Title

Phylogenomic analysis of the Neocallimastigomycota: proposal of Caecomycetaceae fam. nov., Piromycetaceae fam. nov., and emended description of the families Neocallimastigaceae and Anaeromycetaceae.

Journal

International Journal of Systematic and Evolutionary Microbiology, 73(2)

ISSN

1466-5026

Authors

Hanafy, Radwa A

Wang, Yan

Stajich, Jason E

et al.

Publication Date

2023-02-24

DOI

10.1099/ijsem.0.005735

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE

Hanafy *et al.*, *Int. J. Syst. Evol. Microbiol.* 2023;73:005735

DOI 10.1099/ijsem.0.005735

MICROBIOLOGY SOCIETY

OPEN ACCESS

Phylogenomic analysis of the *Neocallimastigomycota*: proposal of *Caecomycetaceae* fam. nov., *Piromycetaceae* fam. nov., and emended description of the families *Neocallimastigaceae* and *Anaeromycetaceae*

Radwa A. Hanafy[1,2]†, Yan Wang[3,4]†, Jason E. Stajich[5], Carrie J. Pratt[1], Noha H. Youssef[1] and Mostafa S. Elshahed[1,*]

Abstract

The anaerobic gut fungi (AGF) represent a coherent phylogenetic clade within the Mycota. Twenty genera have been described so far. Currently, the phylogenetic and evolutionary relationships between AGF genera remain poorly understood. Here, we utilized 52 transcriptomic datasets from 14 genera to resolve AGF inter-genus relationships using phylogenomics, and to provide a quantitative estimate (amino acid identity, AAI) for intermediate rank assignments. We identify four distinct supra-genus clades, encompassing all genera producing polyflagellated zoospores, bulbous rhizoids, the broadly circumscribed genus *Piromyces*, and the *Anaeromyces* and affiliated genera. We also identify the genus *Khoyollomyces* as the earliest evolving AGF genus. Concordance between phylogenomic outputs and RPB1 and D1/D2 LSU, but not RPB2, MCM7, EF1α or ITS1, phylogenies was observed. We combine phylogenomic analysis and AAI outputs with informative phenotypic traits to propose accommodating 14/20 AGF genera into four families: *Caecomycetaceae* fam. nov. (encompassing the genera *Caecomyces* and *Cyllamyces*), *Piromycetaceae* fam. nov. (encompassing the genus *Piromyces*), emend the description of the family *Neocallimastigaceae* to encompass the genera *Neocallimastix*, *Orpinomyces*, *Pecoramyces*, *Feramyces*, *Ghazallomyces*, *Aestipascuomyces* and *Paucimyces*, as well as the family *Anaeromycetaceae* to include the genera *Oontomyces*, *Liebetanzomyces* and *Capellomyces* in addition to *Anaeromyces*. We refrain from proposing families for the deeply branching genus *Khoyollomyces* and for genera with uncertain position (*Buwchfawromyces*, *Joblinomyces*, *Tahromyces*, *Agriosomyces* and *Aklioshbomyces*) pending availability of additional isolates and sequence data; and these genera are designated as 'genera incertae sedis' in the order *Neocallimastigales*. Our results establish an evolutionary-grounded Linnaean taxonomic framework for the AGF, provide quantitative estimates for rank assignments, and demonstrate the utility of RPB1 as an additional informative marker in *Neocallimastigomycota* taxonomy.

INTRODUCTION

Members of the anaerobic gut fungi (AGF) represent a phylogenetically, metabolically and ecologically coherent clade in the kingdom Mycota [1]. Twenty genera and 36 different species have been described so far [2]. A recent review provided detailed descriptions of current genera and resolved historical inaccuracies and synonymies within the *Neocallimastigomycota* [2]. Further, criteria for the identification and characterization, as well as guidelines for genus- and species-level rank assignment for novel AGF isolates have recently been formulated [3]. Despite such progress, the phylogenetic and evolutionary relationships between various genera within the *Neocallimastigomycota* are currently unclear. Similarities in specific microscopic traits (zoospore flagellation, thallus development and rhizoidal growth patterns) across genera have been identified and the significance of using such traits for proposing higher order relationship has been debated [4–6]. As well, phylogenetic analysis using two ribosomal loci: the internal transcribed spacer region 1 (ITS1) and the D1/D2 region of the large

Author affiliations: [1]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA; [2]Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, DE, USA; [3]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada; [4]Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON M1C 1A4, Canada; [5]Department of Microbiology and Plant Pathology, University of California, Riverside, CA, USA.

***Correspondence:** Mostafa S. Elshahed, mostafa@okstate.edu

Keywords: anaerobic gut fungi; Neocallimastigomycota; phylogenomics; rumen.

Abbreviations: AAU, average amino acid identity; AGF, anaerobic gut fungi; ANI, average nucleotide identity; D1/D2 LSU, D1/D2 region of the large ribosomal subunit; EF1α, elongation factor 1-alpha; ITS1, the internal transcribed spacer region 1; MCM7, minichromosome maintenance complex component 7; RPB1, RNA polymerase II large subunit; RPB2, RNA polymerase II second largest subunit.

†These authors contributed equally to this work

Two supplementary figures and two supplementary tables are available with the online version of this article.

Table 1. List of strains used in this study

Genus	species	Strain	Genome BioProject accession no.	Transcriptome BioProject accession no.	SRA accession no.	Assembled transcriptome TSA accession no.	Reference
Aestapascuomyces	*dupliciliberans*	R1		PRJNA847081	SRR19612713		This study
Aklioshbomyces	*papillarum*	WTS1		PRJNA847081	SRR19612712		This study
Anaeromyces	sp.	ABS23		PRJNA847081	SRR19612701		This study
Anaeromyces	*contortus*	C3G		PRJNA489922		GGWR00000000	[67, 68]
Anaeromyces	*contortus*	C3J		PRJNA489922		GGWO00000000	[67, 68]
Anaeromyces	*contortus*	G3G		PRJNA489922		GGWP00000000	[67, 68]
Anaeromyces	*contortus*	NA		PRJNA489922		GGWN00000000	[67, 68]
Anaeromyces	*contortus*	O2		PRJNA489922		GGWQ00000000	[67, 68]
Anaeromyces	*mucronatus*	YE505		PRJNA437872			[38]
Anaeromyces	*robustus*	S4	PRJNA330692	PRJNA250973			[40]
Caecomyces	*communis* var. *churrovis*	A	PRJNA347164	PRJNA393353			[39, 41]
Caecomyces	cf. *communis*	FD27		PRJNA847081	SRR19612700		This study
Caecomyces	sp.	TB33		PRJNA847081	SRR19612699		This study
Caecomyces	cf. *communis*	Iso3		PRJNA489922		GGXE00000000	[67, 68]
Caecomyces	cf. *communis*	Brit4		PRJNA489922		GGWS00000000	[67, 68]
Capellomyces	*foraminis*	Cap2a		PRJNA847081	SRR19612698		This study
Cyllamyces	cf. *aberensis*	TSB2		PRJNA847081	SRR19612697		This study
Feramyces	*austinii*	WSF2		PRJNA489922		GGWT00000000	[67, 68]
Feramyces	*austinii*	WSF3		PRJNA489922		GGWU00000000	[67, 68]
Khoyollomyces	*ramosus*	ZO44		PRJNA847081	SRR19612696		This study
Liebetanzomyces	cf. *polymoprphus*	Orc37		PRJNA847081	SRR19612695		This study
Neocallimastix	cf. *frontalis*	EC30		PRJNA847081	SRR19612694		This study
Neocallimastix	cf. *frontalis*	Hef5		PRJNA489922		GGXJ00000000	[67, 68]
Neocallimastix	*frontalis*	27		PRJNA437872			[38]
Neocallimastix	*cameroonii* var. *californiae*	G1	PRJNA262392	PRJNA251043			[40]
Neocallimastix	*cameroonii* var. *lanati*	sp3	PRJNA658393	PRJNA677809			[43]
Neocallimastix	cf. *cameroonii*	G3		PRJNA489922		GGXC00000000	[67, 68]
Orpinomyces	cf. *joyonii*	AB6		PRJNA847081	SRR19612711		This study
Orpinomyces	cf. *joyonii*	AB3		PRJNA847081	SRR19612710		This study
Orpinomyces	cf. *joyonii*	ABC24		PRJNA847081	SRR19612709		This study
Orpinomyces	cf. *joyonii*	D3A		PRJNA489922		GGWV00000000	[67, 68]
Orpinomyces	cf. *joyonii*	D3B		PRJNA489922		GGWW00000000	[67, 68]
Orpinomyces	cf. *joyonii*	D4C		PRJNA489922		GGWX00000000	[67, 68]
Orpinomyces	*joyonii*	SG4		PRJNA437872			[38]
Paucimyces	*polynucleatus*	BB3		PRJNA847081	SRR19612708		This study
Pecoramyces	*ruminantium*	C1A	PRJNA200719	PRJNA284193			[67, 69]
Pecoramyces	*ruminantium*	S4B		PRJNA489922		GGWY00000000	[67, 68]

Continued

Table 1. Continued

Genus	species	Strain	Genome BioProject accession no.	Transcriptome BioProject accession no.	SRA accession no.	Assembled transcriptome TSA accession no.	Reference
Pecoramyces	cf. *ruminantium*	FS3C		PRJNA489922		GGXF00000000	[67, 68]
Pecoramyces	cf. *ruminantium*	FX4B		PRJNA489922		GGWZ00000000	[67, 68]
Pecoramyces	cf. *ruminantium*	YC3		PRJNA489922		GGXA00000000	[67, 68]
Pecoramyces	cf. *ruminantium*	Orc32		PRJNA847081	SRR19612707		This study
Pecoramyces	cf. *ruminantium*	AS31		PRJNA847081	SRR19612706		This study
Pecoramyces	cf. *ruminantium*	AS32		PRJNA847081	SRR19612705		This study
Pecoramyces	*ruminantium*	F1	PRJNA517297	PRJNA517315			[42]
Piromyces	*finnis*	finn	PRJNA330696	PRJNA268530			[40]
Piromyces	cf. *finnis*	DonB11		PRJNA847081	SRR19612704		This study
Piromyces	sp.	Axs23		PRJNA847081	SRR19612703		This study
Piromyces	sp.	A1		PRJNA489922		GGXB00000000	[67, 68]
Piromyces	sp.	B4		PRJNA489922		GGXH00000000	[67, 68]
Piromyces	sp.	Ors32		PRJNA847081	SRR19612702		This study
Piromyces	sp.	E2	PRJNA82799				[40]
Piromyces	*rhizinflatus*	YM600		PRJNA437872			[38]

ribosomal subunit (D1/D2 LSU) has yielded multiple statistically supported supra-genus groupings, although such topologies were often dependent on the locus examined, region amplified, taxa included in the analysis, and tree-building algorithm employed [5, 7, 8].

