

ORIGINAL ARTICLE

Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores

Audra S Liggenstoffer, Noha H Youssef, MB Couger and Mostafa S Elshahed
Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA

The phylogenetic diversity and community structure of members of the gut anaerobic fungi (AF) (phylum *Neocallimastigomycota*) were investigated in 30 different herbivore species that belong to 10 different mammalian and reptilian families using the internal transcribed spacer region-1 (ITS-1) ribosomal RNA (rRNA) region as a phylogenetic marker. A total of 267 287 sequences representing all known anaerobic fungal genera were obtained in this study. Sequences affiliated with the genus *Piromyces* were the most abundant, being encountered in 28 different samples, and representing 36% of the sequences obtained. On the other hand, sequences affiliated with the genera *Cyllamyces* and *Orpinomyces* were the least abundant, being encountered in 2, and 8 samples, and representing 0.7%, and 1.1% of the total sequences obtained, respectively. Further, 38.3% of the sequences obtained did not cluster with previously identified genera and formed eight phylogenetically distinct novel anaerobic fungal lineages. Some of these novel lineages were widely distributed (for example NG1 and NG3), whereas others were animal specific, being encountered in only one or two animals (for example NG4, NG6, NG7, and NG8). The impact of various physiological and environmental factors on the diversity and community structure of AF was examined. The results suggest that animal host phylogeny exerts the most significant role on shaping anaerobic fungal diversity and community composition. These results greatly expand the documented global phylogenetic diversity of members of this poorly studied group of fungi that has an important function in initiating plant fiber degradation during fermentative digestion in ruminant and non-ruminant herbivores.

The ISME Journal advance online publication, 22 April 2010; doi:10.1038/ismej.2010.49

Subject Category: microbial population and community ecology

Keywords: anaerobic fungi; neocallimastigomycota; culture-independent pyrosequencing surveys

Introduction

Although flagellated zoospores of anaerobic fungi (AF) were observed as early as 1910, definitive proof that AF are an important constituent of the cow rumen came relatively late (Orpin, 1975, 1977, 1994; Ho and Barr, 1995). The accidental discovery and proof that such flagellates were actually spores of a new fungal lineage rather than ciliated protozoa came when vegetative fungal growth was consistently observed while attempting to isolate anaerobic ciliated protozoa from sheep rumen (Orpin, 1975). AF are now classified in a single order (Neocallimastigales) within the recently erected

phylum Neocallimastigomycota (Hibbett *et al.*, 2007). Originally described in sheep, members of the AF have since been shown to exist in the rumen, hindgut, and feces of ruminant and non-ruminant herbivorous mammals, as well as reptilian herbivores (Bauchop, 1979; Teunissen and Op den Camp, 1994; Mackie *et al.*, 2004). Currently, only 6 genera and 20 species have been described (Griffith *et al.*, 2009), although multiple uncharacterized isolates have also been reported (Philips and Gordon, 1989; Ho and Barr, 1995).

The presence of AF in multiple (at least 50) ruminant and non-ruminant herbivorous mammals (Ljungdahl, 2008), as well as reptilian herbivores (Mackie *et al.*, 2004) has been well documented. However, the presence of AF in such habitats has mainly been assessed through isolation of a single or few AF strains (Ho and Barr, 1995; Ozkose *et al.*, 2001; Thareja *et al.*, 2006; Chen *et al.*, 2007) or through microscopic observation of the characteristic zoospores of AF in rumen content

Correspondence: MS Elshahed, Department of Microbiology and Molecular Genetics, Oklahoma State University, 1110 S. Innovation Way, Stillwater, OK 74074, USA.

E-mail: mostafa@okstate.edu

Received 21 January 2010; revised 10 March 2010; accepted 12 March 2010

(Mackie *et al.*, 2004). Collectively, these culture-based and microscopic studies have provided valuable insights on the prevalence and association of specific genera with certain animals. Recently, PCR primers that selectively amplify the internal transcribed spacer region-1 (ITS-1) within the ribosomal RNA (rRNA) of members of the Neocallimastigomycota has been described and validated (Brookman *et al.*, 2000; Edwards *et al.*, 2008). These primers have mainly been used either to identify AF isolates (Brookman *et al.*, 2000; Fliegerova *et al.*, 2004; Tuckwell *et al.*, 2005) or to identify AF community composition using various fingerprinting approaches, for example DGGE, T-RFLP, ARISA, and size-based selection (sephadex) (Edwards *et al.*, 2008; Nicholson *et al.*, 2009). To our knowledge, an examination of the phylogenetic diversity of AF community using a high throughput sequencing approach (either by cloning and sequencing a large number of clones or by pyrosequencing) has not yet been attempted, and only 236 ITS-1 AF sequences from pure cultures and environmental isolates are available in GenBank (as of October 2009). As such, little is currently known regarding the extent of global phylogenetic diversity within the AF, the presence and prevalence of novel yet-uncultured anaerobic fungal genera, the complexity of AF community within a single host, and the influence of various ecological and environmental factors on AF diversity and community composition within various hosts.

As part of a broader effort on exploring the utility of AF in direct fermentation schemes and biofuel production from lignocellulolytic biomass, we sought to explore the diversity of AF in multiple herbivores using a culture-independent sequencing approach. We present the results of a pyrosequencing-based survey of the Neocallimastigomycota from the fecal samples of a wide range of herbivores that belong to 10 different animal families. We document the presence of an extremely diverse AF community in different hosts, identify multiple novel genera, and present evidence that host phylogeny is an important factor in determining the AF diversity and community composition in different samples.

Materials and methods

Sampling

Fecal samples were obtained from domesticated animals from farms surrounding the cities of Stillwater and Cushing in Payne county, OK USA, from non-domesticated animals housed at the Oklahoma City Zoo (Oklahoma City, OK, USA), and from a reptile (Green Iguana) housed within the Learning Resource Center, Department of Zoology at Oklahoma State University (Stillwater, OK, USA) in November and December 2008. Fresh fecal samples were collected from animals in 50 ml sterile falcon

tubes immediately after deposition, stored on ice on-site, promptly transferred and stored in a -20°C freezer, usually within no more than 30 min of collection. Care was taken to avoid cross-contamination between different samples. A detailed description of the animals, locations, feed, and gut type is presented in Supplementary Table 1.

