Horizontal gene transfer as an indispensable 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: Horizontal gene transfer 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 283 distinct incidents of HGT in AGF transcriptomes, with subsequent gene duplication resulting in an HGT frequency of 2.1-3.6% in AGF genomes. The majority of HGT events were AGF specific (91.5%) and wide (70.7%), 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. 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.

Importance

The anaerobic gut fungi (AGF) represent a distinct basal phylum lineage (Neocallimastigomycota) commonly encountered in the rumen and alimentary tracts of herbivores. Survival and growth of anaerobic gut fungi in these anaerobic, eutrophic, and prokaryotes dominated habitats necessitates the acquisition of several traits absent in other fungal lineages. This manuscript assesses the role of horizontal gene transfer as a relatively fast mechanism for trait acquisition by the Neocallimastigomycota post sequestration in the herbivorous gut. Analysis of twenty-seven transcriptomes that represent the broad Neocallimastigomycota diversity identified 283 distinct HGT events, with subsequent gene duplication resulting in an HGT frequency of 2.1-3.6% in AGF genomes. These HGT events have allowed AGF to survive in the herbivorous gut by expanding their substrate utilization range, augmenting their biosynthetic pathway, providing new routes for electron disposal by expanding fermentative capacities, and facilitating their adaptation to anaerobiosis. HGT in the AGF is also shown to be mainly a cross-kingdom affair, with the majority of donors belonging to the bacteria. This work represents 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 (1). 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 (2-6). In prokaryotes, the occurrence, patterns, frequency, and impact of HGT on the genomic architecture (7), metabolic abilities (8, 9), physiological preferences (10, 11), and ecological fitness (12) has been widely investigated, and the process is now regarded as a major driver of genome evolution in bacteria and archaea (13, 14). 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 (5, 15), it is now widely accepted that HGT contributes significantly to eukaryotic genome evolution (16, 17). HGT events have convincingly been documented in multiple phylogenetically disparate eukaryotes ranging from the Excavata (18-21), SAR supergroup (22-25), Algae (26), Plants (27), and Opisthokonta (28-31). Reported HGT frequency in eukaryotic genomes ranges from a handful of genes, e.g. (32), to up to 9.6% in Bdelloid rotifers (30).

The kingdom Fungi represents a phylogenetically coherent clade that evolved ≈ 900-1481 Mya from a unicellular flagellated ancestor (33-35). To date, multiple efforts have been reported on the detection and quantification of HGT in fungi. A survey of 60 fungal genomes reported HGT frequencies of 0-0.38% (29), and similar low values were observed in the genomes of five early-diverging pathogenic Microsporidia and Cryptomycota (36). The osmotrophic lifestyle of

fungi (37) has typically been regarded as less conducive to HGT compared to the phagocytic lifestyle of several microeukaryotes with relatively higher HGT frequency (38).

The anaerobic gut fungi (AGF, Neocallimastigomycota) represent a phylogenetically distinct basal fungal lineage. The AGF appear to exhibit a restricted distribution pattern, being encountered in the gut of ruminant and non-ruminant herbivorous (39). In the herbivorous gut, the life cycle of the AGF (Figure S1) involves the discharge of motile flagellated zoospores from sporangia in response to animal feeding, the chemotaxis and attachment of zoospores to ingested plant material, spore encystment, and the subsequent production of rhizoidal growth that penetrates and digests plant biomass through the production of a wide array of cellulolytic and lignocellulolytic enzymes.

Survival, colonization, and successful propagation of AGF in the herbivorous gut 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 (40, 41). 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 (39), 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 (42, 43), and preliminary BLAST-based queries of a few genomes (41, 44), 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 (45). 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). Samples were immediately transferred to the laboratory and the isolation procedures usually commenced within 24 hours of collection. A second round of isolation was occasionally conducted on samples stored at -20^0 C for several weeks (Table 1).

Isolation was performed using a rumen fluid medium reduced by cysteine-sulfide, supplemented with a mixture of kanamycin, penicillin, streptomycin, and chloramphenicol (50 µg/mL, 50 µg/mL, 20 µg/mL, and 50 µg/mL, respectively), and dispensed under a stream of 100% CO_2 (41, 46). All media were prepared according to the Hungate technique (47), as modified by Balch and Wolfe (48). Cellulose (0.5%), or a mixture of switchgrass (0.5%) and cellobiose (0.5%) were used as carbon sources. Samples were serially diluted and incubated at 39°C for 24-48 h. Colonies were obtained from dilutions showing visible signs of fungal growth using the roll tube technique (49). Colonies obtained were inoculated into liquid media, and a second round of isolation and colony picking was conducted to ensure culture purity. Microscopic examination of thallus growth pattern, rhizoid morphology, and zoospore flagellation, as well LSU rRNA gene D1-D2 domain amplification and sequencing were employed to determine the genus level affiliation of all isolates (46). The cultures were routinely

sub-cultured on rumen fluid medium supplemented with antibiotics (to guard against accidental bacterial contamination) and stored on agar media as described previously (41, 50).

Sequencing and assembly. Transcriptomic sequencing was conducted for twenty-two AGF strains. Sequencing multiple taxa provides stronger evidence for the occurrence of HGT in a target lineage (51), 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 a built-in safeguard against possible prokaryotic contamination, an issue that often plagued eukaryotic genome-based HGT detection efforts (52, 53), as well as to demonstrate that HGT genes identified are transcribed in AGF. Further, sequencing and assembly of a large number of Neocallimastigomycota genomes is challenging due to the extremely high AT content in intergenic regions and the extensive proliferation of microsatellite repeats, often necessitating employing multiple sequencing technologies for successful genomic assembly (41, 44).

RNA extraction was conducted as described previously (54). Briefly, fungal biomass was obtained by vacuum filtration and grounded with a pestle under liquid nitrogen. RNA was extracted using Epicentre MasterPure Yeast RNA Purification kit (Epicentre, Madison, WI, USA) and stored in RNase-free TE buffer. Transcriptomic sequencing using Illumina HiSeq2500 2X150bp paired end technology was conducted using the services of a commercial provider (Novogene Corporation, Beijing, China).

RNA-Seq reads were assembled by the de novo transcriptomic assembly program Trinity (55) using previously established protocols (56). All settings were implemented according to the recommended protocol for fungal genomes, with the exception of the absence of the "–

jaccard_clip" flag due to the low gene density of anaerobic fungal genomes. The assembly process was conducted on the Oklahoma State University High Performance Computing Cluster as well as the XSEDE HPC Bridges at the Pittsburg Super Computing Center. Quantitative levels for all assembled transcripts were determined using Bowtie2 (57). The program Kallisto was used for quantification and normalization of the gene expression of the transcriptomes (58). All final peptide models predicted were annotated using the Trinotate platform with a combination of homology-based search using BLAST+, domain identification using hmmscan and the Pfam 30.0 database 19 (59), and cellular localization with SignalP 4.0 (60). The twenty-two transcriptomes sequenced in this effort, as well as previously published transcriptomic datasets from *Pecoramyces ruminantium* (41), *Piromyces finnis, Piromyces* sp. E2, *Anaeromyces robustus*, and *Neocallimastix californiae* (44) were examined. In each dataset, redundant transcripts were grouped into clusters using CD-HIT-EST with identity parameter of 95% (-c 0.95). The obtained non-redundant transcripts from each analyzed transcriptome were subsequently used for peptide and coding sequence prediction using the TransDecoder with a minimum peptide length of 100 amino acids (http://transdecoder.github.io). Assessment of transcriptome coverage per strain was conducted using BUSCO (61).

HGT identification. A combination of BLAST similarity searches, comparative similarity index (HGT index, h_U), and phylogenetic analyses were conducted to identify HGT events in the analyzed transcriptomic datasets (Figure 1). We define an HGT event as the acquisition of a foreign gene/Pfam by AGF from a single lineage/donor. All predicted peptides were queried against Uniprot databases (downloaded May 2017) each containing both reviewed (Swiss-Prot) and unreviewed (TrEMBL) sequences. The databases encompassed nine different phylogenetic groups; Bacteria, Archaea, Viridiplantae, Opisthokonta-Chaonoflagellida, Opisthokonta-Fungi

(without Neocallimastigomycota representatives), Opisthokonta-Metazoa, Opisthokonta-Nucleariidae and Fonticula group, all other Opisthokonta, and all other non-Opisthokonta-non-Viridiplantae Eukaryota. For each peptide sequence, the bit score threshold and HGT index h_U (calculated as the difference between the bit-scores of the best non-fungal and the best Dikarya fungal matches) were determined. Peptide sequences that satisfied the criteria of having a BLASTP bit-score against a non-fungal database that was >100 (i.e. 2^{-100} chance of random observation) and an HGT index h_U that was ≥30 were considered HGT candidates and subjected to additional phylogenetic analysis. We chose to work with bit-score rather than the raw scores since the bit-score measures sequence similarity independent of query sequence length and database size. This is essential when comparing hits from databases with different sizes (for example, the Bacteria database contained 83 million sequences while the Choanoflagellida database contained 21 thousand sequences). We chose an h_U value of ≥30 (a difference of bit-score of at least 30 between the best non-fungal hit and the best fungal hit to an AGF sequence) previously suggested and validated (62, 63) as the best tradeoff between sensitivity and specificity. Since the bit-score is a logarithmic value that describes sequence similarity, a bit-score > 30 ensure that the sequence aligned much better to the non-fungal hit than it did to the fungal hit.

The identified HGT candidates were modified by removing all CAZyme-encoding sequences (due to their multi-modular nature, see below) and further clustered into orthologues using OrthoMCL (64). Orthologues obtained were subjected to detailed phylogenetic analysis to confirm HGT occurrence as well as to determine the potential donor. Each Orthologue was queried against the nr database using web Blastp (65) under two different settings: once against the full nr database and once against the Fungi (taxonomy ID: 4751) excluding the

Neocallimastigomycetes (Taxonomy ID: 451455). The first 250 hits obtained using these two Blastp searches with an e-value below e^{-10} were downloaded and combined in one fasta file. Datasets were reduced by removing duplicate sequences as well as redundant sequences from one organism. AGF and reference sequences were aligned using the standalone Clustal Omega (66). Alignments were viewed and manually curated in Mega (67). The alignments were used to generate guide trees in FastTree under the LG model (68). Guide trees were in turn used as input to IQ-tree (69) to generate maximum likelihood trees under the posterior mean site frequency method (PMSF), shown before to ameliorate long-branch attraction artifacts (70). Both the (-alrt 1000) option for performing the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT), as well as the (-bb 1000) option for ultrafast bootstrap (UFB) (71) were added to the IQ-tree command line. This resulted in the generation of phylogenetic trees with two support values (SH-aLRT and UFB) on each branch. Candidates that showed a nested phylogenetic affiliation that was incongruent to organismal phylogeny with strong SH-aLRT and UFB supports were deemed horizontally transferred.

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 (41, 44). A single gene can hence harbor multiple CAZyme pfams of different (fungal or non-fungal) origins (41, 44). 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. CAZyme-containing transcripts (Glycoside hydrolases (GHs), Polysaccharide lyases (PLs), and Carbohydrate Esterases (CEs)) were first identified by searching the entire transcriptomic datasets against the dbCAN hidden markov models V5 (72) (downloaded from the dbCAN web server in September 2016) using the command hmmscan in standalone HMMER. For each CAZy family identified, predicted peptides across all transcriptomic datasets were grouped in one fasta file that was then amended with the corresponding Pfam seed sequences (downloaded from the Pfam website (http://pfam.xfam.org/) in March 2017). Sequences were aligned using the standalone Clustal Omega (66) to their corresponding Pfam seeds. Using the Pfam seed sequences as a guide for the start and end of the domain, aligned sequences were then truncated in Jalview (73). Truncated transcripts with an identified CAZy domain were again compared to the pfam database (74) using hmmscan (75) to ensure correct assignment to CAZy families and accurate domain trimming. These truncated peptide sequences were then analyzed to pinpoint incidents of HGT using the approach described above.

