

1 TITLE

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

4

5 AUTHORS

6

7 C. Rune Stensvold<sup>a</sup>, Marianne Lebbad<sup>b</sup>, Emma L. Victory<sup>c</sup>, Jaco J. Verweij<sup>d</sup>, Egbert  
8 Tannich<sup>e</sup>, Mohammed Alfellani<sup>c</sup>, Paulette Legarraga<sup>c,2</sup>, C. Graham Clark<sup>c,1</sup>

9

10 AFFILIATIONS

11 <sup>a</sup>Department of Microbiology and Diagnostics, Statens Serum Institut, Artillerivej 3, DK-  
12 2300 Copenhagen S, Denmark

13 <sup>b</sup>Department of Parasitology, Mycology and Environmental Microbiology, Swedish Institute  
14 for Infectious Disease Control, SE-171 82 Solna, Sweden

15 <sup>c</sup>Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical  
16 Medicine, Keppel Street, WC1E 7HT London, United Kingdom.

17 <sup>d</sup>Department of Parasitology, Leiden University Medical Center, Postbus 9600, 2300 RC  
18 Leiden, The Netherlands

19 <sup>e</sup>Department of Molecular Parasitology, Bernhard Nocht Institute for Tropical Medicine,  
20 Bernhard-Nocht-Str. 74, D-20359 Hamburg, Germany

21 <sup>2</sup>Present Address: Clinical Laboratory Department, Pontificia Universidad Católica de Chile,  
22 Santiago, Chile

23

24

25 <sup>1</sup>Corresponding author; fax: +44 20-7636-8739; email: [Graham.Clark@lshtm.ac.uk](mailto:Graham.Clark@lshtm.ac.uk)

26 ABSTRACT

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

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

42

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

44

45

46

47 INTRODUCTION

48         The genus *Entamoeba* comprises numerous unicellular, parasitic species found in  
49 humans, non-human primates, other vertebrates and invertebrates. Until recently, the  
50 detection, identification and assignment of *Entamoeba* organisms to species relied mainly on  
51 morphology. The introduction of molecular tools such as PCR and sequencing made it clear  
52 that definitive species identification and establishment of taxonomic relationships within the  
53 genus using microscopy only is not always possible (Clark and Diamond, 1997; Clark et al.,  
54 2006). This is due to a combination of factors, including overlap in morphological  
55 characteristics between species, morphological variation within species, the existence of  
56 mixed species *Entamoeba* infections, and limited knowledge of host specificity.

57         Molecular tools enable us to resolve many of the issues related to the identification,  
58 taxonomy, epidemiology and clinical significance of *Entamoeba* species without reliance on  
59 parasite cultures or experimental infections. Small subunit rRNA gene (SSU rDNA) sequence  
60 data are widely used for analysis of phylogenetic relationships between eukaryotic organisms  
61 and are available for several species of *Entamoeba*. Numerous papers have provided insights  
62 into the phylogeny and host specificity of *Entamoeba* based on sequence data (Silberman et  
63 al., 1999; Verweij et al., 2001; Ponce Gordo et al., 2004; Clark et al., 2006; Suzuki et al.,  
64 2007; Kobayashi et al., 2009; Stensvold et al., 2010; Levecke et al., 2010). Nevertheless,  
65 there are many described species of *Entamoeba* for which no molecular data are available.  
66 Conversely, it is equally likely that species of *Entamoeba* exist that have never been noted  
67 due to a lack of morphologically discriminating features.

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

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

73

## 74 RESULTS

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

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

86

### 87 ***Entamoeba* from cattle**

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

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

100           Microscopic examination of faecal concentrates from cattle samples Cow349 and  
101 Cow350 revealed that the vast majority of the cysts were uninucleated, but that there were  
102 also some cysts with 4 nuclei (and rarely even 6 or 8). The uninucleated cysts reacted with a  
103 monoclonal antibody (mAb) known to react with *E. histolytica* and *E. bovis* (Stensvold et al.,  
104 2010) but the tetranucleated cysts did not. The initial sequences obtained from these two  
105 samples were identical to *E. bovis* (Stensvold et al., 2010, and unpublished data). However,  
106 screening these DNAs using primers designed during the sequencing of CO4 revealed the  
107 presence of a second *Entamoeba* sequence. Full length sequences were obtained that proved  
108 to be distinct from both *E. bovis* and *Entamoeba* RL4 (Figs. 1 and 2). Since they cannot be  
109 linked to any valid species name for now, it is proposed that sequences Cow349.2 and  
110 Cow350 are assigned to *Entamoeba* RL2 (Table 1, Figs. 1 and 2).

