mammals. *Sci Adv* **4**:eaar6478.
541 doi:10.1126/sciadv.aar6478

542 Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M,
543 Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein
544 families database: towards a more sustainable future. *Nucleic Acids Res* **44**:D279–D285.
545 doi:10.1093/nar/gkv1344

546 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation
547 sequencing data. *Bioinformatics* **28**:3150–3152. doi:10.1093/bioinformatics/bts565

548 García-Maldonado E, Cano-Sánchez P, Hernandez-Santoyo A. 2017. Molecular and functional
549 characterization of a glycosylated galactose-binding lectin from *Mytilus californianus*. *Fish
550 Shellfish Immunol* **66**:564–574. doi:10.1016/j.fsi.2017.05.057

551 Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, Kazan K,

552 Chakraborty S, McDonald BA, Manners JM. 2012. Comparative pathogenomics reveals
553 horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathog* **8**.
554 doi:10.1371/journal.ppat.1002952

555 Gaujoux R, Seoighe C. 2010. A flexible R package for nonnegative matrix factorization. *BMC*
556 *Bioinformatics* **11**:367. doi:10.1186/1471-2105-11-367

557 Gaut BS. 2002. Evolutionary dynamics of grass genomes. *New Phytol* **154**:15–28.
558 doi:10.1046/j.1469-8137.2002.00352.x

559 Gomez-Cortecero A, Harrison RJ, Armitage AD. 2015. Draft genome sequence of a European
560 isolate of the apple canker pathogen *Neonectria ditissima*. *Genome Announc* **3**:10–11.
561 doi:10.1128/genomeA.01243-15

562 Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X,
563 Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal:
564 gearing up for 1000 fungal genomes. *Nucleic Acids Res* **42**:D699–D704.
565 doi:10.1093/nar/gkt1183

566 Gruninger RJ, Puniya AK, Callaghan TM, Edwards JE, Youssef N, Dagar SS, Fliegerova K,
567 Griffith GW, Forster R, Tsang A, Mcallister T, Elshahed MS. 2014. Anaerobic fungi
568 (phylum Neocallimastigomycota): Advances in understanding their taxonomy, life cycle,
569 ecology, role and biotechnological potential. *FEMS Microbiol Ecol* **90**:1–17.
570 doi:10.1111/1574-6941.12383

571 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
572 Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman
573 R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. *De novo*
574 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference
575 generation and analysis. *Nat Protoc* **8**:1494–1512. doi:10.1038/nprot.2013.084

576 Hackmann TJ, Spain JN. 2010. Ruminant ecology and evolution: perspectives useful to ruminant
577 livestock research and production. *J Dairy Sci* **93**:1320–1334. doi:10.3168/jds.2009-2071

578 Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm
579 U, Dallery JF, Hainaut M, Henrissat B, Lespinet O, Sacristán S, Ver Loren Van Themaat E,
580 Kemen E, McHardy AC, Schulze-Lefert P, O'Connell RJ. 2016. Survival trade-offs in plant
581 roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun*
582 **7**:11362. doi:10.1038/ncomms11362

583 Haitjema CH, Gilmore SP, Henske JK, Solomon K V, De Groot R, Kuo A, Mondo S, Salamov
584 AA, LaButti K, Zhao Z, Chiniquy J, Barry KW, Brewer HM, Purvine SO, Wright AT,
585 Hainaut M, Boxma B, Van Alen T, Hackstein JHP, Henrissat B, Baker SE, Grigoriev I V,
586 O'Malley MA. 2017. A parts list for fungal cellulosomes revealed by comparative
587 genomics. *Nat Microbiol* **2**:1–8. doi:10.1038/nmicrobiol.2017.87

588 Haitjema CH, Solomon K V., Henske JK, Theodorou MK, O'Malley MA. 2014. Anaerobic gut
589 fungi: advances in isolation, culture, and cellulolytic enzyme discovery for biofuel

590 production. *Biotechnol Bioeng* **111**:1471–1482. doi:10.1002/bit.25264

591 Hanafy RA, Elshahed MS, Liggenstoffer AS, Griffith GW, Youssef NH. 2017. Pecoramycetes
592 ruminantium, gen. nov, sp. nov., an anaerobic gut fungus from the feces of cattle and sheep.
593 *Mycologia* **5514**:37–41. doi:10.1080/07474938.2014.956624

594 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV. 2017. UFBoot2: improving the
595 ultrafast bootstrap approximation. *Mol Biol Evol* **35**:518–522. doi:10.1093/molbev/msx281

596 Hughes J, Ward C, Peral B, Aspinwall R, Clark K, San Millán J, Gamble V, Harris P. 1995. The
597 polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell
598 recognition domains. *Nat Genet* **10**:151–160. doi:10.1038/ng0695-151

599 Hungate RE, Macy J. 1973. The roll-tube method for cultivation of strict anaerobes. *Bull from*
600 *Ecol Res Comm* **17**:123–126.

601 James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker
602 E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE,
603 Hosaka K, Sung G-H, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC,
604 Wang Z, Wilson AW, Schüssler A, Longcore JE, O'Donnell K, Mozley-Standridge S,
605 Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR,
606 Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D,
607 Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts
608 RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA,
609 Lücking R, Büdel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R,
610 Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R. 2006a. Reconstructing
611 the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**:818–822.
612 doi:10.1038/nature05110

613 James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D, Powell MJ, Griffith GW,
614 Vilgalys R. 2006b. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and
615 description of a new phylum (Blastocladiomycota). *Mycologia* **98**:860–871.
616 doi:10.3852/mycologia.98.6.860

617 James TY, Porter D, Leander C a., Vilgalys R, Longcore JE. 2000. Molecular phylogenetics of
618 the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Can J*
619 *Bot* **78**:336–350. doi:10.1139/cjb-78-3-336

620 Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS,
621 Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR,
622 Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J,
623 Dawe RK, Hastie A, Rank DR, Ware D. 2017. Improved maize reference genome with
624 single-molecule technologies. *Nature* **546**:524–527. doi:10.1038/nature22971

625 Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder:
626 fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589.
627 doi:10.1038/nmeth.4285

628 Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res* **12**:656–664.
629 doi:10.1101/gr.229202.

630 Krings M, Taylor TN, Martin H. 2016. An enigmatic fossil fungus from the 410 Ma Rhynie chert
631 that resembles *Macrochytrium* (Chytridiomycota) and *Blastocladiella*
632 (Blastocladiomycota). *Mycologia* **108**:303–312. doi:10.3852/15-224

633 Laatu M, Condemine G. 2003. Rhamnogalacturonate lyase rhiE is secreted by the out system in
634 *Erwinia chrysanthemi*. *J Bacteriol* **185**:1642–1649. doi:10.1128/JB.185.5.1642-1649.2003

635 Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of
636 partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*
637 **29**:1695–1701. doi:10.1093/molbev/mss020

638 Liggenstoffer AS, Youssef NH, Couger MB, Elshahed MS. 2010. Phylogenetic diversity and
639 community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant
640 and non-ruminant herbivores. *ISME J* **4**:1225–1235. doi:10.1038/ismej.2010.49

641 Low DHP, Frecer V, Le Saux A, Srinivasan GA, Ho B, Chen J, Ding JL. 2010. Molecular
642 interfaces of the galactose-binding protein tectonin domains in host-pathogen interaction. *J
643 Biol Chem* **285**:9898–9907. doi:10.1074/jbc.M109.059774