DNA extraction, PCR amplification, pyrosequencing, and sequence quality control

DNA was extracted from 0.5 g of fecal material from each sample using the FastDNA SPIN kit for soil (MP Biomedicals, Santa Ana, CA, USA). The extraction was conducted according to the manufacturer's instructions, with the exception that the lysis step was conducted for 30 s thrice, to allow for disruption of fungal tissues (zoospores and vegetative growth) (Cheng *et al.*, 2009). PCR was conducted using forward primers GCCTCCCTCGGCCATCAG-(barcode)-TCCTACCCTTTGTGAATTTG and reverse primer GCCTTGCCAGCCCGCTCAG-CTGCGTTCTT CATCGTTGCG. These primers are a modification of the previously described ITS-1 primers MN100 and MNGM2 (Edwards *et al.*, 2008; Nicholson *et al.*, 2009), with the universal forward pyrosequencing adaptor, and one of 12 barcode sequences (Multiplex Identifiers, 454 Life Sciences, Roche Diagnostics Corp., Branford, CT, USA) attached to the 5' end of the forward primer, and the universal reverse pyrosequencing adaptor added to the 3' end of reverse primers. The utilization of 12 different barcode decamers allows for sequencing of up to 12 different samples in a single plate quadrant, and a total of 48 different samples in a single pyrosequencing run. PCR amplification was conducted in 50 μl reaction mixtures containing: 2 μl of extracted DNA, 2.5 mM MgSO_4 , 0.2 mM dNTPs, 1.5 U of GoTaq Flexi DNA polymerase (Promega, Madison, WI, USA), and 10 μM of each of the forward and reverse primers. PCR amplification was carried out as follows: initial denaturation for 5 min at 95°C , followed by 40 cycles of denaturation at 95°C for 30 s, annealing at 48°C for 30 s, and elongation at 72°C for 1.5 min. PCR products from different samples with different barcodes were pooled and purified using an Invitrogen PureLink PCR Purification kit (Invitrogen Corporation, Carlsbad, CA, USA). In all, 454 FLX LR70 sequencing of pooled, purified, and barcoded PCR products was carried out at the University of South Carolina EnGenCore facility.

Sequences obtained were binned into different host animal groups using a perl script that identifies unique decamer barcodes (available upon request). Sequences with <130 bases and with quality scores of <25 were removed. In addition, sequences with ambiguous bases or homopolymers strings of >10 (the maximum length of homopolymer strings in Sanger sequenced ITS-1 anaerobic fungal sequences in GenBank database) were also removed from the data sets. The remaining sequences were

examined against all available ITS-1 rRNA sequences belonging to AF using BLAST search (Altschul *et al.*, 1997) conducted on a local server. Sequences with no similarity, or partial similarity (that is a segment of the amplicon has similarity to an ITS-1 sequence in the database, whereas the remainder of the amplicon does not have any similarity to ITS-1 sequence) were further removed from the data set.

Phylogenetic analysis

Operational taxonomic unit assignments. ITS rRNA regions within Bacteria, Archaea, and Fungi are known to be more variable than SSU regions, and hence the established putative species (3%) and genus (6%) sequence divergence values that cater to 16S rRNA gene-based diversity surveys of Bacteria and Archaea are unsuitable as universal thresholds for operational taxonomic assignments in ITS-1 diversity surveys (Nilsson *et al.*, 2008). Therefore, to group sequences obtained into operational taxonomic units (OTUs) representing relevant AF species and genera, we used the publicly available ITS-1 AF sequences to empirically determine a putative species and genus cutoff for the ITS-1 sequenced fragment in AF. Sequences derived from AF pure cultures were used to confirm phylogenetic affiliations or to assign uncultured clones or uncharacterized isolates to specific species and genera. In all, 83 sequences were assignable to known genera. These sequences were aligned using ClustalX (Thompson *et al.*, 1997), and a distance matrix was created using PAUP (Version 4.01b10; Sinauer associates, Sunderland, MA, USA). Sequence divergence values between all possible pairs belonging to the same genus were averaged to compute a species-level sequence divergence cutoff. Also, sequence divergence between all possible pairs belonging to different genera was averaged to compute a genus-level sequence divergence cutoff. Using this approach, a species cutoff of 4.80% and a genus cutoff of 16.95% were obtained. Species cutoff value of 0.05 was thus used for estimation of the number of putative species within each sample, and for computing the various diversity estimates and rankings described below. Genus-level cutoff of 0.17 was used in conjunction with phylogenetic analysis (see below) to identify novel genus-level diversity within data sets.

In addition to OTU identification in individual data sets, we identified OTUs shared between different data sets by constructing a single alignment for all sequences obtained in this study, followed by distance matrix generation and OTU assignment using mothur (Schloss *et al.*, 2009). Shared OTU information gained through analysis of the entire data set (267 287 sequences) were used for various comparative diversity approaches between the different data sets. The Petascale Data Analysis Facility, a data-intensive computing cluster part of the Triton Resource located at the San Diego

Supercomputing Center, University of California San Diego (<http://tritonresource.sdsc.edu/pdaf.php>), was used for generating all mothur outputs for the entire data set.

Phylogenetic placement

Classification and identification of AF lineages have mainly been based on pattern of thallus/rhizoid morphology (monocentric or polycentric), and zoospore flagellation (uniflagellate or polyflagellate) (Ho and Barr, 1995; Ozkose *et al.*, 2001; Nicholson *et al.*, 2009). The use of molecular phylogenetic approaches in AF taxonomy has recently been examined (Brookman *et al.*, 2000), and ITS-1-based phylogeny have shown that while multiple AF genera are monophyletic (*Cyllumyces*, *Anaeromyces*, and *Orpinomyces*), members of the genus *Piromyces* seem to be polyphyletic (Brookman *et al.*, 2000), and members of the genus *Caecomyces* cluster as a distinct subgroup within the genus *Neocallimastix* (Fliegerova *et al.*, 2004). Despite such discrepancies between microscopic-based and phylogenetic-based classification of AF, no revisions (for example species reassignment, proposition of new genera) based on molecular taxonomic data have been proposed, and microscopic-based taxonomical schemes are still currently in use.

To determine the phylogenetic affiliation of OTU_{0.05} obtained, representative sequences were searched against all ITS-1 sequences available in public databases. Sequences with high (>94%) sequence similarity to multiple isolates belonging to a single genus were assigned to that genus. On the other hand, OTUs with lower sequence similarity or similarity to multiple sequences from different genera were further probed by examining their phylogenetic position relative to other AF ITS-1 sequences in a PAUP-generated tree using various distance-based and character-based phylogenetic placements. OTUs with >17% sequence divergence that formed distinct phylogenetic lineages with high bootstrap support were judged to constitute a novel AF lineage at the genus level. The effect of filtering hypervariable regions on tree topologies was analyzed using GBLOCKS (Talavera and Castresana, 2007) under multiple stringency conditions. The program JModelTest (Posada, 2008) was used to determine the optimum nucleotide substitution model to be used in constructing phylogenetic trees.

