

1 Running head: Expanding the *Entamoeba* universe

2

3

4 **Expanding the *Entamoeba* Universe: New Hosts Yield Novel Ribosomal Lineages** <sup>1</sup>

5

6 Alison S. Jacob<sup>a</sup>, Eloise J. Busby<sup>a</sup>, Abigail D. Levy<sup>a, 2</sup>, Natasha Komm<sup>a</sup> and C. Graham  
7 Clark<sup>a</sup>

8

9 <sup>a</sup> Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical  
10 Medicine, London WC1E 7HT, England, GB

11

12

13 Key words: *Entamoeba*, Genetic Diversity, Phylogeny, Ribosomal RNA, Sequencing

14

15 1 – This paper is dedicated to Govinda S. Visvesvara on the occasion of his retirement.

16

17 2 – Current Address: Histocompatibility & Immunogenetics Department, Hammersmith  
18 Hospital, Imperial College Healthcare NHS Trust, London W12 0HS, England, GB

18

19 Corresponding Author: C. Graham Clark, Faculty of Infectious and Tropical Diseases,  
20 London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT,  
21 England, GB. Telephone number: +44-20-7927-2351; FAX number: +44-20-7636-8739;  
22 email: graham.clark@lshtm.ac.uk

23 **ABSTRACT**

24  
25 Removing the requirement for cell culture has led to a substantial increase in the number  
26 of lineages of *Entamoeba* recognized as distinct. Surveying the range of potential host  
27 species for this parasite genus has barely been started and it is clear that additional  
28 sampling of the same host in different locations often identifies additional diversity. In  
29 this study, using small subunit ribosomal RNA gene sequencing, we identify four new  
30 lineages of *Entamoeba*, including the first report of *Entamoeba* from an elephant, and  
31 extend the host range of some previously described lineages. Additionally, examination  
32 of microbiome data from a number of host animals suggests that substantial *Entamoeba*  
33 diversity remains to be uncovered.

34  
35 **Keywords:** Diversity; next generation sequencing; ribosomal RNA; phylogeny;

36  
37  
38 Over the past 25 years, our understanding of diversity in the genus *Entamoeba* has  
39 increased significantly as a result of complementary developments in DNA amplification,  
40 purification and sequencing. Traditionally, naming of species in *Entamoeba* was based on  
41 a mixture of host identity and parasite morphology, the latter being rather limited in these  
42 amoeboid organisms and the former being of debatable value due to uncertainty over host  
43 ranges of the parasites. DNA sequencing allows quantitative measurement of similarity  
44 that is not dependent on such characters and although it is not without its own limitations,  
45 it has fundamentally changed our approach to studying diversity in organisms such as  
46 *Entamoeba*.

47 In this time period, the study of *Entamoeba* has gone from being dependent on  
48 stable laboratory cultures of parasites, preferably in the axenic form, to DNA analysis of  
49 organisms directly from stool samples in the absence of even microscopic investigation.  
50 The latter aspect has been problematic as it is not possible to assign new sequences to  
51 previously named species where the original description is dependent on morphology.  
52 For this reason, many new and distinct *Entamoeba* sequences have been assigned to  
53 ‘ribosomal lineages’ rather than species to reflect the absence of morphological  
54 information (Stensvold et al. 2011).

55 The new approach is dependent on the reliability of DNA purification from stool  
56 samples, which are notorious for the presence of enzyme inhibitors, and the specificity of  
57 the primers used for PCR amplification. In addition, investigation of *Entamoeba* in such  
58 samples is largely limited to the ribosomal RNA genes due to the complexity of the  
59 DNAs extracted from stool, which often contains DNA from multiple other parasites and  
60 may include multiple *Entamoeba* species. The elimination of culture dependency has led  
61 to a dramatic expansion in the number of genetically distinct *Entamoeba* organisms being  
62 recognized but also to a greater understanding of sequence variability within species due  
63 to the relative ease with which multiple samples can be studied in parallel. The present  
64 report contributes information on several novel *Entamoeba* lineages as well as variation  
65 within and host range of known species.

66  
67 **MATERIALS AND METHODS**

68

69 **Cultures**

70 *E. bangladeshi* strains 8111 and 8237 were received as xenic cultures from Dr Rashidul  
71 Haque, ICDDR, B, Bangladesh. Partial sequences of their small subunit ribosomal RNA  
72 genes (SSU rDNA) were published in the original species description (Royer et al. 2012)  
73 but as these only covered ca. 20% of the full gene, the complete sequences were obtained  
74 to allow more accurate phylogenetic investigation. The organisms were grown in  
75 LYSGM with 5% adult bovine serum and rice starch at 22 °C and subcultured twice  
76 weekly (Clark and Diamond 2002).

77 *E. invadens* VK-1 was received as an axenic culture from Dr Avelina Espinosa  
78 (Roger Williams University, United States) and grown in LYI-S-2 (Clark and Diamond  
79 2002) with 15% adult bovine serum at 22 °C. The complete SSU rDNA was sequenced to  
80 investigate intra-specific variation in this species.

81

82 **Stool samples**

83 Stool samples came from a variety of sources. Most were collected by MSc students as  
84 part of a parasitology field trip in two consecutive years. Asian elephant (*Elephas*  
85 *maxima*) stool samples were obtained from Amsterdam Zoo, courtesy of Mark J. Hoyer  
86 and Daphne Valk through Dr Bruno Levecke, University of Ghent. Samples were either  
87 extracted as fresh material or stored (at 4 °C) in 70% ethanol at a ratio of 3:1 (v:v) for  
88 later processing. Some of the DNA samples used were from a previously published study  
89 (Alfellani et al. 2013).

90

91 **DNA purification from stool**

92 When present, ethanol was removed from stool samples by washing in phosphate  
93 buffered saline (pH = 8) three times prior to processing. DNA was extracted from stool  
94 using the Qiagen DNA stool minikit (Qiagen, Hilden, Germany) according to the  
95 manufacturer's instructions.

96

97 **Amplification and sequencing**

98 Primers used, amplification and sequencing conditions were all essentially as described  
99 previously unless otherwise stated (Stensvold et al. 2011). Initial analysis of samples  
100 involved amplification of purified DNA using the genus-specific primer pairs Entam 1/2  
101 (Verweij and Stensvold 2014) or Entagen F/R (Stensvold et al. 2011). PCR conditions  
102 consisted of 30 cycles of 30 sec each at 94 °C, 59 °C and 72 °C with a final extension of  
103 72 °C for 10 min.

104

105 **Phylogenetic analysis**

106 New sequences were aligned with reference sequences from Stensvold et al. (2011),  
107 using the alignment tool MUSCLE as implemented in MEGA 5 (Tamura et al., 2011).  
108 The alignment was edited manually to remove regions of ambiguity resulting in an  
109 alignment of 1,447 positions. Phylogenetic analyses were performed using distance  
110 (Neighbor-Joining (NJ); distance matrix in Supplementary Data File S2) and Maximum  
111 Likelihood (ML) algorithms as implemented in MEGA5 and Bayesian analysis (MrBayes  
112 3.1.5; Huelsenbeck and Ronquist, 2001). Bayesian and ML analysis used a General Time  
113 Reversible (GTR) model of nucleotide substitution with four categories of among-site  
114 rate variation and the proportion of invariant sites, the best model selected by ModelTest,

115 implemented in MEGA5. Statistical support for distance and maximum likelihood trees  
116 was evaluated using bootstrapping (1,000 replicates). Bayesian analysis used four  
117 Markov chain Monte Carlo (MCMC) strands, 1,000,000 generations, with trees sampled  
118 every 100 generations. The resulting average standard deviation of split frequencies was  
119 less than 0.01. A consensus tree was produced after excluding an initial burn-in of 25%  
120 of the samples, as recommended.

121 For the analysis of the microbiome sequence data, sequences were extracted from  
122 the curated SILVA 108 database ([http://qiime.org/home\\_static/dataFiles.html](http://qiime.org/home_static/dataFiles.html)) generated  
123 by Parfrey et al. (2014) which are also available as Supplementary Data File S1. After  
124 alignment with relevant reference sequences and editing, as above, the same phylogenetic  
125 methods were employed, except that for Bayesian analysis 2,000,000 generations were  
126 used and the standard deviation of split frequencies stabilized at 0.027.

127

## 128 **Terminology**

129 Because non-standard nomenclature is used to describe *Entamoeba* diversity, we here  
130 define some of the terms used. These explain our working criteria based on the data  
131 available at this time.

132 Species: *Entamoeba* species with Latin binomials have been described primarily  
133 on the basis of morphology and host. More recently some have incorporated molecular  
134 data into the species definition. We will use the term “species” only where a Latin  
135 binomial has been published elsewhere. A sequence identified as belonging to a named  
136 species will exhibit a high percentage identity to sequences derived from morphologically  
137 verified organisms and will cluster specifically with such sequences to the exclusion of  
138 all others with high statistical support.

139 Subtype (ST): DNA sequences that cluster as a discrete clade within the range of  
140 diversity of a defined species. The identification of a new ST must be based on SSU  
141 rDNA sequences where gene coverage is  $\geq 80\%$ . Where STs are defined, all sequences  
142 within the species must be demarcated into STs. Sequence divergence within a defined  
143 ST will not normally be greater than 3%.

144 Ribosomal lineage (RL): Organisms for which  $\geq 80\%$  of the SSU rDNA gene has  
145 been sequenced, that differ from previously known sequences by 5% or more and where  
146 there is no morphological information are assigned RL numbers. It is possible that RLs  
147 could become species in the future if morphological and other relevant data become  
148 available, but we do not think it appropriate to assign names based only on a single SSU  
149 rDNA sequence.

150 Conditional lineage (CL): When a divergent sequence does not meet the criteria  
151 for a new ribosomal lineage because  $\leq 80\%$  coverage of the SSU rDNA has been  
152 obtained, we propose identifying it as a conditional lineage. Such lineages are likely to  
153 represent novel organisms and to be “promoted” to RLs or species when more data  
154 become available.

155

## 156 **RESULTS AND DISCUSSION**

157

### 158 **Molecular species identification and survey overview**

159 In addition to a few cultures, stool samples were obtained from a wide range of hosts in a  
160 number of locations, including farms and zoos, as well as from trapped wild hosts.

161 Culture material gave no problems with either amplification or sequencing. However,  
162 with stool DNA, *Entamoeba* sequences were preferentially amplified when present but,  
163 when absent, non-specific products from the SSU rDNA of fungi and plants were  
164 sometimes produced. It is likely that some of the unreadable sequences obtained were due  
165 to mixed products of multiple origins. PCR positivity varied dramatically among hosts,  
166 with sheep and cattle being the hosts most frequently infected with *Entamoeba*.

167 Where a partial sequence obtained showed 98% or greater identity to reference  
168 sequences, the *Entamoeba* species or RL was considered to have been identified and no  
169 further investigation was undertaken. However samples yielding sequences with lower  
170 identities, 95% or less, were subjected to further amplification and sequencing, usually  
171 involving broad-specificity primers paired with sequence-specific primers designed from  
172 the initial sequence data obtained. DNA sample amount was a significant limitation, in  
173 some cases leading to the DNA being depleted before the complete gene sequence could  
174 be obtained. There were nevertheless sufficient data to allow phylogenetic analyses to be  
175 undertaken in most cases.

176 The results from our survey implementing this approach are presented in Table 1  
177 and include the identification of four new ribosomal lineages (RLs 8, 9, 10 and 11).  
178 Phylogenetic analysis using their SSU rDNA sequences (Fig. 1) identified their  
179 relationships to previously described species and lineages.

180

### 181 ***Entamoeba* from artiodactyls**

182 As the majority of ungulate samples in our study were obtained from ruminant artiodactyl  
183 species it is of little surprise that this group displays the highest number of *Entamoeba*  
184 PCR positive samples, given previous results (Stensvold et al. 2011). We did note  
185 however, that the amplification of non-specific PCR products (fungi, plants and other  
186 protists) was a particular problem when dealing with ruminant faecal material and that  
187 this may have precluded the identification of *Entamoeba* in some samples.

188 *Entamoeba in cattle.* In cattle, an *Entamoeba* positivity rate of 44.4% was  
189 observed. *E. bovis* was the predominant species with single occurrences of *Entamoeba*  
190 RL2, *Entamoeba* RL4 and *E. moshkovskii*. Three samples were found to contain mixed  
191 infections. The most abundant species, as determined from the highest peaks in the  
192 sequence trace chromatograms, was *E. bovis* and the minority species was undetermined.  
193 This observation is in keeping with our previous findings that it is possible for a single  
194 animal to be infected with more than one lineage of *Entamoeba* (Stensvold et al. 2011).

195 It is of particular interest that *E. moshkovskii* was detected in cattle from Devon,  
196 UK. To our knowledge this represents the first example of this species being detected in a  
197 ruminant animal, having been reported previously only in humans (Heredia et al. 2012)  
198 and aquatic turtles (Garcia et al. 2014), in addition to environmental samples. The  
199 detection of *Entamoeba* RL4 in Hertfordshire, UK, further expands the geographic  
200 distribution of this lineage, which has been detected previously only in cattle from Libya  
201 and Estonia (Stensvold et al. 2011).