644 Lutzoni F, Nowak MD, Alfaro ME, Reeb V, Miadlikowska J, Krug M, Arnold AE, Lewis LA,
645 Swofford D, Hibbett D, Hilu K, James TY, Quandt D, Magallón S. 2018. Contemporaneous
646 radiations of fungi and plants linked to symbiosis. *Nat Commun* **9**:1–11.
647 doi:10.1038/s41467-018-07849-9

648 Maddison W, Maddison D. 2007. Mesquite: a modular system for evolutionary analysis Version
649 2.75.

650 Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer
651 RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang
652 Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE:
653 functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*
654 **45**:D200–D203. doi:10.1093/nar/gkw1129

655 Martin G, Baurens FC, Droc G, Rouard M, Cenci A, Kilian A, Hastie A, Doležel J, Aury JM,
656 Alberti A, Carreel F, D'Hont A. 2016. Improvement of the banana “*Musa acuminata*”
657 reference sequence using NGS data and semi-automated bioinformatics methods. *BMC
658 Genomics* **17**:1–12. doi:10.1186/s12864-016-2579-4

659 Marvin-Sikkema FD, Driessen AJM, Gottschal JC, Prins RA. 1994. Metabolic energy generation
660 in hydrogenosomes of the anaerobic fungus *Neocallimastix*: evidence for a functional
661 relationship with mitochondria. *Mycol Res* **98**:205–212. doi:10.1016/S0953-
662 7562(09)80187-1

663 Molina-Hidalgo FJ, Franco AR, Villatoro C, Medina-Puche L, Mercado JA, Hidalgo MA,
664 Monfort A, Caballero JL, Muñoz-Blanco J, Blanco-Portales R. 2013. The strawberry

665 (*Fragaria × ananassa*) fruit-specific *rhamnogalacturonate lyase 1 (FaRLyase1)* gene
666 encodes an enzyme involved in the degradation of cell-wall middle lamellae. *J Exp Bot*
667 **64**:1471–1483. doi:10.1093/jxb/ers386

668 Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, Haridas S, Kuo A,
669 Salamov A, Ahrendt SR, Lau R, Bowen BP, Lipzen A, Sullivan W, Andreopoulos BB,
670 Clum A, Lindquist E, Daum C, Northen TR, Kunde-Ramamoorthy G, Schmitz RJ,
671 Gryganskyi A, Culley D, Magnuson J, James TY, O’Malley MA, Stajich JE, Spatafora JW,
672 Visel A, Grigoriev I V. 2017. Widespread adenine N6-methylation of active genes in fungi.
673 *Nat Genet* **49**:964–968. doi:10.1038/ng.3859

674 Murphy CM, Youssef NH, Hanafy RA, Couger MB, Jason E, Wang Y, Baker K, Dagar SS,
675 Griffith GW, Ibrahim F, Callaghan TM, Elshahed MS. n.d. Horizontal gene transfer forged
676 the evolution of anaerobic gut fungi into a phylogenetically distinct gut-dwelling fungal
677 lineage. *Revis.*

678 Nagpal R, Puniya AK, Griffith GW, Goel G, Puniya M, Sehgal JP, Singh K. 2009. Anaerobic
679 rumen fungi: potential and applications In: Khachatourians GG, Arora DK, Rajendran TP,
680 Srivastava AK, editors. *Agriculturally Important Microorganisms*. Academic World
681 International. pp. 375–393.

682 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
683 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
684 **32**:268–274. doi:10.1093/molbev/msu300

685 Nielsen CB, Friedman B, Birren B, Burge CB, Galagan JE. 2004. Patterns of intron gain and loss
686 in fungi. *PLoS Biol* **2**. doi:10.1371/journal.pbio.0020422

687 Pakarinen A, Zhang J, Brock T, Maijala P, Viikari L. 2012. Enzymatic accessibility of fiber
688 hemp is enhanced by enzymatic or chemical removal of pectin. *Bioresour Technol*
689 **107**:275–281. doi:10.1016/j.biortech.2011.12.101

690 Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G,
691 Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T,
692 Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev I V., Lyons E, Maher CA, Martis
693 M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC,
694 Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R,
695 Peterson DG, Mehboob-Ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar
696 DS. 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**:551–
697 556. doi:10.1038/nature07723

698 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
699 W, Lu Y, Weng Q, Zhou C, Zhang L, Huang T, Zhao Y, Zhu C, Liu X, Yang X, Wang T,
700 Miao K, Zhuang C, Cao X, Tang W, Liu G, Liu Y, Chen J, Liu Z, Yuan L, Liu Z, Huang X,
701 Lu T, Fei B, Ning Z, Han B, Jiang Z. 2013. The draft genome of the fast-growing non-
702 timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat Genet* **45**:456–461.
703 doi:10.1038/ng.2569

704 Picard KT. 2017. Coastal marine habitats harbor novel early-diverging fungal diversity. *Fungal*
705 *Ecol* **25**:1–13. doi:10.1016/j.funeco.2016.10.006

706 Pombert J-F, Selman M, Burki F, Bardell FT, Farinelli L, Solter LF, Whitman DW, Weiss LM,
707 Corradi N, Keeling PJ. 2012. Gain and loss of multiple functionally related, horizontally
708 transferred genes in the reduced genomes of two microsporidian parasites. *Proc Natl Acad*
709 *Sci* **109**:12638–12643. doi:10.1073/pnas.1205020109

710 Ponniah SK, Thimmapuram J, Bhide K, Kalavacharla VK, Manoharan M. 2017. Comparative
711 analysis of the root transcriptomes of cultivated sweetpotato (*Ipomoea batatas* [L.] Lam)
712 and its wild ancestor (*Ipomoea trifida* [Kunth] G. Don). *BMC Plant Biol* **17**:1–14.
713 doi:10.1186/s12870-016-0950-x

714 Prasad V, Strömberg CAE, Alimohammadian H, Sahni A. 2005. Dinosaur coprolites and the
715 early evolution of grasses and grazers. *Science* **310**:1177–1180.
716 doi:10.1126/science.1118806

717 Qiu Q, Zhang G, Ma T, Qian W, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, Auvil L, Capitanu B,
718 Ma J, Lewin HA, Qian X, Lang Y, Zhou R, Wang L, Wang K, Xia J, Liao S, Pan S, Lu X,
719 Hou H, Wang Y, Zang X, Yin Y, Ma H, Zhang J, Wang Z, Zhang Y, Zhang D, Yonezawa
720 T, Hasegawa M, Zhong Y, Liu W, Zhang Y, Huang Z, Zhang S, Long R, Yang H, Lenstra
721 JA, Cooper DN, Wu Y, Wang J, Shi P, Wang J, Liu J, Wang J. 2012. The yak genome and
722 adaptation to life at high altitude. *Nat Genet* **44**:946–949. doi:10.1038/ng.2343

723 Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F,
724 Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki
725 Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A,
726 Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho
727 SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA,
728 Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M,
729 Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu S-H, Stueber
730 K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D,
731 Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev I V, Quatrano RS, Boore
732 JL. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of
733 land by plants. *Science* **319**:64–69. doi:10.1126/science.1150646

734 Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. 2009. Phylogenomic
735 analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants
736 and fungi. *Plant Cell* **21**:1897–1911. doi:10.1105/tpc.109.065805

737 Salem H, Bauer E, Kirsch R, Vogel H, Fukatsu T, Kaltenpoth M, Salem H, Bauer E, Kirsch R,
738 Berasategui A, Cripps M, Weiss B, Koga R, Fukumori K, Vogel H, Fukatsu T, Kaltenpoth
739 M. 2017. Drastic genome reduction in an herbivore's pectinolytic symbiont. *Cell* **171**:1520–
740 1525. doi:10.1016/j.cell.2017.10.029