Diversity estimates, rankings, and evaluation of various factors affecting AF diversity and community structure within individual data sets

Basic diversity estimates, as well as rarefaction curves were computed on OTU_{0.05} outputs using mothur (Schloss *et al.*, 2009). Good's coverage was computed for each sample as described earlier (Good, 1953). Three different approaches were used to rank all data sets obtained according to

diversity: number of genera per sample, rarefaction curve analysis, and diversity rankings approaches. Diversity ranking-based approaches have been widely used in macro ecology (Liu *et al.*, 2007), and only recently introduced to microbial ecology (Youssef and Elshahed, 2009). We used both an information-related diversity ordering method (Renyi generalized entropy), and an expected number of species-related diversity ordering method (Hulbert family of diversity indices) to reach a consensus ranking of fungal diversity for all the 33 animals studied (Liu *et al.*, 2007; Youssef and Elshahed, 2009).

To identify the factors that most affect fungal diversity, we examined the correlation between AF diversity and various multiple measurable factors (Supplementary Table 1) that might influence the AF communities diversity estimates calculated for the 33 data sets. As these factors are nominal, χ^2 Contingency tables was the method of choice for correlation (Chernoff and Lehmann, 1954; Plackett, 1983). However, the dependent variable (ordinal in cases of diversity rankings, and rarefaction curve rankings, and quantitative in case of identified number of fungal genera) had to be converted to nominal variables first. Ordinal variables were grouped into: low–medium–high diversity categories such that; ranks 1–11 were classified as low diversity, ranks 12–22 were classified as medium diversity, and ranks 23–33 were classified as high diversity. As for the quantitative variables, we first ranked these from the least to the most (1–33) then the ordinal ranks were converted to nominal variables as discussed above. With the two variables (dependent and independent) being nominal, χ^2 Contingency correlation was carried out. To measure the degree of association between the two variables, the obtained χ^2 value was used to calculate Cramer's V statistics; $V = \sqrt{\frac{\chi^2}{n \times \min(M-1, N-1)}}$, where χ^2 is the calculated χ^2 value, n is the number of species (33), M is the number of rows (or dependent variables), and N is the number of columns (or independent variables).

β -Diversity estimates

Network-based analysis (Shannon *et al.*, 2003) and non-metric multidimensional scaling plots were used to visualize differences in community structure between various AF data sets (β -diversity). Network graphs were created with Cytoscape 2.6.3 using a spring-embedded algorithm allowing for visualization of species-level OTUs within and between animal hosts (Shannon *et al.*, 2003; Ley *et al.*, 2008). An example of the input file used is shown in Supplementary Table 2. Cytoscape depicts data sets as nodes (animals and OTUs) connected by lines that denote the presence of a specific OTU within or between animal hosts. Animal hosts with more similar AF communities, and therefore, more

OTUs that are shared between them, appear spatially closer on the graph. Animal hosts are depicted as circular nodes, whereas, OTU nodes are represented as squares. Generally, data sets with more shared OTUs are pulled toward each other and toward the center of the graph, whereas, data sets with fewer shared OTUs and/or a higher proportion of unique OTUs remain on the periphery. Non-metric multidimensional scaling plots were generated using Bray–Curtis similarity index matrices (Bray and Curtis, 1957) between the 33 different animals studied. Bray–Curtis similarity indices were calculated in mothur program (Schloss *et al.*, 2009), and the function metaMDS in the Vegan library of R statistical program (<http://www.r-project.org/>)

Nucleotide sequences accession numbers

Sequences generated were deposited under accession numbers GQ576478–GQ843764.

Results

A total of 350 363 sequences were obtained from 33 different samples. In all, 76.3% of the total sequences generated were kept after implementing quality control measures, yielding 267 287 sequences that were used for further phylogenetic analyses. The range of sequence lengths of amplicons included in the analysis was 130–304 bp (average 236). A histogram of sequence read length is provided in Supplementary Figure 1. The average number of sequences per animal sampled was 8100. Coverage estimates (Table 1), as well as rarefaction curve analysis (Supplementary Figure 2), indicate that the sequencing effort was successful in capturing the majority of AF taxa in all samples.

Genus-level taxonomic placement

Monocentric genera. Sequences affiliated with the genus *Piromyces* were the most abundant in the entire data set, being encountered in 28 different samples, and representing 36% of the total number of sequences obtained. Although it is currently assumed that *Piromyces* spp. represent the most abundant sequences in hindgut fermenters (Orpin, 1994), the distribution of *Piromyces* varied greatly within hindgut fermenters depending on the host animal family. Within the family Equidae, *Piromyces*-affiliated sequences were identified in low numbers and were even absent in some horse and Grevy's Zebra replicates (Table 1). On the other hand, within the hindgut fermenter Black Rhinoceros (family Rhinocerotidae), *Piromyces*-affiliated sequences constituted 100% of the AF community. *Piromyces*-affiliated sequences were also encountered in all but two of the ruminants sampled (Rothschild's Giraffe and Greater Kudu).

A total of 22 950 (8.6%) sequences affiliated with the genus *Neocallimastix* were encountered in this

