

ORIGINAL PAPER

DNA Barcoding of *Blastocystis*

Stephanie M. Scicluna², Blessing Tawari, and C. Graham Clark¹

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Submitted October 11, 2005; Accepted December 4, 2005
 Monitoring Editor: Michael Melkonian

We have developed a simple method for subtyping the intestinal protistan parasite *Blastocystis* using an approach equivalent to DNA barcoding in animals. Amplification of a 600 bp region of the small subunit ribosomal RNA gene followed by single primer sequencing of the PCR product provides enough data to assign isolates to specific subtypes unambiguously. We believe that this approach will prove useful in future epidemiological studies.

© 2005 Elsevier GmbH. All rights reserved.

Key words: *Blastocystis*; phylogeny; ribosomal RNA; subtyping.

Introduction

'DNA barcoding' is the term used to describe a method proposed for producing a unique identifier for all living species (Hebert et al. 2003). To date this effort has focussed on identification of animals and the use of a 648 bp segment near the 5' end of the mitochondrial gene encoding cytochrome C oxidase subunit I (COI-5'). The rationale for using this particular sequence is four-fold: mitochondrial DNA is present in multiple copies per cell making PCR amplification of the target DNA relatively easy; there already exists a large database of mitochondrial sequences against which novel sequences can be compared; the gene is usually well conserved within species but sufficiently variable that interspecific differences can be detected easily; and the gene

segment is of a size that can be easily amplified by PCR then sequenced in a single reaction. The resulting sequences can be used for both species identification and for limited phylogenetic analysis to aid in placing those organisms not already represented in the database. This approach has already led to a number of publications representing studies of a variety of animal groups (Hebert et al. 2004; Hogg and Hebert 2004).

The proponents of this approach recognise, however, that this gene is not suitable for studying all groups of organisms (<http://www.barcodinglife.org/>). For example, rates of evolution of COI are too slow to allow this gene to be useful in plant studies and other sequences have been proposed (Kress et al. 2005). In addition, anaerobic eukaryotes do not normally have a mitochondrial genome and therefore lack the target gene altogether. For DNA barcoding in such organisms an alternative target is required, a gene with the requisite characteristics but that is present in the organism of interest.

The protist *Blastocystis hominis* is the only stramenopile parasitic in humans (Silberman

¹Corresponding author;

Fax: +44 20 7636 8739

e-mail graham.clark@lshtm.ac.uk (C.G. Clark)

²Present address: Department of Microbiology, Hammersmith Hospital NHS Trust, Du Cane Road, London W12 0HS, GB, UK.

et al. 1996). It is also not really a true species but rather a species complex consisting of many different variants that can differ by over 7% in their small subunit ribosomal RNA gene (SSU-rDNA) sequence (Clark 1997). Other animals can be infected with the same range of variants making the taxonomy of this group uncertain at best. In the remainder of this article the organisms will be referred to for simplicity as *Blastocystis* irrespective of its host origin.

Blastocystis is of uncertain pathogenicity in humans (Stenzel and Boreham 1996). Most population-based studies find no difference between rates of infection in symptomatic and asymptomatic individuals. In contrast, in individual infections a strong case can be made for the organism being the cause of disease. The reconciling of these two disparate views may lie in understanding more about the different genetic variants. If only one type of *Blastocystis* causes disease in humans, for example, this would not have been uncovered by any population survey to date as they have relied solely on microscopy and/or growth in culture for species identification.

At least 7 major subtypes of *Blastocystis* have been isolated from humans and other mammals (Noël et al. 2005; Yoshikawa et al. 2004). These can be considered to be molecular operational taxonomic units (MOTU; Blaxter 2004). Most of the methods used to detect these subtypes have relied on finding differences in the nuclear SSU-rDNA, either indirectly using restriction enzyme patterns (Clark 1997; Hoevers et al. 2000; Rivera and Tan 2005; Snowden et al. 2000) or by directly sequencing the gene (Abe 2004; Arisue et al. 2003; Noël et al. 2005; Thathaisong et al. 2003). Restriction enzyme patterns may miss or over-emphasise differences in sequence and are quite labour intensive to produce. However, for epidemiological studies sequencing of the complete gene is unrealistic. It is ca. 1800 bp in length and in the most recent publication on diversity the amplified gene was also cloned before being sequenced (Noël et al. 2005). Realistically, a rapid 'DNA barcoding' approach needs to be developed before evidence for a link between subtype and the outcome of infection can be studied in large numbers of *Blastocystis* isolates.

