TITLE

# Increased sampling reveals novel lineages of *Entamoeba*: consequences of genetic diversity and host specificity for taxonomy and molecular detection


AUTHORS

C. Rune Stensvold[a], Marianne Lebbad[b], Emma L. Victory[c], Jaco J. Verweij[d], Egbert Tannich[e], Mohammed Alfellani[c], Paulette Legarraga[c,2], C. Graham Clark[c,1]

AFFILIATIONS

[a]Department of Microbiology and Diagnostics, Statens Serum Institut, Artillerivej 3, DK-2300 Copenhagen S, Denmark

[b]Department of Parasitology, Mycology and Environmental Microbiology, Swedish Institute for Infectious Disease Control, SE-171 82 Solna, Sweden

[c]Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT London, United Kingdom.

[d]Department of Parasitology, Leiden University Medical Center, Postbus 9600, 2300 RC Leiden, The Netherlands

[e]Department of Molecular Parasitology, Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Str. 74, D-20359 Hamburg, Germany

[2]Present Address: Clinical Laboratory Department, Pontificia Universidad Católica de Chile, Santiago, Chile

[1]Corresponding author; fax: +44 20-7636-8739; email: Graham.Clark@lshtm.ac.uk

ABSTRACT

To expand the representation for phylogenetic analysis, ten additional complete *Entamoeba* small-subunit rRNA gene sequences were obtained from humans, non-human primates, cattle and a tortoise. For some novel sequences no corresponding morphological data were available, and we suggest that these organisms should be referred to as ribosomal lineages (RL) rather than being assigned species names at present.

To investigate genetic diversity and host specificity of selected *Entamoeba* species, a total of 91 new partial small subunit rRNA gene sequences were obtained, including 49 from *Entamoeba coli*, 18 from *Entamoeba polecki*, and 17 from *Entamoeba hartmanni*. We propose a new nomenclature for significant variants within established *Entamoeba* species. Based on current data we propose that the uninucleated-cyst-producing *Entamoeba* infecting humans is called *Entamoeba polecki* and divided into four subtypes (ST1-ST4) and that *Entamoeba coli* is divided into two subtypes (ST1-ST2). New hosts for several species were detected and while host specificity and genetic diversity of several species remain to be clarified, it is clear that previous reliance on cultivated material has given us a misleading and incomplete picture of variation within the genus *Entamoeba.*

Keywords: *Entamoeba*, parasite, phylogeny, ribosomal lineage, subtype, diversity

## INTRODUCTION

The genus *Entamoeba* comprises numerous unicellular, parasitic species found in humans, non-human primates, other vertebrates and invertebrates. Until recently, the detection, identification and assignment of *Entamoeba* organisms to species relied mainly on morphology. The introduction of molecular tools such as PCR and sequencing made it clear that definitive species identification and establishment of taxonomic relationships within the genus using microscopy only is not always possible (Clark and Diamond, 1997; Clark et al., 2006). This is due to a combination of factors, including overlap in morphological characteristics between species, morphological variation within species, the existence of mixed species *Entamoeba* infections, and limited knowledge of host specificity.

Molecular tools enable us to resolve many of the issues related to the identification, taxonomy, epidemiology and clinical significance of *Entamoeba* species without reliance on parasite cultures or experimental infections. Small subunit rRNA gene (SSU rDNA) sequence data are widely used for analysis of phylogenetic relationships between eukaryotic organisms and are available for several species of *Entamoeba*. Numerous papers have provided insights into the phylogeny and host specificity of *Entamoeba* based on sequence data (Silberman et al., 1999; Verweij et al., 2001; Ponce Gordo et al., 2004; Clark et al., 2006: Suzuki et al., 2007; Kobayashi et al., 2009; Stensvold et al., 2010; Levecke et al., 2010). Nevertheless, there are many described species of *Entamoeba* for which no molecular data are available. Conversely, it is equally likely that species of *Entamoeba* exist that have never been noted due to a lack of morphologically discriminating features.

This paper expands the *Entamoeba* phylogeny and infers taxonomic relationships from analysis of complete SSU rDNA sequences, many of which are from organisms not available in culture and some of which reveal unexpected diversity in the genus. In addition, the levels of intraspecific genetic diversity are examined for several species, with

implications for host range and the design of molecular detection tools.

## RESULTS

Ten new complete *Entamoeba* SSU rDNA sequences were obtained (Table 1) and phylogenetic analysis of these is presented in Fig. 1, together with 23 previously reported reference sequences. To obtain higher resolution in one part of the tree, a subset of new and reference sequences was aligned to include a larger number of unambiguously aligned positions in the analysis, which is presented in Fig. 2. Information on the origins of the complete *Entamoeba* SSU rRNA gene sequences generated in the study is listed in Table 1 while the primers used for amplification and sequencing are in Table 2.

The 91 partial SSU rDNA sequences obtained included many from the same species, which allowed investigation of intraspecific diversity and relationships. Four phylogenetic trees produced using these partial SSU rRNA gene sequences are displayed in Fig. 3, while Table 3 provides a list of all partial SSU rDNA sequences obtained.

### *Entamoeba* from cattle

To the knowledge of the authors, only uninucleated cysts of *Entamoeba* have been described in cattle (Stensvold et al., 2010) and these are all ascribed to the species *E. bovis*; however, the present study revealed that *Entamoeba* other than *E. bovis* can be found in this host (Figs. 1 and 2). Sequence CO4 was obtained from DNA in a faecal specimen from a Libyan cow and no morphological data were available. A 780 bp sequence differing at only 4 positions from the CO4 sequence was identified in faecal DNA from a cow in Estonia (Table 3), indicating that this *Entamoeba* lineage is probably widespread in cattle and without any geographic restriction. However, we have at this stage no further information on the host specificity or the genetic diversity of this *Entamoeba*, nor do we know what type of cysts it

produces. Hence, to assign the organism a (new) species name is not justifiable, and therefore we propose to use the designation *Entamoeba* Ribosomal Lineage (RL) 4 at present (see Discussion).

Microscopic examination of faecal concentrates from cattle samples Cow349 and Cow350 revealed that the vast majority of the cysts were uninucleated, but that there were also some cysts with 4 nuclei (and rarely even 6 or 8). The uninucleated cysts reacted with a monoclonal antibody (mAb) known to react with *E. histolytica* and *E. bovis* (Stensvold et al., 2010) but the tetranucleated cysts did not. The initial sequences obtained from these two samples were identical to *E. bovis* (Stensvold et al., 2010, and unpublished data). However, screening these DNAs using primers designed during the sequencing of CO4 revealed the presence of a second *Entamoeba* sequence. Full length sequences were obtained that proved to be distinct from both *E. bovis* and *Entamoeba* RL4 (Figs. 1 and 2). Since they cannot be linked to any valid species name for now, it is proposed that sequences Cow349.2 and Cow350 are assigned to *Entamoeba* RL2 (Table 1, Figs. 1 and 2).

***Entamoeba polecki***

Four clades of uninucleated cyst-producing *Entamoeba*s were described in humans by Verweij et al. (2001) from partial SSU rDNA sequences. Two clades consisted of sequences that were very similar to previously reported complete sequences from a pig (identified as *E. polecki*) and a monkey (identified as *E. chattoni*) (Silberman et al., 1999) while a third proved to be very similar to a sequence reported subsequently from an ostrich and named *E. struthionis* (Ponce Gordo et al. 2004). To date a complete SSU rDNA sequence from the fourth clade has not been available. In order to complete the picture and assess fully the nomenclature of this group, a complete SSU rDNA sequence of this variant was obtained

from a human faecal sample (J69; Table 1). With this sequence in hand, we can confirm that there are indeed four lineages of *E. polecki*-like *Entamoeba*s that are separated by high bootstrap support (Fig. 3). The four lineages also group into two pairs (Fig. 1) although without strong bootstrap support for this topology. Since there is very little evidence to support consistent morphological differences among the lineages or host specificity, we suggest that they are renamed *E. polecki* subtypes (ST) 1-4 (Fig. 1; see Discussion).

The relative prevalence of the subtypes in humans and their intra-subtype genetic diversity were incompletely known. To help clarify this, we analysed numerous partial SSU rDNA sequences and can now state that *E. polecki* is characterised by high intra-subtype genetic similarity (Fig. 3B). The majority of the new human samples characterised here belong to ST4; only one belongs to ST1, two are ST3, and none are ST2 (Fig. 3B).

**Other uninucleated-cyst-producing *Entamoeba*s**

Faecal samples from two langur species containing uninucleated cysts yielded two closely related sequences (Hulman and 09/1247) that were phylogenetically distant from *E. polecki*. The langur *Entamoeba* sequences form a lineage emerging from the clade of sequences obtained from cattle described earlier (Figs. 1 and 2) and are most closely related to *E. bovis*, although there is only modest boostrap support for this. In the absence of an existing species name, we propose that these organisms be referred to as *Entamoeba* RL3. Variation among sequences of *Entamoeba* RL3 is significant and when our data are combined with those of Levecke et al. (2010) the sequences appear to fall into two groups, but a larger sample is needed before conclusions regarding potential subtypes are made (Fig. 3C).

A uninucleated-cyst-producing *Entamoeba* in Vietnamese pigs was initially thought to be *E. polecki* based on its practically indistinguishable morphology (Blessmann et al.,

2002), but it was shown to have a distinct SSU rDNA (Clark et al., 2006). Partial sequences from 10 Vietnamese pigs were identical (unpublished data) and since it was not at the time detected in any other hosts, including humans living in close proximity to their infected pigs (Blessmann et al., 2002), the species name *E. suis* was resurrected to separate this *Entamoeba* from *E. polecki* (Clark et al., 2006) and to reflect its apparent host specificity. However, our detection of *E. suis* in a gorilla (Table 3) shows that this parasite is not restricted to pigs; the implications for the species name are not yet clear. The two sequences differed at only one position out of 590. Whether this discovery of *E. suis* in a non-porcine host will prove to be a rare finding remains to be determined.