202 A new lineage was discovered in cattle from Devon (sequence Cow5) which we  
203 identify here as *Entamoeba* RL8 (Fig. 1). Our phylogenetic analysis suggests that  
204 *Entamoeba* RL8 is most closely related to *Entamoeba* RL1, a sequence from Roe Deer,  
205 also a ruminant artiodactyl, albeit with weak support.

206 *Entamoeba in sheep.* The prevalence of *Entamoeba* in sheep (49.1%) was slightly  
207 higher than that found in cattle. Again, *E. bovis* was the predominate species and the  
208 remainder of the positive samples were represented by the single detection of *Entamoeba*  
209 RL2 and three mixed *E. bovis* samples. The identification of *Entamoeba* RL2 in this  
210 group further expands the host range of this lineage, which has previously only been  
211 found in cattle (Stensvold et al. 2011).

212 To date, the molecular sampling of sheep populations has failed to identify a  
213 distinct lineage that could correspond to the species *E. ovis* (Noble and Noble, 1952).  
214 Phylogenetic analysis of two sheep derived DNA sequences (Stensvold et al. 2010) and  
215 the sequence data from the present survey (derived from 27 *Entamoeba*-positive  
216 specimens) show that sequences obtained from sheep do not form a unique host-specific  
217 clade. Based on this evidence, we therefore suggest that *E. ovis* is a junior synonym and  
218 that *E. bovis* infects both cattle and sheep.

219 *Entamoeba in deer.* Nine deer were sampled, with a positivity rate of 22.2%. The  
220 discovery of *E. bovis* in two fallow deer from Mauritius gives a new host for this species,  
221 having only been detected in reindeer (*Rangifer tarandus*) previously (Clark et al. 2006).  
222 Our data confirm that *E. bovis* has a wide geographic distribution and a wide host range  
223 which now encompasses cattle, sheep and deer, all of which are ruminant artiodactyls.

224 *Entamoeba in pigs.* In non-ruminant artiodactyls, without exception the species  
225 present in PCR positive samples (52.4%) derived from pigs was *E. polecki* ST1. Three  
226 specimens were found to have a mixed infection consisting of two closely related  
227 subtypes (ST1 and ST3). It is noteworthy that *E. suis* was not detected. A study of five  
228 pigs taken from a population of 148 displaying hemorrhagic colitis symptoms in Japan  
229 suggested that *E. suis* was responsible for the disease (Matsubayashi et al., 2014).  
230 Conversely, *E. suis* was detected in 28 specimens taken from pigs in Vietnam and no ill  
231 health was reported (Clark et al. 2006).

232 *Entamoeba bovis-related lineages.* The shaded area in Fig. 1 identifies the current  
233 phylogenetic relationships of *E. bovis* and related ribosomal lineages. We find that *E.*  
234 *bovis* and *Entamoeba* RLs 1–3 plus the newly defined *Entamoeba* RL8 form a strongly  
235 supported monophyletic clade. However, within this clade we were unable to resolve the  
236 relationships among the lineages. The grouping of *E. bovis* and *Entamoeba* RLs 1–3  
237 together as a robust monophyletic clade with high bootstrap support and low resolution is  
238 consistent with previous tree reconstructions (Stensvold et al. 2011).

239 In need of further comment is the continued, seemingly incongruous, inclusion of  
240 *Entamoeba* RL3, a sequence isolated from langurs, within a clade otherwise consisting of  
241 sequences from ruminant artiodactyls. Langurs of the genera *Trachypithecus* and  
242 *Semnopithecus* are unique amongst primates for their possession of a ruminant-like  
243 stomach (Bauchop and Martucci 1968). We speculate that the similarity of the langur  
244 ruminant stomach to that found in artiodactyls provides environmental and physiological  
245 conditions in which an *Entamoeba* lineage related to *E. bovis* was able to colonize and  
246 subsequently become host-adapted. It should be pointed out that, at present, no stool  
247 samples of artiodactyls from East, South and Southeast Asia have been investigated.

248 *Entamoeba in microbiome data.* During the course of our survey work, a study  
249 was published that utilised high-throughput 454-amplicon pyrosequencing to  
250 systematically investigate the eukaryotic communities in mammalian gut microbiota,  
251 including *Entamoeba* species (Parfrey et al. 2014). The study included samples obtained

252 from captive herbivores. Some of these were positive for *Entamoeba* and are listed in  
253 Table 2. A preliminary analysis of the sequences from this dataset revealed that one of  
254 the sequences obtained from an Okapi was identical to our newly identified lineage  
255 *Entamoeba* RL8, isolated from cattle. This serendipitous discovery prompted us to  
256 further analyze *Entamoeba* sequences isolated from this dataset and the phylogenetic  
257 reconstruction is presented in Fig. 2.

258 Since the data are derived from the short reads (ca 500 bp) generated by 454-  
259 amplicon pyrosequencing we are unable to assign new ribosomal lineages on the basis of  
260 the guidelines for *Entamoeba* nomenclature (Stensvold et al. 2011). Instead, we refer to  
261 the several potentially new lineages as “conditional lineages” (CL; see Methods section).  
262 The resultant phylogenetic tree shows the relationships of the 454 sequences related to *E.*  
263 *bovis* plus RLs 1-4 plus 8. Five newly defined *Entamoeba* clades, CLs3-7, are identified.

264 The shaded region of the resulting cladogram shows an expanded version of the  
265 *E. bovis* clade, which loosely comprises six internal clusters (Fig. 2). While the  
266 monophyly of the *E. bovis* clade is strongly supported in the ML analysis (95%) it is  
267 poorly supported in the distance-based analysis (67%), and is absent in the Bayesian  
268 analysis. The latter results from the inclusion of a single sequence from an Okapi (Okapi2  
269 92564 in Fig. 2) within the *E. bovis* clade in this analysis only. Otherwise, the clade  
270 contains the same sequences in all analyses and has strong Bayesian posterior probability  
271 support. This sequence appears as a distinct lineage, identified as *Entamoeba* CL3, in  
272 both ML and NJ analyses.

273 A number of sequences were identified in two Okapi, a ruminant artiodactyl,  
274 sampled in this survey and both specimens revealed known and novel lineages of  
275 *Entamoeba*. A total of eight sequences were obtained from the Okapi2 sample and these  
276 included five belonging to the *E. bovis* clade. One corresponds to the newly defined  
277 *Entamoeba* RL8, robustly supported in all three analysis methods, while the remaining  
278 two sequences from this host (Okapi2 92564 and Okapi2 39254) were assigned to the  
279 putative ribosomal clades *Entamoeba* CL3, which has no close relatives, and *Entamoeba*  
280 CL4, respectively. The existence of *Entamoeba* CL4 is supported in ML (92%) and  
281 Bayesian inference, but not NJ analysis. In contrast, just three *Entamoeba* sequences  
282 were obtained from sample Okapi1: two of these fall into *Entamoeba* CL4 and one  
283 (Okapi16235) into the newly defined *Entamoeba* CL5 with high support in both ML and  
284 Bayesian inference analysis.

285 One of the sampled gazelles (Gazelle3) yielded an astonishing 14 distinct  
286 *Entamoeba* sequences, 11 of which cluster within the *E. bovis* clade, two with *Entamoeba*  
287 RL8 and one with *Entamoeba* RL1, which expands the host range of that lineage. A big  
288 horn sheep sample (BigHornSD) yielded 10 distinct sequences of which 7 were found to  
289 cluster within *E. bovis*, one with *Entamoeba* RL1 (also expanding the host range of this  
290 lineage), and two within the newly defined *Entamoeba* CL5.

291 There were three sequences found to have no close relatives in our phylogenetic  
292 reconstruction; sequences from a kangaroo (*Entamoeba* CL7), an okapi (*Entamoeba*  
293 CL3) and a wild ass (*Entamoeba* CL6). This survey also demonstrates that the same host  
294 species at the same location can carry different lineages of *Entamoeba*, in that the  
295 sequences obtained from two okapi living in the same herd showed differences in lineage  
296 representation.

297 Finally, the detection of *E. bovis* in a kangaroo represents a very different new  
298 host for this species. Macropods (kangaroos and wallabies) are foregut fermenters but do  
299 not have the ruminant stomach structure seen in cattle. It is possible that this *Entamoeba*  
300 was transient and not established within the host. Only further sampling of macropods  
301 will help to establish whether *E. bovis* is a normal member of their gut fauna.

302 The data presented in Fig. 2 reveal not only a remarkable degree of diversity  
303 within the known *E. bovis* and related lineages but gives an insight into *Entamoeba*  
304 diversity within a single host. Application of this approach to analyzing the eukaryotic  
305 microbiome in a range of host samples is likely to become the method of choice in the  
306 future for detecting diversity, although the short sequences obtained are not ideal for  
307 phylogenetic analyses.

308

### 309 ***Entamoeba* from elephants.**

310 There have been three parasitological surveys of wild elephants in the past decade. The  
311 first documented nematode and ciliate populations in the stool of African forest elephants  
312 (Kinsella et al. 2004), a second reported helminth and coccidian parasites from African  
313 elephants in Botswana (Baines et al. 2015), while the other selectively concentrated on  
314 nematode eggs in Asian elephants (Hing et al. 2013). None of the studies reported finding  
315 *Entamoeba*.

316 To the best of our knowledge, this is the first report of *Entamoeba* in elephants.  
317 The sequence obtained from an Asian elephant living in Amsterdam Zoo represents a  
318 novel lineage, which we define as *Entamoeba* RL10. *E. moshkovskii* has also been  
319 identified, in an African elephant in a zoo setting (Table 2) (Parfrey et al. 2014). How  
320 widespread *Entamoeba* RL10 is, whether it is found in both Asian and African elephants,  
321 and whether it is specific to elephants will require additional surveys to be undertaken. It  
322 is also important that wild elephants be sampled to rule out a captivity-acquired infection.  
323 It is important that such surveys be molecular in nature, as no cysts were detected in the  
324 *Entamoeba*-positive sample from the Asian elephant. These might have been expected to  
325 be 4-nucleated owing to their relationship of RL10 to *E. hartmanni*. It could be  
326 speculated that the absence of cysts might explain why no *Entamoeba* has previously  
327 been reported from elephants (Kinsella et al. 2004) but raises the question of how it is  
328 transmitted.

329 Our phylogenetic analysis (Fig. 1) demonstrates that *Entamoeba* RL10 is closely  
330 related to *E. hartmanni* (having a pairwise sequence identity of 95.7%) and these two  
331 sequences consistently form a clade with very high bootstrap support in the recovered  
332 trees. The placement of *Entamoeba* RL10 as sister taxon to *E. hartmanni* is significant  
333 because the latter species has not been found to have any close relatives in previous  
334 phylogenetic reconstructions (Stensvold et al. 2011).

335 There is a tendency for NJ analyses to cluster the *Entamoeba* RL10/*E. hartmanni*  
336 clade with the *E. ranarum*/*E. invadens* clade. Support is weak but this finding is in  
337 agreement with previous studies (Stensvold et al. 2010, Clark et al. 2006). In contrast, our  
338 most recent study did not recover this relationship (Stensvold et al. 2011) suggesting that  
339 it may be sensitive to the sequences included in the alignment. It is clear that finding  
340 further lineages related to *E. hartmanni* will help to resolve the relationships of this well-  
341 defined clade.

342



343 ***Entamoeba moshkovskii* and its relatives**

344 The finding of *E. moshkovskii* in elephant and cattle further broadens the host range of  
345 this species. It is clear that this usually free-living species, which exists as a species  
346 complex consisting of multiple variants (Clark and Diamond 1997), is being detected in  
347 animal hosts more frequently, particularly in humans. Since the advent of molecular  
348 detection tools there has been growing interest in *E. moshkovskii* as there is speculation  
349 that it may be a facultative parasite (Heredia et al. 2012). Previous phylogenetic  
350 reconstructions have consistently placed *E. moshkovskii* as a sister lineage to a clade  
351 consisting of *E. dispar*, *E. nuttalli*, *E. histolytica* and *E. ecuadoriensis*. The latter has only  
352 been isolated once, from sewage, and like *E. moshkovskii* is considered to be potentially  
353 free-living (Stensvold et al. 2011, Stensvold et al. 2010). *E. bangladeshi* is the most  
354 recent species to be described from humans. In phylogenetic reconstructions based on  
355 partial SSU rDNA sequences, it was found to branch between *E. moshkovskii* and *E.*  
356 *ecuadoriensis* (Royer et al. 2012). However, our analysis based on the complete SSU  
357 rDNA sequences (KR025411 and KR025412) specifically positions *E. bangladeshi* in a  
358 highly supported clade with the latter species. *E. bangladeshi*, like *E. moshkovskii* and *E.*  
359 *ecuadoriensis*, is able to grow at both body temperature and room temperature,  
360 suggesting it might also be found in the environment in the future.

361

362 ***Entamoeba* from horses**

363 The species name *E. equi* was first used to describe 4-nucleated *Entamoeba* cysts in  
364 horses from South Africa (Fantham, 1921). It was subsequently resurrected for a DNA  
365 sequence obtained from a horse in Aberystwyth, UK (Clark et al. 2006). As the species  
366 name suggests, the host range of this species appears to be confined to the family  
367 Equidae, which also includes donkeys and zebras. The later finding of *E. equi* in a zebra  
368 (*Equus zebra hartmannae*) from a zoo in the UK (Stensvold et al. 2011) appeared to  
369 support this host range and species designation, although no cysts were seen in either  
370 sample.