Table 1 Composition of anaerobic fungal communities in sampled herbivores

Common name	Rumen type	Family	No. of seqs	OTU _{0.05}	Monocentric			Polycentric			Novel								Coverage
					Neocallimastix	Piromyces	Caecomyces	Anaeromyces	Orpino-	Cyella-	NG1	NG2	NG3	NG4	NG5	NG6	NG7	NG8	
Horse individual 1	H	Equidae	12772	41	0.3	0.2	56.7	42.8									99.9		
Horse individual 2	H	Equidae	8305	33	0.3	12.3	68.3										99.9		
Horse individual 3	H	Equidae	3650	22	1.9	3.5	1.1	1.3						0.06			99.8		
Miniature donkey	H	Equidae	3827	15	0.03	0.4	4.7										99.9		
Somali wild ass	H	Equidae	1609	10	44.9	0.1	15.0								0.6		99.8		
Grants zebra	H	Equidae	7591	26		0.01	99.9										99.9		
Grevy's zebra individual 1	H	Equidae	14190	31	0.02		99.9										99.9		
Grevy's zebra individual 2	H	Equidae	8789	27			99.9	0.2									99.9		
Black rhinoceros	H	Rhinocerotidae	49215	49	0.002		99.9										99.9		
White-fronted wallaby	F (N)	Macropodidae	13346	53	49.5	16.9	9.0	0.01	18.8	5.8							99.9		
Red kangaroo	F (N)	Macropodidae	5782	28	30.8	12.4	17.8		25.3							13.7	99.9		
Pygmy hippopotamus	F (P)	Hippopotamidae	7642	48	39.3	38.6	0.03		11.7	10.3							99.9		
Llama	F (P)	Camelidae	11575	58	5.3	63.0	2.4		14.9	0.01							99.9		
Rothschild's giraffe	F (R)	Giraffidae	6583	29	14.4	27.1	0.4		34.4	4.0							99.9		
Okapi	F (R)	Giraffidae	2046	16	8.6	8.6	39.8		15.5	15.2							99.8		
Indo-Chinese sika deer	F (R)	Cervidae	5680	31	52.7	18.3	0.1			13.4							99.9		
Indian hog deer	F (R)	Cervidae	5727	27	12.2	20.3	13.9		39.9	7.3							99.9		
American elk	F (R)	Cervidae	48	4	72.9	6.3	2.1		2.1								97.9		
Pere David's deer	F (R)	Cervidae	4212	29	33.6	9.7	49.5		6.8								99.8		
Western tufted deer	F (R)	Cervidae	3172	20	0.03	0.03	10.1		2.6	27.3							99.8		
Pronghorn	F (R)	Antilocapridae	12950	24	96.7	1.2	0.01		0.01	2.0							99.9		
Bontebok	F (R)	Bovidae	12431	53	47.3	0.01	0.02										99.9		
Grant's gazelle	F (R)	Bovidae	4144	19	21.1	6.7	29.6		49.3								99.9		
Southern reem	F (R)	Bovidae	4215	23	88.6	1.0	0.02		0.02								99.7		
American bison	F (R)	Bovidae	9180	55	37.5	9.5	9.2		3.0								99.9		
Greater kudu	F (R)	Bovidae	4966	31	8.4	31.2	1.3		0.02	99.1							99.9		
Goral	F (R)	Bovidae	3274	23	0.4	0.9	32.1		0.03								99.8		
Sable antelope	F (R)	Bovidae	11395	46	22.1	7.6	8.4		47.5								99.9		
Nile lechwe	F (R)	Bovidae	8768	37	7.4	18.2	0.7		5.2	11.1							99.9		
Domestic cattle	F (R)	Bovidae	5448	60	10.2	17.5	8.5		33.7	5.3							99.9		
Domestic sheep	F (R)	Bovidae	8554	37	5.3	61.0	1.9		2.2	0.09							99.7		
Domestic goat	F (R)	Bovidae	5291	22	44.0	22.1	3.6		8.0	0.02							99.8		
Green iguana	n/a	Iguanidae	910	10	0.11	6.2	1.1		34.8	8.0							99.8		

Abbreviations: F (N), non-ruminant foregut fermenters; F (P), pseudo-ruminant foregut fermenters; F (R), ruminant foregut fermenters; H, Hindgut fermenters; OTU, operational taxonomic unit.

study, and *Neocallimastix*-affiliated sequences were identified in 18 different data sets belonging to seven different animal families (Table 1). This reinforces the notion that *Neocallimastix* spp. are prevalent in foregut fermenters. However, the results also show that the members of the genus *Neocallimastix* are more widely distributed than previously implied (Orpin, 1994), as they also seem to constitute a minor component of AF community identified in multiple hindgut fermenters (0.72% of the sequences from hindgut fermenters). Sequences affiliated with the genus *Caecomyces* were present in both foregut and hindgut fermenters, but were encountered in fewer data sets (14) than *Neocallimastix* and *Piromyces*-affiliated sequences. *Caecomyces* was the most abundant genus only in two data sets (Llama and domestic cattle).

Polycentric genera. Although *Anaeromyces* spp. are generally assumed to be present mainly in foregut ruminants (cattle and water buffalo) as well as non-described species from sheep and goat (Orpin, 1994; Ho and Barr, 1995), they were widely distributed in our data set and were encountered in 26 different samples. However, *Anaeromyces*-affiliated sequences typically represented a minor component (average 11% in samples where they were detected) of a specific population, rarely exceeding 30% and never exceeding 50% within any data set studied (Table 1).

Orpinomyces-affiliated sequences were identified in only eight animal species (llama, giraffe, and six ruminants). In general, *Orpinomyces*-affiliated sequences were present in very low abundance (average of only 3% community composition), and made only 1.1% of the total sequences in this study. *Orpinomyces*-affiliated sequences were not identified in any of the hindgut samples analyzed.

Cyllamyces is the most recently described genus of AF and has so far been isolated only from domestic Cattle (Ozkose *et al.*, 2001). This study suggests that members of the genus *Cyllamyces* are the least widely distributed in nature, being detected only in two data sets (American bison and Sable antelope). *Cyllamyces*-affiliated sequences comprised <10% of the sequences within each of these two data sets, and made up only 0.7% of all the sequences generated in this study. Interestingly, we did not detect any *Cyllamyces*-affiliated sequences in cattle, although this is where it was originally identified (Ozkose *et al.*, 2001), implying that other factors (for example feed type, location) could have an important function in establishing *Cyllamyces* populations in herbivores.

Novel AF groups. In addition to the members of previously described genera, a significant fraction (38.3% of total sequences) could not be assigned to any of these six genera. Phylogenetic analysis suggested that these groups belonged to eight different novel lineages that were designated novel

groups NG1–NG8 (Figure 1). These lineages remained monophyletic regardless of the tree-building algorithm used (Parsimony, Maximum likelihood, distance) or the exclusion of hypervariable base pairs from the analysis using different stringency options in GBlocks (Supplementary Figure 3). Some of these groups, for example NG1 and NG3 were present in high abundance in multiple hindgut and foregut samples. NG1 and NG3 were the second and third most abundant lineages (with 19.8% and 12.0% of the total number of sequences, respectively). These two groups, either individually or together, constituted the majority of sequences in all hindgut Equidae samples and were also co-identified in multiple foregut fermenters. NG2 and NG5 were present in multiple animals (8 and 14, respectively), but typically were present in low abundance in data sets where they were encountered.

