

The rapid assignment of ruminal fungi to presumptive genera using ITS1 and ITS2 RNA secondary structures to produce group-specific fingerprints

Danny S. Tuckwell,¹ Matthew J. Nicholson,^{2,3†} Christopher S. McSweeney,⁴ Michael K. Theodorou² and Jayne L. Brookman²

Correspondence

Michael K. Theodorou
mike.theodorou@bbsrc.ac.uk

¹F2G Ltd, Lankro Way, Eccles, Manchester M30 0BH, UK

²Institute of Grassland and Environmental Research, Plas Gogerddan, Aberystwyth, Ceredigion, SY23 3EB, UK

³School of Biological Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, UK

⁴CSIRO Livestock Industries, Queensland Bioscience Precinct, Carmody Road, St Lucia, Brisbane, Australia

Received 8 October 2004
Revised 24 January 2005
Accepted 26 January 2005

Identification of microbial community members in complex environmental samples is time consuming and repetitive. Here, ribosomal sequences and hidden Markov models are used in a novel approach to rapidly assign fungi to their presumptive genera. The ITS1 and ITS2 fragments from a range of axenic, anaerobic gut fungal cultures, including several type strains, were isolated and the RNA secondary structures predicted for these sequences were used to generate a fingerprinting program. The methodology was then tested and the algorithms improved using a collection of environmentally derived sequences, providing a rapid indicator of the fungal diversity and numbers of novel sequence groups within the environmental sample from which they were derived. While the methodology was developed to assist in investigations involving the rumen ecosystem, it has potential generic application in studying diversity and population dynamics in other microbial ecosystems.

INTRODUCTION

The anaerobic fungi are an unusual group of zoosporic fungi occupying a unique niche in the digestive tract of wild and domesticated ruminants and large monogastric herbivores (Theodorou *et al.*, 1996). Since their discovery in sheep in 1975 (Orpin, 1975), rumen microbiologists have been trying to establish the role played by these fungi in the digestive tract ecosystem. According to some researchers,

ruminal fungi play a pivotal role as the initial (primary) colonizers of plant fibre in the rumen (e.g. Bauchop, 1979; Joblin *et al.*, 2002; Lee *et al.*, 2000). Others consider that their low population density in rumen fluid is sufficient for them to make no more than a negligible contribution to rumen function (Akin & Benner, 1988; Cheng *et al.*, 1991). These contradictory points of view stem from technical difficulties in quantifying population density and diversity of ruminal fungi in colonizing digesta samples using classical culture methodologies because of the interference caused by exceedingly high numbers of ruminal bacteria in rumen fluid.

†Present address: AgResearch Grasslands, Private Bag 11008, Palmerston North, New Zealand.

Abbreviations: HMM, hidden Markov model; IGER, Institute of Grassland and Environmental Research; ITS, internal transcribed spacer.

A representative alignment of 12 ITS1 sequences plus flanking regions from anaerobic gut fungi showing Variable Regions I–IV is available as Supplementary Fig. S1. An alignment of the five *Caecomyces* and nine *Cyllamyces* sequences for ITS1 Variable Regions I–IV, together with the ITS1 Variable Regions II–IV sequences from the eleven of the forty-eight environmental samples that group together as Novel Group 1, is available as Supplementary Fig. S2 with the online version of this paper at <http://mic.sgmjournals.org>.

Historically, the anaerobic gut fungi have been assigned to genera and species based on physiological and developmental characteristics, such as thallus morphology, the location of nuclei in monocentric and polycentric forms and the number of flagella on each zoospore (Munn *et al.*, 1988). These characteristics tend to be pleomorphic, varying with culture condition and particularly carbon source, and are only visible by microscopy, prompting the application of more-recent DNA-based technologies in phylogenetic

analyses (Brookman *et al.*, 2000a; Mennim, 1997). Six genera of fungi belonging to the order *Neocallimastigales* are now recognized, namely: *Anaeromyces*, *Caecomyces*, *Cyllamyces*, *Neocallimastix*, *Orpinomyces* and *Piromyces* (Theodorou *et al.*, 1996; Ozkose *et al.*, 2001).

DNA-based techniques have facilitated the understanding of the phylogenetic relationships and diversity of micro-organisms in natural ecosystems: they introduce considerably fewer biases in sampling than culture-based methodologies, can be generated directly from DNA and are considered more representative of the entire community than culture-derived data alone (Ward *et al.*, 1990). Favoured indicators of genetic diversity are the rRNA encoding gene sequences, particularly the internal transcribed spacers ITS1 and ITS2 and the intervening 5·8S rDNA; these can be used both to identify micro-organisms and to determine phylogenetic relationships within communities (Hausner *et al.*, 2000; Vainio & Hantula, 2000; Ranjard *et al.*, 2001), including the gut fungi (Brookman *et al.*, 2000a).

Although a variety of DNA-based separation and visualization techniques, including RFLP, temperature gradient gel electrophoresis (TGGE) and denaturing gradient gel electrophoresis (DGGE), have been used to examine phylogenetic relationships and to highlight diversity in microbial communities, the identification of population shifts in ecosystems with rapidly changing microbial communities is hampered by the requirement for alignment of large numbers of previously determined ribosomal sequences. This process is particularly laborious when populations are changing in time and space and where only one or a few new sequences are obtained at any given time or in any given space. This situation is prevalent in the rumen ecosystem, where freshly ingested plant biomass is rapidly colonized by a succession of microbial species, and is the reason why the primary colonization hypothesis for the gut fungi has been so difficult to substantiate.

Fingerprinting programs [e.g. Prosite (Hulo *et al.*, 2004), Prints (Attwood *et al.*, 2003), PFAM (Bateman *et al.*, 2002), Smart (Schultz *et al.*, 1998) and Interpro (Mulder *et al.*, 2003)] that rely upon pattern matching as well as sequence are widely used for the identification of proteins. A variety of different methodologies underlie these programs; for example, Prosite uses regular expressions, whilst PFAM uses hidden Markov models (HMMs). In contrast, little use of these pattern-matching methodologies has been made in nucleic acid studies. There are exceptions, however, such as MatInspector, for identifying transcription factor-binding sites (Quandt *et al.*, 1995). In this paper, the development of a fingerprinting (pattern matching) approach is described for the rapid identification and classification of gut fungi from ribosomal ITS1 and ITS2 sequences. The methodology was developed using DNA obtained from axenic cultures and has been used to interrogate the diversity of fungal populations in *in vivo* samples. While the technique has potential generic application in studying diversity and

population dynamics in microbial ecology, it was developed for use in future studies to elucidate the role played by anaerobic fungi in the primary colonization of plant biomass in the rumen ecosystem.