371 A new ribosomal lineage, *Entamoeba* RL9, was detected in three horses from  
372 various locations in Devon. It occupies a position in the phylogenetic tree well removed  
373 from the distinct *E. equi* lineage (Fig. 1). Our phylogenetic reconstruction shows the  
374 placement of *Entamoeba* RL9 as a sister taxon to *Entamoeba* RL4 (a lineage associated  
375 with cattle) but this relationship is not well supported by either NJ or ML analyses. The  
376 exact position of *Entamoeba* RL9 may become clearer with further sampling. An  
377 intriguing observation is that although the multiple *Entamoeba* RL9 DNA sequences  
378 obtained are closely related, they are not identical.

379 In common with the findings for *E. equi*, cysts were not detectable  
380 microscopically in the new equine samples containing *Entamoeba* RL9 using standard  
381 methodologies. This suggests that horse physiology may be responsible for the absence  
382 of cysts rather than this being a species-specific trait of *E. equi*. It is also possible that the  
383 apparent absence of cysts in horses (and elephants) is related to their scarcity or to  
384 periodic shedding. Only more detailed investigation can solve this conundrum.

385 454-amplicon pyrosequencing data (Table 2) also identified *E. hartmanni* in a  
386 zebra (Parfrey et al. 2014), a new host for this species which has previously only been  
387 detected in primates. The latter observation again raises the question of whether zoo  
388 hosts are true natural reservoirs of the *Entamoeba* lineages being detected or if we are

389 observing a captivity-acquired infection, in the same way that wild birds have been  
390 identified as a possible source of *Plasmodium* spp. in captive penguins (Leclerc et al.  
391 2014). To date, information on *Entamoeba* in wild animals is sparse.

392

### 393 ***Entamoeba* from rodents**

394 There have been a number of morphological descriptions of *Entamoeba* spp. from  
395 rodents in the literature. These reports range from the morphologically indistinguishable  
396 4-nucleated cysts of *E. histolytica* and *E. dispar* in rats (Mishra and Gonzalez 1975,  
397 Shafiyah et al. 2012, Neal 1948) to the 8-nucleated cyst former *E. muris*, detectable in  
398 both wild (El-Ridi et al. 1987, Nateghpour et al. 2015) and laboratory rat populations  
399 (Won et al. 2006). Other 8-nucleated cyst forming species such as *E. funambulae* (Ray  
400 and Banik 1964) and *E. citelli* (Davis 1969, Diakou et al. 2015) have been reported in  
401 squirrel populations. Finally, there have been reports of *Entamoeba* spp. which also  
402 formed 8-nucleated cysts in Syrian hamsters (Neal 1947).

403 In contrast, there is a dearth of molecular sequence data relating to *Entamoeba* in  
404 rodents. Only recently has PCR been used to detect and differentiate *E. histolytica* and *E.*  
405 *dispar* in rats, which were shown by phylogenetic analysis of partial SSU rDNA  
406 sequences to cluster with those typically found in primates (Lau et al., 2014).

407 Furthermore there is only one reliable GenBank entry for *E. muris*, isolated from  
408 Mongolian gerbil (Kobayashi et al. 2009); the other sequence listed as *E. muris*  
409 (FN396613), isolated from *Rattus rattus*, shares 100% identity with *E. coli* ST2, also an  
410 8-nucleated cyst former, and is likely to have been misidentified because of the host.

411 There are no GenBank sequence entries for either *E. funambulae* or *E. citelli*.

412 Our finding of an *Entamoeba* sequence in DNA extracted from a stool sample  
413 from a field vole is significant. The full-length SSU rDNA sequence is here defined as  
414 *Entamoeba* RL11 and represents a novel lineage from a host that has not been sampled  
415 previously. There is moderate bootstrap support in both NJ (84%) and ML (82%)  
416 analyses and strong Bayesian support for placing this new sequence in a clade with *E.*  
417 *muris*. The *Entamoeba* RL11/*E. muris* clade is sister to *Entamoeba* RL7 with high  
418 support (> 95%) in all analyses. The clustering of *E. muris* with *Entamoeba* RL7, which  
419 has been found in langurs and humans (A. Vidal-Lapiedra pers. commun.), has been  
420 previously established (Stensvold et al. 2011). Only future microscopic analysis will  
421 confirm if *Entamoeba* RL11 forms 8-nucleated cysts as seen in *E. muris* and *Entamoeba*  
422 RL7.

423 Relatively few rodent hosts have been sampled to date so it is quite likely that  
424 more *Entamoeba* diversity remains to be detected. In our experience, rodent stool  
425 presents a particular problem unless fresh in that fungi growth on the pellet can  
426 potentially mask any *Entamoeba* species sequences present in the extracted DNA, as  
427 mentioned earlier for ungulates.

428

### 429 ***Entamoeba* from reptiles**

430 *E. invadens* produces 4-nucleated cysts and is the most important *Entamoeba* infection of  
431 reptiles since it is the causative agent of invasive amebiasis (Geiman and Ratcliffe 1936)  
432 and amebic myositis (Chia et al. 2009). Our sequencing here of the complete SSU rDNA  
433 of *E. invadens* VK-1 (KR025413), from a Komodo Dragon, identified only a single base  
434 difference when compared to the reference strain IP-1, isolated from a snake. In

435 combination with the other existing sequence data (Garcia et al. 2014) and the results  
436 from restriction enzyme digestion (Clark and Diamond 1997), this would suggest that  
437 intra-specific SSU rDNA sequence diversity in this species is present but low.

438 A number of other *Entamoeba* species have been identified microscopically in  
439 reptiles. These include the 4-nucleated cyst formers *E. insolita* from Galapagos tortoises  
440 (Gieman and Wichterman 1937) and *E. terrapinae* from terrapins. An 8-nucleated cyst  
441 former, *E. barreti*, has also been reported in Snapping turtle (Geiman and Ratcliffe 1936).  
442 In terms of sequence data, the disease-causing agent *E. invadens* constitutes the majority  
443 of DNA sequences deposited in GenBank. *E. insolita* (Silberman et al. 1999); *Entamoeba*  
444 RL5 from Leopard tortoise (Stensvold et al. 2011) and *Entamoeba* RL6 from Iguana  
445 (Silberman et al. 1999) are all represented by single entries. There are no sequence data  
446 for *E. barreti*.

447 In our stool survey, the single python sample proved to be negative for  
448 *Entamoeba*, but a novel conditional lineage (*Entamoeba* CL1) was detected in a Giant  
449 Aldabran tortoise (Table 1). Unfortunately due to the complexity of the specimen, which  
450 also contained *Nyctotherus ovalis* and *E. insolita*, only a partial sequence (470 bp) could  
451 be obtained, and it was not possible to assign a new ribosomal lineage number using the  
452 published nomenclature criteria (Stensvold et al. 2011). Inclusion of this sequence in our  
453 dataset, followed by editing to produce an unambiguous alignment, resulted in a further  
454 reduction in the number of useable characters within this sequence (368bp). This in turn  
455 resulted in an unresolved and unstable branch that was not specifically linked to any of  
456 the other sequences. For this reason, it was decided to exclude this sequence from the  
457 final phylogenetic analyses.

458 A study of 127 aquatic turtles (Garcia et al. 2014) found that over half (58.7%)  
459 were infected with *Entamoeba*. *E. terrapinae* was found to be the most prevalent species  
460 and was present in 63 turtles, followed by *E. invadens* (6 turtles) and *E. moshkovskii* (5  
461 turtles). This represents the first report of *E. moshkovskii* in reptiles. An acknowledged  
462 limitation of this study was that it relied on culturing the amoebae rather than direct  
463 detection in faecal DNA samples. The same study found a novel putative ribosomal clade  
464 represented by three sequences, which we define here as *Entamoeba* CL2.

465 Our phylogenetic analyses show that the *Entamoeba* CL2 sequence is distinct  
466 from *E. terrapinae*, and there is poor bootstrap support in both NJ and ML analyses to  
467 support a sister relationship (Fig. 1). The various lineages of reptilian *Entamoeba* appear  
468 to form a limited number of clusters in the tree at present (*E. insolita* with *Entamoeba*  
469 RL6 and *Entamoeba* RL5, *E. terrapinae* with the newly described lineage) although  
470 without strong support. The present as well as prior phylogenetic reconstructions  
471 (Stensvold et al. 2011, Stensvold et al. 2010) consistently recover *E. invadens* in a sister  
472 relationship with *E. ranarum* (Silberman et al. 1999), the only amphibian-derived  
473 *Entamoeba* sequence available to date. As yet, no sequences from reptiles cluster with the  
474 8- or 1-nucleated cyst-producing lineages from mammals, although cysts with eight  
475 nuclei have been reported from reptiles on several occasions.

476

#### 477 **Concluding remarks**

478 Our current survey and phylogenetic reconstructions have further expanded our  
479 knowledge of the diversity and host range of *Entamoeba* species. We have discovered  
480 new hosts for known *Entamoeba* lineages, the most striking of which is perhaps the

481 reporting of *E. moshkovskii* in cattle and elephants. Furthermore, the finding of four new  
482 ribosomal lineages and a further seven ribosomal clades during the course of this work  
483 show the merit of the continued sampling of livestock and wild animals.

484 A notable observation from the present work is the consistently higher incidence  
485 of *Entamoeba* infection in animals from managed herds and animals in captivity. In  
486 contrast, there was low or no *Entamoeba* positivity in wild animals, even in wild  
487 ungulates. It is not clear at present whether the lower rate of *Entamoeba* infection  
488 detected in wild animals is an artifact, is due to this group being relatively under-sampled  
489 or is real, and only further sampling from wild animal populations can answer this  
490 question.

491

#### 492 **ACKNOWLEDGMENTS**

493 The work described in this manuscript was derived in part from the MSc thesis projects  
494 of EJB, ADL and NK. We are very grateful to Dr. Laura Wegener Parfrey for providing  
495 us with the *Entamoeba* sequences extracted from her dataset, and for encouraging us to  
496 submit the sequences as supplementary information with this manuscript to make them  
497 more easily accessible.

498

499

#### 500 **LITERATURE CITED**

- 501 Alfellani, M. A., Taner-Mulla, D., Jacob, A. S., Imeede, C. A., Yoshikawa, H., Stensvold,  
502 C. R. & Clark, C. G. 2013. Genetic diversity of *Blastocystis* in livestock and zoo  
503 animals. *Protist*, **164**:497-509.
- 504 Baines, L., Morgan, E. R., Ofthile, M. & Evans, K. 2015. Occurrence and seasonality of  
505 internal parasite infection in elephants, *Loxodonta africana*, in the Okavango Delta,  
506 Botswana. *Int J Parasitol: Parasites Wildl*, **4**:43-48.
- 507 Bauchop, T. & Martucci, R. W. 1968. Ruminant-like digestion of the langur monkey.  
508 *Science (New York, N.Y.)*, **161**:698-700.
- 509 Chia, M. Y., Jeng, C. R., Hsiao, S. H., Lee, A. H., Chen, C. Y. & Pang, V. F. 2009.  
510 *Entamoeba invadens* myositis in a common water monitor lizard (*Varanus*  
511 *salvator*). *Vet Pathol*, **46**:673-676.
- 512 Clark, C. G. & Diamond, L. S. 2002. Methods for cultivation of luminal parasitic protists  
513 of clinical importance. *Clin Microbiol Rev*, **15**:329-341.
- 514 Clark, C. G. & Diamond, L. S. 1997. Intraspecific variation and phylogenetic  
515 relationships in the genus *Entamoeba* as revealed by riboprinting. *J Eukaryot*  
516 *Microbiol*, **44**:142-154.
- 517 Clark, C. G., Kaffashian, F., Tawari, B., Windsor, J. J., Twigg-Flesner, A., Davies-Morel,  
518 M. C., Blessmann, J., Ebert, F., Peschel, B., Le Van, A., Jackson, C. J.,  
519 Macfarlane, L. & Tannich, E. 2006. New insights into the phylogeny of  
520 *Entamoeba* species provided by analysis of four new small-subunit rRNA genes.  
521 *Int J Syst Evol Microbiol*, **56**:2235-2239.
- 522 Davis, S. D. 1969. Hibernation: intestinal protozoa populations in ground squirrels. *Exp*  
523 *Parasitol*, **26**:156-165.
- 524 Diakou, A., Kapantaidakis, E. & Youlatos, D. 2015. Endoparasites of the European  
525 ground squirrel (*Spermophilus citellus*) (Rodentia: Sciuridae) in central  
526 Macedonia, Greece. *J Nat Hist*, **49**:359-370.