Other groups had an extremely limited distribution and abundance. NG4 was a minor constituent within American bison and Sable antelope. NG6 comprised all of the sequences within the Greater Kudu data set, and constituted 15% and 34% of AF sequences in Okapi and Rothschild's giraffe, the two animals belonging to the family Giraffidae in our data set. Finally, NG7 and NG8 were each found in only a single animal, Somali wild ass and Red kangaroo, respectively.

Diversity estimates and factors influencing AF community diversity

Diversity estimates for various data sets were elucidated and compared. Diversity estimates used were the number of genera encountered in each data sets, rarefaction curve-based ranking, and diversity ordering-based approaches. The results (Supplementary Table 3) were used as a starting point for diversity correlation using χ^2 methods. Although gut type, ruminant ability, and feed showed low correlation ($r=0.20$ – 0.37) with all three diversity ranking schemes, a higher correlation ($r=0.56$ – 0.63) was observed when correlating animal family to various diversity schemes (Table 2).

Community relatedness and factors influencing community composition

A network graph based on OTUs that were shared between various data set was constructed using Cytoscape 2.6.3 (Shannon *et al.*, 2003), and the graphs were color-coded based on different factors potentially affecting AF community relatedness. The results (Figures 2a–c) indicate that, similar to diversity studies, feed type seems to be the least relevant factor in shaping community structure, as evident by the scattered color distribution in Figure 2a. Gut types (hindgut, foregut-non-ruminant, foregut-pseudoruminant, foregut-ruminant, and Iguana), Figure 2b, provided slightly better explanation of community relatedness, but members of the same gut type belonging to different animal host families had clearly

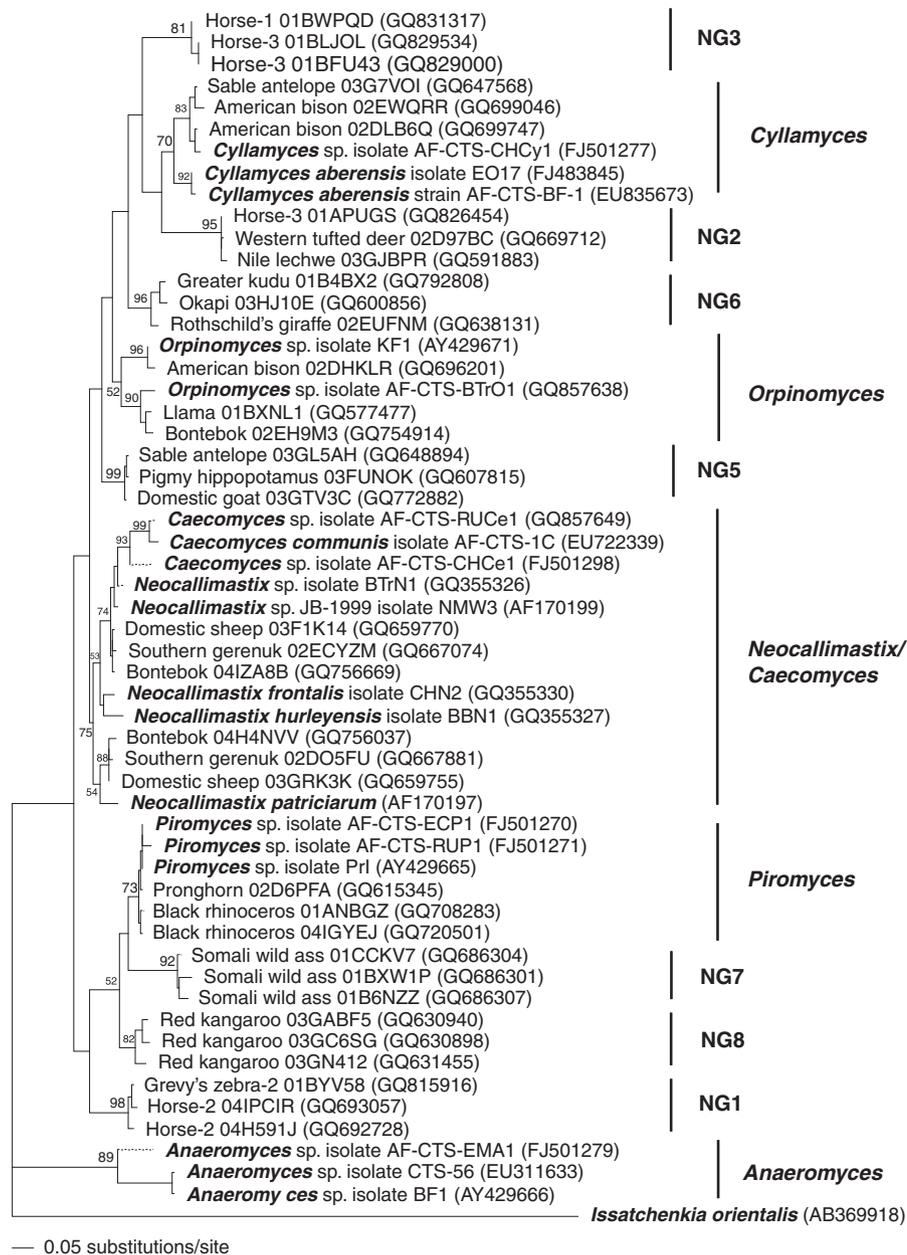


Figure 1 Distance dendrogram highlighting the phylogenetic affiliation of anaerobic gut fungi sequences encountered in this study. Sequences used in tree construction include reference sequences of anaerobic fungal isolates, representative OTUs affiliated with known anaerobic fungal genera encountered in this study, and representatives of novel anaerobic fungal lineages. The tree was constructed using neighbor-joining algorithm with the Hasegawa–Kishino–Yano (HKY) substitution model and a γ -shaped distribution of 0.6190. Bootstrap values are based on 1000 replicates and are shown for branches with > 50% bootstrap support. The corresponding ITS-1 region of the ascomycetous yeast *Issatchenikia orientalis* was used as an outgroup. Genbank accession numbers of reference sequences are given in parentheses.