741 Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Elof-fadrosch EA,
742 Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-espino D, Perry R, Henderson G,
743 Creevey CJ, Terrapon N, Lapebie P, Drula E, Lombard V, Rubin E, Kyriides NC, Henrissat

744 B, Woyke T, Ivanova NN, Kelly WJ. 2018. Cultivation and sequencing of rumen
745 microbiome members from the Hungate1000 Collection. *Nat Biotechnol* **36**:359–367.
746 doi:10.1038/nbt.4110

747 Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, Ooi LCL, Ooi SE,
748 Chan KL, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan
749 M, Budiman MA, Lee EK, Desalle R, Kudrna D, Goicoechea JL, Wing RA, Wilson RK,
750 Fulton RS, Ordway JM, Martienssen RA, Sambanthamurthi R. 2013. Oil palm genome
751 sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **500**:335–
752 339. doi:10.1038/nature12309

753 Solomon K V., Haitjema CH, Henske JK, Gilmore SP, Borges-Rivera D, Lipzen A, Brewer HM,
754 Purvine SO, Wright AT, Theodorou MK, Grigoriev I V., Regev A, Thompson DA,
755 OMalley MA. 2016. Early-branching gut fungi possess a large, comprehensive array of
756 biomass-degrading enzymes. *Science (80-)* **351**:1192–1195. doi:10.1126/science.aad1431

757 Spatafora JW. 2011. 1000 fungal genomes to be sequenced. *IMA Fungus* **2**:41.

758 Spatafora JW, Aime MC, Grigoriev I V, Martin F, Stajich JE, Blackwell M. 2017. The fungal
759 tree of life: from molecular systematics to genome-scale phylogenies. *Microbiol Spectr* **5**:3–
760 34. doi:10.1128/microbiolspec.FUNK-0053-2016

761 Stadler T. 2009. On incomplete sampling under birth–death models and connections to the
762 sampling-based coalescent. *J Theor Biol* **261**:58–66. doi:10.1016/j.jtbi.2009.07.018

763 Stajich JE. 2017. Fungal genomes and insights into the evolution of the Kingdom. *Microbiol
764 Spectr* **5**:1–15. doi:10.1128/microbiolspec.FUNK-0055-2016

765 Stajich JE, Block D, Boulez K, Steven E. Brenner, Chervitz SA, Chris Dagdigian, Fuellen G,
766 Gilbert JGR, Korf I, Lapp H, Lehvastlaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock
767 MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. 2002. The
768 Bioperl toolkit: Perl modules for the life sciences. *Genome Res* **12**:1611–1618.
769 doi:10.1101/gr.361602.the

770 Stajich JE, Dietrich FS, Roy SW. 2007. Comparative genomic analysis of fungal genomes
771 reveals intron-rich ancestors. *Genome Biol* **8**. doi:10.1186/gb-2007-8-10-r223

772 Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ,
773 Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. metagenomic sequencing of the cow
774 rumen. *Nat Commun* **9**:870. doi:10.1038/s41467-018-03317-6

775 Strullu-Derrien C, Goral T, Longcore JE, Olesen J, Kenrick P, Edgecombe GD. 2016. A new
776 chytridiomycete fungus intermixed with crustacean resting eggs in a 407-millionyear-old
777 continental freshwater environment. *PLoS One* **11**:1–14. doi:10.1371/journal.pone.0167301

778 Sun G, Yang Z, Kosch T, Summers K, Huang J. 2011. Evidence for acquisition of virulence
779 effectors in pathogenic chytrids. *BMC Evol Biol* **11**. doi:10.1186/1471-2148-11-195

780 Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer
781 T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E.
782 2008. The *Arabidopsis* Information Resource (TAIR): gene structure and function
783 annotation. *Nucleic Acids Res* **36**:D1009–D1014. doi:10.1093/nar/gkm965

784 The International Brachypodium Initiative, Vogel JP, Garvin DF, Mockler TC, Schmutz J,
785 Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, Tice H, Grimwood J,
786 McKenzie N, Huo N, Gu YQ, Lazo GR, Anderson OD, You FM, Luo MC, Dvorak J,
787 Wright J, Febrer M, Idziak D, Hasterok R, Lindquist E, Wang M, Fox SE, Priest HD,
788 Filichkin SA, Givan SA, Bryant DW, Chang JH, Wu H, Wu W, Hsia AP, Schnable PS,
789 Kalyanaraman A, Barbazuk B, Michael TP, Hazen SP, Bragg JN, Laudencia-Chingcuanco
790 D, Weng Y, Haberer G, Spannagl M, Mayer K, Rattei T, Mitros T, Lee SJ, Rose JKC,
791 Mueller LA, York TL, Wicker T, Buchmann JP, Tanskanen J, Schulman AH, Gundlach H,
792 Beven M, Costa De Oliveira A, Da C. Maia L, Belknap W, Jiang N, Lai J, Zhu L, Ma J, Sun
793 C, Pritham E, Salse J, Murat F, Abrouk M, Bruggmann R, Messing J, Fahlgren N, Sullivan
794 CM, Carrington JC, Chapman EJ, May GD, Zhai J, Ganssmann M, Gurazada SGR, German
795 M, Meyers BC, Green PJ, Tyler L, Wu J, Thomson J, Chen S, Scheller H V., Harholt J,
796 Ulvskov P, Kimbrel JA, Bartley LE, Cao P, Jung KH, Sharma MK, Vega-Sanchez M,
797 Ronald P, Dardick CD, De Bodt S, Verelst W, Inzé D, Heese M, Schnittger A, Yang X,
798 Kalluri UC, Tuskan GA, Hua Z, Vierstra RD, Cui Y, Ouyang S, Sun Q, Liu Z, Yilmaz A,
799 Grotewold E, Sibout R, Hematy K, Mouille G, Höfte H, Micheel T, Pelloux J, O'Connor D,
800 Schnable J, Rowe S, Harmon F, Cass CL, Sedbrook JC, Byrne ME, Walsh S, Higgins J, Li
801 P, Brutnell T, Unver T, Budak H, Belcram H, Charles M, Chalhoub B, Baxter I. 2010.
802 Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*
803 **463**:763–768. doi:10.1038/nature08747

804 The International Sheep Genomics Consortium, Archibald AL, Cockett NE, Dalrymple BP,
805 Faraut T, Kijas JW, Maddox JF, McEwan JC, Hutton Oddy V, Raadsma HW, Wade C,
806 Wang J, Wang W, Xun X. 2010. The sheep genome reference sequence: a work in progress.
807 *Anim Genet* **41**:449–453. doi:10.1111/j.1365-2052.2010.02100.x

808 The Rice Annotation Project. 2007. Curated genome annotation of *Oryza sativa* ssp. *japonica*
809 and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res* **17**:175–183.
810 doi:10.1101/gr.5509507.

811 Uhlenbrück G, Steinhausen G. 1977. Tridacnins: symbiosis-profit or defense-purpose? *Dev
812 Comp Immunol* **1**:183–192.

