Infection, Genetics and Evolution 12 (2012) 263–273

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Levels of genetic diversity vary dramatically between *Blastocystis* subtypes

C. Rune Stensvold [a,*], Mohammed Alfellani [b], C. Graham Clark [b]

[a] Department of Microbiological Diagnostics, Statens Serum Institut, Orestads Boulevard 5, DK-2300 Copenhagen S, Denmark
[b] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT London, United Kingdom

ARTICLE INFO

Article history:
Received 16 September 2011
Received in revised form 5 November 2011
Accepted 7 November 2011
Available online 17 November 2011

Keywords:
Small subunit ribosomal DNA
Mitochondrial DNA
Molecular epidemiology
MLST
Genetic diversity
Zoonosis

ABSTRACT

Blastocystis is a common single-celled parasite of humans and other animals comprising at least 13 genetically distinct small subunit ribosomal RNA lineages (subtypes (STs)). In this study we investigated intra-subtype genetic diversity and host specificity of two of the most common subtypes in humans, namely ST3 and ST4, by analysing and comparing over 400 complete and partial nuclear SSU-rDNAs and data from multilocus sequence typing (MLST) of the mitochondrion-like organelle (MLO) genome of 132 samples. Inferences from phylogenetic analyses of nuclear SSU-rDNA and concatenated MLST sequences were compatible.

Human ST3 infections were restricted to one of four identified MLO clades except where exposure to non-human primates had occurred. This suggests relatively high host specificity within ST3, that human ST3 infections are caused predominantly by human-to-human transmission, and that human strains falling into other clades are almost certainly the result of zoonotic transmission. ST4 from humans belonged almost exclusively to one of two SSU-rDNA clades, and only five MLST sequence types were found among 50 ST4s belonging to Clade 1 (discriminatory index: 0.41) compared to 58 MLST sequence types among 81 ST3s (discriminatory index: 0.99).

The remarkable differences in intra-subtype genetic variability suggest that ST4 has a more recent history of colonising humans than ST3. This is congruent with the apparently restricted geographical distribution of ST4 relative to ST3. The implications of this observation are unclear, however, and the population structure and distribution of ST4 should be subject to further scrutiny in view of the fact ST4 is being increasingly linked with intestinal disease.

1. Introduction

Blastocystis is a parasitic protist and the most common non-yeast eukaryotic organism in the intestinal tract of humans and many other animals (Stenzel and Boreham, 1996; Stensvold et al., 2009c). The genus *Blastocystis* can be divided into at least 13 small subunit ribosomal RNA (SSU-rDNA) lineages, termed subtypes (STs), which are genetically so distinct that they could be considered separate species (Stensvold et al., 2007b, 2009a,c; Parkar et al., 2010).

Humans are mainly colonised by ST1–ST4 and rarely by ST5–ST9; ST10–ST13 have not been found in humans to date (Stensvold et al., 2009a; Parkar et al., 2010). The public health significance of *Blastocystis* and the potential for zoonotic transmission are subjects currently under intense scrutiny (Parkar et al., 2007; Stensvold et al., 2009a–c), and it is possible that differences in clinical outcome of *Blastocystis* infection are related to genetic differences on the subtype- or strain-level (Stensvold et al., 2009c, 2011).

In addition to humans, ST3 is also found in a variety of non-human hosts, including non-human primates (NHPs) and ungulates (Stensvold et al., 2009a; Alfellani et al. (unpublished data), whereas ST4 appears to be restricted to primates and rodents. Moreover, ST3 appears to have a cosmopolitan distribution, whereas ST4 may be restricted primarily to Europe and North America (Malheiros et al., in press; Forsell et al., in press).

The molecular epidemiology of *Blastocystis* is incompletely known and novel subtypes are still being discovered (Stensvold et al., 2009a; Parkar et al., 2010; Alfellani et al. (unpublished data). Very little is known about genetic variation in *Blastocystis* except for the nuclear SSU-rDNA, and no investigations of diversity within subtypes have been reported.

Multilocus sequence typing (MLST) has been central to many studies seeking to unravel the molecular epidemiology of pathogenic microorganisms (Sullivan et al., 2005). Although initially developed for studying haploid organisms (specifically bacteria), diploid sequence types have been described for *Trypanosoma cruzi* (Yeo et al., 2011), and some fungi, such as *Aspergillus fumigatus* and nosocomial *Candida albicans* strains (Bain et al., 2007; Bougnoux et al., 2002), but in diploid organisms the MLST alleles can be difficult to interpret. Therefore, use of a sequence like the mitochondrial DNA

* Corresponding author. Tel.: +45 3268 3604; fax: +45 3268 3033.
E-mail address: RUN@ssi.dk (C.R. Stensvold).

1567-1348/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2011.11.002

(mtDNA) is advantageous as it is equivalent to a haploid genome. *Blastocystis* possesses mitochondrion-like organelles (MLOs) rather than classical mitochondria (Lantsman et al., 2008; Stechmann et al., 2008; Wawrzyniak et al., 2008), and MLO genomes of three subtypes (ST1, ST4 and ST7) have already been published (Pérez-Brocal and Clark, 2008; Wawrzyniak et al., 2008).

In this study, we have developed a MLST scheme for ST3 and ST4 based on MLO genome sequence data and applied it to 132 samples from these subtypes. We have analysed nuclear SSU-rDNAs from GenBank and from our laboratories and compared them with sequence type data obtained by MLST. The results reveal remarkable differences in diversity within subtypes and suggest interesting conclusions on host specificity.

2. Materials and methods

Nuclear SSU-rDNA sequences obtained from GenBank, previous studies and new samples were analysed in order to detect intra-subtype nucleotide sequence variation. Candidate *Blastocystis* isolates for complete sequencing of MLO genomes were selected based on these results. All samples included in the present study are given in Tables 1 and 2 with references.

2.1. GenBank SSU-rDNA sequences and unpublished sequences from previous studies

Complete and partial ST3 and ST4 SSU-rDNAs were downloaded from GenBank. Hosts were recorded for each sequence; where no host was indicated in the entry or associated publication, it was assumed that the sequence was from a human sample. Since hundreds of ST3 sequences have been deposited in GenBank, only sequences that cover the barcode region (Scicluna et al., 2006) were included.

Unpublished sequences from completed or ongoing studies by Stensvold et al. (2011), Alfellani et al. (unpublished data), Rene et al. (2009), Forsell et al. (in press), and Onuoha et al. (unpublished) were also included (Tables 1 and 2).

2.2. Original SSU-rDNA sequences

Genomic DNAs from human and NHP faecal samples or cultures were mainly barcoded (Scicluna et al., 2006), but some sequences were obtained using primers targeting other parts of the gene. Samples from human hosts were obtained mainly from the UK and Denmark and NHP samples were from UK zoos (Tables 1 and 2). Consistent information on the clinical status of patients was not available, but most of the ST4 samples were from patients attending irritable bowel syndrome clinics. *Blastocystis* cultures of samples with the prefix 'MA' and 'DMP' (Tables 1 and 2) were established and maintained according to the method described by Clark and Diamond (2002) and discontinued after use.

DNA from cultures was extracted as follows: cells were harvested by centrifugation, washed 3 times in phosphate-buffered saline, and lysed in 0.25% SDS/0.1 M EDTA pH 8. DNA was purified from lysates using the Puregene Core Kit A (QIAGEN, Hilden, Germany) according to the manufacturer's instructions (including a proteinase K step). DNA was extracted directly from stool using the QIAamp DNA Stool Mini Kit (QIAGEN, Hilden, Germany) according to the recommendations of the manufacturer.

Five SSU-rDNA sequences were submitted to GenBank: DMP/04-872 (HQ909898), DMP/08-1040 (HQ909890), DMP/08-1043 (HQ909891), DMP/10-212 (JN682513), and GP_KVL (JN682512).

2.3. Alignment and analysis of SSU-rDNA sequences

For ST3, the 5′-terminal 500–600 base pair (bp) 'barcode' region (Scicluna et al., 2006) was analysed for SNPs in an alignment of 217 sequences, of which 171 were from humans, 44 from NHPs and two from non-primate hosts (Table 1). For ST4, all available positions in 183 complete and partial sequences were analysed in a similar way; 170 sequences were from humans, three from NHPs (lemurs), seven from rats, and three from guinea pigs (Table 2). Alignments were generated using MultAlin, an alignment program with hierarchical clustering (Corpet, 1988) (Supplementary Figs. 1 and 2).

2.4. Mitochondrion-like organelle (MLO) genome sequences

For ST3, two human isolates (DMP/IH:478 and DMP/08-326) and a NHP isolate with a distinct SSU-rDNA (DMP/08-1043), all still available in culture, were chosen for complete MLO genome sequencing. For ST4, the human sample DMP/10-212 was chosen for whole MLO genome sequencing, since it represented the rarer Clade 2; DMP/02-328 from a human representing Clade 1 was previously sequenced by Pérez-Brocal and Clark (2008) (for an introduction to the clade system in ST4, please refer to Stensvold et al., 2011).