Table 2 Correlation coefficients of various diversity measures

Factor	Correlation		
	Diversity ordering	Rarefaction	No. of genera
Family	0.63	0.60	0.56
Gut type	0.37	0.30	0.21
Ruminance	0.28	0.37	0.32
Feed type	0.20	0.29	0.31

different community structures. For example, although both are hindgut fermenters, members of the Equidae seem to have little shared OTUs with Black Rhinoceros. Similarly, the two foregut pseudo-ruminants belonging to different families had very different community structures, whereas both foregut non-ruminants, both belonging to the family Macro-podidae have more similar community structures.

Compared with feed type and gut type, animal host phylogeny seems to provide better explanation

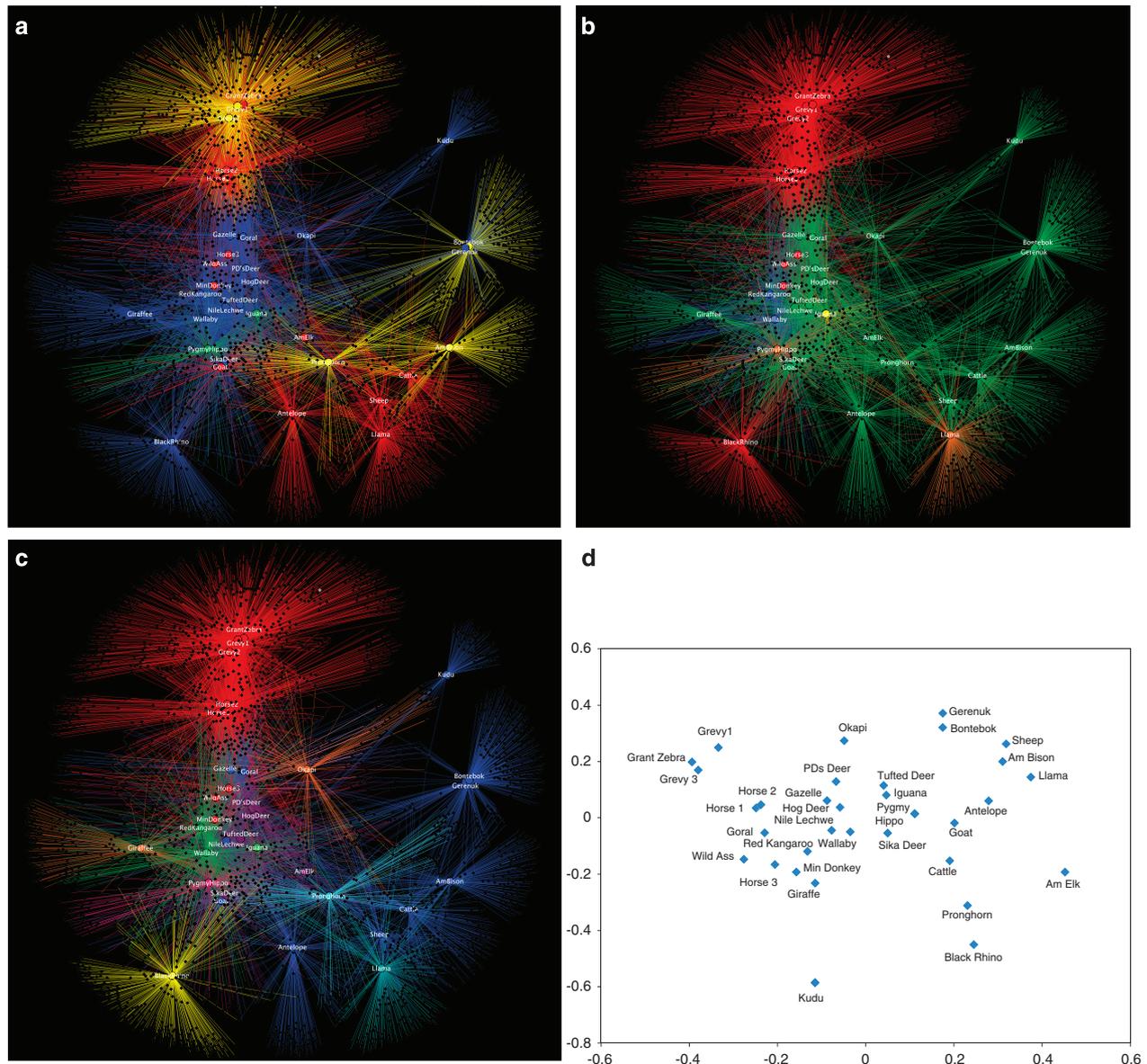


Figure 2 (a–c) Network graph highlighting shared OTUs between different anaerobic fungal communities in different animal hosts. The same graph is coded with three different criteria to ease comparison. (a) Feed type. (b) Gut type. (c) Animal host phylogeny (family). Circular nodes indicate animal data sets, whereas smaller square, grey nodes represent individual OTUs. Data sets with a higher proportion of shared OTUs are pulled to the middle, whereas data sets with a high proportion of unique OTUs remain on the periphery. The distance between any two data sets is a function of the number of shared OTUs between the two. (d) Non-metric multidimensional scaling plot of AF data sets obtained in this study. A colored, higher resolution version of this figure is available at the ISME journal website in html format.

of community relatedness of AF. Members of the family Equidae clustered at the top of the graph (Figure 2c), with replicates of the same animal having highly shared AF community. Both zebras and two of the horse replicates (individuals 1 and 2) had a peripheral position at the top of the graph because such samples, mainly composed of NG1 and NG3, had a very low percentage of shared OTUs with non-Equidae samples (7.82%). On the other hand, Somali wild ass and Miniature donkey had more shared OTUs with non-Equidae samples (22.43%) and on average had fewer unique OTUs (10.28% vs 30.06% in zebra and two

horse replicates), and are therefore more centrally located than the other samples from the Equidae family.

Similar to the Equidae, members of the Cervidae clustered together (Figure 2c), as well as the two samples belonging to the family Macropodidae. However, although both families have a high percentage of shared OTUs (86.78%), many of these OTUs are not family specific (that is encountered only in these families), and have been encountered in other samples. Therefore, the members of both the Cervidae and the Macropodidae clustered toward the center of the graph.

Within the large number of samples belonging to the family Bovidae, multiple trends were observed. Some of these samples had a high proportion of shared OTUs with other members of the Bovidae and non-Bovidae (for example Nile lechwe, domestic goat, Grant's gazelle, Goral) and as such are centrally located. Others had a high proportion of OTUs that seem to be only shared within the certain members of the Bovidae. As such, these animals are collectively located in the periphery of the graph in close proximity to each other (for example American bison, domestic cattle, and sheep). Southern gerenuk had a strikingly similar community to bontebok, sharing 67.77% of its OTUs. Finally, Greater Kudu had a unique peripheral location in the graph, as its community was mainly composed of OTUs belonging to a lineage of limited distribution (NG6), and had a low proportion of shared OTUs with other animals (only with Okapi and Rothschild's giraffe).