813 Van Dongen S. 2000. Graph clustering by flow simulation. University of Utrecht.
814 doi:10.1016/j.cosrev.2007.05.001

815 Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA. 2008. The age of the grasses
816 and clusters of origins of C4 photosynthesis. *Glob Chang Biol* **14**:2963–2977.
817 doi:10.1111/j.1365-2486.2008.01688.x

818 Vogel J. 2008. Unique aspects of the grass cell wall. *Curr Opin Plant Biol* **11**:301–307.
819 doi:10.1016/j.pbi.2008.03.002

820 Voragen AGJ, Coenen GJ, Verhoef RP, Schols HA. 2009. Pectin, a versatile polysaccharide
821 present in plant cell walls. *Struct Chem* **20**:263–275. doi:10.1007/s11224-009-9442-z

822 Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey
823 E, Bellone RR. 2009. Genome sequence, comparative analysis, and population genetics of
824 the domestic horse. *Science* **326**:865–867. doi:10.1126/science.1178158.Genome

825 Wang X, Liu X, Groenewald JZ. 2017. Phylogeny of anaerobic fungi (phylum
826 Neocallimastigomycota), with contributions from yak in China. *Antonie Van Leeuwenhoek*
827 **110**:87–103. doi:10.1007/s10482-016-0779-1

828 Wang Y, White M, Moncalvo J. n.d. Divergence time estimation of the gut fungi *Smittium*
829 (Harpellales) suggests a co-emergence with the complete metamorphosis of their lower
830 Diptera (fly) hosts. *Rev.*

831 Wang Y, White MM, Kvist S, Moncalvo J-M. 2016. Genome-wide survey of gut fungi
832 (Harpellales) reveals the first horizontally transferred ubiquitin gene from a mosquito host.
833 *Mol Biol Evol* **33**:2544–2554. doi:10.1093/molbev/msw126

834 Weston BS, Bagnérés C, Price RG, Stirling JL. 2001. The polycystin-1 C-type lectin domain
835 binds carbohydrate in a calcium-dependent manner, and interacts with extracellular matrix
836 proteins in vitro. *Biochim Biophys Acta - Mol Basis Dis* **1536**:161–176. doi:10.1016/S0925-
837 4439(01)00046-1

838 Weston BS, Malhas AN, Price RG. 2003. Structure-function relationships of the extracellular
839 domain of the autosomal dominant polycystic kidney disease-associated protein, polycystin-
840 1. *FEBS Lett* **538**:8–13. doi:10.1016/S0014-5793(03)00130-3

841 Xiao C, Anderson CT. 2013. Roles of pectin in biomass yield and processing for biofuels. *Front
842 Plant Sci* **4**:1–7. doi:10.3389/fpls.2013.00067

843 Yarlett N, Orpin CG, Munn EA, Yarlett NC, Greenwood CA. 1986. Hydrogenosomes in the
844 rumen fungus *Neocallimastix patriciarum*. *Biochem J* **236**:729–739. doi:10.1042/bj2360729

845 Yoshino-Yasuda S, Karita S, Kato M, Kitamoto N. 2012. Sequence analysis and heterologous
846 expression of rhamnogalacturonan lyase A gene (*AsrglA*) from *Shoyu Koji* Mold,
847 *Aspergillus sojae* KBN1340. *Food Sci Technol Res* **18**:901–909. doi:10.3136/fstr.18.901

848 Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ, Atiyeh
849 HK, Wilkins MR, Elshahed MS. 2013. The genome of the anaerobic fungus orpinomyces
850 sp. strain c1a reveals the unique evolutionary history of a remarkable plant biomass
851 degrader. *Appl Environ Microbiol* **79**:4620–4634. doi:10.1128/AEM.00821-13

852 Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall
853 A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH,
854 Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W,
855 Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A,
856 Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A,

857 Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ,
858 Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates
859 A, Flicek P. 2018. *Nucleic Acids Res* **46**:D754–D761.
860 doi:10.1093/nar/gkx1098

861 Zimin A V., Puiu D, Luo M-C, Zhu T, Koren S, A.Yorke J, Dvorak J, Salzberg SL. 2017. Hybrid
862 assembly of the large and highly repetitive genome of *Aegilops tauschii*, a grogenitor of
863 bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res* **27**:787–792.
864 doi:10.1101/066100

865

866 **Figure Legends**

867 **Figure 1.** Bayesian phylogenomic Maximum Clade Credibility tree of Neocallimastigomycota
868 with divergence time estimation. All clades are fully supported by Bayesian posterior
869 probabilities (BPP). For clarity, mean ages and 95% highest-probability density ranges (blue
870 bars) are denoted on the nodes above the rank of genus.

871 **Figure 2.** Cladogram and heatmap enrichment of the Pfam domains between
872 Neocallimastigomycota and Chytridiomycota. (a) Cladogram showing the phylogenetic
873 relationship of the compared taxa (Neocallimastigomycota genomes are in bold, transcriptomes
874 in standard type); (b) Heatmap plot of natural logarithm of the domain copy numbers showing
875 the ones uniquely gained in Neocallimastigomycota (suggested HGT elements are in blue); (c)
876 Pfam domains highly enriched in Neocallimastigomycota; and (d) Pfam domains absent in
877 Neocallimastigomycota (presented domains are partial; see Table S1 for the full list).

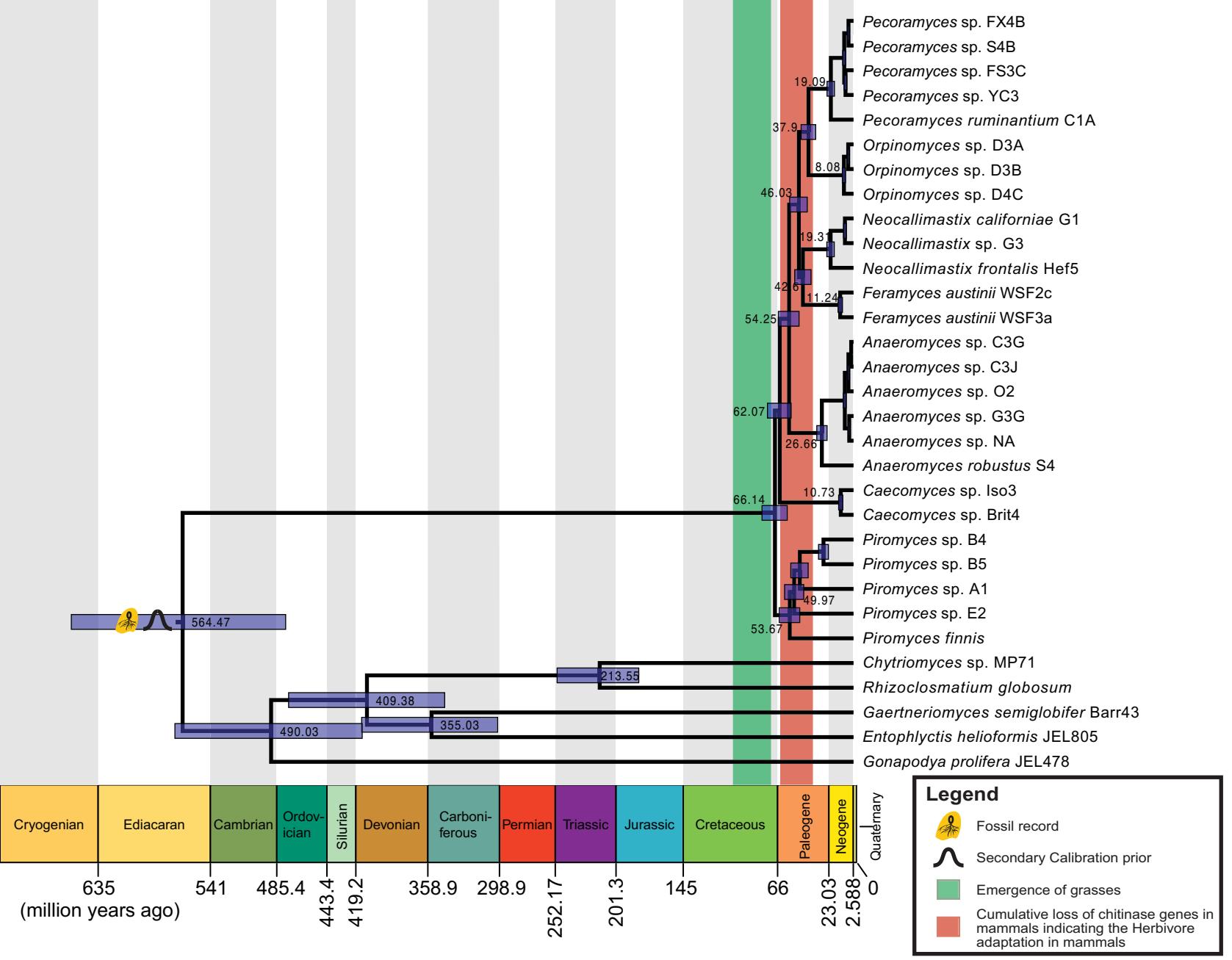
878 **Figure 3.** Mid-point rooted phylogenetic tree of the “Cthe_2159” domain. All 126
879 Neocallimastigomycota (AGF) copies (copies that >90% identities have been removed) form a
880 single clade (red) indicating the HGT donor, *Clostridiales bacterium* C5EMF8 (an obligate
881 rumen bacterium), with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