In the present study we have investigated the usefulness of a 600 bp segment of the *Blastocystis* SSU-rDNA for detecting the established subtypes of *Blastocystis*. We show that this smaller region of the gene can be sequenced in one reaction and that this allows all the subtypes to be identified. Most of the established subtype relationships can

also be recovered in phylogenetic analyses using this region alone. We believe that, using this approach, large-scale epidemiological surveys can now be undertaken with the aim of settling the question of *Blastocystis* pathogenicity in humans once and for all.

Results

Primer design and testing

A selection of 25 *Blastocystis* SSU-rDNA sequences representing all known mammalian subtypes were aligned, including two or more sequences of each subtype and all 'unclassified' sequences. SSU-rDNA sequences representing a range of other human intestinal protozoan parasites were also included. Regions were identified in which conservation across all *Blastocystis* subtypes was seen but that were not well conserved in the other human parasites. One such region is at around 600 bases from the 5' end of the *Blastocystis* SSU-rDNA. A primer complementary to this sequence (BhRDr) was designed and used subsequently in all PCR amplification and sequencing. Although its sequence does not match that of some other eukaryotes (especially stramenopiles) it should not amplify the SSU-rDNA from other organisms likely to be found in the mammalian intestinal tract.

The primer was tested using *Blastocystis* DNA extracted from short-term cultures established as part of the diagnostic routine at two laboratories. Amplification of the 5' 600 bp region of the *Blastocystis* SSU-rDNA was efficient and gave little background. PCR product yield was high and no additional bands were seen. Subsequently, the lack of background was confirmed by our ability to sequence the amplicon directly and obtain unambiguous sequence in most instances.

Phylogenetic testing

All of the unique *Blastocystis* rDNA sequences included in the analysis of Noël et al. (2005) were edited to include only the region under study here and a Bayesian phylogenetic analysis was undertaken. The aligned region was 626 bp in length and had 235 unique site patterns. Bayesian analysis of the 5' one-third of the gene produces a tree similar to that of Noël et al. (2005) but supported by lower posterior probabilities. However, the sequences form clearly defined clades that are the same as those seen with the complete

gene (Fig. 1). Minor differences in the branch order among subgroups were observed. Clustering of subtypes I+II is recovered in both trees as is a specific relationship between subtypes VI and VII, but while the latter pair emerges near the base of our tree it is specifically related to I+II+V in the tree of Noël et al. (2005). Nevertheless, strong bootstrap support was observed for all subtype clades and for two clusters that were identified by Noël et al. as of 'uncertain classification'. For simplicity, we have named these clusters IVa and VIa to indicate their closest well-supported subtype relationship in the tree of Noël et al. (2005) (Fig. 1). In conclusion, because the same clustering of sequences into subtype clades was observed, sufficient phylogenetic signal clearly exists in the 5' one-third of the gene to allow accurate subtype attribution of new sequences.

One significant difference observed in our analysis was with an outlying pig isolate (GenBank Accession number: AF538348). This clusters strongly with subtype V sequences in our tree but is basal to I+II in that of Noël et al. (Fig. 1). Further inspection of this sequence indicates that it is likely to be a chimaera, as although the 5' end clusters with subtype V the 3' end clusters strongly with subtype I (data not shown). Until the correct origin of this sequence is confirmed we have designated it as subtype 'X'.

DNA barcoding

To investigate the stability of the subtype clades when a large number of new sequences are added to the analysis, we sequenced the 5' one-third of the SSU-rDNA from *Blastocystis* isolates obtained from two diagnostic laboratories. DNA was purified from culture lysates and PCR products were obtained using primers RD5 and BhrRDr. Products were gel purified and sequenced with both primers to evaluate the relative efficacy and quality of the sequence produced from each of 53 *Blastocystis* samples (Table 1A). Retrospectively it was found that a number of these had been isolated from non-human primates; however no monkey-specific subtypes have been described to date. DNA barcoding protocols use only a single primer for sequencing in order to reduce time and expense. Of the two primers we used the quality and reliability of sequences obtained using BhrRDr were consistently greater than those generated using RD5. Peaks were higher and background was lower with BhrRDr when traces for the two strands were compared for the same sample. Such differences between

primers are quite commonly observed. We therefore recommend BhrRDr for *Blastocystis* DNA barcoding. Sequences obtained using either primer cover all the polymorphic positions in the region of the gene under study so that no information is lost by using only one primer for sequencing. An additional 20 sequences were generated using primer BhrRDr alone (Table 1B).