Members of the family Giraffidae analyzed in this study (Rothschild's Giraffe and Okapi) did not have any shared OTUs and thus, are not located in proximity to one another. This represents a deviation from the observed importance of animal host phylogeny on community structure. Although Green iguana, represented the only non-mammalian, cold-blooded animal included in the study, and although it had a unique diet, the AF community in iguana had the lowest proportion of unique OTUs and was centrally located in the network graphs.

In addition to network analysis, a non-metric multidimensional scaling plot was generated to visualize the similarities in AF community structure between the various animal hosts. This non-metric multidimensional scaling plot, Figure 2d, shows a striking similarity to the network graph plots and further reinforces the importance of animal host phylogeny in shaping AF community.

Discussion

In this study, we present a detailed survey of phylogenetic diversity, community structure, and comparative diversity of members of the anaerobic gut fungi using rRNA ITS-1 as a phylogenetic marker. To our knowledge, this represents the first wide scale culture-independent sequence analysis of members of the phylum Neocallimastigomycota. In addition, the work represents the first culture-independent survey of AF community in a reptilian host (Green Iguana), and in multiple mammalian species (for example American elk, pronghorn, bontebok, southern gerenuk, Goral, and Nile lechwe).

The high level of AF phylogenetic diversity observed in animal hosts surveyed is evident by the fact the average number of species per sample (31) is higher than the total number of AF species currently described (20). We acknowledge that our estimates are solely based on sequence divergence values of a single amplicon, rather than a thorough

microscopic, biochemical, and sequence analysis. However, this chosen cutoff value (5%) was based on averaging ITS-1 sequence divergence values of known AF isolates. Therefore, although not definitive, this cutoff reflects a reasonable estimate of the number of AF species per sample. Another indicator of the highly diverse nature of AF is the identification of multiple novel AF lineages that represented 38.3% of the total sequences obtained. The presence of novel lineages has previously been speculated (Orpin, 1994), and unclassifiable patterns in finger printing approaches suggestive of novel genera have subsequently been observed in cow rumen (Nicholson *et al.*, 2009). Multiple plausible reasons could account for the inability to previously identify and isolate these novel AF lineages. It is entirely possible that AF affiliated with many of these lineages have thallus and zoospore structures similar to those of well-described AF genera and thus isolates belonging to such lineages would have been classified as members of an already existing AF genus upon isolation. Alternatively, thallus and zoospore morphologies of these novel lineages could possess unique microscopic characteristics that have hence escaped microscopic detection. Finally, regardless of zoospore/thallus morphology, members of such lineages might require unique, yet-unidentified growth media factors or selective substrates for enrichment and growth under laboratory conditions, and are hence unculturable using standard methodologies used for isolating AF (Theodorou *et al.*, 2005). It is interesting to note that the choice of substrate indeed seems to have an important influence on the morphology of isolates obtained (Ho and Barr, 1995; Griffith *et al.*, 2009).

Although this study sheds light on the diversity and distribution of AF, we caution against considering the described patterns of AF diversity a definitive description of global AF communities in herbivore hosts. Rather, this study represents a community snapshot of multiple animals from few locations within a single state in a single country. We reason that only a well-controlled experiment tracking AF community structure in replicates of a single animal species at different age groups, feed regimens, and geographical locations would provide an accurate description of the community dynamics of AF fungi, and factors influencing the community structure within various animal species.

AF are highly fibrolytic microorganisms, producing a wide array of cell-bound and cell-free cellulolytic, hemicellulolytic, glycolytic, and proteolytic enzymes (Lowe *et al.*, 1987; Williams and Orpin, 1987a,b; Raghothama *et al.*, 2001). The anaerobic nature (which deters many mycologists), and eukaryotic affiliation (which deterred anaerobic microbiologists) have limited the number of active research groups investigating these microorganisms to a dedicated but small group of scientists. Bauchop (1989) concluded a review on the biology of AF by asserting that 'The AF also attract attention as a new

group of cellulase- and hemicellulase-producing microorganisms. The challenge of adapting this group of microorganisms in biotechnology will undoubtedly be accepted by scientists in the near future.' With few exceptions (Li *et al.*, 1997a,b; Chen *et al.*, 2006; Ljungdahl, 2008), we believe that this challenge has not sufficiently been met.

