New insights into the phylogeny of *Entamoeba* species provided by analysis of four new small-subunit rRNA genes

C. Graham Clark,1 Farrokh Kaffashian,1 Blessing Tawari,1 Jeffrey J. Windsor,2 Anke Twigg-Flesner,3 Mina C. G. Davies-Morel,3 Joerg Blessmann,4 Frank Ebert,4 Babett Peschel,4 An Le Van,5 Colin J. Jackson,3 Lorna Macfarlane2 and Egbert Tannich4

Correspondence
C. Graham Clark
graham.clark@lshtm.ac.uk

1Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
2National Public Health Service for Wales Aberystwyth, Bronglais Hospital, Aberystwyth, Ceredigion SY23 1ER, UK
3Institute of Rural Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3AL, UK
4Department of Molecular Parasitology, Bernhard Nocht Institute for Tropical Medicine, D-20359 Hamburg, Germany
5Medical College, University of Hué, Hué, Vietnam

Sequences of small-subunit rRNA genes have been obtained for four new isolates of *Entamoeba*. Phylogenetic analyses give new insights into the evolution of these organisms. A novel *Entamoeba* from pigs in Vietnam that produces uninucleate cysts proved to be unrelated to other uninucleated cyst-producing species. Revival of the name *Entamoeba suis* for this organism is proposed. Instead of being related to *Entamoeba polecki*, it shares a recent common ancestor with the non-encysting *Entamoeba gingivalis* in a lineage that is basal to the tetranucleate cyst-producing clade. This suggests that species producing cysts with four nuclei are descended from an ancestor that produced cysts with a single nucleus. An *Entamoeba* from a horse was isolated in culture. No cysts were observed in the original stool sample but the sequence is placed unequivocally within the clade of tetranucleate cyst-producing species with no other sequences being specifically related. Revival of the name *Entamoeba equi* for this organism is proposed. The *Entamoeba ecuadoriensis* sequence was found to be the most closely related to *Entamoeba histolytica* and *Entamoeba dispar*, as predicted, despite the organism having been an environmental isolate originally assigned to *Entamoeba moshkovskii*. Finally, a partial *E. polecki* gene sequence from a pig proved to be virtually identical to that of *Entamoeba struthionis* from an ostrich, suggesting that the latter name is a synonym.

INTRODUCTION

In the first quarter of the 20th century, dozens of species names were assigned to amoebae of the genus *Entamoeba* that were often based largely on the parasite being found in a new host (Dobell, 1919). Morphological descriptions are often incomplete and figures unclear, leaving doubts as to the validity of the names and making comparisons between species described by different individuals virtually impossible in some cases. *Entamoeba* species are not very diverse at the morphological level, with cyst size, number of cyst nuclei and appearance of the chromatoid bars (crystalline arrays of ribosomes) being the main identification criteria. As a result, significant genetic diversity is present among morphologically indistinguishable organisms, and it is also possible that certain morphological variations do not reflect species-level differences (Clark & Diamond, 1997). Not much information is available on the host range of most *Entamoeba* species, but it is clear that some can infect diverse hosts. *Entamoeba dispar* has been isolated from a wide range of Old and New World primates as well as humans (Verweij et al., 2003) (and possibly rodents; Neal, 1948), while *Entamoeba polecki* is found in...
both pigs and humans (Desowitz & Barnish, 1986). Thus it is not known how many of the named species are actually distinct and how many are synonyms.

Comparison of DNA sequences offers a solution to the problem of morphological conservatism. The ability to extract DNA directly from stool samples and amplify genes using PCR allows our inability to find appropriate culture conditions for certain Entamoeba species to be circumvented. To address some of the unresolved questions in the phylogeny of Entamoeba, four new small-subunit (SSU) rRNA gene sequences have been obtained, two without the use of cultures, and the relationships among species have been reanalysed. The results give new insights into the evolution of these organisms.

METHODS

DNA samples. During surveys performed in Hué, Vietnam (Blessmann et al., 2002, 2003), 3.5% of human stool samples were positive for a uninucleated cyst-producing Entamoeba. Such amoebae are usually rare in humans. To investigate whether animal contact was the source of human infection, faecal samples from domestic animals living in close contact with the study population were analysed. All 28 samples from pigs were found to contain a uninucleated cyst-producing Entamoeba. DNA was extracted using the QIAamp DNA Stool Miniprep kit (Qiagen). A small region of the SSU rRNA gene was amplified using primers 542 and 543 (Table 1) (these are extended versions of Entam1 and Entam2 described previously; Verweij et al. 2001) and sequenced. However, while sequence analysis revealed that all the pig amoebae were the same, they were distinct in sequence from that in the human population (which was most closely related to the sequence of Entamoeba chattoni). Specific primer 764 was designed using the 542/543 product sequence, and a large portion of the pig amoeba SSU rRNA gene was amplified using 764 and primer RD3 (Table 1). This 1800 bp product was obtained from three independent pig samples and sequenced by primer walking. The three sequences were identical, as was the region overlapping with the 542/543 sequence (450 bp). The combined consensus sequence is missing only the first few bases of the gene. A second specific primer (765) was designed and used with primer 764 to screen faecal samples from humans and their domestic animals. The expected product was only ever obtained from pig samples, never from human, chicken, goose or duck.

