Horizontal gene transfer as an indispensible driver for Neocallimastigomycota evolution into a distinct gut-dwelling fungal lineage

Chelsea L. Murphy[1¶], Noha H. Youssef[1¶], Radwa A. Hanafy[1], MB Couger[2], Jason E. Stajich[3], Y. Wang[3], Kristina Baker[1], Sumit S. Dagar[4], Gareth W. Griffith[5], Ibrahim F. Farag[1], TM Callaghan[6], and Mostafa S. Elshahed[1*]

[1]Department of Microbiology and Molecular Genetics and [2] High Performance Computing Center, Oklahoma State University, Stillwater, OK. [3]Department of Microbiology and Plant Pathology, Institute for Integrative Genome Biology, University of California-Riverside, Riverside, CA. Bioenergy group, Agharkar Research Institute, Pune, India. [5]Institute of Biological, Environmental, and Rural Sciences (IBERS) Aberystwyth University, Aberystwyth, Wales, UK. [6]Department for Quality Assurance and Analytics, Bavarian State Research Center for Agriculture, Freising, Germany.

Running Title: HGT in the Neocallimastigomycota

[*]Corresponding author: Mailing address: Oklahoma State University, Department of Microbiology and Molecular Genetics, 1110 S Innovation Way, Stillwater, OK 74074. Phone: (405) 744-3005, Fax: (405) 744-1112. Email: **Mostafa@okstate.edu**.

[¶] Both authors contributed equally to this work.

Abstract

Survival and growth of the anaerobic gut fungi (AGF, Neocallimastigomycota) in the herbivorous gut necessitate the possession of multiple abilities absent in other fungal lineages. We hypothesized that horizontal gene transfer (HGT) was instrumental in forging the evolution of AGF into a phylogenetically distinct gut-dwelling fungal lineage. Patterns of HGT were evaluated in the transcriptomes of 27 AGF strains, 22 of which were isolated and sequenced in this study, and 4 AGF genomes broadly covering the breadth of AGF diversity. We identified 327 distinct incidents of HGT in AGF transcriptomes, with subsequent gene duplication resulting in an HGT frequency of 2.9-4.1% in AGF genomes. The majority of HGT events were AGF specific (90.8%) and wide (67.3%), indicating their occurrence at early stages of AGF evolution. The acquired genes allowed AGF to expand their substrate utilization range, provided new venues for electron disposal, augmented their biosynthetic capabilities, and facilitated their adaptation to anaerobiosis. The majority of donors were anaerobic fermentative bacteria prevalent in the herbivorous gut. In addition, acquisition incidents from marine invertebrates provide interesting clues to the habitat of AGF ancestors prior to terrestrialization. This work strongly indicates that HGT indispensably forged the evolution of AGF as a distinct fungal phylum and provides a unique example of the role of HGT in shaping the evolution of a high rank taxonomic eukaryotic lineage.

Introduction

Horizontal gene transfer (HGT) is defined as the acquisition, integration, and retention of foreign genetic material into a recipient organism (Doolittle 1999). HGT represents a relatively rapid process for trait acquisition; as opposed to gene creation either from preexisting genes (via duplication, fission, fusion, or exon shuffling) or through *de-novo* gene birth from non-coding sequences (Andersson et al 2015, Carvunis et al 2010, Innan and Kondrashov 2010, Cai, 2008, Kaessmann 2010). In prokaryotes, the occurrence, patterns, frequency, and impact of HGT on the genomic architecture (Ochman et al 2000), metabolic abilities (Caro-Quintero and Konstantinidis 2015, Youssef et al 2015), physiological preferences (Omelchenko et al 2005, Puigbo et al 2008), and ecological fitness (Wiedenbeck and Cohan 2011) has been widely investigated, and the process is now regarded as a major driver of genome evolution in bacteria and archaea (Philippe and Douady 2003, Syvanen 2012). Although eukaryotes are perceived to evolve principally through modifying existing genetic information, analysis of HGT events in eukaryotic genomes has been eliciting increasing interest and scrutiny. In spite of additional barriers that need to be overcome in eukaryotes, e.g. crossing the nuclear membrane, germ line sequestration in sexual multicellular eukaryotes, and epigenetic nucleic acids modifications mechanisms (Andersson et al 2015, Fitzpatrick 2012), it is now widely accepted that HGT contributes significantly to eukaryotic genome evolution (Husnik and McCutcheon 2017, Keeling and Palmer 2008). HGT events have convincingly been documented in multiple phylogenetically disparate eukaryotes ranging from the Excavata (Eichinger et al 2005, Hirt et al 2002, Nixon et al 2002, Qian and Keeling 2001), SAR supergroup (Eme et al 2017, Kishore et al 2013, Ricard et al 2006, Wisecaver et al 2013), Algae (Schönknecht et al 2013b), Plants (Richardson and Palmer 2007), and Opisthokonta (Gladyshev et al 2008, Danchin, 2010 #72,

64 Marcet-Bouben and Gabaldon 2010, Sun et al 2010). Reported HGT frequency in eukaryotic
65 genomes ranges from a handful of genes, e.g. (McCarthy and Fitzpatrick 2016), to up to 9.6% in
66 Bdelloid rotifers (Gladyshev et al 2008).

67 The kingdom Fungi represents a phylogenetically coherent clade that evolved ≈ 900-1481
68 Mya from a unicellular flagellated ancestor (Douzery et al 2004, Parfrey et al 2011, Taylor and
69 Berbee 2006). To date, multiple efforts have been reported on the detection and quantification of
70 HGT in fungi. A survey of 60 fungal genomes reported HGT frequencies of 0-0.38% (Marcet-
71 Bouben and Gabaldon 2010), and similar low values were observed in the genomes of five early-
72 diverging pathogenic Microsporidia and Cryptomycota (Alexander et al 2016b). As such, the
73 prevailing consensus is that HGT events in fungal genomes are infrequent and sporadic
74 (Fitzpatrick 2012, Schönknecht et al 2013b). This has been attributed to the osmotrophic lifestyle
75 of fungi (Berbee et al 2017), which is less conducive to HGT compared to the phagocytic
76 lifestyle of several microeukaryotes with relatively higher HGT frequency (Doolittle 1998).

77 The anaerobic gut fungi (AGF, Neocallimastigomycota) represent a phylogenetically
78 distinct basal fungal lineage. The AGF appear to exhibit a restricted distribution pattern, being
79 encountered in the gut of ruminant and non-ruminant herbivorous (Gruninger et al 2014). In the
80 herbivorous gut, the life cycle of the AGF (Figure S1) involves the discharge of motile
81 flagellated zoospores from sporangia in response to animal feeding, the chemotaxis and
82 attachment of zoospores to ingested plant material, spore encystment, and the subsequent
83 production of rhizoidal growth that penetrates and digests plant biomass through the production
84 of a wide array of cellulolytic and lignocellulolytic enzymes.

85 Survival, colonization, and successful propagation of AGF in the herbivorous gut
86 necessitate the acquisition of multiple unique physiological characteristics and metabolic abilities

absent in other fungal lineages. These include, but are not limited to, development of a robust plant biomass degradation machinery, adaptation to anaerobiosis, and exclusive dependence on fermentation for energy generation and recycling of electron carriers (Boxma et al 2004, Youssef et al 2013). Therefore, we hypothesized that sequestration into the herbivorous gut was conducive to the broad adoption of HGT as a relatively faster adaptive evolutionary strategy for niche adaptation by the AGF (Figure S1). Further, since no part of the AGF life cycle occurs outside the animal host and no reservoir of AGF outside the herbivorous gut has been identified (Gruninger et al 2014), then acquisition would mainly occur from donors that are prevalent in the herbivorous gut (Figure S1). Apart from earlier observations on the putative bacterial origin of a few catabolic genes in two AGF isolates (Garcia-Vallvé et al 2000, Harhangi et al 2003), and preliminary BLAST-based queries of a few genomes (Haitjema et al 2017, Youssef et al 2013), little is currently known on the patterns, determinants, and frequency of HGT in the Neocallimastigomycota. To address this hypothesis, we systematically evaluated the patterns of HGT acquisition in the transcriptomes of 27 AGF strains and 4 AGF genomes broadly covering the breadth of AGF genus-level diversity. Our results document the high level of HGT in AGF in contrast to HGT paucity across the fungal kingdom. The identity of genes transferred, distribution pattern of events across AGF genera, phylogenetic affiliation of donors, and the expansion of acquired genetic material in AGF genomes highlight the role played by HGT in forging the evolution and diversification of the Neocallimastigomycota as a phylogenetically, metabolically, and ecologically distinct lineage in the fungal kingdom.

Materials and Methods

Organisms. Type strains of the Neocallimastigomycota are unavailable through culture collections due to their strict anaerobic and fastidious nature, as well as the frequent occurrence of senescence in AGF strains (Ho and Barr 1995). As such, obtaining a broad representation of the Neocallimastigomycota necessitated the isolation of representatives of various AGF genera *de novo*. Samples were obtained from the feces, rumen, or digesta of domesticated and wild herbivores around the city of Stillwater, OK and Val Verde County, Texas (Table 1). Isolation procedures are explained in detail in the Supplementary text.

Sequencing and assembly. Transcriptomic sequencing was conducted for 22 AGF strains. Sequencing multiple taxa provides stronger evidence for the occurrence of HGT in a target lineage (Richards and Monier 2017), and allows for the identification of phylum-wide versus genus- and species-specific HGT events. Transcriptomic, rather than genomic, sequencing was chosen for AGF-wide HGT identification efforts since enrichment for polyadenylated (poly(A)) transcripts prior to RNA-seq provides an additional safeguard against possible prokaryotic contamination, an issue that often plagued eukaryotic genome-based HGT detection efforts (Boothby et al 2015, Koutsovoulos et al 2016), as well as to demonstrate that HGT genes identified are transcribed in AGF. Protocols for RNA extraction, transcriptomic sequencing and assembly, and peptide model prediction are described in detail in the supplementary text.

HGT identification. A combination of BLAST similarity searches, comparative similarity index (HGT index, h_U), phylogenetic analyses, and parametric gene composition approaches were conducted to identify HGT events in the analyzed transcriptomic datasets (Fig. 1). We define an HGT event as the acquisition of a foreign gene/pfam by AGF from a single lineage/donor. Details on the criteria used for identification of HGT events are described in the supplementary

text. The GC content, and intron distribution were assessed in all identified events and compared to averages of an equal number of randomly chosen genes from AGF genomes using Student t-test to identify possible deviations in such characteristics as often observed with HGT genes (Soucy et al 2015). As a control, the frequency of HGT occurrence in the genomes of a filamentous ascomycete (*Colletotrichum graminicola*, GenBank Assembly accession number GCA_000149035.1), and a microsporidian (*Encephalitozoon hellem*, GenBank Assembly accession number GCA_000277815.3) were determined using our pipeline (Table S1); and the results were compared to previously published results (Alexander et al 2016a, Jaramillo et al 2015).

Identification of HGT events in carbohydrate active enzymes (CAZymes) transcripts. In AGF genomes, carbohydrate active enzymes (CAZymes) are often encoded by large multi-module genes with multiple adjacent CAZyme or non-CAZyme domains. A single gene can hence harbor multiple CAZyme pfams of different (fungal or non-fungal) origins (Haitjema et al 2017, Youssef et al 2013). As such, our initial efforts for HGT assessment in CAZyme-encoding transcripts using an entire gene/ transcript strategy yielded inaccurate results since similarity searches only identified pfams with the lowest e-value or highest number of copies, while overlooking additional CAZyme pfams in the transcript (Figure S2). To circumvent the multi-modular nature of AGF CAZyme transripts, we opted for the identification of CAZyme HGT events on trimmed domains, rather than entire transcript. Details on the identification of CAZyme-containing transcripts and criteria used for detection of CAZyme HGT events are explained in the supplementary text.

Neocallimastigomycota-specific versus non-specific HGT events. For all HGT events identified in the Neocallimatigomycota, orthologues (30% identity, >100 amino acids alignment)

were extracted from the genomes of other basal fungi, i.e. members of Blastocladiales, Chytridiomycota, Cryptomycota, Microsporidia, Mucoromycota, and Zoopagomycota, and the phylogenetic affiliation of these orthologues was assessed. An HGT event was judged to be Neocallimastigomycota-specific if: 1. orthologues were absent in all basal fungal genomes, 2. orthologues were identified in basal fungal genomes, but these orthologues were of clear fungal origin or displayed an affiliation different from that observed in the Neocallimastigomycota. On the other hand, events were judged to be non-specific to the Neocallimastigomycota if phylogenetic analysis of basal fungal orthologues indicated a non-fungal origin with a donor affiliation similar to that observed in the Neocallimastigomycota (Figure 1).

Data accession. Sequences of individual transcripts identified as horizontally transferred are deposited in GenBank under the accession number MH043627-MH043936, and MH044722-MH044724. The whole transcriptome shotgun sequences were deposited in GenBank under the BioProject PRJNA489922, and Biosample accession numbers SAMN09994575-SAMN09994596. Transcriptomic assemblies were deposited in the SRA under project accession number SRP161496. Alignments, as well as Newick tree files for all HGT genes are provided as a supplementary file (Supp. files 1 and 2). Trees of specific HGT events discussed in the results and discussion sections are presented in the supplementary document (S5-S56).