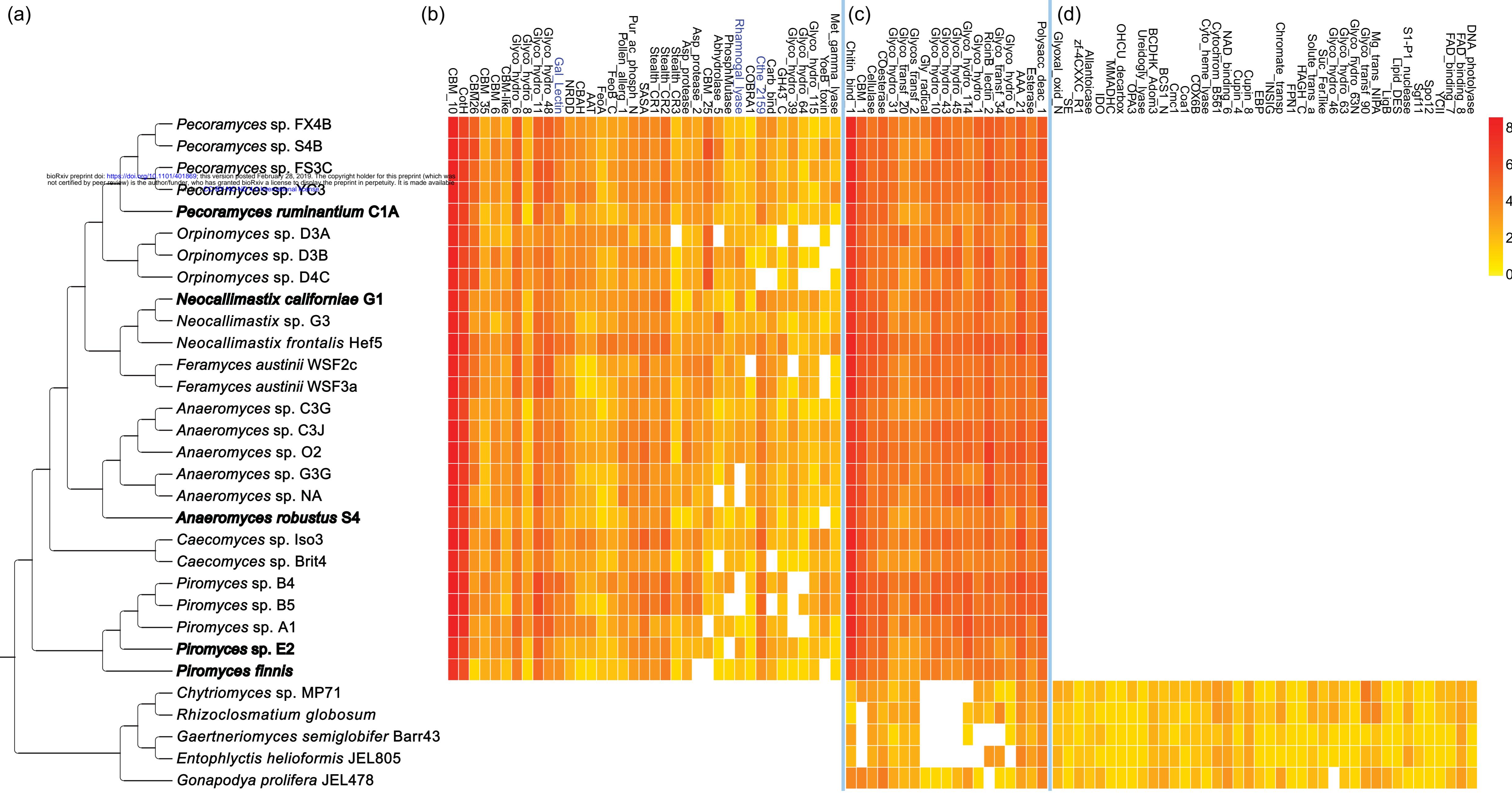
885 **Figure 4.** Phylogenetic tree of the 83 “Cthe_2159” domains identified in five AGF genomes
886 based on protein sequences (rooted with the closest related bacterial homolog). Domain maps on
887 the right shows the conserved domains produced by the “Cthe_2159” containing genes.

888 **Figure 5.** Phylogenetic tree of the animal-like “Gal-Lectin” domain identified in
889 Neocallimastigomycota. (a) Collapsed phylogenetic tree based on protein sequences (rooted with
890 the bacterial outgroup), including clades of Neocallimastigomycota (red), Animals (blue; three
891 clades are labeled as 1-3), plants (green), and bacteria (brown) (A complete tree with all tip
892 information is in the Figure S4); (b) Schematic diagrams showing the “Gal_Lectin” and other
893 conserved domains on the same protein in each clade individually (dotted box highlights the
894 aligned region used to produce the phylogenetic tree).

895 **Figure 6.** Radial phylogenetic tree of the “Rhamnogal_lyase” domain encoded by the
896 Neocallimastigomycota (red). Plant copies are colored in green and other homologous fungal
897 genes are colored in brown. Oomycetes are in cyan and animal copies only known in the
898 mountain pine beetles *Dendroctonus ponderosae* are in blue. Bacterial branches remain in black.
899 The tree also included homologs of “Rhgb_N” and “Rhamnogalacturonan lyase A, B, and C”.
900 Domain names are suggested using NCBI’s conserved domain search tool (cutoff 1E⁻⁵) with
901 unaligned FASTA sequences (refer to the Figure S6 for a tree with detailed information).