Therefore, while phenotypic and phylogenetic analyses suggest the existence of supra-genus relationships within the *Neocallimastigomycota*, the exact nature of such groupings is yet unclear. Approaches that utilize whole genomic and/or transcriptomic (henceforth referred to as –omics) datasets represent a promising tool towards resolving such relationships [9–13]. Comparative genomics approaches (e.g. calculation of shared Kmer (Kmer overlap) [14, 15], average nucleotide identity (ANI) [16], identification of genomic syntenic blocks [17]) have been increasingly utilized in taxonomic studies, aided by the development of lower cost high throughput sequencing technologies and the wider availability of bioinformatic analysis tools. More importantly, the development and implementation of phylogenomic approaches have been crucial in resolving high-rank [12] and intra-clade (e.g. [18]) phylogenies within fungi. Phylogenomic analysis involves the identification of groups of single-copy orthologous genes in the group of interest followed by multiple alignments of each orthologous gene. Analysis to determine a species tree can then be performed on either the concatenated alignment of all genes to obtain a single phylogeny of the group in question, or on the individual alignments via coalescence of individual gene trees. In addition, the inferred gene trees output from such approaches could also be compared to single gene phylogenies to assess their value and potential utility for taxonomic assessment and ecological surveys.

Within a Linnaean taxonomic framework, taxonomic associations between genera are accommodated in the intermediate ranks of families, orders and classes. Currently, AGF genera are recognized in a single family (*Neocallimastigaceae*), order (*Neocallimastigales*) and class (*Neocallimastigomycetes*) in the phylum *Neocallimastigomycota* [19, 20]. It is interesting to note that a nomenclature novelty entry in the Index Fungorum database (IF550425) proposes an additional family (*Anaeromycetaceae*) with the genus *Anaeromyces* as its sole member, although no detailed justification for such a proposal was provided. Indeed, all current genera in the AGF, including *Anaeromyces*, are assigned to the family *Neocallimastigaceae* in recent publications [2, 3] and databases (Mycobank and Index Fungorum). Regardless, it is clear that the current intermediate rank taxonomic outline of AGF genera has not been proposed based on a detailed comparative phenotypic and phylogenetic analysis of relationships between genera. Rather, it reflects the cumulative and progressive recognition of the phylogenetic and phenotypic distinction of the *Neocallimastigomycota* when compared to all other fungal clades. The earliest studies on AGF taxonomy [21] proposed accommodating them into a family (*Neocallimastigaceae*) within the chytrid order *Spizellomycetales*, a reflection of zoospore ultrastructure similarity, and emended the description of *Spizellomycetales* order to include zoopsores with multiple flagella. Ten years later, Li *et al.* [22] used cladistic analysis of 42 morphological and ultrastructural characters to demonstrate the distinction of the AGF when compared to members of the *Chytridiomycetes*, hence elevating the anaerobic gut fungi from a family to an order (*Neocallimastigales*). Molecular analysis using concatenated protein-coding genes as well as rRNA genes [20, 23, 24], and several morphological and ultrastructural differences from other *Chytridiomycetes* [25] necessitated their recognition as a phylum (*Neocallimastigomycota*) with one class (*Neocallimastigomycetes*), a view that has recently

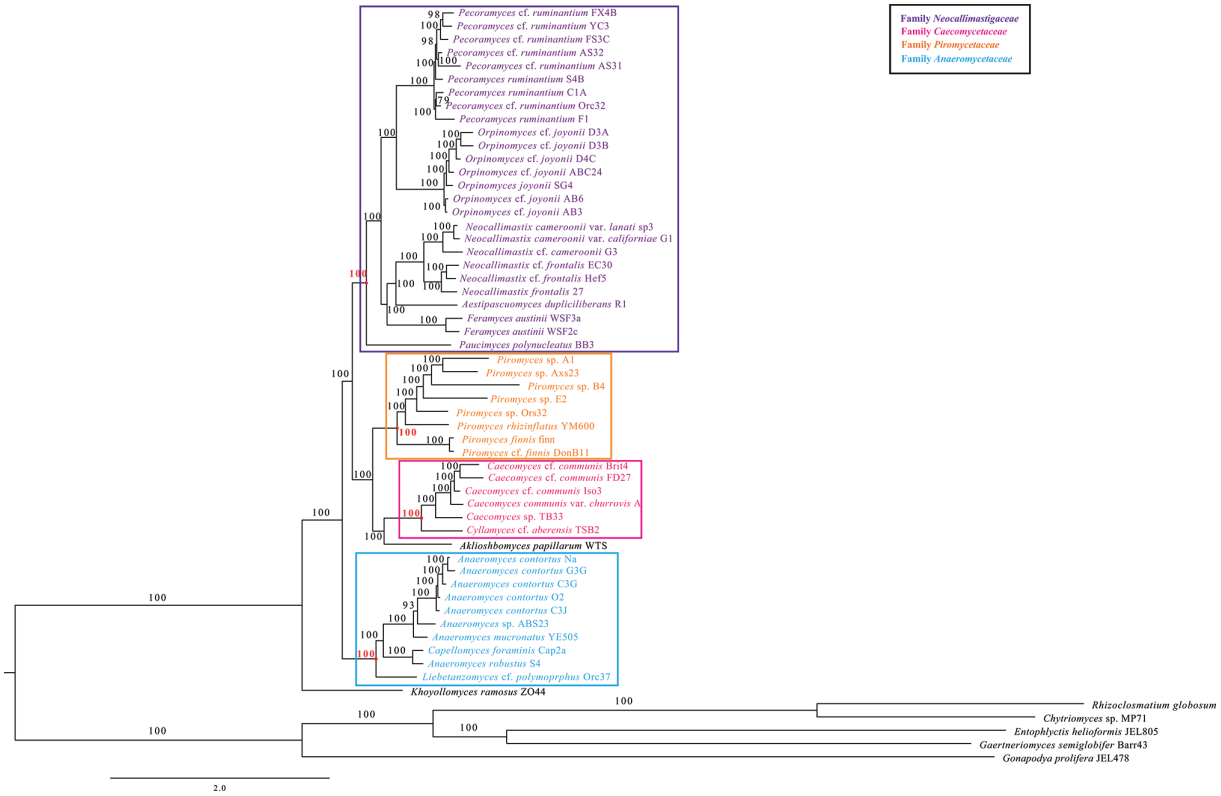

Fig. 1. Phylogenomic tree of *Neocallimastigomycota* based on 670 genome-wide markers highlighting the family-level relationships within the phylum. The tree was reconstructed using the maximum-likelihood approach implemented in the IQ-TREE package. The numbers at the nodes represent the ultrafast bootstrap values suggesting robustness of the taxa joining. The scale bar at the bottom indicates the number of substitutions per site in the analysis. Isolate names at tree tips are colour coded by family as shown in the key (family *Neocallimastigaceae*, clade 1, purple; family *Caecomycetaceae*, clade 2, lavender; family *Piromycetaceae*, clade 3, orange; and family *Anaeromycetaceae*, clade 4, light blue), and boxes with the same colours are used to delimit each family. The support values at the nodes corresponding to each family are shown in bold red text, and the node itself is shown as a red dot.

been corroborated via phylogenomic analysis [12]. Indeed, currently published taxonomic outlines, e.g [19], and databases (e.g. GenBank [26], and Mycocosm [27]) recognize the AGF at the rank of phylum within the Mycota.

The last decade has witnessed a rapid expansion in the number of described genera within the *Neocallimastigomycota* [4, 28–34]. Due to such expansion, as well as the continuous recognition of the value of genome-based taxonomy in resolving relationships and circumscribing ranks in fungal taxonomy [9, 12, 13]; we posit that a lineage-wide phylogenomic assessment is warranted to resolve inter-genus relationships and explore the need for intermediate ranks to establish a proper Linnaean-based outline for the phylum. Here, we conducted transcriptomic sequencing on multiple additional AGF genera isolated and characterized in our laboratory, and combined these datasets with previously available AGF transcriptomes and genomes to resolve the inter-genus relationships within the *Neocallimastigomycota*. Based on our results, we propose accommodating described AGF genera into four distinct families to reflect the observed inter-genus relationships. In addition, we provide quantitative amino acid identity (AAI) values for circumscribing such families, and test the utility of multiple single genes/loci as additional markers for resolving AGF phylogenetic relationships.

METHODS

Cultures

Transcriptomes and genomes from 52 strains representing 14 AGF genera were analysed (Table 1). Of these, transcriptomes of twenty strains, representing six genera for which no prior sequence data were available were sequenced as part of this study. Many of the analysed strains have previously been described as novel genera or species by the authors [29–31, 33, 34] (Table 1). Others possessed identical features to previously described type strains and were designated as conferre (cf.) strains (Table 1). Few were identified to the genus level and given an alphanumeric strain name designation (Table 1).

Table 2. Clades circumscribed in this study using phylogenomics and AAI values

Clades	Genera	AAI			Phenotype*
		Average intra-genus (range)	Average inter-genus intra-clade (range)	Average inter-clade (range)	
Clade 1	*Pecoramyces*, *Orpinomyces*, *Neocallimastix*, *Paucimyces*, *Aestipascuomyces*, *Feramyces*	96.89 (87.78–99.49)	79.22 (74.61–82.95)	73.21 (65.27–76.64)	Filamentous rhizoidal growth pattern, polyflagellated zoospores (except for *Pecoramyces* and *Paucimyces*), monocentric thalli (except for *Orpinomyces* and *Paucimyces*)
Clade 2	*Cyllamyces*, *Caecomyces*	94.01 (88.02–98.37)	83.67 (83.08–84.05)	72.8 (67.39–74.91)	Bulbous rhizoidal growth pattern
Clade 3	*Piromyces*	79.35 (72.58–99.06)	79.35 (72.58–99.06)	72.61 (65.27–75.61)	Filamentous rhizoidal growth pattern, monoflagellated zoospores, monocentric thalli
Clade 4	*Anaeromyces*, *Liebetanzomyces*, *Capellomyces*	96.55 (93.07–99.6)	84.41 (83.58–85.48)	73.75 (67.25–76.64)	Filamentous rhizoidal growth pattern, monoflagellated zoospore, monocentric thalli (except *Anaeromyces*)
	Aklioshbomyces	NA†	NA†	73.54 (69.1–75.42)	
	Khoyollomyces	NA†	NA†	71.88 (66.47–73.41)	

*Detailed features for every AGF genus have been described in detail in a recent review [2].
†NA, not applicable.

RNA extraction, sequencing, quality control and transcript assembly

Isolates were grown in rumen fluid medium with cellobiose as the sole carbon source [35] to the late log/early stationary phase (approximately 48–60 h post inoculation). Cultures were vacuum filtered to obtain fungal biomass then ground with a pestle in a mortar under liquid nitrogen. Total RNA was extracted using an Epicentre MasterPure yeast RNA purification kit according to manufacturer's instructions and stored in RNase-free TE buffer (10 mM Tris pH 8, 1 mM EDTA). Transcriptomic sequencing using an Illumina HiSeq2500 platform and 2×150 bp paired-end library was conducted using the services of a commercial provider (Novogene Corporation, Beijing, PR China) or at the Oklahoma State University Genomics and Proteomics centre. The RNA-seq data were quality trimmed and *de novo* assembled with Trinity (version 2.6.6) using default parameters. For each data set, redundant transcripts were clustered using CD-HIT [36] with identity parameter of 95% (–c 0.95). The obtained nonredundant transcripts were subsequently used for peptide and coding sequence prediction using the TransDecoder (version 5.0.2) (https://github.com/TransDecoder/TransDecoder) with a minimum peptide length of 100 amino acids. Assessment of transcriptome completeness per strain was conducted using BUSCO [37] with the fungi_odb10 dataset modified to remove 155 mitochondrial protein families as previously suggested [38].