111

## 112 ***Entamoeba polecki***

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

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

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

132

### 133 **Other uninucleated-cyst-producing *Entamoebas***

134 Faecal samples from two langur species containing uninucleated cysts yielded two  
135 closely related sequences (Hulman and 09/1247) that were phylogenetically distant from *E.*  
136 *polecki*. The langur *Entamoeba* sequences form a lineage emerging from the clade of  
137 sequences obtained from cattle described earlier (Figs. 1 and 2) and are most closely related  
138 to *E. bovis*, although there is only modest bootstrap support for this. In the absence of an  
139 existing species name, we propose that these organisms be referred to as *Entamoeba* RL3.  
140 Variation among sequences of *Entamoeba* RL3 is significant and when our data are  
141 combined with those of Levecke et al. (2010) the sequences appear to fall into two groups,  
142 but a larger sample is needed before conclusions regarding potential subtypes are made (Fig.  
143 3C).

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

146 2002), but it was shown to have a distinct SSU rDNA (Clark et al., 2006). Partial sequences  
147 from 10 Vietnamese pigs were identical (unpublished data) and since it was not at the time  
148 detected in any other hosts, including humans living in close proximity to their infected pigs  
149 (Blessmann et al., 2002), the species name *E. suis* was resurrected to separate this *Entamoeba*  
150 from *E. polecki* (Clark et al., 2006) and to reflect its apparent host specificity. However, our  
151 detection of *E. suis* in a gorilla (Table 3) shows that this parasite is not restricted to pigs; the  
152 implications for the species name are not yet clear. The two sequences differed at only one  
153 position out of 590. Whether this discovery of *E. suis* in a non-porcine host will prove to be a  
154 rare finding remains to be determined.

155

#### 156 **Tetranucleated-cyst-producing *Entamoebas***

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

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

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

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

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

196 is still lacking.

197

### 198 ***Entamoebas* producing octonucleate cysts**

199 Two complete SSU rDNA sequences from octonucleated-cyst-producers were obtained  
200 (09/1246 and S2702) that are significantly different from those previously reported (Fig 1).

201 *Entamoeba coli* comprises two major clades; the first (here named *E. coli* ST1) is represented  
202 by S2702 and GenBank accession number AF149915, and the second (*E. coli* ST2) by  
203 accession numbers AF149914 and AB444953. The S2702 sequence adds to the extensive  
204 genetic diversity seen in this species, which is estimated at around 13%. Indeed, based on  
205 sequence divergence it would be reasonable to consider *E. coli* ST1 and ST2 to be distinct  
206 species but at present there is no other justification for such a radical step.

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

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

219 In contrast to S2702, the SSU rDNA sequence from the Phayre's Leaf Monkey  
220 (09/1246) clustered not with *E. coli* but with the *E. muris* sequence from a Mongolian Gerbil

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

223

## 224 DISCUSSION

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

230 Therefore, until morphological data are available to establish a valid taxonomic name,  
231 we propose to use identification tags constructed as follows: 1. Well supported phylogenetic  
232 clusters within a defined species are assigned Arabic numerals identifying them as specific  
233 subtypes (STs). All sequences from a species must clearly fall into one of the STs. 2.  
234 Branches within phylogenetic trees that do not show a strong affinity with previously  
235 described species are assigned Arabic numerals identifying them as distinct ribosomal  
236 lineages (RLs). It is difficult to generalise about what constitutes ‘well supported’ and ‘strong  
237 affinity’, as these criteria will vary according to the amount of data available (partial or whole  
238 gene) and included in the alignment, and the method of analysis employed. Different  
239 boundaries may be appropriate in different circumstances. In general, subtypes will be  
240 defined using partial gene sequences, as in our cases, while we strongly suggest that  
241 assignment of new ribosomal lineages should be made using complete SSU rDNA sequences  
242 only. In our data, ‘well supported’ clusters designated as subtypes all have bootstrap support  
243 of 95% or more in PhyML, while a complete gene sequence showing bootstrap support of  
244 less than 80% in PhyML for affinity to another lineage should be considered for identification  
245 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