METHODS

Isolate collection and sequence preparation. Genomic DNA was extracted from 14 anaerobic fungi. Twelve of the fungi were isolated from the rumen contents or faeces of wild herbivores (addax, banteng cattle, gembok, kudu, nilgai, oryx, sable and zebra) grazing native pastures at the Tipperary Sanctuary in the Northern Territory, Australia, and two (isolated from a sheep and an African elephant) were a gift from Dr Colin Orpin. Each isolate was assigned to a genus based on culture morphology, and the ribosomal ITS1 and ITS2 fragments amplified by PCR using primer set: 18S forward primer JB206 (GGA AGT AAA AGT CGT AAC AAG G) and the 28S reverse primer JB205 (TCC TCC GCT TAT TAA TAT GC). The ribosomal fragments were cloned into pGEMT-Easy (Promega) and multiple clones for each isolate were selected by restriction digest patterns to maximize the chance of polymorphic sequences; where only a single sequence is given, no polymorphisms were observed. Clones were sequenced to at least 99% accuracy. Fifty sequences were obtained from the 14 different isolates with sequence names shown in parentheses: TSB1 (3-1, 3-2 and 3-5); TAP F9 (7-1, 7-2, 7-4, 7-5 and 7-10); TAP F11 (9-1, 9-2 and 9-6); TAP F8 (13-4 and 13-6); TNL1 (27-2, 27-3 and 27-6); TNL2 (28-3, 28-4, 28-5 and 28-6); TZB1 (29-1, 29-2, 29-5 and 29-6); TZB2 (30-2, 30-4, 30-5 and 30-6); TOX1 (53-1, 53-2, 53-3 and 53-4); TAX1 (55-1 and 55-6); TBT2 (56-1, 56-3, 56-4, 56-5, 56-6 and 56-8); TBT3 (62-1, 62-2, 62-3, 62-4, 62-5 and 62-6); PN1 (72-2); AE1 (74-2, 74-3 and 74-6). Nine additional sequences were added from isolates in the Institute of Grassland and Environmental Research (IGER) culture collection; six of these were polymorphic ITS1 sequences from previously characterized isolates: PLA1, NMW3, NMW5, NCS1, OUS1 and PCS1 (Brookman *et al.*, 2000a); the new sequences are suffixed 2, whereas the originally published sequence is now suffixed 1; for example, the previously published sequence from PLA1 with accession number AF170207 is given as PLA1-1 and the second newly described sequence from this isolate is given as PLA1-2. The three remaining sequences were from strains AUC3 and AUC4, isolated from the rumen of a cow in the UK and characterized morphologically as *Anaeromyces*, and from strain NMZ4, isolated from the faeces of a Malaysian zebu by Dr Michelle Lawrence and characterized as belonging to the genus *Neocallimastix*.

Forty-eight ITS1 sequences from environmental samples, namely faeces of wild and domesticated animals collected in Australia and Africa, were also analysed using the HMM-based program in this study. The ITS1 sequences from the environmental samples were amplified using primers specific for the gut fungi, forward primer MN100 (TCCTACCCTTTGTGAATTTG) and reverse primer MNGM2 (CTGCGTCTTCATCGTTGCG), to give gut-fungal-specific amplification of a truncated ITS1 sequence without Variable Region I at the 5' end (see Supplementary Fig. S1, available as supplementary data with the online version of this paper at <http://mic.sgmjournals.org>, for region nomenclature). This primer set was necessary, despite the loss of some sequence information compared with Brookman *et al.* (2000a), to prevent contamination of the amplicon population with aerobic fungal sequences from the environmental samples. The samples and the sequences isolated from them form part of a wider study described elsewhere (M. J. Nicholson and others, unpublished data).

Sequence analysis. Multiple alignments of sequences were generated using CLUSTAL (Thompson *et al.*, 1997) and manually edited using Align (<http://science.do-mix.de/>). RNA secondary-structure

predictions were carried out using MFOLD 3.1 (Zuker *et al.*, 1999; Mathews *et al.*, 1999) at <http://www.bioinfo.rpi.edu/applications/mfold/old/rna/>, with a preset folding temperature of 37 °C, and the following default options: ionic conditions, 1 M NaCl, no divalent ions; percentage suboptimality number, 5; upper bound on number of computed foldings, 50. The 18S and 5·8S sequences were removed manually before folding. Predictions were carried out for the following representative sequences: 3-1, 7-4, 28-3, 29-1, 56-5, NMW5-2, NCS1-2 (this study), AUC1 (Brookman *et al.*, 2000b), PCS1-1, PLA1-1, OUC1A (OUC1-1, AF170189), NMW4, *Neocallimastix patriciarum*, PAC1, PAK1, PCG1, *Piromyces* (Brookman *et al.*, 2000a). Energies of structures ranged from ~ -35 kcal mol⁻¹ to ~ -93 kcal mol⁻¹. RNA structures were drawn using RNAviz (<http://rrna.uia.ac.be/rrnaviz/>).

The pattern matching programs were written in PERL 5.6.1 with BioPERL 1.0 modules and run in Cygwin on a PC. HMMER2.2 was downloaded from <http://hmmer.wustl.edu/> (Eddy, 1998, 2001). HMMs were built and run according to the accompanying documentation using the following clusters to generate the group HMMs (Table 1): the *Anaeromyces* model incorporates individual HMMs for I3, II6, III4 and IV4; *Neocallimastix* for I2, II2, II3, II4, II5, III2, III3 and an HMM for IV2 and IV3 together; the *Piromyces* I model has individual HMMs for I1, II1, III1 and IV1; *Orpinomyces/Piromyces* II for I5, I5a, II8, III6, III7, III8 and IV5; and finally *Piromyces* III has individual HMMs for I4, I6, II7, III5, a joint HMM for III9 and III10 together, and IV6. A copy of the program is available from the authors on request.

RESULTS

Sequence and structure of ITS1

A multiple alignment of 50 ITS1 sequences derived from 14 Australian anaerobic gut fungal isolates, nine new sequences from the IGER culture collection, together with the 25 ITS1 sequences reported in Brookman *et al.* (2000a) was generated. A condensed version of this alignment showing the ITS1 sequences for 12 representative isolates can be found in Supplementary Fig. S1. Within the ITS1 sequences, four regions of major variation (Variable Regions I–IV) were seen, flanked by regions of largely conserved sequence. To improve the alignment and provide structural explanations for the conservation and variation seen, RNA secondary-structure predictions were carried out for a selection of 17 sequences, each representative of a set of similar sequences in the alignment.

Secondary-structure predictions indicated that despite significant sequence variation between the different ITS1 sequences, the overall secondary structure for ITS1 was broadly conserved for all isolates. Variable Region I formed a stem and loop (Fig. 1), consistent with the complementary sequences in the ‘stem’ in this region. Although stem length and terminal bulb sequence varied, this structure was the consensus prediction for 11 of the 17 sequences and was seen in 65 of 113 of the predicted structures. The consensus prediction was not seen for the ITS1 structure from *Piromyces* isolates PAC1 and PAK1, and this region of sequence (nucleotides 61–97) does not align with the other gut fungal sequences. The sequence from Variable Region II to Variable Region IV formed a long stem and bulb. This

was the consensus prediction for all sequences, including PAC1 and PAK1, and was seen in $\sim 80\%$ of structures (Fig. 2). The conserved regions between Variable Regions II and III and between Variable Regions III and IV were found midway along the stem and were complementary to one another, as illustrated by the boxed region in Fig. 2. The stem regions flanking this section varied in length for the different isolate types (Fig. 2).

The conserved sequences between Variable Regions I and II and between Variable Region IV and the 5·8S gene sequence did not appear to form a conserved structure within the ITS1 region as analysed here. However, the complementarity of sequences at the 3' end of the 18S region (nucleotides 46–55) and at the 5' end of the conserved region between Variable Regions I and II (nucleotides 99–108) suggest that there may be structural influences exerted upon these sequences from outside of the immediate ITS1 region, as seen in the yeast ITS1 structure (van Nues *et al.*, 1994).