Phylogenomic analysis

The phylogenomic analysis includes 20 newly sequenced and 32 existing AGF genomic and transcriptomic sequences (Table 1) [38–43]. Five *Chytridiomycota* genomes were also included as the outgroup (*Chytriomyces* sp. strain MP 71, *Entophlyctis helioformis* JEL805, *Gaertneriomyces semiglobifer* Barr 43, *Gonapodya prolifera* JEL478 and *Rhizoclosmatium globosum* JEL800 [44, 45]). The 'fungi_odb10' dataset including 758 phylogenomic markers for kingdom Fungi was retrieved from the BUSCO version 4.0 package [37], and used in our analysis. Profile hidden-Markov-models of these markers were created and used to identify homologues in all included 58 fungal proteomes using hmmer3 (version 3.1b2) employed in the PHYling pipeline (https://doi.org/10.5281/zenodo.1257002). A total of 670 out of the 758 'fungi_odb10' markers were identified with conserved homologs in the 57 AGF and chytrids genomes, which were then aligned and concatenated for subsequent phylogenomic analyses. The final input data included 491 301 sites with 421 690 distinct patterns. IQ-TREE version 1.7 package [46, 47] was used to find the best-fit substitution model and reconstruct the phylogenetic tree with the maximum-likelihood approach.

Average amino acid identity

We calculated average amino acid identity (AAI) values for all possible pairs in the dataset using the predicted peptides output from TransDecoder.LongOrfs. AAI values were generated using the aai.rb script available as part of the enveomics collection [48]. Through reciprocal all versus all protein BLAST, AAI values represent indices of pairwise genomic relatedness [49]. Since its introduction in 2005 [49] as a means for standardizing taxonomy at ranks higher than species, AAI has been extensively used in bacterial and archaeal genome-based taxonomic studies [50–52]. However, AAI has been utilized only sparsely in the fungal world (e.g. [53, 54]), with genome-based quantitative comparisons (e.g. Jaccard index of genomic distance (the fraction of shared k-mers), identification of syntenic blocks and average nucleotide identity (ANI) [14, 17]) being more heavily utilized and often for delineating lower taxonomic level (e.g. species) boundaries. AAI, however, has the advantage

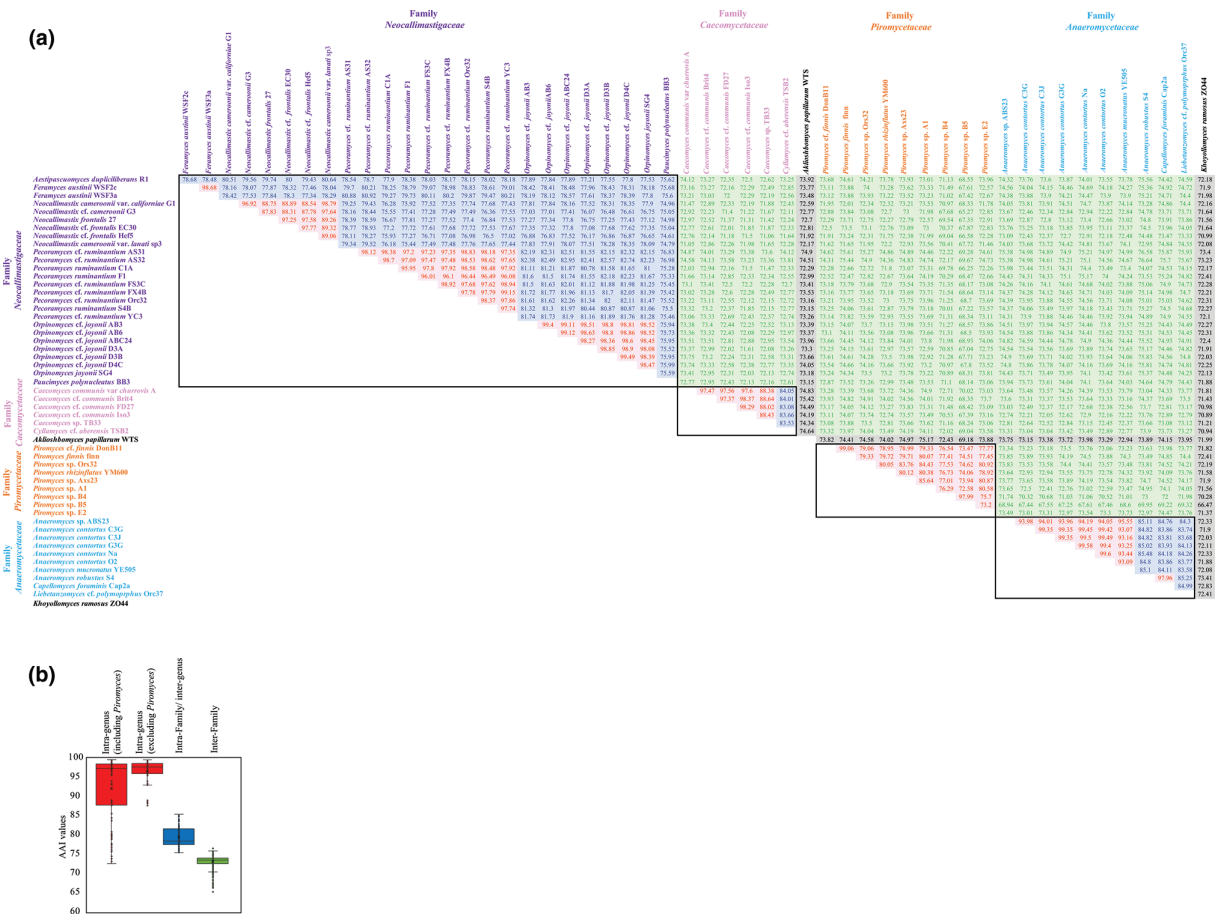

Fig. 2. Upper triangle matrix (a) and box and whisker plots (b) for the AAI values obtained for all possible pairwise comparisons of the datasets analysed in this study. (a) Isolate names in rows and columns are coloured coded by family (family *Neocallimastigaceae*, clade 1, purple; family *Caecomycetaceae*, clade 2, lavender; family *Piromycetaceae*, clade 3, orange; and family *Anaeromycetaceae*, clade 4, light blue). The AAI values for each family are shown within a thick border. Intra-genus values are shown in red text with pink highlight, intra-family/ inter-genus values are shown in blue text with light blue highlight, while inter-family values are shown in green text with light green highlight. Values for the two genera unaffiliated with the four families (genera incertae sedis) are highlighted in grey. (b) Box and whisker plots constructed using the values in (a). Intra-genus values (red) are shown both including and excluding the genus *Piromyces*. Intra-family/ inter-genus values are shown in blue. Inter-family values are shown in green. Each box plot spans the region between the 25th to 75th percentile, while the whiskers limit the minimum and maximum scores excluding the outliers. The thick line inside the box marks the median, while the 'x' corresponds to the average value.

of being readily conducted on the predicted peptides from transcriptomic datasets, as it uses amino acid sequences. The ease of obtaining transcriptomic rather than genomic sequences for AGF (mostly 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) makes the use of AAI for delineation of taxonomic boundaries more appealing.

Single gene phylogenetic analysis

Two ribosomal loci (D1/D2 LSU and ITS1) and four protein-coding gene trees (RNA polymerase II large subunit (RPB1), RNA polymerase II second largest subunit (RPB2), minichromosome maintenance complex component 7 (MCM7) and elongation factor 1-alpha (EF1α)) were evaluated. Sequences for ITS1 and D1/D2 LSU were either obtained from prior studies [5, 29–31, 33, 34, 55] or were bioinformatically extracted from genomic assemblies [56]. Amino acids sequences of RPB1, RPB2, MCM7 and EF1α were obtained from the *Anaeromyces robustus* genome (GenBank assembly accession number: GCA_002104895.1), and used as bait for BLASTp searches against all predicted proteomes in all transcriptomic datasets. Sequences for each protein, as well as for the rRNA loci were aligned using MAFFT [57] with default parameters. The alignments were used as inputs to IQ-TREE [46, 47], first to predict the best substitution model (using the lowest Bayesian information criteria (BIC)) and to generate maximum-likelihood trees under the predicted best model. The '–alrt 1000' option for performing the Shimodaira–Hasegawa approximate-likelihood ratio test (SH-aLRT), '-abayes' option for performing

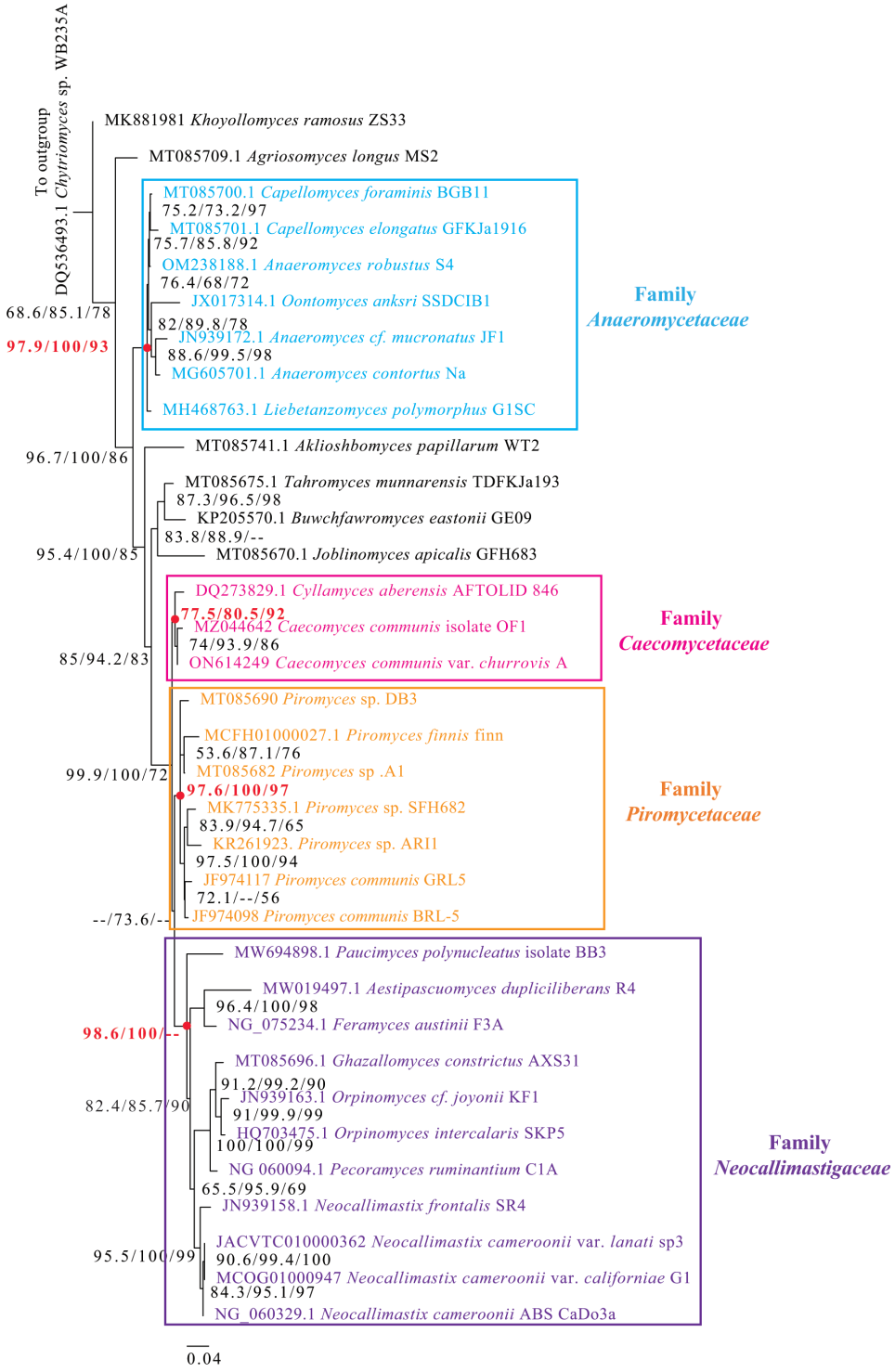

Fig. 3. Maximum-likelihood phylogenetic tree reconstructed using the D1/D2 regions of the LSU rRNA genes of all cultured and described *Neocallimastigomycota* genera. Sequences were either obtained from prior studies [5, 29–31, 33, 34, 55] or were bioinformatically extracted from genomic assemblies [56], and GenBank accession numbers are shown for each branch label. Sequences were aligned using MAFFT with default parameters. IQ-TREE [46, 47] was used to choose the best substitution model (TN+F+G4 was chosen using the lowest BIC) and to generate the maximum-likelihood tree. Support values at each node correspond to SH-aLRT, aBayes and ultrafast bootstrap. Colour code as in Fig. 1 and boxes with the same colours are used to delimit each family. The support values at the nodes corresponding to each family are shown in bold red text and the node itself is shown as a red dot. The tree was rooted (root not shown) using the D1/D2 region of the LSU rRNA gene from *Chytriomyces* sp. WB235A (GenBank accession number DQ536493.1).