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

274 Uninucleated-cyst-producing *Entamoeba* infections have been reported in humans  
275 across the world but with greatly varying prevalence (Desowitz and Barnish, 1986; McMillan  
276 and Kelly, 1970; Chacín-Bonilla, 1992; Blessmann et al., 2002). Studies have reported a 70%  
277 prevalence in the wild in both chimpanzees and baboons (Jackson et al., 1990; Muehlenbein,  
278 2005). Uninucleated-cyst-producing *Entamoebas* from humans and non-human primates have  
279 usually been assigned to *E. polecki* and *E. chattoni*, respectively. *Entamoeba chattoni* was  
280 first described by Chatton (1912) and named by Swellengrebel (1914) who found it in a  
281 rhesus monkey (*Macaca rhesus*), while *E. polecki* was originally described in pigs (von  
282 Prowazek, 1911). Kessel and Johnstone (1949) reported finding cysts of *E. chattoni* in both  
283 rhesus monkeys and humans, but concluded that the morphologies of *E. chattoni* and *E.*  
284 *polecki* were so similar they might represent the same species. Likewise, Sumardjo and Joe  
285 (1953) found the morphology of *E. chattoni* and *E. polecki* to be almost identical, except that  
286 the karyosome of *E. polecki* is commonly large and granulated whereas it is smaller and more  
287 delicate in *E. chattoni*. Sargeant et al. (1992) found seven cases of human infection all but  
288 one of which had contact with non-human primates, but in many cases no patient contact with  
289 infected pigs or non-human primates is found (Chacín-Bonilla, 1992; Blessmann et al., 2002).

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

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

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

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

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

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

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

346 1919; Dobell, 1919). The *E. coli* in the present study most likely belong primarily to what  
347 would be the “small races” of *E. coli* (Matthews, 1919; Dobell, 1919), since most samples  
348 were originally mistaken for *E. histolytica/E. dispar*. Many such sequences belong to *E. coli*  
349 ST1, although a few are ST2. It is possible that cyst size is not related to subtype, but only  
350 accumulation of sequence and cyst size data for the same samples plus examination of “large  
351 race” *E. coli* will help clarify this point.

352         Of the protists found in the human intestinal tract, *E. coli* is one of the most  
353 commonly found, and it is generally considered non-pathogenic. A few reports have drawn  
354 the attention to cases of gall bladder disease (Kalk and Wildhirt, 1954; Witte, 1956; Geyer,  
355 1959; von Meyenfeldt et al., 2007) and diarrhoea (Corcoran et al., 1991; Wahlgren, 1991)  
356 that might be attributable to *E. coli*. The ability of *E. coli* to phagocytose erythrocytes has  
357 been documented and varies among strains (Dobell, 1936). Recently, *E. coli* in Colombian  
358 school children appeared to be an indicator of poor nutritional status (Boeke et al., 2010).  
359 Because of the degree of genetic diversity in this species, future studies on its potential role in  
360 disease should note the subtype of the organism so that any links between phenotype and  
361 subtype can be explored.

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

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

375

### 376 **Concluding remarks**

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

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

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

396 with the new data.

397

## 398 METHODS

### 399 **Samples and sequences**

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

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

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

421

## 422 **DNA extraction and DNA sequencing**

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

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

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

442

## 443 **Samples yielding complete *Entamoeba* SSU rDNA sequences**

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

446 4 years.