Clustering of ITS1 sequences and generation of ITS1 fingerprints

By clustering similar sequences, several different motifs with differing levels of degeneracy were identified within each of the four variable regions of the ITS1 sequences (Table 1). The motifs were based upon primary sequence alignment and predicted structure, and were generated from isolates previously classified by phylogenetic methodologies as references.

Almost all of the sequences in each variable region contained one of the recurring motifs for that region; for example, of the 84 ITS1 sequences, 79 contained one of only six motifs described for Variable Region I (Table 2). Furthermore, the combination of these motifs across the four variable regions could be used to define a fingerprint for each ITS1 sequence. Groupings made in this way correspond well with the generic assignments of the originating isolates made according to morphological criteria and/or sequence-based phylogenetic analyses. Thus, each of the groups were defined by the ITS1 fingerprints within that group; for example, *Anaeromyces* have the following fingerprint for the four variable regions: Region I, Motif 3; Region II, Motif 6; Region III, Motif 4 and Region IV, Motif 4 (I3,II6,III4,IV4), whilst *Neocallimastix* have fingerprints made up of a combination of the following motifs: Region I Motif 2, Region II Motif 2, 3, 4 or 5, Region 3 Motif 2 or 3, and Region IV, Motif 2, 3 or 5 (I2,II2/3/4/5,III2/3, and IV2/3/5; Table 2).

The 28 sequences from 16 *Neocallimastix* isolates formed a single group when sorted by their fingerprints; similarly, the five sequences from the five *Anaeromyces* isolates formed a single group. The 31 sequences from 13 *Piromyces* isolates formed three groups (*Piromyces* I, II and III) plus a fourth collection of individual sequences designated ‘miscellaneous *Piromyces*’ due to their lack of coherent grouping (PAC1, PAK1, PCG1). In cases when different polymorphic

Table 1. Variable Region sequence motifs for the ITS1 sequence groups

Motif sequences are given in the form for PERL regular expressions, such that the symbols | or . represent any character or space, and where variable numbers of residues are shown as e.g. T{2,4}, representing any number of T residues between 2 and 4. Motifs were drawn from sequences 62–96 for Variable Region I; 130–196 for Variable Region II; 198–266 for Variable Region III and 264–318 for Variable Region IV, as per alignment in Supplementary Fig. S1.

Region	Motif	Motif sequence	Group
I	1	GGAGTG(G)CCGTTGTTTTGGTCA(CCITT)	<i>Piromyces</i> I
	2	AT(TI)C(GIA)(GI)CT(CIT)GATTGAGAGTGATT(TIA)	<i>Neocallimastix</i>
	3	TTTCCAACAAATGTGTT(GI)GATAAT	<i>Anaeromyces</i>
	4	GGTCTACGTTTT(TI)AAGT(TI)AGAC(CGCI)(GIA)CT	<i>Piromyces</i> III
	5	AATCAC(AIT)GT(TIA)TTGTGATTTT	<i>Orpinomyces/Piromyces</i> II
	5a	ATCAC(AIT)TTT(TI)GT(AIT)TGTGATTT	<i>Orpinomyces/Piromyces</i> II
II	6	GGCTCGATTTT(TI)AGAGCA(TIC)T	<i>Piromyces</i> III
	1	TTGTGTAATAATTTTTTCCAATGAAAAAATTATTTGTCTAT	<i>Piromyces</i> I
	2	TTGTTGTTGAAATAATTTTTTTTTCACGTGGAATAAAAAAAAAA(AI)TTATTTATG-GTTTTGTCTAT	<i>Neocallimastix</i>
	3	TTGT((TTG)) (AIT)AATAATTTTTTTT(TI)(TI)GATTTTAAAAAAAAA(AI)TTATTTATG-(GIT)T(TIG)TT(GIAC)(TI)(CIGIT)TAT	<i>Neocallimastix</i>
	4	T(TIG)GTTTGAATATTTATTTATG(GTTI)TTGTCTAT	<i>Neocallimastix</i>
	5	TTGTTTGTAAATGTTATTTATG(GTTI)TTGTCTAT	<i>Neocallimastix</i>
	6	TTGTTTGTTTGAAATAATTTTTTTT(TI)AAAAAATTATTTAGAATGTCTTT	<i>Anaeromyces</i>
III	7	TTGTTT(TI)GAAA(AI)(TIC)AATTTTTTTT(TI)(TI)(TI)AAAA(AI)TTATTGA(AIT)TGTC-TTT	<i>Piromyces</i> III
	8	TTG(TI)TTTTGT((AAA) (TT) (TAA))GAAA(AI)TTATGT(TIC)GTCTAT	<i>Orpinomyces/Piromyces</i> II
	1	CCCTGAAGAA((AA))GTAGA(AIT)(TIA)(AIT)TACTTTTTTTAGGG	<i>Piromyces</i> I
	2	CCC(TIA)(. {0,3})AAA(. {0,2})TTG(GI)(T{2,4})(. {0,2})GTT(GI)TGTTG(TIC)AA(AIG)(AI)-(AI)(GIC)ACAA(GI)((TG) (GT) (TA))A(. {2,4})A(T{2,4})(AIT)GG	<i>Neocallimastix</i>
	2a	CCCTAAA(AI)TTG(GI)TTTGTT((TG) (GT) (TI)(GI)AAA((GACAAG))TGA(AI)TCAA(AI)-TTTTAGG	<i>Neocallimastix</i>
	3	CC(CI)T((TGG)) AAA(GI)TT(TIG)GTT(TI)CATTTAAACCC((AAGTGAATCATTT)) A-(GIA)GG	<i>Neocallimastix</i>
	4	CCCTGAAATTGTGAGGAGCTTGTAATGGT(GIA)A(AIC)TTGCAATTCAGGG	<i>Anaeromyces</i>
	5	CCCTGTTGAATAGAGTGTTCG((AA) (TT))CA((TGA) (GGG))AATTCGACAGGG	<i>Orpinomyces/Piromyces</i> II
	6	CCCTG(GIA)A((TA) (AT))TG(AIG)AAGAAA(AI)GTCGT(TI)AAGATTTTTTTT(TI)CAT-TC(TI)AGGG	<i>Orpinomyces/Piromyces</i> II
	7	CCCT(TI)GGATT(TI)ATTTGGAAAAGTTGAAAGACTTTTTCATTT(AI)ACCA(AI)GGG	<i>Orpinomyces/Piromyces</i> II
IV	8	CCCTGGTGGTTTT(TI)(TI)GGAAAAGTTGAAAACCTTTTTCGAATCC(CI(CT) (TC))A-GGG	<i>Orpinomyces/Piromyces</i> II
	9	CCCTGTTGAATTGGTTGTC(TI)TTAATG(AIG)CAA(AIG)TGATTCGCAGGG	<i>Piromyces</i> III
	10	CCCTGTTGAATT(GIA)GTT(GI)GTCGTAAGAC(TI)AAGTGATTC(TIG)CAGGG	<i>Piromyces</i> III
	1	GGGA(GIA)TAGATTTAATAATTTTTTCAGTGAAAAGACC(TIA)T	<i>Piromyces</i> I
	2	GG(GI)AATAGGCTTTCCTAAATAATTGATTTTTTTT(TI)(CI)(CI)CAA(TI)GTCTTAAA-AG(AI)CCTT	<i>Neocallimastix</i>
	3	GG(GI)(AIC)ATAG(GI)CTTCA(TIC)AA(AI)(TI)AAG(TI)(TI)(TI)(TI)TTTTTT(AI)AAG-TCTT(AI)AAAG	<i>Neocallimastix</i>
	4	GGGGATAGATTTTC(TIG)AGTAATTTTTTAGGAAAAGACCTT	<i>Anaeromyces</i>
	5	GG(GI)A(AIC)TAGGCTTTCATAGAC(AIC)TT	<i>Orpinomyces/Piromyces</i> II
	6	GGGGA((GT) (TA))GTC(AIG)(GIA)TAATTTTT(AI)AAAAAGACCTT	<i>Piromyces</i> III