- 527 El-Ridi, A. M. S., Ismail, M. M., El-Gamal, R. L. R. & Abdel-Fattah, M. E. M. 1987.  
528 Protozoal infections in rodents in Sharkiya Governorate, Egypt. *J Egypt Soc*  
529 *Parasitol*, **17**:275-280.
- 530 Fantham, H. B. 1921. Some parasitic protozoa found in South Africa: IV. *South Afr J Sci*,  
531 **17**:164–170.
- 532 Garcia, G., Ramos, F., Perez, R. G., Yanez, J., Estrada, M. S., Mendoza, L. H., Martinez-  
533 Hernandez, F. & Gaytan, P. 2014. Molecular epidemiology and genetic diversity  
534 of *Entamoeba* species in a chelonian collection. *J Med Microbiol*, **63**:271-283.
- 535 Geiman, Q. M. & Ratcliffe, H. L. 1936. Morphology and life-cycle of an amoeba  
536 producing amoebiasis in reptiles. *Parasitology*, **28**:208-228.
- 537 Geiman, Q. M. & Wichterman, R. 1937. Intestinal protozoa from Galapagos tortoises (with  
538 descriptions of three new species). *J Parasitol*, **23**:331-347.
- 539 Heredia, R. D., Fonseca, J. A. & López, M. C. 2012. *Entamoeba moshkovskii*:  
540 perspectives of a new agent to be considered in the diagnosis of amebiasis. *Acta*  
541 *Tropica*, **123**:139-145.
- 542 Hing, S., Othman, N., Nathan, S. K. S. S., Fox, M., Fisher, M. & Goossens, B. 2013. First  
543 parasitological survey of endangered Bornean elephants *Elephas maximus*  
544 *borneensis*. *Endangered Species Research*, **21**:223-230.
- 545 Huelsenbeck, J. P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic  
546 trees. *Bioinformatics*, **17**:754–755.
- 547 Kinsella, J. M., Deem, S. L., Blake, S. & Freeman, A. 2004. Endoparasites of African  
548 forest elephants (*Loxodonta africana cyclotis*) from the Republic of Congo and  
549 Central African Republic. *Comp Parasitol*, **71**:104-110.
- 550 Kobayashi, S., Suzuki, J. & Takeuchi, T. 2009. Establishment of a continuous culture  
551 system for *Entamoeba muris* and analysis of the small subunit rRNA gene.  
552 *Parasite*, **16**:135-139.
- 553 Lau, Y.L., Jamaiah, I., Rohela, M., Fong, M.Y., Siti, C.O. & Siti, F.A. 2014. Molecular  
554 detection of *Entamoeba histolytica* and *Entamoeba dispar* infection among wild  
555 rats in Kuala Lumpur, Malaysia. *Trop Biomed*, **31**:721-727.
- 556 Leclerc, A., Chavatte, J. M., Landau, I., Snounou, G. & Petit, T. 2014. Morphologic and  
557 molecular study of hemoparasites in wild corvids and evidence of sequence  
558 identity with *Plasmodium* DNA detected in captive Black-Footed Penguins  
559 (*Spheniscus demersus*). *J Zoo Wildl Med*, **45**:577-588.
- 560 Matsubayashi, M., Suzuta, F., Terayama, Y., Shimojo, K., Yui, T., Haritani, M., &  
561 Shibahara, T. 2014. Ultrastructural characteristics and molecular identification of  
562 *Entamoeba suis* isolated from pigs with hemorrhagic colitis: implications for  
563 pathogenicity. *Parasitol Res*, **113**:3023-3028.
- 564 Mishra, G. S. & Gonzalez, J. P. 1975. Account of a study on the endo parasites if the rat,  
565 *Rattus norvegicus* Berkenhout, 1769 , in Tunis *Arch l'Inst Pasteur Tunis*, **52**:71-  
566 88.
- 567 Nateghpour, M., Motevalli-Haghi, A., Akbarzadeh, K., Akhavan, A. A., Mohebali, M.,  
568 Mobedi, I. & Farivar, L. 2015. Endoparasites of wild rodents in southeastern Iran.  
569 *J Arthropod-Borne Dis*, **9**:1-6.
- 570 Neal, R. A. 1948. *Entamoeba histolytica* in wild rats caught in London. *J Hyg (Lond)*,  
571 **46**:90-93.

- 572 Neal, R. A. 1947. *Entamoeba* sp. from the Syrian hamster (*Cricetus auratus*). *Nature*,  
573 **159**:502.
- 574 Noble, G.A. & Noble, E.R. 1952. Entamoebae in farm mammals. *J Parasitol*, **38**:571-  
575 595.
- 576 Parfrey, L. W., Walters, W. A., Lauber, C. L., Clemente, J. C., Berg-Lyons, D., Teiling,  
577 C., Kodira, C., Mohiuddin, M., Brunelle, J., Driscoll, M., Fierer, N., Gilbert, J. A.  
578 & Knight, R. 2014. Communities of microbial eukaryotes in the mammalian gut  
579 within the context of environmental eukaryotic diversity. *Frontiers in*  
580 *Microbiology*, **5**:298.
- 581 Ray, H. N. & Banik, D. C. 1964. *Entamoeba funambulae* n.sp., from Indian Palm  
582 Squirrel, *Funambulus palmarum*. *Bull Calcutta Sch Trop Med*, **12**:114-115.
- 583 Royer, T. L., Gilchrist, C., Kabir, M., Arju, T., Ralston, K. S., Haque, R., Clark, C. G. &  
584 Petri, W. A., Jr. 2012. *Entamoeba bangladeshi* nov sp., Bangladesh. *Emerg Infect*  
585 *Dis*, **18**:1543-1545.
- 586 Shafiyah, C. O. S., Jamaiah, I., Rohela, M., Lau, Y. L. & Aminah, F. S. 2012.  
587 Prevalence of intestinal and blood parasites among wild rats in Kuala Lumpur,  
588 Malaysia. *Trop Biomed*, **29**:544-550.
- 589 Silberman, J. D., Clark, C. G., Diamond, L. S. & Sogin, M. L. 1999. Phylogeny of the  
590 genera *Entamoeba* and *Endolimax* as deduced from small-subunit ribosomal RNA  
591 sequences. *Mol Biol Evol*, **16**:1740-1751.
- 592 Stensvold, C. R., Lebbad, M. & Clark, C. G. 2010. Genetic characterisation of  
593 uninucleated cyst-producing *Entamoeba* spp. from ruminants. *Int J Parasitol*,  
594 **40**:775-778.
- 595 Stensvold, C. R., Lebbad, M., Victory, E. L., Verweij, J. J., Tannich, E., Alfellani, M.,  
596 Legarraga, P. & Clark, C. G. 2011. Increased sampling reveals novel lineages of  
597 *Entamoeba*: consequences of genetic diversity and host specificity for taxonomy  
598 and molecular detection. *Protist*, **162**:525-541.
- 599 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5:  
600 molecular evolutionary genetics analysis using maximum likelihood, evolutionary  
601 distance, and maximum parsimony methods. *Mol Biol Evol*, **28**:2731–2739.
- 602 Verweij, J. J. & Stensvold, C. R. 2014. Molecular testing for clinical diagnosis and  
603 epidemiological investigations of intestinal parasitic infections. *Clin Microbiol*  
604 *Rev*, **27**:371-418.
- 605 Won, Y. S., Jeong, E. S., Park, H. J., Lee, C. H., Nam, K. H., Kim, H. C., Hyun, B. H.,  
606 Lee, S. K. & Choi, Y. K. 2006. Microbiological contamination of laboratory mice  
607 and rats in Korea from 1999 to 2003. *Exp Anim*, **55**:11-16.

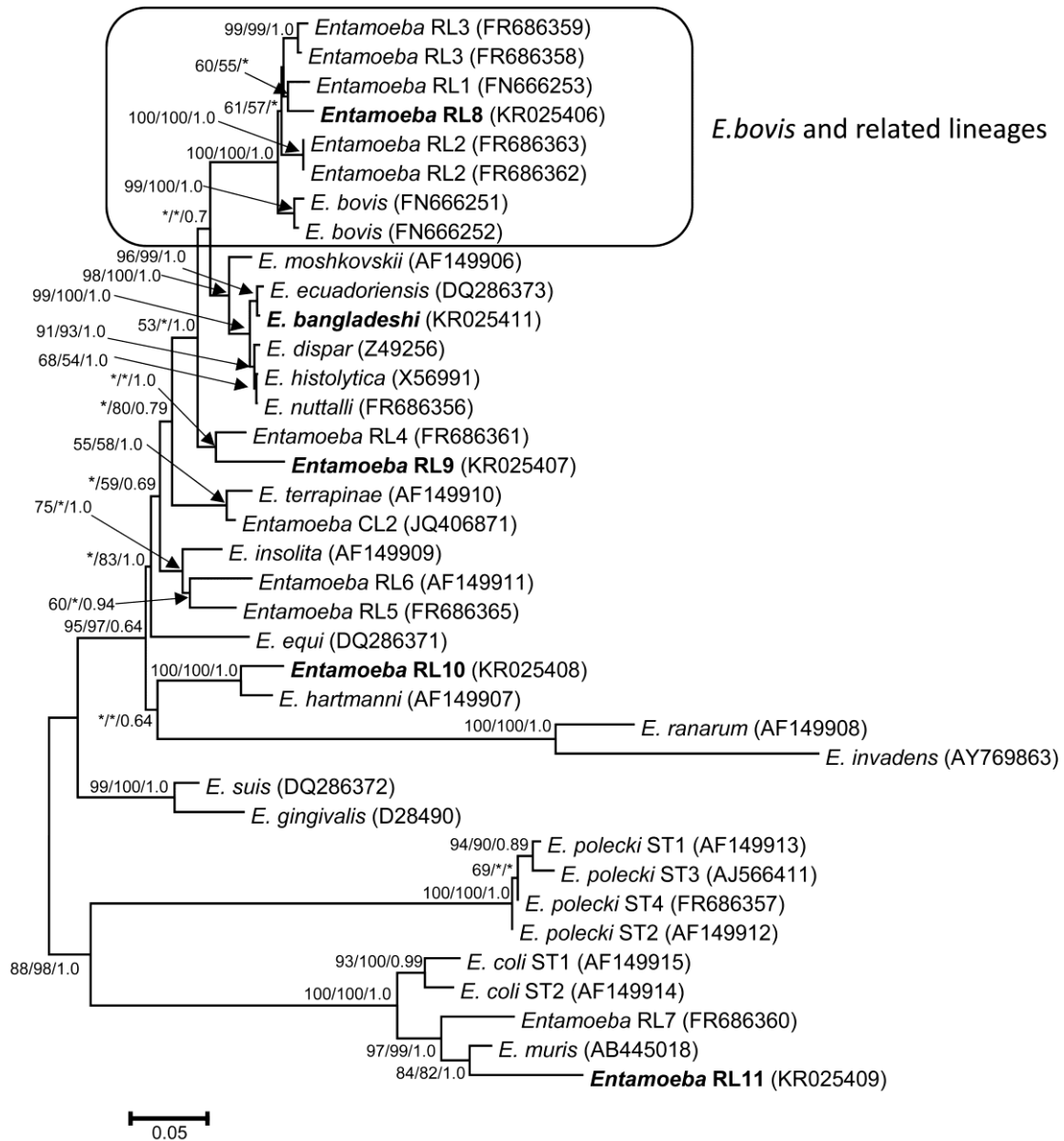
## 609 **FIGURE LEGENDS**

610

611 **Figure 1.** The phylogenetic relationships of *Entamoeba* species as inferred from SSU  
612 rDNA sequences. New sequences are indicated in bold text. The unrooted distance-based  
613 (Neighbor-Joining) tree is shown. Bootstrap proportions and Bayesian posterior  
614 probabilities are shown at each node in the order: Neighbor-Joining/Maximum  
615 Likelihood/Bayesian analysis. An asterisk indicates a value of less than 50% and if three  
616 analyses gave a value of lower than 50% no values are shown for that node. Accession  
617 numbers are listed in parentheses. The scale bar represents 0.05 substitutions per site.