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

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

456 *Entamoeba* sp. SSU rDNA sequences 09/1246 and 09/1247 were obtained using DNA  
457 extracted from faecal samples from a Phayre's Leaf Monkey (*Trachypithecus phayrei*)  
458 (09/1246) and either a Javan Langur (*Trachypithecus auratus*) or a Silvery Lutung  
459 (*Trachypithecus cristatus*) (09/1247) from Twycross Zoo, Warwickshire, UK, respectively.  
460 The latter two animals were housed together. Examination of faecal concentrates revealed  
461 octonucleated cysts (09/1246) and mixed *Entamoeba* infections with uni-, tetra-, and  
462 octonucleated cysts as well as *Blastocystis* (09/1247).

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

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

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

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

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

479

#### 480 **Samples for partial SSU rDNA sequencing**

481 49 *E. coli* partial sequences from humans (n=45), non-human primates (n=3) and a  
482 chinchilla (n=1), 17 partial *E. hartmanni* sequences from humans (n=9) and non-human  
483 primates (n=8), and 18 partial *E. polecki* sequences from humans (n=12), a non-human  
484 primate (n=1), pigs (n=3) and a nandu (n=2, from one sample) were obtained. Other partial  
485 *Entamoeba* sequences were obtained from an Estonian cow (n=1), langurs (n=4), a zebra  
486 (n=1), and a gorilla (n=1).

487

#### 488 **Sequence assembly, alignment and phylogenetic analyses**

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

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

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

498 Phylogenetic analyses were performed as described previously (Stensvold et al.,  
499 2010) using distance (Neighbor Joining; MEGA 4.0), maximum likelihood (PhyML 3.0;  
500 Guindon and Gascuel, 2003) and Bayesian (MrBayes 3.1.2; Huelsenbeck and Ronquist,  
501 2001) methods. Bayesian and maximum likelihood analysis used a General Time Reversible  
502 (GTR) model of nucleotide substitution with four categories of among-site rate variation and  
503 the proportion of invariant sites, as in previous phylogenetic analyses of *Entamoeba* SSU  
504 rDNA sequences. Statistical support for distance and maximum likelihood trees was  
505 evaluated using bootstrapping (1000 replicates). Bayesian analysis used four Markov chain  
506 Monte Carlo (MCMC) strands, 1,000,000 generations, with trees sampled every 100  
507 generations. In every case the average standard deviation of split frequencies was less than  
508 0.01. A consensus tree was produced after excluding an initial burn-in of 25% of the samples,  
509 as recommended.

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

517

#### 518 ACKNOWLEDGEMENTS

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

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

530

#### 531 REFERENCES

- 532 **Blessmann J, Van Linh P, Nu PA, Thi HD, Muller-Myhsok B, Buss H, Tannich E** (2002)  
533 Epidemiology of amebiasis in a region of high incidence of amebic liver abscess in central  
534 Vietnam. *Am J Trop Med Hyg* **66**:578–583
- 535 **Boeke CE, Mora-Plazas M, Forero Y, Villamor E** (2010) Intestinal protozoan infections in  
536 relation to nutritional status and gastrointestinal morbidity in Colombian school children. *J*  
537 *Trop Pediatr* (In Press)
- 538 **Bradford CM, Denver MC, Cranfield MR** (2008) Development of a polymerase chain  
539 reaction test for *Entamoeba invadens*. *J Zoo Wildl Med* **39**:201-207
- 540 **Chacín-Bonilla L** (1992) *Entamoeba polecki*: human infections in Venezuela. *Trans R Soc*  
541 *Trop Med Hyg* **86**:634
- 542 **Chatton, E** (1912) Sur quelques genres d'amibes libres et parasites. Synonymies,  
543 homonymie, impropriété. *Bull Soc Zool France* **37**:109-115
- 544 **Clark CG, Diamond LS** (1997) Intraspecific variation and phylogenetic relationships in the  
545 genus *Entamoeba* as revealed by riboprinting. *J Euk Microbiol* **44**:142-154

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

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

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

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

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

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

560 **Elwood HJ, Olsen GJ, Sogin ML** (1985) The small-subunit ribosomal RNA gene sequences  
561 from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. *Mol Biol Evol*  
562 **2**:399-410

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

565 **Geyer E** (1959) [Is there an *Entamoeba coli* cholecystopathy?] *Z Gesamte Inn Med.* **14**:968-  
566 972