sequences were obtained from the same *Piromyces* isolate, both sequences grouped together in one of the three *Piromyces* groups. All 15 sequences from the seven *Orpinomyces* isolates fell within a single group together with the *Piromyces* II group of sequences, which contains a deposited *Piromyces* sequence, and both ITS1 sequences from TAX1 (55-1, 55-6;

this study), another *Piromyces* isolate (*Orpinomyces/Piromyces* II; Table 2). These data parallel very closely the phylogenetic distribution of the 25 sequences determined using more-conventional sequence-based phylogenetic analyses (Brookman *et al.*, 2000a). For the 50 new ITS1 sequences from the 14 Australian isolates, all grouped within the

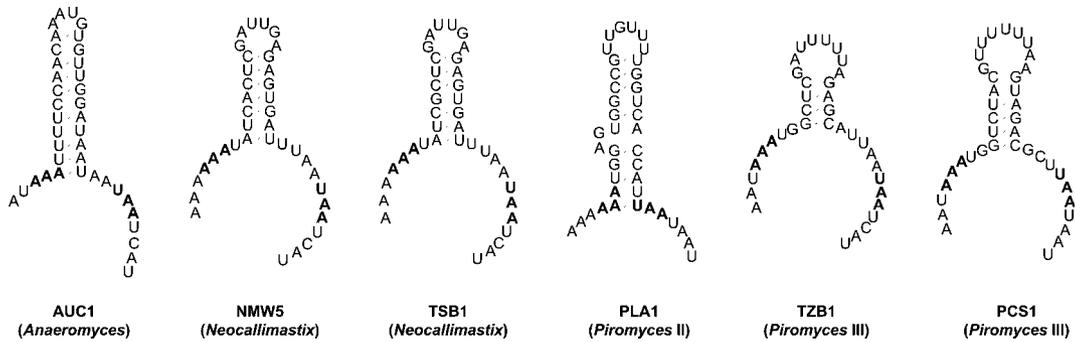


Fig. 1. RNA structure for Variable Region I (predictions were carried out for the full-length ITS1 sequence). Bases given in bold are conserved, the border sequences AAA, nt 58–60, and U(T)AA, nt 111–113, are boxed in Supplementary Fig. S1 (available as supplementary data with the online version of this paper at <http://mic.sgmjournals.org>). Structures shown are from sequences AUC1, NMW5-2, 3-1, PLA1-1, 29-1 and PCS1-2, reading from left to right (see also Table 2).

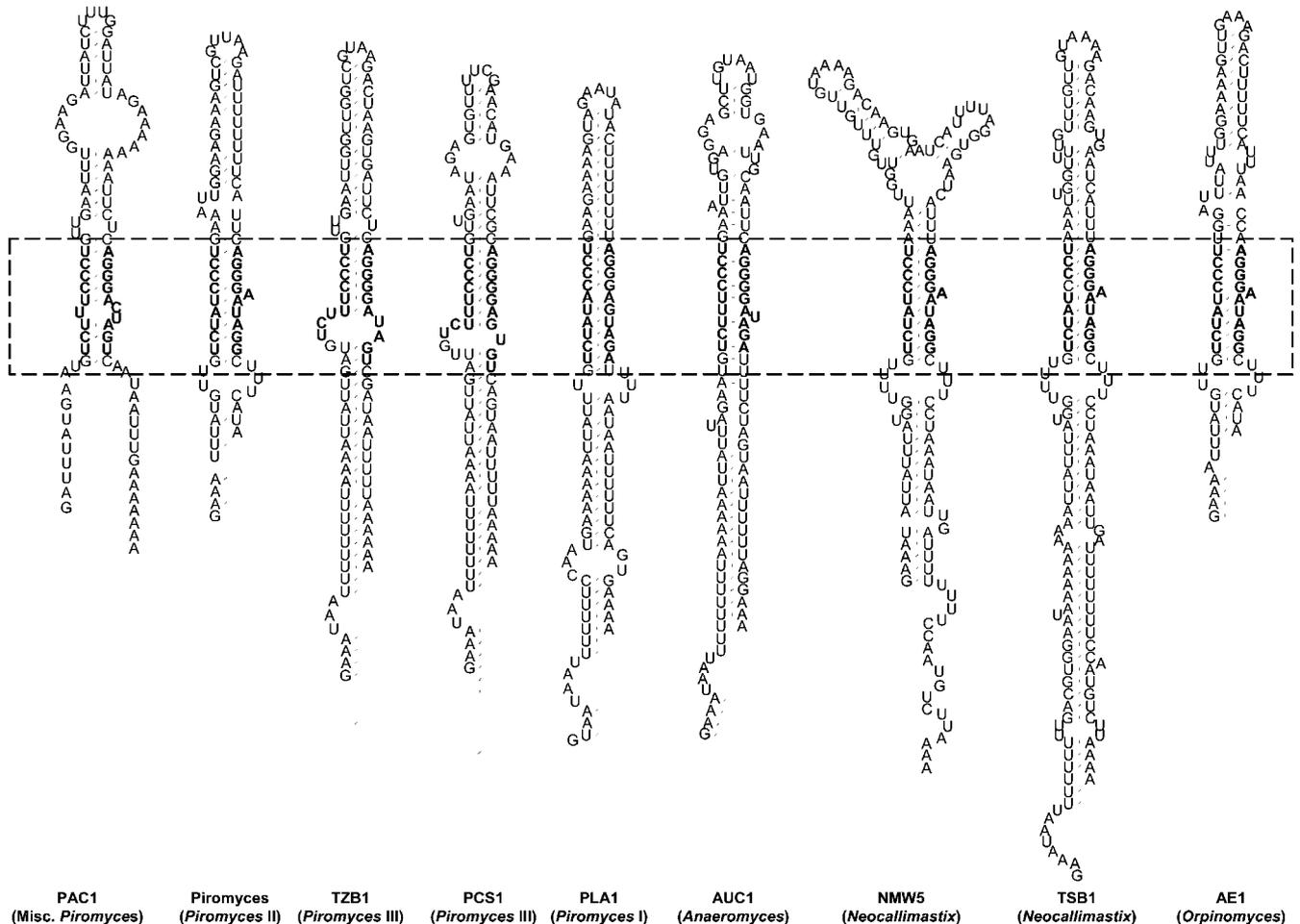


Fig. 2. RNA structure for the ITS1 sequences from Variable Regions II–IV (predictions were carried out for the full-length ITS1 sequence with 18S and 5.8S sequences removed). Structures correspond to the sequence in positions 148–339 in Fig. 1. Bases given in bold are conserved and correspond to positions 202–212 and 290–299. Note the differences in stem length either side of the conserved regions. Structures shown are from sequences PAC1, *Piromyces*, 29-1, PCS1-2, PLA1-1, AUC1, NMW5-2, 3-1 and 7-4, reading from left to right (see also Table 2). Misc., miscellaneous.