All four MLO genomes were assembled using the Staden software package (Staden et al., 2000). Several primer pairs used in a previous study of the ST1 and ST4 MLO genomes (Pérez-Brocal and Clark, 2008) were re-used to obtain partial sequences of the ST3 genome, and the remaining sequence was covered by "primer walking". PCR conditions (Biomix, Bioline, London, UK) and sequencing procedures were similar to those described previously (Pérez-Brocal and Clark, 2008). Briefly, the amplification profile comprised an initial denaturing step at 94 °C for 2 min, followed by 10 cycles of touch-down PCR (denaturation at 94 °C for 30 s, annealing at 60 °C, decreasing by 0.5 °C per cycle, and extension at 68 °C for 4 min) followed by 20 cycles of conventional PCR, including a denaturation step at 94 °C for 30 s, an annealing step at 55 °C for 30 s, and an extension step at 68 °C for 4 min. Purification of PCR products was performed using the GeneJET™ PCR Purification Kit (Fermentas, York, UK). Sequencing was performed on an ABI3730 with ABI Prism BigDye® Terminator v3.1 reagents (Applied Biosystems, Warrington, UK) using the PCR primers as sequencing primers.

Further details and analyses of the MLO genomes of *Blastocystis* will be published separately.

2.5. Selection of gene targets for MLST schemes

Complete MLO genomes obtained for ST3 isolates DMP/IH:478, DMP/08-326 and DMP/08-1043 were aligned using MultAlin (Corpet, 1988) to locate clustered nucleotide sequence differences. Seven regions covering 300–600 bp with at least 2–3 polymorphisms in each between the three isolates were chosen for initial investigation as ST3 MLST locus candidates. Similarly, the ST4 MLO genomes of DMP/02-328 and DMP/10-212 were aligned to identify seven regions of polymorphism. For both MLST schemes, loci were chosen without regard to whether the regions to be sequenced were coding or non-coding.

2.6. MLST PCR and sequencing of gene targets from multiple samples

Initial screening and validation of MLST candidates was performed in individual tubes. The majority of the samples, however, were processed in PCR plates (Life Science Products, Scientific Laboratory Supplies, Ltd., Nottingham, UK) using the same PCR conditions indicated above. PCR products in plates were purified using the SureClean protocol (Bioline, London, UK). Bidirectional sequencing of PCR products was performed as above using the amplification primers in 96 well sequencing plates (Micro-Amp®,

Table 1
Blastocystis ST3 samples analysed in the study.

SSU clade[a]	MLO genome clade	Sequence type	Host	Geographic origin	DNA sample and/or GenBank Accession. No.	References
1	1	ST3.1	*Homo sapiens*	Sweden	JF2815	Forsell et al. (in press)
		ST3.2	*Homo sapiens*	UK	MA4, MA9, MA279	Alfellani et al. (unpublished data)
		ST3.2	*Homo sapiens*	Sweden	JF375	Forsell et al. (in press)
		ST3.3	*Homo sapiens*	Denmark	25548, 25556, 25562	Present study
		ST3.4	*Homo sapiens*	Denmark	FD5	Present study
		ST3.5	*Homo sapiens*	Denmark	44010	Present study
		ST3.6	*Homo sapiens*	UK	MA62	Alfellani et al. (unpublished data)
		ST3.7	*Homo sapiens*	UK	MA108	Alfellani et al. (unpublished data)
		ST3.8	*Homo sapiens*	UK	MA38	Alfellani et al. (unpublished data)
		ST3.9	*Homo sapiens*	UK	MA126	Alfellani et al. (unpublished data)
		ST3.10	*Homo sapiens* (NHP keeper)	UK	MA32, MA132	Alfellani et al. (unpublished data)
		ST3.11	*Homo sapiens*	UK	MA18	Alfellani et al. (unpublished data)
		ST3.12	*Homo sapiens*	Denmark	M30515	Present study
		ST3.13	*Homo sapiens*	UK	MA110, MA142	Alfellani et al. (unpublished data)
		ST3.14	*Homo sapiens*	UK	MA45	Alfellani et al. (unpublished data)
		ST3.15	*Homo sapiens*	UK	MA266, MA274	Alfellani et al. (unpublished data)
		ST3.16	*Homo sapiens*	UK	MA29	Alfellani et al. (unpublished data)
		ST3.17	*Homo sapiens*	UK	MA42, MA268	Alfellani et al. (unpublished data)
		ST3.18	*Homo sapiens*	UK	DMP/IH:478	Present study
		ST3.19	*Homo sapiens*	UK	MA118	Alfellani et al. (unpublished data)
		ST3.20	*Homo sapiens*	UK	MA41	Alfellani et al. (unpublished data)
		ST3.21	*Homo sapiens*	UK	MA262	Alfellani et al. (unpublished data)
		ST3.22	*Homo sapiens*	UK	MA278	Alfellani et al. (unpublished data)
		ST3.23	*Homo sapiens*	UK	MA66	Alfellani et al. (unpublished data)
		ST3.24	*Homo sapiens*	UK	MA86, MA98	Alfellani et al. (unpublished data)
		ST3.25	*Homo sapiens*	UK	MA92	Alfellani et al. (unpublished data)
		ST3.26	*Homo sapiens*	Denmark	51702	Present study
		ST3.27	*Nomascus gabriellae*	UK (zoo)	MA141	Alfellani et al. (unpublished data)
		ST3.28	*Homo sapiens*	UK	MA15, MA54	Alfellani et al. (unpublished data)
		ST3.29	*Homo sapiens*	UK	MA50	Alfellani et al. (unpublished data)
		ST3.30	*Homo sapiens*	UK	MA317	Alfellani et al. (unpublished data)
		ST3.31	*Homo sapiens*	UK	MA261	Alfellani et al. (unpublished data)
		ST3.32	*Homo sapiens*	UK	MA79, MA80	Alfellani et al. (unpublished data)
		ST3.32	*Homo sapiens*	Denmark	46288, 57438	Present study
		ST3.33	*Lagothrix lagotricha*	UK (zoo)	MA140	Alfellani et al. (unpublished data)
		ST3.34	*Homo sapiens*	UK	DMP/08-326	Present study
		ST3.35	*Homo sapiens*	UK	MA81, MA130	Alfellani et al. (unpublished data)
		ST3.35	*Callithrix jacchus*	UK (zoo)	MA87	Alfellani et al. (unpublished data)
		ST3.36	*Homo sapiens*	UK	MA14, MA270	Alfellani et al. (unpublished data)
		ST3.37	*Homo sapiens*	UK	MA284	Alfellani et al. (unpublished data)
		ST3.38	*Homo sapiens*	UK	MA302	Alfellani et al. (unpublished data)
		ST3.39	*Homo sapiens*	UK	MA20, MA83–4	Alfellani et al. (unpublished data)
		ST3.40	*Homo sapiens*	UK	MA30, MA134	Alfellani et al. (unpublished data)
		ST3.41	*Homo sapiens*	UK	MA75	Alfellani et al. (unpublished data)
		ST3.42	*Homo sapiens*	UK	MA282	Alfellani et al. (unpublished data)
Mixed ST3 and ST1[b]		ST3.42	*Homo sapiens*	UK	MA135	Alfellani et al. (unpublished data)
		ST3.43	*Homo sapiens*	UK	MA287	Alfellani et al. (unpublished data)
		ST3.44	*Homo sapiens*	UK	MA280	Alfellani et al. (unpublished data)
		ST3.45	*Homo sapiens* (NHP keeper)	UK	MA16	Alfellani et al. (unpublished data)
		ST3.46	*Homo sapiens*	UK	MA311	Alfellani et al. (unpublished data)
		ST3.47	*Homo sapiens*	UK	MA312	Alfellani et al. (unpublished data)
		ST3.48	*Homo sapiens*	UK	MA313	Alfellani et al. (unpublished data)
		ST3.49	*Colobus* sp.	UK (zoo)	MA291	Alfellani et al. (unpublished data)
Mixed ST3[b]		ST3.50	*Homo sapiens*	UK	MA25	Alfellani et al. (unpublished data)
2	2	ST3.51	*Erythrocebus patas*	UK (zoo)	MA299	Alfellani et al. (unpublished data)
		ST3.52	*Homo sapiens* (NHP keeper)	UK	DMP/04-872	Alfellani et al. (unpublished data)
3	3	ST3.53	*Macaca sylvanus*	UK (zoo)	MA119	Alfellani et al. (unpublished data)
		ST3.53	*Pan troglodytes*	UK (zoo)	MA65	Alfellani et al. (unpublished data)
		ST3.54	*Colobus abyssinicus*	UK (zoo)	DMP/08-1043	Alfellani et al. (unpublished data)

(*continued on next page*)

Table 1 (*continued*)