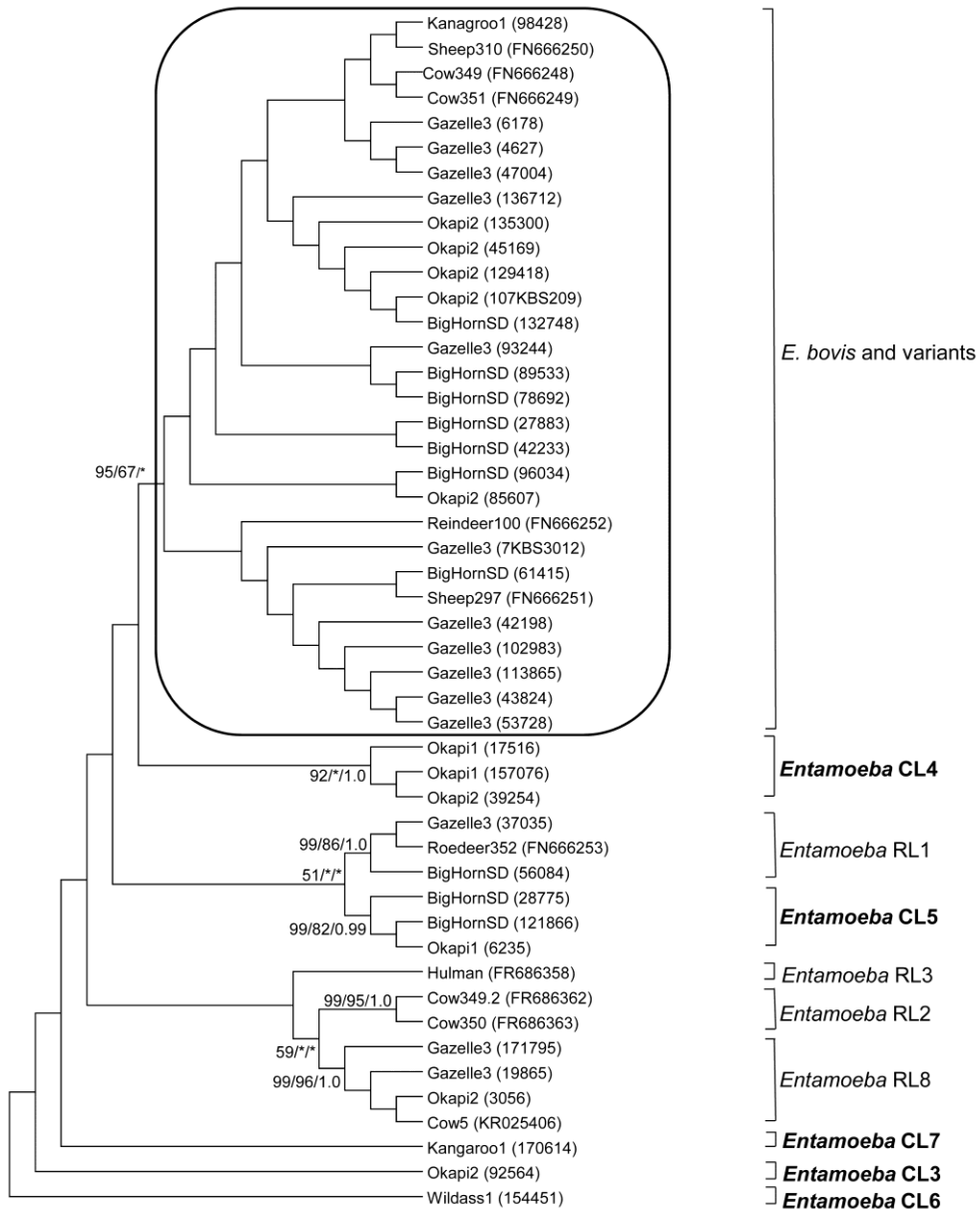
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630

**Figure 2.** A cladogram depicting the phylogenetic relationships among partial *Entamoeba* SSU rDNA sequences from the study of Parfrey et al. (2014). Also included in the cladogram are RL8 from the present study (Cow5) and relevant reference sequences available from GenBank (Cow349, Cow 349.2, Cow350, Cow351, Sheep297, Reindeer100, RoeDeer352, Hulman). The bootstrap consensus maximum likelihood tree is shown as a cladogram for clarity. The corresponding tree showing branch lengths is available as Supplementary Figure S1. Bootstrap support and Bayesian posterior probabilities are shown at each node in the order: Maximum Likelihood/Neighbor-Joining/Bayesian analysis. Only nodes corresponding to existing known species, ribosomal lineages (RL) or newly proposed conditional ribosomal lineages (CL) are labelled. Accession numbers or unique identifier codes are listed in parentheses.



631  
632





633  
634

635  
636

**Table 1.** Animal stool samples analyzed during the present study.

| Host (species name)                      | Location           | No. of samples | <i>Entamoeba</i> -positive samples | <i>Entamoeba</i> identified (number of sequences) <sup>a</sup>                              |
|--|--------------------|----------------|------------------------------------|---|
| Cattle ( <i>Bos taurus</i> )             | Devon, UK          | 15             | 4                                  | <i>E. bovis</i> (3); <i>E. moshkovskii</i> (1)  |
|  | Hertfordshire, UK  | 21             | 12                                 | <i>E. bovis</i> (10); <b><i>Entamoeba</i> RL8<sup>b</sup></b> (1); <i>Entamoeba</i> RL4 (1) |
| Sheep ( <i>Ovis aries</i> )              | Devon, UK          | 36             | 13                                 | <i>E. bovis</i> (9); <i>Entamoeba</i> RL2 (1) ; <i>Entamoeba</i> Mixed (3)                  |
|  | Hertfordshire, UK  | 19             | 14                                 | <i>E. bovis</i> (14)  |
| Pig ( <i>Sus scrofa domestica</i> )      | Devon, UK          | 9              | 0                                  |   |
|  | Vietnam            | 12             | 11                                 | <i>E. polecki</i> ST1 (7); <i>E. polecki</i> mixed ST1 & ST3 (4)                            |
| Horse ( <i>Equus ferus caballus</i> )    | Devon, UK          | 15             | 3                                  | <b><i>Entamoeba</i> RL9<sup>b</sup></b> (3)   |
| Donkey ( <i>Equus africanus asinus</i> ) | Devon, UK          | 2              | 0                                  |   |
| Roe deer ( <i>Capreolus capreolus</i> )  | Devon, UK          | 3              | 0                                  |   |
| Red deer ( <i>Cervus elaphus</i> )       | Devon, UK          | 4              | 0                                  |   |
| Fallow deer ( <i>Dama dama</i> )         | Mauritius          | 2              | 2                                  | <i>E. bovis</i> (2)   |
| Bank vole ( <i>Myodes glareolus</i> )    | Devon, UK          | 7              | 0                                  |   |
| Field vole ( <i>Microtus agrestis</i> )  | Northumberland, UK | 12             | 1                                  | <b><i>Entamoeba</i> RL11<sup>b</sup></b> (1)  |

|  |                            |     |                |   |
|--|----------------------------|-----|----------------|---|
| Wood mouse<br>( <i>Apodemus sylvaticus</i> )                 | Devon, UK                  | 5   | 0              |   |
| Stoat<br>( <i>Mustela erminea</i> )                          | Devon, UK                  | 2   | 0              |   |
| European Badger<br>( <i>Meles meles</i> )                    | Devon, UK                  | 1   | 0              |   |
| Chicken<br>( <i>Gallus gallus domesticus</i> )               | Devon, UK                  | 1   | 0              |   |
| Fox ( <i>Vulpes vulpes</i> )                                 | Devon, UK                  | 1   | 0              |   |
| Rabbit ( <i>Oryctolagus cuniculus</i> )                      | Devon, UK                  | 1   | 0              |   |
| Asian Elephant ( <i>Elephas maximus</i> )                    | Amsterdam, The Netherlands | 4   | 1              | <b><i>Entamoeba RL10<sup>b</sup></i></b> (1)                        |
| Goose ( <i>Anser domesticus</i> )                            | Devon, UK                  | 2   | 0              |   |
| Royal python ( <i>Python regius</i> )                        | London, UK                 | 1   | 0              |   |
| Aldabran Giant tortoise<br>( <i>Aldabrachelys gigantea</i> ) | Mauritius                  | 3   | 1 <sup>c</sup> | <b><i>Entamoeba CL1<sup>b</sup></i></b> (1); <i>E. insolita</i> (1) |
| Total  |                            | 178 | 62             |   |

637 Numbers in parentheses indicate the number of samples corresponding to particular species/ribosomal lineages.

638 <sup>a</sup>Bold typeface indicates either a new ribosomal lineage (RL) or conditional lineage (CL) identified during in this study.

639 <sup>b</sup>Accession numbers for newly described RLs and CL are as follows: *Entamoeba* RL8 (KR025406), *Entamoeba* RL9 (KR025407),  
640 *Entamoeba* RL10 (KR025408), *Entamoeba* RL11 (KR025409) and *Entamoeba* CL1 (KR025410).

641 <sup>c</sup>Two distinct *Entamoebas* were present in this one sample.

642

643 **Table 2. List of animals surveyed by Parfrey et al. (2014) in which *Entamoeba* sequences were found**

| Host (species name)                                      | Location             | Number of samples | <i>Entamoeba</i> -positive samples <sup>a</sup> | <i>Entamoeba</i> identified (number of sequences) <sup>b</sup>  |
|--|----------------------|-------------------|---|---|
| <b>Wild ass (<i>Equus asinus</i>)</b>                    | Saint Louis zoo, USA | 1                 | 1   | <b><i>Entamoeba</i> CL6</b> (1)   |
| <b>Zebra (<i>Equus grevyi</i>)</b>                       | Saint Louis zoo, USA | 2                 | 1   | <i>E. hartmanni</i> (1)   |
| <b>Gazelle (<i>Gazella spekei</i>)</b>                   | Saint Louis zoo, USA | 2                 | 2   | <i>E. bovis</i> (11); <i>Entamoeba</i> RL1 (1); <i>Entamoeba</i> RL8 (2)  |
| <b>African bush elephant (<i>Loxodonta africana</i>)</b> | Namibia              | 1                 | 1   | <i>E. moshkovskii</i>   |
| <b>Red kangaroo (<i>Macropus rufus</i>)</b>              | Saint Louis zoo, USA | 2                 | 1   | <i>E. bovis</i> (1); <b><i>Entamoeba</i> CL7</b> (1)  |
| <b>Okapi (<i>Okapia johnstoni</i>)</b>                   | Saint Louis zoo, USA | 3                 | 2   | <i>E. bovis</i> (5); <b><i>Entamoeba</i> CL3</b> (1); <b><i>Entamoeba</i> CL4</b> (3); <b><i>Entamoeba</i> CL5</b> (1); <i>Entamoeba</i> RL8 (1); |
| <b>Bighorn sheep (<i>Ovis canadensis</i>)</b>            | Saint Louis zoo, USA | 2                 | 1   | <i>E. bovis</i> (7); <i>Entamoeba</i> RL1 (1); <b><i>Entamoeba</i> CL5</b> (2)  |
| <b>Baboon (<i>Papio hamadryas</i>)</b>                   | Namibia              | 1                 | 1   | <i>E. hartmanni</i>   |
| <b>Sumatran oranutan (<i>Pongo abelii</i>)</b>           | Saint Louis zoo, USA | 1                 | 1   | <i>E. hartmanni</i>   |
| <b>Total</b>   |                      | <b>15</b>         | <b>11</b>                                       |   |

644 <sup>a</sup>Several samples contained more than one *Entamoeba* species or lineage

645 <sup>b</sup>Bold typeface in this column indicates a conditional lineage (CL). These are defined as distinct sequences that are too short to meet  
646 the criteria as established for ribosomal lineages (Stensvold et al. 2011)

647

648 **SUPPORTING INFORMATION**

649 Supplementary Data File S1. Partial *Entamoeba* SSU rDNA sequences from the study of  
650 Parfrey et al. (2014). *Entamoeba* sequences extracted from the curated SILVA 108  
651 database ([http://qiime.org/home\\_static/dataFiles.html](http://qiime.org/home_static/dataFiles.html)) generated in Parfrey et al. (2014)  
652 and used to obtain Fig. 2 are listed.

653  
654 Supplementary Figure S1. A phylogram depicting the phylogenetic relationships among  
655 partial *Entamoeba* SSU rDNA sequences from the study of Parfrey et al. (2014). This is  
656 the same tree as shown in fig. 2, except with branch lengths shown.

657  
658 Supplementary Data File S2. Distance matrix. The estimated distances between  
659 sequences in the Neighbor-Joining tree shown in fig. 1 were computed using the  
660 Maximum Composite Likelihood method.  
661

662 **Supplementary data file S1.**

663 Partial *Entamoeba* SSU rRNA gene sequences from the study of Parfrey et al. (2014).  
664 *Entamoeba* sequences extracted from the curated SILVA 108 database  
665 ([http://qiime.org/home\\_static/dataFiles.html](http://qiime.org/home_static/dataFiles.html)) generated in Parfrey et al. (2014) and used  
666 to obtain Fig. 2 are listed.

667 **>Okapi2 (45169)**

668 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
669 TAAAGCGCTTAGCTTGC GG GTGCCCTGCTCTGCGGGGAGGAAGCTTGC GATAAACGGCGC  
670 GGAGGCGATGCCGTTTTCCGGCCGGTGTCACTACTTTGAAAAAATAGGGTGTTC AAAGCA  
671 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCTTTCGGGATTT  
672 CGGAAAAGGATTAAGAGGAACAATTGGGGTGATTTCAGAAAATGACGGGAGAGGTA AAA  
673 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAA  
674 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
675 ACGATGTCAACCAAGGATTGGATTAGTTTTAGAGTGACAGAAGTCCGGTAACGCTGTTA  
676 CTGGGTTGACGGATCTCGCTTCCACCTTATTCAGAACTTAAAGAGAAATCTTGAGTT

677 **>Gazelle3 (42198)**

678 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
679 TAAAGCGCTTAGCGTATGGGTGCCCTGCTCGTTCGGGGAGGAAATGCGCATCAAACGGTGC  
680 GAAGGCGATGCCGTTTTCGATCGGTGTCACTACTTTGAAAAAATAGGGTGTTC AAAGCA  
681 AATCTTATGTTAATGCATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
682 CGGGATAAGGATTAAGAGGAACAATTGGGGTGATTTCAGAAAATGACGGGAGAGGTA AAA  
683 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAA  
684 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
685 ACGATGTCAACCAAGGATTGGATTAGTTTAAAGAGGGACAGAATTCGGCAACGCTGTTG  
686 TTGGGATGACGGACCTCGCTTCCACCTTATTCAGAACTT

687 **>Gazelle3 (37035)**

688 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
689 TAATGTTTCCTAGTATATGGAGGTTCTGCTTGCAGGGCTGATATGTACAATTAATGGAA  
690 CTTAGGCGATGCTGCTCACGCGGTGTCACTACTTTGAAAAAATAGGGTGTTTAAAGCAA  
691 ATCTTATGTTAATGAATAATGAAGCATGGGATAATATTGAGGAGATTCTTCGGGATTTTC  
692 GAGAGAAGGATTAAGAGGAACAATTGGGGTGATTTCAGAAAATAACGGGAGAGGTA AAA  
693 TTCCATGATCGCTATAGGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAA  
694 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
695 ACGATGTCAACCAAGGATTGGATTAGTATTTGGACGACAGAGGTTAGATAACATTGTTA  
696 TTTGACTAACGGAGT