Results

Isolates. The transcriptomes of 22 different isolates were sequenced. These isolates belonged to six out of the nine currently described AGF genera: *Anaeromyces* (n=5), *Caecomyces* (n=2), *Neocallimastix* (n=2), *Orpinomyces* (n=3), *Pecoramyces* (n=4), *Piromyces* (n=4), as well as the recently proposed genus *Feramyces* (n=2) (Hanafy et al 2018) (Table 1, Supplementary Figure 3). Out of the three AGF genera not included in this analysis, two are currently represented by a single strain that was either lost (genus *Oontomyces* (Dagar et al 2015)), or appears to exhibit an extremely limited geographic and animal host distribution (genus *Buwchfawromyces* (Callaghan et al 2015)). The third unrepresented genus (*Cyllamyces*) has recently been suggested to be phylogenetically synonymous with *Caecomyces* (Wang et al 2017). As such, the current collection is a broad representation of currently described AGF genera.

Sequencing. Transcriptomic sequencing yielded 15.2-110.8 million reads (average, 40.87) that were assembled into 31,021-178,809 total transcripts, 17,539-132,141 distinct transcripts (clustering at 95%), and 16,500-70,061 predicted peptides (average 31,611) (Table S2). Assessment of transcriptome coverage using BUSCO (Simão et al 2015) yielded high completion values (82.76-97.24%) for all assemblies (Table S2). For strains with a sequenced genome (*Pecoramyces ruminantium*, *Piromyces finnis*, *Piromyces* sp. E2, *Anaeromyces robustus*, and *Neocallimastix californiae*), genome coverage (percentage of genes in a strain's genome for which a transcript was identified) ranged between 70.9-91.4% (Table S2).

HGT events. A total of 327 distinct HGT events were identified in the Neocallimastigomycota pantranscriptome analyzed (Table S3). The average number of events per genus was 251±16 and ranged between 232 in the genus *Caecomyces* to 276 in the genus *Pecoramyces* pantranscriptomes (Fig. 2A). The majority of HGT acquisition events identified (297, 90.83%)

193 appear to be Neocallimastigomycota-specific, i.e. identified only in genomes belonging to the
194 Neocallimastigomycota, but not in other basal fungal genomes (Table S3), strongly suggesting
195 that such acquisitions occurred post, or concurrent with, the evolution of Neocallimastigomycota
196 as a distinct fungal lineage. As well, the majority of these identified genes were
197 Neocallimastigomycota-wide, being identified in strains belonging to at least six out of the seven
198 examined genera (220 events, 67.3%), suggesting the acquisition of such genes prior to genus
199 level diversification within the Neocallimastigomycota. Only 47 events (14.4%) were genus-
200 specific, with the remainder (60 events, 18.3%) being identified in the transcriptomes of 3-5
201 genera (Table S3, Figure S4, and Fig. 2b).
202 The absolute majority (89.9%) of events were successfully mapped to at least one of the
203 four AGF genomes (Table S4), with a fraction (8/33) of the unmapped transcripts being specific
204 to a genus with no genome representative (*Feramyces*, *Caecomyces*). Compared to a random
205 subset of 327 genes in each of the sequenced genomes, horizontally transferred genes in AGF
206 genomes exhibited significantly ($P<0.0001$) fewer introns (1.55±3.67 vs 3.32±0.83), as well as
207 higher GC content (30.94±4.6 vs 27.7±5.5) (Table S4). Further, HGT genes/pfams often
208 displayed high levels of gene/ pfam duplication and expansion within the genome (Table S4),
209 resulting in an HGT frequency of 2.88% in *Pecoramyces ruminantium* (470 HGT genes out of
210 16,347 total genes), 3.74% in *Piromyces finnis* (429 HGT genes out of 11,477 total genes),
211 4.00% in *Anaeromyces robustus* (517 HGT genes out of 12,939 total genes), and 4.13% in
212 *Neocallimastix californiae* (864 HGT genes out of 20,939 total genes).
213 **Donors.** A bacterial origin was identified for the majority of HGT events (82.6%), with four
214 bacterial phyla (Firmicutes, Proteobacteria, Bacteroidetes, and Spirochaetes) identified as donors
215 for 203 events (62.1% of total, 75.2% of bacterial events) (Fig. 3A). Specifically, the

contribution of members of the Firmicutes (142 events) was paramount, the majority of which were most closely affiliated with members of the order Clostridiales (119 events). In addition, minor contributions from a wide range of bacterial phyla were also identified (Fig. 3A). The majority of the putative donor taxa are strict/ facultative anaerobes, and many of which are also known to be major inhabitants of the herbivorous gut and often possess polysaccharide-degradation capabilities (He et al 2018, Stewart et al 2018). Archaeal contributions to HGT were extremely rare (5 events). On the other hand, multiple (50) events with eukaryotic donors were identified. Remarkably, eukaryotic marine lineages, e.g. Cnidaria (stony corals and sea anemonae), Arthropoda (crustaceans and horse shoe crabs), Mollusca (Oysters and Scallops), Osteichthyes (bony fish), Brachiopoda, Echinodermata (sea urchins, sea cucumber, and starfish), Porifera (sponges), and *Trichoplax* (the only extant member of the Placozoa; known to inhabit marine environments especially on substrates such as stony corals and mollusk shells (Pearse and Voigt 2007)) contributed 10 out of the 50 eukaryotic HGT events, despite their physical separation from the current AGF habitat (Table S3, supplementary document). In few instances, a clear non-fungal origin was identified for a specific event, but the precise inference of the donor based on phylogenetic analysis was not feasible (Table S3).

Metabolic characterization. Functional annotation of HGT genes/pfams indicated that the majority (61.8%) of events encode metabolic functions such as extracellular polysaccharide degradation and central metabolic processes. Bacterial donors were slightly overrepresented in metabolic HGT events (88.1% of the metabolism-related events, compared to 82.6% of the total events). Genes involved in cellular processes and signaling represent the second most represented HGT events (11%), while genes involved in information storage and processing only

made up 6.12% of the HGT events identified (Figs 3b-e). Below we present a detailed description of the putative abilities and functions enabled by HGT transfer events.

Central catabolic abilities. Multiple HGT events encoding various central catabolic processes were identified in AGF transcriptomes and successfully mapped to the genomes (Fig. 4, Table S3, Figs S5-S17). A group of events appear to encode enzymes that allow AGF to channel specific substrates into central metabolic pathways. For example, genes encoding enzymes of the Leloir pathway for galactose conversion to glucose-1-phosphate (galactose-1-epimerase, galactokinase (Fig. 5A), galactose-1-phosphate uridylyltransferase, and UDP-glucose-4-epimerase) were identified, in addition to genes encoding ribokinase, as well as xylose isomerase and xylulokinase for ribose and xylose channeling into the pentose phosphate pathway. As well, genes encoding deoxyribose-phosphate aldolase (DeoC) enabling the utilization of purines as carbon and energy sources were also horizontally acquired in AGF. Further, several of the glycolysis/gluconeogenesis genes, e.g. phosphoenolpyruvate synthase, as well as phosphoglycerate mutase were also of bacterial origin. Fungal homologues of these glycolysis/gluconeogenesis genes were not identified in the AGF transcriptomes, suggesting the occurrence of xenologous replacement HGT events.

In addition to broadening substrate range, HGT acquisitions provided additional venues for recycling reduced electron carriers via new fermentative pathways in this strictly anaerobic and fermentative lineage. The production of ethanol, D-lactate, and hydrogen appears to be enabled by HGT (Fig. 4). The acquisition of several aldehyde/alcohol dehydrogenases, and of D-Lactate dehydrogenase for ethanol and lactate production from pyruvate was identified. Although these two enzymes are encoded in other fungi as part of their fermentative capacity (e.g. *Saccharomyces* and *Schizosaccharomyces*), no homologues of these fungal genes were

identified in AGF pantranscriptomes. Hydrogen production in AGF, as well as in many anaerobic eukaryotes with mitochondria-related organelles, involves pyruvate decarboxylation to acetyl CoA, followed by the use of electrons generated for hydrogen formation via an anaerobic Fe-Fe hydrogenase. In AGF, while pyruvate decarboxylation to acetyl CoA via pyruvate-formate lyase and the subsequent production of acetate via acetyl-CoA:succinyl transferase appear to be of fungal origin, the Fe-Fe hydrogenase and its entire maturation machinery (HydEFG) seem to be horizontally transferred being phylogenetically affiliated with similar enzymes in Thermotogae, Clostridiales, and the anaerobic jakobid excavate, *Stygiella incarcerate* (Fig. 5B). It has recently been suggested that *Stygiella* acquired the Fe-Fe hydrogenase and its maturation machinery from bacterial donors including Thermotogae, Firmicutes, and Spirochaetes (Leger et al 2016), suggesting either a single early acquisition event in eukaryotes, or alternatively independent events for the same group of gene have occurred in different eukaryotes.

Anabolic capabilities. Multiple anabolic genes that expanded AGF biosynthetic capacities appear to be horizontally transferred (Fig. S18-S31). These include several amino acid biosynthesis genes e.g. cysteine biosynthesis from serine; glycine and threonine interconversion; and asparagine synthesis from aspartate. In addition, horizontal gene transfer allowed AGF to de-novo synthesize NAD via the bacterial pathway (starting from aspartate via L-aspartate oxidase (NadB; Fig. 5C) and quinolinate synthase (NadA) rather than the five-enzymes fungal pathway starting from tryptophan (Lin et al 2010)). HGT also allowed AGF to salvage thiamine via the acquisition of phosphomethylpyrimidine kinase. Additionally, several genes encoding enzymes in purine and pyrimidine biosynthesis were horizontally transferred (Fig. 4). Finally, horizontal gene transfer allowed AGF to synthesize phosphatidyl-serine from CDP-diacylglycerol, and to convert phosphatidyl-ethanolamine to phosphatidyl-choline.

284 ***Adaptation to the host environment.*** Horizontal gene transfer also appears to have provided
285 means of guarding against toxic levels of compounds known to occur in the host animal gut (Fig.
286 S32-S38). For example, methylglyoxal, a reactive electrophilic species (Lee and Park 2017), is
287 inevitably produced by ruminal bacteria from dihydroxyacetone phosphate when experiencing
288 growth conditions with excess sugar and limiting nitrogen (Russell 1993). Genes encoding
289 enzymes mediating methylglyoxal conversion to D-lactate (glyoxalase I and glyoxalase II-
290 encoding genes) appear to be acquired via HGT in AGF. Further, HGT allowed several means of
291 adaptation to anaerobiosis. These include: 1) acquisition of the oxygen-sensitive ribonucleoside-
292 triphosphate reductase class III (Fig. 5D) that is known to only function during anaerobiosis to
293 convert ribonucleotides to deoxyribonucleotides (Jordan and Reichard 1998), 2) acquisition of
294 squalene-hopene cyclase, which catalyzes the cyclization of squalene into hopene, an essential
295 step in biosynthesis of the cell membrane steroid tetrahymanol that replaced the molecular O_2-
296 requiring ergosterol in the cell membranes of AGF, 3) acquisition of several enzymes in the
297 oxidative stress machinery including Fe/Mn superoxide dismutase, glutathione peroxidase,
298 rubredoxin/rubrerythrin, and alkylhydroperoxidase.

299 In addition to anaerobiosis, multiple horizontally transferred general stress and repair
300 enzymes were identified (Fig S39-S46). HGT-acquired genes encoding 2-phosphoglycolate
301 phosphatase, known to metabolize the 2-phosphoglycolate produced in the repair of DNA lesions
302 induced by oxidative stress (Pellicer et al 2003) to glycolate, were identified in all AGF
303 transcriptomes studied (Fig. 4, Table S3). Surprisingly, two genes encoding antibiotic resistance
304 enzymes, chloramphenicol acetyltransferase and aminoglycoside phosphotransferase, were
305 identified in all AGF transcriptomes, presumably to improve its fitness in the eutrophic rumen
306 habitat that harbors antibiotic-producing prokaryotes (Table S3). While unusual for eukaryotes to

express antibiotic resistance genes, basal fungi such as *Allomyces, Batrachochytrium,* and *Blastocladiella* were shown to be susceptible to chloramphenicol and streptomycin (Bishop et al 2009, Rooke and Shattock 1983). Other horizontally transferred repair enzymes include DNA-3-methyladenine glycosylase I, methylated-DNA--protein-cysteine methyltransferase, galactoside and maltose O-acetyltransferase, and methionine-R-sulfoxide reductase (Table S3).

HGT transfer in AGF carbohydrate active enzymes machinery. Within the analyzed AGF transcriptomes, CAZymes belonging to 39 glycoside hydrolase (GHs), 5 polysaccharide lyase (PLs), and 10 carbohydrate esterase (CEs) families were identified (Fig. 6). The composition of the CAZyomes of various AGF strains examined were broadly similar, with the following ten notable exceptions: Presence of GH24 and GH78 transcripts only in *Anaeromyces* and *Orpinomyces*, the presence of GH28 transcripts only in *Pecoramyces*, *Neocallimastix*, and *Orpinomyces*, the presence of GH30 transcripts only in *Anaeromyces*, and *Neocallimastix*, the presence of GH36 and GH95 transcripts only in *Anaeromyces*, *Neocallimastix*, and *Orpinomyces*, the presence of GH97 transcripts only in *Neocallimastix*, and *Feramyces*, the presence of GH108 transcripts only in *Neocallimastix*, and *Piromyces*, and the presence of GH37 predominantly in *Neocallimastix*, GH57 transcripts predominantly in *Orpinomyces*, GH76 transcripts predominantly in *Feramyces*, and CE7 transcripts predominantly in *Anaeromyces* (Fig. 6).