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

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

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

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

575 **Jackson TFHG, Sargeant PG, Visser PS, Gathiram V, Suparsad S, Anderson CB**  
576 (1990) *Entamoeba histolytica*: naturally occurring infections in baboons. Arch Invest Med  
577 (Mex) **21** Suppl **1**:153-156

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

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

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

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

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

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

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

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

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

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

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

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

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

603 **Ponce Gordo F, Martinez Diaz RA, Herrera S** (2004) *Entamoeba struthionis* n. sp.  
604 (Sarcomastigophora: Endamoebidae) from ostriches (*Struthio camelus*). Vet Parasitol  
605 **119**:327-335

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

608 **Rodhain J, van Hoof MT** (1947) *Entamoeba knowlesi* n.sp. parasite de deux tortues:  
609 *Terrapina cinosternoides* et *Platysternum megacephalum*. Ann Parasitol Hum Comp **22**:129-  
610 137

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

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

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

619 **Stensvold R, Brillowska-Dabrowska A, Nielsen HV, Arendrup MC** (2006) Detection of  
620 *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 Tijdschr 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 Geneesk **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

660 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	<i>Homo sapiens</i>	N/A	N/A	<i>Entamoeba polecki</i> ST4	FR686357
Hulman	<i>Semnopithecus entellus</i>	9.2—15.4 $\mu\text{m}$ (12.34 $\mu\text{m} \pm 1.83$ $\mu\text{m}$ , 50 cysts)	1	<i>Entamoeba</i> RL3	FR686358#
09/1247	<i>Trachypithecus auratus</i> or <i>T. cristatus</i> <sup>1</sup>	N/A	N/A	<i>Entamoeba</i> RL3	FR686359
09/1246	<i>Trachypithecus phayrei</i>	N/A	N/A	<i>Entamoeba</i> RL7	FR686360
CO4	<i>Bos taurus</i>	N/A	N/A	<i>Entamoeba</i> RL4	FR686361
Cow349.2	<i>Bos taurus</i>	N/A	1 or 4 <sup>2</sup>	<i>Entamoeba</i> RL2	FR686362*
Cow350	<i>Bos taurus</i>	N/A	1 or 4 <sup>2</sup>	<i>Entamoeba</i> RL2	FR686363*
S2702	<i>Homo sapiens</i>	N/A	8	<i>Entamoeba coli</i> ST1	FR686364
Oedla	<i>Geochelone pardalis</i>	12.0—19.5 $\mu\text{m}$ (14.7 $\mu\text{m} \pm 1.33$ $\mu\text{m}$ , 100 cysts)	4 <sup>3</sup>	<i>Entamoeba</i> RL5	FR686365*
360	<i>Colobus guereza</i> <i>kikuyuensis</i>	N/A	4	<i>Entamoeba nuttalli</i>	FR686356

661 <sup>1</sup>These two hosts were housed together so the source of the sample is not identifiable.

662 <sup>2</sup>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.

663 <sup>3</sup>The sample contained cysts with varying number of nuclei, mostly 4 or 2 nuclei were seen with a few being uninucleate.

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

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

666

667

668

669

670

671

672  
673

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

Primer name	Primer sequence (5'—3')	Reference	PCR/Sequencing use <sup>1</sup>
RD5	ATCTGGTTGATCCTGCCAGT	Clark et al., 2006	1,2,3,4,5,6,7,8,9
RD3	ATCCTTCCGCAGGTTACCTAC	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	CAAGAGACACCAAAAAGGCATC	Present study	3

