TITLE

Last of the Human Protists: The Phylogeny and Genetic Diversity of *Iodamoeba*

RUNNING HEAD

Diversity and Phylogeny of *Iodamoeba.*

AUTHORS

C. Rune Stensvold[1*], Marianne Lebbad[2], C. Graham Clark[3].

CURRENT AFFILIATIONS

[1]Department of Microbiological Diagnostics, Statens Serum Institut, Orestads Boulevard 5, DK-2300 Copenhagen S, Denmark

[2]Department of Diagnostics and Vaccinology, Swedish Institute for Communicable Infectious Disease Control, SE-171 82 Solna, Sweden

[3]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

***Corresponding author:** Email: RUN@ssi.dk

Type of article: Letter

Institution at which the work was done: Statens Serum Institut, London School of Hygiene and Tropical Medicine.

Title length (characters including spaces): 76.

Abstract length: 75 words

Total length of text, including all legends and methods, but not Abstract (in characters including spaces): 8,682 (not incl refs)

Total page requirement for all items (expressed as 0.7 pages, 0.5 pages, etc.): 2.5

Number of references: 23

ABSTRACT

Iodamoeba is the last genus of obligately parasitic human protist whose phylogenetic position is unknown. *Iodamoeba* SSU-rDNA sequences were obtained using samples from three host species and phylogenetic analyses convincingly placed *Iodamoeba* as a sister taxon to *Endolimax*. This clade in turn branches among free-living amoeboflagellates of the genus *Mastigamoeba*. Two *Iodamoeba* ribosomal lineages (RL1 and RL2) were detected whose sequences differ by 31%, each of which is found in both human and non-human hosts.

Keywords:

Iodamoeba, protist, parasite, genetic diversity, phylogeny, evolution

Iodamoeba is a genus of intestinal parasitic protist found in humans, non-human primates and other animals. The genus was described by Dobell (1919) who also gave the name *Iodamoeba bütschlii* to the human parasite, and *Iodamoeba* from humans has been assigned to this species ever since. The name *Iodamoeba* derives from the conspicuous iodophilic glycogen mass present in *Iodamoeba* cysts (Supplementary Fig. 1A), often called a vacuole although it is not membrane-bound (Zaman 1972). Cysts are noticeably irregularly shaped, vary in diameter with a mean of ca. 10 µm (Dobell 1919; Taliaferro and Becker 1922), and usually have a single, vesicular nucleus with a large, spherical karyosome. Although mitochondrial structures were reported by Brown (1958) and Dutta (1962), ultrastructural studies did not confirm their presence (Zaman 1972). The life cycle comprises a trophozoite stage, found in the colon where it ingests bacteria and multiplies by binary fission (Dobell 1919; Rodenhuis 1919), and a cyst stage responsible for transmission. Although originally placed in the family Entamoebidae together with *Entamoeba, Dientamoeba* and *Endolimax* (Chatton 1925), to date DNA sequence data have not been available for *Iodamoeba* and therefore its phylogenetic relationships remain unconfirmed. It is also not known whether humans and non-humans are hosts for the same or different species. In this report we finally answer most of the outstanding questions regarding this, the last genus of human parasitic protist to be investigated.

DNA was extracted from purified *Iodamoeba* cysts (Lebbad et al. 2008; Supplementary Fig. 1B), directly from faeces, or from primary culture (Table 1). Complete and partial *Iodamoeba* SSU-rDNA sequences were obtained directly from PCR products, or from clones thereof, using a wide range of primers (Supplementary Table 1). Our results indicate a remarkable degree of genetic diversity within *Iodamoeba*. The sequences obtained fall into one of two ribosomal lineages (RLs) (Table 1, Fig. 1) with a genetic divergence of 31%. Even within each RL a substantial degree of diversity exists (Supplementary Fig. 2).