Neocallimastigomycota-specific versus non-specific HGT events. To determine whether an identified HGT event (i.e. foreign gene acquisition from a specific donor) is specific to the phylum Neocallimastigomycota; the occurrence of orthologues (30% identity, >100 amino acids alignment) of the identified HGT genes in basal fungi, i.e. members of Blastocladiales, Chytridiomycota, Cryptomycota, Microsporidia, Mucormycota, and Zoopagomycota, as well as the putative phylogenetic affiliation of these orthologues, when encountered, were assessed. HGT events were 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 3. orthologues were identified in basal fungal

248 genomes and showed a non-fungal phylogenetic affiliation, but such affiliation was different
249 from that observed in the Neocallimastigomycota. On the other hand, events were judged to be
250 non-specific to the Neocallimastigomycota if phylogenetic analysis of basal fungal orthologues
251 indicated a non-fungal origin with a donor affiliation similar to that observed in the
252 Neocallimastigomycota (Figure 1).
253 **Mapping HGT events to available AGF genomes.** HGT events identified in AGF datasets
254 examined (both CAZy and non-CAZy events) were mapped onto currently available AGF
255 genome assemblies (41, 44) (Genbank accession numbers ASRE00000000.1,
256 MCOG00000000.1, MCFG00000000.1, MCFH00000000.1). The duplication and expansion
257 patterns, as well as GC content, and intron distribution were assessed in all identified genes.
258 Averages were compared to AGF genome average using Student t-test to identify possible
259 deviations in such characteristics as often observed with HGT genes (76). To avoid any bias the
260 differences in the number of genes compared might have on the results, we also compared the
261 GC content, codon usage, and intron distribution averages for the identified genes to a subset of
262 an equal number of randomly chosen genes from AGF genomes. We used the MEME Suite's
263 fasta-subsample function (http://meme-suite.org/doc/fasta-subsample.html) to randomly select an
264 equal number of genes from the AGF genomes.
265 **Validation of HGT-identification pipeline using previously published datasets.** As a control,
266 the frequency of HGT occurrence in the genomes of a filamentous ascomycete (*Colletotrichum*
267 *graminicola*, GenBank Assembly accession number GCA_000149035.1), and a microsporidian
268 (*Encephalitozoon hellem*, GenBank Assembly accession number GCA_000277815.3) were
269 determined using our pipeline (Table S1); and the results were compared to previously published
270 results (36, 77).

Guarding against false positive HGT events due to contamination. Multiple safeguards were taken to ensure that the frequency and incidence of HGT reported here are not due to bacterial contamination of AGF transcripts. These included: 1. Application of antiobiotics in all culturing procedures as described above. 2. Utilization of transcriptomes rather than genomes selects for eukaryotic polyadenylated (poly(A)) transcripts prior to RNA-seq as a built-in safeguard against possible prokaryotic contamination. 3. Mapping HGT transcripts identified to genomes generated in prior studies and confirming the occurrence of introns in the majority of HGT genes identified. 4. Applying a threshold where only transcripts identified in >50% of transcriptomic assemblies from a specific genus are included and 5. The exclusion of HGT events showing suspiciously high (>90%) sequence identity to donor sequences.

In addition, recent studies have demonstrated that GenBank-deposited reference genomes (52) and transcriptomes (78) of multicellular organisms are often plagued by prokaryotic contamination. The occurrence of prokaryotic contamination in reference donors' genomes/transcriptomes could lead to false positive HGT identification, or incorrect HGT assignments. To guard against any false positive HGT event identification due to possible contamination in reference datasets, sequence data from potential donor reference organisms were queried using blast, and their congruence with organismal phylogeny was considered a prerequisite for inclusion of an HGT event.

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

294 number SRP161496. Alignments, as well as Newick tree files for all HGT genes are provided as
295 Supplementary datasets. Trees of HGT events discussed in the results and discussion sections are
296 presented in the supplementary document (S5-S45).

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) (79) (Table 1, Supplementary Fig. 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* (80)), or appears to exhibit an extremely limited geographic and animal host distribution (genus *Buwchfawromyces* (81)). The third unrepresented genus (*Cyllamyces*) has recently been suggested to be phylogenetically synonymous with *Caecomyces* (82). 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 (61) yielded high completion values (82.76-97.24%) for all assemblies (Table S1). For strains with a sequenced genome, 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 12,786 orthologues with a non-fungal bit score > 100, and an HGT index > 30 were identified. After removing orthologues occurring only in a single strain or in less than 50% of isolates belonging to the same genus, 2147 events were further evaluated. Phylogenetic analysis could not confirm the HGT nature (e.g. single long branch that could either be attributed to HGT or gene loss in all other fungi, unstable phylogeny, and/or low bootstrap) of 1846 orthologues and so were subsequently removed. Of the remaining 291 orthologues, 8 had

suspiciously high (>90%) first hit amino acid identity. Although the relatively recent divergence and/or acquisition time could explain this high level of similarity, we opted to remove these orthologues as a safeguard against possible bacterial contamination of the transcriptomes. Ultimately, a total of 283 distinct HGT events that satisfied the criteria described above for HGT were identified (Table S3). The average number of events per genus was 223±13 and ranged between 210 in the genus *Orpinomyces* to 242 in the genus *Pecoramyces* pantranscriptomes (Fig. 2A). The majority of HGT acquisition events identified (259, 91.52%) appear to be Neocallimastigomycota-specific, i.e. identified only in genomes belonging to the Neocallimastigomycota, but not in other basal fungal genomes (Table S4), strongly suggesting that such acquisitions occurred post, or concurrent with, the evolution of Neocallimastigomycota as a distinct fungal lineage. As well, the majority of these identified genes were Neocallimastigomycota-wide, being identified in strains belonging to at least six out of the seven examined genera (200 events, 70.7%), suggesting the acquisition of such genes prior to genus level diversification within the Neocallimastigomycota. Only 33 events (11.7%) were genus-specific, with the remainder (50 events, 17.7%) being identified in the transcriptomes of 3-5 genera (Table S4, Figure S4, and Fig. 2b).

The absolute majority (89%) of events were successfully mapped to at least one of the four AGF genomes (Table S5), with a fraction (7/31) of the unmapped transcripts being specific to a genus with no genome representative (*Feramyces*, *Caecomyces*). Compared to a random subset of 283 genes in each of the sequenced genomes, horizontally transferred genes in AGF genomes exhibited significantly ($P<0.0001$) fewer introns (1.1±031 vs 3.32±0.83), as well as higher GC content (31±4.5 vs 27.7±5.5) (Table S5). Further, HGT genes/pfams often displayed high levels of gene/ pfam duplication and expansion within the genome (Table S5), resulting in

an HGT frequency of 2.13% in *Pecoramyces ruminantium* (348 HGT genes out of 16,347 total genes), 3.07% in *Piromyces finnis* (352 HGT genes out of 11,477 total genes), 3.27% in *Anaeromyces robustus* (423 HGT genes out of 12,939 total genes), and 3.60% in *Neocallimastix californiae* (753 HGT genes out of 20,939 total genes).

Donors. A bacterial origin was identified for the majority of HGT events (84.8%), with four bacterial phyla (Firmicutes, Proteobacteria, Bacteroidetes, and Spirochaetes) identified as donors for 177 events (62.5% of total, 73.8% of bacterial events) (Fig. 3A). Specifically, the contribution of members of the Firmicutes (125 events) was paramount, the majority of which were most closely affiliated with members of the order Clostridiales (106 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 (83, 84). Archaeal contributions to HGT were extremely rare (6 events). On the other hand, multiple (34) events with eukaryotic donors were identified. 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 S4).

Metabolic characterization. Functional annotation of HGT genes/pfams indicated that the majority (63.96%) of events encode metabolic functions such as extracellular polysaccharide degradation and central metabolic processes. Bacterial donors were slightly overrepresented in metabolic HGT events (87.3% of the metabolism-related events, compared to 84.8% of the total events). Genes involved in cellular processes and signaling represent the second most represented HGT events (10.95%), while genes involved in information storage and processing

only made up 4.95% 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 S4, Figs S5-S16). 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), and galactose-1-phosphate uridylyltransferase) 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 (85), 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. S17-S30). 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 (86)). 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.

411 ***Adaptation to the host environment.*** Horizontal gene transfer also appears to have provided
412 means of guarding against toxic levels of compounds known to occur in the host animal gut (Fig.
413 S31-S37). For example, methylglyoxal, a reactive electrophilic species (87), is inevitably
414 produced by ruminal bacteria from dihydroxyacetone phosphate when experiencing growth
415 conditions with excess sugar and limiting nitrogen (88). Genes encoding enzymes mediating
416 methylglyoxal conversion to D-lactate (glyoxalase I and glyoxalase II-encoding genes) appear to
417 be acquired via HGT in AGF. Further, HGT allowed several means of adaptation to
418 anaerobiosis. These include: 1) acquisition of the oxygen-sensitive ribonucleoside-triphosphate
419 reductase class III (Fig. 5D) that is known to only function during anaerobiosis to convert
420 ribonucleotides to deoxyribonucleotides (89), 2) acquisition of squalene-hopene cyclase, which
421 catalyzes the cyclization of squalene into hopene, an essential step in biosynthesis of the cell
422 membrane steroid tetrahymanol that replaced the molecular O_2-requiring ergosterol in the cell
423 membranes of AGF, 3) acquisition of several enzymes in the oxidative stress machinery
424 including Fe/Mn superoxide dismutase, glutathione peroxidase, rubredoxin/rubrerythrin, and
425 alkylhydroperoxidase.

426 In addition to anaerobiosis, multiple horizontally transferred general stress and repair
427 enzymes were identified (Fig. S38-S45). HGT-acquired genes encoding 2-phosphoglycolate
428 phosphatase, known to metabolize the 2-phosphoglycolate produced in the repair of DNA lesions
429 induced by oxidative stress (90) to glycolate, were identified in all AGF transcriptomes studied
430 (Fig. 4, Table S4). Surprisingly, two genes encoding antibiotic resistance enzymes,
431 chloramphenicol acetyltransferase and aminoglycoside phosphotransferase, were identified in all
432 AGF transcriptomes, presumably to improve its fitness in the eutrophic rumen habitat that
433 harbors antibiotic-producing prokaryotes (Table S4). 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 (91, 92). 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 S4).

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 CAZymes 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 75 events (26.5% of total HGT events) were identified, with 40% occurring in all AGF genera examined (Fig. 6, Table S4). In 36.5% of GH families, 37.5% of CE families, and 28.6% of PL families, a single event (i.e. attributed to one donor) was observed (Fig. 6, Table S4).

Duplication of these events in AGF genomes was notable, with 134, 322, 161, and 135 copies of

457 HGT CAZyme pfams identified in *Anaeromyces*, *Neocallimastix*, *Piromyces* and *Pecoramyces*
458 genomes, representing 34.1%, 38.2%, 41.7%, and 25.6% of the overall organismal CAZyme
459 machinery (Table S5). The contribution of Viridiplantae, Fibrobacteres, and Gamma-
460 Proteobacteria was either exclusive to CAZyme-related HGT events or significantly higher in
461 CAZyme, compared to other, events (Fig. 3A).
462 Transcripts acquired by HGT represented >50% of transcripts in anywhere between 13
463 (*Caecomyces*) to 20 (*Anaeromyces*) GH families; 3 (*Caecomyces*) to 5 (*Anaeromyces*,
464 *Neocallimastix*, *Orpinomyces*, and *Feramyces*) CE families; and 2 (*Caecomyces* and *Feramyces*)
465 to 3 (*Anaeromyces*, *Pecoramyces*, *Piromyces*, *Neocallimastix*, and *Orpinomyces*) PL families
466 (Fig. 6). It is important to note that in all these families, multiple transcripts appeared to be of
467 bacterial origin based on BLAST similarity search but did not meet the strict criteria
468 implemented for HGT determination in this study. As such, the contribution of HGT transcripts
469 to overall transcripts in these families is probably an underestimate. Only GH9, GH20, GH37,
470 GH45, and PL3 families appear to lack any detectable HGT events. A PCA biplot comparing
471 CAZyomes in AGF genomes to other basal fungal lineages strongly suggests that the acquisition
472 and expansion of many of these foreign genes play an important role in shaping the
473 lignocellulolytic machinery of AGF (Fig. 7). The majority of CAZyme families defining AGF
474 CAZyome were predominantly of non-fungal origin (Fig. 7). This pattern clearly attests to the
475 value of HGT in shaping AGF CAZyome via acquisition and extensive duplication of acquired
476 gene families.
477 Collectively, HGT had a profound impact on AGF plant biomass degradation
478 capabilities. The AGF CAZyome encodes enzymes putatively mediating the degradation of
479 twelve different polysaccharides (Fig. S46). In all instances, GH and PL families with >50%

480 horizontally transferred transcripts contributed to backbone cleavage of these polymers; although
481 in many polymers, e.g. cellulose, glucoarabinoxylan, and rhamnogalactouronan, multiple
482 different GHs can contribute to backbone cleavage. Similarly, GH, CE, and PL families with
483 >50% horizontally transferred transcripts contributed to 10 out of 13 side-chain-cleaving
484 activities, and 3 out of 5 oligomer-to-monomer breakdown activities (Fig. S46).

Discussion

Here, we present a systematic analysis of HGT patterns in 27 transcriptomes and 4 genomes belonging to the Neocallimastigomycota. Our analysis identified 283 events, representing 2.1-3.6% of genes in examined AGF genomes. Further, we consider these values to be conservative estimates due to the highly stringent criteria and 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.