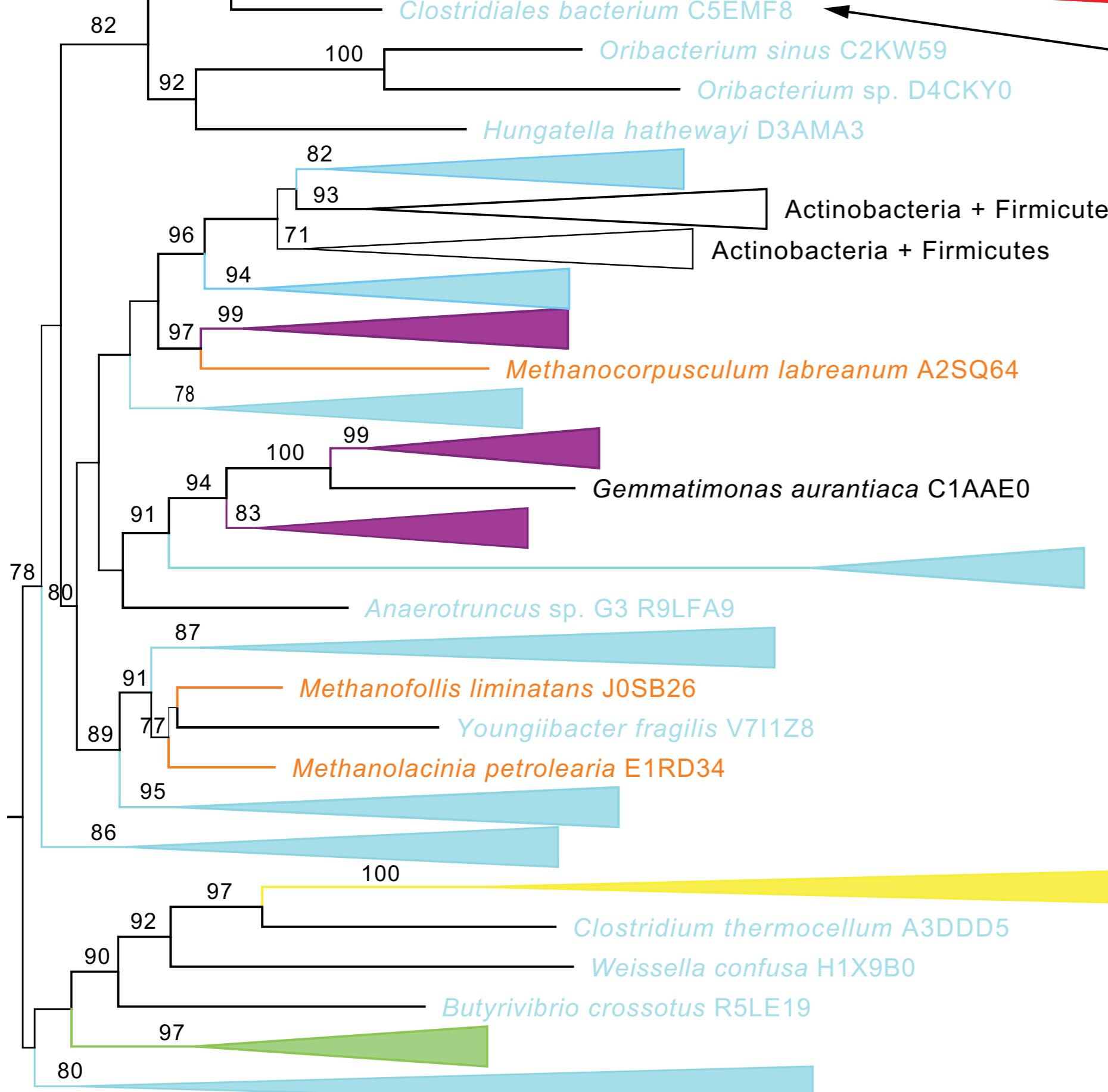
902





Heatmap of the genome-wide enriched Pfam domains (natural logarithm)

126 AGF copies (from both genomes and transcriptomes)



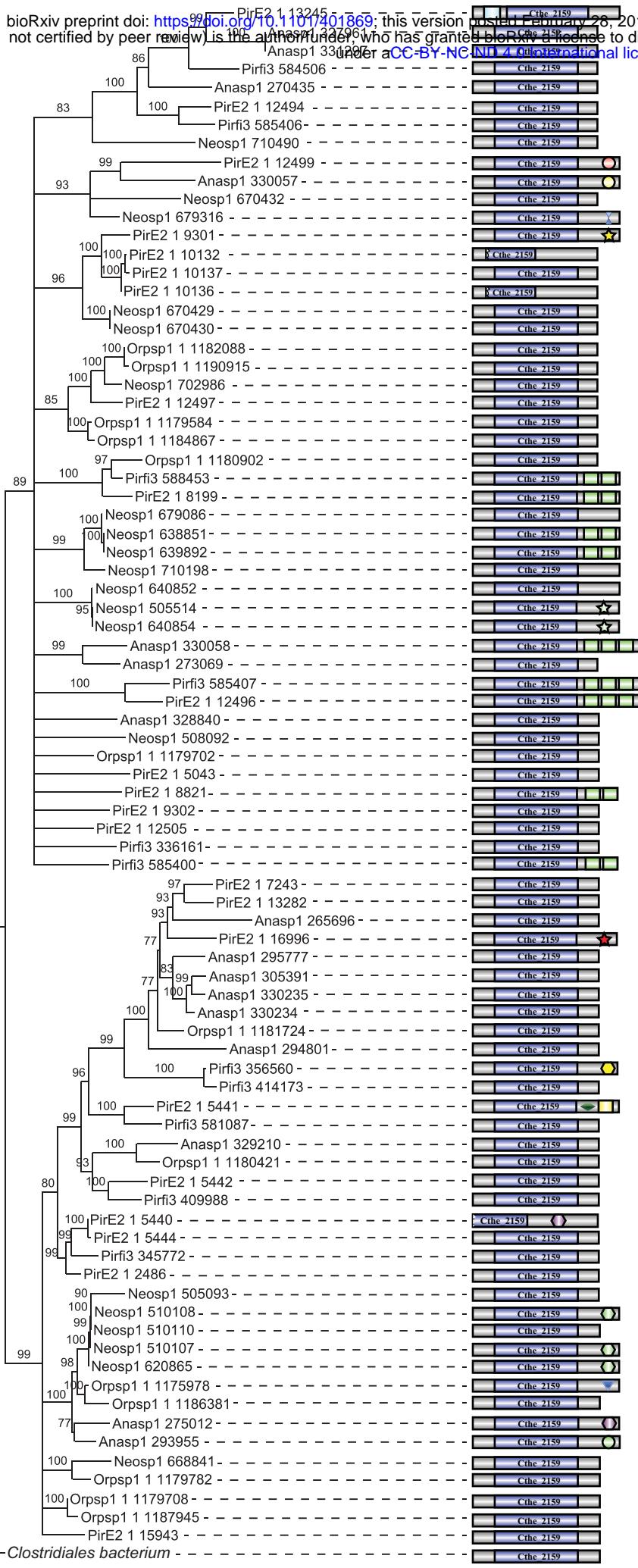
HGT donor: rumen bacterium

Legend

- Neocallimastigomycota
- Firmicutes
- Actinobacteria
- Bacteroidetes
- Fibrobacteres
- Archaea
- Gemmatimonadetes
- Mixed phyla