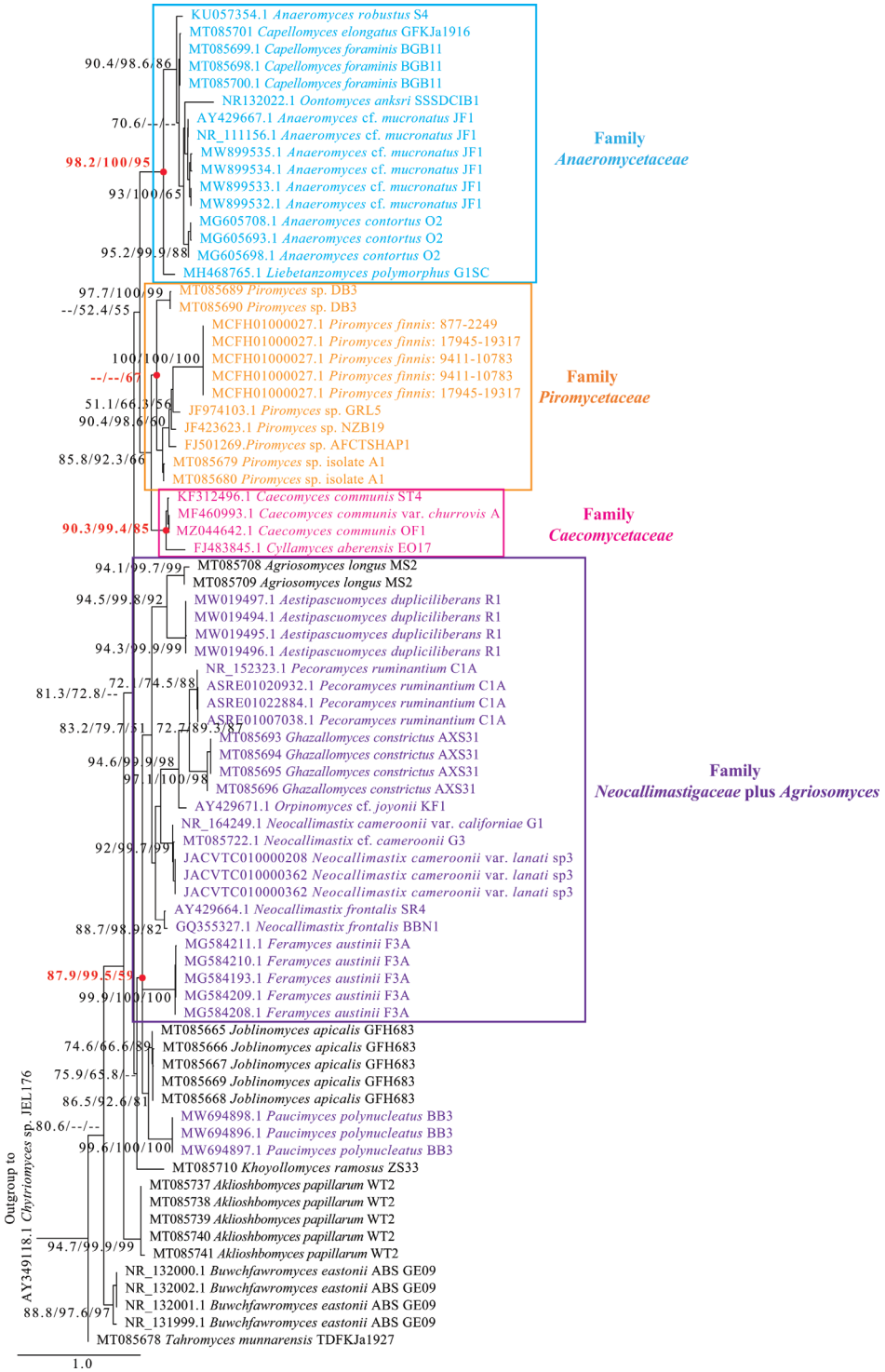

Fig. 4. Maximum-likelihood phylogenetic tree reconstructed using the ITS1 regions of all cultured and described *Neocallimastigomycota* genera. Sequences were either obtained from prior studies [5, 29–31, 33, 34, 55] or were bioinformatically extracted from genomic assemblies [56], and GenBank accession numbers are shown for each branch label. Sequences were aligned using MAFFT with default parameters. IQ-TREE [46, 47] was used to choose the best substitution model (TN+F+G4 was chosen using the lowest BIC) and to generate the maximum-likelihood tree. Support values at each node correspond to SH-aLRT, aBayes and ultrafast bootstrap. Colour code as in Fig. 1 and and boxes with the same colours are used to delimit each family. The support values at the nodes corresponding to each family are shown in bold red text and the node itself is shown as a red dot. The tree was rooted (root not shown) using the ITS1 region from *Chytriomyces* sp. JEL176 (GenBank accession number AY349118.1).

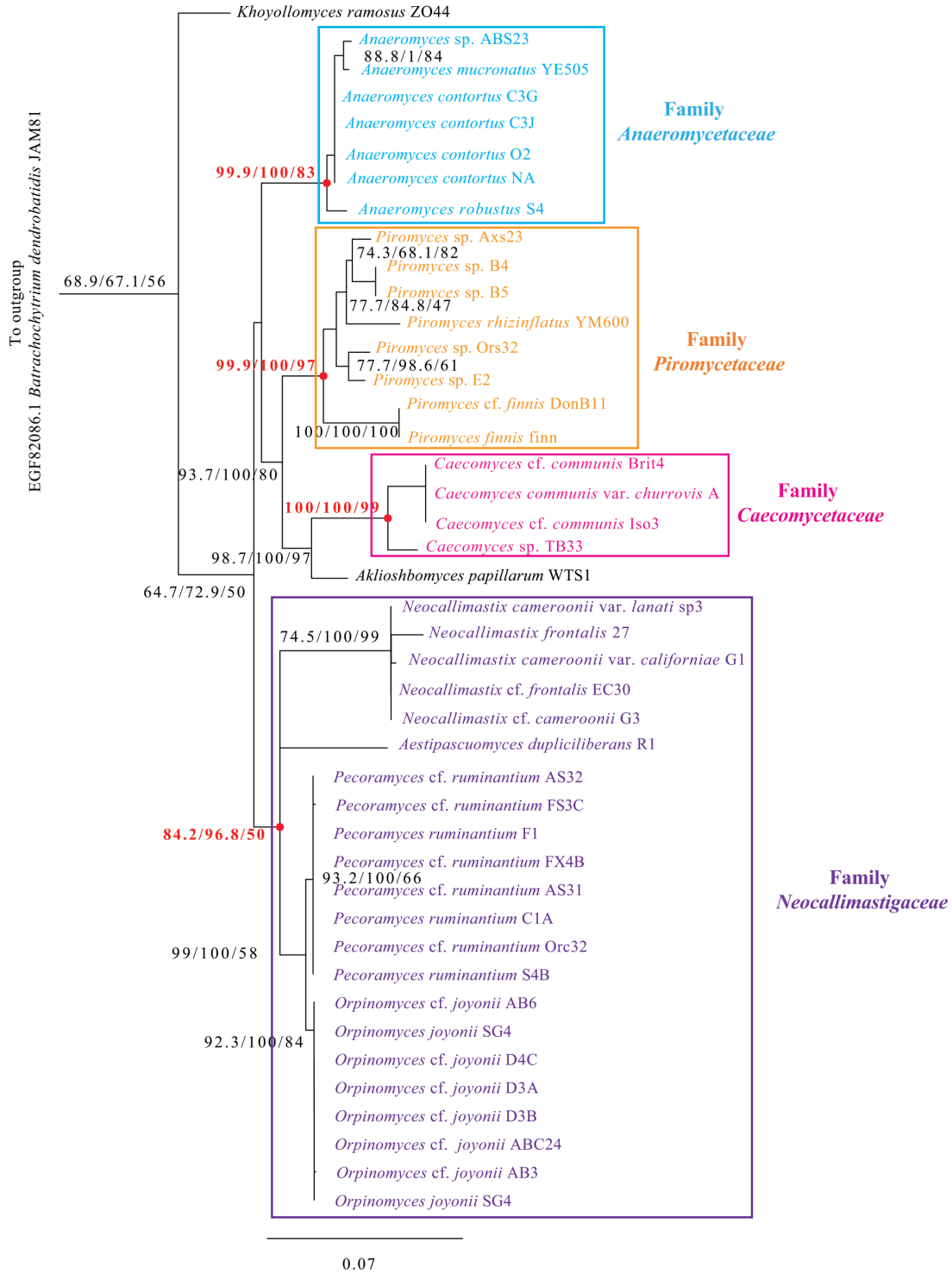

Fig. 5. Maximum-likelihood phylogenetic tree reconstructed using RPB1 amino acid sequences. The amino acid sequence of RPB1 was obtained from the *Anaeromyces robustus* genome (GenBank assembly accession number: GCA_002104895.1) and used as bait for BLASTp searches against all predicted proteomes in all transcriptomic datasets. Sequences were aligned using MAFFT with default parameters. IQ-TREE [46, 47] was used to choose the best substitution model (LG+R2 was chosen using the lowest BIC) and to generate the maximum-likelihood tree. Support values at each node correspond to SH-aLRT, aBayes and ultrafast bootstrap. Colour code as in Fig. 1 and boxes with the same colours are used to delimit each family. The support values at the nodes corresponding to each family are shown in bold red text and the node itself is shown as a red dot. The tree was rooted (root not shown) using the RPB1 sequence from *Batrachochytrium dendrobatidis* JAM81 (GenBank accession number EGF82086.1).