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 (16, 23, 93, 94). 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 (95-97). Finally, AGF release asexual motile free zoospores into the herbivorous gut as part of their life cycle (39). According to the weak-link model (98), 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 (50).

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 (70.7%) 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. S46). 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 S4), 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 S4).

A smaller number of observed events (n=33) were genus-specific (Fig. 2, Table S4). This group was characterized by being significantly enriched in CAZymes (60.6% of genus-specific horizontally transferred events have a predicted CAZyme function, as opposed to 26.5% in the overall HGT dataset), and being almost exclusively acquired from donors that are known to

inhabit the herbivorous gut (99) (26 out of the 33 events were acquired from the orders Clostridiales, Bacillales, and Negativicutes within Firmicutes, Burkholderiales within the Beta-Proteobacteria, Flavobacteriales and Bacteroidales within Bacteroidetes, and the Spirochaetes, Actinobacteria, and Lentisphaerae), or from Viridiplantae (4 out of the 33 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.

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 (22). 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 (100-102). 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

554 of the gut microbiome and survived in locations with relatively higher oxygen concentration in
555 the alimentary tract e.g. mouth, saliva, esophagus or in micro-niches in the rumen where
556 transient oxygen exposure occurs. Subsequently, HGT acquisition has enabled the expansion of
557 their niche, improved their competitiveness and their relative abundance in the herbivorous gut to
558 the current levels.

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

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

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

575 **References:**

576 1. Doolittle WF. 1999. Lateral Genomics. Trends Cell Biol 9:M5-M8.
577 2. Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and
578 distinguishing between models. Nat Rev Genet 11:97-10.
579 3. Cai J, Zhao R, Jiang H, Wang W. 2008. De Novo origination of a new protein-coding
580 gene in Saccharomyces cerevisiae. Genetics 179:487-496.
581 4. Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes2. Genome
582 Res 20:1313-1326.
583 5. Andersson DI, Jerlström-Hultqvist J, Näsvall J. 2015. Evolution of new functions de
584 novo and from preexisting genes. Cold Spring Harb Perspect Biol 7:a017996.
585 6. Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N,
586 Charloteaux B, Hidalgo CA, Barbette J, Santhanam B, Brar GA, Weissman JS, Regev A,
587 Thierry-Mieg N, Cusick ME, Vidal M. 2010. Proto-genes and de novo gene birth. Nature
588 487:370-374.
589 7. Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of
590 bacterial innovation. Nature 405:299-304.
591 8. Caro-Quintero A, Konstantinidis K. 2015. Inter-phylum HGT has shaped the metabolism
592 of many mesophilic and anaerobic bacteria. ISME J 9:958-967.
593 9. Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS. 2015. Insights
594 into the metabolism, lifestyle and putative evolutionary history of the novel archaeal
595 phylum 'Diapherotrites'. ISME J 9:447-460.
596 10. Puigbo P, Pasamontes A, Garcia-Vallve S. 2008. Gaining and losing the thermophilic
597 adaptation in prokaryotes. Trends Genet 24:10-14.
598 11. Omelchenko MV, Wolf YI, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M,
599 Daly MJ, Koonin EV, Makarova KS. 2005. Comparative genomics of Thermus
600 thermophilus and Deinococcus radiodurans: divergent routes of adaptation to
601 thermophily and radiation resistance BMC Evol Biol 5:57.
602 12. Wiedenbeck J, Cohan FM. 2011. Origins of bacterial diversity through horizontal genetic
603 transfer and adaptation to new ecological niches. FEMS Microbiol Rev 35:957-976.
604 13. Syvanen M. 2012. Evolutionary implications of horizontal gene transfer. Annu Rev
605 Genet 46:341-358.
606 14. Philippe H, Douady CJ. 2003. Horizontal gene transfer and phylogenetics. Curr Opin
607 Microbiol 6:498-505.
608 15. Fitzpatrick DA. 2012. Horizontal gene transfer in fungi. FEMS Microbiol Lett 2011:1-8.
609 16. Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev
610 Genet 9:605-618.
611 17. Husnik F, McCutcheon JP. 2017. Functional horizontal gene transfer from bacteria to
612 eukaryotes. Nat Rev Microbiol doi:10.1038/nrmicro.2017.137.
613 18. Qian Q, Keeling PJ. 2001. Diplonemid glyceraldehyde–3-phosphate dehydrogenase
614 (GAPDH) and prokaryoteto-eukaryote lateral gene transfer. Protist 152:193–201.
615 19. Hirt RP, Harriman N, Kajava AV, Embley TM. 2002. A novel potential surface protein in
616 Trichomonas vaginalis contains a leucine-rich repeat shared by micro-organisms from all
617 three domains of life. Mol Biochem Parasitol 125:195–199.
618 20. Nixon JEJ, Wang A, Field J, Morrison HG, McArthur AG, Sogin ML, al. e. 2002.
619 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.

21. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, al. e. 2005. The genome of the social amoeba Dictyostelium discoideum. Nature 435:43-57.
22. Eme L, Gentekaki E, Curtis B, Archibald JM, Roger AJ. 2017. Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite Blastocystis to the Gut. Curr Biol 27:807-820.
23. Ricard G, McEwan NR, Dutilh BE, Jouany J-P, Macheboeuf D, Mitsumori M, McIntosh FM, Michalowski T, Nagamine T, Nelson N, Newbold CJ, Nsabimana E, A.Takenaka, Thomas NA, Ushida K, Hackstein JH, Huynen MA. 2006. Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. BMC Genomics 7:22.
24. Kishore SP, Stiller JW, Deitsch KW. 2013. Horizontal gene transfer of epigenetic machinery and evolution of parasitism in the malaria parasite Plasmodium falciparum and other apicomplexans. BMC Evol Biol 13:37.
25. Wisecaver JH, Brosnahan ML, Hackett JD. 2013. Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. Genome Biol Evol 5:2368-2381.
26. Schönknecht G, Weber AP, Lercher MJ. 2013. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. Bioassays 36:9-20.
27. Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. J Exp Bot 58:1-9.
28. Sun GL, Yang ZF, Ishwar A, Huang JL. 2010. Algal genes in the closest relatives of animals. Mol Biol Evol 27:2879-2889.
29. Marcet-Bouben M, Gabaldon T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet 26:5-8.
30. Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid rotifers. Science 320:1210-1213.
31. Danchin EG, Rosso MN, Vieira P, Almeida-Engler Jd, Coutinho PM, Henrissat B, Abad P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc Nat Acad Sci USA 107:17651-17656.
32. McCarthy CGP, Fitzpatrick DA. 2016. Systematic Search for Evidence of Interdomain Horizontal Gene Transfer from Prokaryotes to Oomycete Lineages. mSphere 1:e00195-16.
33. Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? . Proc Nat Acad Sci USA, 101:15386-15391.
34. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks Proc Nat Acad Sci USA 108:13624- 13629.
35. Taylor JW, Berbee ML. 2006. Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia 98:838-49.
36. Alexander WG, Wisecaver JH, Rokas A, Httinger CT. 2016. Horizontally aquired gene in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. Proc Nat Acad Sci USA 113:4116-4121.
37. Berbee ML, James TY, Strullu-Derrien C. 2017. Early Diverging Fungi: Diversity and Impact at the Dawn of Terrestrial Life. Annu Rev Microbiol 71:41-60.
38. Doolittle WF. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 14:307–311.

666 39. Gruninger RJ, Puniyab AK, Callaghanc TM, Edwardsc JE, Youssef N, Dagare SS,
667 Fliegerova K, Griffith GW, Forster R, Tsang A, McAllister T, Elshahed MS. 2014.
668 Anaerobic Fungi (Phylum Neocallimastigomycota): Advances in understanding of their
669 taxonomy, life cycle, ecology, role, and biotechnological potential. FEMS Microbiol Ecol
670 Under Review.
671 40. Boxma B, Voncken F, Jannink S, Alen TV, Akhmanova A, Weelden SWHV, Hellemond
672 JJV, Ricard G, Huynen M, Tielens AGM, Hackstein JHP. 2004. The anaerobic
673 chytridiomycete fungus Piromyces sp. E2 produces ethanol via pyruvate:formate lyase
674 and an alcohol dehydrogenase E Mol Microbiol 51:1389 -1399.
675 41. Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ,
676 Atiyeh HK, Wilkins MR, Elshahed MS. 2013. Genome of the anaerobic fungus
677 Orpinomyces sp. C1A reveals the unique evolutionary history of a remarkable plant
678 biomass degrader Appl Environ Microbiol 79:4620-4634.
679 42. Harhangi HR, Akhmanova AS, Emmens R, Drift Cvd, Laat WTAMd, Dijken JPv, Jetten
680 MSM, Pronk JT, Camp HJMOd. 2003. Xylose metabolism in the anaerobic fungus
681 Piromyces sp. strain E2 follows the bacterial pathway. Arch Microbiol 180:134-142.
682 43. Garcia-Vallvé S, Romeu A, Palau J. 2000. Horizontal gene transfer of glycosyl
683 hydrolases of the rumen Fungi. Mol Biol Evol 17:352-361.
684 44. Haitjema CH, Gilmore SP, Henske JK, Solomon KV, Groot Rd, Kuo A, Mondo SJ,
685 Salamov AA, LaButti K, Zhao Z, Chiniquy J, Barry K, Brewer HM, Purvine SO, Wright
686 AT, Hainaut M, Boxma B, Alen Tv, Hackstein JHP, Henrissat B, Baker SE, Grigoriev IV,
687 O'Malley MA. 2017. A parts list for fungal cellulosomes revealed by comparative
688 genomics. Nature Microbiol 2:17087.
689 45. Ho YW, Barr DJS. 1995. Classification of anaerobic gut fungi from herbivores with
690 emphasis on rumen fungi from malaysia. mycologia 87:655-677.
691 46. Hanafy RA, Elshahed MS, Liggenstoffer AS, Griffith GW, Youssef NH. 2017.
692 Pecoramyces ruminantium, gen. nov., sp. nov., an anaerobic gut fungus from the feces of
693 cattle and sheep. Mycologia 109:231-243.
694 47. Bryant M. 1972. Commentary on the Hungate technique for culture of anaerobic bacteria.
695 Am J Clin Nutr 25:1324-1328.
696 48. Balch WE, Wolfe R. 1976. New approach to the cultivation of methanogenic bacteria: 2-
697 mercaptoethanesulfonic acid (HS-CoM)-dependent growth of Methanobacterium
698 ruminantium in a pressureized atmosphere. Appl Environ Microbiol 32:781-791.
699 49. Hungate RE. 1969. A roll tube method for cultivation of strict anaerobes. Meth Microbiol
700 3:117-132.
701 50. Calkins S, Elledge NC, Hanafy RA, Elshahed MS, Youssef NH. 2016. A fast and reliable
702 procedure for spore collection from anaerobic fungi: Application for RNA uptake and
703 long-term storage of isolates. J Microbiol Methods 127:206-213.
704 51. Richards TA, Monier A. 2017. A tale of two tradigrades. Proc Nat Acad Sci USA
705 113:4892-4894.
706 52. Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, Tintori SC,
707 Li Q, Jones CD, Yandell M, Messina DN, Glasscock J, Goldstein B. 2015. Evidence for
708 extensive horizontal gene transfer from the draft genome of a tardigrade. Proc Nat Acad
709 Sci USA 112:15976-15981.
710 53. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H,
711 Thomas F, Aboobaker AA, Blaxter M. 2016. No evidence for extensive horizontal gene

transfer in the genome of the tardigrade Hypsibius dujardini. Proc Nat Acad Sci USA 113:5053-5058.

54. Couger MB, Youssef NH, Struchtemeyer CG, Liggenstoffer AS, Elshahed MS. 2015. Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal isolate Orpinomyces sp. strain C1A. Bitechnol for Biofuels 8:208.

55. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma Fd, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. . Nat Biotechnol 29:644–652.

56. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger M, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols 8:1494-1512.

57. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.

58. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525-527.

59. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. 44:D279-D285.

60. Petersen TN, Brunak S, Heijne Gv, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786.

61. 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.

62. Boschetti C, Carr A, Crisp A, Eyres I, Wang-Koh Y, Lubzens E, Barraclough TG, Micklem G, Tunnacliffe A. 2012. Biochemical diversification through foreign gene expression in bdelloid rotifers. PLOS Genet 8:e1003035.

63. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. 2015. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biology 16:50.

64. Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Research 13:2178-2189.

65. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

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

67. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-4.

68. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

69. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32:268-274.