HGT appears to be rampant in the AGF pan-CAZyome: A total of 90 events (27.5% of total HGT events) were identified, with the majority occurring in all AGF genera examined (Fig. 6, Table S3). In 48% of GH families, 55% of CE families, and 20% of PL families, a single event (i.e. attributed to one donor) was observed (Fig. 6, Table S3).

Duplication of these events in AGF genomes was notable, with 152, 373, 201, and 191 copies of

HGT CAZyme pfams identified in *Anaeromyces*, *Neocallimastix*, *Piromyces* and *Pecoramyces* genomes, representing 40.7%, 45.3%, 52.9%, and 41.2% of the overall organismal CAZyme machinery (Table S4). The contribution of Viridiplantae, Flavobacteriales, Fibrobacteres, and Gamma-Proteobacteria was either exclusive to CAZyme-related HGT events or significantly higher in CAZyme, compared to other, events (Fig. 3A).

Transcripts acquired by HGT represented >50% of transcripts in anywhere between 13 (*Caecomyces*) to 20 (*Anaeromyces*) GH families; 3 (*Caecomyces*) to 5 (*Anaeromyces*, *Neocallimastix*, *Orpinomyces*, and *Feramyces*) CE families; and 2 (*Caecomyces* and *Feramyces*) to 3 (*Anaeromyces*, *Pecoramyces*, *Piromyces*, *Neocallimastix*, and *Orpinomyces*) PL families (Fig. 6). It is important to note that in all these families, multiple transcripts appeared to be of bacterial origin based on BLAST similarity search but did not meet the strict criteria of h_U >30, and so were deemed not horizontally transferred. As such, the contribution of HGT transcripts to overall transcripts in these families is probably an underestimate. Only GH9, GH20, GH37, GH45, and PL3 families appear to lack any detectable HGT events. A PCA biplot comparing CAZyome in AGF genomes to other basal fungal lineages strongly suggests that the acquisition and expansion of many of these foreign genes play an important role in shaping the lignocellulolytic machinery of AGF (Fig. 7). The majority of CAZyme families defining AGF CAZyome were predominantly of non-fungal origin (Fig. 7). This pattern clearly attests to the value of HGT in shaping AGF CAZyome via acquisition and extensive duplication of acquired gene families.

Collectively, HGT had a profound impact on AGF plant biomass degradation capabilities. The AGF CAZyome encodes enzymes putatively mediating the degradation of twelve different polysaccharides (Fig. S57). In all instances, GH and PL families with >50% horizontally

353 transferred transcripts contributed to backbone cleavage of these polymers; although in many
354 polymers, e.g. cellulose, glucoarabinoxylan, and rhamnogalactouronan, multiple different GHs
355 can contribute to backbone cleavage. Similarly, GH, CE, and PL families with >50%
356 horizontally transferred transcripts contributed to 10 out of 13 side-chain-cleaving activities, and
357 3 out of 5 oligomer-to-monomer breakdown activities (Fig. S57).

Discussion

Here, we present a systematic analysis of HGT patterns in 27 transcriptomes and 4 genomes belonging to the Neocallimastigomycota. Our analysis identified 327 events, representing 2.9-4.1% of genes in examined AGF genomes. Further, we consider these values to be conservative estimates due to the highly stringent criteria employed. Only events with h_U of >30 were considered, and all putative events were further subjected to manual inspection and phylogenetic tree construction to confirm incongruence with organismal evolution and bootstrap-supported affiliation to donor lineages. Further, events identified in less than 50% of strains in a specific genus were excluded, and parametric gene composition approaches were implemented in conjunction with sequence-based analysis. Nevertheless, the observed HGT frequency in this study is in contrast to the reported paucity in HGT events across the fungal kingdom (Alexander et al 2016b, Marcet-Bouben and Gabaldon 2010), and hence the Neocallimastigomycota represent a notable exception within the Mycota. Multiple factors could be postulated to account for the observed high HGT frequency in AGF. The sequestration of AGF into the anaerobic, prokaryotes-dominated herbivorous gut necessitated the implementation of the relatively faster adaptive mechanisms for survival in this new environment, as opposed to the slower strategies of neofunctionalization and gene birth. Indeed, niche adaptation and habitat diversification events are widely considered important drivers for HGT in eukaryotes (de Koning et al 2000, Keeling and Palmer 2008, Ricard et al 2006, Schönknecht et al 2013a). Further, AGF are constantly exposed to a rich milieu of cells and degraded DNA in the herbivorous gut. Such close physical proximity between donors/ extracellular DNA and recipients is also known to greatly facilitate HGT (Beiko et al 2005, Moliner et al 2010, Shterzer and Mizrahi 2015). Finally, AGF release asexual motile free zoospores into the herbivorous gut as part of their life cycle (Gruninger et al

2014). According to the weak-link model (Huang 2013), these weakly protected and exposed structures provide excellent entry point of foreign DNA to eukaryotic genomes. It is important to note that AGF zoospores also appear to be naturally competent, capable of readily uptaking nucleic acids from their surrounding environment (Calkins et al 2016).

The distribution of HGT events across various AGF taxa (Fig. 2), identities of HGT donors (Fig. 3), and abilities imparted (Figs. 4-5) could offer important clues regarding the timing and impact of HGT on Neocallimastigomycota evolution. The majority of events (67.3%) were Neocallimastigomycota-wide and were mostly acquired from lineages known to inhabit the herbivorous gut, e.g. Firmicutes, Proteobacteria, Bacteroidetes, and Spirochaetes (Figs. 2-3). This pattern strongly suggests that such acquisitions occurred post (or concurrent with) AGF sequestration into the herbivorous gut, but prior to AGF genus level diversification. Many of the functions encoded by these events represented novel functional acquisitions that impart new abilities, e.g. galactose metabolism, methyl glyoxal detoxification, pyruvate fermentation to d-lactate and ethanol, and chloramphenicol resistance (Fig. 3). Others represented acquisition of novel genes or pfams augmenting existing capabilities within the AGF genomes, e.g. acquisition of GH5 cellulases to augment the fungal GH45, acquisition of additional GH1 and GH3 beta gluco- and galactosidases to augment similar enzymes of apparent fungal origin in AGF genomes (Fig. 6-7, Fig. S47). Novel functional acquisition events enabled AGF to survive and colonize the herbivorous gut by: 1. Expanding substrate-degradation capabilities (Fig. 5a, 6, 7, S5-S17, Table S3), hence improving fitness by maximizing carbon and energy acquisition from available plant substrates, 2. Providing additional venues for electron disposal via lactate, ethanol, and hydrogen production, and 3. Enabling adaptation to anaerobiosis (Fig. 4, S32-S38, Table S3).

A smaller number of observed events (n=47) were genus-specific (Fig. 2, Table S3). This group was characterized by being significantly enriched in CAZymes (53.2% of genus-specific horizontally transferred events have a predicted CAZyme function, as opposed to 27.5% in the overall HGT dataset), and being almost exclusively acquired from donors that are known to inhabit the herbivorous gut (Creevey et al 2014) (35 out of the 47 events were acquired from the orders Clostridiales, Bacillales, Lactobacillales and Negativicutes within Firmicutes, Burkholderiales, and Vibrionales within the Beta- and Gamma-Proteobacteria, Flavobacteriales and Bacteroidales within Bacteroidetes, and the Spirochaetes, Actinobacteria, and Lentisphaerae), or from Viridiplantae (5 out of the 47 events). Such pattern suggests the occurrence of these events relatively recently, in the herbivorous gut post AGF genus level diversification. We reason that the lower frequency of such events is a reflection of the relaxed pressure for acquisition and retention of foreign genes at this stage of AGF evolution.

Finally, few Neocallimastiogomycota-wide HGT events were characterized by donors that are not typically encountered in the herbivorous gut (Comtet-Marre et al 2017, Creevey et al 2014, Neves et al 2017, Qi et al 2011, Ricard et al 2006, Romero-Pérez et al 2011, Tapio et al 2016). Remarkably, many of these donors are marine inhabitants (Table S3, Fig. 3, Fig. S47-S56). In general, these HGT events of marine origin were Neocallimastigomycota-wide (identified in 5-7 genera), and encoded functions that are beneficial for survival in a wide range of habitats (e.g. DNA repair, motility, and signal transduction, with only one metabolism-related event). We reason that this observation could offer interesting clues regarding the pre-gut sequestration ancestor of AGF. The presence of such genes could be a marker of an ancient symbiotic relationship between AGF ancestors and marine eukaryotes prior to fungal terrestrialization. The marine ancestral origin of the kingdom Fungi is currently undisputed

(Berbee et al 2017), but the nature and mechanism of the process, e.g. single ancestral terrestrialization followed by diversification, or multiple independent events of terrestrialization, is still unclear. It is notable that the majority (80%) of these HGT events were Neocallimastigomycota-specific, reflecting either a unique symbiotic relationship between AGF ancestors and these marine hosts or their loss in all other currently known basal fungal lineages.

Gene acquisition by HGT necessitates physical contact between donor and recipient organisms. Many of the HGT acquired traits by AGF are acquired from prokaryotes that are prevalent in the herbivorous gut microbiota (Fig. 3). However, since many of these traits are absolutely necessary for survival in the gut, the establishment of AGF ancestors in this seemingly inhospitable habitat is, theoretically, unfeasible. This dilemma is common to all HGT processes enabling niche adaptation and habitat diversification (Eme et al 2017). We put forth two evolutionary scenarios that could explain this dilemma not only for AGF, but also for other gut-dwelling anaerobic microeukaryotes, e.g. *Giardia*, *Blastocystis*, and *Entamoeba*, where HGT was shown to play a vital role in enabling survival in anaerobic conditions (Andersson et al 2003, Eme et al 2017, Grant and Katz 2014). The first is a coevolution scenario in which the progressive evolution of the mammalian gut from a short and predominantly aerobic structure characteristic of carnivores/insectivores to the longer, more complex, and compartmentalized structure encountered in herbivores was associated with a parallel progressive and stepwise acquisition of genes required for plant polymers metabolism and anaerobiosis by AGF ancestors, hence assuring its survival and establishment in the current herbivorous gut. The second possibility is that AGF ancestors were indeed acquired into a complex and anaerobic herbivorous gut, but initially represented an extremely minor component of the gut microbiome and survived in locations with relatively higher oxygen concentration in the alimentary tract e.g. mouth,

saliva, esophagus or in micro-niches in the rumen where transient oxygen exposure occurs. Subsequently, HGT acquisition has enabled the expansion of their niche, improved their competitiveness and their relative abundance in the herbivorous gut to the current levels.

In conclusion, our survey of HGT in AGF acquisition demonstrates that the process is absolutely crucial for the survival and growth of AGF in its unique habitat. This is not only reflected in the large number of events, massive duplication of acquired genes, and overall high HGT frequency observed in AGF genomes, but also in the nature of abilities imparted by the process. HGT events not only facilitated AGF adaptation to anaerobiosis, but also allowed them to drastically improve their polysaccharide degradation capacities, provide new venues for electron disposal via fermentation, and acquire new biosynthetic abilities. As such, we reason that the process should not merely be regarded as a conduit for supplemental acquisition of few additional beneficial traits. Rather, we posit that HGT enabled AGF to forge a new evolutionary trajectory, resulting in Neocallimastigomycota sequestration, evolution as a distinct fungal lineage in the fungal tree of life, and subsequent genus and species level diversification. This provides an excellent example of the role of HGT in forging the formation of high rank taxonomic lineages during eukaryotic evolution.

Conflict of Interest. The authors declare no conflict of interest.

Acknowledgments. This work has been funded by the NSF-DEB Grant numbers 1557102 to N.Y. and M.E. and 1557110 to J.E.S.

469 **References:**

470 Alexander WG, Wisecaver JH, Rokas A, Httinger CT (2016). Horizontally aquired gene in early-
471 diverging pathogenic fungi enable the use of host nucleosides and nucleotides. *Proc Nat Acad*
472 *Sci USA* **113:** 4116-4121.
473
474 Andersson DI, Jerlström-Hultqvist J, Näsvall J (2015). Evolution of new functions de novo and
475 from preexisting genes. *Cold Spring Harb Perspect Biol* **7:** a017996.
476
477 Andersson JO, Sjögren AM, Davis LA, Embley TM, Roger AJ (2003). Phylogenetic analyses of
478 diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr Biol* **13:** 94-
479 104.
480
481 Beiko RG, Harlow TJ, Ragan MA (2005). Highways of gene sharing in prokaryotes. *Proc Nat*
482 *Acad Sci USA* **102:** 14332-14337.
483
484 Berbee ML, James TY, Strullu-Derrien C (2017). Early Diverging Fungi: Diversity and Impact
485 at the Dawn of Terrestrial Life. *Annu Rev Microbiol* **71:** 41-60.
486
487 Bishop PJ, Speare R, Poulter R, Butler M, Speare BJ, Hyatt A *et al* (2009). Elimination of the
488 amphibian chytrid fungus *Batrachochytrium dendrobatidis* by Archey's frog *Leiopelma archeyi.*
489 *Dis Aquatic Organ* **84:** 9-15.
490

Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO *et al* (2015). Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Nat Acad Sci USA* **112:** 15976-15981.