SSU clade[a]	MLO genome clade	Sequence type	Host	Geographic origin	DNA sample and/or GenBank Accession. No.	References
3	4	ST3.55	*Papio* sp.	UK (zoo)	MA257	Alfellani et al. (unpublished data)
		ST3.56			DMP/08-1040	Alfellani et al. (unpublished data)
		ST3.56	*Macaca nigra*	UK (zoo)	MA314	Alfellani et al. (unpublished data)
		ST3.57	*Homo sapiens* (NHP keeper)	UK	MA94	Alfellani et al. (in preparation)
		ST3.58	*Macaca sylvanus*	UK (zoo)	MA320	Alfellani et al. (unpublished data)
1	–	–	*Homo sapiens*	Japan	AB070986, AB070988, AB0701233–5	Arisue et al. (2003)
				Thailand	AY618268	Thathaisong et al. (unpublished)
				UK	DQ232780, DQ232793, DQ232798, DQ232801–4, DQ232811, DQ232817, DQ232819, DQ232820, DQ232822, DQ232844DQ232840, DQ232839, DQ232827, DQ232825,	Scicluna et al. (2006)
					MA2, MA217, MA219, MA300B, MA303B, MA308, MA310, MA370, MA387, MA389, MA397, MA404, MA406, MA412–3, MA418, MA421, MA430–1, MA435, MA437	Alfellani et al. (unpublished data)
				Philippines	EU4454936	Rivera (2008)
				France	AY135402	Noël et al. (2003)
					FJ666842, FJ666848–51, FJ666853, FJ666862, FJ666866, FJ666870–2, FJ666877, FJ666889, FJ666892–4, FJ666896	Souppart et al. (2009)
				Egypt	GU130223, GU130225, GU130232–4, GU130236–7, GU130243, GU130246–7	Souppart et al. (2010)
				Italy	JF274669, JF274680, JF274682–3, JF274685–6, JF274696–9	Meloni et al. (2011)
				Libya	MALI6	Alfellani et al. (in preparation)
				Vietnam	VIET-DK227, VIET-DK233–4, VIET-DK236–7	Present study
				Tanzania	JF792494	Petrášová et al. (2011)
			Homo sapiens (NHP keeper)	UK	DQ232823, DQ232834	Scicluna et al. (2006)
					MA12, MA13, MA367	Alfellani et al. (unpublished data)
			Colobus polykomos	UK (zoo)	08/1016	Alfellani et al. (unpublished data)
			Macaca fuscata	Italy (zoo)	A740	Alfellani et al. (unpublished data)
			Chlorocebus aethiops pygerythrus	Tanzania	HQ286908	Petrášová et al. (2011)
			Lagothrix lagotricha	UK (zoo)	DQ462722, DQ462724	Scicluna et al. (2006)
			Sus scrofa	Japan	AB107963	Abe (2004)
			Bos taurus	Japan	AB107965	Abe (2004)
2	–	–	*Homo sapiens*	UK	DQ232784	Scicluna et al. (2006)
				Tanzania	JF792495	Petrášová et al. (2011)
			Erythrocebus patas	UK (zoo)	MA399	Alfellani et al. (unpublished data)
			Pan troglodytes	UK (zoo)	MA116	Alfellani et al. (unpublished data)
			unidentified primate	UK (zoo)	MA372	Alfellani et al. (unpublished data)
3	–	–	*Homo sapiens*	UK	MA214	Alfellani et al. (unpublished data)
			Cercocebus torquatus	UK (zoo)	09/0805	Alfellani et al. (unpublished data)
			Macaca nigra	UK (zoo)	09/0493	Alfellani et al. (unpublished data)
			Macaca sylvanus	UK (zoo)	09/1070	Alfellani et al. (unpublished data)
				Italy (zoo)	A796	Alfellani et al. (unpublished data)
			Macaca arctoides	UK (zoo)	DQ232797	Scicluna et al. (2006)
			Macaca sp.	UK (zoo)	MA369, MA380	Alfellani et al. (unpublished data)
			Trachypithecus francoisi	UK (zoo)	09/1259	Alfellani et al. (unpublished data)
			Allenopithecus nigroviridis	UK (zoo)	09/1327, MA433	Alfellani et al. (unpublished data)
			Lagothrix lagotricha	UK (zoo)	09/1620, 09/1624, MA429	Alfellani et al. (unpublished data)
				UK (zoo)	DQ462716	Scicluna et al. (2006)
			Erythrocebus patas	UK (zoo)	MA405	Alfellani et al. (unpublished data)
			Semnopithecus sp.	UK (zoo)	MA424, MA426	Alfellani et al. (unpublished data)
			unidentified primate	UK (zoo)	DQ232788–DQ232792	Scicluna et al. (2006)
				Philippines	EU445489	Rivera (2008)

Table 1 (*continued*)

SSU clade[a]	MLO genome clade	Sequence type	Host	Geographic origin	DNA sample and/or GenBank Accession. No.	References
4	–	–	*Homo sapiens*	Japan France	AB070992 FJ666873	Arisue et al. (2003) Souppart et al. (2009)
5	–	–	*Colobus guereza*	Tanzania	HQ286916	Petrášová et al. (2011)

–, Data not available.
[a] SSU Clade based on SNP configuration at position 130–132; see text for details.
[b] Chromatograms showed mixed sequences and therefore no unambiguous sequence was available for alignment and phylogenetic analyses.

Applied Biosystems, Cheshire, UK). DNA samples processed by MLST are given in Tables 1 and 2.

For ST3, MLST sequences were submitted to GenBank in batches as follows: locus 1: HQ909892–HQ909974, locus 2: HQ909975–HQ910056, locus 3: HQ910057–HQ910138, locus 4: HQ910139–HQ910221, and locus 5: HQ909804–HQ909885 (for loci 1 and 4, 83 sequences were submitted, however, complete data across all loci were available for 81 samples only). For ST4, MLST sequences were given the following Accession Nos.: locus 1: JN682212–JN682261, locus 2: JN682262–JN682311, locus 3: JN682312–JN682361, locus 4: JN682362–JN682411, locus 5: JN682412–JN682461, and locus 6: JN682462–JN682511.

Sequences were also uploaded to a sequence typing database (Jolley and Maiden, 2010) at http://pubmlst.org/blastocystis/, which is now open for ST3 and ST4 MLST sequence submission as well as SSU-rDNA sequence submission (18S database).

2.7. Sequence editing, sequence type identification and discriminatory index

Sequences were edited, assembled and analysed using Chromas version 2.33 (Technelysium Pty. Ltd., Australia) and the Staden software package and entered into locus-specific files. Multiple sequence alignments were performed using the ClustalW algorithm with default parameters in MEGA5 (Tamura et al., 2011), and alleles calculated by haplotype analysis using DNAsp v5 (Librado and Rozas, 2009) including sites with alignment gaps. Concatenated sequences from all loci were aligned, and sequence type identification and discriminatory index were calculated by haplotype analysis as above.

2.8. Annotation of sequence types

Sequence types in MLST data sets are usually identified by the acronym "ST". However, since "ST" is a widely accepted acronym for "subtype" in the *Blastocystis* literature, we chose not to change this. Instead, we propose that sequence types are annotated by numbers following the subtype (e.g. ST3.1, ST3.2, etc.).

2.9. Phylogenetic analysis

Concatenated nucleotide sequences from all MLST loci were produced for each sample and aligned. Maximum Likelihood (ML) analysis of aligned concatenated sequences was performed in Phyml v.2.4.5 (Guindon and Gascuel, 2003), using the General Time Reversible (GTR) model of nucleotide substitution with four categories of among-site rate variation, the proportion of invariant sites estimated from the data, and 1000 bootstrap replicates. Bayesian inference analysis was carried out using MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001), the GTR model, four Markov chain Monte Carlo (MCMC) strands, and 1,000,000 generations with trees sampled every 100 generations, after which the average standard deviation of split frequencies stabilised below 0.01. A consensus tree was produced after excluding an initial burn-in of 25% of the samples (Fig. 1A).

In a similar way an alignment of partial ST3 SSU-rDNA sequences corresponding to those samples for which MLST data were available was submitted to manual editing and subsequent phylogenetic analysis using Bayesian and ML analysis as described above (Fig. 1B).

Finally, an alignment of 217 ST3 sequences (Table 1) was submitted to Maximum Likelihood analysis as above (Supplementary Fig. 3).

3. Results

Data obtained by using the MLST assay and results of studies of intra-subtype variability will be described below separately for the two subtypes. Due to extensive genetic divergence in the *Blastocystis* MLO genome between subtypes, it proved impossible for us to identify a common set of primers for MLST analysis that could be applied to all subtypes, and so it seems likely that a distinct set of MLST loci will need to be developed for analysis of each subtype. However, one advantage of developing subtype-specific markers is that the problem of analysing samples containing mixtures of subtypes will be overcome.

3.1. ST3 MLST assay and intra-subtype variability

Two of the initial seven MLST locus candidates were discarded; one (*rps11*) due to a homopolymer of 11–12 adenine bases present in the middle of the locus that frequently hampered successful bidirectional sequencing, and the other (*rps4*) due to reliability problems with amplification. None of the 81 DNAs tested failed to amplify at any of the remaining five loci. This indicates high assay sensitivity for ST3 from primates, whether using DNA extracted directly from stool or from cultures. MLST primer sequences, locus sizes and genome positions are given in Table 3.

A total of 185 polymorphic sites were identified among the 1448 positions in the concatenated alignment (12.9% diversity). Fifty-eight sequence types were detected and the overall discriminatory power based on analysis of the 81 sequences was 0.99. A discriminatory index of >0.90 is necessary for MLST data to be interpreted with confidence (Hunter, 1990; Hunter and Gaston, 1988). Even within MLO Clade 1 (in which most of the human sequences were found, see below) the discriminatory index was 0.99. Sixteen sequence types were present in more than one sample. Three sequences included in the study were retrospectively identified as being from the same individual sampled at different times and sequence type ST3.3 was obtained for all three samples, indicating high MLST assay reproducibility. A list of alleles linked to sample IDs and sequence types is available in Supplementary Table 1.