70. Wang H-C, Minh BQ, Susko E, Roger AJ. 2018. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Systematic Biology 67:216-235.
71. Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30:1188-95.
72. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 40:W445-W451.
73. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189-91.
74. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222-30.
75. Prakash A, Jeffryes M, Bateman A, Finn RD. 2017. The HMMER Web Server for Protein Sequence Similarity Search. Curr Protoc Bioinformatics 60:3.15.1-3.15.23.
76. Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. Nat Rev Genet 16:472-82.
77. 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.
78. Stairs CW, Eme L, Muñoz-Gómez SA, Cohen A, Dellaire G, Shepherd JN, Fawcett JP, Roger AJ. 2018. Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. eLife 7:e34292.
79. Hanafy RA, Elshahed MS, Youssef NH. 2018. Feramyces austinii, gen. nov, sp. nov., an anaerobic gut fungus from rumen and fecal samples of wild Barbary sheep and fallow deer Submitted.
80. Dagar SS, Kumar S, Griffith GW, Edwards JE, Callaghan TM, Singh R, Nagpal AK, Puniya AK. 2015. A new anaerobic fungus (Oontomyces anksri gen. nov., sp. nov.) from the digestive tract of the Indian camel (Camelus dromedarius). Fungal Biol 19:731-737.
81. Callaghan TM, Podmirseg SM, Hohlweck D, Edwards JE, Puniya AK, Dagar SS, Griffith GW. 2015. Buwchfawromyces eastonii gen. nov., sp. nov.: a new anaerobic fungus (Neocallimastigomycota) isolated from buffalo faeces. Mycokeys 9:11-28.
82. Wang X, Liu X, Groenewald JZ. 2017. Phylogeny of anaerobic fungi (phylum Neocallimastigomycota), with contributions from yak in China. Antonie Van Leeuwenhoek 110:87-103.
83. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nature Communications 9:870.
84. He J, Yi L, Hai L, Ming L, Gao W, Ji R. 2018. Characterizing the bacterial microbiota in different gastrointestinal tract segments of the Bactrian camel. Sci Rep 8:654.
85. Leger MM, Eme L, Hug LA, Roger AJ. 2016. Novel Hydrogenosomes in the Microaerophilic Jakobid Stygiella incarcerata. Molecular Biology and Evolution 33:2318-2336.

803 86. Lin H, Kwan AL, Dutcher SK. 2010. Synthesizing and Salvaging NAD+: Lessons
804 Learned from Chlamydomonas reinhardtii. PLOS Genetics 6:e1001105.
805 87. Lee C, Park C. 2017. Bacterial Responses to Glyoxal and Methylglyoxal: Reactive
806 Electrophilic Species. International Journal of Molecular Sciences 18:169.
807 88. Russell JB. 1993. Glucose toxicity in Prevotella ruminicola: methylglyoxal accumulation
808 and its effect on membrane physiology. Applied and Environmental Microbiology
809 59:2844-2850.
810 89. Jordan A, Reichard P. 1998. Ribonucleotide Reductases. Annual Review of Biochemistry
811 67:71-98.
812 90. Pellicer MT, Nuñez MF, Aguilar J, Badia J, Baldoma L. 2003. Role of 2-
813 Phosphoglycolate Phosphatase of Escherichia coli in Metabolism of the 2-
814 Phosphoglycolate Formed in DNA Repair. Journal of Bacteriology 185:5815-5821.
815 91. Rooke DM, Shattock RC. 1983. Effect of Chloramphenicol and Streptomycin on
816 Developmental Stages of Phytophthom infestans. Microbiology 129:3401-3410.
817 92. Bishop PJ, Speare R, Poulter R, Butler M, Speare BJ, Hyatt A, Olsen V, Haigh A. 2009.
818 Elimination of the amphibian chytrid fungus Batrachochytrium dendrobatidis by Archey's
819 frog Leiopelma archeyi. Dis Aquat Organ 84:9-15.
820 93. de Koning AP, Brinkman FS, Jones SJ, Keeling PJ. 2000. Lateral gene transfer and
821 metabolic adaptation in the human parasite Trichomonas vaginalis. . Mol Biol Evol
822 17:1769-1773.
823 94. Schönknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Bräutigam
824 A, Baker BJ, Banfield JF, Garavito RM, Carr K, Wilkerson C, Rensing SA, Gagneul D,
825 Dickenson NE, Oesterhelt C, Lercher MJ, Weber AP. 2013. Gene transfer from bacteria
826 and archaea facilitated evolution of an extremophilic eukaryote. Science 339:1207-1210.
827 95. Shterzer N, Mizrahi I. 2015. The animal gut as a melting pot for horizontal gene transfer.
828 Can J Microbiol 61:603-605.
829 96. Moliner C, Fournier PE, Raoult D. 2010. Genome analysis of microorganisms living in
830 amoebae reveals a melting pot of evolution FEMS Microbiol Rev 34:281-294
831 97. Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proc
832 Nat Acad Sci USA 102:14332-14337.
833 98. Huang JL. 2013. Horizontal gene transfer in eukaryotes: the weak-link model. Bioassays
834 35:868-875.
835 99. Creevey CJ, Kelly WJ, Henderson G, Leahy SC. 2014. Determining the culturability of
836 the rumen bacterial microbiome. Microbial Biotechnology 7:467-479.
837 100. Andersson JO, Sjögren AM, Davis LA, Embley TM, Roger AJ. 2003. Phylogenetic
838 analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes.
839 Curr Biol 13:94-104.
840 101. Eme L, Gentekaki E, Curtis B, Archibald J, Roger AJ. 2017. Lateral gene transfer in the
841 adaptation of the anaerobic parasite Blastocystis to the gut. Curr Biol 27: 807–820
842 102. Grant JR, Katz LA. 2014. Phylogenomic study indicates widespread lateral gene transfer
843 in Entamoeba and suggests a past intimate relationship with parabasalids. Genome Biol
844 Evol 6:2350-2360.
845 103. Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ,
846 Atiyeh HK, Wilkins MR, Elshahed MS. 2013. The genome of the anaerobic fungus
847 Orpinomyces sp. strain C1A reveals the unique evolutionary history of a remarkable
848 plant biomass degrader. Appl Environ Microbiol 79:4620-34.

849 104. Couger MB, Youssef NH, Struchtemeyer CG, Liggenstoffer AS, Elshahed MS. 2015.
850 Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal
851 isolate Orpinomyces sp. strain C1A. Biotechnol Biofuels 8:208.
852 105. Teunissen MJ, Op den Camp HJ, Orpin CG, Huis in 't Veld JH, Vogels GD. 1991.
853 Comparison of growth characteristics of anaerobic fungi isolated from ruminant and non-
854 ruminant herbivores during cultivation in a defined medium. J Gen Microbiol 137:1401-8.
855 106. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK,
856 Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N,
857 Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017.
858 CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.
859 Nucleic Acids Research 45:D200-D203.
860 107. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and
861 interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109-14.
862

863

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, CA	NA*	(44)
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 lervia*)	Feces	Stillwater, OK	MG605675	This study
Feramyces	*austinii*	F3a	Aoudad sheep (*Ammotragus lervia*)	Feces	Stillwater, OK	MG584226	This study
Neocallimastix	*californiae*	G1	Goat (*Capra aegagrus hircus*)	Feces	Santa Barbara, CA	Genomic sequence**	(44)
Neocallimastix	cf. *cameroonii*	G3	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992493	This study
Neocallimastix	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	(103, 104)
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, CA	Genomic sequence**	(44)
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 maximus*)	Feces	London, UK	NA	(44, 105)

*NA: Not available

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

Figure Legends

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

Figure 2. (A) Distribution pattern of HGT events in AGF transcriptomes demonstrating that the majority of events were Neocallimastigomycota-wide i.e. identified in all seven AGF genera examined. (B) Total Number of HGT events identified per AGF genus.

Figure 3. Identity of HGT donors and their contribution to the various functional classes. The X-axis shows the absolute number of events belonging to each of the functional classes shown in the legend. The tree is intended to show the relationship between the donors' taxa and is not drawn to scale. Bacterial donors are shown with red branches depicting the phylum-level, with the exception of Firmicutes and Bacteroidetes donors, where the order-level is shown, and Proteobacteria, where the class-level is shown. Archaeal donors are shown with green branches and all belonged to the Methanobacteriales order of Euryarchaeota. Eukaryotic donors are shown with blue branches. Only the 230 events from a definitive-taxon donor are shown in the figure. The other 53 events were clearly nested within a non-fungal clade, but a definitive donor taxon could not be ascertained. Functional classification of the HGT events, determined by searching the Conserved Domain server (106) against the COG database are shown in B. For events with no COG classification, a search against the KEGG orthology database (107) was performed. For the major COG/KEGG categories (metabolism, cellular processes and signaling, and Information storage and processing), sub-classifications are shown in C, D, and E, respectively.

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

887 highlighted in grey, pathways for lipid metabolism are highlighted in orange, fermentation
888 pathways are highlighted in purple, while pathways for detoxification are highlighted in brown.
889 The double black lines depict the hydrogenosomal outer and inner membrane. Arrows
890 corresponding to enzymes encoded by horizontally transferred transcripts are shown with thicker
891 dotted lines and are given numbers 1 through 48 as follows. Sugar metabolism (1-11): 1. Xylose
892 isomerase, 2. Xylulokinase, 3. Ribokinase, 4. 2,3-bisphosphoglycerate-independent
893 phosphoglycerate mutase, 5. 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase, 6.
894 Phosphoenolpyruvate synthase, 7. Phosphoenolpyruvate carboxykinase (GTP), 8. Aldose-1-
895 epimerase, 9. Galactokinase, 10. Galactose-1-phosphate uridyltransferase. Amino acid
896 metabolism (11-19): 11. Aspartate-ammonia ligase, 12. Tryptophan synthase (TrpB), 13.
897 Tryptophanase, 14. Monofunctional prephenate dehydratase, 15. Serine-O-acetyltransferase, 16.
898 Cysteine synthase, 17. Low-specificity threonine aldolase, 18. 5'-methylthioadenosine
899 nucleosidase/5'-methylthioadenosine phosphorylase (MTA phosphorylase), 19. Arginase.
900 Cofactor metabolism (20-27): 20. Pyridoxamine 5'-phosphate oxidase, 21. L-aspartate oxidase
901 (NadB), 22. Quinolate synthase (NadA), 23. NH(3)-dependent NAD(+) synthetase (NadE), 24.
902 2-dehydropantoate 2-reductase, 25. dephosphoCoA kinase, 26. Dihydrofolate reductase (DHFR)
903 family, 27. Dihydropteroate synthase. Nucleotide metabolism (28-35): 28. GMP reductase, 29.
904 Trifunctional nucleotide phosphoesterase, 30. deoxyribose-phosphate aldolase (DeoC), 31.
905 Oxygen-sensitive ribonucleoside-triphosphate reductase class III (NrdD), 32.
906 nucleoside/nucleotide kinase family protein, 33. Cytidylate kinase-like family, 34. thymidylate
907 synthase, 35. thymidine kinase. Pyruvate metabolism (fermentation pathways) (36-40): 36. D-
908 lactate dehydrogenase, 37. bifunctional aldehyde/alcohol dehydrogenase family of Fe-alcohol
909 dehydrogenase, 38. Butanol dehydrogenase family of Fe-alcohol dehydrogenase, 39. Zn-type

910 alcohol dehydrogenase, 40. Fe-only hydrogenase. Detoxification reactions (41-44): 41.
911 Phosphoglycolate phosphatase, 42. Glyoxal reductase, 43. Glyoxalase I, 44. Glyoxalase II. Lipid
912 metabolism (45-47): 45. CDP-diacylglycerol--serine O-phosphatidyltransferase, 46.
913 lysophospholipid acyltransferase LPEAT, 47. methylene-fatty-acyl-phospholipid synthase.
914 Abbreviations: CDP-DAG, CDP-diacylglycerol; 7,8 DHF, 7,8 dihydrofolate; EthA,
915 ethanolamine; Gal, galactose; GAP, glyceraldehyde-3-P; Glu, glucose; GSH, glutathione; I,
916 complex I NADH dehydrogenase; NaMN, Nicotinate D-ribonucleotide; Orn, ornithine; PEP,
917 phosphoenol pyruvate; Phenyl-pyr, phenylpyruvate; PRPP, phosphoribosyl-pyrophosphate; Ptd,
918 phosphatidyl; SAM; S-adenosylmethionine; THF, tetrahydrofolate.
919 **Figure 5.** (A) Maximum likelihood tree showing the phylogenetic affiliation of AGF
920 galactokinase. AGF genes highlighted in light blue clustered within the Flavobacteriales order of
921 the Bacteroidetes phylum and were clearly nested within the bacterial domain (highlighted in
922 green) attesting to their non-fungal origin. Fungal galactokinase representatives are highlighted
923 in pink. (B) Maximum likelihood tree showing the phylogenetic affiliation of AGF Fe-only
924 hydrogenase. AGF genes highlighted in light blue clustered within the Thermotogae phylum and
925 were clearly nested within the bacterial domain (highlighted in green) attesting to their non-
926 fungal origin. *Stygiella incarcerata* (anaerobic Jakobidae) clustered with the Thermotogae as
927 well, as has recently been suggested (85). Fe-only hydrogenases from *Gonopodya prolifera*
928 (Chytridiomycota) (shown in orange text) clustered with the AGF genes. This is an example of
929 one of the rare occasions (n=24) where a non-AGF basal fungal representative showed an HGT
930 pattern with the same donor affiliation as the Neocallimastigomycota. Other basal fungal Fe-only
931 hydrogenase representatives are highlighted in pink and clustered outside the bacterial domain.
932 (C) Maximum likelihood tree showing the phylogenetic affiliation of AGF L-aspartate oxidase