Boxma B, Voncken F, Jannink S, Alen TV, Akhmanova A, Weelden SWHV *et al* (2004). The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase *E. Mol Microbiol* **51:** 1389 -1399.

Calkins S, Elledge NC, Hanafy RA, Elshahed MS, Youssef NH (2016). A fast and reliable procedure for spore collection from anaerobic fungi: Application for RNA uptake and long-term storage of isolates. *J Microbiol Methods* **127:** 206-213.

Callaghan TM, Podmirseg SM, Hohlweck D, Edwards JE, Puniya AK, Dagar SS *et al* (2015). *Buwchfawromyces eastonii* gen. nov., sp. nov.: a new anaerobic fungus (Neocallimastigomycota) isolated from buffalo faeces. *Mycokeys* **9:** 11-28.

Caro-Quintero A, Konstantinidis K (2015). Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. *ISME J* **9:** 958-967.

Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N *et al* (2010). Proto-genes and de novo gene birth. *Nature* **487:** 370-374.

513 Comtet-Marre S, Parisot N, Lepercq P, Chaucheyras-Durand F, Mosoni P, Peyretaillade E *et al*
514 (2017). Metatranscriptomics Reveals the Active Bacterial and Eukaryotic Fibrolytic
515 Communities in the Rumen of Dairy Cow Fed a Mixed Diet. *Front Microbiol* **8:** 67.
516
517 Creevey CJ, Kelly WJ, Henderson G, Leahy SC (2014). Determining the culturability of the
518 rumen bacterial microbiome. *Microb. Biotechnol.* **7:** 467-479.
519
520 Dagar SS, Kumar S, Griffith GW, Edwards JE, Callaghan TM, Singh R *et al* (2015). A new
521 anaerobic fungus (*Oontomyces anksri* gen. nov., sp. nov.) from the digestive tract of the Indian
522 camel (*Camelus dromedarius*). *Fungal Biol* **19:** 731-737.
523
524 de Koning AP, Brinkman FS, Jones SJ, Keeling PJ (2000). Lateral gene transfer and metabolic
525 adaptation in the human parasite *Trichomonas vaginalis*. *Mol Biol Evol* **17:** 1769-1773.
526
527 Doolittle WF (1998). You are what you eat: a gene transfer ratchet could account for bacterial
528 genes in eukaryotic nuclear genomes. *Trends Genet* **14:** 307–311.
529
530 Doolittle WF (1999). Lateral Genomics. *Trends Cell Biol* **9:** M5-M8.
531
532 Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H (2004). The timing of eukaryotic
533 evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc Nat Acad Sci USA*,
534 **101:** 15386-15391.
535

536 Eichinger L, Pachebat JA, Glockner G, Rajandream MA, al. e (2005). The genome of the social
537 amoeba *Dictyostelium discoideum*. *Nature* **435:** 43-57.
538
539 Eme L, Gentekaki E, Curtis B, Archibald J, Roger AJ (2017). Lateral gene transfer in the
540 adaptation of the anaerobic parasite *Blastocystis* to the gut. *Curr Biol* **27:** 807–820
541
542 Fitzpatrick DA (2012). Horizontal gene transfer in fungi. *FEMS Microbiol Lett* **2011:** 1-8.
543
544 Garcia-Vallvé S, Romeu A, Palau J (2000). Horizontal gene transfer of glycosyl hydrolases of
545 the rumen Fungi. *Mol Biol Evol* **17:** 352-361.
546
547 Gladyshev EA, Meselson M, Arkhipova IR (2008). Massive horizontal gene transfer in bdelloid
548 rotifers. *Science* **320:** 1210-1213.
549
550 Grant JR, Katz LA (2014). Phylogenomic study indicates widespread lateral gene transfer in
551 Entamoeba and suggests a past intimate relationship with parabasalids. *Genome Biol Evol* **6:**
552 2350-2360.
553
554 Gruninger RJ, Puniyab AK, Callaghanc TM, Edwardsc JE, Youssef N, Dagare SS *et al* (2014).
555 Anaerobic Fungi (Phylum Neocallimastigomycota): Advances in understanding of their
556 taxonomy, life cycle, ecology, role, and biotechnological potential. *FEMS Microbiol Ecol* **90:** 1-
557 17.
558

559 Haitjema CH, Gilmore SP, Henske JK, Solomon KV, Groot Rd, Kuo A *et al* (2017). A parts list
560 for fungal cellulosomes revealed by comparative genomics. *Nature Microbiol* **2:** 17087.
561
562 Hanafy RA, Elshahed MS, Youssef NH (2018). *Feramyces austinii*, gen. nov, sp. nov., an
563 anaerobic gut fungus from rumen and fecal samples of wild Barbary sheep and fallow deer **110:**
564 513-525
565
566 Harhangi HR, Akhmanova AS, Emmens R, Drift Cvd, Laat WTAMd, Dijken JPv *et al* (2003).
567 Xylose metabolism in the anaerobic fungus *Piromyces* sp. strain E2 follows the bacterial
568 pathway. *Arch Microbiol* **180:** 134-142.
569
570 He J, Yi L, Hai L, Ming L, Gao W, Ji R (2018). Characterizing the bacterial microbiota in
571 different gastrointestinal tract segments of the Bactrian camel. *Sci Rep* **8:** 654.
572
573 Hirt RP, Harriman N, Kajava AV, Embley TM (2002). A novel potential surface protein in
574 *Trichomonas vaginaliscontains* a leucine-rich repeat shared by micro-organisms
575 from all three domains of life. . *Mol Biochem Parasitol* **125:** 195–199.
576
577 Ho YW, Barr DJS (1995). Classification of anaerobic gut fungi from herbivores with emphasis
578 on rumen fungi from malaysia. *Mycologia* **87:** 655-677.
579
580 Huang JL (2013). Horizontal gene transfer in eukaryotes: the weak-link model. *Bioassays* **35:**
581 868-875.

Husnik F, McCutcheon JP (2017). Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol* **16:** 67-79.

Innan H, Kondrashov F (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* **11:** 97-10.

Jaramillo VDA, Sukno SA, Thon MR (2015). Identification of horizontally transferred genes in the genus *Colletotrichum* reveals a steady tempo of bacterial to fungal gene transfer. *BMC Genomics* **16:** 2.

Jordan A, Reichard P (1998). Ribonucleotide Reductases. *Annu Rev Biochem* **67:** 71-98.

Kaessmann H (2010). Origins, evolution, and phenotypic impact of new genes2. *Genome Res* **20:** 1313-1326.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40:** D109-114.

Keeling PJ, Palmer JD (2008). Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9:** 605-618.

604 Kishore SP, Stiller JW, Deitsch KW (2013). Horizontal gene transfer of epigenetic machinery
605 and evolution of parasitism in the malaria parasite Plasmodium falciparum and other
606 apicomplexans. *BMC Evol Biol* **13:** 37.
607
608 Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C *et al* (2016). No evidence
609 for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. *Proc*
610 *Nat Acad Sci USA* **113:** 5053-5058.
611
612 Lee C, Park C (2017). Bacterial Responses to Glyoxal and Methylglyoxal: Reactive Electrophilic
613 Species. *Int J Mol Sci* **18:** 169.
614
615 Leger MM, Eme L, Hug LA, Roger AJ (2016). Novel Hydrogenosomes in the Microaerophilic
616 Jakobid Stygiella incarcerata. *Mol Biol Evol* **33:** 2318-2336.
617
618 Lin H, Kwan AL, Dutcher SK (2010). Synthesizing and Salvaging NAD+: Lessons Learned
619 from Chlamydomonas reinhardtii. *PLOS Genet* **6:** e1001105.
620
621 Marcet-Bouben M, Gabaldon T (2010). Acquisition of prokaryotic genes by fungal genomes.
622 *Trends Genet* **26:** 5-8.
623
624 Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S *et al* (2017). CDD/SPARCLE:
625 functional classification of proteins via subfamily domain architectures. *Nucleic acids research*
626 **45:** D200-D203.

McCarthy CGP, Fitzpatrick DA (2016). Systematic Search for Evidence of Interdomain Horizontal Gene Transfer from Prokaryotes to Oomycete Lineages. *mSphere* **1:** e00195-00116.

Moliner C, Fournier PE, Raoult D (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution *FEMS Microbiol Rev* **34:** 281-294

Neves ALA, Li F, Ghoshal B, McAllister T, Guan LL (2017). Enhancing the Resolution of Rumen Microbial Classification from Metatranscriptomic Data Using Kraken and Mothur. *Front Microbiol* **8:** 2445.

Nguyen L-T, 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.

Nixon JEJ, Wang A, Field J, Morrison HG, McArthur AG, Sogin ML *et al* (2002). Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. *Eukaryot Cell* **1:** 181–190.

Ochman H, Lawrence JG, Groisman EA (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299-304.

650 Omelchenko MV, Wolf YI, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M *et al* (2005).
651 Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes
652 of adaptation to thermophily and radiation resistance *BMC Evol Biol* **5:** 57.
653
654 Parfrey LW, Lahr DJG, Knoll AH, Katz LA (2011). Estimating the timing of early eukaryotic
655 diversification with multigene molecular clocks *Proc Nat Acad Sci USA* **108:** 13624- 13629.
656
657 Pearse VB, Voigt O (2007). Field biology of placozoans (Trichoplax): distribution, diversity,
658 biotic interactions. *Integr Comp Biol* **47:** 677-692.
659
660 Pellicer MT, Nuñez MF, Aguilar J, Badia J, Baldoma L (2003). Role of 2-Phosphoglycolate
661 Phosphatase of Escherichia coli in Metabolism of the 2-Phosphoglycolate Formed in DNA
662 Repair. *J Bacteriol* **185:** 5815-5821.
663
664 Philippe H, Douady CJ (2003). Horizontal gene transfer and phylogenetics. *Curr Opin Microbiol*
665 **6:** 498-505.
666
667 Puigbo P, Pasamontes A, Garcia-Vallve S (2008). Gaining and losing the thermophilic
668 adaptation in prokaryotes. *Trends Genet* **24:** 10-14.
669
670 Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB *et al* (2011). Snapshot of the
671 Eukaryotic Gene Expression in Muskoxen Rumen-A Metatranscriptomic Approach. *PloS one* **6:**
672 e20521.

673

674 Qian Q, Keeling PJ (2001). Diplonemid glyceraldehyde–3-phosphate dehydrogenase (GAPDH)
675 and prokaryote to-eukaryote lateral gene transfer. *Protist* **152:** 193–201.

676

677 Ricard G, McEwan NR, Dutilh BE, Jouany J-P, Macheboeuf D, Mitsumori M *et al* (2006).
678 Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic,
679 carbohydrates-rich environment. *BMC Genomics* **7:** 22.

680

681 Richards TA, Monier A (2017). A tale of two tradigrades. *Proc Nat Acad Sci USA* **113:** 4892-
682 4894.

683

684 Richardson AO, Palmer JD (2007). Horizontal gene transfer in plants. *J Exp Bot* **58:** 1-9.

685

686 Romero-Pérez GA, Ominski KH, McAllister TA, Krause DO (2011). Effect of Environmental
687 Factors and Influence of Rumen and Hindgut Biogeography on Bacterial Communities in Steers.
688 *Appl Environ Microbiol* **77:** 258-268.

689

690 Rooke DM, Shattock RC (1983). Effect of Chloramphenicol and Streptomycin on
691 Developmental Stages of Phytophthom infestans. *Microbiol* **129:** 3401-3410.

692

693 Russell JB (1993). Glucose toxicity in Prevotella ruminicola: methylglyoxal accumulation and its
694 effect on membrane physiology. *Appl Environ Microbiol* **59:** 2844-2850.

695

Schönknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, Stanke M *et al* (2013a). Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339:** 1207-1210.

Schönknecht G, Weber AP, Lercher MJ (2013b). Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *Bioassays* **36:** 9-20.

Shterzer N, Mizrahi I (2015). The animal gut as a melting pot for horizontal gene transfer. *Can J Microbiol* **61:** 603-605.

Sievers F, Higgins DG (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27:** 135-145.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31:** 3210-3212.

Soucy SM, Huang J, Gogarten JP (2015). Horizontal gene transfer: building the web of life. *Nature Rev Genet* **16:** 472-482.

Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW *et al* (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat Commun* **9:** 870.