**Tetranucleated-cyst-producing *Entamoeba*s**

The species name *E. nuttalli* was resurrected by Tachibana et al. (2007) to separate a pathogenic, but genetically distinct, tetranucleated-cyst-producing *Entamoeba* found in non-human primates from *E. histolytica*. The absence of morphological differences makes it impossible to distinguish between *E. nuttalli* and *E. histolytica* by microscopy either in stool or in tissue, so it is not possible to identify the agent responsible for invasive amoebiasis in non-human primates documented in earlier reports (Tachibana et al., 2007). Only little is known about the genetic diversity and host specificity of this parasite. *Entamoeba nuttalli* is being reported in an increasing number of non-human primate hosts (Levecke et al., 2010; Suzuki et al., 2007, 2008; Tachibana et al., 2007, 2009, Takano et al., 2007). The complete SSU rDNA sequence reported here from a colobus monkey was identical to that isolated from a rhesus macaque (Fig. 1), although the short tandem repeats in the tRNA gene arrays were found to differ between the two strains (unpublished observations).

At least 7 species of tetranucleated-cyst-producing *Entamoeba*s have been identified in chelonian hosts (Rodhain and Hoof, 1947; Ghosh, 1968; Philbey, 2006), but to our

knowledge only one species of *Entamoeba* has to date been described in leopard tortoises, namely *E. invadens* (Bradford et al., 2008). The sequence presented here (Oedla) clustered with *E. insolita* and an unnamed sequence from an iguana (Fig. 1). Cysts of *E. insolita* measure 12.8—19.5 µm (mean 15.7 µm) (Geiman and Wichterman, 1937) which is virtually the exact size range of the cysts found in the present tortoise. However, cysts of that size range could also be attributable to *E. invadens* and there are other named *Entamoeba*s for which cyst data are not available, for instance *E. testudinis* (Hartmann, 1910) and *E. barreti* (Taliaferro and Holmes, 1924). The bootstrap value uniting the new sequence with *E. insolita* is relatively low and dependent on the type of analysis, with a substantial 15% divergence in the SSU rDNA sequences. We propose that the Oedla sequence is assigned the name *Entamoeba* RL5 until more is known about the genetic diversity and host range of reptilian *Entamoeba*s. For the same reason, we suggest that *Entamoeba* sp. NIH:1091:1 from the iguana (AF149911) be renamed *Entamoeba* RL6 (Fig. 1).

Clark and Diamond (1997) investigated six isolates of *Entamoeba hartmanni* (four from humans and two from non-human primates) and found no evidence of variation using restriction fragment length polymorphism, suggesting that the genetic diversity of this species might be low. The present data support this hypothesis. All 17 partial sequences fell into a single clade and some sequences from humans were identical to those from non-human primates (Fig. 3D). Given that this parasite has been encountered only in primate hosts, we conclude that the genetic diversity within this species is low and that, for now, *E. hartmanni* should be considered a valid and well-defined species.

A 723 bp *Entamoeba* SSU rDNA sequence from a zebra (Table 3) showed almost complete identity to the *E. equi* sequence in GenBank from a horse, differing at only 1 position (data not shown). Although fixed stools from the zebra were available, no cysts were detected by microscopy, and therefore a morphological description of the cyst in this species

is still lacking.

### *Entamoeba*s producing octonucleate cysts

Two complete SSU rDNA sequences from octonucleated-cyst-producers were obtained (09/1246 and S2702) that are significantly different from those previously reported (Fig 1). *Entamoeba coli* comprises two major clades; the first (here named *E. coli* ST1) is represented by S2702 and GenBank accession number AF149915, and the second (*E. coli* ST2) by accession numbers AF149914 and AB444953. The S2702 sequence adds to the extensive genetic diversity seen in this species, which is estimated at around 13%. Indeed, based on sequence divergence it would be reasonable to consider *E. coli* ST1 and ST2 to be distinct species but at present there is no other justification for such a radical step.

No absolute subtype-related host-specificity was evident. However, 28 partial sequences falling into *E. coli* ST1 are identical and all of them are from humans (Fig. 3A, Table 3). Sequence variation is common in *E. coli* ST2, where both human and non-human sequences are present (Fig. 3A, Table 3). This suggests that clonal expansion of ST1 has happened relatively recently in humans. Only a few sequences of *E. coli* from non-human primates are available, and more data from this host group are needed to further establish whether subtype host specificity exists. To date the Drill1 sequence (Table 3) is the only *E. coli* ST1 from a non-human source.

Information on ethnicity and travel activity were available for only some of the human samples, but the sequences of human origin belonging to ST2 were primarily from individuals with a recent history of travelling to or living in Africa, Asia or South America. It is possible therefore, that ST2 is not common in Europe.

In contrast to S2702, the SSU rDNA sequence from the Phayre's Leaf Monkey (09/1246) clustered not with *E. coli* but with the *E. muris* sequence from a Mongolian Gerbil

(Kobayashi et al, 2009), although sequence identity was only 84%. At present we propose to assign 09/1246 to *Entamoeba* RL7.

## DISCUSSION

In this study we have faced a situation that probably will become more and more typical: the discovery – sometimes by pure serendipity – of new species or lineages based on molecular data in the absence of morphological data. This predicament makes it impossible to assess whether a valid species name for the newly identified organism is already available and precludes assigning a new one to the organism in question.

Therefore, until morphological data are available to establish a valid taxonomic name, we propose to use identification tags constructed as follows: 1. Well supported phylogenetic clusters within a defined species are assigned Arabic numerals identifying them as specific subtypes (STs). All sequences from a species must clearly fall into one of the STs. 2. Branches within phylogenetic trees that do not show a strong affinity with previously described species are assigned Arabic numerals identifying them as distinct ribosomal lineages (RLs). It is difficult to generalise about what constitutes ‘well supported’ and ‘strong affinity’, as these criteria will vary according to the amount of data available (partial or whole gene) and included in the alignment, and the method of analysis employed. Different boundaries may be appropriate in different circumstances. In general, subtypes will be defined using partial gene sequences, as in our cases, while we strongly suggest that assignment of new ribosomal lineages should be made using complete SSU rDNA sequences only. In our data, ‘well supported’ clusters designated as subtypes all have bootstrap support of 95% or more in PhyML, while a complete gene sequence showing bootstrap support of less than 80% in PhyML for affinity to another lineage should be considered for identification as a new ribosomal lineage. However the latter need not always be the case. For example,

246 *Entamoeba* RL7 was designated a new lineage even though it showed bootstrap support of
247 96% for a specific relationship with *E. muris*. The rationale for assigning it to a new
248 ribosomal lineage rather than a new subtype of *E. muris* is that they showed 16% sequence
249 divergence and there is very low sampling in this region of the tree. Overall, this means that,
250 unfortunately, designation of STs and RLs will be somewhat subjective and context
251 dependent, and supporting arguments will have to be provided. As an additional illustration,
252 we recently published complete SSU rDNA sequences and a phylogenetic analysis for *E.*
253 *bovis* (Stensvold et al., 2010). It was clear that *E. bovis* could be found in various hosts, such
254 as cattle, sheep and reindeer. However, a SSU rDNA sequence from a uninucleated-cyst-
255 producing *Entamoeba* infecting a roe deer appeared to represent a separate lineage based on
256 sequence divergence and cyst size, and we excluded it from *E. bovis.* We now suggest that
257 this lineage is assigned the name *Entamoeba* RL1 (Fig. 1). Future studies of morphology,
258 genetic diversity and host specificity of this organism, or the others identified by a RL
259 number, may eventually allow assignment of a (new) species name.

260 Mixed infections also make species assignment difficult, if not impossible, as
261 illustrated by samples Cow349 and Cow350. Initially, *E. bovis* sequences were obtained from
262 these samples using broad specificity primers. These were easily readable and gave no
263 indication of a mixed infection; only a small amount of “background” was present under the
264 *E. bovis* peaks in chromatograms. Yet the use of alternative primers revealed the presence of
265 a distinct SSU rDNA, which we propose to call *Entamoeba* RL2. The sample contained
266 primarily uninucleated cysts but also scant cysts containing 4 or more nuclei. The
267 predominance of the uninucleated cysts in the sample and the *E. bovis* sequence in the
268 chromatograms supports the hypothesis that the new RL2 sequences may be attributable to
269 the tetranucleated cysts. It is not possible to prove this link at present; however, screening of
270 cow samples in which no tetranucleated cysts were seen gave no evidence of this novel

*Entamoeba* sequence. Future combined molecular and morphological studies of *Entamoeba* in ruminant hosts are needed in order to establish the relationship between the cysts observed and the sequences obtained from samples.

Uninucleated-cyst-producing *Entamoeba* infections have been reported in humans across the world but with greatly varying prevalence (Desowitz and Barnish, 1986; McMillan and Kelly, 1970; Chacín-Bonilla, 1992; Blessmann et al., 2002). Studies have reported a 70% prevalence in the wild in both chimpanzees and baboons (Jackson et al., 1990; Muehlenbein, 2005). Uninucleated-cyst-producing *Entamoeba*s from humans and non-human primates have usually been assigned to *E. polecki* and *E. chattoni*, respectively. *Entamoeba chattoni* was first described by Chatton (1912) and named by Swellengrebel (1914) who found it in a rhesus monkey (*Macaca rhesus*), while *E. polecki* was originally described in pigs (von Prowazek, 1911). Kessel and Johnstone (1949) reported finding cysts of *E. chattoni* in both rhesus monkeys and humans, but concluded that the morphologies of *E. chattoni* and *E. polecki* were so similar they might represent the same species. Likewise, Sumardjo and Joe (1953) found the morphology of *E. chattoni* and *E. polecki* to be almost identical, except that the karyosome of *E. polecki* is commonly large and granulated whereas it is smaller and more delicate in *E. chattoni.* Sargeaunt et al. (1992) found seven cases of human infection all but one of which had contact with non-human primates, but in many cases no patient contact with infected pigs or non-human primates is found (Chacín-Bonilla, 1992; Blessmann et al., 2002).