Table 2
Blastocystis ST4 samples analysed in the study.

SSU-rDNA clade[a]	MLO genome clade	SQT	Host (Species or common name)	Geographic origin	DNA sample/GenBank Accession. No.	References
1	1	ST4.1	*Cavia porcellus*	Denmark	GP_KVL	Present study
			Homo sapiens	Denmark	F4130, T66888	Stensvold et al. (2011)
					T51586, W54277	Rene et al. (2009)
				UK	DMP/02-328	Pérez-Brocal and Clark (2008)
					MA24, MA52, MA61, MA70, MA100, MA136–7, MA164, MA167, MA179, MA181–2, MA187, MA192, MA321, MA328, MA335, MA341, MA355–6, MA366, MA368, MA371, MA375–7, MA388, MA392, MA401, MA415, MA420	Alfellani et al. (unpublished data)
				Nigeria	SL3	Onuoha et al. (in preparation)
		ST4.2	*Homo sapiens*	UK	MA49, MA72, MA93, MA96, MA114	Alfellani et al. (unpublished data)
		ST4.3	*Homo sapiens*	UK	MA145	Alfellani et al. (unpublished data)
		ST4.4	*Homo sapiens*	UK	MA333	Alfellani et al. (unpublished data)
		ST4.5	*Homo sapiens*	Denmark	T8428	Stensvold et al. (2011)
				UK	MA59, MA112, MA144, MA158	Alfellani et al. (unpublished data)
2	2	ST4.6	*Homo sapiens*	USA	DMP/10-212	Present study
1	–	–	*Homo sapiens*	Sweden	454, 560, 575, 885, 1542, 1842, 1859, 2042, 2109, 3025, 4321	Forsell et al. (in press)
				Denmark	1922, 3058, 4009, 5002, 5023, 8024, 8032, 26825, 26861, 36582, 66842, 68507	Stensvold et al. (2011)
					AM118079, AM275389–93	Stensvold et al. (2006)
					MAUMR	Alfellani et al. (unpublished data)
				Ireland	AM992465, AM992467–8	Scanlan and Marchesi (2008)
				Germany	AY244619–20	Yoshikawa et al., 2004
				Japan	AY244621	Yoshikawa et al. (2004)
				UK	DQ232781, DQ232812, DQ232813, DQ232815, DQ232816, DQ232818, DQ232826, DQ232831, DQ232835, DQ232837, DQ232838, DQ232841, DQ232846	Scicluna et al. (2006)
					MA001, MA3, MA010, MA19, MA22, MA51, MA82, MA85, MA88, MA107, MA131, MA147, MA151, MA155, MA165, MA189, MA197, MA205, MA208, MA217, MA219–20, MA231, MA245, MA265, MA267, MA269, MA272, MA283, MA285, MA295, MA300, MA301, MA329, MA332, MA343, MA347, MA436, MA438–9, MA441, MA447, MAAB114, MAAB135, MAAB138, MAAB151, MAAB153, MAAB161, MAAB163, MAAB170, MAAB175, MAAB183, MAAB186, MAAB190–91, MAG1	Alfellani et al. (unpublished data)
				USA	HQ641622 (ATCC 50608)	Santín et al. (2011)
					EU679347	Whipps et al. (2010)
					EU482085, EU482087 (ATCC 50608, ATCC 50753)	Jones et al. (2008)
					EF494741 (ATCC 50608)	Stechmann et al. (2008)
				France	FJ666840, FJ666852, FJ666868, FJ666869, FJ666885	Souppart et al. (2009)
				Australia	RJT51AUSTRALIA	Traub et al. (unpublished data)
			Lemur catta	Spain	HQ641652	Santín et al. (2011)
			Rattus norvegicus	Singapore	AY590111, AY590113–4	Noël et al. (2005)
			Cavia porcellus	USA	U51152	Silberman et al. (1996)
2	–	–	*Homo sapiens*	Denmark	AM712466	Stensvold et al. (2007a)
				Turkey	AM778994	Ozyurt et al. (2008)
			Lemur catta	Denmark (zoo)	RL081_Ringtailedlemur, RL083_Ringtailedlemur	Stensvold et al. (2009a)
			Rattus norvegicus	France	AY135407–8	Noël et al. (2003)
				Japan	AB071000, AB091251	Arisue et al. (2003)
			Cavia porcellus	USA	U26177 (ATCC 50578)	Leipe et al. (1996)

–, Data not available.
[a] Based on SNP configuration throughout the gene; see text for details.

Phylogenetic analysis of concatenated nucleotide sequences of the 81 samples revealed the existence of four MLO clades (Fig. 1A, Table 1). MLO Clade 1 comprised four sequences from NHPs and all but two of the human sequences; these two exceptions were from NHP keepers. MLO Clade 2 comprised only two sequences, one from a patas monkey and one from a NHP keeper; the keeper was not affiliated with the zoo that hosts the patas monkey. MLO Clades 3 and 4 each included four sequences; Clade 3 comprised sequences from a colobus monkey, two macaques and a chimpanzee, while Clade 4 comprised sequences from two baboons, a macaque and a NHP keeper, who again was not affiliated with the zoo hosting the three NHPs in this clade.

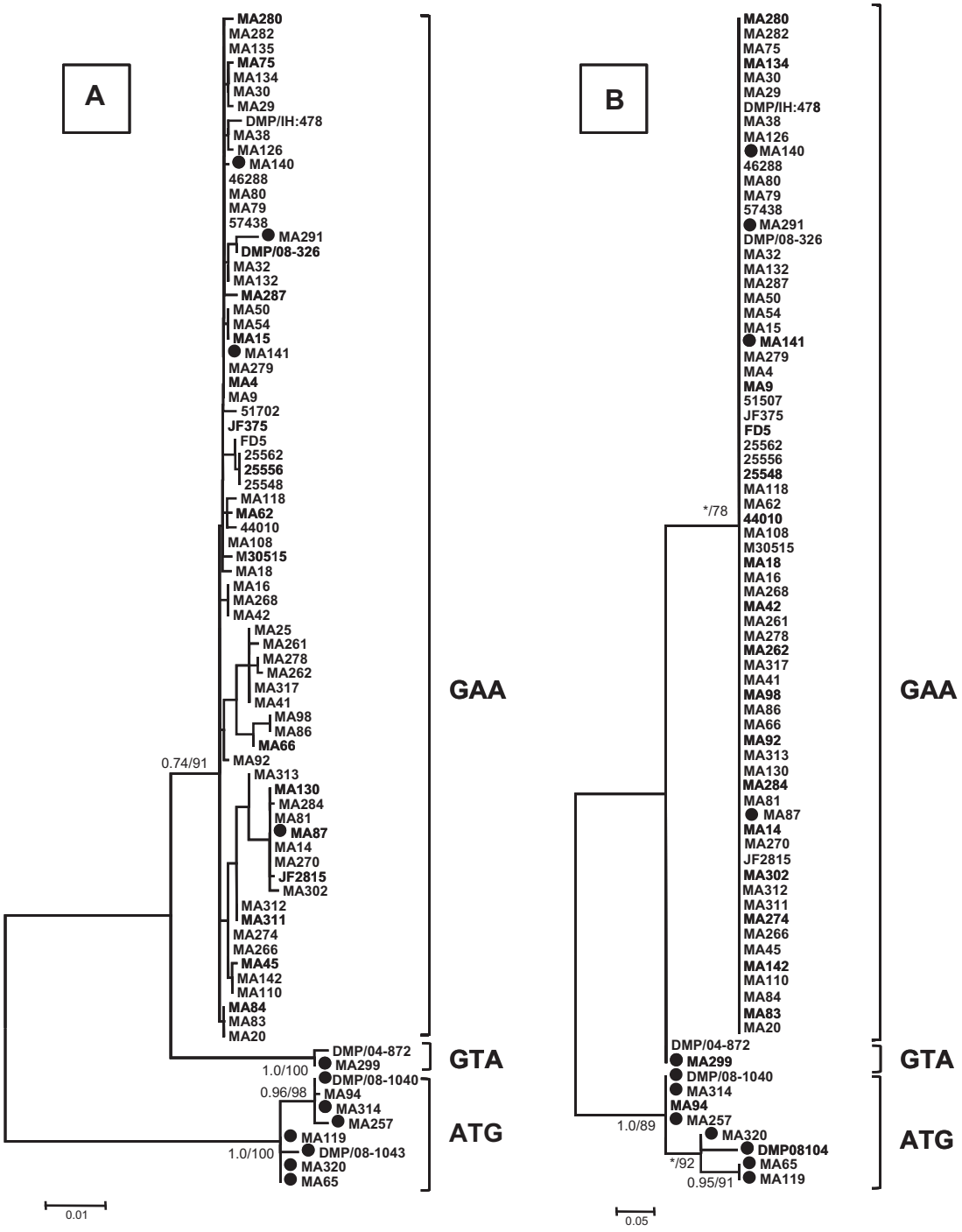

Fig. 1. Phylogenetic analyses of 81 concatenated sequences (1448 bp) obtained by MLST of *Blastocystis* ST3 (A) and their corresponding SSU-rDNAs (309 bp) (B) from humans ($n = 69$) and non-human primates ($n = 12$); NHP samples are indicated by solid black circles. Maximum Likelihood trees are shown. Statistical support is given only for relevant nodes, and posterior probabilities/bootstrap values <0.85/85 are not shown (Bayesian/Maximum Likelihood). In B, 'MA25' and 'MA135' have not been included (see footnote in Table 1). Diagnostic SSU-rDNA SNP configurations (three consecutive bases at positions 131–133) are given for each of the clades (see text for details).