719 Sun GL, Yang ZF, Ishwar A, Huang JL (2010). Algal genes in the closest relatives of animals.
720 *Mol Biol Evol* **27:** 2879-2889.
721
722 Syvanen M (2012). Evolutionary implications of horizontal gene transfer. *Annu Rev Genet* **46:**
723 341-358.
724
725 Tapio I, Shingfield KJ, McKain N, Bonin A, Fischer D, Bayat AR *et al* (2016). Oral Samples as
726 Non-Invasive Proxies for Assessing the Composition of the Rumen Microbial Community. *PloS*
727 *one* **11:** e0151220.
728
729 Taylor JW, Berbee ML (2006). Dating divergences in the Fungal Tree of Life: review and new
730 analyses. *Mycologia* **98:** 838-849.
731
732 Wang X, Liu X, Groenewald JZ (2017). Phylogeny of anaerobic fungi (phylum
733 Neocallimastigomycota), with contributions from yak in China. *Antonie Van Leeuwenhoek* **110:**
734 87-103.
735
736 Wiedenbeck J, Cohan FM (2011). Origins of bacterial diversity through horizontal genetic
737 transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* **35:** 957-976.
738
739 Wisecaver JH, Brosnahan ML, Hackett JD (2013). Horizontal gene transfer is a significant driver
740 of gene innovation in dinoflagellates. *Genome Biol Evol* **5:** 2368-2381.
741

742 Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ *et al*
743 (2013). Genome of the anaerobic fungus Orpinomyces sp. C1A reveals the unique evolutionary
744 history of a remarkable plant biomass degrader *Appl Environ Microbiol* **79:** 4620-4634.
745
746 Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS (2015). Insights into the
747 metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum
748 'Diapherotrites'. *ISME J* **9:** 447-460.
749
750

751 Figure Legends

752 Fig. 1. Workflow diagram describing the procedure employed for identification HGT events in
753 Neocallimastigomycota datasets analyzed in this study.

754 Fig. 2. (A) Total Number of HGT events identified per AGF genus. (B) Distribution pattern of
755 HGT events in AGF transcriptomes in all seven AGF genera examined.

756 Fig. 3. Identity of HGT donors and their contribution to the various functional classes. The X-
757 axis shows the absolute number of events belonging to each of the functional classes shown in
758 the legend. The tree is intended to show the relationship between the donors taxa and is not
759 drawn to scale. Bacterial donors are shown with red branches depicting the phylum-level, with
760 the exception of Firmicutes and Bacteroidetes donors, where the order-level is shown, and
761 Proteobacteria, where the class-level is shown. Archaeal donors are shown with green branches
762 and all belonged to the Methanobacteriales order of Euryarchaeota. Eukaryotic donors are shown
763 with blue branches. Only the 262 events from a definitive-taxon donor are shown in the figure.
764 The other 65 events were clearly nested within a non-fungal clade, but a definitive donor taxon
765 could not be ascertained. Functional classification of the HGT events, determined by searching
766 the Conserved Domain server (Marchler-Bauer et al 2017) against the COG database are shown
767 in B. For events with no COG classification, a search against the KEGG orthology database
768 (Kanehisa et al 2012) was performed. For the major COG/KEGG categories (metabolism,
769 cellular processes and signaling, and Information storage and processing), sub-classifications are
770 shown in C, D, and E, respectively.

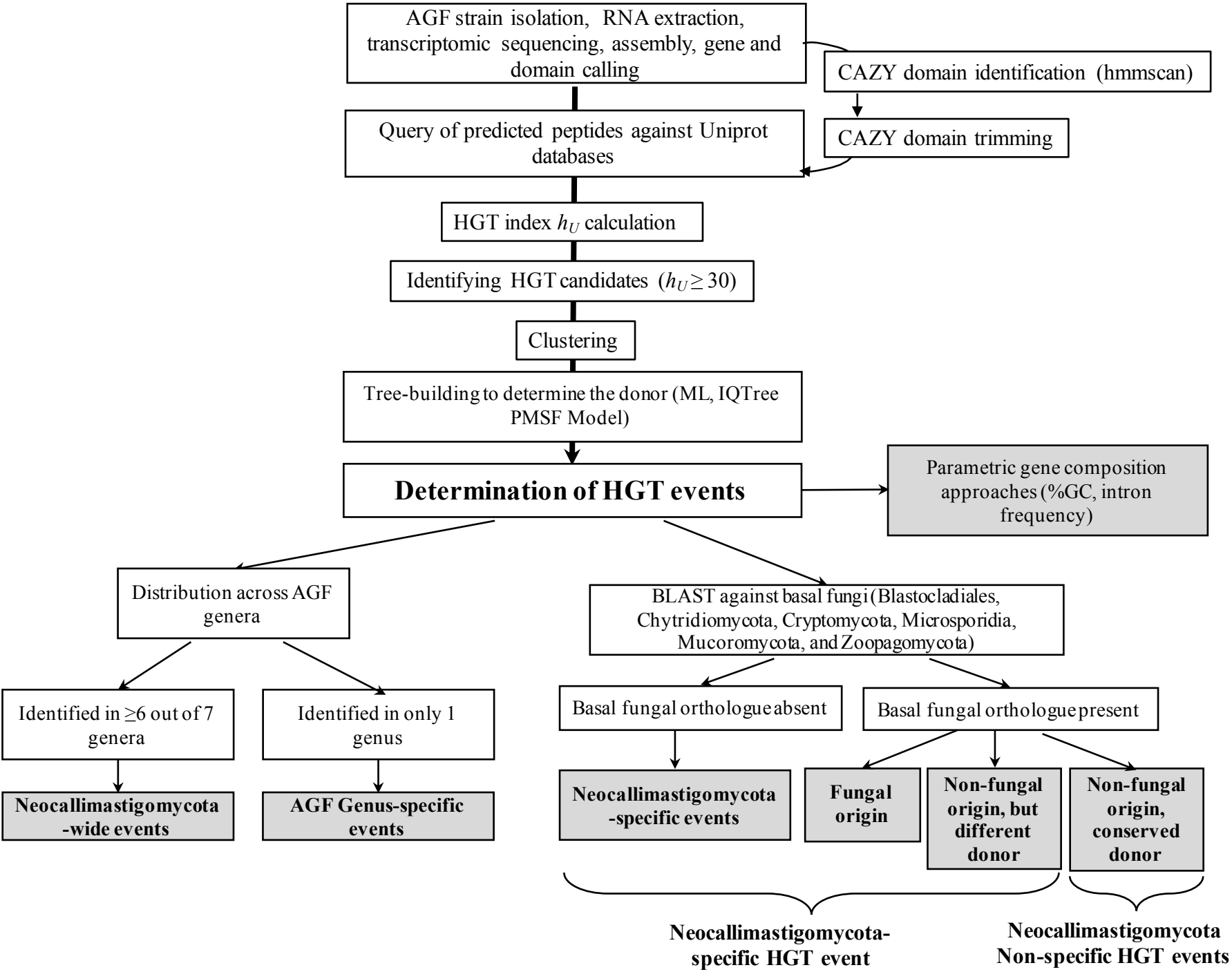
771 Fig. 4. HGT impact on AGF central metabolic abilities. Pathways for sugar metabolism are
772 highlighted in blue, pathways for amino acid metabolism are highlighted in red, pathways for
773 cofactor metabolism are highlighted in green, pathways for nucleotide metabolism are

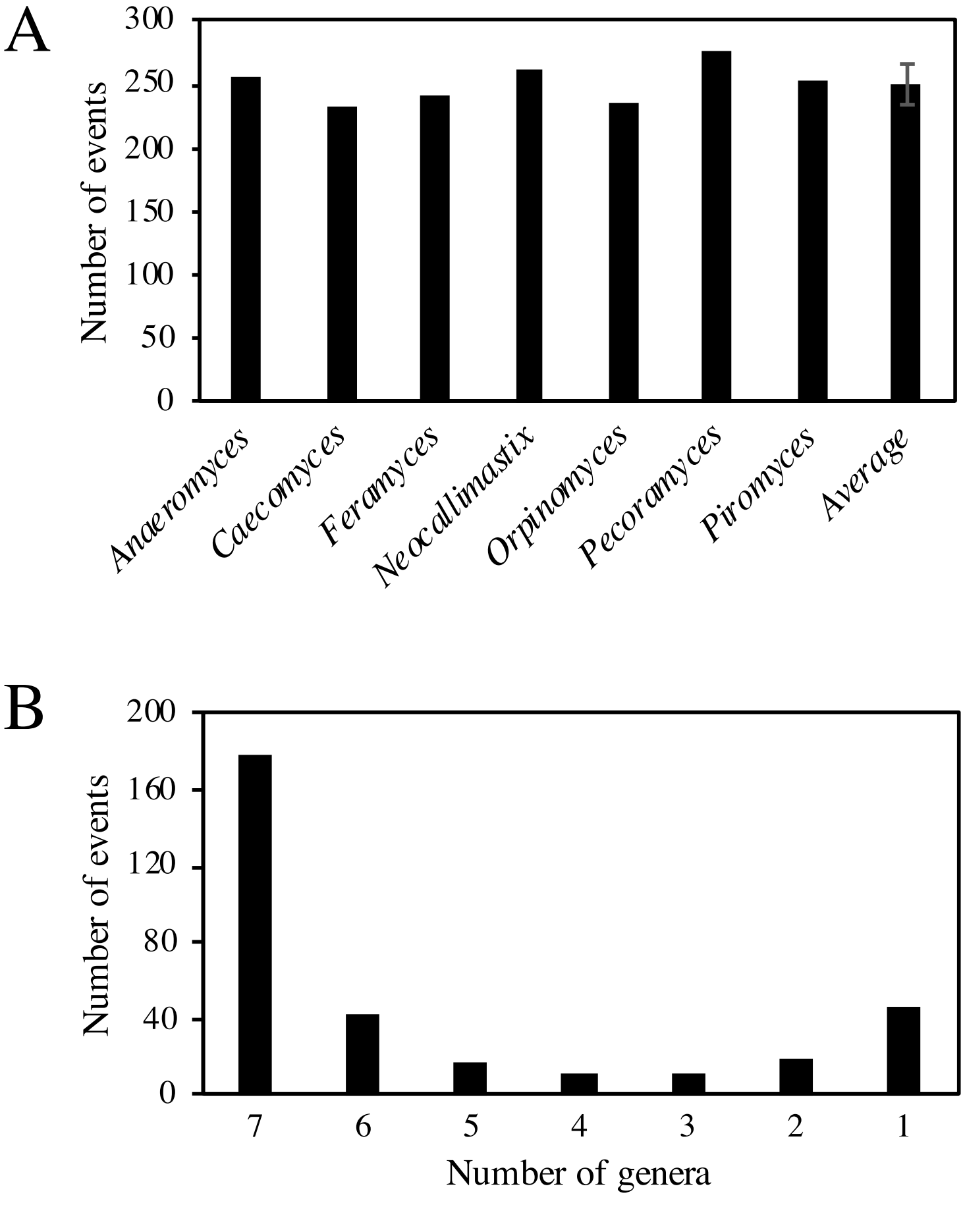
774 highlighted in grey, pathways for lipid metabolism are highlighted in orange, fermentation
775 pathways are highlighted in purple, while pathways for detoxification are highlighted in brown.
776 The double black lines depict the hydrogenosomal outer and inner membrane. Arrows
777 corresponding to enzymes encoded by horizontally transferred transcripts are shown with thicker
778 dotted lines and are given numbers 1 through 48 as follows. Sugar metabolism (1-11): 1. Xylose
779 isomerase, 2. Xylulokinase, 3. Ribokinase, 4. 2,3-bisphosphoglycerate-independent
780 phosphoglycerate mutase, 5. 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase, 6.
781 Phosphoenolpyruvate synthase, 7. Phosphoenolpyruvate carboxykinase (GTP), 8. Aldose-1-
782 epimerase, 9. Galactokinase, 10. Galactose-1-phosphate uridyltransferase, 11. UDP-glucose-4-
783 epimerase. Amino acid metabolism (12-20): 12. Aspartate-ammonia ligase, 13. Tryptophan
784 synthase (TrpB), 14. Tryptophanase, 15. Monofunctional prephenate dehydratase, 16. Serine-O-
785 acetyltransferase, 17. Cysteine synthase, 18. Low-specificity threonine aldolase, 19. 5'-
786 methylthioadenosine nucleosidase/5'-methylthioadenosine phosphorylase (MTA phosphorylase),
787 20. Arginase. Cofactor metabolism (21-28): 21. Pyridoxamine 5'-phosphate oxidase, 22. L-
788 aspartate oxidase (NadB), 23. Quinolate synthase (NadA), 24. NH(3)-dependent NAD(+)
789 synthetase (NadE), 25. 2-dehydropantoate 2-reductase, 26. dephosphoCoA kinase, 27.
790 Dihydrofolate reductase (DHFR) family, 28. Dihydropteroate synthase. Nucleotide metabolism
791 (29-36): 29. GMP reductase, 30. Trifunctional nucleotide phosphoesterase, 31. deoxyribose-
792 phosphate aldolase (DeoC), 32. Oxygen-sensitive ribonucleoside-triphosphate reductase class III
793 (NrdD), 33. nucleoside/nucleotide kinase family protein, 34. Cytidylate kinase-like family, 35.
794 thymidylate synthase, 36. thymidine kinase. Pyruvate metabolism (fermentation pathways) (37-
795 41): 37. D-lactate dehydrogenase, 38. bifunctional aldehyde/alcohol dehydrogenase family of Fe-
796 alcohol dehydrogenase, 39. Butanol dehydrogenase family of Fe-alcohol dehydrogenase, 40. Zn-