In 2001, Verweij et al. investigated genetic variation within human uninucleated-cyst-producing *Entamoeba*s using partial SSU rDNAs and reported that four clades existed. They concluded that all were variants of *E. polecki* and should be called *E. polecki*-like. Those results have been confirmed in the present work and we propose to call the four clades *E. polecki* subtypes 1-4. Three of the subtypes have also been found in other host species, namely those formerly assigned to *E. polecki* (pigs; ST1), *E. chattoni* (non-human primates;

ST2) and *E. struthionis* (pigs and ostriches; ST3), but the fourth *E. polecki* subtype (ST4, represented by sequence J69) has so far only been found in humans. ST2 has been found only in primates, but it appears that the host specificity of ST1 and ST3 is low. Indeed, a sample from a Rhea was shown to be a mixed infection with ST1 and ST3 (Table 3), the first time a mixed-subtype *E. polecki* infection has been documented. The apparent restriction of *E. polecki* ST4 to humans implies that infections due to ST4 are unlikely to be of zoonotic origin.

It is important to note that all the new human *E. polecki* samples analysed are from Europe and our results may not reflect the subtype distribution in other regions. However, although diagnosed in Sweden those individuals with ST4 for whom data were available had all been travelling in Asia or Africa (Table 3). In addition, the human infections reported by Blessmann et al. (2002) in a Vietnamese population were all ST4 also (based on 10 unpublished partial sequences). Taken together with the data of Verweij et al. (2001), the evidence suggests that this subtype is the most common in humans and is widely distributed.

Traditionally, uninucleated-cyst-producing *Entamoeba*s from non-human primates have been assigned to *E. chattoni* (now *E. polecki* ST2). We here have reported a new, uninucleated-cyst-producing *Entamoeba* lineage in non-human primates – *Entamoeba* RL3. Since there is considerable size overlap between the cysts from langurs and cysts reported in the literature as *E. chattoni*, it is not possible to know whether previous morphology-based reports of *E. chattoni* were in fact *E. polecki* ST2 or *Entamoeba* RL3. We do not believe that *Entamoeba* RL3 is a subtype of *E. bovis*: the sequence divergence is substantial, the size of the cysts does not match the description of *E. bovis* cysts (Stensvold et al., 2010), and this lineage has been found only in langurs (several species) and a colobus monkey (all Subfamily Colobinae). Further assessment of host specificity, morphology and genetic variation is needed before assigning a species name.

Numerous octonucleated-cyst-producing *Entamoeba*s have been described in the literature, e.g. *E. caviae*, *E. cuniculi*, and *E. wenyoni*, but sequence data are available only for *E. coli* and *E. muris* and therefore it is not possible to assign a species name to *Entamoeba* RL7 (09/1246). If *E. muris* proves to be a complex of subtypes, as in *E. coli*, the 09/1246 sequence could very well represent a second *E. muris* subtype with different host specificity. Hence, whether the taxon *E. muris* will need re-structuring depends on future molecular analyses of octonucleated-cyst-producing *Entamoeba*s, especially in primate and rodent hosts.

Octonucleated cysts of *Entamoeba* found in humans and non-human primates have exclusively been assigned to *E. coli* and, as a result, *E. coli* has been reported only in primates. In addition to a monkey sequence clustering with *E. muris* (*Entamoeba* RL7), we also found a rodent sequence, from a chinchilla, that differs from a gorilla *E. coli* ST2 sequence at only one position out of almost 1000. In 1950, Neal reported that mice and rats could be infected experimentally with *E. muris*, but not with cysts or trophozoites of *E. coli* from human faeces or cultures. This was used as part of the justification for *E. muris* being a separate species. However, on the basis of the present data one could speculate that host specificity is not absolute. Therefore, the experimental infections with *E. coli* in the study by Neal (1950) should be interpreted cautiously, since the choice of *E. coli* isolate might be influencing the outcome of the study – perhaps only *E. coli* ST2 can infect rodents for example. Indeed, before Neal's work, Kessel (1923) and Regendanz (1929) both reported successful experimental infection of rodents with *E. coli*, so it might be conjectured that they were working with *E. coli* ST2 while Neal was using *E. coli* ST1. Wider sampling of rodent *Entamoeba*s and other octonucleated-cyst-producing infections may clarify the situation.

Cyst size variation has been described for *E. coli* in several studies, and bi- or tri-modal distributions of cyst size have been reported on more than one occasion (Matthews,

1919; Dobell, 1919). The *E. coli* in the present study most likely belong primarily to what would be the "small races" of *E. coli* (Matthews, 1919; Dobell, 1919), since most samples were originally mistaken for *E. histolytica*/*E. dispar*. Many such sequences belong to *E. coli* ST1, although a few are ST2. It is possible that cyst size is not related to subtype, but only accumulation of sequence and cyst size data for the same samples plus examination of "large race" *E. coli* will help clarify this point.

Of the protists found in the human intestinal tract, *E. coli* is one of the most commonly found, and it is generally considered non-pathogenic. A few reports have drawn the attention to cases of gall bladder disease (Kalk and Wildhirt, 1954; Witte, 1956; Geyer, 1959; von Meyenfeldt et al., 2007) and diarrhoea (Corcoran et al., 1991; Wahlgren, 1991) that might be attributable to *E. coli*. The ability of *E. coli* to phagocytose erythrocytes has been documented and varies among strains (Dobell, 1936). Recently, *E. coli* in Colombian school children appeared to be an indicator of poor nutritional status (Boeke et al., 2010). Because of the degree of genetic diversity in this species, future studies on its potential role in disease should note the subtype of the organism so that any links between phenotype and subtype can be explored.

The new lineages of *Entamoeba* detected here have implications for correct speciation by microscopy and suggest that molecular tools are the only way to accurately identify the organisms present in a sample. Nevertheless, molecular tools are not without their problems. Accumulation of information on intra-specific sequence variation is necessary in order to design sensitive and specific of primer for PCR-based detection. This is exemplified by our findings on *E. hartmanni.* Primers for detecting *E. hartmanni* were published recently and used by Suzuki et al. (2008). The authors found that 5/47 non-human primates were positive for *E. hartmanni* and the sequences they obtained were all similar to the only reference sequence in GenBank (AF149907), which was therefore also the sequence used to design the

primers. Some of the *E. hartmanni* SSU rDNA sequence variants detected in the present study might not have been amplifiable using those primers as sequence variation exists in both of the primer binding regions, raising questions about the prevalence data based on those primers or others based on only a single sequence.

**Concluding remarks**

To further expand our understanding of the taxonomy and epidemiology of *Entamoeba,* future studies should focus on PCR-based screening of faecal samples from various hosts. However, faecal samples subject to DNA extraction should also be fixed for subsequent microscopic examination in order to allow correlation of molecular and morphological data. This good intention may still not always provide the data required, as illustrated by *E. equi*, but for classical species descriptions such information is still essential.

Our results highlight the need for molecular data in order to investigate the epidemiology of *Entamoeba*, since observation of cyst and trophozoite morphology can lead to erroneous species identification and conclusions regarding host specificity. Our recent data on *E. bovis* showed that the grouping of *Entamoeba* species based on cyst nuclear number does not always reflect phylogenetic relationships (Stensvold et al., 2010), and the present data give further support to this assertion. We also show that organisms in the same host with morphological identity may in fact be hiding substantial cryptic diversity.

It is important to emphasise that the generation, re-assignment or resurrection of species names should be based on extensive studies of host specificity and genetic diversity, preferably supported by morphological information also. At present our data provide evidence of both host-specificity and a lack thereof for different species, subtypes and lineages of *Entamoeba*. We feel certain that many novel species of *Entamoeba* remain to be identified and we hope that our proposed nomenclature approach will be useful in dealing

with the new data.

## METHODS

### Samples and sequences

Samples included in the study were from humans, non-human primates, other mammals and a few non-mammalian hosts. Since samples were collected in different ways, from different populations and for different initial purposes it is not possible to generate prevalence data. Most of the human samples had been shown to be microscopy-positive for *Entamoeba* during routine laboratory analysis. All of the *E. coli*, *E. hartmanni* and *E. polecki* isolates with the prefix EM or UNE were from humans and represent cysts initially mistaken for *E. histolytica* or *E. dispar* in routine parasitological analyses in local laboratories in Sweden; these were subsequently re-evaluated at Smittskyddsinstitutet in Stockholm and definitively identified to species level. For some of the isolates, information on recent travel activity was available (Table 3). All human samples had been anonymised prior to inclusion in the study so that only anamnestic details were available.

The non-human samples originated from samples either submitted to routine screening for potential pathogens or obtained during prospective studies looking for parasitic protists, including *Entamoeba*; not all of them were submitted for microscopic analysis.

Most of the information available on the sample origins of complete and partial *Entamoeba* SSU rDNA sequences obtained during this study is displayed in Tables 1 and 3. Additional information on those used for complete gene sequencing is given below. All DNAs tested were from single individual or animal samples, apart from DNA from pig faeces. Genomic DNAs extracted from pig stool samples used for a previous study (Stensvold et al., 2009) were pooled in groups of five and tested by PCR. Specific PCR products were sequenced for three of the pools (Table 3).

## DNA extraction and DNA sequencing

Most of the sequences were obtained from PCR products amplified using DNA extracted directly from faecal samples with the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany). Some sequences were obtained using DNA extracted from purified cyst suspensions as described (Table 3; Lebbad et al., 2008). The *E. nuttalli* 360 sequence was obtained using DNA extracted from cultures with the PureGene core kit A (Qiagen). The organisms were grown in medium LYSGM (Stechmann et al., 2008) with 5% adult bovine serum. Purification and sequencing of PCR products was as previously described (Stensvold et al., 2006; Stensvold et al., 2010).

In most cases, initial sequence data were obtained using broad specificity primers designed to amplify all *Entamoeba* SSU rDNAs: ENTAM 1/2, 542/3, and ENTAGEN F/R (Table 2). In a few cases, where mixed *Entamoeba* infections were present, PCR products were cloned using the TOPO-TA Cloning® Kit (Invitrogen) before being sequenced; these exceptions are noted in Tables 1 and 3. From this initial sequence information, and where it was though important to obtain the complete gene sequence, specific primers were designed and primer walking used to obtain the complete sequence (Table 2).

Partial SSU rDNA sequences from a large number of samples (Table 3) were obtained using the broad specificity primer pairs mentioned above, and in some cases these were supplemented by sequencing of additional gene regions using other primers in Table 2 as indicated.