During a project to isolate Blastocystis hominis from domestic animals, an Entamoeba was by chance grown in Robinson’s medium (Clark & Diamond, 2002) inoculated with horse faeces. Careful microscopic examination of the faecal specimen failed to identify any Entamoeba cysts and unfortunately the culture quickly died out. However, sediment from one culture was harvested and DNA was extracted by the CTAB method as described by Ali et al. (2005). The SSU rRNA gene sequence was obtained by PCR using broad-specificity primers in two pieces of about 1-5 kb each that overlap by about 1 kb. The 5' end was obtained using RD5 in combination with AEMH pool 3 and the 3' end using RD3 with AEMH pool 5 (Table 1). As two unrelated sequences were present in each product, they were cloned into pGEM-T Easy (Promega) before being sequenced by primer walking. The two partial Entamoeba genes differed at five positions in the approx. 1 kb region of overlap (<0.5%). The second sequence identified in the PCR product was from a Colpodella species, which was probably an environmental contaminant.

An Entamoeba SSU rRNA gene sequence was obtained from a pig in England. DNA was extracted from an amoebal sample as for the Vietnamese pig samples and partial SSU rRNA gene sequences were obtained by amplification using the AEMH 5 and 3 pools of primers. The products were again cloned into pGEM-T Easy as there were several other eukaryotic sequences present in addition to that from the Entamoeba. One clone was sequenced by primer walking but, as the result showed close similarity to E. polecki, the complete gene sequence was not obtained.

Entamoeba ecuadoriensis EC was obtained from the American Type Culture Collection (ATCC 50261; listed as Entamaoba sp.). It was grown at room temperature in TYSGM-9 medium (Clark & Diamond, 2002) supplemented with 5% heat-inactivated adult bovine serum and rice starch. Subsequently, it was grown monoxenically at 37 °C in L1Y-S-2 (Clark & Diamond, 2002) supplemented with 15% heat-inactivated adult bovine serum and antibiotic-inhibited Escherichia coli. The SSU rRNA gene was amplified using primers RD5 and RD3 and sequenced by primer walking, both directly on the PCR product and using a cloned copy in the vector pGEM-T Easy.

Phylogenetic analysis. The novel Entamoeba sequences from the horse and Vietnamese pig and that of E. ecuadoriensis were added to a previously constructed alignment (Silberman et al., 1999). The Entamoeba struthionis sequence, from an ostrich, which had been described since that publication, was also incorporated. The alignment of 18 taxa included 1572 positions after ambiguously aligned regions were removed. The alignment file is available by anonymous FTP from ftp://ftp.ebi.ac.uk/pub/databases/embl/align or from the EMBALIGN database via SRS at http://srs.ebi.ac.uk under accession number ALIGN_000957.

The data were analysed with distance, parsimony, likelihood (all using the PHYLP 3.6 package; Felsenstein, 1989) and Bayesian (Mr Bayes 3.0; Huelsenbeck & Ronquist, 2001) methods. Maximum-likelihood analysis (DNAML) used a model with four categories of among-site rate variation and the proportion of invariant sites. These parameters and the transition/transversion ratio were estimated using Tree-Puzzle 5.0 (Schmidt et al., 2002). Distance analysis (UNADIST and NEIGHBOR) was based on LogDet-transformed distances. Statistical support for the tree topology was evaluated using bootstrapping [1000 replicates for distance and parsimony (TIPARS) analysis, 500 for maximum-likelihood analysis].