One DNA sample was clearly derived from a mixed infection containing two subtypes of *Blastocystis* and was excluded from further analysis; the sequence was unreadable in the polymorphic regions. For all other samples, once again assignment of strains to subtypes was straightforward. In some samples certain positions in the sequence were unresolved in the traces with two bases appearing to be present in equal amounts, similar to what is seen when there is heterozygosity at single copy gene loci. This was equally true of those samples sequenced on both strands and those sequenced on one. Two explanations are possible—a coinfection with two variants within the same subtype could be present, or alternatively variation could exist between SSU-rDNA copies within the same isolate. It is not possible at present to differentiate the two possibilities.

Phylogenetic analysis

All 72 new sequences were aligned with a subset of 30 sequences selected to represent the full range of variation observed in the study by Noël et al. (Fig. 1). Phylogenetic analysis was performed using Bayesian, Maximum Likelihood and Distance approaches. All of the subtype groupings and a number of the minor clusters observed in Fig. 1 were recovered with strong support in all three analyses (Fig. 2). A majority of the new sequences belonged to subtypes III and IV. All sequences were able to be attributed with confidence to a previously established specific subtype.

Discussion

It is clear that the complete SSU-rDNA sequence of *Blastocystis* is not required for assignment of an isolate to its appropriate subtype. We believe that the amplicon described here when sequenced with primer BhrRDr will provide sufficient information for accurate subtyping of new samples in a manner equivalent to the DNA barcoding approach already described for animal species. Although not formally described as species, the subtypes of *Blastocystis* are OTUs with a substantial degree of divergence,

Table 1. Origin of *Blastocystis* samples used for sequencing in this study. **(A)** Samples for which the 600 bp region was sequenced on both strands. **(B)** Samples for which the 600 bp region was used for single primer sequencing.

Identification no.	Host	Subtype	Accession no.
(A)			
01/893	Woolly monkey	I	DQ232800
01/905	Woolly monkey	I	DQ232807
JJW16	Human	I	DQ232777
01/676	Woolly monkey	II	DQ232806
01/970	Human [#]	II	DQ232814
02/309	Human	II	DQ232805
02/521	Woolly monkey	II	DQ232799
02/1024	Human	II	DQ232782
05/387	Human	II	DQ232824
01/898	Human	II	DQ232802
02/532	Human	II	DQ232794
JJW5	Human	II	DQ232775
JJW18	Human	II	DQ232778
02/027	Unidentified primate	III	DQ232788
02/028	Unidentified primate	III	DQ232789
02/029	Unidentified primate	III	DQ232790
02/030	Unidentified primate	III	DQ232791
02/033	Unidentified primate	III	DQ232792
02/444	Stump tailed macaque	III	DQ232797
02/1002	Woolly monkey	III	DQ232785
02/682	Human [#]	III	DQ232823
02/284	Human	III	DQ232803
02/311	Human	III	DQ232798
02/440	Human	III	DQ232809
02/550	Human	III	DQ232801
02/604	Human	III	DQ232817
02/615	Human	III	DQ232804
02/638	Human	III	DQ232819
02/640	Human	III	DQ232820
02/668	Human	III	DQ232822
02/1086	Human	III	DQ232780
05/266	Human	III	DQ232825
05/386	Human	III	DQ232827
02/796	Human	III	DQ232810
02/039	Human	III	DQ232793
02/826	Human	III	DQ232811
02/1020	Human	III	DQ232784
JJW19	Human	III	DQ232779
02/919	Human	IV	DQ232787
02/1017	Human	IV	DQ232781
02/636	Human	IV	DQ232818
05/379	Human	IV	DQ232826
02/797	Human	IV	DQ232812
02/813	Human	IV	DQ232815
02/949	Human	IV	DQ232786
02/951	Human	IV	DQ232816
02/127	Human	IV	DQ232813
02/825	Human	IV	DQ232808
JJW6	Human	IV	DQ232776
02/393	Woolly monkey	IVa	DQ232795

Table 1. (continued)