Acknowledgements

We thank Sue Haley and Dr Jennifer D'Agostino for help with sampling. This work was supported by grants from the Oklahoma Bioenergy Center, and NSF EPSCoR award EPS 0814361.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Bauchop T. (1979). Rumen anaerobic fungi of cattle and sheep. *Appl Environ Microbiol* **38**: 148–158.
- Bauchop T. (1989). Biology of gut anaerobic fungi. *BioSystems* **23**: 53–64.
- Bray JR, Curtis JT. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol Monog* **27**: 325–349.
- Brookman JL, Mennim G, Trinci APJ, Theodorou MK, Tuckwell DS. (2000). Identification of characterization of anaerobic gut fungi using molecular methodologies based on ribosomal ITS1 and 18S rRNA. *Microbiology* **146**: 393–403.
- Chen H, Hopper SL, Li XL, Ljungdahl LG, Cerniglia CE. (2006). Isolation of extremely AT-rich genomic DNA and analysis of genes encoding carbohydrate-degrading enzymes from *Orpinomyces* sp. strain PC-2. *Curr Microbiol* **53**: 396–400.
- Chen Y-C, Tsai S-D, Cheng H-L, Chien C-Y, Hu C-Y, Cheng T-Y. (2007). *Caecomyces sympodialis* sp. nov., a new rumen fungus isolated from *Bos indicus*. *Mycologia* **99**: 125–130.
- Cheng YF, Edwards JE, Allison GG, Zhu W-Y, Theodorou MK. (2009). Diversity and activity of enriched ruminal cultures of anaerobic fungi and methanogens grown together on lignocellulose in consecutive batch culture. *Bioresour Technol* **100**: 4821–4828.
- Chernoff H, Lehmann EL. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness-of-fit. *Ann Math Stat* **25**: 579–586.
- Edwards JE, Kingston-Smith AH, Jimenez HR, Huws SA, Skot KP, Griffith GW *et al.* (2008). Dynamics of initial colonization of nonconserved perennial ryegrass by anaerobic fungi in the bovine rumen. *FEMS Microbiol Ecol* **66**: 537–545.
- Fliegerova K, Hodrova B, Voigt K. (2004). Classical and molecular approaches as a powerful tool for the characterization of rumen polycentric fungi. *Folia Microbiol* **49**: 157–164.
- Good IJ. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**: 237–264.
- Griffith GW, Ozkose E, Theodorou MK, Davies DR. (2009). Diversity of anaerobic fungal populations in cattle revealed by selective enrichment culture using different carbon sources. *Fungal Ecol* **2**: 87–97.
- Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PE, Eriksson OE *et al.* (2007). A higher-level phylogenetic classification of the Fungi. *Mycol Res* **111**: 509–547.
- Ho YW, Barr DJS. (1995). Classification of anaerobic gut fungi from herbivores with emphasis on rumen fungi from Malaysia. *Mycologia* **87**: 655–677.
- Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS *et al.* (2008). Evolution of mammals and their gut microbes. *Science* **320**: 1647–1651.
- Li X-L, Chen H, Ljungdahl LG. (1997a). Two cellulases, CelA and CelC, from the polycentric anaerobic fungus *Orpinomyces* strain PC-2 contain N-terminal docking domains for a cellulase-hemicellulase complex. *Appl Environ Microbiol* **63**: 4721–4728.
- Li XL, Chen H, Ljungdahl LG. (1997b). Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. *Appl Environ Microbiol* **63**: 628–635.
- Liu C, Whittaker RJ, Ma K, Malcolm JR. (2007). Unifying and distinguishing diversity ordering methods for comparing communities. *Popul Ecol* **49**: 89–100.
- Ljungdahl LG. (2008). The cellulase/hemicellulase system of the anaerobic fungus *orpinomyces* PC-2 and aspects of its use. *Ann NY Acad Sci* **1125**: 308–321.
- Lowe SE, Theodorou MK, Trinci APJ. (1987). Cellulases and xylanase of an anaerobic rumen fungus grown on wheat straw, wheat straw holocellulose, cellulose, and xylan. *Appl Environ Microbiol* **53**: 1216–1223.
- Mackie R, Rycyk M, Reummler R, Aminov R, Wikelski M. (2004). Biochemical and microbiological evidence for fermentative digestion in free-living land iguanas and marine iguanas on the Galapagos archipelago. *Physiol Biochem Zool* **77**: 127–138.
- Nicholson MJ, McSweeney CS, Mackie RI, Brookman JL, Theodorou MK. (2009). Diversity of anaerobic gut fungal populations analysed using ribosomal ITS1 sequences in faeces of wild and domesticated herbivores. *Anaerobe*. In Press: doi:10.1016/j.anaerobe.2009.1005.1003.
- Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson K-H. (2008). Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evol Bioinform Online* **4**: 193–201.
- Orpin CG. (1975). Studies on the rumen flagellate *Neocallimastix frontalis*. *J Gen Microbiol* **91**: 249–262.
- Orpin CG. (1977). The occurrence of chitin in the cell walls of the rumen organism *Neocallimastix frontalis*, *Piromonas communis*, and *Sphaeromonas communis*. *J Gen Microbiol* **99**: 215–218.
- Orpin CG. (1994). Anaerobic fungi: taxonomy, biology, and distribution in nature. In: Mountfort DO, Orpin CG (eds). *Anaerobic Fungi: Biology, Ecology, and Function*. Marcel Dekker Inc.: New York, USA, pp 1–45.
- Ozkose E, Thomas BJ, Davies DR, Griffith GW, Theodorou MK. (2001). *Cyllamyces aberensis* gen. nov. sp. nov., a new anaerobic gut fungus with branched sporangio-phores isolated from cattle. *Can J Bot* **79**: 666–673.
- Philips MW, Gordon GLR. (1989). Growth characteristics on cellobiose of three different anaerobic fungi isolated from the ovine rumen. *Appl Environ Microbiol* **55**: 1695–1702.

- Plackett RL. (1983). Karl Pearson and the Chi-squared test. *Int Stat Rev* **51**: 59–72.
- Posada D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**: 1253–1256.
- Raghothama S, Eberhardt RY, Simpson P, Wigelsworth D, White P, Hazlewood GP *et al.* (2001). Characterization of a cellulose dockerin domain from the anaerobic fungus *Piromyces equi*. *Nat Struct Biol* **8**: 775–778.
- Schloss PD, Westcott SL, Thomas R, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504.
- Talavera G, Castresana J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**: 564–577.
- Teunissen MJ, Op den Camp HJ. (1994). Anaerobic fungi and their cellulolytic and xylanolytic enzymes. *Antonie van Leeuwenhoek* **63**: 63–76.
- Thareja A, Puniya AK, Goel G, Nagpal R, Sehgal J, Singh PK *et al.* (2006). *In vitro* degradation of wheat straw by anaerobic fungi from small ruminants. *Arch Anim Nutr* **60**: 412–417.
- Theodorou MK, Brookman J, Trinci A. (2005). Anaerobic fungi. In: Makkar HP, McSweeney CS (eds). *Methods in Gut Microbial Ecology for Ruminants*. Springer: Dordrech, The Netherlands.
- Thompson JD, Gibson TJ, Pleiomiak F, Jeanmougin F, Higgins DG. (1997). The ClustalX interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. *Nucleic Acids Res* **25**: 4876–4882.
- Tuckwell DS, Nicholson MJ, McSweeney CS, Theodorou MK, Brookman JL. (2005). The rapid assignment of ruminal fungi to presumptive genera using ITS1 and ITS2 RNA secondary structures to produce group-specific fingerprints. *Microbiology* **151**: 1557–1567.
- Williams AG, Orpin CG. (1987a). Polysaccharide-degrading enzymes formed by three species of anaerobic rumen fungi grown on a range of carbohydrate substrates. *Can J Microbiol* **33**: 418–426.
- Williams AG, Orpin CG. (1987b). Glycoside hydrolase enzymes present in the zoospores and vegetative growth stages of the rumen fungi *Neocallimastix patriciarum*, *Piromonas communis*, and an unidentified isolate grown on a range of carbohydrates. *Can J Microbiol* **33**: 427–434.
- Youssef NH, Elshahed MS. (2009). Diversity rankings among bacterial lineages in soil. *ISME J* **3**: 305–313.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)