## Samples yielding complete *Entamoeba* SSU rDNA sequences

*Entamoeba polecki* SSU rDNA sequence J69 was obtained using DNA extracted from a faecal specimen submitted by a 7-year-old Somali girl who had lived in the Netherlands for

4 years.

*Entamoeba coli* SSU rDNA sequence S2702 was obtained using DNA extracted from a non-mucoid, non-bloody faecal specimen submitted by a Nigerian HIV-positive female patient with diarrhoea. This sample was positive by microscopy for *Entamoeba coli*, *Chilomastix mesnili* and *Blastocystis* sp.

*Entamoeba* sp. SSU rDNA sequence Hulman was obtained using DNA extracted from a faecal sample from one of several captive Hanuman Langurs (syn. Grey Langur) (*Semnopithecus entellus*) in the Zoologischer Garten Neunkirchen (Germany). All five animals were microscopy-positive for uninucleated cysts, as well as octonucleated cysts, and all sequences obtained were identical.

*Entamoeba* sp. SSU rDNA sequences 09/1246 and 09/1247 were obtained using DNA extracted from faecal samples from a Phayre's Leaf Monkey (*Trachypithecus phayrei*) (09/1246) and either a Javan Langur (*Trachypithecus auratus*) or a Silvery Lutung (*Trachypithecus cristatus*) (09/1247) from Twycross Zoo, Warwickshire, UK, respectively. The latter two animals were housed together. Examination of faecal concentrates revealed octonucleated cysts (09/1246) and mixed *Entamoeba* infections with uni-, tetra-, and octonucleated cysts as well as *Blastocystis* (09/1247).

The *Entamoeba nuttalli* SSU rDNA sequence 360 was obtained using DNA extracted from cultures established using faeces of a Mantled Guereza (*Colobus guereza kikuyuensis*) in “La Vallée des Singes”, Romagne, France, suffering from non-dysenteric diarrhoea.

*Entamoeba* SSU rDNA sequence CO4 was obtained using DNA extracted from faeces of a Libyan cow. No morphological data are available.

*Entamoeba* SSU rDNA sequences Cow349.2 and Cow350 were detected in DNA extracted from cyst preparations from two Swedish cows. These were positive for mainly uninucleated cysts but also, conspicuously, a few tetranucleated cysts were seen. Both cows

were also positive for *E. bovis* as detected by PCR and sequencing.

*Entamoeba* SSU rDNA sequence Oedla was obtained using DNA extracted from purified cysts from a leopard tortoise (*Geochelone pardalis*) in Eskilstuna Zoo, Sweden. The tetranucleated cysts observed also did not react with a monoclonal antibody (mAb) known to react with *E. histolytica* and *E. bovis* (Stensvold et al., 2010). A few of the cysts only had one nucleus, quite a few had 2, but most were tetranucleate and generally the nuclei were eccentrically located. The sequence obtained was partly from cloned products and partly from direct sequencing of PCR products; no differences were seen in the regions of overlap.

**Samples for partial SSU rDNA sequencing**

49 *E. coli* partial sequences from humans (n=45), non-human primates (n=3) and a chinchilla (n=1), 17 partial *E. hartmanni* sequences from humans (n=9) and non-human primates (n=8), and 18 partial *E. polecki* sequences from humans (n=12), a non-human primate (n=1), pigs (n=3) and a nandu (n=2, from one sample) were obtained. Other partial *Entamoeba* sequences were obtained from an Estonian cow (n=1), langurs (n=4), a zebra (n=1), and a gorilla (n=1).

**Sequence assembly, alignment and phylogenetic analyses**

PCR products were sequenced in both directions using dideoxynucleotide chain terminator methods. In most cases, sequences were edited and assembled in sample-specific databases using the Staden software package (http://staden.sourceforge.net/). Sequences were deposited in the NCBI nucleotide database with Accession Nos. FR868356-FR868456.

Complete sequences were aligned with reference sequences from GenBank using the online alignment tool MUSCLE (Edgar, 2004; http://www.ebi.ac.uk/Tools/muscle). The output was imported into MEGA 4.0 (Kumar et al., 2008) and edited manually to produce an

alignment of 1,446 unambiguous positions for all 32 taxa. Sequence divergence percentages were calculated using MEGA using the aforementioned alignment.

Phylogenetic analyses were performed as described previously (Stensvold et al., 2010) using distance (Neighbor Joining; MEGA 4.0), maximum likelihood (PhyML 3.0; Guindon and Gascuel, 2003) and Bayesian (MrBayes 3.1.2; Huelsenbeck and Ronquist, 2001) methods. Bayesian and maximum likelihood analysis used a General Time Reversible (GTR) model of nucleotide substitution with four categories of among-site rate variation and the proportion of invariant sites, as in previous phylogenetic analyses of *Entamoeba* SSU rDNA sequences. Statistical support for distance and maximum likelihood trees was evaluated using bootstrapping (1000 replicates). Bayesian analysis used four Markov chain Monte Carlo (MCMC) strands, 1,000,000 generations, with trees sampled every 100 generations. In every case the average standard deviation of split frequencies was less than 0.01. A consensus tree was produced after excluding an initial burn-in of 25% of the samples, as recommended.

To obtain a more precise view of relationships among the cattle and langur isolates, a separate alignment including 1,794 unambiguous positions was generated for a selection of the complete sequences and analysis was performed using the same algorithms. Likewise, the same approach was used to generate alignments and trees for some species using partial sequences, to investigate the population structure. Not all partial sequences covered the same region of the gene and so although all sequences could be assigned unambiguously to species/lineage/subtype not all were able to be included in the phylogenetic analyses.

ACKNOWLEDGEMENTS

Lis Lykke Wassmann, Department of Microbiology and Diagnostics, Statens Serum Institut, Copenhagen, Denmark and Heidrun von Thien, Bernhard Nocht Institute for Tropical

Medicine, are thanked for excellent technical assistance.

Bitte Ljungström, Parasitology Unit, Vidilab, Enköping, Sweden is thanked for providing animal faecal samples. Dr. Sunday Eme Onuoha, College of Medicine, University of Lagos, Nigeria, Dr. Simone Caccio, Department of Infectious, Parasitic and Immunomediated Diseases, Istituto Superiore di Sanità, Rome, Italy, Dr. Brian Lassen, Section of Parasitology, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, Tartu, Estonia and Derya Taner-Mulla, London School of Hygiene and Tropical Medicine, are all thanked for providing purified faecal DNA. A portion of this work formed part of an MSc thesis submitted by PL.

REFERENCES

**Blessmann J, Van Linh P, Nu PA, Thi HD, Muller-Myhsok B, Buss H, Tannich E** (2002) Epidemiology of amebiasis in a region of high incidence of amebic liver abscess in central Vietnam. Am J Trop Med Hyg **66**:578–583

**Boeke CE, Mora-Plazas M, Forero Y, Villamor E** (2010) Intestinal protozoan infections in relation to nutritional status and gastrointestinal morbidity in Colombian school children. J Trop Pediatr (In Press)

**Bradford CM, Denver MC, Cranfield MR** (2008) Development of a polymerase chain reaction test for *Entamoeba invadens*. J Zoo Wildl Med **39:**201-207

**Chacín-Bonilla L (**1992) *Entamoeba polecki*: human infections in Venezuela. Trans R Soc Trop Med Hyg **86**:634

**Chatton, E** (1912) Sur quelques genres d'amibes libres et parasites. Synonymies, homonymie, impropriété. Bull Soc Zool France **37:**109-115

**Clark CG, Diamond LS** (1997) Intraspecific variation and phylogenetic relationships in the genus *Entamoeba* as revealed by riboprinting. J Euk Microbiol **44:**142-154

**Clark CG, Kaffashian F, Tawari B, Windsor JJ, Twigg-Flesner A, Davies-Morel MCG, Blessmann J, Ebert F, Peschel B, Van AL, Jackson CJ, Macfarlane L, Tannich E** (2006) New insights into the phylogeny of *Entamoeba* species provided by analyses of four new small-subunit rRNA genes. Int J Syst Evol Microbiol **56:**2235-2239

**Corcoran GD, O'Connell B, Gilleece A, Mulvihill TE** (1991) *Entamoeba coli* as possible cause of diarrhoea. Lancet **338:**254

**Desowitz RS, Barnish G** (1986) *Entamoeba polecki* and other intestinal protozoa in Papua New Guinea highland children. Ann Trop Med Parasitol **80:**339-402

**Dobell C** (1919) The Amoebae Living in Man. A Zoological Monograph. J. Bale, Sons, and Danielson, London.

**Dobell C** (1936) Researches on the intestinal protozoa of monkeys and man. VIII. An experimental study of some simian strains of "*Entamoeba coli*". Parasitology **28:**541-593

**Edgar RC** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-7

**Elwood HJ, Olsen GJ, Sogin ML** (1985) The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. Mol Biol Evol **2**:399-410

**Geiman QM, Wichtermann R** (1937) Intestinal protozoa from Galapagos tortoises (with description of three new species). J Parasitol **23:**331-347

**Geyer E** (1959) [Is there an *Entamoeba coli* cholecystopathy?] Z Gesamte Inn Med. **14:**968-972

**Ghosh TN** (1968) Observations on the type specimen of *Entamoeba serpentis* (Cunha and Fonseca, 1917). J Protozool **15:**164-166

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

**Hartmann M** (1910) Ueber eine neue Darmamoebe, *Entamoeba testudinis*, n. sp. Mem Inst Oswaldo Cruz **2:**3-10

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

**Jackson TFHG, Sargeaunt PG, Visser PS, Gathiram V, Suparsad S, Anderson CB** (1990) *Entamoeba histolytica*: naturally occurring infections in baboons. Arch Invest Med (Mex) **21** Suppl **1:**153-156

**Kalk H, Wildhirt E** (1954) Ueber das Vorkommen von Amöben im Duodenalsaft und in der Galle. Med Klin **49:**1466-1468

**Kessel JF, Johnstone HG** (1949) The occurrence of *Endamoeba polecki*, Prowazek 1912, in *Macaca mulatta* and in man. Am J Trop Med Hyg **29:**311-317

**Kessel JF** (1923) Experimental infection of rats and mice with the common intestinal amoebae of man. Univ Calif Publ Zool **20:**409-430