Table 2. ITS1 sequence fingerprints enabling differentiation of fungal types

The numbers in each of the four Variable Regions denote the motif sequence identified in the respective region, as given in Table 1 ('–' indicates that the sequence did not match any of the recognized motifs). Bracketed sequences were identical over the sequence alignment.

Sequence name	Variable Region				Group
	I	II	III	IV	
PUC1, PLA1-1	1	1	1	1	<i>Piromyces</i> I
PLA1-2	1	–	1	1	
3-1, 56-1, 56-3, 56-4, 56-6, [56-8, 62-2, 62-6], [62-1, 62-4, 62-5], 62-3	2	2	2	2	<i>Neocallimastix</i>
NMZ4	2	3	2	2	
NCS2	2	–	2	2	
NMW3-2	2	5	2	2	
NMW4	2	5	3	2	
NMW5-2	2	4	2	2	
<i>N. frontalis</i> , 3-5, 3-2, <i>Neocallimastix hurleyensis</i> , 56-5, NCS1-2, NMG2, NMW5-1, NUC1, NCS1-1	2	3	2	3	
NMW2	2	4	2	3	
NMW1	2	5	3	3	
<i>N. patriciarum</i>	2	5	3	5	
AUC1, [AUC2, AUC4], AUC3, <i>An.</i>	3	6	4	4	<i>Anaeromyces</i>
55-1, 55-6, <i>Piromyces</i>	5a	8	6	5	<i>Orpinomyces/Piromyces</i> II
[13-4, 13-6, 7-1], 7-4, 7-5, 7-10, 9-1, 9-2, 9-6, OR	5	8	7	5	
7-2, OUC1-1, OUC1-2, OUS1-2, OUS1-1	5	8	8	5	
PCS1-2, 28-4	4	7	5	6	<i>Piromyces</i> III
[27-2, 27-3, 27-6], 28-3, 28-5, 28-6	4	7	9	6	
29-1, [29-2, 29-5, 30-2, 30-4, 30-5], 29-6, 30-6, [53-1, 53-2, 53-3, 53-4], 72-2	6	7	10	6	
[74-2, 74-3, 74-6]	–	7	10	6	
PCS1-1	5a	7	5	6	
PAC1	–	–	–	–	Miscellaneous <i>Piromyces</i>
PAK1	–	–	–	–	
PCG1	4	–	2	–	

Neocallimastix, *Orpinomyces/Piromyces* II or *Piromyces* III clusters.

A pattern-matching program for ITS1 fingerprints

A pattern-matching program was written incorporating the sequence motifs described in Table 1 as explicit, regular expressions. All 84 ITS1 sequences were analysed using the pattern-matching program, and most matched just one motif in each of the four variable regions, that is, no single ITS1 sequence gave multiple matches to fingerprints within the same region, suggesting an appropriate level of condensation of consensus sequences within the fingerprints.

The pattern-matching program was also used to analyse a set of 48 ITS1 sequences amplified directly from faeces rather than from isolated, axenic cultures. Due to the complex mixture of eukaryotic DNA in these environmental samples, PCR was performed using gut fungal-specific primers producing a slightly shorter amplicon spanning Variable Regions II–IV of the ITS1. The program initially provided annotation for just five of the 48 sequences (matches to motifs in all three variable regions), with

probable annotation for four others (matches to motifs in two of the three variable regions). Examination of these sequences indicated that in many cases the lack of a match was due to a simple difference, such as single base change between the regular expression and the experimental sequence. To counter this problem, the pattern-matching program was redesigned to use HMMs via the HMMER suite of programs (Eddy, 1998, 2001). HMMs permits scoring of matches to motif sequences and thus offer flexibility compared to regular expressions. HMMs were generated for each of the motif sequences detailed in Table 1, and these HMMs were then combined to give one for each group.

The modified program was then used to annotate the ITS1 sequences from characterized fungal isolates, and in all cases the sequences were assigned to groups in accordance with the original annotations. Examples of program outputs for sequences from isolates TAX1, AN and OUC1 are shown in Fig. 3. Matches to multiple HMMs, where a sequence showed similarity to more than one motif for any given variable region, could be resolved on the basis of E-value scores, such that the lower E value was taken as the best match. For example, the sequence for *Anaeromyces* isolate AN showed a fingerprint of I3/II6/III4/IV4, but for Variable

Region II, the sequence also matched Motif 7 (II7), but at a much lower stringency (1.7×10^{-5} for Motif 7 cf. 1.9×10^{-12} for Motif 6).

When the HMM-based program was used to annotate the 48 ITS1 sequences amplified from faecal samples, the analysis was extremely successful overall. In the first instance, 17 of the 48 sequences were unambiguously assigned to one of the established sequence groups on the basis of matches to sequence motifs in each of the three variable regions (II–IV), and six sequences were tentatively assigned on the basis of matches for two out of three variable regions. Of the remaining 25 sequences, 14 matched motifs in only one variable region and 11 did not match motifs in any of the variable regions. However, further examination of these 25 sequences revealed that they could be divided into four discrete novel sequence groups that did not match any of the existing fingerprints plus two orphan sequences that matched neither the old nor the new groupings, nor one another.

The development of this fingerprinting methodology was designed around the four anaerobic gut fungal genera for which reliable data were available (*Anaeromyces*, *Neocallimastix*, *Piromyces* and *Orpinomyces*). Although one deposited ITS1 sequence from a fungal isolate designated *Caecomyces* (accession no. AF492020) was available, it showed over 99% sequence identity to the *Neocallimastix* group of isolates and thus, as it was probably wrongly assigned, it was removed from the database. To extend the fingerprinting technique to cover all of the known anaerobic gut fungal genera, sequences representative of the missing genera have since been obtained (Ozkose, 2001), representing four undeposited ITS1 sequences from a single *Caecomyces* isolate and nine sequences from two *Cyllumyces* isolates. These, together with a single sequence from

a *Caecomyces* isolate, NZB7 (M. J. Nicholson, unpublished data), were used to produce alignments and sequence fingerprints for the ITS1 Variable Regions I–IV for *Caecomyces* and *Cyllumyces* (Table 3). The sequences from the four novel groups identified from the faecal samples were then compared with these additional fingerprints. One of these novel groups, which consisted of 11 sequences, was identified as representing the genus *Cyllumyces* (Supplementary Fig. S2, available as supplementary data with the online version of this paper at <http://mic.sgmjournals.org>). The consensus from the 11 environmental sample sequences did not match the fingerprints exactly, often showing single base pair differences or differences in the numbers of repeated bases, as reported earlier for the other genera. Alignment of sequences in Variable Region III showed a larger difference in a subgroup of the faecal samples; nine of the 11 clustered together in this region and required a separate sequence motif (Variable Region III, Motif 13; Table 3) to describe them. The two remaining sequences, J162 and J163, clustered with the sequences from the isolated *Cyllumyces* strains in this region with the Variable Region III Motif 12 fingerprint (see Supplementary Fig. S2, available as supplementary data with the online version of this paper at <http://mic.sgmjournals.org>). Motif sequences for Variable Regions II–IV for the *Cyllumyces* genus fingerprint have been modified to incorporate the additional sequence data (Table 3). None of the novel groups identified in the faecal sequences represented the genus *Caecomyces*.