Identification no.	Host	Subtype	Accession no.
02/517	Woolly monkey	IVa	DQ232796
02/1003	Woolly monkey	IVa	DQ232783
02/642	Human	VII	DQ232821
(B)			
00/267	Human [#]	I	DQ232832
00/469	Human	I	DQ232843
00/396	Human	II	DQ232829
00/1005	Unidentified primate	II	DQ232845
00/117	Human	III	DQ232844
00/299	Human [#]	III	DQ232834
00/389	Human	III	DQ232840
00/440	Human	III	DQ232839
00/368	Human	IV	DQ232837
00/472	Human	IV	DQ232846
00/369	Human [#]	IV	DQ232831
00/501	Human	IV	DQ232835
00/400	Human	IV	DQ232838
00/387	Human	IV	DQ232841
00/266	Human [#]	IVa	DQ232833
00/266A	Human [#]	IVa	DQ232830
00/392	Human	IVa	DQ232828
00/1009	Unidentified primate	IVa	DQ232842
Mabel	Pig	V	DQ232836
00/136A	Human	Mixed	—

[#]Monkey handler; Unidentified primate = species information not supplied. Isolates with the prefix JJW were provided by Jeffrey J. Windsor (Bronglais Hospital), the rest are from the Diagnostic Parasitology Laboratory (London School of Hygiene and Tropical Medicine).

and it is likely that they will be elevated to specific status at some time in the future.

Phylogenetic analysis of the sort described above is not necessary for correct assignment of sequences to subtypes. Certainly, less computationally complex programmes produce the same clusters seen in [Figures 1 and 2](#). Indeed, in most cases visual inspection of an alignment containing new sequences and a mixture of sequences of known subtype will allow correct subtype identification to be made. This greatly simplifies the process. However, it is likely that subtyping of some sequences by alignment alone will be ambiguous; in this case phylogenetic analysis may well resolve the problem.

Although we have tested a limited number of samples to date, we have uncovered one clear case of a mixed subtype infection. How prevalent this situation will prove to be is at this stage unknown, but it appears based on the current data to be relatively infrequent. Unless our samples are not representative of *Blastocystis* infections across the world it is unlikely that mixed infections

would greatly hamper epidemiological investigations. The need for cloning of PCR products prior to sequencing would eliminate the advantages in time and expense that DNA barcoding offers. However, if apparently new subtypes are identified in the future it will still be possible for the entire gene to be sequenced and a more complete phylogenetic analysis undertaken.

The aim of this work was not to re-examine the phylogenetic relationships among *Blastocystis* but rather to develop a more rapid and accurate method for classifying new isolates into subtypes. The subtype identification method described here will, we believe, allow much larger epidemiological studies to be undertaken in the future as it dramatically reduces the time and expense involved when compared to methods in use at the present time.

Methods

DNA samples: Most *Blastocystis* culture lysates were provided by the staff of the Diagnostic