**Kobayashi S, Suzuki J, Takeuchi T** (2009) Establishment of a continuous culture system for *Entamoeba muris* and analysis of the small subunit rRNA gene. Parasite **16:**135-139

**Kumar S, Nei M, Dudley J, Tamura K (**2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform **9:**299–306

**Lebbad M, Ankarklev J, Tellez A, Leiva B, Andersson JO, Svärd S** (2008) Dominance of *Giardia* assemblage B in León, Nicaragua. Acta Trop **106:**44-53

**Levecke B, Dreesen L, Dorny P, Verweij JJ, Vercammen F, Casaert S, Vercruysse J, Geldhof P** (2010) Molecular identification of *Entamoeba* spp. in captive nonhuman primates. J Clin Microbiol **48:**2988-2990

**Matthews JR** (1919) A mensurative study of the cysts of *Entamoeba coli*. Ann Trop Med Parasitol **12:**259-272

**McMillan B, Kelly A** (1970) *Entamoeba polecki* von Prowazek, 1912 in New Guinea. Trans

R Soc Trop Med Hyg **64:**792-793

**Muehlenbein MP** (2005) Parasitological analyses of the male chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo, Kibale National Park, Uganda. Am J Primatol **65:**167-179

**Neal RA** (1950) An experimental study of *Entamoeba muris* (Grassi, 1879); its morphology, affinities and host-parasite relationship. Parasitology **40:**343-365

**Philbey AW** (2006) Amoebic enterocolitis and acute myonecrosis in leopard tortoises (*Geochelone pardalis*). Vet Rec **22:**567-569

**Ponce Gordo F, Martinez Diaz RA, Herrera S** (2004) *Entamoeba struthionis* n. sp. (Sarcomastigophora: Endamoebidae) from ostriches (*Struthio camelus*). Vet Parasitol **119:**327-335

**Regendanz P** (1929) Über die Übertragung der *Entamoeba histolytica*, *Entamoeba coli* und *Dientamoeba fragilis* auf Ratten. Zentralbl Bakteriol Orig **111:**412-419

**Rodhain J, van Hoof MT** (1947) *Entamoeba knowlesi* n.sp. parasite de deux tortues: *Terrapina cinosternoides* et *Platysternum megacephalum*. Ann Parasitol Hum Comp **22:**129-137

**Sargeaunt PG, Patrick S, O'Keeffe D** (1992) Human infections of *Entamoeba chattoni* masquerade as *Entamoeba histolytica*. Trans R Soc Trop Med Hyg **86**:633-634

**Silberman JD, Clark CG, Diamond LS, Sogin ML** (1999) Phylogeny of the genera *Entamoeba* and *Endolimax* as deduced from small-subunit ribosomal RNA sequences. MolBiol Evol **16:**1740-175

**Stechmann A, Hamblin K, Pérez-Brocal V, Gaston D, Richmond GS, van der Giezen M, Clark CG, Roger AJ** (2008) Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr Biol **18**:580-585 (supplementary information)

**Stensvold R, Brillowska-Dabrowska A, Nielsen HV, Arendrup MC** (2006) Detection of *Blastocystis hominis* in unpreserved stool specimens by using polymerase chain reaction. J

621 Parasitol **92:**1081-1087

622 **Stensvold CR, Lebbad M, Clark CG** (2010) Genetic characterisation of uninucleated cyst-
623 producing *Entamoeba* spp. from ruminants. Int J Parasitol **40:**775-778

624 **Stensvold CR, Alfellani MA, Norskov-Lauritsen S, Prip K, Victory EL, Maddox C,**
625 **Nielsen HV, Clark CG** (2009) Subtype distribution of *Blastocystis* isolates from
626 synanthropic and zoo animals and identification of a new subtype. Int J Parasitol **39:**473-479

627 **Sumardjo B, Joe LK** (1953) Uni- and binuclear cysts, morphologically resembling
628 *Entamoeba polecki* Prowazek, 1912, found in an Indonesian boy. Doc Med Georg Trop **5:**1-4

629 **Suzuki J, Kobayashi S, Murata R, Yanagawa Y, Takeuchi T** (2007) Profiles of a
630 pathogenic *Entamoeba histolytica*-like variant with variations in the nucleotide sequence of
631 the small subunit ribosomal RNA isolated from a primate (De Brazza's guenon). J Zoo Wildl
632 Med **38:**471-474

633 **Suzuki J, Kobayashi S, Murata R, Tajima H, Hashizaki F, Yanagawa Y, Takeuchi T.**
634 (2008) A survey of amoebic infections and differentiation of an *Entamoeba histolytica*-like
635 variant (JSK2004) in nonhuman primates by a multiplex polymerase chain reaction. J Zoo
636 Wildl Med. **39**:370-379

637 **Swellengrebel NH** (1914) Dierlijke entamoeben uit Deli. Geneesk Tijdsschr v Nederl.-Indië
638 **54:**420-426

639 **Tachibana H, Yanagi T, Pandey K, Cheng XJ, Kobayashi S, Sherchand JB, Kanbara H**
640 (2007) An *Entamoeba* sp. strain isolated from rhesus monkey is virulent but genetically
641 different from *Entamoeba histolytica*. Mol Biochem Parasitol **153:**107-114

642 **Tachibana H, Yanagi T, Akatsuka A, Kobayashi S, Kanbara H, Tsutsumi V (**2009)
643 Isolation and characterization of a potentially virulent species *Entamoeba nuttalli* from
644 captive Japanese macaques. Parasitology **136**:1169-1177

645 **Takano J, Narita T, Tachibana H, Terao K, Fujimoto K** (2007) Comparison of

646 *Entamoeba histolytica* DNA isolated from a cynomolgus monkey with human isolates.
647 Parasitol Res **101**:539-456
648 **Taliaferro WH, Holmes FO** (1924) *Endamoeba barreti*, n. sp., from the turtle, *Chelydra*
649 *serpentina*; a description of the amoeba from the vertebrate host and from Barret and Smith's
650 cultures. Am J Hyg **4:**160-168
651 **Verweij JJ, Polderman AM, Clark CG** (2001) Genetic variation among human isolates of
652 uninucleated cyst-producing *Entamoeba* species. J Clin Microbiol **39:**1644-1646
653 **von Meyenfeldt FM, Mantel SF, Gouma DJ, Gulik TM van** (2007) [Tumors in the
654 gallbladder: a possible differentiation between malignant and benign tumors]. Ned Tijdschr
655 Geneeskd **151:**1049-1054
656 **von Prowazek S** (1911) *Entamoeba*. Arch Protistenkd **25**:273–274
657 **Witte S** (1956) Der praktische Wert der zytologischen Untersuchung des Duodenalinhaltes.
658 Mat Med Nordm **8:**23-30
659 **Wahlgren M** (1991) *Entamoeba coli* as cause of diarrhoea? Lancet **337:**675

Table 1. Information on the complete *Entamoeba* SSU rDNA sequences generated in the study.

| Sequence ID | Host | Cyst size (mean) | Number of nuclei in cysts | New nomenclature | Accession no. |
|---|---|---|---|---|---|
| J69 | *Homo sapiens* | N/A | N/A | *Entamoeba polecki* ST4 | FR686357 |
| Hulman | *Semnopithecus entellus* | 9.2—15.4 µm (12.34 µm ± 1.83 µm, 50 cysts) | 1 | *Entamoeba* RL3 | FR686358# |
| 09/1247 | *Trachypithecus auratus* or *T. cristatus*[1] | N/A | N/A | *Entamoeba* RL3 | FR686359 |
| 09/1246 | *Trachypithecus phayrei* | N/A | N/A | *Entamoeba* RL7 | FR686360 |
| CO4 | *Bos taurus* | N/A | N/A | *Entamoeba* RL4 | FR686361 |
| Cow349.2 | *Bos taurus* | N/A | 1 or 4[2] | *Entamoeba* RL2 | FR686362* |
| Cow350 | *Bos taurus* | N/A | 1 or 4 [2] | *Entamoeba* RL2 | FR686363* |
| S2702 | *Homo sapiens* | N/A | 8 | *Entamoeba coli* ST1 | FR686364 |
| Oedla | *Geochelone pardalis* | 12.0—19.5 µm (14.7 µm ± 1.33 µm, 100 cysts) | 4[3] | *Entamoeba* RL5 | FR686365* |
| 360 | *Colobus guereza kikuyuensis* | N/A | 4 | *Entamoeba nuttalli* | FR686356 |

[1] These two hosts were housed together so the source of the sample is not identifiable.

[2] The sample was a mixture of predominantly uninucleated cysts and a few tetra-nucleated cysts. Rarely, cysts with 2, 6 or 8 nuclei were seen.

[3] The sample contained cysts with varying number of nuclei, mostly 4 or 2 nuclei were seen with a few being uninucleate.

N/A = information not available; in primate cases this is because of mixed *Entamoeba* infections.