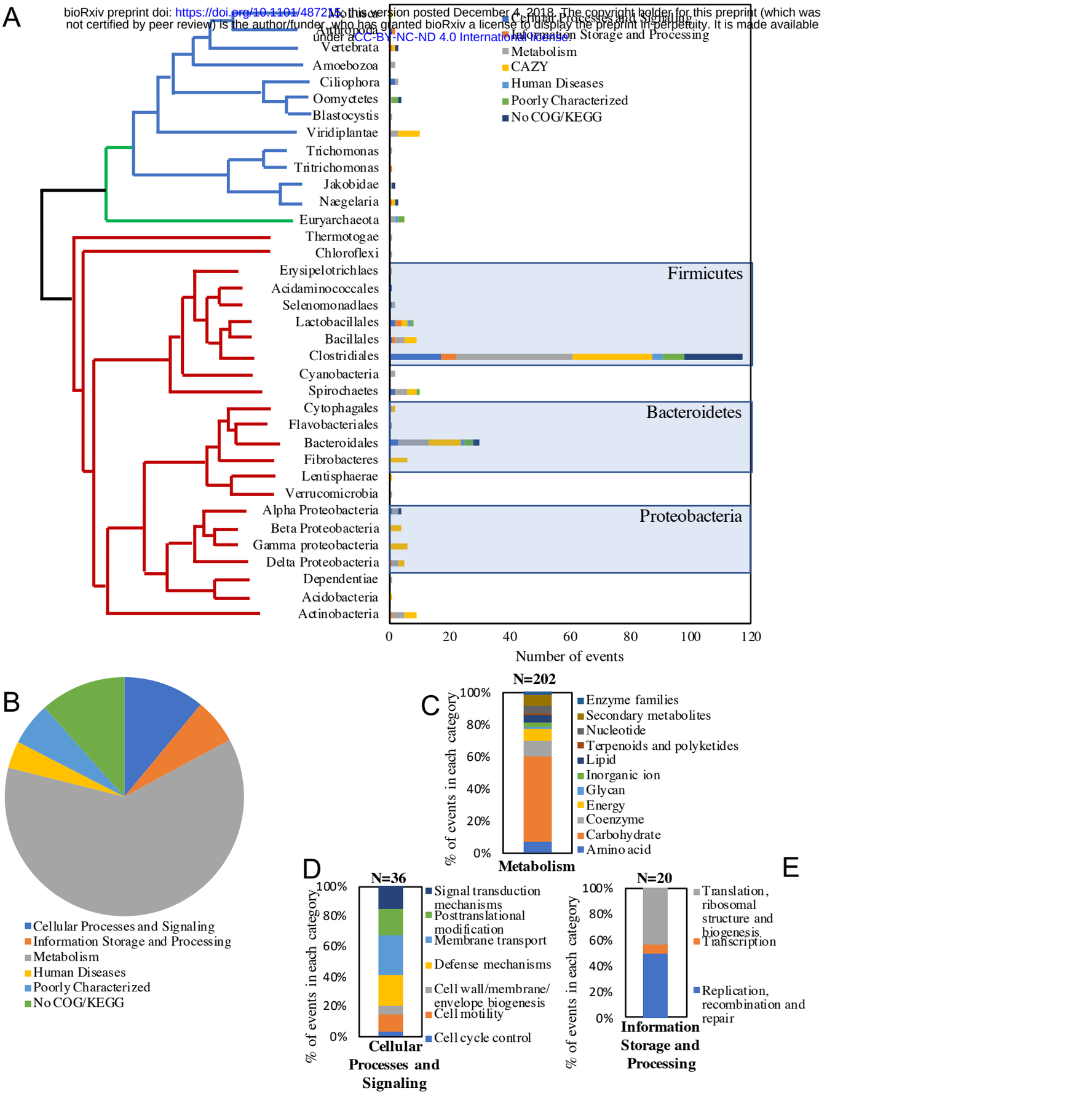
797 type alcohol dehydrogenase, 41. Fe-only hydrogenase. Detoxification reactions (42-45): 42.
798 Phosphoglycolate phosphatase, 43. Glyoxal reductase, 44. Glyoxalase I, 45. Glyoxalase II. Lipid
799 metabolism (46-48): 46. CDP-diacylglycerol--serine O-phosphatidyltransferase, 47.
800 lysophospholipid acyltransferase LPEAT, 48. methylene-fatty-acyl-phospholipid synthase.
801 Abbreviations: CDP-DAG, CDP-diacylglycerol; 7,8 DHF, 7,8 dihydrofolate; EthA,
802 ethanolamine; Gal, galactose; GAP, glyceraldehyde-3-P; Glu, glucose; GSH, glutathione; I,
803 complex I NADH dehydrogenase; NaMN, Nicotinate D-ribonucleotide; Orn, ornithine; PEP,
804 phosphoenol pyruvate; Phenyl-pyr, phenylpyruvate; PRPP, phosphoribosyl-pyrophosphate; Ptd,
805 phosphatidyl; SAM; S-adenosylmethionine; THF, tetrahydrofolate.
806 Fig. 5. (A) Maximum likelihood tree showing the phylogenetic affiliation of AGF galactokinase.
807 AGF genes highlighted in light blue clustered within the Flavobacteriales order of the
808 Bacteroidetes phylum and were clearly nested within the bacterial domain (highlighted in green)
809 attesting to their non-fungal origin. Fungal galactokinase representatives are highlighted in pink.
810 (B) Maximum likelihood tree showing the phylogenetic affiliation of AGF Fe-only hydrogenase.
811 AGF genes highlighted in light blue clustered within the Thermotogae phylum and were clearly
812 nested within the bacterial domain (highlighted in green) attesting to their non-fungal origin.
813 *Stygiella incarcerata* (anaerobic Jakobidae) clustered with the Thermotogae as well as has
814 recently been suggested (Leger et al 2016). Fe-only hydrogenases from *Gonopodya prolifera*
815 (Chytridiomycota) clustered with the AGF genes. This is an example of one of the rare occasions
816 (n=31) where a non-AGF basal fungal representative showed an HGT pattern with the same
817 donor affiliation as the Neocallimastigomycota. Other basal fungal Fe-only hydrogenase
818 representatives are highlighted in pink and clustered outside the bacterial domain. (C) Maximum
819 likelihood tree showing the phylogenetic affiliation of AGF L-aspartate oxidase (NadB). AGF

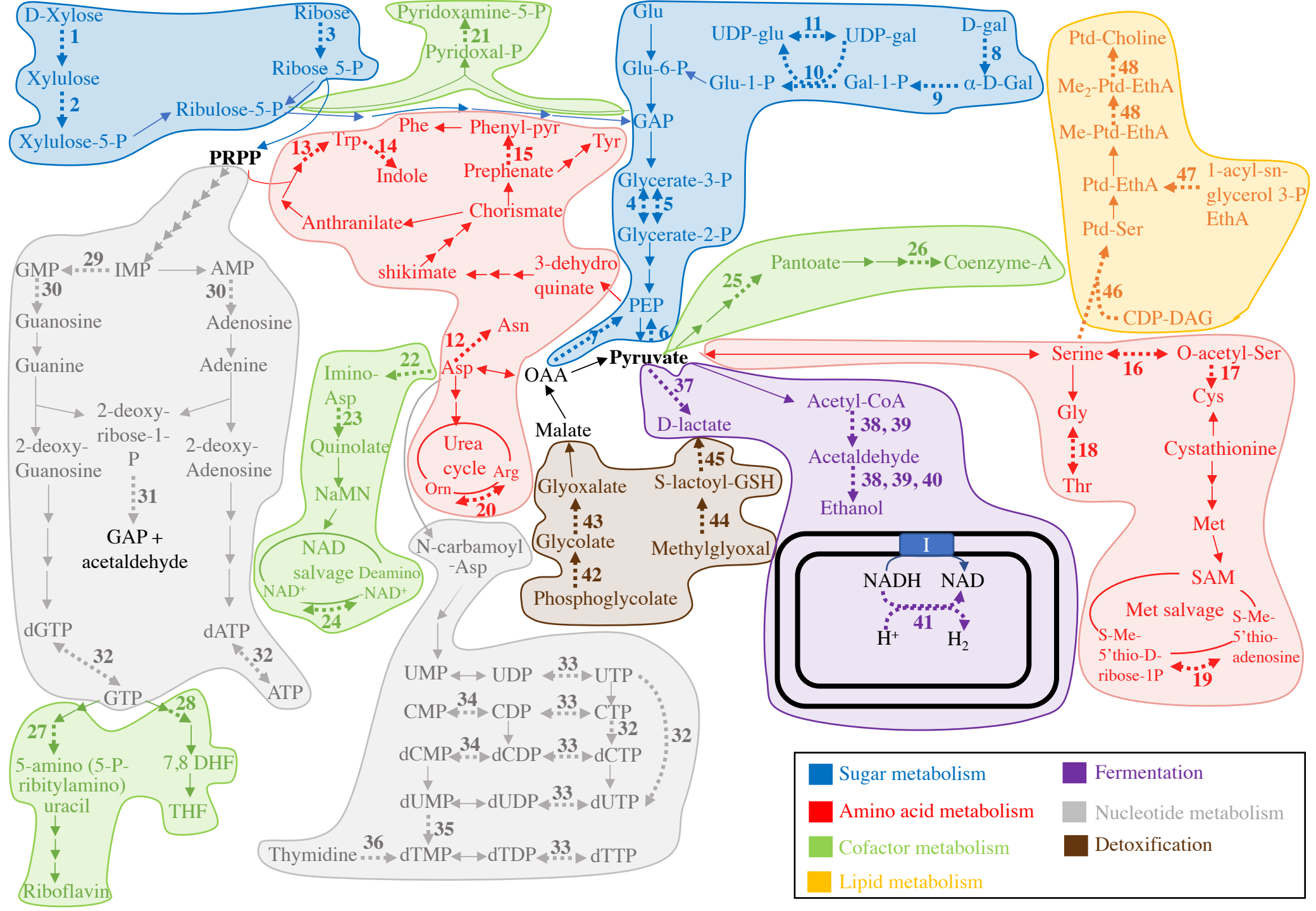
820 genes highlighted in light blue clustered within the Delta-Proteobacteria class and were clearly
821 nested within the bacterial domain (highlighted in green) attesting to their non-fungal origin. As
822 de-novo NAD synthesis in fungi usually follow the five-enzyme pathway starting from
823 tryptophan, as opposed to the two-enzyme pathway from aspartate, no NadB were found in non-
824 AGF fungi and hence no fungal cluster is shown in the tree. (D) Maximum likelihood tree
825 showing the phylogenetic affiliation of AGF oxygen-sensitive ribonucleotide reductase (NrdD).
826 AGF genes highlighted in light blue clustered with representatives from Candidate phylum
827 Dependentiae and were clearly nested within the bacterial domain (highlighted in green) attesting
828 to their non-fungal origin. Fungal NrdD representatives are highlighted in pink. GenBank
829 accession numbers are shown in parentheses. Alignment was done using the standalone Clustal
830 Omega (Sievers and Higgins 2018) and trees were constructed using IQ-tree (Nguyen et al
831 2015).
832 Fig. 6. HGT in the AGF CAZyome shown across the seven genera studied. Glycosyl Hydrolase
833 (GH), Carboxyl Esterase (CE), and Polysaccharide Lyase (PL) families are shown to the left.
834 The color of the cells depicts the prevalence of HGT within each family. Red indicates that
835 100% of the CAZyme transcripts were horizontally transferred. Shades of red-orange indicate
836 that HGT contributed to > 50% of the transcripts belonging to that CAZy family. Blue indicates
837 that 100% of the CAZyme transcripts were of fungal origin. Shades of blue indicate that HGT
838 contributed to < 50% of the transcripts belonging to that CAZy family. The numbers in each cell
839 indicate the affiliation of the HGT donor as shown in the key to the right.
840 Fig. 7. Principal-component analysis biplot of the distribution of CAZy families in AGF
841 genomes (★), compared to representatives of other basal fungi belonging to the
842 Mucoromycotina (⬣), Chytridiomycota (●), Blastocladiomycota (■), Entomophthoromycotina