To test further the utility of the HMM program, a database of 21 416 sequences was generated from GenBank (<http://www.ncbi.nlm.nih.gov/>) using the text query 'ITS1'. This included bacterial, fungal, animal and plant ITS1 sequences. Probing this database with the HMM program gave matches to the 20 gut fungal sequences deposited to date only, and

TAX1 (Sequence 55-1)

```
HMM name I5a Domain start 61 end 83  evalue 5.6e-07
HMM name II8 Domain start 112 end 137  evalue 2.3e-07
HMM name III6 Domain start 136 end 184  evalue 6.7e-11
HMM name IV5 Domain start 183 end 204  evalue 2.6e-07
Orpinomyces/Piromyces II - 4/4 fingerprints
```

Anaeromyces strain (Sequence AN)

```
HMM name I3 Domain start 60 end 84  evalue 3e-07
HMM name II6 Domain start 112 end 159  evalue 1.9e-12
HMM name II7 Domain start 112 end 159  evalue 1.7e-05*
HMM name III4 Domain start 158 end 204  evalue 2.9e-15
HMM name IV4 Domain start 203 end 239  evalue 3.4e-11
Anaeromyces - 4/4 fingerprints
```

OUC1 (Sequence OUC1-1)

```
HMM name I5 Domain start 61 end 82  evalue 2.8e-06
HMM name II8 Domain start 111 end 137  evalue 1.7e-07
HMM name III7 Domain start 136 end 183  evalue 3.1e-08
HMM name III8 Domain start 136 end 183  evalue 2.2e-14**
HMM name IV5 Domain start 182 end 203  evalue 2.6e-07
Orpinomyces/Piromyces II - 4/4 fingerprints
```

Fig. 3. Example output for HMM searches of ITS1 sequences showing different results and their interpretation. Each region and fingerprint sequence is represented by an HMM, thus Variable Region I Motif 5a is represented by 'HMM name I5a'. The group assignment is given in bold for each sequence: sequence 55-1 from isolate TAX1 matches only the correct motifs (Table 1); the deposited *Anaeromyces* sequence, AN, shows two matches in Variable Region II, a correct match to II6 and a false-positive match to II7(*). The false-positive match has a considerably lower E value. OUC1-1 matches III7 and III8; both profiles are for the same group and so are positives, but the optimal match to III8 (**) has the lower E value.

Table 3. Variable Region sequence motifs for the *Caecomyces* and *Cyllamyces* ITS1 sequences

Motif sequences are given in the form for PERL regular expressions, such that the symbols | or . represent any character or space, and where variable numbers of residues are shown as e.g. T{2,4}, representing any number of T residues between 2 and 4. Motifs were drawn from sequences 62–96 for Variable Region I; 130–196 for Variable Region II; 198–266 for Variable Region III and 264–318 for Variable Region IV, as per alignments in Supplementary Figs S1 and S2.

Variable Region	Motif	Motif sequence	Sequence group
I	7	GAGACCTC(T{2,4})GAAGGTC	<i>Caecomyces</i>
	8	GAGACCT(CI)GATT(T{1,3})GAAGGTC	<i>Cyllamyces</i>
II	9	TGTTTGAATAATTTTTTT(TI)CCAAAAAAA(AI)TTATTGAATGTC-TTT	<i>Caecomyces</i>
	10	((T{1,2})(G)(T{2,5})(A{1,2})(T{1,3})(A{1,3})(C)(T{1,2})(G)(TI)GT-C(T{2,3}))	<i>Cyllamyces</i>
III	11	CCCTGTTTGGTGGAGCTTGTAAGTGAACCATCAGGG	<i>Caecomyces</i>
	12	CCCTGT(TIC)(TI)GGTTTGAAGCTTGTAAAAAGTTTCG(AIG)ACCTTCAGGG	<i>Cyllamyces</i>
	13	CCCTG(GIT)TTTT(TI)GATGGAGCTTGT((TIC)I)(TI)(A{3,5})G(CIT)-(GIA)AAATC(A{2,3})(T{1,2})(C{1,2})A(G{2,3})	<i>Cyllamyces</i>
IV	7	GGGGAATGGCAAACAATAA(TI)TTATTTTAAAAA(AI)GACCTT	<i>Caecomyces</i>
	8	GGGGAATGGC(AIG)AAAC(TIC)T	<i>Cyllamyces</i>

not to any other fungal, animal or plant sequence; the only weak matches made were to four AT-residue-rich *Plasmodium falciparum* genome sequences. Therefore the 'gut-fungal trained' version of the program is specific for anaerobic gut fungal ITS1 sequences and represents an effective way of identifying and annotating these ITS1 sequences.

ITS2 sequence fingerprints

The ITS2 sequences of the Australian anaerobic fungal isolates were aligned together with sequences from five isolates previously characterized phylogenetically from their ITS1 sequence (Brookman *et al.*, 2000a) plus the sequence from the *Anaeromyces* isolate AUC4 and a *Neocallimastix* isolate, NMZ4, from the IGER culture collection (data not shown). The alignment was examined to identify regions of sequence that might be phylogenetically useful. As for the ITS1 sequences, regions of variation were found, flanked by regions of highly conserved sequence. Three variable regions were identified (Variable Regions V, VI and VII). The first of these (Variable Region V) was not suitable for generic delineation; however, the two downstream variable regions (VI and VII) were good discriminators, and the ITS2 sequences could be clustered on the basis of sequence similarity within these regions (Table 4).

Several different sequence motifs were identified in Variable Regions VI and VII, and the motifs partitioned with the genus groupings identified with the ITS1 analysis (Table 2). The single representative sequence of the *Piromyces* I group, PLA1, defined a different sequence cluster for the Variable Regions V, VI and VII, as would be expected, but as it is a single sequence no motifs have yet been assigned.

Subsequent to this analysis, several ITS2 sequences were deposited into the public database by Fliegerova *et al.* (2004) and were added to the alignment, including three *Orpinomyces* (AY429671, AY429672, AY429673) and four *Anaeromyces* (AY429666, AY429667, AY429669, AY429670) sequences. The additional *Orpinomyces* sequences clustered with those from this study designated as belonging to the *Orpinomyces* group (Table 4). The *Anaeromyces* sequences clustered with the single *Anaeromyces* sequence, AUC4, giving rise to the definition of motifs Variable Region VI Motif 4 and Variable Region VII Motif 7 (Table 4).

The Variable Region VI motifs were less variable than those from Variable Region VII, with an almost unique sequence for each genus, whereas the Variable Region VII identified subgroups within the *Piromyces* III and *Neocallimastix* clusters (Table 4).