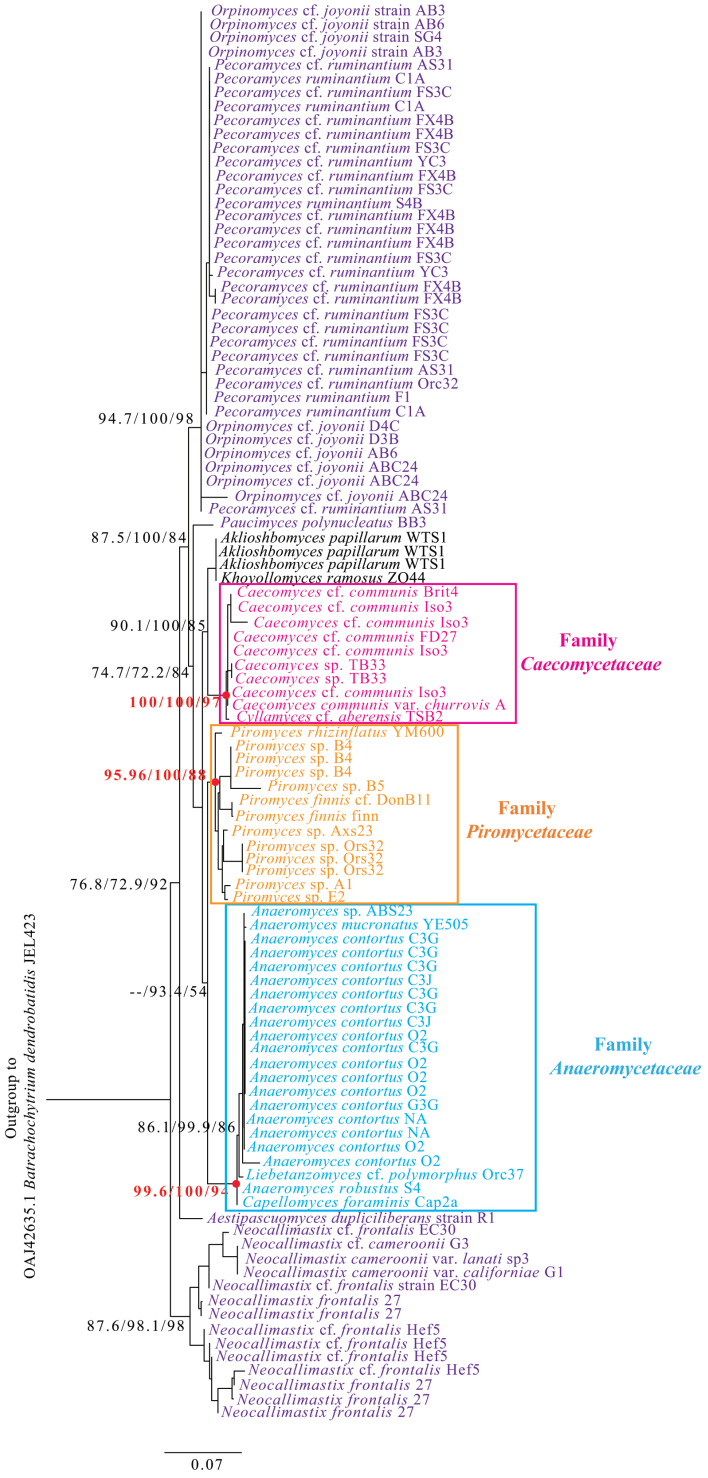

Fig. 6. Maximum-likelihood phylogenetic tree reconstructed using RPB2 amino acid sequences. The amino acid sequence of RPB2 was obtained from the *Anaeromyces robustus* genome (GenBank assembly accession number: GCA_002104895.1) and used as bait for BLASTp searches against all predicted proteomes in all transcriptomic datasets. Sequences were aligned using MAFFT with default parameters. IQ-TREE [46, 47] was used to choose the best substitution model (LG+R3 was chosen using the lowest BIC) and to generate the maximum-likelihood tree. Support values at each node correspond to SH-aLRT, aBayes and ultrafast bootstrap. Colour code as in Fig. 1 and boxes with the same colours are used to delimit each family. The support values at the nodes corresponding to each family are shown in bold red text and the node itself is shown as a red dot. The tree was rooted (root not shown) using the RPB2 sequence from *Batrachochytrium dendrobatidis* JEL423 (GenBank accession number OAJ42635.1).

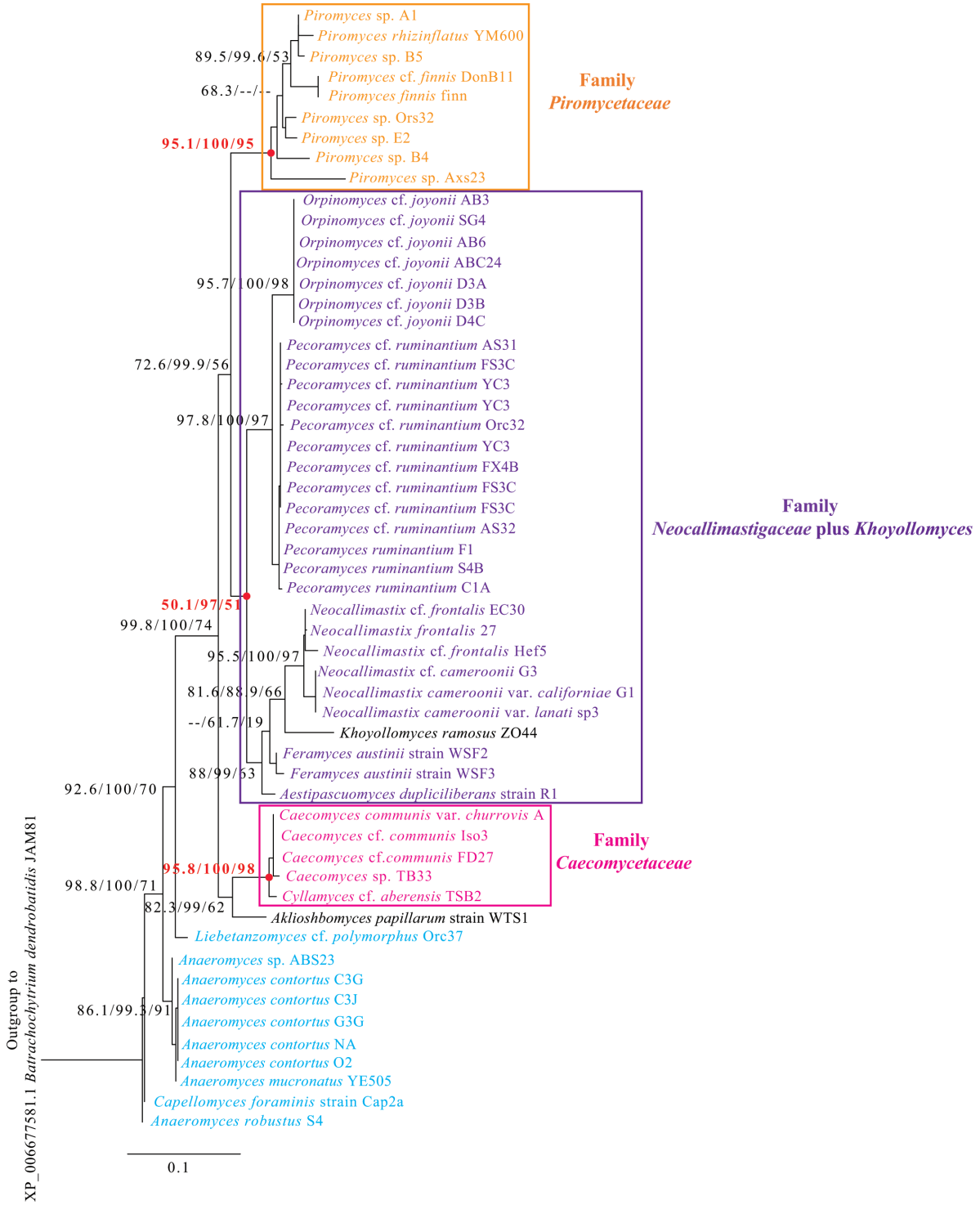

Fig. 7. Maximum-likelihood phylogenetic tree reconstructed using MCM7 amino acid sequences. The amino acid sequence of MCM7 was obtained from the *Anaeromyces robustus* genome (GenBank assembly accession number: GCA_002104895.1) and used as bait for BLASTp searches against all predicted proteomes in all transcriptomic datasets. Sequences were aligned using MAFFT with default parameters. IQ-TREE [46, 47] was used to choose the best substitution model (LG+R3 was chosen using the lowest BIC) and to generate the maximum-likelihood tree. Support values at each node correspond to SH-aLRT, aBayes and ultrafast bootstrap. Colour code as in Fig. 1 and boxes with the same colours are used to delimit each family. The support values at the nodes corresponding to each family are shown in bold red text and the node itself is shown as a red dot. The tree was rooted (root not shown) using the MCM7 sequence from *Batrachochytrium dendrobatidis* JAM81 (GenBank accession number XP_006677581.1).