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'–3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>542</td>
<td>GTCATCCTGCAAGTATTATGCT</td>
</tr>
<tr>
<td>543</td>
<td>GACTTTGGAAGTCAATTACAG</td>
</tr>
<tr>
<td>764</td>
<td>ATCAAACGTTAGCCAACTTA</td>
</tr>
<tr>
<td>542</td>
<td>ATCCGGTGGTACCTGCGCAT</td>
</tr>
<tr>
<td>RD3</td>
<td>ATCCCTCCGCTTCCTCACAT</td>
</tr>
<tr>
<td>AEMH pool 5</td>
<td></td>
</tr>
<tr>
<td>AEMH 5.1</td>
<td>TCCACGCGTGGCAGCACGC</td>
</tr>
<tr>
<td>AEMH 5.2</td>
<td>TTACAGAAAGGCAAGACG</td>
</tr>
<tr>
<td>AEMH 5.3</td>
<td>TCCAGGAAAGGCAAGACG</td>
</tr>
<tr>
<td>AEMH pool 3</td>
<td></td>
</tr>
<tr>
<td>AEMH 3.1</td>
<td>AAGGGGATACGCGACCTGT</td>
</tr>
<tr>
<td>AEMH 3.2</td>
<td>AAGGGGATACGCGACCTGT</td>
</tr>
<tr>
<td>AEMH 3.3</td>
<td>AAGGGGATACGCGACCTGT</td>
</tr>
</tbody>
</table>
Bayesian analysis used four MCMC strands, 100,000 generations and an initial burn-in of 10,000, at which point the likelihood values had stabilized. The GTR model with proportion of invariant bases and four categories of among-site rate variation were used and trees were sampled every 10 generations. Trees were rooted with the Entamoeba coli sequences as this lineage had previously been shown to emerge at the base of the Entamoeba tree (Silberman et al., 1999).

RESULTS AND DISCUSSION

Isolation and analysis of new sequences

The Entamoeba species from the Vietnamese pigs was in most respects morphologically indistinguishable from E. polecki. The cyst size ranges from 9.5 to 15.5 μm (mean 12.85 ± 1.72 μm based on 100 cysts) and only a single nucleus is ever seen (Fig. 1). Nevertheless, it is quite clear that this species is not E. polecki and is not even closely related to the other uninucleated species studied previously (Fig. 2), being specifically related to Entamoeba gingivalis instead. This means that it is impossible to know which organism was being observed in previously published studies where microscopy was the only method of identification used. A variety of names have been given to such Entamoeba from pigs. Although previously considered by many to be a synonym of E. polecki, it is proposed that the name Entamoeba suis Hartmann, 1913 be resurrected and used for this novel uninucleated cyst-producing species, rather than create a new species name, as it is morphologically consistent with the species description of E. suis (Noble & Noble, 1952). The absence of this organism from humans and other domestic animals in the same location supports the host-restriction implied by the species name.

The Entamoeba from the horse is problematic, as no morphological data exist that can be compared to existing species descriptions. Only two species have been erected for Entamoeba from the horse. Gedoelst in 1911 proposed the name Entamoeba intestinalis for an organism he observed in horses (and other animals), and this was also used by Fantham (1920, 1921). No cysts were described and indeed little morphological information of any sort is provided. Hsiung (1930) (cited by Noble & Noble, 1952) described a small amoeoba and no cysts from a horse and suggested the name Entamoeba gedoelsi to be given as E. intestinalis had been previously used. Whether Hsiung was observing the same organism as Gedoelst and Fantham is impossible to determine. Fantham (1921) observed a second species that produced cysts with four nuclei and which he named Entamoeba equi.

As Entamoeba cysts were not identified in the stool sample, comparisons cannot be made between the organism isolated here and E. equi. However, as this seems to be the only reasonably well-described species with an accepted and logical name, it is proposed that E. equi be used as the species name of the organism from which the gene originated. It is to be hoped that it will not be 75 years before another horse Entamoeba is isolated. The phylogenetic position of E. equi in the Entamoeba phylogeny is unresolved, but it clearly belongs in the clade of tetranucleated cyst-producing species. Indeed, there seems to be a region of the tree where rapid diversification of Entamoeba has occurred, as no resolution of branch order has been possible (Fig. 2).

The isolate named E. ecuadoriensis (Clark & Diamond, 1997) was isolated from sewage in Ecuador in 1964 and was originally described as being a strain of Entamoeba moshkovskii. However, isoenzyme analysis (Sargeaunt et al., 1980) and later SSU rRNA gene restriction enzyme analysis (riboprinting; Clark & Diamond, 1991) showed it to be...
distinct from all other isolates of the latter species. It remains a unique isolate and it cannot be proven that it is truly free-living. The organism can certainly grow over a wide temperature range when in xenic culture. In contrast to *E. moshkovskii*, however, it will not grow in monoxenic or axenic culture at room temperature, only at 37°C. As predicted by riboprinting, *E. ecuadoriensis* is the species most closely related to *Entamoeba histolytica* and *Entamoeba dispar* (Fig. 2), making it of interest for comparative purposes.