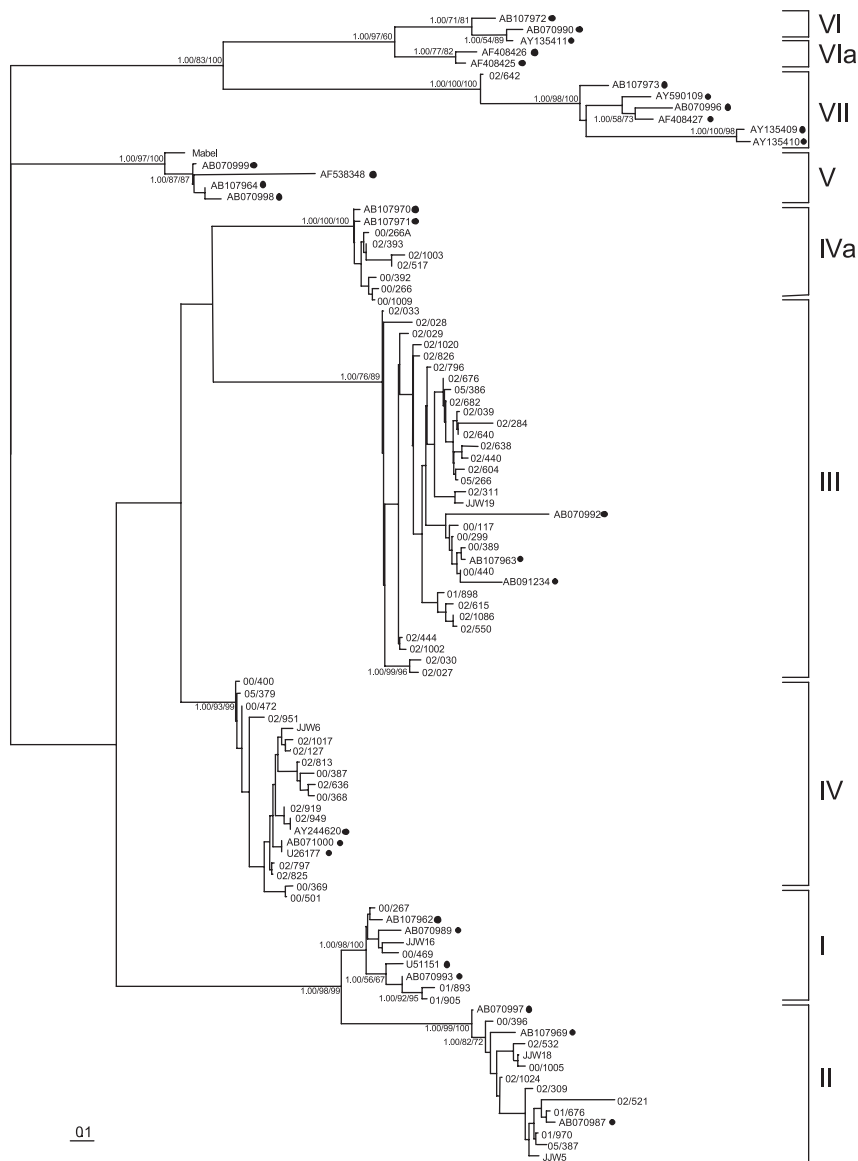


Figure 2. Unrooted tree showing relationships among 102 *Blastocystis* sequences. The numbers adjacent to branchpoints indicate the Bayesian posterior probabilities and percentage of support for the clade based on bootstrap resampling in maximum likelihood and distance analyses (from left to right, respectively). Only where all three analyses agreed on topology and support (greater than 50% in the case of maximum likelihood and distance) are the numbers given. Those sequences selected from the data set of Noël et al are indicated by ● to aid identification and the Roman numerals indicate the subtypes to which the sequences are assigned.

Parasitology Laboratory, London School of Hygiene and Tropical Medicine, while a few were provided by Jeffrey J. Windsor, Bronglais General Hospital, Aberystwyth, Wales. They all derive from stool samples sent for ova and parasite diagnosis (Table 1). Such samples are routinely inoculated into Robinson's medium at both institutions and those cultures in which *Blastocystis*

grew were harvested by centrifugation and lysed in 0.25% SDS/0.1M EDTA pH 8 before being discontinued. DNA was purified by the modified cetyl trimethylammonium bromide (CTAB) extraction method (Ali et al. 2005; Clark and Diamond 1991). Briefly, lysates were treated with Proteinase K before incubation with CTAB under high salt conditions. The CTAB—carbohydrate complex

was precipitated with chloroform and the aqueous phase extracted with phenol–chloroform–isoamyl alcohol before precipitation of the DNA with ethanol. After washing in 70% ethanol and resuspension in water, the DNA was passed over a Sephacryl S200 spin column (GE Healthcare, UK) before being used in PCR.

Primers and sequencing: The *Blastocystis*-specific primer BhRDr (GAGCTTTT-TAACTGCAACAACG) and the broad-specificity eukaryote-specific primer RD5 (ATCTGGTT-GATCCTGCCAGT; Clark 1997) were used in a standard PCR reaction with Taq DNA polymerase (BIOTAQ, Bionline, UK). Amplification conditions consisted of 30 cycles of 1 min each at 94, 59 and 72 °C, with an additional 2 min final extension. PCR products were separated in 1.2% agarose gels and purified using the Qiaquick[®] gel extraction kit (Qiagen) according to the manufacturer's instructions. Purified products were sequenced using the amplification primer(s) and ABI BigDye[™] sequencing kit version 3.1 (Applied Biosystems, Inc.), and analysed on an ABI 3730 sequencer. Sequences were deposited in GenBank with the accession numbers DQ232775–DQ232846.

Phylogenetic analysis: Sequences downloaded from GenBank as well as those generated in this study were aligned using Multalin (Corpet 1988) and the alignments edited manually (available on request). Only the unique sequences from the study by Noël et al. were included.

Bayesian analysis used MrBayes 3.0 (Huelsenbeck and Ronquist 2001) with 4 MCMC strands, 100,000 generations and an initial burn in of 1000. Four categories of among site rate variation were used and trees were sampled every 10 generations. In the reanalysis of the Noël data set, *Proteromonas lacertae* was used to root the tree.