A large number of the DNA polymorphisms are silent and do not affect the amino acid sequences. However some polymorphisms cause quite significant protein changes. Sample MA320 shows a 15 bp deletion in *rps3* (locus 2), which maintains the same reading frame but shortens the Rps3 protein by five amino acids. Additionally, variation in the number of adenine bases in a homopolymer within the *rps12* gene (locus 5) was observed. A homopolymer of either 9, 10 or 11 adenines is found towards the 3′ end of the gene leading to the Rps12 protein differing in length by up to 10 amino acids at the C-terminus due to the varying position of the stop codon; in all three variants the latter is located within a tRNA-Asn gene. This variation was not linked to host or geographic origin (data not shown). Similar variation has not been reported for other *Blastocystis* STs to date, but may also be found when intra-subtype diversity is investigated; a homopolymer in the same location is indeed present in the ST1, ST4 and ST7 MLO genomes sequenced so far. Interestingly, the *rps12* coding region of ST7 (CU914152) is 441 bp long, whereas it is 378 bp in both ST1 (EF494740) and

Table 3
Description of MLST loci and primers for amplification and sequencing of 81 ST3 and 51 ST4 samples and display of differences in polymorphism between the subtypes.

Blastocystis sp. subtype	Locus No.	PCR and sequencing primers		Size of PCR product (kbp)	Locus size (bp)[a]	Locus coordinates[a]	No./% polymorphic sites[b]	No. of alleles[b]	MLO target
		Primer ID	Sequence (5′–3′)						
3	1	ST3_2583F	TTAAGCTCTGTACACACCGC	0.35	180	2635–2814	28/15.6	15	SSU-rRNA
		ST3_2935R	AAATATCTAGGTATCCATCTAATGCTC						
3	2	ST3_7492F	CGTTTGTCTTTCTTCTAGTGTTAATTTAC	0.34	195	7565–7759	33/17.0	20	rps3
		ST3_7832R	AACGAGCTAAAAATTTATCATTATAGC						
3	3	ST3_10354F	AATCAACTATTTATTTCCAAGGTTTAC	0.51	319	10,430–10,748	24/7.5	13	nad4
		ST3_10867R	GGTAACATTGGAACTTTAACAGC						
3	4	ST3_22786F	AAAATTTGAGTAACATTTAAGTTTAAAGC	0.6	385	22,865–23,249	78/20.3	31	nad2, nad11
		ST3_23388R	GATCAGCACCTATGATACCTAAAC						
3	5	ST3_27599F	TGGTATTGGACATAATTTACAAGAAC	0.52	357	27,673–28,029	22/6.2	21	rps12, t-RNA-Asn, tRNA-Leu
		ST3_28117R	CTATAACTTCACAGTTATGAGCAGC						
4	1	ST4_2500F	AAGTTGCAATTATGAGAATAAGAGC	0.49	386	2529–2915	38 (1)/ 9.8 (0.3)	4 (3)	SSU-rRNA, tRNA-Met
		ST4_2987R	AAATATCAAGGTATCCATCTAATGC						
4	2	ST4_5657F	CCATGTCAACAGTTCGAATCTG	0.63	528	5706–6234	72 (3)/ 13.6 (0.6)	3 (2)	tRNA-Lys, rps13
		ST4_6287R	TTTTACCTTGAGAATTATGACCAC						
4	3	ST4_15015F	ATTTATTATTACTAAATATGGAATTGCAC	0.79	305	15,392–15,697	61 (1)/ 20.0 (0.3)	3 (2)	rps8, rpl6
		ST4_15800R	CAAATTTATAACCAATTCCTTG						
4	4	ST4_15800F	TTTACAAGGAATTGGTTATAAATTTG	0.74	337	16,078–16,415	75 (1)/ 22.3 (0.3)	3 (2)	orf192
		ST4_16538R	GATTGATTATTTAATATTAATGTATACATAGTTTG						
4	5	ST4_25637F	TTATTTTGTGGAGGTTGGTTAC	0.45	327	25,689–26,016	25 (1)/ 7.6 (0.3)	3 (2)	nad1, nad4L
		ST4_26084R	TAATCCAATAGAAGTTTCAACTG						
4	6	ST4_27017F	CGTCGAAAAGTTGCTAAAGTG	0.56	435	27,091–27,526	41 (5)/ 9.4 (1.1)	3 (2)	rps12, tRNA-Asn, tRNA-Leu
		ST4_27580R	GTTATGAGCAGCACATTCTAACC						

[a] Relative to DMP/IH:478 (ST3) and DMP/02-328 (ST4), respectively. Locus sizes only included the part of the PCR product for which SNPs were present and for which bidirectional sequencing was consistently obtained.
[b] Data shown in parentheses are calculated on the basis of ST4 Clade 1 strains only, which represent 50/51 ST4 strains analysed by MLST.

ST4 (EF494739) and 408 bp in ST3 (DMP/IH:478), consistent with similar variation existing.

Phylogenetic analyses (Bayesian and ML) of the corresponding nuclear SSU-rDNAs showed a topology compatible with the one obtained for MLO data (Fig. 1B). SSU Clades 1 and 2 were congruent with MLO Clades 1 and 2, respectively, and SSU Clade 3 corresponded to MLO Clades 3 and 4. No samples representing SSU Clades 4 and 5 were available for MLST analysis. At position 131–133 in the ST3 alignment, five different SNP configurations were observed (Table 1; Fig. 1; Supplementary Fig. 1). The base triplet GAA (SSU Clade 1) was the most common and seen in humans (n = 164) from the UK, Denmark, Sweden, France, Italy, Libya, Egypt, Tanzania, Vietnam, Japan and the Philippines, NHPs (n = 11), and two large mammals (pig and cattle). Seven sequences from five NHPs and two humans had GTA (SSU Clade 2). ATG (SSU Clade 3) was seen in 32 sequences from 30 NHPs and two humans, one of whom was known to be a NHP keeper, and two human sequences from France and Japan had ATC (SSU Clade 4). A colobus monkey from Tanzania had AAA (SSU Clade 5) (for a phylogenetic tree illustration of all ST3 SSU-rDNAs included in Table 1 see Supplementary Fig. 3). The data indicate that the short fragment of the SSU-rRNA gene including the base pair triplet in position 131–133 is prognostic for intra-subtype variation in ST3 MLOs.

3.2. ST4 MLST assay and intra-subtype variability

All 51 ST4 DNAs tested amplified consistently across all loci. The overall discriminatory power of the assay was only 0.43 and only six sequence types were detected (Table 2; Supplementary Table 2). The two clades identified at SSU-rDNA level were reflected in the MLO genome sequences (data not shown). In Clade 1, only five sequence types could be identified, and 38/51 (74.5%) ST4s included in the study were sequence type ST4.1, one of which was from a rodent host (GP-KVL). Since the seventh locus targeting the *orf143* did not provide further discrimination, the number of loci in the assay was kept to 6. MLST primer sequences, locus sizes and positions are given in Table 3. No variation was found across the SSU-rRNA gene among the 50 ST4 Clade 1 sequences.

DMP/10-212 was the only DNA sample available representing ST4 Clade 2, and when excluded from the dataset, the discriminatory power of the assay was reduced to 0.41, and only 12/2318 positions in the concatenated alignment (0.5%) exhibited polymorphism (Tables 2 and 3; Supplementary Fig. 4). Of these, only three were in coding regions: in locus 2, a lysine was present at amino acid position 79 in *rps13* of sequence type ST4.5, whereas ST4.1 had a glutamine. In locus 4, at position 15 in *orf192* sequence type ST4.5 had an alanine and ST4.1 a valine. At locus 6, one SNP was present in a tRNA-Met: at position 47 a 'C' was present in sequence type ST4.1, whereas ST4.5 had 'T'. Sequence type ST4.5 was represented by five DNAs with 12 polymorphisms relative to sequence type ST4.1. Sequence types 4.2–4.4 were represented by only one DNA sample each, and each of these shared their polymorphisms with either ST4.1 or ST4.5.

Including DMP/10-212 (sequence type ST4.6) remarkably raised the amount of polymorphism to 13.5% across the six loci (Table 3), which was comparable to the amount of variation seen within ST3. SNPs in sequence type ST4.6 resulted in a protein with 14 amino acid changes in *rps13* (locus 2) and that was nine amino acids longer than that of the remaining sequence types. Due to the divergence of DMP/10-212, substantial differences in amino acid sequences in the other MLST loci were also observed (data not shown).