933 (NadB). AGF genes highlighted in light blue clustered within the Delta-Proteobacteria class and
934 were clearly nested within the bacterial domain (highlighted in green) attesting to their non-
935 fungal origin. As de-novo NAD synthesis in fungi usually follow the five-enzyme pathway
936 starting from tryptophan, as opposed to the two-enzyme pathway from aspartate, no NadB were
937 found in non-AGF fungi and hence no fungal cluster is shown in the tree. (D) Maximum
938 likelihood tree showing the phylogenetic affiliation of AGF oxygen-sensitive ribonucleotide
939 reductase (NrdD). AGF genes highlighted in light blue clustered with representatives from
940 Candidate phylum Dependentiae and were clearly nested within the bacterial domain
941 (highlighted in green) attesting to their non-fungal origin. Fungal NrdD representatives are
942 highlighted in pink. GenBank accession numbers are shown in parentheses. Alignment was done
943 using the standalone Clustal Omega (66) and trees were constructed using IQ-tree (69).
944 **Figure 6**. HGT in the AGF CAZyome shown across the seven genera studied. Glycosyl
945 Hydrolase (GH), Carboxyl Esterase (CE), and Polysaccharide Lyase (PL) families are shown to
946 the left. The color of the cells depicts the prevalence of HGT within each family. Red indicates
947 that 100% of the CAZyme transcripts were horizontally transferred. Shades of red-orange
948 indicate that HGT contributed to > 50% of the transcripts belonging to that CAZy family. Blue
949 indicates that 100% of the CAZyme transcripts were of fungal origin. Shades of blue indicate
950 that HGT contributed to < 50% of the transcripts belonging to that CAZy family. The numbers in
951 each cell indicate the affiliation of the HGT donor as shown in the key to the right.
952 **Figure 7**. Principal-component analysis biplot of the distribution of CAZy families in AGF
953 genomes (★), compared to representatives of other basal fungi belonging to the
954 Mucoromycotina (⬣), Chytridiomycota (●), Blastocladiomycota (■), Entomophthoromycotina
955 (⬬), Mortierellomycotina (△), Glomeromycota (✚), Kickxellomycotina (▽), and

956 Zoopagomycotina (✖). CAZy families are shown as colored dots. The color code used was as
957 follows: green, CAZy families that are absent from AGF genomes; black, CAZy families present
958 in AGF genomes and with an entirely fungal origin; blue, CAZy families present in AGF
959 genomes and for which HGT contributed to < 50% of the transcripts in the examined
960 transcriptomes; red, CAZy families present in AGF genomes and for which HGT contributed to
961 > 50% of the transcripts in the examined transcriptomes. The majority of CAZyme families
962 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 (pfam seed)

Query of predicted peptides against Uniprot databases, identifying candidates with non-fungal bit score ≥ 100

HGT index h_U calculation

Identifying HGT candidates ($h_U \geq 30$)

Clustering (OrthoMCL)

Tree-building to determine the donor (ML, IQ-Tree PMSF Model, SH-aLRT test and UF bootstrap)

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)