697 **>Wildass1 (154451)**

698 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
699 AAACGTGTATTTGTTTTAGGAAAGTTCTCAGAACTGGAATAAAACGCTTAAAGTACACC  
700 GAAGGAGATGAAATGAGCAATCATTTTATCATTACTTTGAAAAAATAGAGTGTTC AAG  
701 CAAAACATTTTGTTAATGAATAATGAAGCATGGGACAATGCTGAGGAGACGGTCTTCGG  
702 ACTGTTTCGAGATAAGGATTAAGGAATACTTGGG

703 **>BigHornSD (96034)**

704 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
705 TAAAGCGCTTAGCATATGGGTGCCCTGCTCTGTGGGGAGGAAGTATGCGCTGAACGGTGC  
706 GAAGGCGACGTCCGTTTTCCGGCCGATGTCACTACTTTGAAAAAATAGGGTGTTC AAAGCA  
707 AATCTTATGTTAATGCATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT

708 CGGGAAAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAACGACGGGAGAGGTAAAA  
709 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCGTTCATTAA  
710 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
711 ACGATGTCAACCAAGGATTGGATTAGTTTTAGGGTGACAGAAGTCCGGTAACGCTGTTA  
712 CTGGGTTGACGGATCTCGCTTCC

713 **>Okapi2 (135300)**

714 TTCCAGCTCCAATAGTGTATATTAAGTTCGCTGTGATTAACGCTCGTAGTTGAATTAT  
715 GAAGCGCTTAGCTTGC GGTTGCCCTGCTCTGCGGGGAGGAAGCTTGC GATAAACGGCGCG  
716 GAGGCGATGCCGTTTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTTC AAAGCAA  
717 ATCTTATGTTAATGAATAATGAAGCATGGGGCAATATCGAGGAGATCTTTCGGGATTTCC  
718 GGGAAAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGTAAAAT  
719 TCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCGTTCATTAAAT  
720 CAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAAA  
721 CGATGTCAACCAAGGATTGGATTAGTTTTAGAGTGACAGAAGTCCGGTAACGCTGTTAC  
722 TGGGTTGACGGATCTCGCTTCCACCTTATTCAGACTTAAAGAGAATC

723 **>Gazelle3 (102983)**

724 TTCCAGCTCCAATAGTGTATATTAAGTTCGCTGTGATTAACGCTCGTAGTTGAATT  
725 ATAAAGCGCTTAGCGTATGGGTGCCCTGCTCGGTCGGGGAGGAAATGCGCATCAAACGGT  
726 GCGAAGGCGATGCCGTTTTCGATCGGTGTCACTTTAGAAAAAATAGGGTGTTC AAA  
727 GCAAATCTTATGTTAATGCATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGA  
728 TTTCCGGGATAAGGATTAAGAGGAGCAATTGGGGTGATTCAGAAAATGACGGGAGAGGTA  
729 AAATTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCGCTCAATTGCGTTCAT  
730 TAATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTA  
731 TAAACGATGTCAACCAAGGATTGGATTAGTTTTAGAGGAAGTGTGACTACAAATCATTG  
732 ATTTGCTGGAACACCGACTCGCTACTACCTCATTTCAGAACTTAATGAGAAA

733 **>BigHornSD (89533)**

734 TTCCAGCTCCAATAGTGTATATTAAGTTCGCTGTGATTAACGCTCGTAGTTGAATTA  
735 TAGAGCGCTTAGCATGCGGGTGCCCTGCTCTGCGGGGAGGAAGTATGCGGTAACGGCGC  
736 GGAGGCGATGCCGTTTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTTC AAAGCA  
737 AATCTTATGTTAATGCATAATGAAGCATGGGACAATATTGAGGGGTTCTTTCGGGAATT  
738 CGAGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTAAAA  
739 TTCCATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCGTTCATTAA  
740 TCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCCTAACTATAA  
741 ACGATGTCATCCAAGGATTGGATTAGTTTTAGAGAGACAGATGCTCGGTAACATCGTTA  
742 TTGAGCGGACGGATCTCGCCTCCACCTTATTCAGAACTTAAAG

743 **>Okapi2 (92564)**

744 TTCCAGCTCCAATAGTGTATATTAAGTTCGCTGTGATTAACGCTCGTAGTTGAATTG  
745 AAAAGTGTCTAGCTTGGGAAAGCCGGTAACCGGGGACTCTTGC GTTAACGGACATGA  
746 AGGTTATGGTCGCAAGGCCGTAATTACTTTGAAAAAATAGGGTGTTC AAAGCAAATCTT  
747 ATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGATCCCTTTTGGGATGTCGAG  
748 ATAAGGATTAACGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTGAAAATCC  
749 ATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCGTTCATTAAATCA  
750 AGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCCTAACTATAAACG  
751 ATGTCAACCAAGGATTGGATTAGTTTTA

752 **>BigHornSD (56084)**

753 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
754 TAATGTTTCCTAGTATATGGAGGTTCTGCTTGCAGGGCTGATATGTACAATTAATGGAA  
755 CTTAGGCGATGCTGCTCACGCGGTGTCATTACTTTGAAAAAATAGGGTGTTTAAAGCAA  
756 ATCTTATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGATTCTTCGGGATTTTC  
757 GAGAGAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTAAAA  
758 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAA  
759 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
760 ACGATGTCAACCAAGGATTGGATTAGTTTAAAGAGGGACAGAAGTCCGGCAACGCTGTTG  
761 TTGGGCTGACGGACCTCGCT

762 **>Gazelle3 (7KBS3012)**

763 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
764 TAAAGCGCTTAGCGTATGGGTGCCCTGCTCGTCGGGGAGGAAATGCGCATCAAACGGTGC  
765 GAAGGCGATGCCGGTTTCGATCGGTGTCATTACTTTGAAAAAATAGGGTGTTCAAAGCA  
766 AATCTTATGTTAATGAATAATGAAGCATGGGATAATATTGAGGAGATTCTTTGGGATTT  
767 TCGAGAGAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTAAA  
768 ATTCCATGATCGCTATAGGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTA  
769 ATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATA  
770 AACGATGTCAACCAAGGATTGGAT

771 **>BigHornSD (121866)**

772 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
773 TAATGGACCCTGTATCTGGAAGCTCTGCTTGCGGGGTTGAACGATACGTTATACGGTCT  
774 TAGGCGATGCTCGTTAACGCGAGTGTCATTACTTTGAAAAAATAGGGTGTTTAAAGCAA  
775 ATCTTATGTTAATGAATAATGAAGCATGGGACAATATTGAGGGGTCTTTTCGGGAATTC  
776 GAGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTAAAAT  
777 TCCATGATCGCTATAAGATGCACGAGAACGAAAGCATTTCACTCAATTGCGTTCATTAAT  
778 CAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCCTAACTATAAA  
779 CGATGTCAACCAAGGATTGGATTAGTTTAAAGAGAGACAGATGCTCGGTAACATCGTTAT  
780 TGAGCGGACGGATCTCGCCTCCACCTTATTCAGAACTTAAAGAGAAATCTT

781 **>Okapi1 (17516)**

782 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
783 TAAAGCGTTTGTAGTGCCTGGAGCACTGCTTGCAGTGTGAAGGGCACAATCAAACAACGCCT  
784 TAGGTGATGCTCTTCGGAGTGTCATTACTTTGAAAAAATAGGGTGTTTAAAGCAAATCT  
785 TATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGATCCTTCGGGATTTTCGAGA  
786 AAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTAAAATTCCA  
787 TGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAATCAA  
788 GAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCCTAACTATAAACGA  
789 TGTCAACCAAGGATTGGATTAGTTTAAAGAGTGACAGATGCCGAATAGTATGGCTATTCT  
790 GGCGGACGGAACTCG

791 **>Gazelle3 (19865)**

792 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
793 TAAGACACTAGAATGGGCAGAAGCTTTTCGAAGTGGAGCTCTATTTCGCTAAAGTGTTTA  
794 GGTGATACTTACTTCGGTTAGTGTCATTACTTTGAAAAAATAGGGTGTTTAAAGCAAAT  
795 CTTATGTTAATGAATAATGAAGCATGGGACAATATTGAAGGAGATCCTTTATGGATTTTC  
796 GAGGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTAAA  
797 ATTCCATGATCGCTATACGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTA  
798 ATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATA



799 AACGATGTCAACCAAGGATTGGATTAGTTT TAGGAGGGACAGAATTCCGGCAACGCTGTT  
800 GTTGGGATGACGGACCTCGCTTCCACCTTATTCAGAACTTAAAGAGAAATCTCGAGTT  
801 **>Okapi2 (85607)**  
802 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
803 TAAAGCGCTTAGCATATCGGTGCCCTACTCTGTGGGGAGGAAGTATGCGCTGAACGGCGC  
804 GAAGGCGATGTCGGTTTCGGCCGATGTCATTACTTTGAAAAAATAGGGTGTCAAAGCA  
805 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
806 CGGGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGTAATA  
807 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTA  
808 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATA  
809 ACGATGTCAACCAAGGATTGGATTAGTCATAGAGGGACAGAAGTCCGGCAACGCTGTTG  
810 TCGGGCTGACGGACCTCGCTTCCACCTTATTCAGAACTTAAAGAGAAAT  
811 **>Okapi1 (157076)**  
812 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
813 TAAAGCATTTAGTTCTTGGAGGCTCTGTTCGCAGGGTTGAAGAGGACAACAAACGAATG  
814 CCTTAGGTGATGCTCTTCGGAGTGCCACTTTGAAAAAATAGGGTGTAAAGCAAA  
815 TCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCTTTCGGGATTTCCG  
816 AGAAAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTAATA  
817 TCCATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTA  
818 CAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCCTAACTATA  
819 CGATGTCAACCAAGGATTGGATTAGTTATAGAGTGACAGATGCCGGGTAGCATGGCTAT  
820 CAGGCGGACGGAGCTCGC  
821 **>Gazelle3 (43824)**  
822 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
823 TAAAGCGCTTAGCGTATGGGTGCCCTGCTTGTTCGGGGAGGAAATGCGCATCAAACGGTGC  
824 GAAGGCGATGCCGGTTTCGATCGGTGTCATTACTTTGAAAAAATAGGGTGTCAAAGCA  
825 AATCTTATGTTAATGCATAATGAAGCATGGGACAATATCGAGGAGACCCTTCGGGTTTC  
826 GGGTAGTAA  
827 **>Gazelle3 (113865)**  
828 TTCCAGCTCCAATAGTGTATATTAAGTTAGCTGTGATTAAAACGACTCGTAGTTGAAT  
829 TATAAAGCGCTTAGCGTATGGGTGCCCTGCTTGTTCGGGGAGGGAAATGCGCATCAAACGG  
830 TGCGAAGGCGATGCCGGTTTCGATCGGTGTCATTACTTTGAAAAAATAAGGGTGTCAA  
831 AGCAAATCTTATGTTAATGCATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGG  
832 ATTTTCGGGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGT  
833 AAAATTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCG  
834 TTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACT  
835 ATAAACGATGTCAACCAAGGATTGGATTAGTTTTAA  
836 **>Kangaroo1 (170614)**  
837 TTCCAGCTCCAATAGTGTATATTAAGTTGTTGTGATTAAAACGCTCGTAGTTGAATTA  
838 TTAGACAGTCTTATTTAACGGATTTTGAGGTTCCGCCGAAAATGAAGTTAATACCAAATG  
839 GACTGTTGAAGGCGATATGTCGTTTACGACATGTCATTACTTTGAATAAATTAAGGTGT  
840 TTAAAGCAAACCTTTTGTTAATGAATAATTAAGCATAGGACAATATTAAGGAGATACTT  
841 AGGTATTTTCGAGAGAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAG  
842 AGGTAATAATTCTAAGATCGCTATAAGATGAACGAGAGCGAAAGCATTTCACTCAATTGT  
843 GTCCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCC  
844 TAACTATAAACGATGTC

845 **>Okapi2 (3056)**  
846 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
847 TAAGACACTAGAATGGGCAGAAGCTTTTCGAAGTGGAGCTCTATTCGCTAAAGTGTTA  
848 GGTGATACTTACTTCGGTTAGTGTCATTACTTTGAAAAAATAGGGTGTTTAAAGCAAAT  
849 CTTATGTTAATGAATAATGAAGCATGGGACAATATTGAAGGAGATCCTTTATGGATTTTC  
850 GAGGATAAGGATTAAGAGGAACAATTGGGGTGATTTCAGAAAATAACGGGAGAGGTAAA  
851 ATTCCATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCGTTCATTA  
852 ATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATA  
853 AACGATGTCAACCAAGGATTGGATTAGTTTTTGGAGTACTAAGAGTTAATAGCATTGCT  
854 ATTGACTTGAAGGAACTCGCTTTCACCTTATTCAGAACTTAAAGAGAAATCTT