Maximum likelihood and distance analyses used the phylogenetic software package PHYLIP v. 3.6 (Felsenstein 1989) and a model with 4 categories of among site rate variation and the proportion of invariant sites. These parameters ($\alpha = 0.16$; pinvar = 0.00) and the transition/transversion ratio (1.10) were estimated using Tree-Puzzle 5.0 (Schmidt et al. 2002). The alignment file was resampled 100 times (SEQBOOT) before analysis using DNAML for maximum likelihood analysis. For distance analysis the alignment file was resampled 1000 times before analysis using DNADIST (F84 model) and NEIGHBOR. The majority rule consensus trees were produced using CONSENSE. *Blastocystis* subtype nomenclature follows that of Yoshikawa et al. (2004) (also used by Noël et al. (2005)).

Acknowledgements

We thank John E. Williams and the staff of the Diagnostic Parasitology Laboratory (London School of Hygiene and Tropical Medicine) and Jeff Windsor (Bronglais Hospital) for providing the samples for analysis, and Konstantina Eleme and Alida Javier Molina for doing some of the DNA extractions. The majority of this work formed part of the thesis submitted as part of the M.Sc. in Medical Microbiology by SMS.

References

- Abe N** (2004) Molecular and phylogenetic analysis of *Blastocystis* isolates from various hosts. *Vet Parasitol* **120**: 235–242
- Ali IKM, Zaki M, Clark CG** (2005) A genotyping method for *Entamoeba histolytica* based on PCR amplification of tRNA gene-linked short tandem repeats. *J Clin Microbiol* **43**: 5842–5847
- Arisue N, Hashimoto T, Yoshikawa H** (2003) Sequence heterogeneity of the small subunit ribosomal RNA genes among *Blastocystis* isolates. *Parasitology* **126**: 1–9
- Blaxter ML** (2004) The promise of a DNA taxonomy. *Philos Trans R Soc London B* **359**: 669–679
- Clark CG** (1997) Extensive genetic diversity in *Blastocystis hominis*. *Mol Biochem Parasitol* **87**: 79–83
- Clark CG, Diamond LS** (1991) The Laredo strain and other *Entamoeba histolytica*-like amoebae are *Entamoeba moshkovskii*. *Mol Biochem Parasitol* **46**: 11–18
- Corpet F** (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**: 10881–10890
- Felsenstein J** (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164–166
- Hebert PD, Cywinska A, Ball SL, de Waard JR** (2003) Biological identifications through DNA barcodes. *Proc R Soc London B* **270**: 313–321
- Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W** (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci USA* **101**: 14812–14817
- Hoevers J, Holman P, Logan K, Hommel M, Ashford R, Snowden K** (2000) Restriction-fragment-length polymorphism analysis of small-subunit

rRNA genes of *Blastocystis hominis* isolates from geographically diverse human hosts. *Parasitol Res* **86**: 57–61

Hogg ID, Hebert PDN (2004) Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. *Can J Zool* **82**: 749–754

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* **102**: 8369–8374

Noël C, Dufernez F, Gerbod D, Edgcomb VP, Delgado-Viscogliosi P, Ho LC, Singh M, Wintjens R, Sogin ML, Capron M, Pierce R, Zenner L, Viscogliosi E (2005) Molecular phylogenies of *Blastocystis* isolates from different hosts: implications for genetic diversity, identification of species, and zoonosis. *J Clin Microbiol* **43**: 348–355

Rivera WL, Tan MA (2005) Molecular characterization of *Blastocystis* isolates in the Philippines by riboprinting. *Parasitol Res* **96**: 253–257

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**: 502–504

Silberman JD, Sogin ML, Leipe DD, Clark CG (1996) Human parasite finds taxonomic home. *Nature* **380**: 398

Snowden K, Logan K, Blozinski C, Hoervers J, Holman P (2000) Restriction fragment-length polymorphism analysis of small-subunit rRNA genes of *Blastocystis* isolates from animal hosts. *Parasitol Res* **86**: 62–66

Stenzel DJ, Boreham PFL (1996) *Blastocystis hominis* revisited. *Clin Microbiol Rev* **9**: 563–584

Thathaisong U, Worapong J, Mungthin M, Tan-Ariya P, Viputtigul K, Sudatis A, Noonai A, Leelayoova S (2003) *Blastocystis* isolates from a pig and a horse are closely related to *Blastocystis hominis*. *J Clin Microbiol* **41**: 967–975

Yoshikawa H, Morimoto K, Wu Z, Singh M, Hashimoto T (2004) Problems in speciation in the genus *Blastocystis*. *Trends Parasitol* **20**: 251–255

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®