≥6 genera

only 1 genus

Basal fungal orthologue absent

Basal fungal orthologue present

AGF-wide events

Genus-specific events

AGF-specific events

Fungal origin

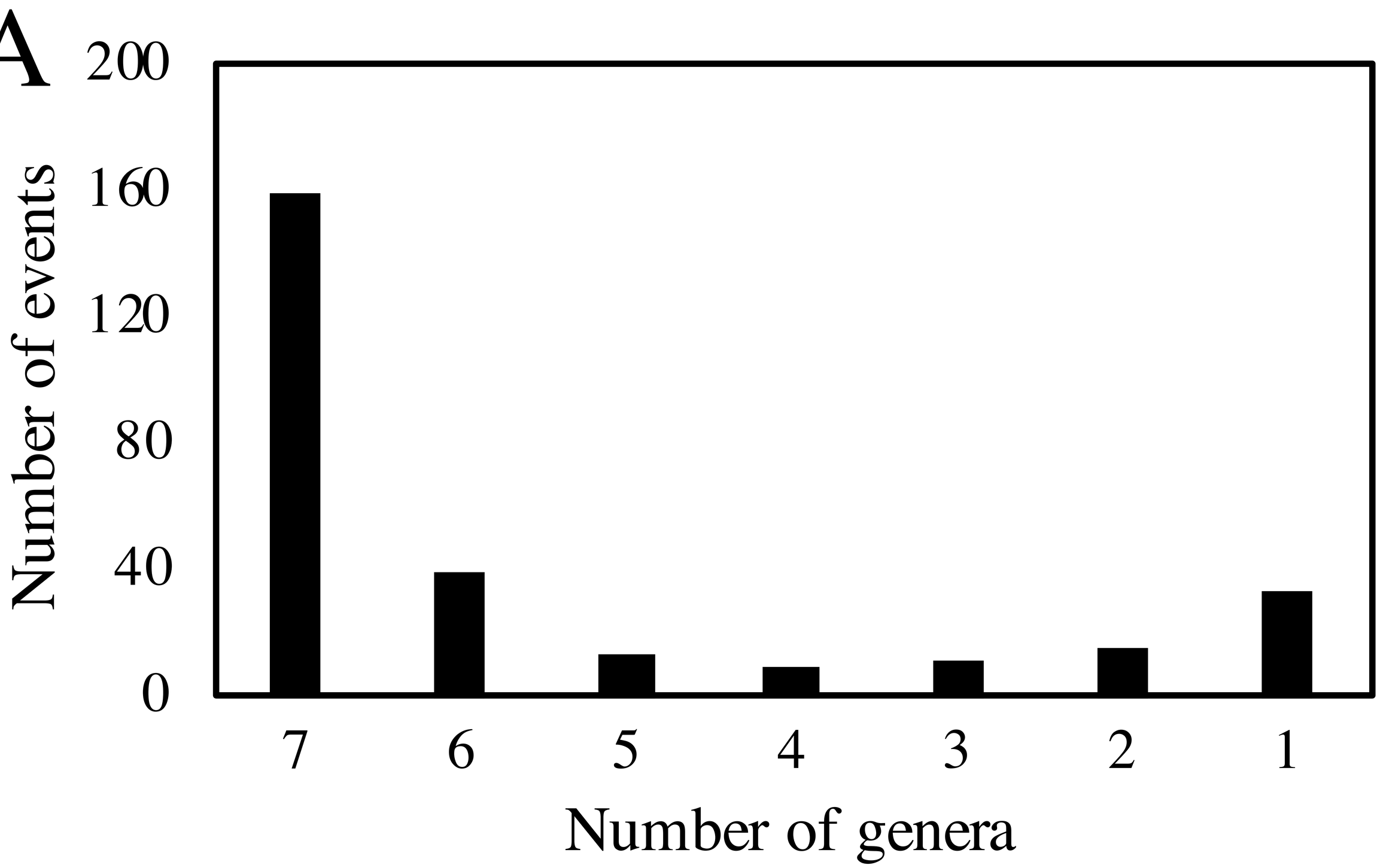
Non-fungal origin, but different donor

Non-fungal origin, conserved donor

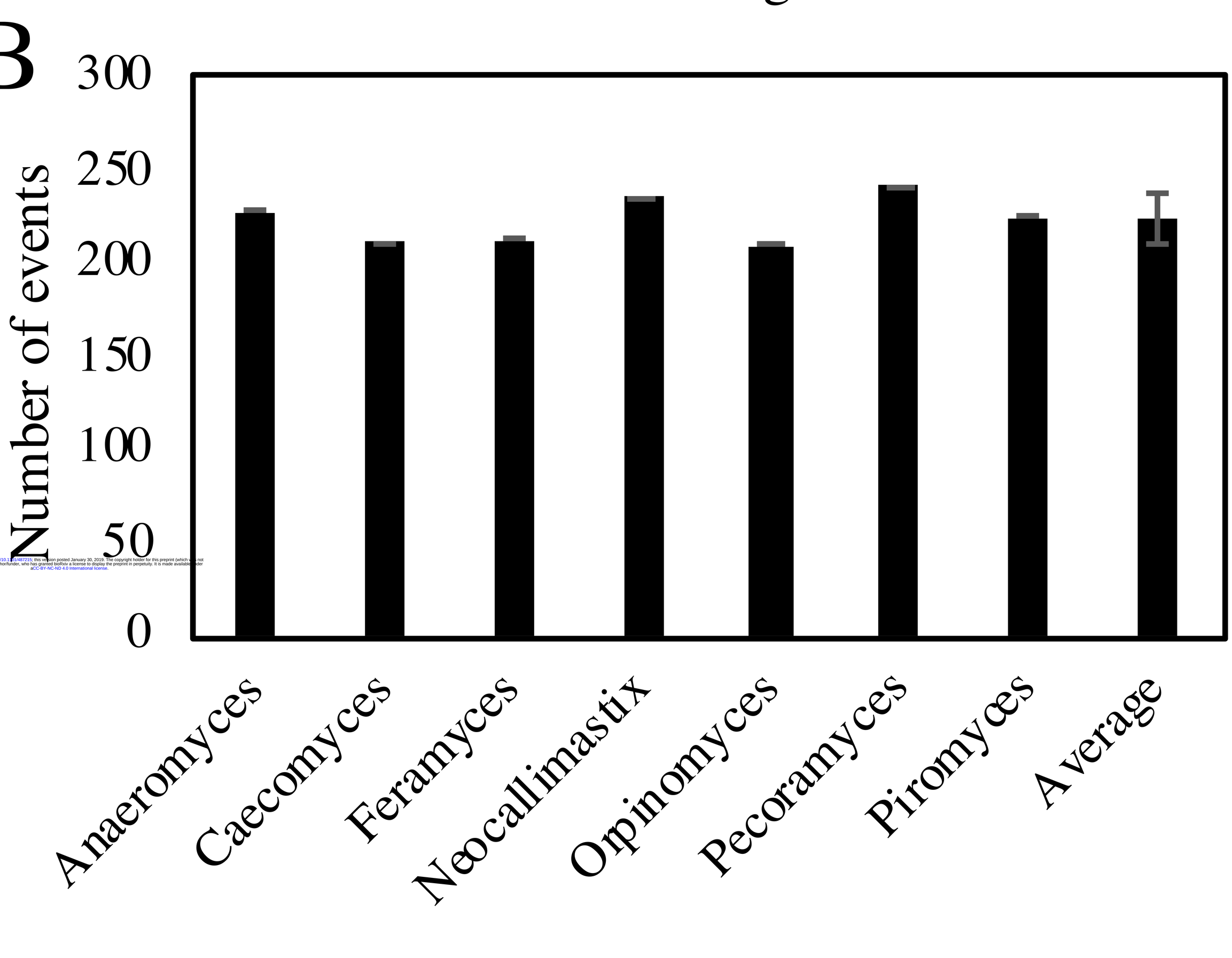
AGF-specific HGT event

AGF non-specific HGT events

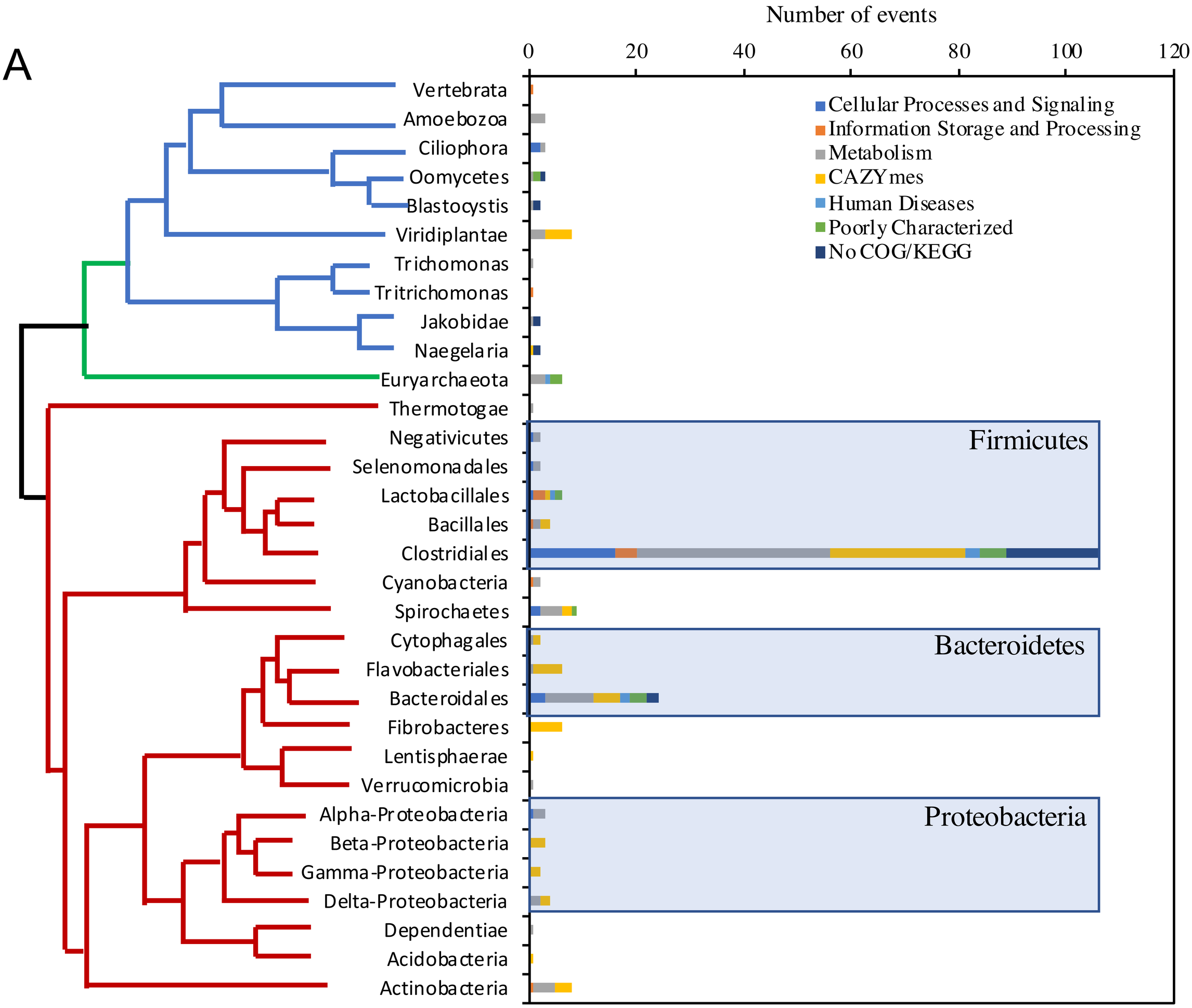
A



B

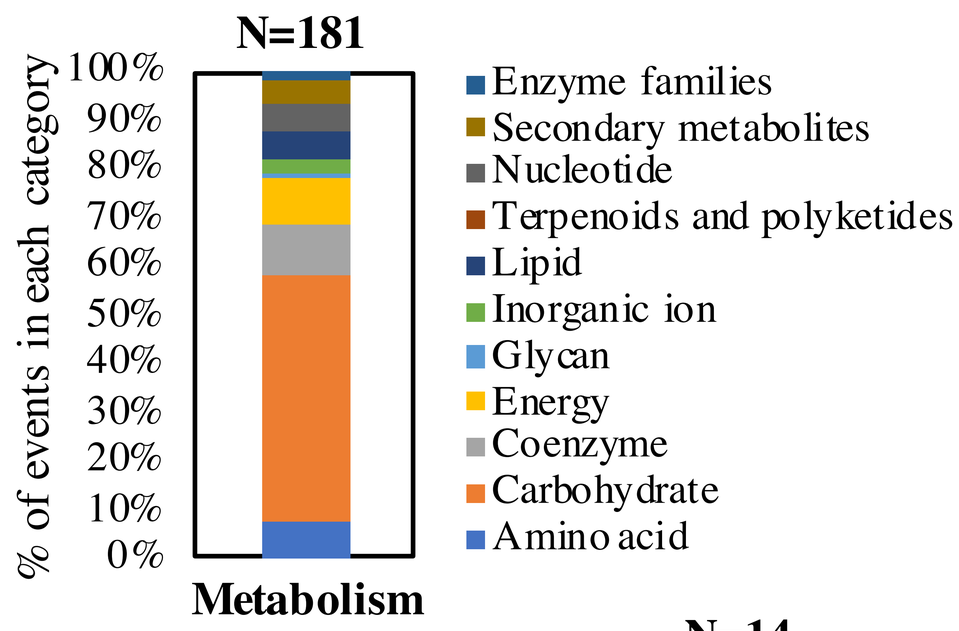


A

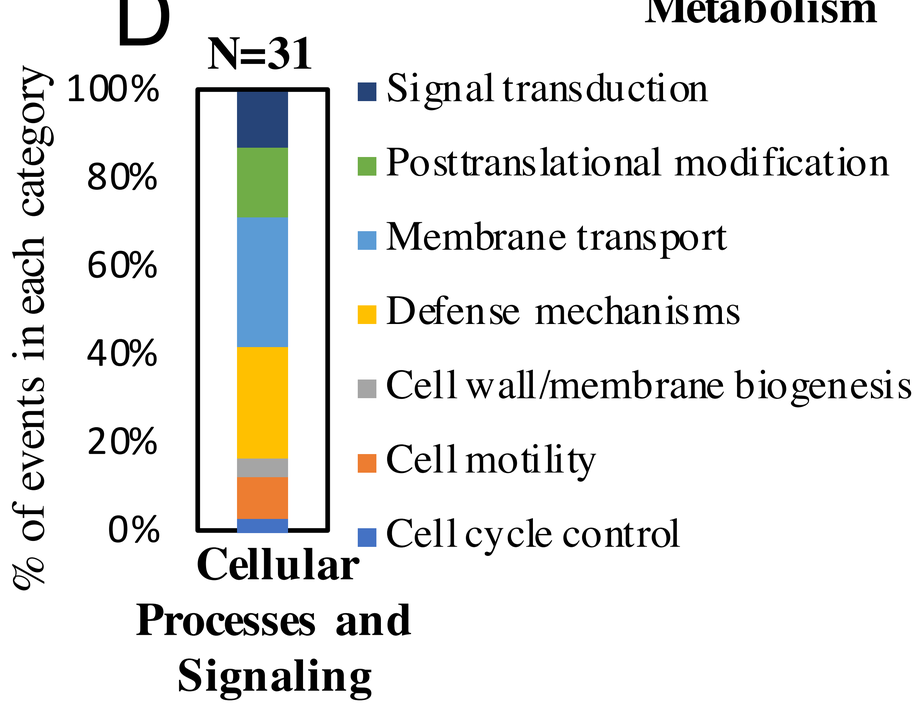


B

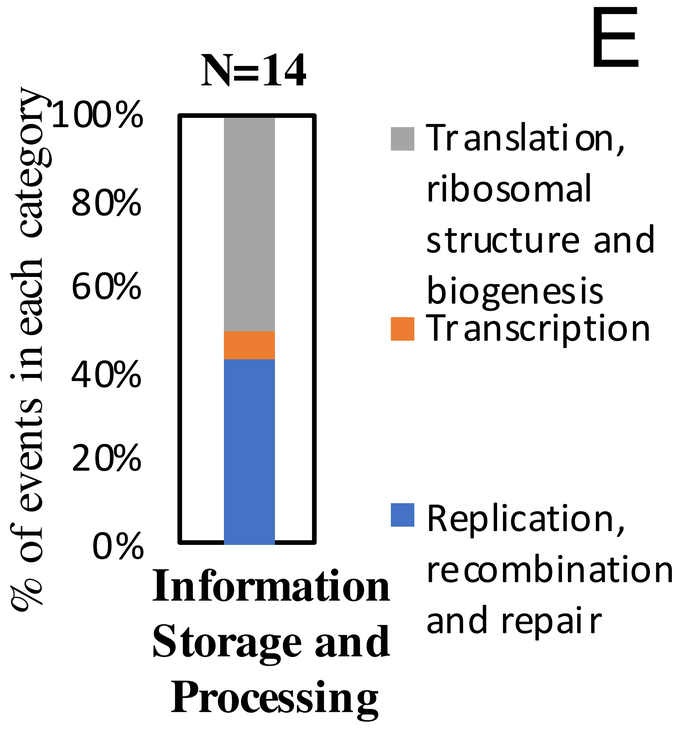
C

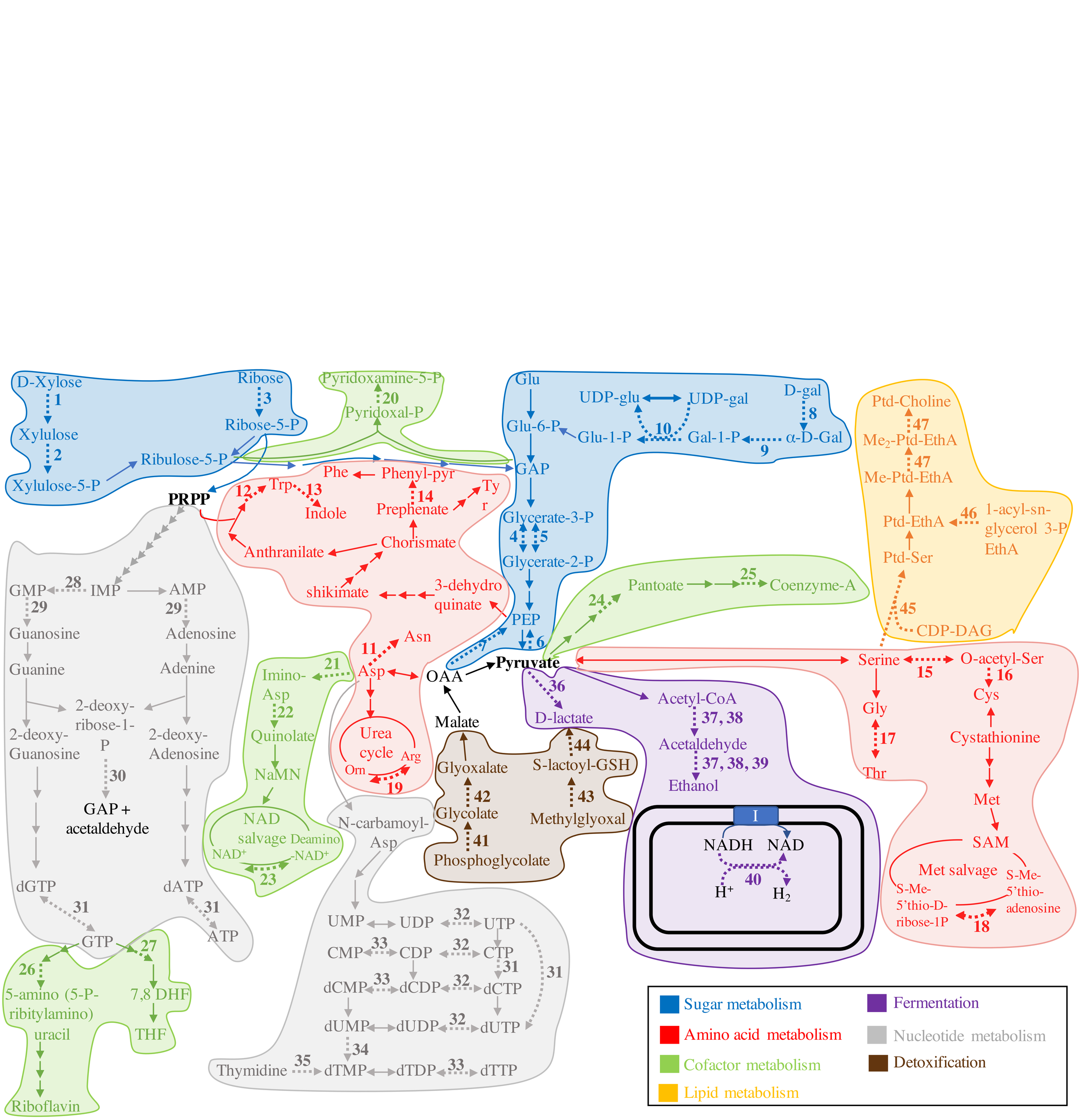


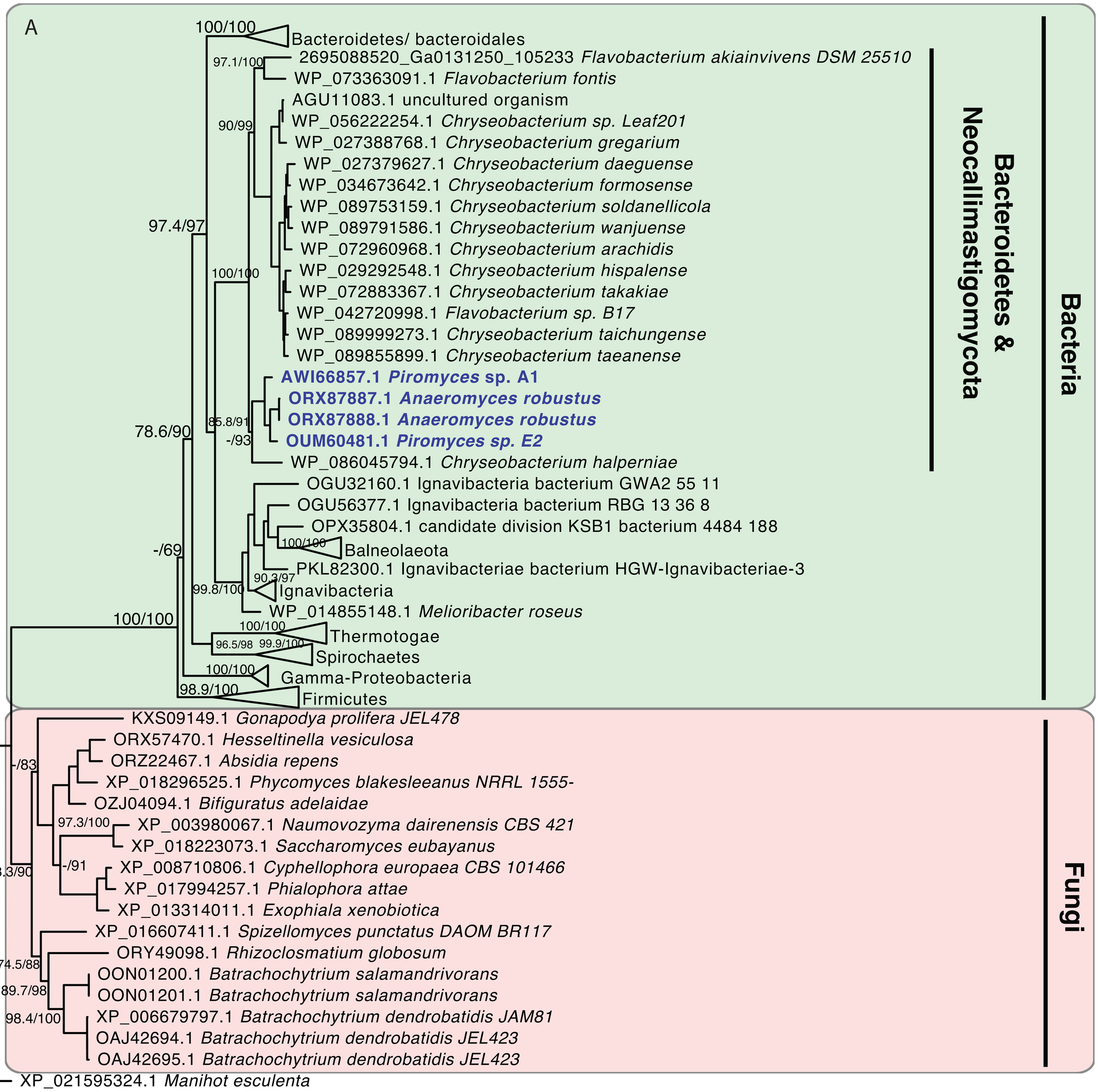
D

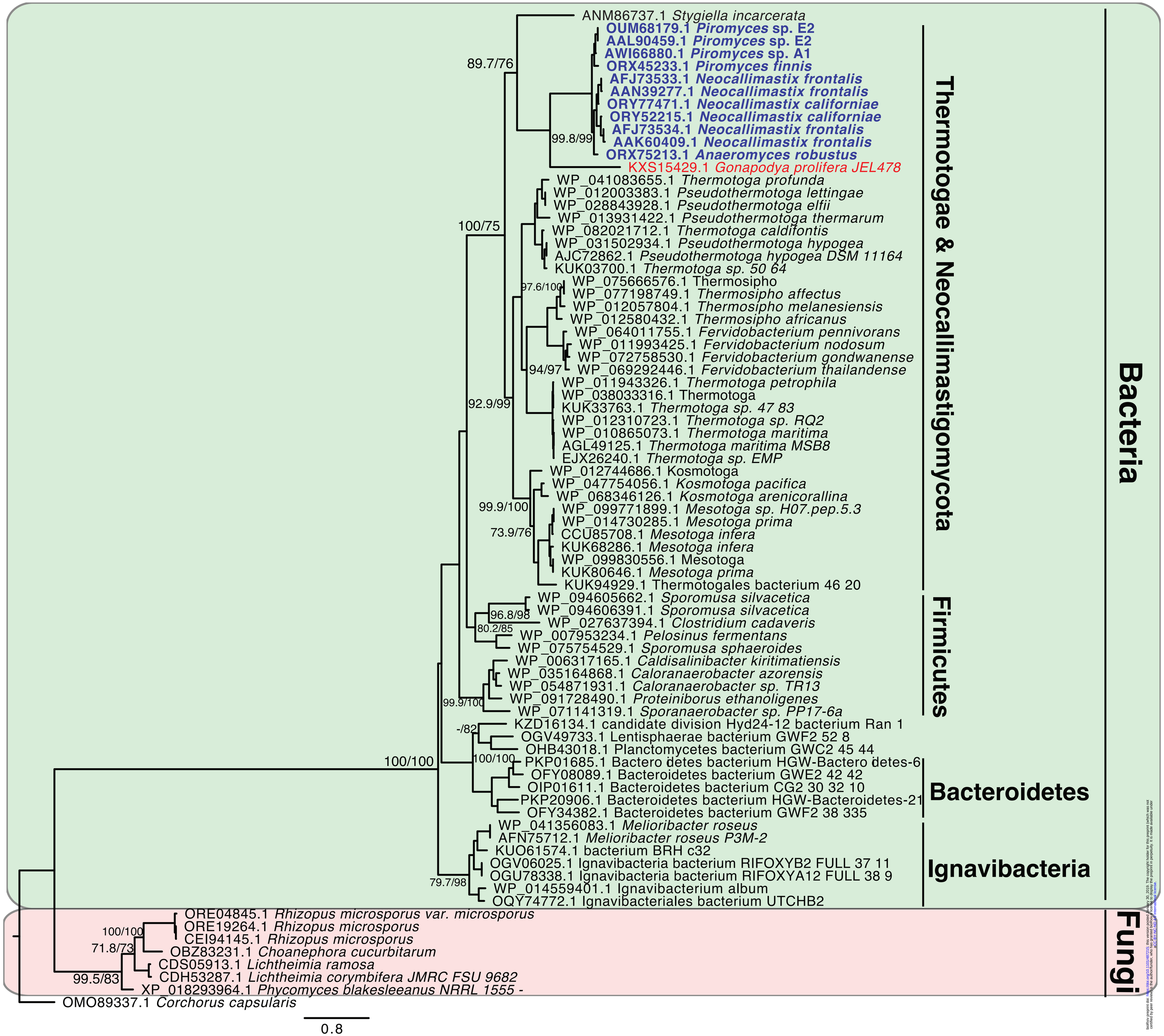


E

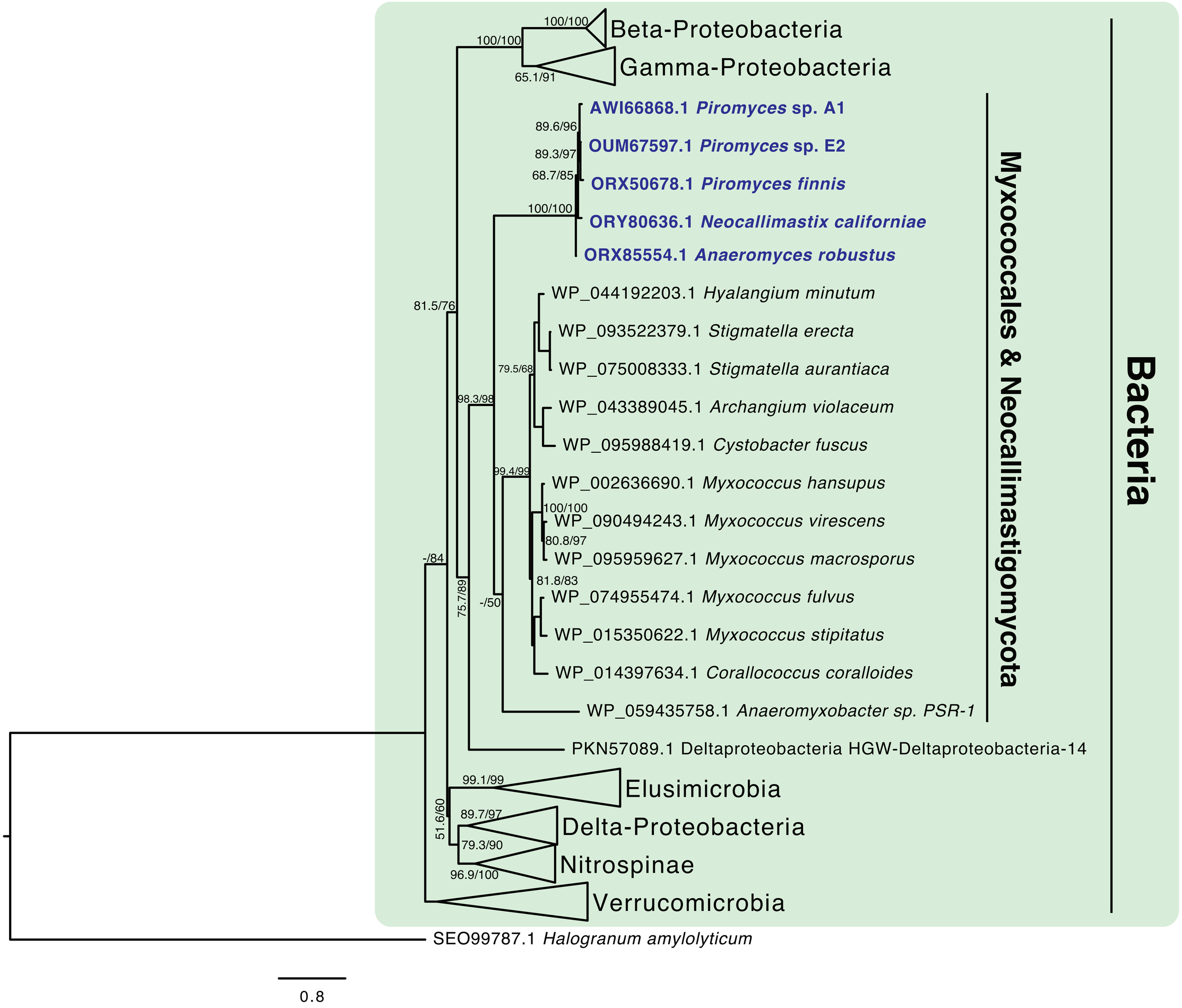




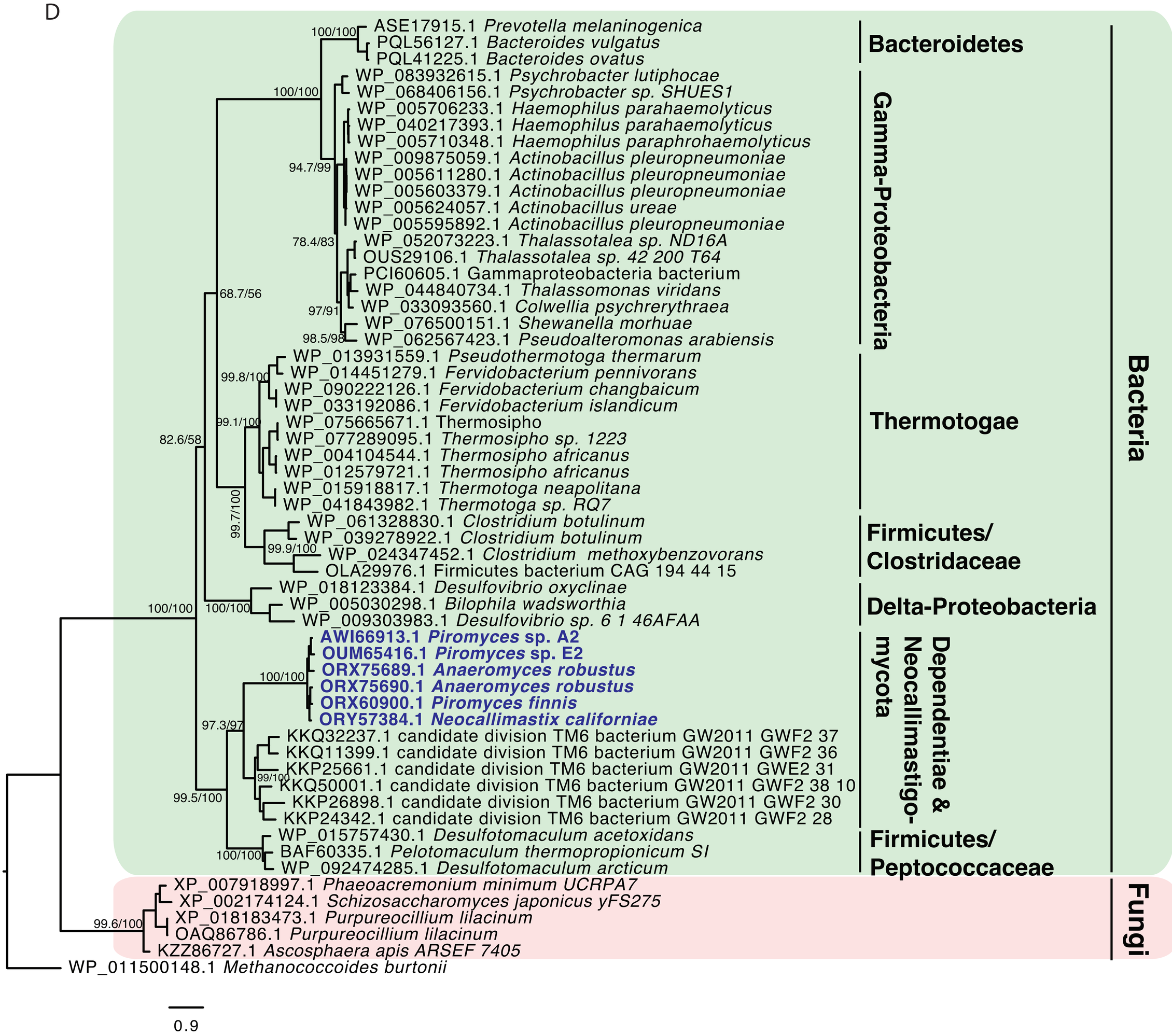



ANM86737.1 Stygiella incarcerata
OUM68179.1 Piromyces sp. E2
AAL90459.1 Piromyces sp. E2
AWI66880.1 Piromyces sp. A1
ORX45233.1 Piromyces finnis
AFJ73533.1 Neocallimastix frontalis
AAN39277.1 Neocallimastix frontalis
ORY77471.1 Neocallimastix californiae
ORY52215.1 Neocallimastix californiae
AFJ73534.1 Neocallimastix frontalis
AAK60409.1 Neocallimastix frontalis
ORX75213.1 Anaeromyces robustus
KXS15429.1 Gonapodya prolifera JEL478
89.7/76
99.8/99
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
100/75
WP_075666576.1 Thermosipho
97.6/100
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
94/97
WP_069292446.1 Fervidobacterium thailandense
WP_011943326.1 Thermotoga petrophila
WP_038033316.1 Thermotoga
KUK33763.1 Thermotoga sp. 47 83
92.9/99
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
99.9/100
WP_099771899.1 Mesotoga sp. H07.pep.5.3
WP_014730285.1 Mesotoga prima
73.9/76
CCU85708.1 Mesotoga infera
KUK68286.1 Mesotoga infera
WP_099830556.1 Mesotoga
KUK80646.1 Mesotoga prima
KUK94929.1 Thermotogales bacterium 46 20
Thermotogae & Neocallimastigomycota