855  
856 **>BigHornSD (27883)**  
857 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATGA  
858 TAAAGCGCTTAGCATACGGGTGCTCTGCTCTGCGGGGAGGAAGTATGCAGTAGACGGCGC  
859 GGAGGCGATGCCGTTCCGGTCCGGTGTGTCATTACTTTGAAAAAATAGTGTGTTCAAAGCA  
860 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
861 CGGAAAAGGATTAAGAGGAACAATTGGGGTGATTTCAGAAAATGACGGGAGAGGTAAAA  
862 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCATTCATTAA  
863 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
864 ACGATGTCAACCAAGGATTGGATTAGTTTTAGAGTGACAGAGGTCCGTAACACAGTTA  
865 CTGGGCTGACGGATCTCGCTTCCACCTTATTCAGAACTTAA

866 **>BigHornSD (42233)**  
867 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATGA  
868 TAAAGCGCTTAGCATACGGGTGCTCTGCTCTGCGGGGAGGAAGTATGCAGTAGACGGCGC  
869 GGAGGCGATGCCGATTCCGGTCCGGTGTGTCATTACTTTGAAAAAATAGTGTGTTCAAAGCA  
870 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
871 CGGGATAAGGATTAAGAGGAACAATTGGGGTGATTTCAGAAAATAACGGGAGAGGTAAAA  
872 TTCCATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCGTTCATTAA  
873 TCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCCTAACTATAA  
874 ACGATGTCAACCAAGGATTGGATTAGTTTTAGAGAGACAGATGCTCGGTAACATCGTTA  
875 TTGAGCGGACGGATCTCGCCTCCACCTTATTCAGAACTTAAAGAGAAATCTTGAGTTT

876 **>Gazelle3 (53728)**  
877 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
878 TAAAGCGCTTAGCGTATGGGTGCCCTGCTCGTTCGGGGAGGAAATGCCATCAAACGGTGC  
879 GAAGGCGATGCCGTTTCGATCGGTGTGTCATTACTTTGAAAAAATAGGGTGTTCAAAGCA  
880 AATCTTATGTTAATGCATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
881 CGGGATAA

882 **>BigHornSD (78692)**  
883 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
884 TAAAGCGCTTAGCATGCGGGTGCCCTGCTCTGCGGGGAGGAAGTATGCGGTAACGGCGC  
885 GGAGGCGATGCCGTTTCGGCCGGTGTGTCATTACTTTGAAAAAATAGGGTGTTCAAAGCA  
886 AATCTTATGTTAATGCATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
887 CGGAAAAGGATTAAGAGGAACAATTGGGGTGATTTCAGAAAATGACGGGAGAGGTAAAA  
888 TTCCATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCCTCAATTGCGTTCATTAA  
889 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA

890 ACGATGTCAACCAAGGATTGGATTAGTATTTGGACGACAGAGGTTAGATAACATTGTTA  
891 TTTGACTAACGGAGTCCGCTT  
892 **>Okapi2 (107KBS209)**  
893 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
894 TAAAGCGCTTAGCTTGC GGGTGCCTGCTCTGCGGGAGGAAGCTTGC GATAAACGGCGCG  
895 GAGGCGATGCCGTTTTCGGCCGGTGTCACTTTGAAAAAATAAGGGTGTTC AAAGCA  
896 AATCTTATGTTAATGAATAATGAAGCATGGGACAATAATCGAGGAGATCTTTCGGGATT  
897 TTCGGGAAAAGGAT  
898 **>Gazelle3 (6178)**  
899 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATGA  
900 TAAAGCGCTTAGCATGCGGGTGC GCTGCTTGCAGTGAGGAAGTATGC GATTAATGGCGC  
901 GAAGGCGATGCTGGTTTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTTC AAAGCA  
902 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGATCCTTCGGGATTT  
903 CGGAAAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGTA AAA  
904 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGC GTTCATTA  
905 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATA  
906 ACGATGTCAACCAAGGATTGGATTAGTTTTAGAGTGACAGAAGTCCGGCAACGCTGTTG  
907 TTGGGCTGACGGA  
908 **>Okapi1 (6235)**  
909 TTCCAGCTCCAATAGTGTATATTTAAAGTTGCTGTGATTAAAACGCTCGCAGTTGAATTAT  
910 AATGGACCCTGTATCCGGAGGCCCTGCTTGC GGGTTAAAGATACGTTATACGGTTCTTA  
911 GCGATGCTCGTTAACGCGAGTGTCATTACTTTGAAAAAATAAGGGTGTTTAAAGCAA  
912 TCTTATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGTTCCTTCGGGAATTCG  
913 AGATAAGGATTAAGAGGAACAATTGGGTGATTCAGAAAATAACGGGAGAGGTA AAACTC  
914 CATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAATCA  
915 AGAACGAAAGTTAGGGGATCGAAGACGATCGGATACCGTCGTAGTCCTAACTATAAACG  
916 ATGTCAACCAAGGATTGGATTAGTTTTAGAGAGACAGACGCTCGGTAGCATTGTTATTG  
917 AGCTGACGGATCTCGCCTCCACCTTATTCAGA ACTTAAAGAGAAATCTTGAAGTTTATGG  
918 A  
919 **>Gazelle3 (4627)**  
920 TTCCAGCTCCAATAGTGTATATTTAAAGTTAGCTGTGATTAAAACGACTCGTAGTTGAAT  
921 GATAAAGACGCTTAGCATGCGAGTGC GCTGCTTCGTCAGTGAGGAAGTATGC GATTAAT  
922 GCGCGAAGGCGATGCTGGTTTTCGGTCCGGTGTCACTTTGAAAAAATA  
923  
924 **>Gazelle3 (171795)**  
925 TTCCAGCTCCAATAGTGTATATTTAAAGTTAGCTGTGATTAAAACGACTCGTAGTTGAAT  
926 TATAAGACACTAGAATGGGCAGAAGCTTTTTCGTAAGTGGAGCTCTATTCGCTAAAGTGT  
927 TTAGGTGATACTTACTTCGGTTAGTGTCACTTTGAAAAAATAAGGGTGTTTAAAGC  
928 AAATCTTATGTTAATGAATAATGATGCATGGGACAATATTGAAGGAGATCCTTTATGGA  
929 TTTTCGAGGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATAACGGGAGAGGTA  
930 AAATTCATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCACTCGATTGCGTTCAT  
931 TAATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTA  
932 TAAACGATGTCAACCAAGGATTGGATTAGTATTTGGACGACAGAGGTTAGATAACATTG  
933 TTATT  
934 **>Okapi2 (39254)**

935 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
936 TAAAGCATTTAGTGCTTGGAGGCTCTGTTGCGAGGGTTGAAGAGTACAACAAACGAATG  
937 CCTTAGGTGATGCTCTTCGGAGTGTCACTTTGAAAAAATAGAGTGTAAAGCAAA  
938 TCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCTTTCGGGATTTCG  
939 AGAAAAGGATTAAGAGGAACAATTGGGGTGATTACAGAAAATAACGGGAGAGGTA AAAAT  
940 TCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAAT  
941 CAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAAAA  
942 CGATGTCAACCAAGGATTGGATTAGTTTTAGAGTGACAGAAGTCCGGTAACGCTGTTAC  
943 TGGGTTGACGGATCTCGCTTCCACCTTATTCAGAACTTAAAGAGAAATCTTGAG

944 **>BigHornSD (28775)**

945 TTCCAGCTCCAATAGTGCATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTAT  
946 AATGGACCCTGTATCTGGAAGCTCTGCTTGGGGGTTGAACGATACGTTATACGGTCTT  
947 AGGCGATGCTCGTTAACGCGAGTGTCACTTTGAAAAAATAGGGTGTAAAGCAAA  
948 TCTTATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGATTCTTCGGGATTTCG  
949 AGAGAAGGATTAAGAGGAACAATTGGGGTGATTACAGAAAATAACGGGAGAGGTA AAAAT  
950 CCATGATCGCTATAGGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAATC  
951 AAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAAAC  
952 GATGTCAACCAAGGATTGGATTAGTATTTGGACGACA

953 **>Kangaroo1 (98428)**

954 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
955 TAAAGCGCTTAGTATGCGGGAGCTCCGCTTTGCGGGGAAGAAGTATGCGATAAACGGCG  
956 CGGAGGCGATGCCGTTTCGGTTCGGTGTCACTTTGAAAAAATAGGGTGTCAAAGC  
957 AAATCTTATGTTAATGAACAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATT  
958 TCGGGAAAAGGATTAAGAGGAACAATTGGGGTGATTACAGAAAATGACGGGAGAGGTA AAA  
959 ATTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTA  
960 ATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATA  
961 AACGATGTCAACCAAGGATTGGATAAATTTTAAAAACTGAGAGTGATTAAAGCATTGCT  
962 TTAGGATCTTAAAGTTTTATTACTACC

963 **>Okapi2 (129418)**

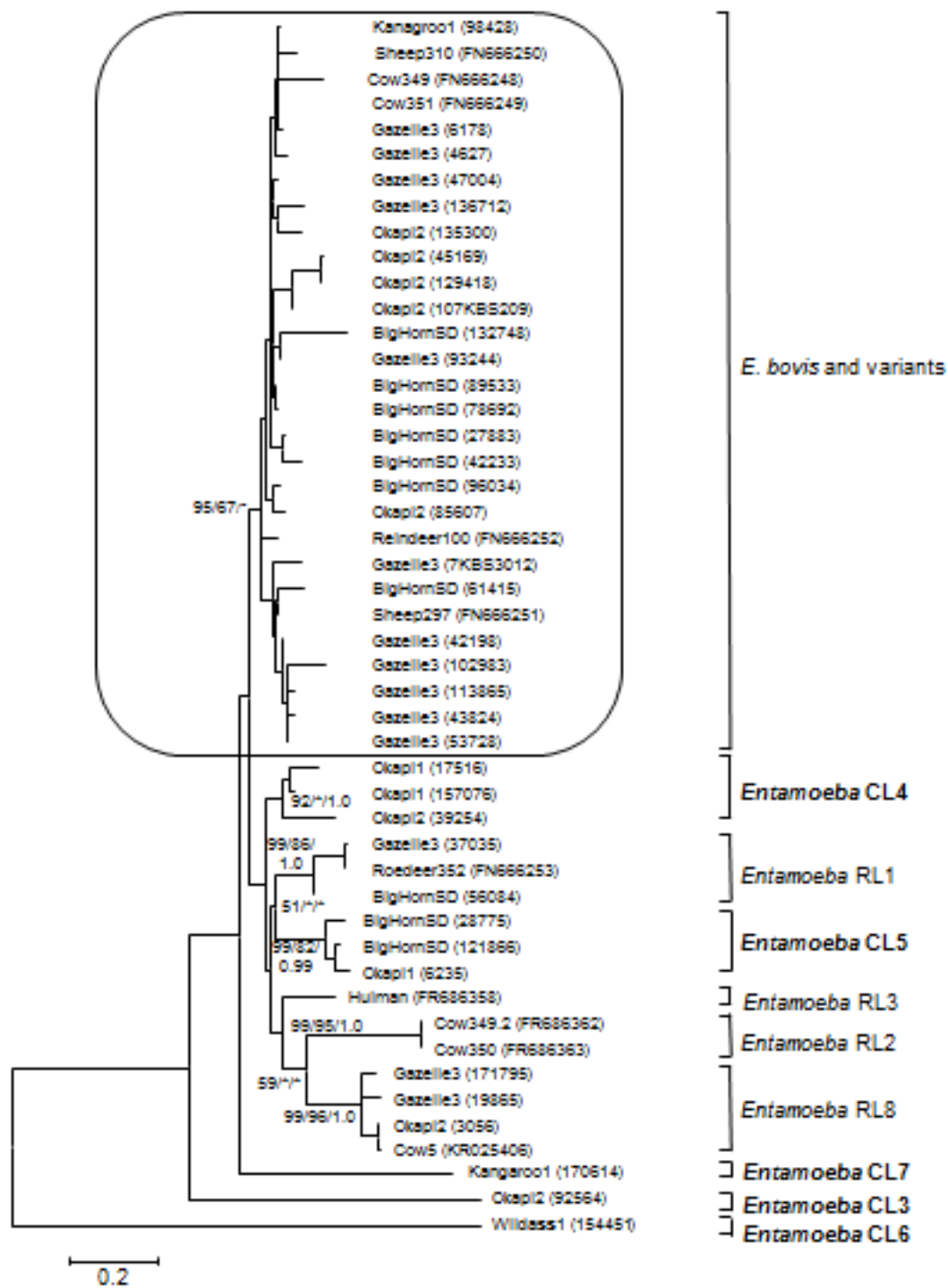
964 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
965 TAAAGCGCTTAGTATGCGGGTGCCTGCTCTGCGGGGAGGAAGCTTGCATAAACGGCGC  
966 GGAGGCGATGCCGTTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTCAAAGCA  
967 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCTTTCGGGATTT  
968 CGGAAAAGGATTAAGAGGAACAATTGGGGTGATTACAGAAAATGACGGGAGAGGTA AAA  
969 TTCCATGATCGCCATAAGATGCACGAGGGCGAAAGCATTTCACTCAATTGCGTTCATTA  
970 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATA  
971 ACGATGTCAACCAAGGATTGGATTAGTTTTAGAGGAAGTGTGACTACAAATCATTGATT  
972 TGCTGGAACACCGACTCGCTACTACCTCATTCAGAACTTAAAGAGAAATCA