The subgroups identified using ITS2 sequences were broadly the same groups as seen with ITS1 fingerprints in Table 2. There were instances in which sequences that were identical across ITS1 were non-identical in their ITS2 sequences: two of the sequences from isolates TBT3, namely 62-1 and 62-4, were identical in their ITS1 sequences (ITS1 fingerprint I2,II2,III2,IV2), but sequence 62-1 had an ITS2 fingerprint of V4,VI3a,VII5, whilst sequence 62-4 had a fingerprint of V4,VI3b,VII6 (Table 4). Similarly, the sequences from the *Piromyces* III group isolate AE1 (74-2 and 74-3/74-6) have different ITS2 fingerprints but identical ITS1 sequence and fingerprints (Table 4). Conversely, sequences 28-3 and 28-4 from isolate TNL2 were identical in their ITS2 sequence but not in their ITS1 sequences. This variability within single isolates warns against over-interpretation of ITS1/ITS2 sequence differences between different fungi.

Table 4. ITS2 fingerprints enabling differentiation of fungal types

The numbers in each of the three ITS2 Variable Regions denote the motif sequence identified in the respective region (‘-’ indicates that the sequence did not match any of the recognized motifs). Bracketed sequences were identical over the sequence alignment.

Sequence name	Variable Region			Group	ITS1 fingerprint
	V	VI	VII		
PCS1-2	1	1	1	<i>Piromyces</i> III	I4,II7,III5,IV6
27-3, 28-5, [27-2, 27-6], 28-6, [28-4, 28-3]	1	1	1		I4,II7,III9,IV6*
74-3, 74-6	3	1	2		I-,II7,III10,IV6
74-2	2	1	3		I-,II7,III10,IV6
[29-1, 29-2, 29-6, 30-5], 29-5, [30-2, 30-4], 30-6, [53-1, 53-2], 53-3, 72-2, 53-4	2	1	3		I6,II7,III10,IV6
7-1, 9-2, 7-4, 9-6, 7-5, 13-6, 13-4, 7-10, 9-1	1	2	4	<i>Orpinomyces/Piromyces</i> II	I5,II8,III7,IV5
[OUS1-2, 7-2] AY429671, AY429672, AY429673	1	2	4		I5,II8,III8,IV5
[55-1, 55-6]	2	2	—		I5a,II8,III6,IV5
3-1, 56-1, 56-3, 56-4, 56-6, 56-8, 62-1, 62-2, 62-3	4	3a	5	<i>Neocallimastix</i>	I2,II2,III2,IV2
3-2	4	3a	5		I2,II3,III2,IV3
62-4	4	3b	6		I2,II2,III2,IV2
NMZ4	4	3b	6		I2,II3,III2,IV2
[NMW5-1, 56-5], NCS1-1	4	3b	6	<i>Anaeromyces</i>	I2,II3,III2,IV3
AUC4, AY429666, AY429667, AY429669, AY429670	5	4	7		I3,II6,III4,IV4

*The motif ITS1 fingerprint for sequence 28-4 only in this subgroup is I4,II7,III5,IV6.

DISCUSSION

New ribosomal sequence data from environmental studies are usually aligned with existing sequences from databases and then analysed phylogenetically. These comparisons are often based on pair-wise comparisons of parameters such as parsimony or distance to generate a series of phylogenetic trees, which can then be analysed statistically. This is a time-consuming process, particularly as datasets grow, and for each new sequence entry the relationships between the previously characterized sequence data and the newly generated data are re-examined.

Secondary-structure prediction is a tool used in protein studies, and more recently for ribosomal ITS data in plants, to improve multiple sequence alignments (Bateman *et al.*, 2002; Coleman, 2003). By using secondary-structure prediction to aid alignment of sequences and then highlight areas of sequence useful for constructing fingerprints for phylogenetic assignment of new sequences, this study has extended the approach into a new area for molecular ecology.

The inclusion of HMMs for comparison of sequence similarities gives flexibility when matching new sequences to fingerprints and allows for the observed level of variability in the ribosomal sequences. Use of HMMs with sequence and secondary-structure predictions has been pioneered in similarity scoring for protein comparisons and provides a method for rapid comparison of protein structures (Bateman *et al.*, 2002; Pearl *et al.*, 2003). Similarly,

combining the sequence alignment, secondary-structure prediction and pattern-matching approaches for the ITS DNA sequences gives a route for rapid clustering of new sequences with predetermined fingerprints from characterized genera of gut fungi.

The analysis of rDNA data described here has generated clusters of related sequences outside of the fungal genera characterized to date. This has also been the case for protein studies, where databases containing outputs from the large genome sequencing datasets have identified families of related proteins with unknown functions (Bateman *et al.*, 2004). In both cases, new information will be generated that is likely to alter the description of the clustered protein or DNA sequences. For example, novel groups identified in the fungal ITS data may be assigned to newly defined genera as characterization of isolated strains progresses.

The utility of the approach described here in general terms is that it provides for environmental studies a rapid means of assigning new sequences to one of the previously characterized fungal genera or to the novel groups of related sequences. The fingerprinting approach provides a predetermined, discrete set of parameters for assignment of the sequences with the flexibility to update the information as new fingerprints are found.

On a more specific level, this study has shown a variety of polymorphic patterns of ITS1 and ITS2 sequences from the same isolate, for example, isolate TBT2 (sequences 56-1,

56-3, 56-4, 56-5, 56-6 and 56-8). As rRNA-encoding sequences are present as long tandem arrays in the genome of most eukaryotes, and it is thought that sequence conservation between individual repeats is preserved by concerted evolution (Liao, 1999; Wendel *et al.*, 1995), this polymorphism may be considered unusual. However, two ITS2 lineages have been demonstrated in a single *Fusarium* species (O'Donnel & Cigelnik, 1997) and ITS polymorphisms have been reported elsewhere (Hausner *et al.*, 2000; Vogler & DeSalle, 1994).

These sequence polymorphisms illustrate the limitations of this and any other approach for clustering of these fungal strains using ITS sequences at a level below that of genus. For example, the *Neocallimastix* sequences were tightly grouped in our previous analysis (Brookman *et al.*, 2000a), although three broad clusters were observed, and the present fingerprint analysis supports at least one of these clusters. However, in this study, a sequence from isolate NMW5 which grouped away from *N. frontalis* in Brookman *et al.* (2000a) is sorted into the *Neocallimastix frontalis* subgroup, suggesting that at least some of the intra-genus groupings are due to polymorphisms and do not represent different species. This uncertainty is compounded by the lack of agreement between subgroupings defined on the basis of ITS1 and ITS2, as also seen for the *Piromyces* III group in Table 4, and indicates that the ITS sequences are unlikely to be of use in understanding the relationships at the species level.

In conclusion, we have sequenced ITS1 and ITS2 DNA from a number of anaerobic fungi, determined the relationship between these sequences and ITS1 secondary structure, and used this information to derive a fingerprinting program by which the genus of a given isolate can be determined. This will facilitate high-throughput approaches for the determination of gut fungal flora and its relationship to microbial succession in the colonization of feed boli. Furthermore, training the program with sequences from other fungal genera would broaden the utility of the approach. We suggest that a similar pattern-matching approach with other fungal ITS sequences would provide a useful tool for rapid identification of new sequences from environmental studies.