# = sequence obtained from cloned DNA. *Sequences obtained from purified cyst preparations.

672 Table 2. Primers used for amplification and/or sequencing.
673

| Primer name | Primer sequence (5'—3') | Reference | PCR/Sequencing use[1] |
|---|---|---|---|
| RD5 | ATCTGGTTGATCCTGCCAGT | Clark et al., 2006 | 1,2,3,4,5,6,7,8,9 |
| RD3 | ATCCTTCCGCAGGTTCACCTAC | Clark et al., 2006 | 1,2,3,4,5,6,7,8,9 |
| ENTAM1 | GTTGATCCTGCCAGTATTATATG | Verweij et al., 2001 | 1,2,4,5,8,9 |
| ENTAM2 | CACTATTGGAGCTGGAATTAC | Verweij et al., 2001 | 1,2,4,5,8,9 |
| 542 | GTTGATCCTGCCAAGTATTATATGCT | Clark et al., 2006 | 3 |
| 543 | GACTATTGGAGCTGGAATTACCG | Clark et al., 2006 | 3 |
| ENTAGEN_F | ACTTCAGGGGGAGTATGGTCAC | Present study | 6 |
| ENTAGEN_R | CAAGATGTCTAAGGGCATCACAG | Present study | 6 |
| Uninuc_400F | AGGTAGTGACGATAATTAATAG | Present study | 1 |
| Uninuc_1630R | TTAATCCCAGTCATGTACACC | Present study | 1 |
| Uninuc_1500F | GCTACAATGGAATTTATAGAGAGT | Present study | 1 |
| Uninuc_1050F | ATTGTTACTCTCTTATTCAGGA | Present study | 1 |
| Entcoli_100F | GAAGCTGCGAACGGCTCATTAC | Present study | 2 |
| Entcoli_500F | GGCGCGAAAATTACCCAATC | Present study | 2,4 |
| Entcoli_390R | CACCTTGGTAAGCCACTACC | Present study | 2 |
| Entcoli_800F | CAAAATCAAGGCGCTTAAAGC | Present study | 4 |
| Entcoli_1000R | CCACCTCTCCCGTTCCTATC | Present study | 2,4 |
| Entcoli_1000F | GGAATTCCATGATCGTTTCGA | Present study | 2,4 |
| Entcoli_1700R | ACAGACCTGTTATTGCTTGAC | Present study | 2,4 |
| Entcoli_NIG | GACACATCTTTAATCTTTCCGGG | Present study | 2 |
| hulman-S21 | TTTATACTTCACGGCCATCAG | Present study | 3 |
| hulman-AS21 | CAAGAGACACCAAAAGGCATC | Present study | 3 |

| | | | |
|---|---|---|---|
| 1247hulman_1700F | CTCTGTTGGAGTGGTAAGAATTCTC | Present study | 5 |
| 1247hulman_1550F | GTTAATTTGTGTTTATGATTTCGGTC | Present study | 5 |
| 1247hulman_430F | AGGAGATGCCGTATGGTATTTC | Present study | 5 |
| EstCowEnt_1690R | ATTCCAATCATTTATCCCTGTC | Present study | 5 |
| EntOv_1200F | GAAAACTTACCAAGACCGAACAG | Stensvold et al., 2010 | 5,6,9 |
| CO4_1050F | CGAAAGCATTTCACTCAATTATGTC | Present study | 6 |
| CO4_950R | ATTATTCCTCTTAATCCTTCTCTTGC | Present study | 6 |
| CO4_700R | GCTTCCAGACGTCTTTCCAC | Present study | 6 |
| CO4_800R | TTTCTGAATCACCCCAATTAATTC | Present study | 6 |
| EST34_1100R | CTACTGTTCGGTCTTGGTAAGTTTTC | Present study | 6 |
| EST34_1230R | AGAACCATTAATCTGTCATTCCTAC | Present study | 6,7 |
| Ent350_1200F | TAGAAATTTCTCGGTCTGGTATCTTC | Present study | 7 |
| Ent350_730R | GCGAATTATCCACTTTACAAAGTAAAG | Present study | 7 |
| Ent350_740R | GCCTAAACATTAAATAGCGAATTATC | Present study | 7 |
| Oedla_1700R | TTCCTAAACTATTTCAGTCTTGGTC | Present study | 8 |
| Oedla_1300F | GACTGAAACCTATTAATTAGTTCGC | Present study | 8 |
| Oedla_470R | TTGTCGTCACTACCTCTCCGC | Present study | 8 |
| Oedla_480R | TCCTACTCATTCCTTCAAGGCTC | Present study | 8 |
| Oedla_1550F | CTACAATGGAGTTACTAGAGAGTAATAC | Present study | 8 |
| Oedla_1600F | CTGTATCAATATGTCGAGCCTCTTGC | Present study | 8 |
| EntMLTURT_550F | GAATGAGTAGGAAGCAAAGTATCC | Present study | 8 |
| EntMLTURT_300F | CCAAGACAATTGTAGAACACGC | Present study | 8 |
| AEMH3.1 | AAGGGCATCACGGACCTGTT | Clark et al., 2006 | 5,8 |
| AEMH3.3 | AAGGGCATCACAGACCTGCT | Clark et al., 2006 | 8 |
| 528F | CGGTAATTCCAGCTCC | Elwood et al., 1985 | 7,9 |

| | | | |
|---|---|---|---|
| 528R | GAGCTGGAATTACCGC | Present study | 9 |
| 1055F | GTGGTGCATGGCCGT | Elwood et al., 1985 | 9 |
| 1055R | ACGGCCATGCACCAC | Elwood et al., 1985 | 9 |
| EmidF | TAGGGGATCGAAGACGA | Present study | 9 |
| EmidR | TCGTCTTCGATCCCCTA | Present study | 9 |
| 1200F | CAGGTCTGTGATGCCC | Elwood et al., 1985 | 9 |

[1]Numbers refer to complete SSU rRNA gene sequences (1=J69; 2=S2702; 3=Hulman; 4=09/1246; 5=09/1247; 6=CO4; 7=349.2/350; 8=Oedla; 9=360).

675 Table 3. Previously unpublished, partial SSU rRNA gene sequences included in the study. Sequences marked in bold were included in phylogenetic analyses (Figs. 1-3).
676 #Sequences obtained from cloned DNA. *Sequences obtained from purified cyst preparations.

677

| Sample ID | Host species | Travel history, ethnicity or other information[1] | SSU rDNA regions available for analysis[2] | *Entamoeba* species/lineage | Fig. | Accession no. |
|---|---|---|---|---|---|---|
| 967 | *Homo sapiens* | n/a | 497-1002 | *E. coli* ST1 | - | FR686401 |
| **3954** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686402** |
| **3968** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686403** |
| **12093** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686404** |
| 21790 | *Homo sapiens* | no info | 466-925 | *E. coli* ST1 | - | FR686409 |
| **28287** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686407** |
| **28305** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686406** |
| 79739 | *Homo sapiens* | n/a | 720-990 | *E. coli* ST1 | - | FR686405 |
| Drill1 | *Mandrillus leucophaeus* | Zoo Saarbruecken, Germany | 32-613# | *E. coli* ST1 | - | FR686410 |
| **EM044** | ***Homo sapiens*** | **n/a** | **46-1420** | ***E. coli* ST1** | **3A** | **FR686411** |
| **EM045** | ***Homo sapiens*** | **n/a** | **47-1420** | ***E. coli* ST1** | **3A** | **FR686412** |
| **EM049** | ***Homo sapiens*** | **n/a** | **51-1420** | ***E. coli* ST1** | **3A** | **FR686413** |
| **EM050** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686414** |
| **EM051** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686415** |
| **EM052** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686416** |
| **EM053** | ***Homo sapiens*** | **n/a** | **160-1419** | ***E. coli* ST1** | **3A** | **FR686417** |
| **EM054** | ***Homo sapiens*** | **n/a** | **48-1420** | ***E. coli* ST1** | **3A** | **FR686418** |
| **EM055** | ***Homo sapiens*** | **n/a** | **160-1419** | ***E. coli* ST1** | **3A** | **FR686419** |
| **EM056** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686420** |
| **EM057** | ***Homo sapiens*** | **n/a** | **41-1420** | ***E. coli* ST1** | **3A** | **FR686421** |
| **EM064** | ***Homo sapiens*** | **Brazil** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686423** |

| | | | | | | |
|---|---|---|---|---|---|---|
| EM065 | *Homo sapiens* | n/a | 795-957 | *E. coli* ST1 | - | FR686424 |
| **EM066** | ***Homo sapiens*** | **Cyprus** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686425** |
| **EM067** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686426** |
| **EM069** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686427** |
| **EM073** | ***Homo sapiens*** | **Lebanon** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686428** |
| **EM074** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686429** |
| **EM075** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686430** |
| **EM077** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686431** |
| **EM078** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686432** |
| EThue2 | *Homo sapiens* | Vietnam | 32-614# | *E. coli* ST2 | - | FR686433 |
| J10 | *Homo sapiens* | n/a | 491-1420 | *E. coli* ST2 | - | FR686434 |
| **J134** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686435** |
| **J147** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686436** |
| **J52** | ***Homo sapiens*** | **n/a** | **160-1420** | ***E. coli* ST1** | **3A** | **FR686437** |
| **J65** | ***Homo sapiens*** | **n/a** | **160-1004** | ***E. coli* ST1** | **3A** | **FR686438** |
| **19885** | ***Homo sapiens*** | **n/a** | **160-1421** | ***E. coli* ST2** | **3A** | **FR686408** |
| A841 | *Chinchilla lanigera* | Pet shop, Belgium | 1110-2047 | *E. coli* ST2 | - | FR686439 |
| **EM047** | ***Homo sapiens*** | **n/a** | **49-1421** | ***E. coli* ST2** | **3A** | **FR686440** |
| **EM061** | ***Homo sapiens*** | **n/a** | **160-1421** | ***E. coli* ST2** | **3A** | **FR686422** |
| **EM063** | ***Homo sapiens*** | **n/a** | **160-1421** | ***E. coli* ST2** | **3A** | **FR686441** |
| **EM068** | ***Homo sapiens*** | **Rwanda** | **160-1000; 1205-1421** | ***E. coli* ST2** | **3A** | **FR686442** |
| **EM070** | ***Homo sapiens*** | **Peru/Mexico** | **160-1421** | ***E. coli* ST2** | **3A** | **FR686443** |
| **EM071** | ***Homo sapiens*** | **Tanzania** | **7-64; 159-1421** | ***E. coli* ST2** | **3A** | **FR686444** |
| **EM072** | ***Homo sapiens*** | **Malawi** | **160-1421** | ***E. coli* ST2** | **3A** | **FR686445** |
| **EM076** | ***Homo sapiens*** | **Ecuador** | **160-1420** | ***E. coli* ST2** | **3A** | **FR686446** |
| ETgor | *Gorilla gorilla* | Allwetter Zoo, Muenster, Germany | 32-614# | *E. coli* ST2 | - | FR686447 |