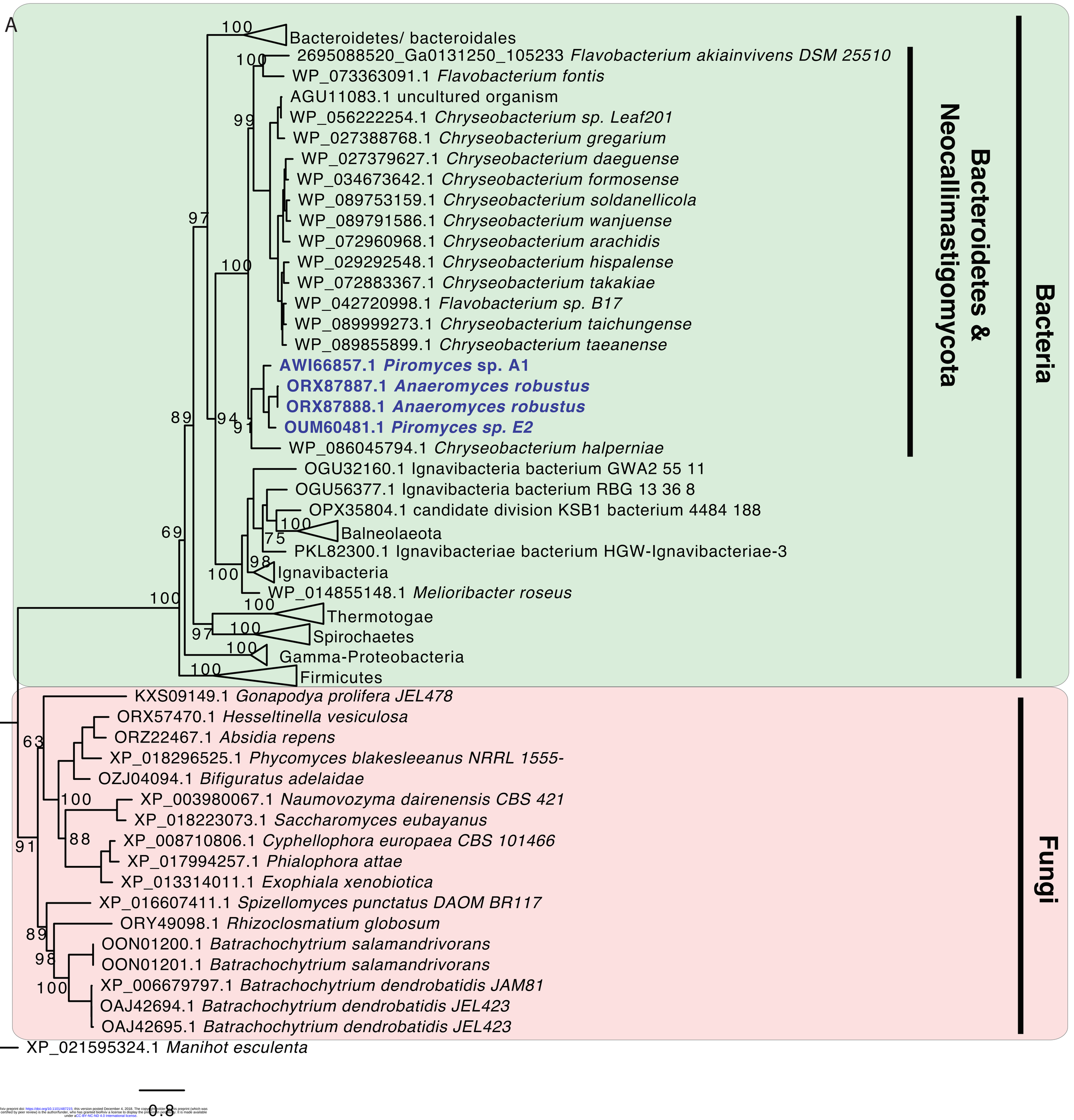
843 (⬬), Mortierellomycotina (△), Glomeromycota (✚), Kickxellomycotina (▽), and
844 Zoopagomycotina (✖). CAZy families are shown as colored dots. The color code used was as
845 follows: green, CAZy families that are absent from AGF genomes; black, CAZy families present
846 in AGF genomes and with an entirely fungal origin; blue, CAZy families present in AGF
847 genomes and for which HGT contributed to < 50% of the transcripts in the examined
848 transcriptomes; red, CAZy families present in AGF genomes and for which HGT contributed to
849 > 50% of the transcripts in the examined transcriptomes. The majority of CAZyme families
850 defining the AGF CAZyome were predominantly of non-fungal origin (red and blue dots).


AGF strain isolation, RNA extraction, transcriptomic sequencing, assembly, gene and domain calling
CAZY domain identification (hmmscan)
CAZY domain trimming
Query of predicted peptides against Uniprot databases
HGT index h_U calculation
Identifying HGT candidates ($h_U \geq 30$)
Clustering
Tree-building to determine the donor (ML, IQTree PMSF Model)
Determination of HGT events
Parametric gene composition approaches (%GC, intron frequency)
Distribution across AGF genera
BLAST against basal fungi (Blastocladiales, Chytridiomycota, Cryptomycota, Microsporidia, Mucoromycota, and Zoopagomycota)
Identified in ≥6 out of 7 genera
Identified in only 1 genus
Basal fungal orthologue absent
Basal fungal orthologue present
Neocallimastigomycota -wide events
AGF Genus-specific events
Neocallimastigomycota -specific events
Fungal origin
Non-fungal origin, but different donor
Non-fungal origin, conserved donor
Neocallimastigomycota-specific HGT event
Neocallimastigomycota Non-specific HGT events


A
300
250
200
150
100
50
0
Number of events
Anaeromyces
Caecomyces
Feramyces
Neocallimastix
Orpinomyces
Pecoramyces
Piromyces
Average
B
200
160
120
80
40
0
Number of events
7
6
5
4
3
2
1
Number of genera




D-Xylose
1
Xylulose
2
Xylulose-5-P
Ribose
3
Ribose 5-P
Ribulose-5-P
Pyridoxamine-5-P
21
Pyridoxal-P
PRPP
Glu
UDP-glu
11
UDP-gal
D-gal
8
Glu-6-P
Glu-1-P
10
Gal-1-P
9
α-D-Gal
GAP
Glycerate-3-P
4
5
Glycerate-2-P
PEP
6
7
Pyruvate
OAA
Malate
Phe
Phenyl-pyr
Tyr
13
Trp
14
Indole
15
Prephenate
Anthranilate
Chorismate
shikimate
3-dehydro quinate
Asn
12
Asp
Urea cycle
Arg
Orn
20
Pantoate
26
Coenzyme-A
25
Ptd-Choline
48
Me2-Ptd-EthA
48
Me-Ptd-EthA
Ptd-EthA
47
1-acyl-sn-glycerol 3-P EthA
Ptd-Ser
46
CDP-DAG
Serine
16
O-acetyl-Ser
17
Gly
18
Thr
Cys
Cystathionine
Met
SAM
Met salvage
S-Me-5'thio-adenosine
S-Me-5'thio-D-ribose-1P
19
GMP
29
IMP
AMP
30
30
Guanosine
Adenosine
Guanine
Adenine
2-deoxy-ribose-1-P
2-deoxy-Guanosine
2-deoxy-Adenosine
31
GAP + acetaldehyde
dGTP
dATP
32
32
GTP
ATP
Imino-Asp
22
23
Quinolate
NaMN
NAD salvage
Deamino
NAD+
-NAD+
24
N-carbamoyl-Asp
37
D-lactate
Acetyl-CoA
38, 39
Acetaldehyde
38, 39, 40
Ethanol
Glyoxalate
45
S-lactoyl-GSH
43
Glycolate
44
Methylglyoxal
42
Phosphoglycolate
I
NADH
NAD
41
H+
H2
UMP
UDP
33
UTP
CMP
34
CDP
33
CTP
32
dCMP
34
dCDP
33
dCTP
32
dUMP
dUDP
33
dUTP
35
Thymidine
36
dTMP
dTDP
33
dTTP
27
28
5-amino (5-P-ribitylamino) uracil
7,8 DHF
THF
Riboflavin
Sugar metabolism
Amino acid metabolism
Cofactor metabolism
Lipid metabolism
Fermentation
Nucleotide metabolism
Detoxification



ANM86737.1 *Stygiella incarcerata*
OUM68179.1 *Piromyces* sp. E2
AAL90459.1 *Piromyces* sp. E2
AWI66880.1 *Piromyces* sp. A1
ORX45233.1 *Piromyces finnis*
AFJ73533.1 *Neocallimastix frontalis*
AAN39277.1 *Neocallimastix frontalis*
ORY77471.1 *Neocallimastix californiae*
ORY52215.1 *Neocallimastix californiae*
AFJ73534.1 *Neocallimastix frontalis*
AAK60409.1 *Neocallimastix frontalis*
ORX75213.1 *Anaeromyces robustus*
KXS15429.1 *Gonapodya prolifera JEL478*
WP_041083655.1 *Thermotoga profunda*
WP_012003383.1 *Pseudothermotoga lettingae*
WP_028843928.1 *Pseudothermotoga elfii*
WP_013931422.1 *Pseudothermotoga thermarum*
WP_082021712.1 *Thermotoga caldifontis*
WP_031502934.1 *Pseudothermotoga hypogea*
AJC72862.1 *Pseudothermotoga hypogea DSM 11164*
KUK03700.1 *Thermotoga sp. 50 64*
WP_075666576.1 Thermosipho
WP_077198749.1 *Thermosipho affectus*
WP_012057804.1 *Thermosipho melanesiensis*
WP_012580432.1 *Thermosipho africanus*
WP_064011755.1 *Fervidobacterium pennivorans*
WP_011993425.1 *Fervidobacterium nodosum*
WP_072758530.1 *Fervidobacterium gondwanense*
WP_069292446.1 *Fervidobacterium thailandense*
WP_011943326.1 *Thermotoga petrophila*
WP_038033316.1 Thermotoga
KUK33763.1 *Thermotoga sp. 47 83*
WP_012310723.1 *Thermotoga sp. RQ2*
WP_010865073.1 *Thermotoga maritima*
AGL49125.1 *Thermotoga maritima MSB8*
EJX26240.1 *Thermotoga sp. EMP*
WP_012744686.1 Kosmotoga
WP_047754056.1 *Kosmotoga pacifica*
WP_068346126.1 *Kosmotoga arenicorallina*
WP_099771899.1 *Mesotoga sp. H07.pep.5.3*
WP_014730285.1 *Mesotoga prima*
CCU85708.1 *Mesotoga infera*
KUK68286.1 *Mesotoga infera*
WP_099830556.1 Mesotoga
KUK80646.1 *Mesotoga prima*
KUK94929.1 Thermotogales bacterium 46 20
WP_094605662.1 *Sporomusa silvacetica*
WP_094606391.1 *Sporomusa silvacetica*
WP_027637394.1 *Clostridium cadaveris*
WP_007953234.1 *Pelosinus fermentans*
WP_075754529.1 *Sporomusa sphaeroides*
WP_006317165.1 *Caldisalinibacter kiritimatiensis*
WP_035164868.1 *Caloranaerobacter azorensis*
WP_054871931.1 *Caloranaerobacter sp. TR13*
WP_091728490.1 *Proteiniborus ethanoligenes*
WP_071141319.1 *Sporanaerobacter sp. PP17-6a*
KZD16134.1 candidate division Hyd24-12 bacterium Ran 1
OGV49733.1 Lentisphaerae bacterium GWF2 52 8
OHB43018.1 Planctomycetes bacterium GWC2 45 44
PKP01685.1 Bacteroidetes bacterium HGW-Bacteroidetes-6
OFY08089.1 Bacteroidetes bacterium GWE2 42 42
OIP01611.1 Bacteroidetes bacterium CG2 30 32 10
PKP20906.1 Bacteroidetes bacterium HGW-Bacteroidetes-2
OFY34382.1 Bacteroidetes bacterium GWF2 38 335
WP_041356083.1 *Melioribacter roseus*
AFN75712.1 *Melioribacter roseus P3M-2*
KUO61574.1 bacterium BRH c32
OGV06025.1 Ignavibacteria bacterium RIFOXYB2 FULL 37 11
OGU78338.1 Ignavibacteria bacterium RIFOXYA12 FULL 38 9
WP_014559401.1 Ignavibacterium album
OQY74772.1 Ignavibacteriales bacterium UTCHB2
ORE04845.1 *Rhizopus microsporus var. microsporus*
ORE19264.1 *Rhizopus microsporus*
CEI94145.1 *Rhizopus microsporus*
OBZ83231.1 *Choanephora cucurbitarum*
CDS05913.1 *Lichtheimia ramosa*
CDH53287.1 *Lichtheimia corymbifera JMRC FSU 9682*
XP_018293964.1 *Phycomyces blakesleeanus NRRL 1555 -*
OMO89337.1 *Corchorus capsularis*

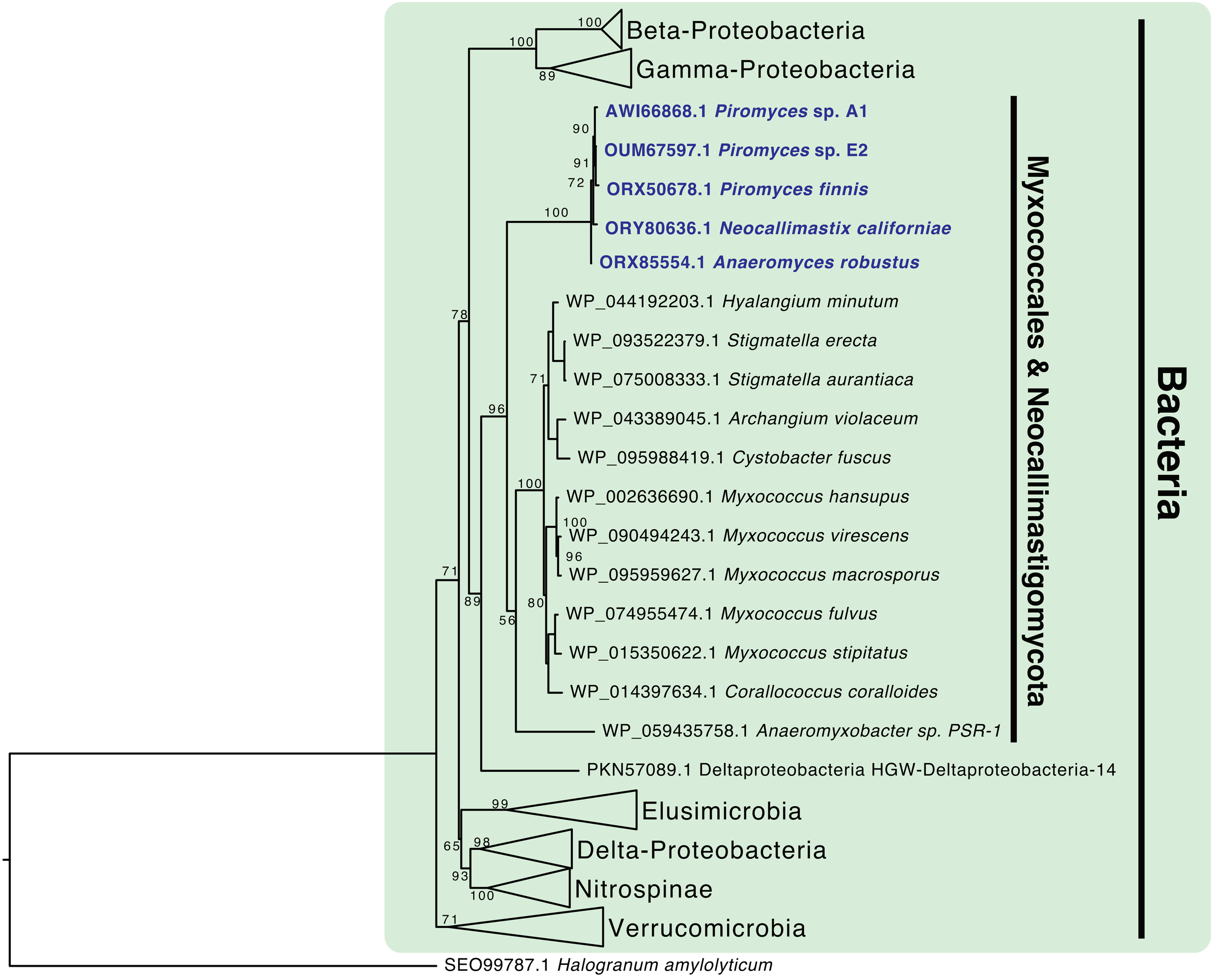
79
99
82
99
97
98
100
71
99
89
100
85
100
100
85
72
85
85