ACKNOWLEDGEMENTS

Matthew Nicholson was a BBSRC-sponsored CASE PhD student with Manchester University and IGER, Aberystwyth. Jayne Brookman gratefully acknowledges the Stapledon Memorial Trust for a Travelling Fellowship. We would like to thank Dr Janet Taylor (IGER) for ongoing assistance with the PERL script and bioinformatics interpretation. We are grateful to Keith Joblin at AgResearch, New Zealand, for permission to use sequence obtained from a fungal isolate, NZB7, from the AgResearch culture collection. Australian fungal isolates were presumptively identified by Geoff Gordon using morphological characteristics.

REFERENCES

- Akin, D. E. & Benner, R. (1988). Degradation of polysaccharides and lignin by ruminal bacteria and fungi. *Appl Environ Microbiol* **54**, 1117–1125.
- Attwood, T. K., Bradley, P., Flower, D. R. & 9 other authors (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res* **31**, 400–402.
- Bateman, A., Birney, E., Cerruti, L. & 7 other authors (2002). The Pfam Protein Families Database. *Nucleic Acids Res* **30**, 276–280.
- Bateman, A., Coin, L., Durbin, R. & 10 other authors (2004). The Pfam Protein Families Database. *Nucleic Acids Res Database Issue* **32**, D138–D141.
- Bauchop, T. (1979). Rumen anaerobic fungi of cattle and sheep. *Appl Environ Microbiol* **38**, 148–158.
- Brookman, J. L., Mennim, G., Trinci, A. P. J., Theodorou, M. K. & Tuckwell, D. S. (2000a). Identification and characterization of anaerobic gut fungi using molecular methodologies based on ribosomal ITS1 and 18S rRNA. *Microbiology* **146**, 393–403.
- Brookman, J. L., Ozkose, E., Rogers, S., Trinci, A. P. J. & Theodorou, M. K. (2000b). Identification of spores in the polycentric anaerobic gut fungi which enhance their ability to survive. *FEMS Microbiol Ecol* **31**, 261–267.
- Cheng, K.-J., Forsberg, C. W., Minato, H. & Costerton, J. W. (1991). Microbial ecology and physiology of feed degradation within the rumen. In *Physiological Aspects of Digestion and Metabolism in Ruminants*, pp. 595–624. Edited by T. Tsuda, Y. Sasaki & R. Kawashima. Toronto, Ontario, Canada: Academic Press.
- Coleman, A. W. (2003). ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet* **19**, 370–375.
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755–763.
- Eddy, S. R. (2001). HMMER: Profile hidden Markov models for biological sequence analysis (<http://hmmer.wustl.edu/>).
- Fliegerova, K., Hodrova, B. & Voigt, K. (2004). Classical and molecular approaches as a powerful tool for the characterization of rumen polycentric fungi. *Folia Microbiol (Praha)* **49**, 157–164.
- Hausner, G., Inglis, G., Yanke, L. J., Kawchuk, L. M. & McAllister, T. A. (2000). Analysis of restriction fragment length polymorphisms in the ribosomal DNA of a selection of anaerobic chytrids. *Can J Bot* **78**, 917–927.
- Hulo, N., Sigrist, C. J. A., Le Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P. & Bairoch, A. (2004). Recent improvements to the PROSITE database. *Nucleic Acids Res* **32**, D134–D137.
- Joblin, K. N., Matsui, H., Naylor, G. E. & Ushida, K. (2002). Degradation of fresh ryegrass by methanogenic co-cultures of ruminal fungi grown in the presence or absence of *Fibrobacter succinogenes*. *Curr Microbiol* **45**, 46–53.
- Lee, S. S., Ha, J. K. & Cheng, K. J. (2000). Relative contributions of bacteria, protozoa, and fungi to in vitro degradation of orchard grass cell walls and their interactions. *Appl Environ Microbiol* **66**, 3807–3813.
- Liao, D. (1999). Concerted evolution: molecular mechanisms and biological implications. *Am J Hum Genet* **64**, 24–30.
- Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**, 911–940.
- Mennim, G. (1997). *The application of ribosomal DNA sequence data and other molecular approaches to the study of anaerobic gut fungi*.

PhD thesis. Faculty of Science and Engineering, University of Manchester, UK.

Mulder, N. J., Apweiler, R., Attwood, T. K. & 34 other authors (2003). The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res* **31**, 315–318.

Munn, E. A., Orpin, C. G. & Greenwood, C. A. (1988). The ultrastructure and possible relationships of four obligate anaerobic chytridiomycete fungi from the rumen of sheep. *Biosystems* **22**, 67–81.

O'Donnel, K. & Cigelnik, E. (1997). Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol Phylogenet Evol* **7**, 103–116.

Orpin, C. G. (1975). Studies on the rumen flagellate *Neocallimastix frontalis*. *J Gen Microbiol* **98**, 423–430.

Ozkose, E. (2001). *Morphology and molecular ecology of anaerobic fungi*. PhD thesis. University of Wales, Aberystwyth.

Ozkose, E., Thomas, B. J., Davies, D. R., Griffith, G. W. & Theodorou, M. K. (2001). *Cyllamyces aberensis* gen.nov. sp.nov., a new anaerobic gut fungus with branched sporangiophores isolated from cattle. *Can J Bot* **79**, 666–673.

Pearl, F. M. G., Bennett, C. F., Bray, J. E., Harrison, A. P., Martin, N., Shepherd, A., Sillitoe, I., Thornton, J. & Orengo, C. A. (2003). The CATH database: an extended protein family resource for structural and functional genomics. *Nucleic Acids Res* **31**, 452–455.

Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* **23**, 4878–4884.

Ranjard, L., Poly, F., Lata, J. C., Mougel, C., Thioulouse, J. & Nazaret, S. (2001). Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability. *Appl Environ Microbiol* **67**, 4479–4487.

Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**, 5857–5864.

Theodorou, M. K., Mennim, G., Davies, D., Zhu, W.-Y., Trinci, A. P. J. & Brookman, J. (1996). Anaerobic fungi in the digestive tract of mammalian herbivores and their potential for exploitation. *Proc Natl Acad Sci U S A* **55**, 913–926.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**, 4876–4882.

Vainio, E. J. & Hantula, J. (2000). Direct analysis of wood-inhabiting fungi using denaturing gradient gel electrophoresis of amplified ribosomal DNA. *Mycol Res* **104**, 927–936.

van Nues, R. W., Rientjes, J. M. J., van der Sande, C. A. F. M., Zerp, S. F., Sluiter, C., Venema, J., Planta, R. J. & Raue, H. A. (1994). Separate structural elements within internal transcribed spacer 1 of *Saccharomyces cerevisiae* precursor ribosomal RNA direct the formation of 17S and 26S rRNA. *Nucleic Acids Res* **22**, 912–919.

Vogler, A. P. & DeSalle, R. (1994). Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. *Mol Biol Evol* **11**, 393–405.

Ward, D. M., Weller, R. & Bateson, M. M. (1990). 16S ribosomal-RNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**, 63–65.

Wendel, J. F., Schnabel, A. & Seelanan, T. (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc Natl Acad Sci U S A* **92**, 280–284.

Zuker, M., Mathews, D. H. & Turner, D. H. (1999). Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA Biochemistry and Biotechnology*, pp. 11–43. Edited by J. Barciszewski & B. F. C. Clark. NATO ASI Series. Dordrecht: Kluwer.