In the alignment of the 183 ST4 SSU-rDNAs (Table 2), it appeared that each belonged to one of two clades, and 19 consistent SNPs across the entire gene separated the two clades (Supplementary Fig. 2). A total of 177 (97%) of the sequences analysed had the Clade 1 SNP configuration (Table 2). Only 3/170 sequences from humans (from Denmark, Turkey and USA) had the Clade 2 configuration, one of which was represented by the North-American sample DMP/10-212. Variation among the few SSU Clade 2 sequences included additional polymorphisms shared by more than one sequence, suggesting that Clade 2 might be more genetically diverse than Clade 1, in which only few sporadic SNPs were detected across

167 sequences (Supplementary Fig. 2). The 167 human ST4 sequences belonging to Clade 1 were from Denmark, Sweden, UK, Ireland, France, Spain, Nigeria, Australia, Japan, and North America.

4. Discussion

Following the introduction of a consensus subtyping system for *Blastocystis* (Stensvold et al., 2007b) dozens of studies have aimed to characterise the distribution of *Blastocystis* subtypes in humans and other animals. This study, however, is the first to thoroughly investigate intra-subtype diversity of *Blastocystis* using genetic markers other than the SSU-rRNA gene. By analysing comprehensive and complex sequence data sets, our results clearly highlight the importance and value of investigating intra-subtype diversity in epidemiological and evolutionary studies of *Blastocystis*.

Although lower in amount, the SSU-rDNA variation seen in each subtype mirrors the variation seen in MLO genome sequences. Moreover, the phylogenetic inferences are more or less the same no matter which of the two datasets is used, although only low to modest statistical support is obtained in the analysis of partial SSU-rDNAs.

4.1. Substantial intra-subtype diversity in ST3 and indications of frequent human-to-human transmission

The uncovering of phylogenetically distinct clades within ST3 is important. Although more data are needed, they clearly indicate that human infections are restricted to MLO Clade 1 except where exposure to NHPs has occurred (Fig. 1A). This suggests relatively high host specificity within ST3 and that human strains falling into MLO Clades 2–4 (=SSU Clades 2 and 3) are almost certainly the result of zoonotic transmission (Fig. 1A, Table 1).

This is the first study to report cryptic host specificity within a *Blastocystis* subtype. Relatively few DNAs from NHPs were analysed in the study, but the fact that they segregated into four different clades suggests that the diversity among strains from NHPs may be even more extensive than observed here. In addition to humans and NHPs, ST3 is hosted by a variety of other mammals, including pigs, cattle and dogs (Stensvold et al., 2009a) and rodents (Alfellani et al. (unpublished data)). Assuming the present MLST assay is applicable to all ST3s of non-primate origin, data from the analysis of such strains will assist in identifying whether further cryptic host specificity can be identified. If the MLST primers do not amplify such samples, this could be indicative of significant divergence between primate and non-primate ST3 sequences and make zoonotic transmission an even less likely contributor to human *Blastocystis* infection. Until more SSU-rDNA and MLST data are available for non-primate hosts, we can only conclude that most ST3 infections in humans are the result of human-to-human transmission. The hypothesis introduced by Noël et al. (2005) about ST3 being of human origin cannot be supported in view of our data. Incidentally, Petrášová et al. (2011) concluded based on the analysis of SSU-rDNAs that zoonotic transmission of ST1 and ST2 was unlikely among syntopic human and NHPs on the Rubondo Island, Tanzania; hence evidence is growing for anthroponotic transmission of *Blastocystis* in general.

4.2. Homogeneity of ST4 and aspects of host spectrum and geographical distribution

The host range of ST4 identified so far is restricted to humans, a few rodents, NHPs, and one Australian opossum (Stensvold et al., 2009a; Parkar et al., 2007). To date, no major differences in the host spectrum of the two clades within ST4 have been identified; however, very few sequences are available for ST4 SSU Clade 2; three of the samples are from humans, two from NHPs and five from rodents (Table 1). The much more common Clade 1 is mainly seen in humans, but NHPs and rodents are known hosts as well (Table 1) and, moreover, the Clade 1 sequence from an opossum apparently showed no divergence when compared to reference sequences (Parkar et al., 2007). Conspicuously, genetic variation within ST4 Clade 1 appears to be practically absent across the globe irrespective of the host.

Surprisingly, ST4 is rarely reported in several Asian and Middle Eastern *Blastocystis* subtype surveys, while ST4 appears to be common in Europe, at least in patients with intestinal symptoms (Forsell et al., in press; Stensvold et al., 2011; Souppart et al., 2009; Domínguez-Márquez et al., 2009). Many of the studies from the Middle East and Asia have been based on the methodology introduced by Yoshikawa et al. (1998, 2000, 2003) which makes use of subtype-specific sequence tagged site (STS) primers (Dogruman-Al et al., 2008, 2009a,b; Eroglu et al., 2009; Eroglu and Koltas, 2010; Hussein et al., 2008; Iguchi et al., 2007; Li et al., 2007a,b; Tan et al., 2008, 2009; Yakoob et al., 2010; Yan et al., 2006, 2007; Yoshikawa et al., 2009). The application of a STS primer panel is theoretically advantageous since mixed infections are more easily detected when compared to other methods, including direct sequencing of PCR products amplified by genus-specific primers as used here; however, the primer pair SB337 used to amplify ST4 in the STS method does not amplify ST4 Clade 1 in our hands (unpublished observations). It did, however, amplify our Clade 2 sample DMP/10-212. SB337 was originally developed using the RN94-9 strain (Yoshikawa et al., 1998) from the ST4 Clade 2, and later validated using NIH:1295:1 (Accession No. U51152), a Clade 1 strain from a guinea pig (Yoshikawa et al., 2003), and therefore, theoretically, neither of the two clades should be missed using this primer pair. However, the failure of Clade 1 amplification in our lab using the SB337 primer pair means that Clade 1 could be overlooked where these primers are used. It may be that the few ST4s reported in Asian studies using the STS primers all belong to Clade 2. Interestingly, in a study from Japan, Kaneda et al. (2001) used RFLP on PCR products amplified by general eukaryotic primers (RD5/RD3) and reported quite a few ST4s (RLFP nomenclature: Ribodeme 3; Stensvold et al., 2007b). What is more, Noël et al. (2005) published three ST4 sequences from Singaporean rats; all three belonged to ST4 Clade 1, which is evidence of this clade being present in Asia. We have recently reported the absence of ST4 in Brazilian indigenous people in the Mato Grosso region (Malheiros et al., in press) using the barcoding method. Studies of samples from Colombia, Philippines and Thailand using a similar methodology also do not report ST4 (Leelayoova et al., 2008; Rivera, 2008; Santín et al., 2011). Hence, while ST4 appears to be absent or very rare in certain sampled regions, in South America, the Middle East and Asia, the reliability of the STS primers should be scrutinised; the absence of ST4 in some regions should be validated by using sequencing methods.

4.3. Implications of differences in intra-subtype genetic variability for parasite epidemiology and evolution

The low genetic variation in ST4 SSU Clade 1 is reflected in the MLO genome. Among 50 Clade 1 DNAs, only five sequence types were detected, and a total of only 12 SNPs could be identified across six loci covering more than 2300 bp. This has several implications. Practically, this means that the current MLST assay is not an appropriate tool for investigating ST4 Clade 1. It may be that SSU Clade 1 strains are genetically very similar across the entire nuclear genome also, in which case the search for useful genetic variation is a futile quest. As yet, no studies have been done on microsatellites in *Blastocystis*, and it may be so that such studies will prove valuable in terms of identifying variation in ST4. Since only one DNA representing Clade 2 was available for analysis, it

is impossible at present to comment on the discriminatory power of the assay for analysis of Clade 2 strains.

More generally, the present data and observations allow us to speculate on a number of points: not only are subtypes of *Blastocystis* separated by great genetic distance (Stensvold et al., 2007b, 2009a; Parkar et al., 2010), they also appear very different in terms of intra-subtype variation. The almost clonal population structure of ST4 Clade 1 combined with its high prevalence relative to ST4 Clade 2 is consistent with this clade having expanded in humans relatively recently compared to ST3. This is also supported by the emerging data on the infrequency or total absence of human ST4 in some parts of the world. Assuming faecal-oral transmission for all subtypes of *Blastocystis*, it is interesting that a parasite widespread in Europe is rare or absent in parts of the world where faecal–orally transmitted parasites are much more prevalent than in Europe. Importantly, ST4 has been associated with intestinal disease and pathogenicity in a number of recent surveys (Stensvold et al., 2011; Domínguez-Márquez et al., 2009) and in-vitro studies (summarised by Stensvold et al., 2009c). Since ST4 Clade 1 sequences from rats, guinea pigs, opossums and most humans appear genetically identical it is impossible to discard a hypothesis of zoonotic transmission of ST4. The data at present support the theory that rodents may be a reservoir for human *Blastocystis* infections, as also proposed by Noël et al. (2005).