1247hulman_1700F	CTCTGTTGGAGTGGTAAGAATTCTC	Present study	5
1247hulman_1550F	GTTAATTTGTGTTTATGATTTTCGGTC	Present study	5
1247hulman_430F	AGGAGATGCCGTATGGTATTTTC	Present study	5
EstCowEnt_1690R	ATTCCAATCATTTATCCCTGTC	Present study	5
EntOv_1200F	GAAAACCTACCAAGACCGAACAG	Stensvold et al., 2010	5,6,9
CO4_1050F	CGAAAGCATTCACTCAATTATGTC	Present study	6
CO4_950R	ATTATTCCTCTTAATCCTTCTCTTGC	Present study	6
CO4_700R	GCTTCCAGACGCTTTCCAC	Present study	6
CO4_800R	TTTCTGAATCACCCCAATTAATTC	Present study	6
EST34_1100R	CTACTGTTCGGTCTTGGTAAGTTTTTC	Present study	6
EST34_1230R	AGAACCATTAATCTGTCATTCCCTAC	Present study	6,7
Ent350_1200F	TAGAAATTTCTCGGTCTGGTATCTTC	Present study	7
Ent350_730R	GCGAATTATCCACTTACAAAGTAAAG	Present study	7
Ent350_740R	GCCTAAACATTAATAGCGAATTATC	Present study	7
Oedla_1700R	TTCCTAAACTATTTTCAGTCTTGGTC	Present study	8
Oedla_1300F	GACTGAAACCTATTAATTAGTTTCGC	Present study	8
Oedla_470R	TTGTCGTCACTACCTCTCCGC	Present study	8
Oedla_480R	TCCTACTCATTTCCTTCAAGGCTC	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	CAGGTCTGTGATGCC	Elwood et al., 1985	9