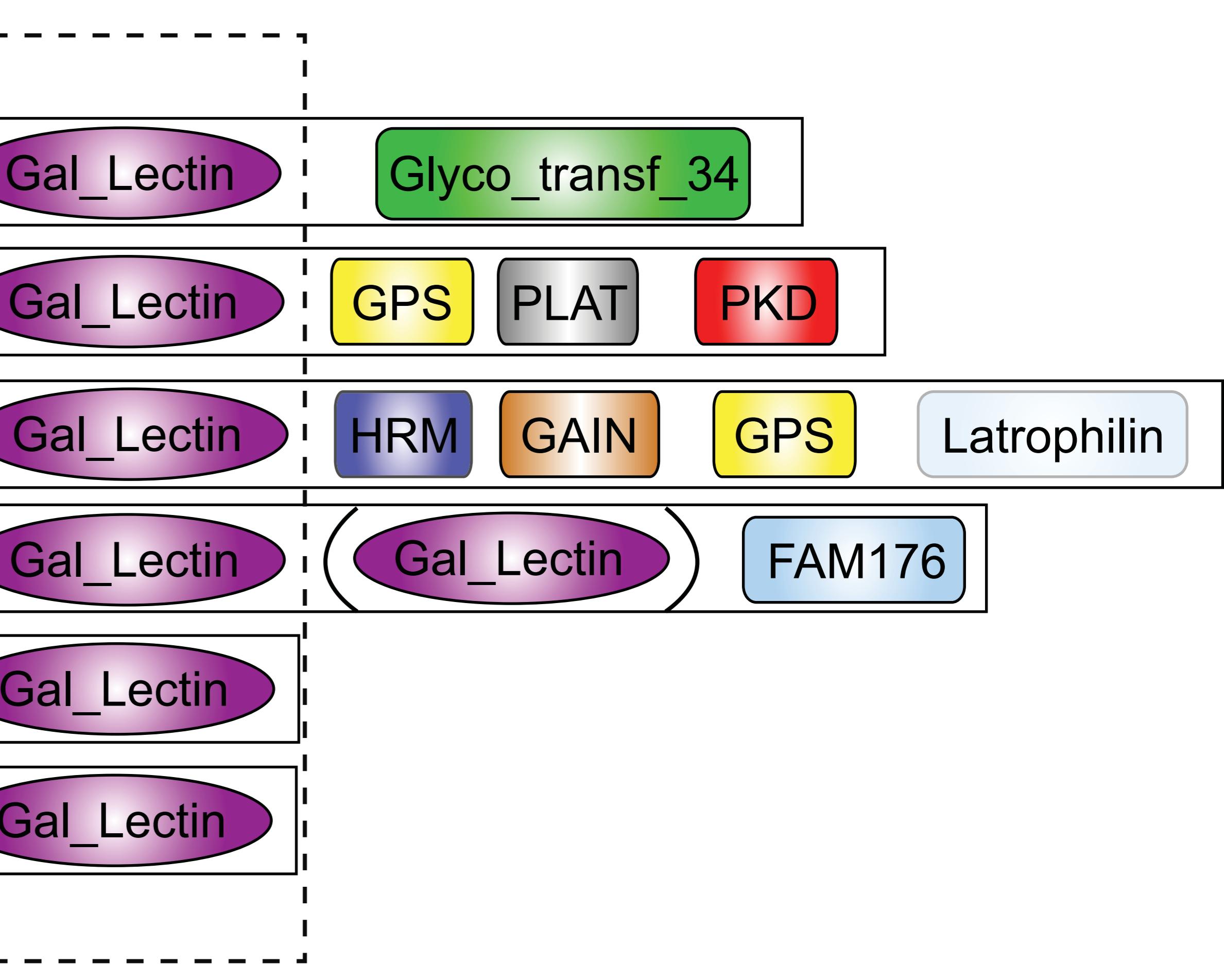
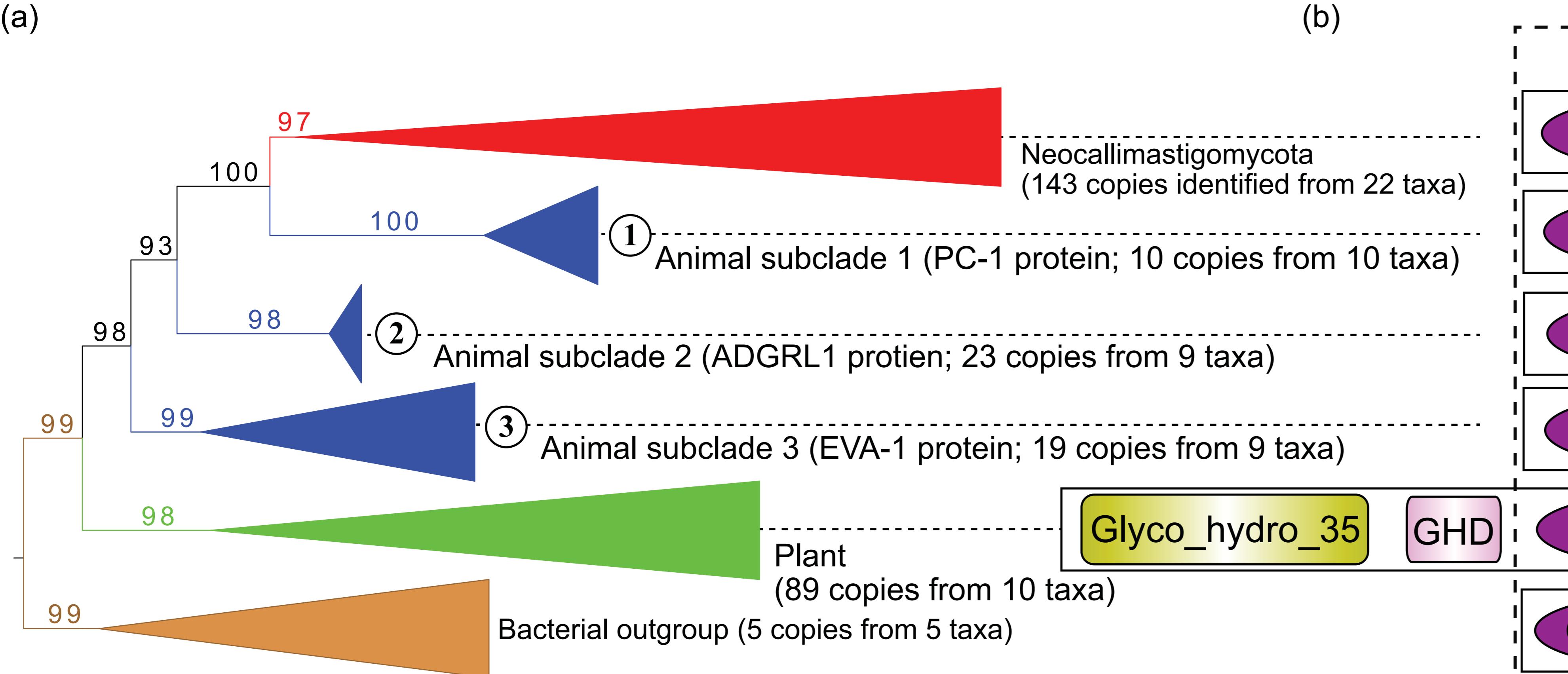
Domain map