| | | | | | | |
|---|---|---|---|---|---|---|
| EThue1 | *Homo sapiens* | Vietnam | 32-614# | *E. coli* ST2 | - | FR686448 |
| 739 | *Macaca fuscata* | Animal rescue centre, Rieti, Italy | 1263-2068 | *E. coli* ST2 | - | FR686449 |
| A2 | *Equus zebra hartmannae* | Paignton Zoo, UK | 1127-1849 | *E. equi* | - | FR686450 |
| **09/1070** | ***Macaca sylvanus*** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686369** |
| **09/1260** | ***Lagothrix lagotricha*** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686368** |
| **09/1620** | ***Lagothrix lagotricha*** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686366** |
| **09/1624** | ***Lagothrix lagotricha*** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686367** |
| **08/1113** | ***Pongo pygmaeus*** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686370** |
| **08/1040** | ***Papio* sp.** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686371** |
| **08/1157** | ***Macaca sylvanus*** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686372** |
| **09/1140** | ***Erythrocebus patas*** | **Twycross Zoo, UK** | **55-584** | ***E. hartmanni*** | **3D** | **FR686373** |
| **EM042** | ***Homo sapiens*** | **n/a** | **55-1954** | ***E. hartmanni*** | **3D** | **FR686374** |
| EM043 | *Homo sapiens* | n/a | 464-1954 | *E. hartmanni* | - | FR686375 |
| **EM046** | ***Homo sapiens*** | **n/a** | **55-1954** | ***E. hartmanni*** | **3D** | **FR686376** |
| **EM059** | ***Homo sapiens*** | **n/a** | **55-1954** | ***E. hartmanni*** | **3D** | **FR686377** |
| **EM060** | ***Homo sapiens*** | **n/a** | **55-1954** | ***E. hartmanni*** | **3D** | **FR686378** |
| **EM061a** | ***Homo sapiens*** | **n/a** | **55-1954** | ***E. hartmanni*** | **3D** | **FR686379** |
| EM062 | *Homo sapiens* | n/a | 464-1954 | *E. hartmanni* | - | FR686380 |
| EM065a | *Homo sapiens* | n/a | 464-1954 | *E. hartmanni* | - | FR686381 |
| **J92** | ***Homo sapiens*** | **n/a** | **55-584** | ***E. hartmanni*** | **3D** | **FR686382** |
| **J136** | ***Homo sapiens*** | **n/a** | **115-1572** | ***E. polecki* ST1** | **3B** | **FR686383** |
| **Swine pool 9** | ***Sus scrofa domesticus*** | **Denmark** | **500-1047** | ***E. polecki* ST1** | **3B** | **FR686384** |
| Nandu1 | *Rhea americana* | Kolmårdens Djurpark, Sweden | 7-584#* | *E. polecki* ST1 | 3B | FR686387 |
| **UNE214** | ***Macaca fascicularis*** | **n/a** | **501-1051** | ***E. polecki* ST2** | **3B** | **FR686389** |
| Nandu2 | *Rhea americana* | Kolmårdens Djurpark, Sweden | 49-585#* | *E. polecki* ST3 | 3B | FR686388 |
| **Swine pool 11** | ***Sus scrofa domesticus*** | **Denmark** | **490-1051** | ***E. polecki* ST3** | **3B** | **FR686385** |

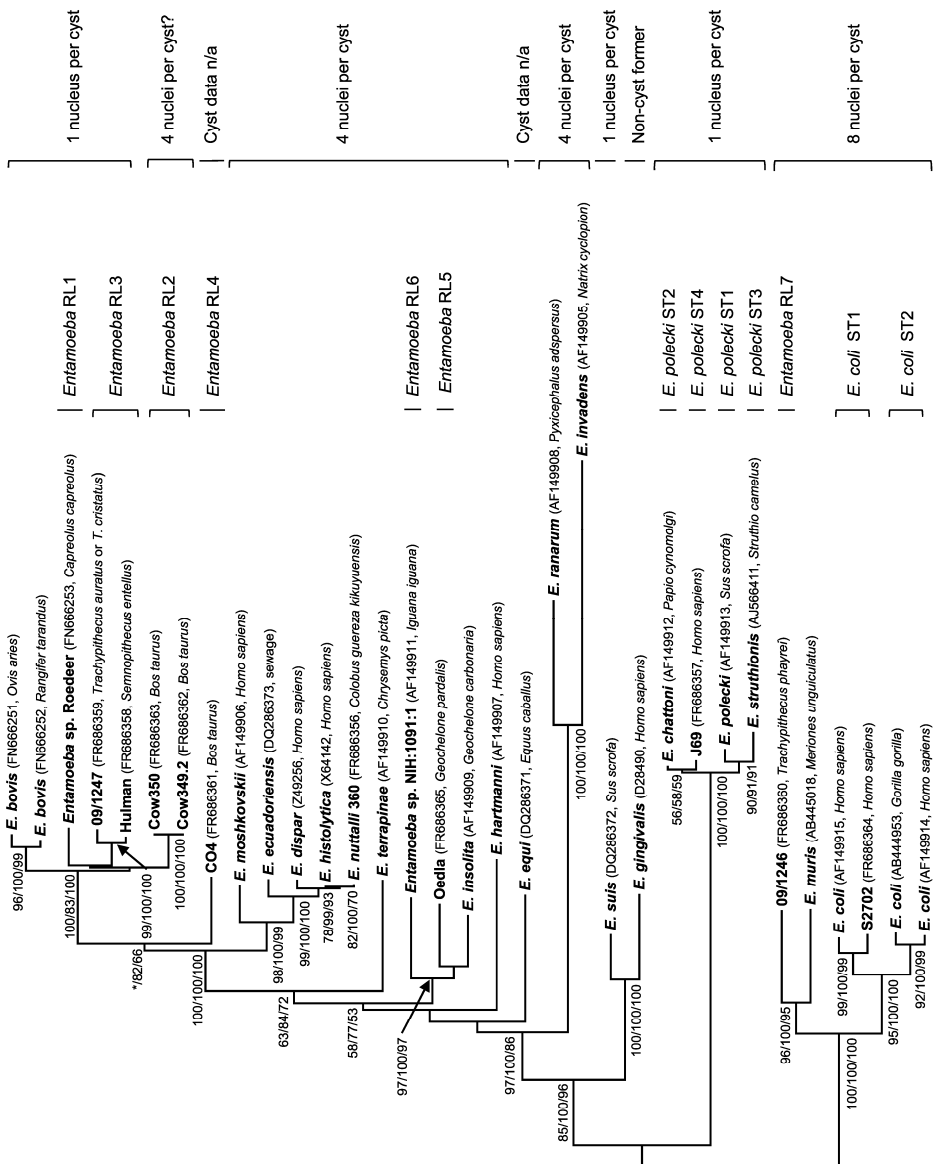
| | | | | | | |
|---|---|---|---|---|---|---|
| **Swine pool 5** | ***Sus scrofa domesticus*** | **Denmark** | **591-1051** | ***E. polecki* ST3** | **3B** | **FR686386** |
| **UNE6** | ***Homo sapiens*** | **Sweden** | **502-1054; 1071-1721** | ***E. polecki* ST3** | **3B** | **FR686390** |
| **UNE755** | ***Homo sapiens*** | **Nigeria** | **502-1039** | ***E. polecki* ST3** | **3B** | **FR686391** |
| **UNE1** | ***Homo sapiens*** | **Somalia, Ethiopia** | **28-1838** | ***E. polecki* ST4** | **3B** | **FR686392** |
| **UNE10** | ***Homo sapiens*** | **n/a** | **29-1850** | ***E. polecki* ST4** | **3B** | **FR686393** |
| **UNE11** | ***Homo sapiens*** | **Ethiopia** | **39-1850** | ***E. polecki* ST4** | **3B** | **FR686394** |
| **UNE2** | ***Homo sapiens*** | **Sudan** | **34-1838** | ***E. polecki* ST4** | **3B** | **FR686395** |
| **UNE2024** | ***Homo sapiens*** | **n/a** | **489-1048** | ***E. polecki* ST4** | **3B** | **FR686396** |
| **UNE5** | ***Homo sapiens*** | **Viet Nam** | **40-1850** | ***E. polecki* ST4** | **3B** | **FR686397** |
| **UNE7** | ***Homo sapiens*** | **Kenya, South Africa** | **32-1850** | ***E. polecki* ST4** | **3B** | **FR686398** |
| **UNE8** | ***Homo sapiens*** | **Iraq** | **501-1047** | ***E. polecki* ST4** | **3B** | **FR686399** |
| **UNE9** | ***Homo sapiens*** | **n/a** | **489-1034** | ***E. polecki* ST4** | **3B** | **FR686400** |
| 09/1464 | *Gorilla gorilla* | Twycross Zoo, UK | 1-565 | *E. suis* | - | FR686456 |
| **09/1618** | ***Trachypithecus francoisi*** | **Twycross Zoo, UK** | **19-543** | ***Entamoeba* RL3** | **3C** | **FR686452** |
| **09/1621** | ***Trachypithecus auratus*** | **Twycross Zoo, UK** | **59-581** | ***Entamoeba* RL3** | **3C** | **FR686453** |
| **09/1622** | ***Trachypithecus phayrei*** | **Twycross Zoo, UK** | **21-546** | ***Entamoeba* RL3** | **3C** | **FR686454** |
| **09/1248** | ***Trachypithecus auratus or T. cristatus*** | **Twycross Zoo, UK** | **22-542** | ***Entamoeba* RL3** | **3C** | **FR686455** |
| EST34 | *Bos taurus* | Estonia | 1094-1893 | *Entamoeba* RL4 | - | FR686451 |

[1]n/a = information not available

[2]positions based on the following reference sequences: AF149915 (*E. coli* ST1), AB444953 (*E. coli* ST2), DQ286371 (*E. equi*), AF149907 (*E. hartmanni*), AF149913 (*E. polecki* ST1), AF149912 (*E. polecki* ST2), AJ566411 (*E. polecki* ST3), FR686357 (*E. polecki* ST4), DQ286372 (*E. suis*), FR686358 (*Entamoeba* RL3) and FR686361 (*Entamoeba* RL4).