Thermotogae & Neocallimastigomycota
Firmicutes
Bacteroidetes
Ignavibacteria
Bacteria
Fungi

B

0.8

C



D

Family	Genus						
	Anaeromyces	*Caecomyces*	*Pecoramyces*	*Piromyces*	*Neocallimastix*	*Feramyces*	*Orpinomyces*
GH1	11, 18	18	6, 18	18	11, 18	18	7, 18
GH2	7		7	7	7	7	7
GH3	7, 15	7, 15	7, 15	7, 15	7		7
GH5	2, 11	2, 7, 11	2, 7, 11	2, 7, 11	2, 7, 11	2, 7, 11	2, 7, 11
GH6	13	13	13	13	13	13	13
GH8	5, 6	6	5, 6	5, 6	5, 6	6	5, 6
GH9							
GH10	4, 7	4, 7	4, 7	4, 7	4, 7	4, 7	4, 7
GH11	5, 7	5	5	5	5	5	5, 7
GH13	3, 7, 8,11	3, 7, 8	3, 7, 8, 11	3, 8, 11	3, 7, 8,11	3, 7, 8,11	3, 7, 8
GH16	3, 7		3	7	3, 7, 19	3, 7	3, 7
GH18		3	6, 12	6	6	6	
GH20							
GH24	7, 12						12
GH25	7	7	7	7	7	7	
GH26	5, 15	5, 15	5, 15	5, 15	5, 15	15	15
GH28			3		3, 19		3
GH30	7				7		
GH31			2	19			
GH32	7		7		3, 19	7	7
GH36	6				6		
GH37							
GH39	3, 7	7	3, 7	7	3, 7		7
GH43	7	7, 12	7	7, 11, 12	7, 12	7, 14	7
GH45							
GH47	16		16	16		16	
GH48	15	15	15	15	15	15	15
GH53	7	7	7	7	7	7	7
GH57							
GH64	8	8	8		8	8	
GH67	3	3	3	3			
GH76						3, 7	
GH78							7
GH88			15	15	15	15	15
GH95	3						3
GH97					3	3	
GH108				20			
GH114	12	12	12	12	12	12	12
GH115	7	7	7	7	7	7	
CE1	5, 7	5, 7	5, 7	5, 7	5, 7	5, 7	5, 7
CE2	7, 14	7, 14	7, 14	7, 14	7, 14	3, 7	7
CE3	7	7	7	7	7	7	7
CE4	2, 5, 8, 12	2, 5, 8, 15	2, 5, 8	2, 5, 8, 12, 15	2, 5, 8	2,5,8,15	5, 8, 15
CE6	7	7	7	7	7	7	7
CE7	7						
CE8	19	19	19	19	19	19	19
CE12	3, 13, 17		7	7, 13	7	3, 7	7, 17
CE15	3	3, 5	3	3	3, 5	3	3
CE16					19		
PL1	2	2	2, 12	2, 12	2	2, 12	2, 12
PL3							
PL4	19	19	19	19	19	19	19
PL9	7, 14		7, 14	9, 14	7		14
PL11	1				7		

KEY

1	Acidobacteria		Bacteria
2	Actinobacteria		
3	Bacteroidetes		
4	Deinococcus		
5	Fibrobacter		
6	Bacillales	Firmicutes	
7	Clostridiales		
8	Unclassified Firmicutes		
9	Lentisphaerae		
10	Alpha-Proteobacteria	Proteobacteria	
11	Beta-Proteobacteria		
12	Gamma-Proteobacteria		
13	Delta-Proteobacteria		
14	Spirochaetes		
15	Bacteria (unnested)		
16	Arthropoda		Eukaryota
17	Metazoa		
18	Mollusca		
19	Viridiplantae		
20	Neagelaria/Cyanobacteria		

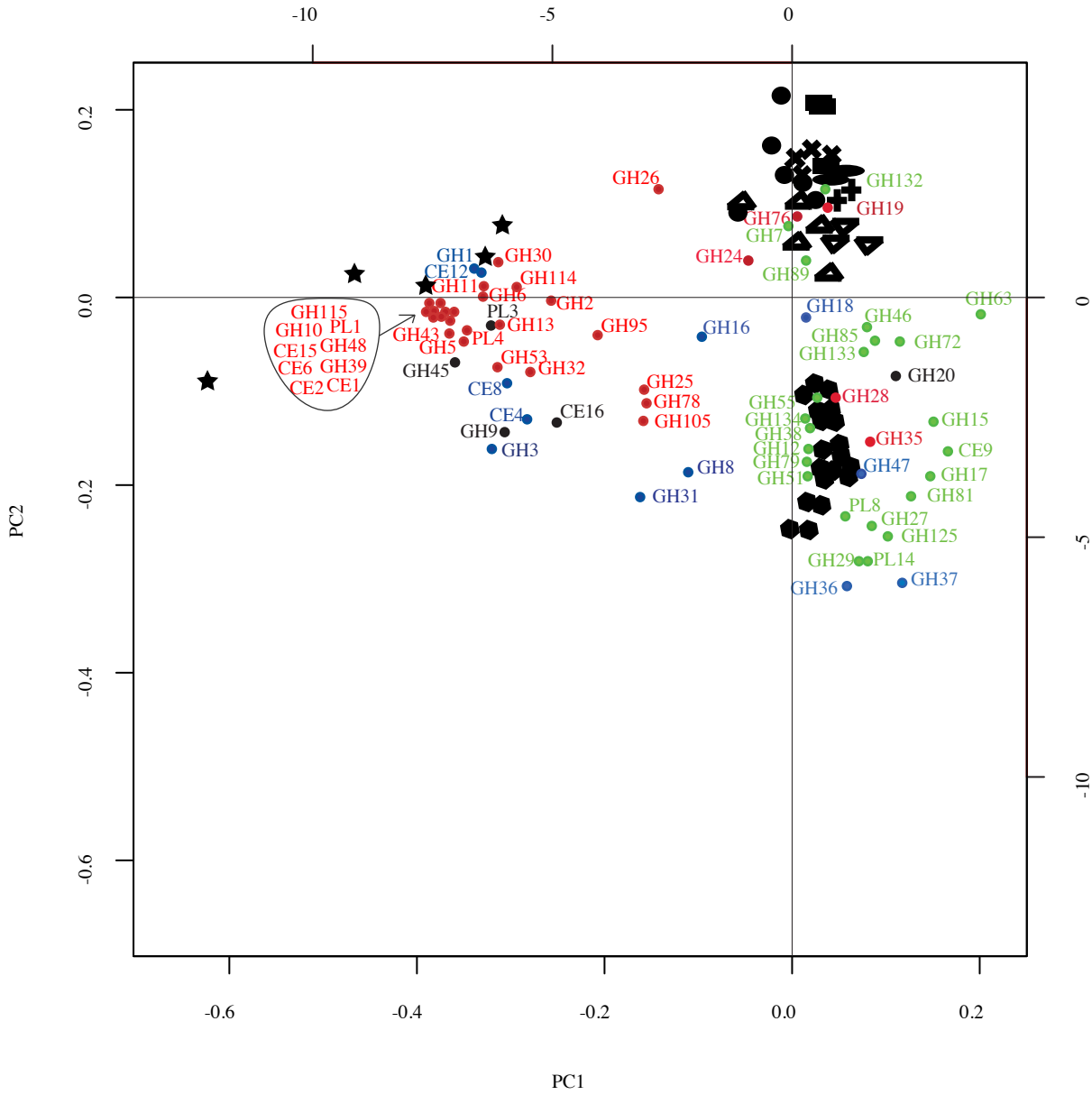

-10
-5
0
0.2
0.0
-0.2
-0.4
-0.6
PC2
-0.6
-0.4
-0.2
0.0
0.2
PC1
0
-5
-10
GH115
GH10 PL1
CE15 GH48
CE6 GH39
CE2 CE1
GH26
GH132
GH19
GH76
GH7
GH24
GH89
GH1
GH30
CE12
GH114
GH11
GH6
GH2
PL3
GH43
GH13
GH95
GH16
PL4
GH5
GH53
GH45
GH32
CE8
GH25
GH78
GH105
CE4
CE16
GH9
GH3
GH8
GH31
GH18
GH63
GH46
GH85
GH72
GH133
GH20
GH28
GH53
GH134
GH15
GH38
GH35
GH12
CE9
GH79
GH47
GH51
GH17
GH81
PL8
GH27
GH125
GH29
PL14
GH36
GH37

Table 1: Neocallimastigomycota strains analyzed in this study.

Genus	Species	Strain	Host	Isolation source	Location	LSU Genbank accession number	Reference
Anaeromyces	*contortus*	C3G	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121936	This study
Anaeromyces	*contortus*	C3J	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121942	This study
Anaeromyces	*contortus*	G3G	Goat (*Capra aegagrus hircus*)	Feces	Stillwater, OK	MF121935	This study
Anaeromyces	*contortus*	Na	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121943	This study
Anaeromyces	*contortus*	O2	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121931	This study
Anaeromyces	*robustus*	S4	Sheep (*Ovis aries*)	Feces	Santa Barbara, C.	NA*	{Haitjema, 2017 #19}
Caecomyces	sp.	Iso3	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992499	This study
Caecomyces	sp.	Brit4	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992500	This study
Feramyces	*austinii*	F2c	Aoudad sheep (*Ammotragus ler*	Feces	Stillwater, OK	MG605675	This study
Feramyces	*austinii*	F3a	Aoudad sheep (*Ammotragus ler*	Feces	Stillwater, OK	MG584226	This study
Neocallimasti	*californiae*	G1	Horse (*Equus caballus*)	Feces	Santa Barbara, C.	Genomic sequence**	{Haitjema, 2017 #19}
Neocallimasti	cf. *cameroor*	G3	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992493	This study
Neocallimasti	cf. *frontalis*	Hef5	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992494	This study
Orpinomyces	cf. *joyonii*	D3A	Cow (*Bos taurus*)	Digesta	Stillwater, OK	MG992487	This study
Orpinomyces	cf. *joyonii*	D3B	Cow (*Bos taurus*)	Digesta	Stillwater, OK	MG992488	This study
Orpinomyces	cf. *joyonii*	D4C	Cow (*Bos taurus*)	Digesta	Stillwater, OK	MG992489	This study
Pecoramyces	*ruminantium*	C1A	Cow (*Bos taurus*)	Feces	Stillwater, OK	JN939127	{Youssef, 2013 #96;Couger, 2015 #95}
Pecoramyces	*ruminantium*	S4B	Sheep (*Ovis aries*)	Feces	Stillwater, OK	KX961618	This study
Pecoramyces	*ruminantium*	FS3C	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992492	This study
Pecoramyces	*ruminantium*	FX4B	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992491	This study
Pecoramyces	*ruminantium*	YC3	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992490	This study
Piromyces	*finnis*	finn	Horse (*Equus caballus*)	Feces	Santa Barbara, C.	Genomic sequence**	{Haitjema, 2017 #19}
Piromyces	sp.	A1	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992496	This study
Piromyces	sp.	A2	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992495	This study
Piromyces	sp.	B4	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992497	This study
Piromyces	sp.	B5	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992498	This study
Piromyces	sp.	E2	Indian Elephant (*Elephas maxim*	Feces	London, UK	NA	{Haitjema, 2017 #19;Teunissen, 1991 #120}

*NA: Not available

** LSU sequence was extracted from the genomic assembly. No LSU accession number was available.