Our analyses, combined with phylogenetic studies of nucleotide sequences coding for elongation factors, hsp70 and an ATPase (Ho et al., 2001; Arisue et al., 2002), provide evidence for the SSU-rRNA gene being a robust and highly informative genetic marker for phylogenetic inferences in *Blastocystis*. However, MLST will prove a powerful tool in studies aiming to characterise ST3 transmission patterns and to survey strains in a host over a prolonged period (e.g. pre- and post-treatment). Currently, MLST schemes for ST1 and ST2 are under development (Alfellani et al. (unpublished data)). Our study shows that the inclusion of SSU-rDNA and MLO MLST analyses of *Blastocystis* from various cohorts (symptomatic and asymptomatic individuals) and hosts (humans, NHP and non-primate hosts) from different parts of the world will be crucial in future attempts to establish the epidemiology and clinical significance of the parasite.

Acknowledgements

This work was funded by the Danish Council for Independent Research – Medical Sciences (Grant No. 271-09-0251). The Council had no influence or impact on study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

The following people are thanked for providing DNA sample or sequence material: Rebecca J. Traub, Emma Victory, Heidi Larsen Enemark, Simone Caccio, Derya Taner-Mulla, Joakim Forsell and Sunday Eme Onuoha.

This publication made use of the *Blastocystis* Sequence Typing website (http://pubmlst.org/blastocystis/) developed by Keith Jolley and sited at the University of Oxford (Jolley and Maiden, 2010).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2011.11.002.

References

Abe, N., 2004. Molecular and phylogenetic analysis of *Blastocystis* isolates from various hosts. Vet. Parasitol. 25, 235–242.

Arisue, N., Hashimoto, T., Yoshikawa, H., Nakamura, Y., Nakamura, F., Yano, T.A., Hasegawa, M., 2002. Phylogenetic position of *Blastocystis hominis* and of stramenopiles inferred from multiple molecular sequence data. J. Eukaryot. Microbiol. 49, 42–53.

Arisue, N., Hashimoto, T., Yoshikawa, H., 2003. Sequence heterogeneity of the small subunit ribosomal RNA genes among *Blastocystis* isolates. Parasitology 126, 1–9.

Bain, J.M., Tawanti, A., Davidson, A.D., Jacobsen, M.D., Shaw, D., Gow, N.A., Odds, F.C., 2007. Multilocus sequence typing of the pathogenic fungus *Aspergillus fumigatus*. J. Clin. Microbiol. 45, 1469–1477.

Bougnoux, M.E., Morand, S., d'Enfert, C., 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. J. Clin. Microbiol. 40, 1290–1297.

Clark, C.G., Diamond, L.S., 2002. Methods for cultivation of luminal parasitic protists of clinical importance. Clin. Microbiol. Rev. 15, 329–341.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16, 10881–10890.

Dogruman-Al, F., Dagci, F., Yoshikawa, H., Kurt, O., Demirel, M., 2008. A possible link between subtype 2 and asymptomatic infections of *Blastocystis hominis*. Parasitol. Res. 103, 685–689.

Dogruman-Al, F., Kustimur, S., Yoshikawa, H., Tuncer, C., Simsek, Z., Tanyuksel, M., Araz, E., Boorom, K., 2009a. *Blastocystis* subtypes in irritable bowel syndrome and inflammatory bowel disease in Ankara, Turkey. Mem. Inst. Oswaldo Cruz 104, 724–727.

Dogruman-Al, F., Yoshikawa, H., Kustimur, S., Balaban, N., 2009b. PCR-based subtyping of *Blastocystis* isolates from symptomatic and asymptomatic individuals in a major hospital in Ankara, Turkey. Parasitol. Res. 106, 263–268.

Domínguez-Márquez, M.V., Guna, R., Muñoz, C., Gómez-Muñoz, M.T., Borrás, R., 2009. High prevalence of subtype 4 among isolates of *Blastocystis hominis* from symptomatic patients of a health district of Valencia (Spain). Parasitol. Res. 105, 949–955.

Eroglu, F., Koltas, I.S., 2010. Evaluation of the transmission mode of *B. Hominis* by using PCR method. Parasitol. Res. 107, 841–845.

Eroglu, F., Genc, A., Elgun, G., Koltas, I.S., 2009. Identification of *Blastocystis hominis* isolates from asymptomatic and symptomatic patients by PCR. Parasitol. Res. 105, 1589–1592.

Forsell, J., Granlund, M., Stensvold, C.R., Clark, C.G., Evengård, B., in press. Subtype analysis of *Blastocystis* isolates in Swedish patients. Eur. J. Clin. Microbiol. Infect. Dis.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Ho, L.C., Jeyaseelan, K., Singh, M., 2001. Use of the elongation factor-1 alpha gene in a polymerase chain reaction-based restriction-fragment-length-polymorphism analysis of genetic heterogeneity among *Blastocystis* species. Mol. Biochem. Parasitol. 112, 278–291.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Hunter, P., 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J. Clin. Microbiol. 28, 1903–1905.

Hunter, P.R., Gaston, M.A., 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26, 2465–2466.

Hussein, E.M., Hussein, A.M., Eida, M.M., Atwa, M.M., 2008. Pathophysiological variability of different genotypes of human *Blastocystis hominis* Egyptian isolates in experimentally infected rats. Parasitol. Res. 102, 853–860.

Iguchi, A., Ebisu, A., Nagata, S., Saitou, Y., Yoshikawa, H., Iwatani, S., Kimata, I., 2007. Infectivity of different genotypes of human *Blastocystis hominis* isolates in chickens and rats. Parasitol. Int. 56, 107–112.

Jolley, K.A., Maiden, M.C., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11, 595.

Jones 2nd, M.S., Ganac, R.D., Hiser, G., Hudson, N.R., Le, A., Whipps, C.M., 2008. Detection of *Blastocystis* from stool samples using real-time PCR. Parasitol. Res. 103, 551–557.

Kaneda, Y., Horiki, N., Cheng, X.J., Fujita, Y., Maruyama, M., Tachibana, H., 2001. Ribodemes of *Blastocystis hominis* isolated in Japan. Am. J. Trop. Med. Hyg. 65, 393–396.

Lantsman, Y., Tan, K.S., Morada, M., Yarlett, N., 2008. Biochemical characterization of a mitochondrial-like organelle from *Blastocystis* sp. subtype 7. Microbiology 154, 2757–2766.

Leelayoova, S., Siripattanapipong, S., Thathaisong, U., Naaglor, T., Taamasri, P., Piyaraj, P., Mungthin, M., 2008. Drinking water: a possible source of *Blastocystis* spp. subtype 1 infection in schoolchildren of a rural community in central Thailand. Am. J. Trop. Med. Hyg. 79, 401–406.

Leipe, D.D., Tong, S.M., Goggin, C.L., Slemenda, S.B., Pieniazek, N., Sogin, M.L., 1996. 16S-like rDNA sequences from *Developayella elegans*, *Labyrinthuloides haliotidis*, and *Proteromonas lacertae* confirm that the stramenopiles are a primarily heterotrophic group. Eur. J. Protistol. 32, 449–458.

Li, L.H., Zhang, X.P., Lv, S., Zhang, L., Yoshikawa, H., Wu, Z., Steinmann, P., Utzinger, J., Tong, X.M., Chen, S.H., Zhou, X.N., 2007a. Cross-sectional surveys and subtype classification of human *Blastocystis* isolates from four epidemiological settings in China. Parasitol. Res. 102, 83–90.

Li, L.H., Zhou, X.N., Du, Z.W., Want, X.Z., Wang, L.B., Jiang, J.Y., Yoshikawa, H., Steinmann, P., Utzinger, J., Wu, Z., Chen, J.X., Chen, S.H., Zhang, L., 2007b. Molecular epidemiology of human *Blastocystis* in a village in Yunnan province, China. Parasitol. Int. 56, 281–286.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Malheiros, A.F., Stensvold, C.R., Clark, C.G., Braga, G.B., Shaw, J.J., in press. Molecular characterisation of *Blastocystis* obtained from members of the indigenous Tapirapé Ethnic Group from the Brazilian Amazon Region, Brazil. Am. J. Trop. Med. Hyg.

Meloni, D., Sanciu, G., El Alaoui, H., Chabé, M., Delhaes, L., Dei-Cas, E., Delbac, F., Luigi Fiori, P., Di Cave, D., Viscogliosi, E., 2011. Molecular subtyping of *Blastocystis* sp. isolates from symptomatic patients in Italy. Parasitol. Res. 109, 613–619.

Noël, C., Peyronnet, C., Gerbod, D., Edgcomb, V.P., Delgado-Viscogliosi, P., Sogin, M.L., Capron, M., Viscogliosi, E., Zenner, L., 2003. Phylogenetic analysis of *Blastocystis* isolates from different hosts based on the comparison of small-subunit rRNA gene sequences. Mol. Biochem. Parasitol. 126, 119–123.

Noël, C., Dufernez, F., Gerbod, D., Edgcomb, V.P., Delgado-Viscogliosi, P., Ho, L.C., Singh, M., Wintjens, R., Sogin, M.L., 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.

Ozyurt, M., Kurt, O., Mølbak, K., Nielsen, H.V., Haznedaroglu, T., Stensvold, C.R., 2008. Molecular epidemiology of *Blastocystis* infections in Turkey. Parasitol. Int. 57, 300–306.