---

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

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

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

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

678

679 <sup>1</sup>n/a = information not available

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

683

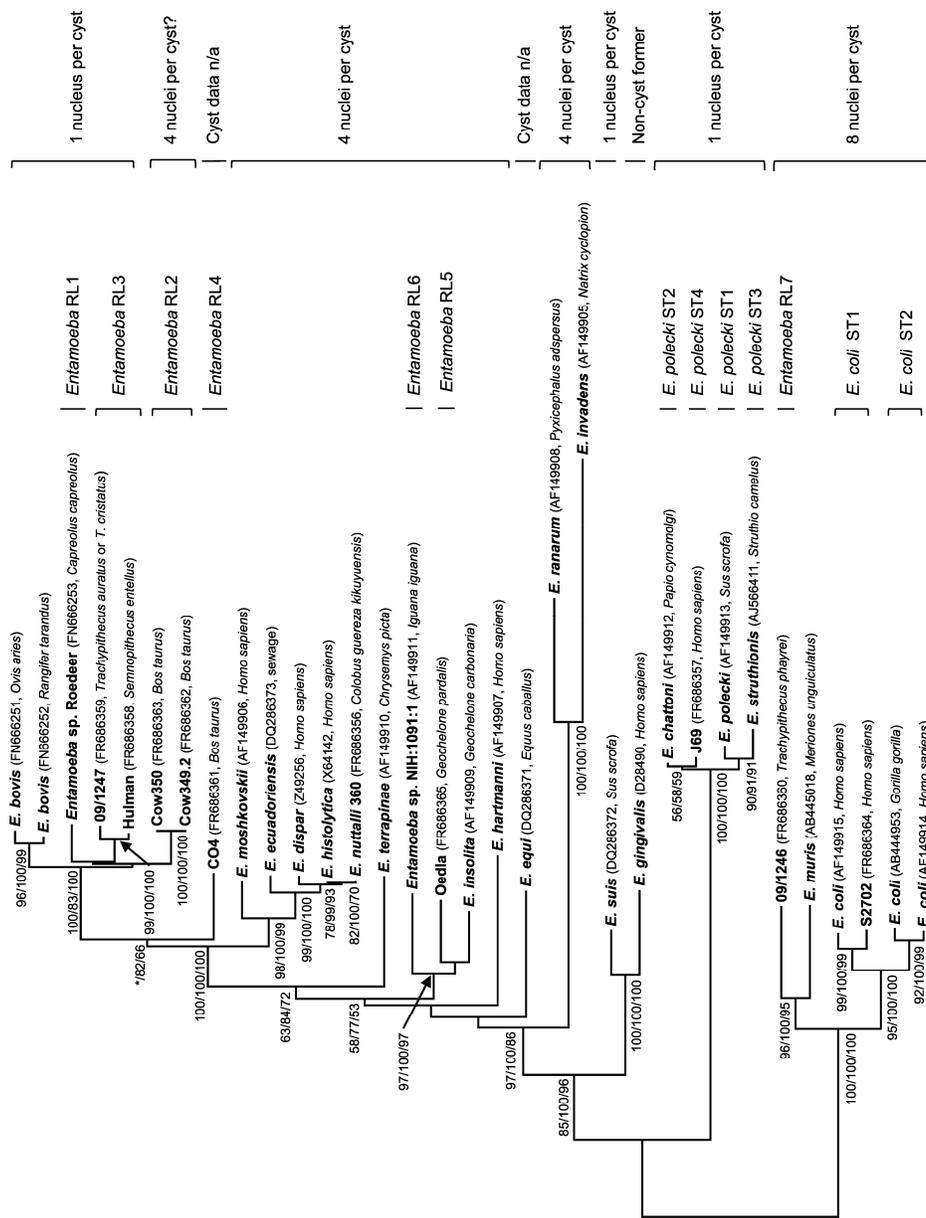
684

685 FIGURE LEGENDS

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

698

699



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

*Entamoeba* RL1

*Entamoeba* RL3

*Entamoeba* RL2

*Entamoeba* RL4

*Entamoeba* RL6

*Entamoeba* RL5

*E. invadens* (AF149905, *Natrix cyclopion*)

*E. polecki* ST2

*E. polecki* ST4

*E. polecki* ST1

*E. polecki* ST3

*Entamoeba* RL7

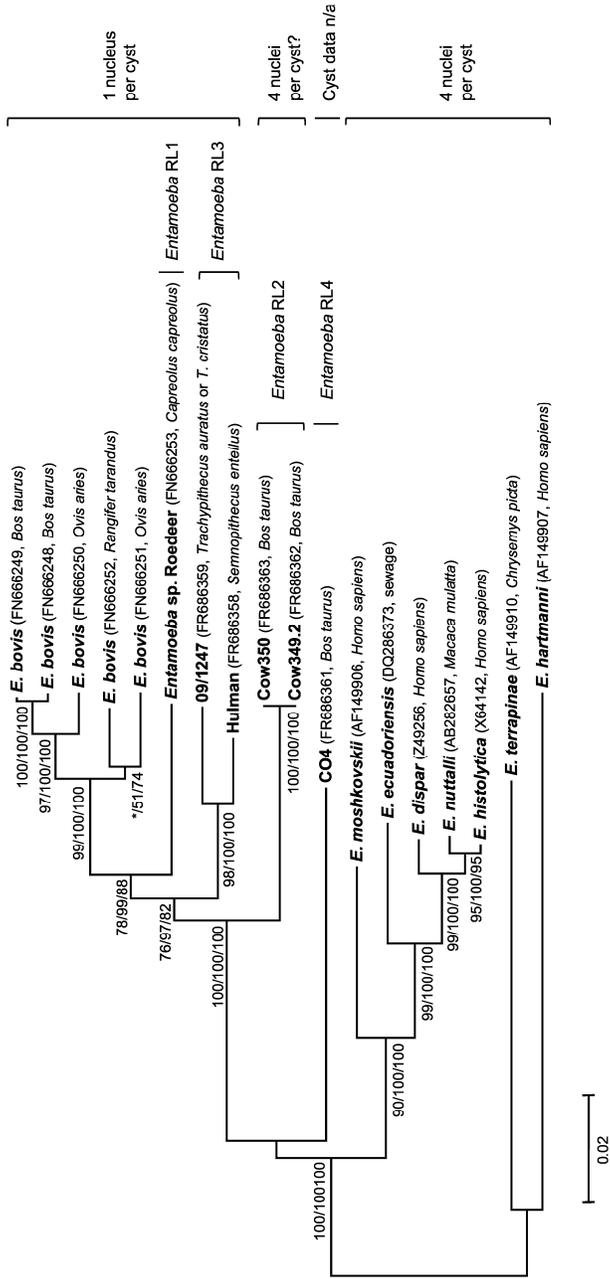
*E. coli* ST1

*E. coli* ST2

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.

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