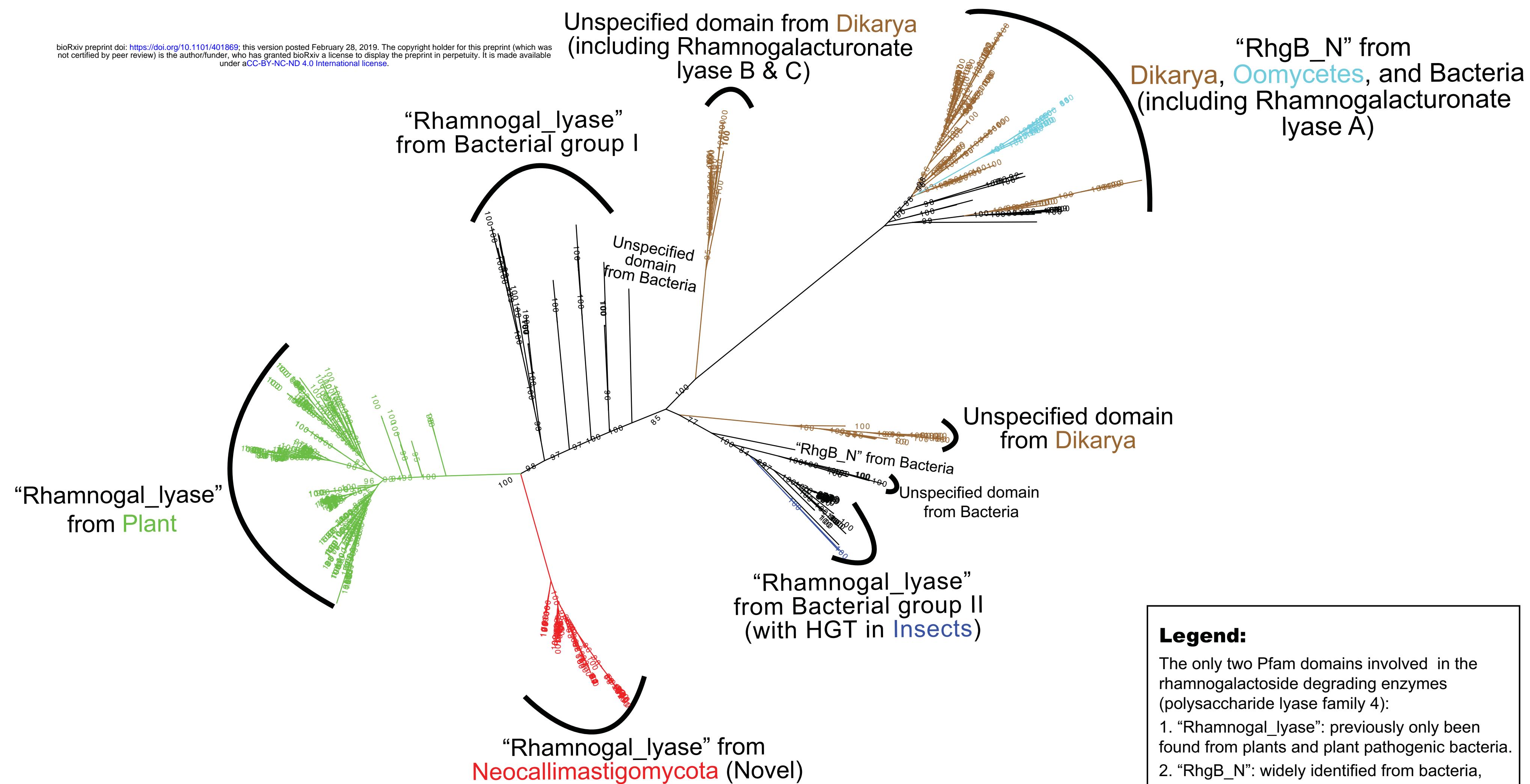
bioRxiv preprint doi: <https://doi.org/10.1101/301869>; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Legend

- Cthe_2159
- Incomplete Cthe_2159
- CBM_10
- Pro-rich
- rne
- elf-3_zeta
- TPH
- PTZ00341
- PT
- Nop14
- Mak16
- hsdR
- Herpes_BLLF1
- DUF390
- Atrophin-1
- PTZ00110
- K+-dependent Na+/Ca+ exchanger





Legend:

The only two Pfam domains involved in the rhamnogalactoside degrading enzymes (polysaccharide lyase family 4):

1. "Rhamnogal_lyase": previously only been found from plants and plant pathogenic bacteria.
2. "RhgB_N": widely identified from bacteria, fungi, and oomycetes.

Table 1. Information for the AGF strains included in this study.

Genus	Species	Strain	Accession No.	Type	Host
<i>Anaeromyces</i>	<i>Anaeromyces contortous</i>	Na	GGWN00000000	Transcriptome	Cow
<i>Anaeromyces</i>	<i>Anaeromyces contortous</i>	C3J	GGWO00000000	Transcriptome	Cow
<i>Anaeromyces</i>	<i>Anaeromyces contortous</i>	G3G	GGWR00000000	Transcriptome	Goat
<i>Anaeromyces</i>	<i>Anaeromyces contortous</i>	O2	GGWQ00000000	Transcriptome	Cow
<i>Anaeromyces</i>	<i>Anaeromyces contortous</i>	C3G	GGWR00000000	Transcriptome	Cow
<i>Anaeromyces</i>	<i>Anaeromyces robustus</i>	S4	MCFG00000000.1	Genome	Sheep
<i>Caecomyces</i>	<i>Caecomyces</i> sp.	Brit4	GGWS00000000	Transcriptome	Cow
<i>Caecomyces</i>	<i>Caecomyces</i> sp.	Iso3	GGXE00000000	Transcriptome	Cow
<i>Feramyces</i>	<i>Feramyces austini</i>	WSF3a	GGWU00000000	Transcriptome	Aoudad
<i>Feramyces</i>	<i>Feramyces austini</i>	WSF2c	GGWT00000000	Transcriptome	Aoudad
<i>Orpinomyces</i>	<i>Orpinomyces</i> sp.	D3A	GGWV00000000	Transcriptome	Cow
<i>Orpinomyces</i>	<i>Orpinomyces</i> sp.	D3B	GGWW00000000	Transcriptome	Cow
<i>Orpinomyces</i>	<i>Orpinomyces</i> sp.	D4C	GGWX00000000	Transcriptome	Cow
<i>Pecoramycetes</i>	<i>Pecoramycetes ruminantium</i>	C1A	ASRE00000000.1	Genome	Cow
<i>Pecoramycetes</i>	<i>Piromyces</i> sp.	S4B	GGWY00000000	Transcriptome	Sheep
<i>Pecoramycetes</i>	<i>Piromyces</i> sp.	FX4B	GGWZ00000000	Transcriptome	Cow
<i>Pecoramycetes</i>	<i>Piromyces</i> sp.	FS3c	GGXF00000000	Transcriptome	Cow
<i>Pecoramycetes</i>	<i>Piromyces</i> sp.	YC3	GGXA00000000	Transcriptome	Cow
<i>Piromyces</i>	<i>Piromyces finnis</i>	Pirfi3	MCFH00000000.1	Genome	Horse
<i>Piromyces</i>	<i>Piromyces</i> sp.	E2	MCNC00000000.1	Genome	Elephant
<i>Piromyces</i>	<i>Piromyces</i> sp.	A1	GGXB00000000	Transcriptome	Sheep
<i>Piromyces</i>	<i>Piromyces</i> sp.	B4	GGXH00000000	Transcriptome	Cow
<i>Piromyces</i>	<i>Piromyces</i> sp.	B5	GGXI00000000	Transcriptome	Cow
<i>Neocallimastix</i>	<i>Neocallimastix californiae</i>	G1	MCOG00000000.1	Genome	Goat
<i>Neocallimastix</i>	<i>Neocallimastix frontalis</i>	Hef5	GGXJ00000000	Transcriptome	Cow
<i>Neocallimastix</i>	<i>Neocallimastix</i> sp.	G3	GGXC00000000	Transcriptome	Sheep

Table 2. Distribution of the three studied domains in the fungal kingdom.

	Number of examined Genomes	No. "Cthe_2159"	No. "Gal_Lectin"	No. "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
Blastocladiomycota	4	0	0	0
Cryptomycota	1	0	0	0
Microsporidia	22	0	0	0
Neocallimastigomycota	5	95	67	26
Total	1121			