973 **>Gazelle3 (47004)**

974 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATGA  
975 TAAAGCGCTTAGCATGCGGGTGCCTGCTTGCAGTGAGGAAGTATGCGATTAATGGCGC  
976 GAAGGCGATGCTGGTTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTCAAAGCA  
977 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGATCCTTCGGGATTT  
978 CGGAAAAGGATTAAGAGGAACAATTGGGGTGATTACAGAAAATGACGGGAGAGGTA AAA  
979 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTA  
980 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCCAACTATAA

981 ACGATGTCAACCAAGGATTGGATTAGTTTTTTGAGTGACTAAGAGTTAATAGCATTGTTA  
982 TTTGACTAACGGAG  
983 **>Gazelle3 (93244)**  
984 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGACTCGTAGTTGAATT  
985 ATAAAGCGCTTAGCATGCGGGTGCCCTACTCTGCGGGGAGGAAGTATGCGGTAAACGGCG  
986 CGGAGGCGATGCCGTTTCGGCCGGTGTCACTTTGAAAAAATAAGGGTGTTCAAAG  
987 CAAATCTTATGTTAATGAATAATGAAGCATGGGACAATATTGAGGAGATCCTTCGGGAT  
988 TTCGGGAAAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGTAA  
989 AATCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATT  
990 AATCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTAT  
991 AAACGATGTCAACCAAGGATTGGATTAGTTTTTTAGAGTGACAGAAGTCCGGCAACGCTG  
992 TTGTTGGGCTGACGGATCTCGCTTCCACCTTATTCAGAACTTAAAGAGAAATCTCGAGTT  
993 TATG  
994 **>BigHornSD (61415)**  
995 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
996 TAAAGCGCTTAGCGTATGGGTGCCCTGCTCGTCGGGGAGGAAATACGCATCAAACGGTGC  
997 GAAGGCGATGCCGTTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTTCAAAGCA  
998 AATCTTATGTTAATGCATAGTGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
999 CGGGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGTAAAA  
1000 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAA  
1001 TCAAGAACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
1002 ACGATGTCAACCAAGGATTGGATTAGTATTTGGACGACAGAGGTTAGATAACAATTGTT  
1003 ATTTGACTAA  
1004 **>Gazelle3 (136712)**  
1005 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
1006 TAAAGCGCTTAGCTTGCGGGTGCCCTGCTCTGCGGGGAGGAAGCTTGCGATAAACGGCGC  
1007 GGAGGCGATGCCGTTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTTCAAAGCA  
1008 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCCTTCGGGATTT  
1009 CGGGATAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGTAAAA  
1010 TTCCATGATCGCCATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAA  
1011 TCAAGGACGAAAGTTAGGGGATCGAAGACGATTAGATACCGTCGTAGTCCTAACTATAA  
1012 ACGATGTCAACCAAGGATTGGATTAGTTTAAAGAGGGACAGAATTCCGGCAACGCTGTTG  
1013 TTGGGATGACGGACCTCGCTTCCACCTTATTCAGAACTTAAAGAGAAAT  
1014 **>BigHornSD (132748)**  
1015 TTCCAGCTCCAATAGTGTATATTAAGTTGCTGTGATTAAAACGCTCGTAGTTGAATTA  
1016 TAAAGCGCTTAGCTTGCGGGTGCCCTGCTCTGCGGGGAGGAAGCTTGCGATAAACGGCGC  
1017 GGAGGCGATGCCGGCTTCGGCCGGTGTCACTTTGAAAAAATAGGGTGTTCAAAGCA  
1018 AATCTTATGTTAATGAATAATGAAGCATGGGACAATATCGAGGAGATCTTTCGGGATTT  
1019 CGGGAAAAGGATTAAGAGGAACAATTGGGGTGATTCAGAAAATGACGGGAGAGGTAAAA  
1020 TTCCATGATCGCTATAAGATGCACGAGAGCGAAAGCATTTCACTCAATTGCGTTCATTAA  
1021 TCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCCTAACTATAA  
1022 ACGATGTCAACCAAGGATTGGATTAGTTTTAGAGAGACAGATGCTCGGTAACATCGTTA  
1023 TTGAGCGGACGGATCTCGCCTCCACCTTATTCAGAACTTAAAGAG  
1024  
1025 Parfrey, L. W., Walters, W. A., Lauber, C. L., Clemente, J. C., Berg-Lyons, D., Teiling,  
1026 C., Kodira, C., Mohiuddin, M., Brunelle, J., Driscoll, M., Fierer, N., Gilbert, J. A. &

1027 Knight, R. 2014. Communities of microbial eukaryotes in the mammalian gut within the  
1028 context of environmental eukaryotic diversity. *Frontiers in Microbiology*, **5**.  
1029  
1030



1031  
1032

|                            |   |
|----------------------------|---|
| Entamoeba_RL10_(KR025408)  |   |
| E._hartmanni_(AF149907)    | 0.048   |
| E._ranarum_(AF149908)      | 0.382 0.376   |
| E._invadens_(AY769863)     | 0.520 0.510 0.221   |
| Entamoeba_RL7_(FR686360)   | 0.462 0.453 0.691 0.754   |
| E._muis_(AB445018)         | 0.447 0.443 0.690 0.768 0.090   |
| Entamoeba_RL11_(KR025409)  | 0.537 0.525 0.773 0.880 0.129 0.087   |
| E._coli_ST1_(AF149915)     | 0.440 0.429 0.755 0.806 0.114 0.109 0.158   |
| E._coli_ST2_(AF149914)     | 0.424 0.427 0.712 0.752 0.109 0.109 0.150 0.041   |
| E._polecki_ST2_(AF149912)  | 0.447 0.452 0.697 0.795 0.538 0.506 0.574 0.534 0.526   |
| E._polecki_ST4_(FR686357)  | 0.454 0.461 0.701 0.777 0.548 0.512 0.588 0.537 0.535 0.005   |
| E._polecki_ST1_(AF149913)  | 0.475 0.497 0.705 0.795 0.527 0.498 0.559 0.514 0.512 0.015 0.017   |
| E._polecki_ST3_(AJ566411)  | 0.463 0.491 0.740 0.821 0.581 0.546 0.589 0.563 0.565 0.020 0.020 0.018   |
| E._suis_(DQ286372)         | 0.195 0.203 0.437 0.539 0.434 0.415 0.481 0.394 0.377 0.372 0.379 0.375 0.393   |
| E._gingivalis_(D28490)     | 0.215 0.224 0.482 0.553 0.477 0.460 0.536 0.427 0.393 0.389 0.394 0.396 0.411 0.061   |
| E._equi_(DQ286371)         | 0.161 0.161 0.355 0.492 0.437 0.424 0.490 0.405 0.396 0.428 0.428 0.444 0.463 0.201 0.225   |
| E._terrapinae_(AF149910)   | 0.157 0.146 0.377 0.480 0.443 0.423 0.475 0.419 0.412 0.404 0.402 0.422 0.439 0.178 0.203 0.149   |
| Entamoeba_RL2_(FR686363)   | 0.194 0.182 0.402 0.559 0.441 0.430 0.474 0.412 0.416 0.467 0.475 0.493 0.490 0.205 0.237 0.164 0.159   |
| Entamoeba_RL2_(FR686362)   | 0.194 0.182 0.402 0.559 0.441 0.430 0.474 0.412 0.416 0.467 0.476 0.493 0.490 0.205 0.237 0.164 0.159 0.000   |
| Entamoeba_RL3_(FR686359)   | 0.207 0.183 0.405 0.559 0.445 0.414 0.455 0.395 0.403 0.417 0.419 0.449 0.449 0.220 0.253 0.176 0.156 0.029 0.029   |
| Entamoeba_RL3_(FR686358)   | 0.209 0.180 0.401 0.553 0.422 0.402 0.445 0.379 0.393 0.421 0.423 0.446 0.450 0.219 0.259 0.170 0.152 0.029 0.029 0.008   |
| Entamoeba_RL1_(FN666253)   | 0.200 0.191 0.421 0.547 0.471 0.416 0.458 0.394 0.387 0.447 0.449 0.478 0.474 0.225 0.257 0.161 0.160 0.035 0.035 0.032 0.031   |
| E._bovis_(FN666251)        | 0.204 0.182 0.428 0.557 0.445 0.424 0.461 0.389 0.412 0.448 0.456 0.494 0.493 0.210 0.242 0.161 0.153 0.033 0.033 0.035 0.031 0.032   |
| E._bovis_(FN666252)        | 0.199 0.179 0.414 0.536 0.437 0.421 0.458 0.386 0.400 0.442 0.450 0.488 0.487 0.207 0.247 0.157 0.147 0.033 0.033 0.032 0.030 0.031 0.008   |
| Entamoeba_RL8_(KR025406)   | 0.204 0.182 0.480 0.585 0.439 0.435 0.479 0.392 0.397 0.477 0.483 0.487 0.497 0.238 0.266 0.163 0.163 0.031 0.031 0.032 0.029 0.030 0.032 0.033   |
| Entamoeba_RL4_(FR686361)   | 0.164 0.160 0.390 0.552 0.420 0.392 0.437 0.372 0.362 0.450 0.459 0.465 0.479 0.177 0.224 0.137 0.107 0.089 0.089 0.094 0.090 0.091 0.085 0.083 0.096   |
| E._moshkovskii_(AF149906)  | 0.155 0.152 0.376 0.538 0.451 0.408 0.442 0.396 0.398 0.439 0.445 0.465 0.468 0.182 0.217 0.126 0.106 0.088 0.088 0.090 0.085 0.097 0.087 0.083 0.095 0.061   |
| E._ecuadonensis_(DQ286373) | 0.168 0.159 0.361 0.491 0.438 0.418 0.499 0.398 0.395 0.428 0.434 0.454 0.463 0.186 0.222 0.123 0.118 0.093 0.093 0.096 0.090 0.098 0.096 0.092 0.101 0.075 0.037   |
| E._bangladeshi_(KR025411)  | 0.164 0.155 0.364 0.488 0.434 0.414 0.493 0.391 0.380 0.435 0.440 0.460 0.466 0.188 0.221 0.121 0.119 0.090 0.090 0.091 0.084 0.095 0.093 0.089 0.097 0.073 0.032 0.005   |
| E._dispar_(Z49256)         | 0.161 0.157 0.374 0.493 0.429 0.425 0.496 0.398 0.390 0.441 0.440 0.472 0.466 0.186 0.228 0.133 0.110 0.093 0.093 0.094 0.087 0.099 0.094 0.090 0.101 0.074 0.033 0.014 0.011   |
| E._histolytica_(X56991)    | 0.157 0.150 0.377 0.497 0.421 0.416 0.488 0.398 0.390 0.429 0.434 0.462 0.463 0.179 0.218 0.130 0.104 0.089 0.089 0.090 0.084 0.095 0.090 0.086 0.096 0.068 0.032 0.013 0.012 0.005                                     |
| E._nuttalli_(FR686356)     | 0.159 0.152 0.379 0.500 0.423 0.418 0.491 0.400 0.392 0.431 0.437 0.465 0.465 0.181 0.220 0.131 0.105 0.090 0.090 0.092 0.085 0.096 0.091 0.087 0.097 0.069 0.033 0.013 0.011 0.004 0.001                               |
| Entamoeba_RL6_(AF149911)   | 0.134 0.141 0.361 0.490 0.448 0.446 0.485 0.383 0.389 0.430 0.438 0.455 0.456 0.166 0.200 0.133 0.117 0.155 0.155 0.163 0.158 0.160 0.157 0.156 0.150 0.127 0.117 0.128 0.128 0.121 0.119 0.120                         |
| Entamoeba_RL5_(FR686365)   | 0.141 0.135 0.373 0.471 0.468 0.448 0.530 0.429 0.422 0.424 0.435 0.458 0.458 0.187 0.206 0.120 0.117 0.146 0.146 0.149 0.151 0.148 0.148 0.143 0.152 0.110 0.108 0.113 0.107 0.109 0.106 0.107 0.070                   |
| E._insolita_(AF149909)     | 0.133 0.122 0.350 0.460 0.449 0.443 0.507 0.414 0.411 0.412 0.423 0.441 0.444 0.174 0.213 0.113 0.103 0.126 0.126 0.131 0.129 0.127 0.123 0.120 0.127 0.105 0.100 0.104 0.100 0.107 0.100 0.101 0.076 0.054             |
| Entamoeba_CL2_(JQ406871)   | 0.147 0.178 0.375 0.500 0.498 0.469 0.569 0.408 0.420 0.465 0.442 0.501 0.448 0.294 0.230 0.084 0.022 0.150 0.150 0.166 0.157 0.170 0.168 0.168 0.159 0.096 0.079 0.091 0.098 0.098 0.091 0.098 0.116 0.071 0.084       |
| Entamoeba_RL9_(KR025407)   | 0.164 0.159 0.457 0.648 0.403 0.428 0.475 0.432 0.427 0.468 0.478 0.511 0.481 0.217 0.267 0.173 0.145 0.121 0.121 0.120 0.111 0.117 0.105 0.103 0.118 0.063 0.095 0.102 0.102 0.102 0.094 0.094 0.148 0.150 0.122 0.028 |