Parkar, U., Traub, R.J., Kumar, S., Mungthin, M., Vitali, S., Leelayoova, S., Morris, K., Thompson, R.C., 2007. Direct characterization of *Blastocystis* from faeces by PCR and evidence of zoonotic potential. Parasitology 134, 359–367.

Parkar, U., Traub, R.J., Vitali, S., Elliot, A., Levecke, B., Robertson, I., Geurden, T., Steele, J., Drake, B., Thompson, R.C., 2010. Molecular characterization of *Blastocystis* isolates from zoo animals and their animal-keepers. Vet. Parasitol. 169, 8–17.

Pérez-Brocal, V., Clark, C.G., 2008. Analysis of two genomes from the mitochondrion-like organelle of the intestinal parasite *Blastocystis*: complete sequences, gene content, and genome organization. Mol. Biol. Evol. 25, 2475–2482.

Petrášová, J., Uzlíková, M., Kostka, M., Petrželková, K.J., Huffman, M.A., Modrý, D., 2011. Diversity and host specificity of *Blastocystis* in syntopic primates on Rubondo Island, Tanzania. Int. J. Parasitol. 41, 1113–1120.

Rene, B.A., Stensvold, C.R., Badsberg, J.H., Nielsen, H.V., 2009. Subtype analysis of *Blastocystis* isolates from *Blastocystis* cyst excreting patients. Am. J. Trop. Med. Hyg. 80, 588–592.

Rivera, W.L., 2008. Phylogenetic analysis of *Blastocystis* isolates from animal and human hosts in the Philippines. Vet. Parasitol. 156, 178–182.

Santín, M., Gómez-Muñoz, M.T., Solano-Aquilar, G., Fayer, R., 2011. Development of a new PCR protocol to detect and subtype *Blastocystis* spp. from humans and animals. Parasitol. Res. 109, 205–212.

Scanlan, P.D., Marchesi, J.R., 2008. Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. ISME J. 2, 1183–1193.

Scicluna, S.M., Tawari, B., Clark, C.G., 2006. DNA barcoding of *Blastocystis*. Protist 157, 77–85.

Silberman, J.D., Sogin, M.L., Leipe, D.D., Clark, C.G., 1996. Human parasite finds taxonomic home. Nature 380, 398.

Souppart, L., Sanciu, G., Cian, A., Wawrzyniak, I., Delbac, F., Capron, M., Dei-Cas, E., Boorom, K., Delhaes, L., Viscogliosi, E., 2009. Molecular epidemiology of human *Blastocystis* isolates in France. Parasitol. Res. 105, 413–421.

Souppart, L., Moussa, H., Cian, A., Sanciu, G., Poirier, P., El Alaoui, H., Delbac, F., Boorom, K., Delhaes, L., Dei-Cas, E., Viscogliosi, E., 2010. Subtype analysis of *Blastocystis* isolates from symptomatic patients in Egypt. Parasitol. Res. 106, 505–511.

Staden, R., Beal, K., Bonfield, J.K., 2000. The Staden package, 1998. Methods Mol. Biol. 132, 115–130.

Stechmann, A., Hamblin, K., Pérez-Brocal, V., Gaston, D., Richmond, G.S., van der Giezen, M., Clark, C.G., Roger, A.J., 2008. Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr. Biol. 18, 580–585.

Stensvold, R., Brillowska-Dabrowska, A., Nielsen, H.V., Arendrup, M.C., 2006. Detection of *Blastocystis hominis* in unpreserved stool specimens using polymerase chain reaction. J. Parasitol. 92, 1081–1087.

Stensvold, C.R., Arendrup, M.C., Jespersgaard, C., Mølbak, K., Nielsen, H.V., 2007a. Detecting *Blastocystis* using parasitologic and DNA-based methods: a comparative study. Diagn. Microbiol. Infect. Dis. 59, 303–307.

Stensvold, C.R., Suresh, G.K., Tan, K.S., Thompson, R.C., Traub, R.J., Viscogliosi, E., Yoshikawa, H., Clark, C.G., 2007b. Terminology for *Blastocystis* subtypes – a consensus. Trends Parasitol. 23, 93–96.

Stensvold, C.R., Alfellani, M.A., Nørskov-Lauritsen, S., Prip, K., Victory, E.L., Maddox, C., Nielsen, H.V., Clark, C.G., 2009a. Subtype distribution of *Blastocystis* isolates from synanthropic and zoo animals and identification of a new *Blastocystis* sp. Subtype. Int. J. Parasitol. 39, 473–479.

Stensvold, C.R., Lewis, H.C., Hammerum, A.M., Porsbo, L.J., Nielsen, S.S., Olsen, K.E., Arendrup, M.C., Nielsen, H.V., Mølbak, K., 2009b. *Blastocystis*: unravelling potential risk factors and clinical significance of a common but neglected parasite. Epidemiol. Infect. 137, 1655–1663.

Stensvold, C.R., Nielsen, H.V., Mølbak, K., Smith, H.V., 2009c. Pursuing the clinical significance of *Blastocystis* – diagnostic limitations. Trends Parasitol. 25, 23–29.

Stensvold, C.R., Christiansen, D., Olsen, K.E.O., Nielsen, H.V., 2011. *Blastocystis* sp. subtype 4 is common in Danish *Blastocystis*-positive patients presenting with acute diarrhea. Am. J. Trop. Med. Hyg. 84, 883–885.

Stenzel, D.J., Boreham, P.F., 1996. *Blastocystis hominis* revisited. Clin. Microbiol. Rev. 9, 563–584.

Sullivan, C.B., Diggle, M.A., Clarke, S.C., 2005. Multilocus sequence typing: data analysis in clinical microbiology and public health. Mol. Biotechnol. 29, 245–254.

Tamura, K., Peterson, D., Perterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Tan, T.C., Suresh, K.G., Smith, H.V., 2008. Phenotypic and genotypic characterisation of *Blastocystis hominis* isolates implicates subtype 3 as a subtype with pathogenic potential. Parasitol. Res. 104, 85–93.

Tan, T.C., Ong, S.C., Suresh, K.G., 2009. Genetic variability of *Blastocystis* sp. isolates obtained from cancer and HIV/AIDS patients. Parasitol. Res. 105, 1283–1286.

Wawrzyniak, I., Roussel, M., Diogon, M., Couloux, A., Texier, C., Tan, K.S., Vivarès, C.P., Delbac, F., Wincker, P., El Alaoui, H., 2008. Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*. Int. J. Parasitol. 38, 1377–1382.

Whipps, C.M., Boorom, K., Bermudez, L.E., Kent, M.L., 2010. Molecular characterization of *Blastocystis* species in Oregon identifies multiple subtypes. Parasitol. Res. 106, 827–832.

Yakoob, J., Jafri, W., Beg, M.A., Abbas, Z., Naz, S., Islam, M., Khan, R., 2010. Irritable bowel syndrome: is it associated with genotypes of *Blastocystis hominis*. Parasitol. Res. 106, 1033–1038.

Yan, Y., Su, S., Lai, R., Liao, H., Ye, J., Li, X., Luo, X., Chen, G., 2006. Genetic variability of *Blastocystis hominis* isolates in China. Parasitol. Res. 99, 597–601.

Yan, Y., Su, S., Ye, J., Lai, X., Liao, H., Chen, G., Zhang, R., Hou, Z., Luo, X., 2007. *Blastocystis* sp. subtype 5: a possible zoonotic genotype. Parasitol. Res. 101, 1527–1532.

Yeo, M., Mauricio, I.L., Messenger, L.A., Lewis, M.D., Llewellyn, M.S., Acosta, N., Bhattacharyya, T., Diosque, P., Carrasco, H.J., Miles, M.A., 2011. Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. PLoS Negl. Trop. Dis. 5, e1049.

Yoshikawa, H., Nagano, I., Wu, Z., Yap, E.H., Singh, M., Takahashi, Y., 1998. Genomic polymorphism among *Blastocystis hominis* strains and development of subtype-specific diagnostic primers. Mol. Cell. Probes 12, 153–159.

Yoshikawa, H., Abe, N., Iwasawa, M., Kitano, S., Nagano, I., Wu, Z., Takahashi, Y., 2000. Genomic analysis of *Blastocystis hominis* strains isolated from two long-term health care facilities. J. Clin. Microbiol. 38, 1324–1330.

Yoshikawa, H., Abe, N., Wu, Z., 2003. Genomic polymorphism among *Blastocystis* isolates and development of PCR-based identification of zoonotic isolates. J. Eukaryot. Microbiol. 50 (Suppl.), 710–711.

Yoshikawa, H., Wu, Z., Kimata, I., Iseki, M., Ali, I.K., Hossain, M.B., Zaman, V., Haque, R., Takahashi, Y., 2004. Polymerase chain reaction-based genotype classification among human *Blastocystis hominis* populations isolated from different countries. Parasitol. Res. 92, 22–29.

Yoshikawa, H., Wu, Z., Pandey, K., Pandey, B.D., Sherchand, J.B., Yanagi, T., Kanbara, H., 2009. Molecular characterization of *Blastocystis* isolates from children and rhesus monkeys in Kathmandu, Nepal. Vet. Parasitol. 160, 295–300.