WP_094605662.1 Sporomusa silvacetica
WP_094606391.1 Sporomusa silvacetica
96.8/98
WP_027637394.1 Clostridium cadaveris
80.2/85
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
99.9/100
WP_071141319.1 Sporanaerobacter sp. PP17-6a
Firmicutes
KZD16134.1 candidate division Hyd24-12 bacterium Ran 1
-/82
OGV49733.1 Lentisphaerae bacterium GWF2 52 8
OHB43018.1 Planctomycetes bacterium GWC2 45 44
100/100
100/100
PKP01685.1 Bactero detes bacterium HGW-Bactero detes-6
OFY08089.1 Bacteroidetes bacterium GWE2 42 42
OIP01611.1 Bacteroidetes bacterium CG2 30 32 10
PKP20906.1 Bacteroidetes bacterium HGW-Bacteroidetes-21
OFY34382.1 Bacteroidetes bacterium GWF2 38 335
Bacteroidetes
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
79.7/98
WP_014559401.1 Ignavibacterium album
OQY74772.1 Ignavibacteriales bacterium UTCHB2
Ignavibacteria
Bacteria
ORE04845.1 Rhizopus microsporus var. microsporus
ORE19264.1 Rhizopus microsporus
100/100
CEI94145.1 Rhizopus microsporus
71.8/73
OBZ83231.1 Choanephora cucurbitarum
CDS05913.1 Lichtheimia ramosa
CDH53287.1 Lichtheimia corymbifera JMRC FSU 9682
99.5/83
XP_018293964.1 Phycomyces blakesleeanus NRRL 1555 -
Fungi
OMO89337.1 Corchorus capsularis
B
0.8

C



D



Family	Genus						
	Anaeromyces	*Caecomyces*	*Pecoramyces*	*Piromyces*	*Neocallimastix*	*Feramyces*	*Orpinomyces*
GH1	11		6		11		7
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	7, 15	7, 15	7, 15	7, 15	7, 15	7, 15	7, 15
GH11	5, 7	5	5	5	5	5	5, 7
GH13	3, 7, 9	3, 7, 9	3, 7, 9	3, 9	3, 7, 9	3, 7, 9	3, 7, 9
GH16	3, 7		3	7	3, 7, 16	3, 7	3, 7
GH18		15	15	15	15	15	
GH20							
GH24	7, 12						12
GH25	7	7	7	7	7	7	
GH26	5	5	5	5	5		
GH28			3		3, 16		3
GH30	7				7		
GH31			2				
GH32	8		8		3, 16	8	8
GH36							
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	18		18	18		18	
GH48	15	15	15	15	15	15	15
GH53	7	7	7	7	7	7	7
GH57							