FIGURE LEGENDS

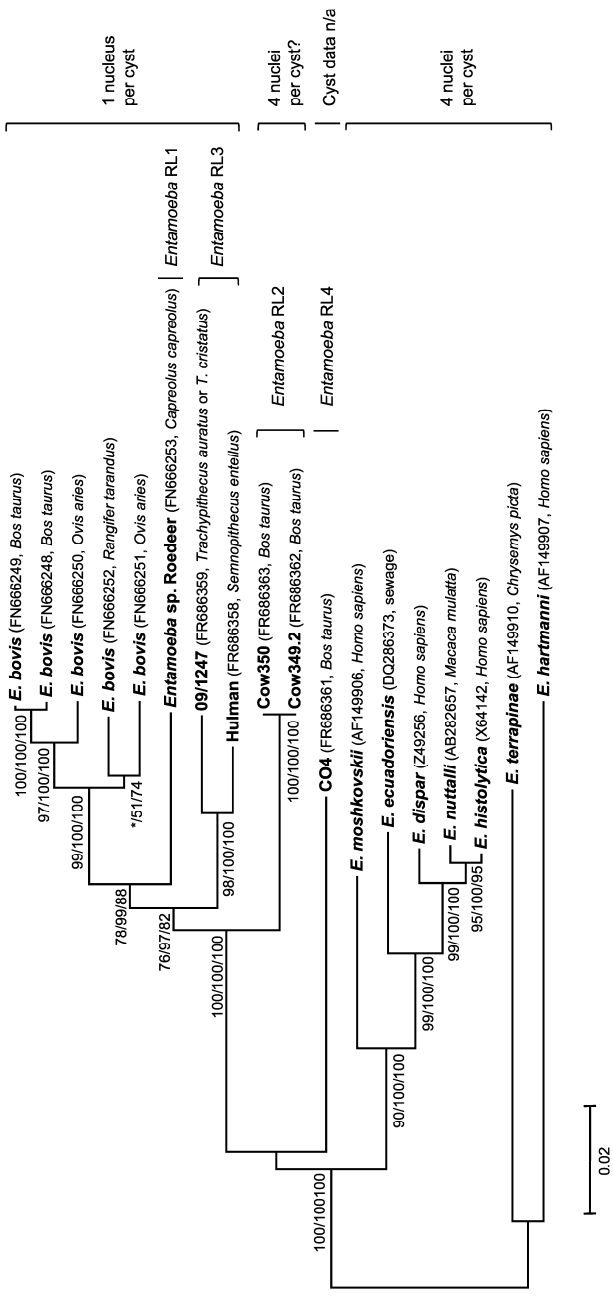
Fig. 1. Phylogenetic relationships among SSU rRNA gene sequences of *Entamoeba* species. The tree shown is the one inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Maximum Composite Likelihood method with rate variation among sites modelled using a gamma distribution (shape parameter = 0.5). The percentage of trees clustered together in the bootstrap test (1,000 replicates) and the posterior probabilities (expressed as a percentage) are shown next to the branch nodes in the order PhyML/MrBayes/Neighbor Joining. An asterisk indicates a value of less than 50% and if two or three analyses gave a value of lower than 50% no values are shown for that node. Accession numbers for the sequences generated in this study and reference sequences are listed parentheses with the Latin name of the host. n/a = not available. Bar = estimated number of substitutions per site. Accession numbers and host species are indicated in brackets for each sequence.


E. bovis (FN666251, Ovis aries)
E. bovis (FN666252, Rangifer tarandus)
Entamoeba sp. Roedeer (FN666253, Capreolus capreolus)
09/1247 (FR686359, Trachypithecus auratus or T. cristatus)
Hulman (FR686358, Semnopithecus entellus)
Cow350 (FR686363, Bos taurus)
Cow349.2 (FR686362, Bos taurus)
CO4 (FR686361, Bos taurus)
E. moshkovskii (AF149906, Homo sapiens)
E. ecuadoriensis (DQ286373, sewage)
E. dispar (Z49256, Homo sapiens)
E. histolytica (X64142, Homo sapiens)
E. nuttalli 360 (FR686356, Colobus guereza kikuyuensis)
E. terrapinae (AF149910, Chrysemys picta)
Entamoeba sp. NIH:1091:1 (AF149911, Iguana iguana)
Oedla (FR686365, Geochelone pardalis)
E. insolita (AF149909, Geochelone carbonaria)
E. hartmanni (AF149907, Homo sapiens)
E. equi (DQ286371, Equus caballus)
E. ranarum (AF149908, Pyxicephalus adspersus)
E. invadens (AF149905, Natrix cyclopion)
E. suis (DQ286372, Sus scrofa)
E. gingivalis (D28490, Homo sapiens)
E. chattoni (AF149912, Papio cynomolgi)
J69 (FR686357, Homo sapiens)
E. polecki (AF149913, Sus scrofa)
E. struthionis (AJ566411, Struthio camelus)
09/1246 (FR686360, Trachypithecus phayrei)
E. muris (AB445018, Mentones unguiculatus)
E. coli (AF149915, Homo sapiens)
S2702 (FR686364, Homo sapiens)
E. coli (AB444953, Gorilla gorilla)
E. coli (AF149914, Homo sapiens)
96/100/99
100/83/100
*/82/66
99/100/100
100/100/100
100/100/100
98/100/99
99/100/100
78/99/93
82/100/70
58/77/53
97/100/97
63/84/72
97/100/86
85/100/96
100/100/100
100/100/100
56/58/59
100/100/100
90/91/91
96/100/95
99/100/99
100/100/100
95/100/100
92/100/99
0.05
Entamoeba RL1
Entamoeba RL3
Entamoeba RL2
Entamoeba RL4
Entamoeba RL6
Entamoeba RL5
E. polecki ST2
E. polecki ST4
E. polecki ST1
E. polecki ST3
Entamoeba RL7
E. coli ST1
E. coli ST2
1 nucleus per cyst
4 nuclei per cyst?
Cyst data n/a
4 nuclei per cyst
Cyst data n/a
4 nuclei per cyst
1 nucleus per cyst
Non-cyst former
1 nucleus per cyst
8 nuclei per cyst

698

699

700 Fig. 2. Phylogenetic analysis of cattle and langur amoebae. The distance-based tree of
701 selected complete SSU rRNA sequences generated to further resolve the relationship between
702 *E. bovis* and RL1, RL2, RL3 and RL4 is shown. Analysis and labelling is as in Fig. 1.
703 Sequences from *E. terrapinae* and *E. hartmanni* were included as an outgroup. Accession
704 numbers and host species are indicated in brackets for each sequence.


E. bovis (FN666249, Bos taurus)
E. bovis (FN666248, Bos taurus)
E. bovis (FN666250, Ovis aries)
E. bovis (FN666252, Rangifer tarandus)
E. bovis (FN666251, Ovis aries)
Entamoeba sp. Roedeer (FN666253, Capreolus capreolus)
09/1247 (FR686359, Trachypithecus auratus or T. cristatus)
Hulman (FR686358, Semnopithecus entellus)
Cow350 (FR686363, Bos taurus)
Cow349.2 (FR686362, Bos taurus)
CO4 (FR686361, Bos taurus)
E. moshkovskii (AF149906, Homo sapiens)
E. ecuadoriensis (DQ286373, sewage)
E. dispar (Z49256, Homo sapiens)
E. nuttalli (AB262657, Macaca mulatta)
E. histolytica (X64142, Homo sapiens)
E. terrapinae (AF149910, Chrysemys picta)
E. hartmanni (AF149907, Homo sapiens)
100/100/100
97/100/100
99/100/100
*/51/74
78/99/88
76/57/82
98/100/100
100/100/100
100/100/100
90/100/100
99/100/100
99/100/100
95/100/95
0.02
Entamoeba RL1
Entamoeba RL3
Entamoeba RL2
Entamoeba RL4
1 nucleus per cyst
4 nuclei per cyst?
Cyst data n/a
4 nuclei per cyst

705

706

707 Fig. 3. Phylogenetic analysis of partial SSU rDNA sequences. Distance-based trees showing
708 intra-specific variation in *E. coli* (A), *E. polecki* (B), *Entamoeba* RL3 (C) and *E. hartmanni*
709 (D) were obtained using partial SSU rDNA sequences as in Fig. 1. A total of 854, 540, 575,
710 and 519 base pair positions were aligned unambiguously and analysed for (A), (B), (C) and
711 (D), respectively. The regions included correspond to the 5' two-thirds (A), 5' one-third (C
712 and D) and the central third (B) of the gene. The trees in (A), (B) and (C) are unrooted, but
713 shown with the same topology as in Fig. 1. Sequences from *E. ranarum* and *E. invadens* were
714 included as an outgroup in (D). All sample IDs beginning with 08/ or 09/ are from non-
715 human primates; samples from humans are marked with an asterisk in (B) and (D). In (C),
716 GU****** is the accession number of a sequence obtained by Levecke et al. (2010).
717


A
E. coli ST1
EM069
3968
EM073
J52
3954
28305
EM045
EM056
EM075
EM054
J65
EM078
EM074
EM057
EM055
EM044
12093
28237
EM051
EM66
J147
EM053
EM049
EM050
EM052
J134
EM067
E. coli (AF149915)
EM064
EM077
S2702
63/56/55
62/*/63
99/100/98
E. coli ST2
EM068
87/100/80
E. coli (AF149914)
EM047
EM070
EM072
EM071
EM076
EM061
19885
E. coli (AB449953)
EM063
100/100/100
90/97/87
73/*/62
63/*/67
0.1
B
ST4
ST2
ST1
ST3
UNE8*
UNE2024*
UNE9*
UNE5*
UNE1*
UNE2*
UNE10*
UNE11*
UNE7*
J69*
98/71/99
E. chattoni (AF149912)
UNE214
99/100/100
98/100/100
J136*
Swine pool 9
E. polecki (AF149913)
96/85/94
UNE6*
E. struthionis (AJ566411)
E. polecki (DQ286374)
Swine pool 11
Swine pool 5
UNE755*
99/100/100
61/*/76
0.005
C
09/1621
09/1248
09/1247
09/1622
GU437823
GU437824
09/1618
Hulman
GU437826
GU437825
36/*/94
100/100/100
98/100/100
86/98/89
*/82/64
98/99/99
0.01
D
09/1624
EM061a*
E. hartmanni (AF149907)*
091140
09/1260
09/1620
J92*
08/1113
EM042*
08/1157
EM060*
EM 046*
EM059*
08/1040
09/1070
68/75/70
60/*/8*
62/95/66
100/100/99
77/88*
E. ranarum (AF149908)
E. invadens (AF149905)
0.05

718