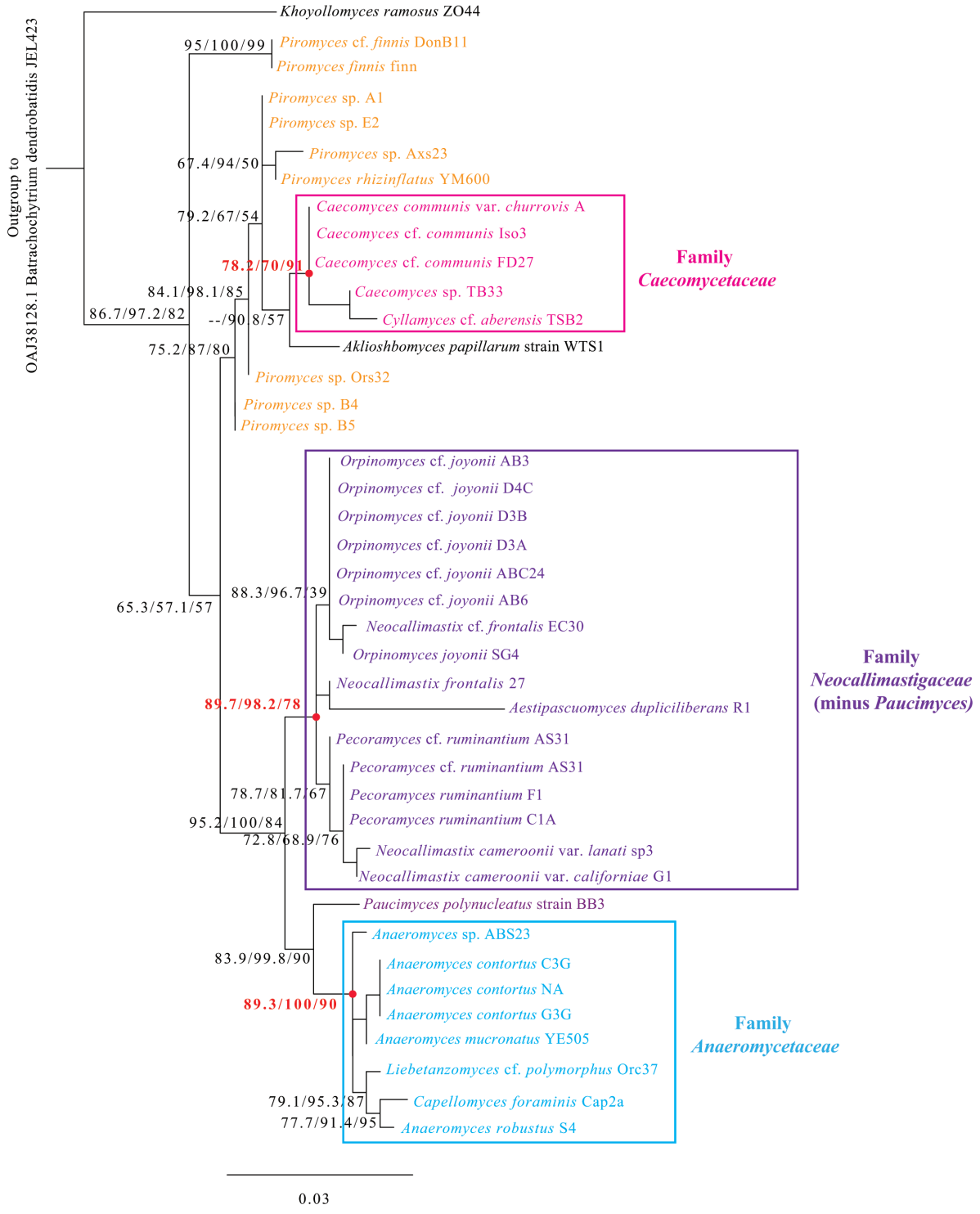

Fig. 8. Maximum-likelihood phylogenetic tree reconstructed using EF-1A amino acid sequences. The amino acid sequence of EF-1A was obtained from the *Anaeromyces robustus* genome (GenBank assembly accession number: GCA_002104895.1) and used as bait for BLASTp searches against all predicted proteomes in all transcriptomic datasets. Sequences were aligned using MAFFT with default parameters. IQ-TREE [46, 47] was used to choose the best substitution model (LG+R2 was chosen using the lowest BIC) and to generate the maximum-likelihood tree. Support values at each node correspond to SH-aLRT, aBayes and ultrafast bootstrap. Colour code as in Fig. 1 and boxes with the same colours are used to delimit each family. The support values at the nodes corresponding to each family are shown in bold red text and the node itself is shown as a red dot. The tree was rooted (root not shown) using the EF-1A sequence from *Batrachochytrium dendrobatidis* JEL423 (GenBank accession number OAJ38128.1).

approximate Bayes tests and the '–bb 1000' option for ultrafast bootstrap (UFB) were added to the IQ-TREE command line, which resulted in the generation of phylogenetic trees with three support values (SH-aLRT, aBayes and UFB) on each branch.

Nucleotide sequencing accession number

Raw Illumina RNA-seq read sequences are deposited in GenBank under the BioProject accession number PRJNA847081 and BioSample accessions numbers SAMN28920465–SAMN28920484. Individual SRA accessions are provided in Table 1.

RESULTS

Sequencing

Transcriptomic sequencing yielded 15.6–23.8 (average 19.82) million reads that were assembled into 22 649–106 687 total transcripts, 20 599–103 405 distinct transcripts (clustering at 95%; average 40 099) and 13 858–28 405 predicted peptides (average 19 667) (Table S1 available with the online version of this article). Assessment of transcriptome completion using BUSCO yielded high values (73.63–99.5%) for all assemblies (Table S1).

Resolving inter-genus relationships in the *Neocallimastigomycota*

Multiple supra-genus relationships were well supported in all phylogenomic outputs. Four distinct clades were observed (Fig. 1 and Table 2). Clade 1 constituted members of the genera *Pecoramyces*, *Orpinomyces*, *Neocallimastix*, *Feramyces*, *Paucimyces* and *Aestipascuomyces*. Within this large clade, a strong support for *Pecoramyces* and *Orpinomyces* association, as well as for *Neocallimastix*, *Aestipascuomyces* and *Feramyces* association was observed (Fig. 1). The genus *Paucimyces* formed a distinct branch at the base of the clade. Phenotypically, all members of the clade, except for the genera *Pecoramyces* and *Paucimyces*, produce polyflagellated zoospores. Clade 2 constituted members of the genera *Cyllamyces* and *Caecomyces*. Phenotypically, this clade encompasses the two genera exhibiting a bulbous rhizoidal growth pattern in the *Neocallimastigomycota*. Clade 3 constituted members of the genus *Piromyces*. Compared to all other AGF genera, the genus *Piromyces* currently exhibits high intra-genus sequence divergence based on ITS1 and D1/D2 LSU analysis [3]. The genus was first defined to encompass all phenotypes with monocentric thalli, a filamentous rhizoidal system and monoflagellated (1–4 flagella) zoospores [58]. However, subsequent isolation efforts clearly demonstrated that such phenotype is prevalent in a wide range of phylogenetically disparate genera across the *Neocallimastigomycota* [4, 28, 34]. Currently, *Piromyces* encompasses all taxa phylogenetically affiliated with the first described monocentric, monoflagellated and filamentous isolate (*Piromyces communis* [58]). Clade 4 constituted members of the genera *Anaeromyces*, *Liebetanzomyces* and *Capellomyces*. The clade encompasses genera with a filamentous rhizoidal system and monoflagellated (1–4 flagella) zoospores. The genus *Anaeromyces* produces polycentric thalli, while the genera *Liebetanzomyces* and *Capellomyces* produce monocentric thalli.

Few genera clustered outside these four clades described above. The genus *Aklioshbomyces* formed a distinct branch at the base of clade 2 (Fig. 1). Finally, the position of the genus *Khoyollomyces* was unique and solitary, being consistently located at the base of the tree, suggesting its deep-branching and relatively ancient origin.

Estimating AAI identities

AAI values were estimated using the entire dataset of predicted peptides (Fig. 2). Intra-genus AAI values ranged between 72.58–99.6% (average 92.16±8.55). However, the low intra-genus divergence estimates were only confined to the broadly circumscribed genus *Piromyces*. Indeed, excluding *Piromyces* from this analysis, intra-genus AAI values ranged between 87.78–99.6% (average 95.67±3.41). Pairwise AAI values for members of different genera within the same clade (intra-clade inter-genus AAI values) ranged between 74.61–85.48% (average 79.22±2.63). Maximum intra-clade inter-genus divergence was observed between members of the genus *Paucimyces* and other genera in clade 1 (average 75.55±0.55), while minimal intra-clade inter-genus divergence were observed between *Caecomyces* and *Cyllamyces* in clade 2 (83.7%±0.4), as well as between the genera *Anaeromyces* and *Capellomyces* (average 84.5±0.57), the genera *Anaeromyces* and *Liebetanzomyces* (average 83.9±0.3), and the genera *Capellomyces* and *Liebetanzomyces* (average 85.1±0.18) in clade 4. Inter-clade AAI values averaged 73.15±1.57, and ranged between 65.27% (between members of the genera *Piromyces* and *Neocallimastix*) and 76.64% (between members of the genera *Capellomyces* and *Pecoramyces*).

Single gene phylogenetic analysis for resolving AGF inter-genus relationships

We tested whether supra-genus clades topology as well as within clades inter-genus relationships observed in phylogenomic analysis were retained in single gene phylogenies (Figs 3–8). One ribosomal locus (D1/D2 LSU) and one protein-coding gene (RPB1) retained the monophyly of all four clades described above (Figs 3 and 5, Table S2). As well, both D1/D2 LSU and RPB1 phylogenies resolved all inter-genus relationships within all clades in the *Neocallimastigomycota* (Figs 3 and 5). On the other hand, ITS1, RPB2, MCM7 and EF1α phylogenies each recovered two or three out of the four supra-genus clades delineated above. The monophyly of clade one was not retained in ITS1 and RPB2 phylogenies (Figs 4 and 6, Table S2), the monophyly of

clade four was not retained in the MCM7 phylogeny (Fig. 7) and the monophyly of clades 1 and 3 was not retained in the EF1α phylogeny (Fig. 8). Further, within the clades that were supported, few inter-genus relationships were compromised in ITS1 (genus *Anaeromyces*) and EF1α (genera *Neocallimastix*, *Orpinomyces* and *Pecoramyces*) phylogenies.

DISCUSSION

Identifying and circumscribing supra-genus relationships within the *Neocallimastigomycota*

Our phylogenomic analysis identified four distinct statistically supported supra-genus clades in the *Neocallimastigomycota* (Table 2, Fig. 1). Clade boundaries were based on phylogenomic tree topologies, while taking taxonomically informative morphological characteristics into account. For example, phylogenomic analyses placed the genus *Aklioshbomyces* at the base of clade 2. Exclusion of *Aklioshbomyces* from clade 2 was based on its filamentous rhizoidal growth pattern, which contrasts the bulbous growth pattern exclusive to both genera (*Caecomyces* and *Cyllamyces*) constituting clade 2. Therefore, we propose conferring an 'incertae sedis' status on *Aklioshbomyces*, pending availability of additional isolates and description of additional novel species and closely related genera.

AAI values were further examined to quantitatively circumscribe these clades. A clear delineation of the clade boundaries was evident using AAI values (Fig. 2). Within the genus, AAI values ranged between 87.78–99.6% (or 72.58–99.6% if including values for the broadly circumscribed genus *Piromyces*). Inter-genus/ Intra-clade AAI estimates ranged between 74.61–85.48% (*Piromyces* excluded), while inter-clade values ranged between 65.27–76.64% (genera incertae sedis excluded, Fig. 2). These values are similar to AAI values estimated for delineating the *Ascomycetes* family *Hypoxylaceae* [53], but are higher than the arbitrary cutoffs used for delineating taxa in the prokaryotic world (~45–65% for family, ~65–95% for genus [50]). Therefore, we suggest using 85.0 and 75.0% AAI cutoff values as a guide for circumscribing genera, and families, respectively, in the *Neocallimstigomycota*. Currently, the genus *Piromyces* represents the sole genus in clade 3. AAI estimates using the currently available *Piromyces* species –omics datasets suggest a broader intra-genus AAI range when compared to other genera (Fig. 2). This is a reflection of the fact that the genus was originally circumscribed based on phenotypic, rather than a combination of phenotypic and molecular data. Future detailed analysis of available members of the genus is required to fully resolve relationships between members of this broadly circumscribed genus.

Up to this point, only ITS1 and D1/D2 LSU loci have been evaluated for the assessment of phylogenetic positions of genera within the *Neocallimastigomycota*, as well as for ecological culture-independent surveys [5, 7]. To test the utility of other phylomarkers commonly utilized in fungal taxonomy, we assessed four additional protein-coding genes, and examined concordance between each of the six loci (ribosomal ITS1 and D1/D2 LSU, and RPB1, RPB2, MCM7, and EF-1α) and phylogenomic trees topologies. Our results demonstrate that D1/D2 LSU, currently regarded as the phylomarker of choice for genus-level delineation [5, 59] and utilized as a marker in culture-independent diversity surveys [5], is equally useful in resolving supra-genus clades delineated by phylogenomics (Figs 3 and S1). As well, our results add the protein-coding gene RPB1 to the list of phylomarkers that could be used for inter-genus, and supra-clade delineation (Figs 5 and S2). As such, identity values of 91.5 and 97.9% for LSU and RPB1, respectively (these values correspond to the 75th percentile value for intra-clade inter-genus divergence based on the distance matrix from the alignments used to generate the maximum-likelihood trees in Figs 3 and 5) seem to circumscribe these clades. The high sequence similarity in the protein-coding gene RPB1 is quite surprising since, typically, higher levels of divergence are usually observed in protein-coding genes when compared to the non-protein-coding ribosomal RNA genes [60]. Other phylomarkers tested here were only successful in resolving two to three of the four clades, and some also compromised intra- and inter-genus relationships (Figs 4, 6 and 7). Such failure to resolve genus-level relationships appears to be a function of high sequence similarities in these genes. For example, the inter-genus divergence values between *Orpinomyces* and *Pecoramyces* RPB2 sequences ranged between 0–1.8%, which are comparable to the values within the genus *Orpinomyces*. This has resulted in failure of RPB2 to resolve the *Orpinomyces–Pecoramyces* relationship. The unreliability of the ITS1 locus for clade delineation has been described before and is mainly due to length variability between genera and high within-strain sequence divergence [5, 7].