GH64	7	7	7		7	7	
GH67	3	3	3	3			
GH76						3	
GH78							7
GH88			15	15	15	15	15
GH95	3						3
GH97					3	3	
GH108				17			
GH114							
GH115	7	7	7	7	7	7	
CE1	5, 7	5, 7	5, 7	5, 7	5, 7	5, 7	5, 7
CE2	7, 15	7, 15	7, 15	7, 15	7, 15	3, 7	7
CE3	7	7	7	7	7	7	7
CE4	5	5	5	5	5	5	5
CE6	7	7	7	7	7	7	7
CE7	7						
CE8	16	16	16	16	16	16	16
CE12	3, 13		7	7, 13	7	3, 7	7
CE15	15	5, 15	15	15	5, 15	15	15
CE16					19		
PL1	2	2	2	2	2	2	2
PL3							
PL4	15	15	15	15	15	15	15
PL9	7, 14		7, 14	10, 14	7		14
PL11	1				7		

Donor Key			
1	Acidobacteria		Bacteria
2	Actinobacteria		
3	Bacteroidetes		
4	Deinococcus		
5	Fibrobacter		
6	Bacillales	Firmicutes	
7	Clostridiales		
8	Lactobacillales		
9	Unclassified Firmicutes		
10	Lentisphaerae		
11	Beta-Proteobacteria	Proteobacteria	
12	Gamma-Proteobacteria		
13	Delta-Proteobacteria		
14	Spirochaetes		
15	Bacteria (unnested)		
16	Viridplantae		Eukaryota
17	Neagelaria		
18	Unclassified Eukaryotes		

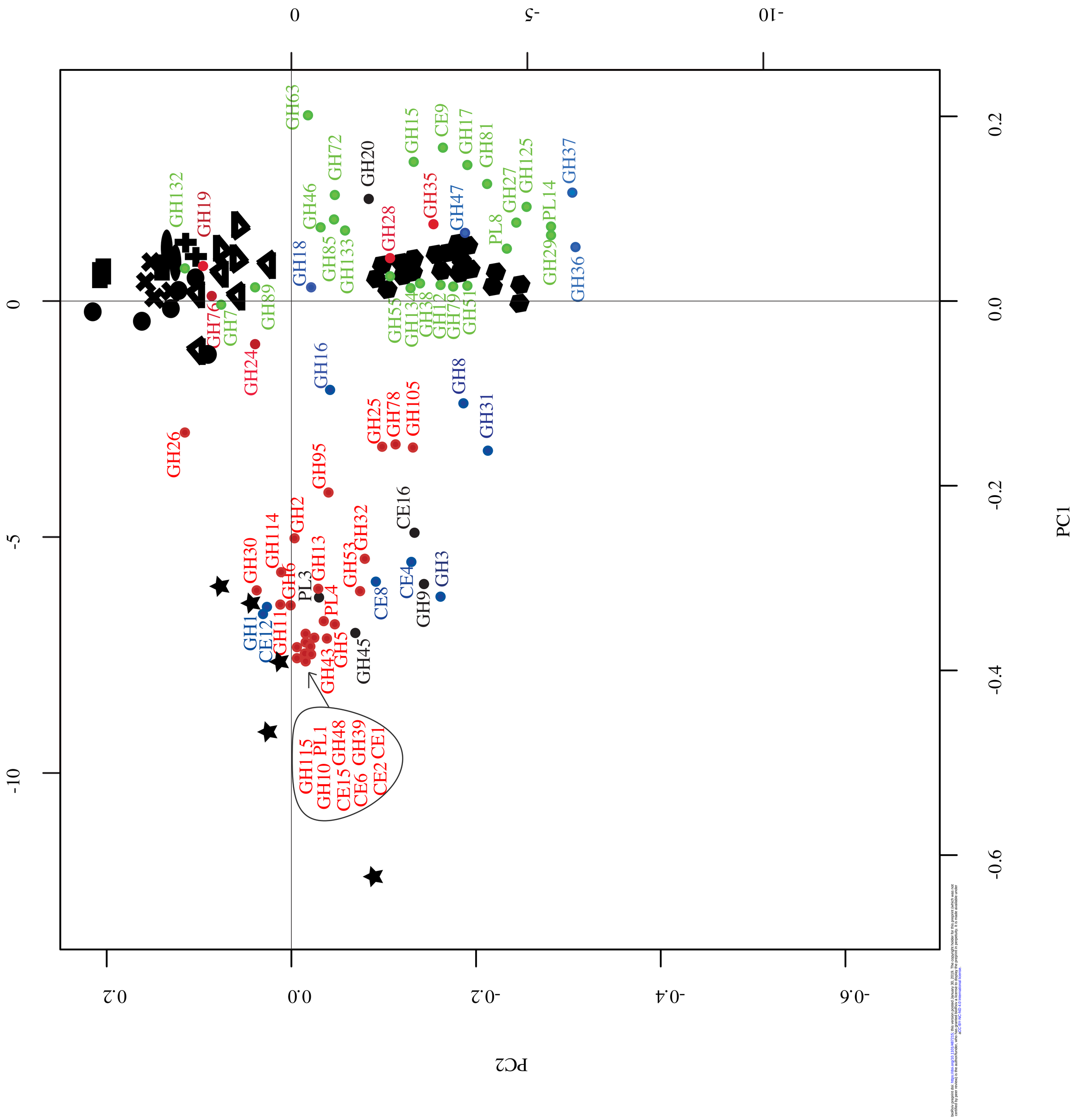


Table 1: Neocallimastigomycota strains analyzed in this study.

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, CA	NA*	(44)
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 lervia*)	Feces	Stillwater, OK	MG605675	This study
Feramyces	*austinii*	F3a	Aoudad sheep (*Ammotragus lervia*)	Feces	Stillwater, OK	MG584226	This study
Neocallimastix	*californiae*	G1	Goat (*Capra aegagrus hircus*)	Feces	Santa Barbara, CA	Genomic sequence**	(44)
Neocallimastix	cf. *cameroonii*	G3	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992493	This study
Neocallimastix	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	(103, 104)
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, CA	Genomic sequence**	(44)
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 maximus*)	Feces	London, UK	NA	(44, 105)

*NA: Not available

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