Since the original phylogenetic analysis of 14 SSU rRNA gene sequences was published (Silberman *et al.*, 1999), it has become clear that numerous genetic variants of *E. chattoni*/*E. polecki* exist and can infect humans (Verweij *et al.*, 2001). Indeed, it seems that such infections can be locally common, as was the case in Vietnam (Blessmann *et al.*, 2002). There is now no clear distinction between these two species on the basis of host or morphology and so one name should be regarded as a synonym of the other. Despite the original description being inadequate (von Prowazek, 1911), the name *E. polecki* has precedence and should probably be the one retained (Verweij *et al.*, 2001).

The partial sequence of *E. polecki* from a pig in England was obtained several years ago, before the description of *E. struthionis* from farmed ostriches in Spain (Ponce Gordo *et al.*, 2004). Remarkably, there is only one base difference between this pig *Entamoeba* sequence and that deposited as *E. struthionis* in the over 1000 bases sequenced for the former. This suggests that *E. polecki* is not restricted to pigs and humans but can infect birds also, and that the name *E. struthionis* is a synonym of *E. polecki*. This is further supported by the observation that the *E. struthionis* sequence is not basal to the available *E. polecki* and *E. chattoni* sequences in the phylogenetic tree but rather is specifically related to that of *E. polecki* in most analyses (Fig. 2).

**Implications for Entamoeba evolution**

The simplest morphological feature to use in *Entamoeba* species identification is the number of nuclei per cyst. Species produce cysts with one, four or eight nuclei and a few do not encyst. A previous phylogenetic study of *Entamoeba* species (Silberman *et al.*, 1999) suggested that this morphological feature reflected the phylogenetic relationships among organisms, with the species producing cysts with different numbers of nuclei forming distinct clades. The lineage of octanucleate cyst-producing species (represented by *Entamoeba coli*) emerged first, and the non-encysting *E. gingivalis* emerged at the base of the tetranucleate cyst-producing clade.

The sequence of *E. suis* reported here sheds more light on the evolution of cyst nuclear number. Remarkably, the uninucleate cyst-producing *E. suis* is specifically related to the non-encysting *E. gingivalis*. This has two implications. The first is that *E. gingivalis* has secondarily lost the ability to form cysts and is most likely derived from a uninucleated cyst-producing ancestor. The second implication is that the tetranucleated cyst-producing species are descended from a uninucleated cyst-producing ancestor rather than the other way round. This also means that the uninucleated cyst-producers are a paraphyletic group. These are the most parsimonious explanations based on the tree in Fig. 2. A less parsimonious possibility is that *E. suis* has independently
evolved uninucleated cysts from a tetranucleated cyst-producing ancestor.

Another curiosity is *E. eucadoriensis*. This species and *E. moshkovskii* are both thought to be free-living, although the latter does appear to infect humans on occasion (Ali *et al.*, 2003; Parija & Khairnar, 2005). If they are both truly free-living, this characteristic must have arisen twice, unless *E. histolytica* and *E. dispar* are descended from a recent free-living ancestor. As *E. eucadoriensis* has only been isolated once, this question remains unresolved.

The organism here called *E. equi* was not available for morphological analysis. This illustrates a problem that is being encountered more and more, namely correlating molecular data with morphological descriptions. Most similar instances have involved environmental surveys, and no link to the organism is usually possible. In the case of *E. equi*, the amoeba was observed in culture but not in the original sample. Even if the culture had survived, it would not have been possible to link the organism unambiguously to an existing species description as size in culture need not reflect size in vivo. In this case, the genus to which the organism belonged was clear and potential species names existed, but this is not always the case.

The phylogenetic analysis performed here is based on a single marker, the SSU rRNA gene. This was chosen as it is the only sequence available for most *Entamoeba* species, and it is possible that analysis of other genes could resolve the uncertainty surrounding the relationship of *E. equi* to other *Entamoeba* species, but it is unlikely they would alter the main conclusions of this study.

Many additional species of *Entamoeba* have been described for which no molecular data are available. The ability to circumvent the need for parasite culture, using direct PCR amplification of the SSU rRNA gene from DNA extracted from faeces, should allow the phylogenetic relationships of many more *Entamoeba* species to be studied in the future.

**ACKNOWLEDGEMENTS**

Part of this work was supported by a grant from the Biotechnology and Biological Sciences Research Council (UK) to C.G.C. (grant #G18391). A portion of this work formed part of an MSc thesis submitted by F.K. J.W. was supported by a grant from the Institute of Biomedical Sciences (IBMS).

**REFERENCES**