Phylogenetic position of taxa currently lacking genome or transcriptome sequences

The 52 transcriptomic datasets examined cover 14 out of the 20 currently described AGF genera. The remaining six genera (*Oontomyces*, *Buwchfawromyces*, *Agriosomyces*, *Ghazallomyces*, *Tahromyces* and *Joblinomyces*) are all currently represented by a single species. Further, most of these genera appear to exhibit extremely limited geographic and animal host distribution patterns [4, 5, 28, 34]. The phylogenetic position of these six genera could hence be only evaluated using available D1/D2 LSU (and to some extent ITS1) sequence data from taxa description publications. D1/D2 LSU phylogeny strongly supports placement of the genus *Ghazallomyces* as a member of clade 1 (Figs 3 and 4) [34]. Further, members of this genus produce polyflagellated zoospores (a trait observed in four out of the six genera of clade 1), a filamentous rhizoid (similar to all taxa in clade 1) and monocentric thalli (a trait observed in four out of the six genera of clade 1), further supporting its recognition as member of clade 1 [34]. Similarly, phylogenetic analysis using D1/D2 LSU supports the placement of the genus *Oontomyces* as a member of clade 4 (Figs 3 and 4).

Members of the genus *Oontomyces* exhibit similar phenotypes (monocentric thalli, monoflagellated zoospores and filamentous rhizoidal growth patterns) to two (*Liebetanzomyces* and *Capellomyces*) out of the three genera in the clade [28].

Interestingly, phylogenetic analysis using the D1/D2 region of LSU rRNA genes places three of the genera for which no –omics data is available (*Buwchfawromyces*, *Tahromyces* and *Joblinomyces*) in a single distinct monophyletic clade (Fig. 3). Future availability of –omics data is needed to confirm such topology. Finally, while the genus *Agriosomyces* has a distinct position in both LSU and ITS1 phylogenies (Figs 3 and 4), no clear association to any of the clades was apparent. As such, –omics data is needed to resolve the position of such genus.

Rank assignment for supra-genus clades in the *Neocallimastigomycota*

We propose retaining all currently described AGF genera in a single order (*Neocallimastigales*) and a single class (*Neocallimastigomycetes*) in the phylum *Neocallimastigomycota*. This proposition is based on the lack of fundamental differences in cellular structures, metabolic capabilities, ecological distribution and lifecycle phases in all currently described genera, coupled with the observed AAI values, when compared to the few studies utilizing this approach in fungi [53].

Beyond the four clades described above, we refrain from proposing additional families for the genus *Agriosomyces*, or the D1/D2 LSU-defined and well-supported clade encompassing the genera *Buwchfawromyces*, *Tahromyces* and *Joblinomyces* pending the availability of confirmatory phylogenomic data. In addition, we refrain from proposing new families for the genera *Khoyollomyces* and *Aklioshbomyces* due to their current solitary positions in phylogenomic trees (Fig. 1), although such proposition would be justified by the isolation and characterization of additional novel taxa and the availability of –omics data from such taxa. Such genera should be designated as 'genera incertae sedis' within the order *Neocallimastigales* for the present time. The proposed novel families would be named after the first described genus within the clade: clade 1, *Neocallimastigaceae* comprising the genera *Neocallimastix* [21, 61, 62], *Ghazallomyces* [34, 63, 64], *Pecoramyces* [29, 30], *Aestipascuomyces* [33] and *Paucimyces* [31]; clade 2, *Caecomycetaceae* fam. nov., comprising the genera *Caecomyces* [58] and *Cyllamyces* [32]; clade 3, *Piromycetaceae* fam. nov., comprising the genus *Piromyces* [58]; and clade 4, *Anaeromycetaceae*, comprising the genera *Anaeromyces* [65], *Capellomyces* [34], *Liebetanzomyces* [66] and *Oontomyces* [28]. Such arrangement would necessitate amending the description of the family *Neocallimastigaceae*, currently encompassing all 20 genera, to include only the seven genera stated above, rather than all 20 currently described AGF genera, as well as assigning the genera *Anaeromyces*, *Capellomyces* [34], *Liebetanzomyces* [66] and *Oontomyces* to the previously proposed (IF550425) nomenclature novelty family *Anaeromycetaceae*.

EMENDED DESCRIPTION OF FAMILY *NEOCALLIMASTIGACEAE*

Obligate anaerobic fungi with monocentric (genera *Neocallimastix*, *Ghazallomyces*, *Pecoramyces*, *Feramyces* and *Aestipascuomyces*) or polycentric (genera *Orpinomyces* and *Paucimyces*) thalli and filamentous rhizoidal system. Zoospores are polyflagellated (genera *Neocallimastix*, *Ghazallomyces*, *Orpinomyces*, *Feramyces* and *Aestipascuomyces*) or monoflagellated (genera *Pecoramyces* and *Paucimyces*). The clade is circumscribed by phylogenomic analysis and AAI values, and confirmed by LSU and RPB1 phylogenetic analyses, as well as morphological characteristics, and currently accommodates the genera *Neocallimastix* [21, 61, 62], *Ghazallomyces* [34, 63, 64], *Pecoramyces* [29], *Feramyces* [30], *Aestipascuomyces* [33] and *Paucimyces* [31].

The emended description of the family *Neocallimastigaceae* is generally similar to that provided for the family *Neocallimastigaceae* [21] and order *Neocallimastigales* [22], with amendments to exclude bulbous rhizoidal growth and to circumscribe its boundaries to encompass a monophyletic clade of seven genera.

Type genus: *Neocallimastix*

Mycobank ID: MB25486

DESCRIPTION OF *CAECOMYCETACEAE* FAM. NOV.

Obligate anaerobic fungi that produce monoflagellated zoospores, monocentric or polycentric thalli that are either uni- or multisporangiate, and a bulbous rhizoidal system with spherical holdfasts. The clade is circumscribed by phylogenomic analysis and AAI values, and confirmed by LSU and RPB1 phylogenetic analyses, as well as morphological characteristics. Currently accommodates the genera *Caecomyces* [58] and *Cyllamyces* [32].

Type genus: *Caecomyces* [58]

Mycobank ID: MB844401

DESCRIPTION OF *PIROMYCETACEAE* FAM. NOV.

Obligate anaerobic fungi that produce monoflagellated zoospores, monocentric thalli and a filamentous rhizoidal system. The clade is circumscribed by phylogenomic analysis and AAI values, and confirmed by LSU and RPB1 phylogenetic analyses, as well as morphological characteristics. Currently accommodates the genus *Piromyces* [58].

Type genus: *Piromyces* [58]

Mycobank ID: MB844402

DESCRIPTION OF *ANAEROMYCETACEAE* FAM. NOV.

Obligate anaerobic fungi that produce monoflagellated zoospores, monocentric or polycentric thalli, and a filamentous rhizoidal system. The clade is circumscribed by phylogenomic analysis and AAI values, and confirmed by LSU and RPB1 phylogenetic analyses, as well as morphological characteristics. Currently accommodates the genera *Anaeromyces* [65], *Capellomyces* [34], *Liebetanzomyces* [66] and *Oontomyces* [28].

Type genus: *Anaeromyces* [65]

Mycobank ID: MB550425

GENERA INCERTAE SEDIS

The following genera are assigned as 'genera incertae sedis' within the order *Neocallimastigales*: *Buwchfawromyces* ([4]; Mycobank ID: MB550797), *Joblinomyces* ([34]; Mycobank ID: MB830867), *Tahromyces* ([34]; Mycobank ID: MB830865), *Agriosomyces* ([34]; Mycobank ID: MB830737), *Aklioshbomyces* ([34]; Mycobank ID: MB830735) and *Khoyollomyces* ([34]; Mycobank ID: MB830741).

Funding information
This work has been supported by NSF grant 2029478 to N.H.Y. and M.S.E.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Gruninger RJ, Puniya AK, Callaghan TM, Edwards JE, Youssef N, *et al*. Anaerobic fungi (phylum *Neocallimastigomycota*): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiol Ecol* 2014;90:1–17.
2. Hanafy RA, Dagar SS, Griffith GW, Pratt CJ, Youssef NH, *et al*. Taxonomy of the anaerobic gut fungi (*Neocallimastigomycota*): a review of classification criteria and description of current taxa. *Int J Syst Evol Microbiol* 2022;72.
3. Elshahed MS, Hanafy RA, Cheng Y, Dagar SS, Edwards JE, *et al*. Characterization and rank assignment criteria for the anaerobic fungi (*Neocallimastigomycota*). *Int J Syst Evol Microbiol* 2022;72.
4. Callaghan TM, Podmirseg SM, Hohlweck D, Edwards JE, Puniya AK, *et al*. *Buwchfawromyces eastonii* gen. nov., sp. nov.: a new anaerobic fungus (*Neocallimastigomycota*) isolated from buffalo faeces. *MC* 2015;9:11–28.
5. Hanafy RA, Johnson B, Youssef NH, Elshahed MS. Assessing anaerobic gut fungal diversity in herbivores using D1/D2 large ribosomal subunit sequencing and multi-year isolation. *Environ Microbiol* 2020;22:3883–3908.
6. Wang X, Liu X, Groenewald JZ. Phylogeny of anaerobic fungi (phylum *Neocallimastigomycota*), with contributions from yak in China. *Antonie van Leeuwenhoek* 2017;110:87–103.
7. Edwards JE, Hermes GDA, Kittelmann S, Nijsse B, Smidt H. Assessment of the accuracy of high-throughput sequencing of the ITS1 region of *Neocallimastigomycota* for community composition analysis. *Front Microbiol* 2019;10:2370.
8. Stabel M, Schweitzer T, Haack K, Gorenflo P, Aliyu H, *et al*. Isolation and biochemical characterization of six anaerobic fungal strains from zoo animal feces. *Microorganisms* 2021;9:1655.
9. Choi J, Kim S-H. A genome tree of life for the fungi kingdom. *Proc Natl Acad Sci*2017;114:9391–9396.
10. Montoliu-Nerin M, Sánchez-García M, Bergin C, Kutschera VE, Johannesson H, *et al*. In-depth phylogenomic analysis of arbuscular mycorrhizal fungi based on a comprehensive set of de novo genome assemblies. *Front Fungal Biol* 2021;2.
11. Galindo LJ, López-García P, Torruella G, Karpov S, Moreira D. Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. *Nat Commun* 2021;12:4973.
12. Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, *et al*. A genome-scale phylogeny of the kingdom fungi. *Curr Biol* 2021;31:1653–1665..
13. James TY, Stajich JE, Hittinger CT, Rokas A. Toward a fully resolved fungal tree of life. *Annu Rev Microbiol* 2020;74:291–313.
14. Gostinčar C. Towards genomic criteria for delineating fungal species. *J Fungi (Basel)* 2020;6:246.
15. Pornputtapong N, Acheampong DA, Patumcharoenpol P, Jenjaroenpun P, Wongsurawat T, *et al*. KITSUNE: a tool for identifying empirically optimal k-mer length for alignment-free phylogenomic analysis. *Front Bioeng Biotechnol* 2020;8:556413.
16. Gramaje D, Berlanas C, Martínez-Diz MDP, Diaz-Losada E, Antonielli L, *et al*. Comparative genomic analysis of *Dactylonectria torresensis* strains from grapevine, soil and weed highlights potential mechanisms in pathogenicity and endophytic lifestyle. *J Fungi (Basel)* 2020;6:255.
17. Li C, Guo Z, Zhou S, Han Q, Zhang M, *et al*. Evolutionary and genomic comparisons of hybrid uninucleate and nonhybrid *Rhizoctonia* fungi. *Commun Biol* 2021;4:201.
18. Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A. A robust phylogenomic time tree for biotechnologically and medically

important fungi in the genera *Aspergillus* and *Penicillium mBio* 2019;10:e00925-19.

19. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, *et al*. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol* 2019;66:4–119.
20. Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, *et al*. A higher-level phylogenetic classification of the fungi. *Mycol Res* 2007;111:509–547.
21. Heath IB, Bauchop T, Skipp RA. Assignment of the rumen anaerobe Neocallimastix frontalis to the *Spizellomycetales* (*Chytridiomycetes*) on the basis of its polyflagellate zoospore ultrastructure. *Can J Bot* 1983;61:295–307.
22. Li J, Heath IB, Packer L. The phylogenetic relationships of the anaerobic chytridiomycetous gut fungi (*Neocallimasticaceae*) and the *Chytridiomycota*. II. Cladistic analysis of structural data and description of *Neocallimasticales* ord.nov. *Can J Bot* 1993;71:393–407.
23. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V *et al*. Reconstructing the early evolution of fungi using a six-gene phylogeny. Nature 2006;443:818-822.
24. Tedersoo L, Sánchez-Ramírez S, Kõljalg U, Bahram M, Döring M, *et al*. High-level classification of the fungi and a tool for evolutionary ecological analyses. *Fungal Diversity* 2018;90:135–159.
25. Powell MJ, Letcher PM. *Chytridiomycota*, *Monoblepharidomycota*, and *Neocallimastigomycota*. In: McLaughlin DJ and Spatafora JW (eds). *Systematics and Evolution*. Berlin, Heidelberg: Springer; 2014. pp. 141–175.
26. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, *et al*. GenBank. *Nucl Acids Res* 2018;46:D41–D47.
27. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, *et al*. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucl Acids Res* 2014;42:D699–D704.
28. Dagar SS, Kumar S, Griffith GW, Edwards JE, Callaghan TM, *et al*. A new anaerobic fungus (*Oontomyces anksri* gen. nov., sp. nov.) from the digestive tract of the Indian camel (*Camelus dromedarius*). *Fungal Biol* 2015;119:731–737.
29. Hanafy RA, Elshahed MS, Liggenstoffer AS, Griffith GW, Youssef NH. *Pecoramyces ruminantium*, gen. nov., sp. nov., an anaerobic gut fungus from the feces of cattle and sheep. *Mycologia* 2017;109:231–243.
30. Hanafy RA, Elshahed MS, Youssef NH. *Feramyces austinii*, gen. nov., sp. nov., an anaerobic gut fungus from rumen and fecal samples of wild Barbary sheep and fallow deer. *Mycologia* 2018;110:513–525.
31. Hanafy RA, Youssef NH, Elshahed MS. *Paucimyces polynucleatus* gen. nov, sp. nov., a novel polycentric genus of anaerobic gut fungi from the faeces of a wild blackbuck antelope. *Int J Syst Evol Microbiol* 2021;71.
32. Ozkose E, Thomas BJ, Davies DR, Griffith GW, Theodorou MK. *Cyllamyces aberensis* gen.nov. sp.nov., a new anaerobic gut fungus with branched sporangiophores isolated from cattle. *Can J Bot* 2001;79:666–673.
33. Stabel M, Hanafy RA, Schweitzer T, Greif M, Aliyu H, *et al*. *Aestipascuomyces dupliciliberans* gen. nov, sp. nov., the first cultured representative of the uncultured SK4 clade from aoudad sheep and alpaca. *Microorganisms* 2020;8:1734.
34. Hanafy RA, Lanjekar VB, Dhakephalkar PK, Callaghan TM, Dagar SS, *et al*. Seven new *Neocallimastigomycota* genera from wild, zoo-housed, and domesticated herbivores greatly expand the taxonomic diversity of the phylum. *Mycologia* 2020;112:1212–1239.
35. Calkins S, Elledge NC, Hanafy RA, Elshahed MS, Youssef N. A fast and reliable procedure for spore collection from anaerobic fungi: application for RNA uptake and long-term storage of isolates. *J Microbiol Methods* 2016;127:206–213.
36. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;28:3150–3152.
37. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 2021;38:4647–4654.
38. Gruninger RJ, Nguyen TTM, Reid ID, Yanke JL, Wang P, *et al*. Application of transcriptomics to compare the carbohydrate active enzymes that are expressed by diverse genera of anaerobic fungi to degrade plant cell wall carbohydrates. *Front Microbiol* 2018;9:1581.
39. Brown JL, Swift CL, Mondo SJ, Seppala S, Salamov A, *et al*. Co-cultivation of the anaerobic fungus *Caecomyces churrovis* with *Methanobacterium bryantii* enhances transcription of carbohydrate binding modules, dockerins, and pyruvate formate lyases on specific substrates. *Biotechnol Biofuels* 2021;14:234.
40. Haitjema CH, Gilmore SP, Henske JK, Solomon KV, de Groot R, *et al*. A parts list for fungal cellulosomes revealed by comparative genomics. *Nat Microbiol* 2017;2:17087.
41. Henske JK, Gilmore SP, Knop D, Cunningham FJ, Sexton JA, *et al*. Transcriptomic characterization of *Caecomyces churrovis*: a novel, non-rhizoid-forming lignocellulolytic anaerobic fungus. *Biotechnol Biofuels* 2017;10:305.
42. Li Y, Li Y, Jin W, Sharpton TJ, Mackie RI, *et al*. Combined genomic, transcriptomic, proteomic, and physiological characterization of the growth of *Pecoramyces* sp. F1 in monoculture and co-culture with a syntrophic methanogen. *Front Microbiol* 2019;10:435.
43. Wilken SE, Monk JM, Leggieri PA, Lawson CE, Lankiewicz TS, *et al*. Experimentally validated reconstruction and analysis of a genome-scale metabolic model of an anaerobic *Neocallimastigomycota* fungus. *mSystems* 2021;6:e00002-21.
44. Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, *et al*. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biol Evol* 2015;7:1590–1601.
45. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, *et al*. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet* 2017;49:964–968.
46. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
47. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
48. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 2016. DOI: 10.7287/peerj.preprints.1900v1.
49. Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 2005;187:6258–6264.
50. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. *ISME J* 2017;11:2399–2406.
51. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, *et al*. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017;2:1533–1542.
52. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, *et al*. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.
53. Wibberg D, Stadler M, Lambert C, Bunk B, Spröer C, *et al*. High quality genome sequences of thirteen *Hypoxylaceae* (*Ascomycota*) strengthen the phylogenetic family backbone and enable the discovery of new taxa. *Fungal Diversity* 2021;106:7–28.
54. Wang K, Sipilä T, Overmyer K. A novel *Arabidopsis* phyllosphere resident *Protomyces* species and a re-examination of genus *Protomyces* based on genome sequence data. *IMA Fungus* 2021;12:8.
55. Hanafy RA, Johnson B, Elshahed MS, Youssef NH. *Anaeromyces contortus*, sp. nov., a new anaerobic gut fungal species (*Neocallimastigomycota*) isolated from the feces of cow and goat. *Mycologia* 2018;110:502–512.
56. Solomon KV, Haitjema CH, Henske JK, Gilmore SP, Borges-Rivera D, *et al*. Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. *Science* 2016;351:1192–1195.
57. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.

58. Gold JJ, Heath IB, Bauchop T. Ultrastructural description of a new chytrid genus of caecum anaerobe, *Caecomyces equi* gen. nov., sp. nov., assigned to the *Neocallimasticaceae*. *Biosystems* 1988;21:403–415.
59. Dagar SS, Kumar S, Mudgil P, Singh R, Puniya AK. D1/D2 domain of large-subunit ribosomal DNA for differentiation of *Orpinomyces* spp. *Appl Environ Microbiol* 2011;77:6722–6725.
60. Kandawatte Wedaralalage TC, Jayawardena RS, Hyde KD. Hurdles in fungal taxonomy: effectiveness of recent methods in discriminating taxa. *MT* 2020;1:114–122.
61. Braune R. Untersuchungen uber die im wiederkauermagen vorkommenden protozoen (investigations into the protozoa occurring in the ruminant stomach). *Archiv fiir Protistenkunde* 1913;32:1–170.
62. Vavra J, Joyon L. Etude sur la morphologie, le cycle ivolutif et la position systematique de *Callimastix cyclopsis* Weissenberg 1912. (study on the morphology, the evolutionary cycle and the systematic position of *Callimastix cyclopsis* Weissenberg 1912). *Protistologica* 1966;2:5–15.
63. Breton A, Bernalier A, Bonnemoy F, Fonty G, Gaillard B, *et al*. Morphological and metabolic characterization of a new species of strictly anaerobic rumen fungus: *Neocallimastix joyonii* . *FEMS Microbiol Lett* 1989;58:309–314.
64. Barr DJ, KudO H, Jakober KD, Cheng KJ. Morphology and development of rumen fungi: *Neocallimastix* sp., *Piromyces communis*, and *Orpinomyces bovis* gen.nov., sp.nov. Can J Bot 1989;67:2815–2824
65. Breton A, Bernalier A, Dusser M, Fonty G, Gaillard-Martinie B, *et al*. *Anaeromyces mucronatus* nov. gen., nov. sp. A new strictly anaerobic rumen fungus with polycentric thallus. *FEMS Microbiol Lett* 1990;70:177–182.
66. Joshi A, Lanjekar VB, Dhakephalkar PK, Callaghan TM, Griffith GW, *et al*. *Liebetanzomycespolymorphus* gen. et sp. nov., a new anaerobic fungus (Neocallimastigomycota) isolated from the rumen of a goat. *MycoKeys* 2018;40:89–110.
67. Murphy CL, Youssef NH, Hanafy RA, Couger MB, Stajich J, *et al*. Horizontal gene transfer forged the evolution of anaerobic gut fungi into a phylogenetically distinct gut-dwelling fungal lineage. *Appl Environ Microbiol* 2019;85:e00988–00919.
68. Wang Y, Youssef N, Couger M, Hanafy R, Elshahed M, *et al*. Comparative genomics and divergence time estimation of the anaerobic fungi in herbivorous mammals. *msystems* 2019:00247-19.
69. Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, *et al*. The genome of the anaerobic fungus *Orpinomyces* sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader. *Appl Environ Microbiol* 2013;79:4620–4634.