No two sequences from *Iodamoeba* DNA samples investigated in the study were identical. Substantial genetic diversity (8%) is seen among six clones from EM080 (Table 1; Supplementary Fig. 3) and the divergence between clones EM081-6 and EM081-3.1 in a 1,416 bp overlapping region is 6.7% (not shown). High levels of variation in the SSU-rDNA within strains is uncommon but has been reported previously in, for example, *Dientamoeba fragilis* (Silberman et al. 1996) and *Vannella simplex* (Nassonova et al. 2010). However, in this situation we cannot differentiate between two possibilities: each *Iodamoeba* cell may encode several distinct SSU-rDNA variants (intra-genome variation) or most *Iodamoeba* infections are mixtures of multiple strains, each of which has a single SSU-rDNA variant. Whatever the underlying basis of the variation, the remarkable levels of genetic diversity within single *Iodamoeba* infections has implications for the interpretation of boundaries between Operational Taxonomic Units (OTUs). Caron et al. (2009) used a 95% identity level as their boundary between eukaryotic microbial OTUs. Our data indicate that *Iodamoeba* genes can exceed this 5% divergence value even within an individual infection.

Iodamoeba is well known from pigs and non-human primates, and other examples of natural hosts include rodents, camels and birds (Wenyon 1926; Kessel 1928; Mackinnon and Dibb 1938; Levine 1962; Ray and Banik 1964; Sano et al. 1980; Ponce Gordo et al. 2002; Howells et al. 2011). The fact that *Iodamoeba* sequence 215 from *Macaca fascicularis* is closely related to human RL1 sequences (data not shown) and that RL2 is found in both human and pig suggests that existing *Iodamoeba* species names linked to specific hosts may not be valid. More data are needed to clarify the number and host range of RLs in *Iodamoeba,* and until such data are available we suggest that the two lineages identified in the present study be referred to as *Iodamoeba* RL1 and RL2 rather than allocating species names to each, a similar approach to that recently suggested for novel lineages of *Entamoeba* (Stensvold et al. 2011).

In our phylogenetic analyses, *Iodamoeba*, *Endolimax* and all mastigamoebids always cluster together to the exclusion of the remaining Amoebozoa with strong support, confirming the placement of *Iodamoeba* within this group (Fig. 1). The respective lengths of the SSU-rDNAs of *Iodamoeba* and *Endolimax* are comparable (2.2—2.4 kbp) and in the range of typical mastigamoebid SSU-rDNAs, giving additional credence to the relationship. However, support for the well established taxon Archamoebae as a whole is only moderate except in Bayesian analysis.

The sister taxon relationship of the two genera *Iodamoeba* and *Endolimax* is highly supported but, surprisingly, while monophyly of the two *Iodamoeba* sequences was supported by a high bootstrap value in distance-based analyses, statistical support in Bayesian and maximum likelihood analyses was absent. Manual comparison of the two *Iodamoeba* sequences with the *Endolimax* sequence revealed that shared SNPs were much more frequent between the two *Iodamoeba* RLs than were shared by *Endolimax* and either of the two *Iodamoeba* sequences.

In all our analyses *Endolimax* and *Iodamoeba* share a specific common ancestor (Fig. 1) and their branch emerges from within the free-living amoeboflagellate mastigamoebids rather than clustering with the parasitic *Entamoeba* spp. This indicates that adaptation to parasitism occurred independently at least twice in the Archamoebae, in the ancestor of *Entamoeba* and in the *Iodamoeba*+*Endolimax* branch; we cannot be sure whether the common ancestor of *Iodamoeba* and *Endolimax* was a parasite or not.

We set out to finally resolve the identity of the last genus of human parasitic protist to be studied at the molecular level – *Iodamoeba.* To fully resolve the phylogenetic position and taxonomic status of *Iodamoeba* and *Endolimax* based on SSU-rDNA, more data on intrageneric diversity for both *Endolimax* and *Iodamoeba*, but also *Mastigamoeba,* are needed. For now, we can conclude: 1) that the genus *Iodamoeba* comprises at least two

distinct ribosomal lineages, both of which are found in humans and also occur in non-human hosts, 2) that substantial genetic variation is common in *Iodamoeba* from a single infection, 3) that *Iodamoeba* and *Endolimax* share a most recent common ancestor, and 4) that the genera *Iodamoeba* and *Endolimax* have arisen from within the mastigamoebids.

Acknowledgements

Jaco Verweij and Egbert Tannich are both thanked for providing DNA from *Iodamoeba*-positive samples.

Supplementary Material

A supplementary table and figures are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

REFERENCES

Brown RL. 1958. Cytochemistry of enterozoic protozoan parasites in man. *Abstract Dissertations, Univ. South Calif. Press* p. 9.

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, Dennett MR, Moran DM, Jones AC. 2009. Defining DNA-based Operational Taxonomic Units for microbial-eukaryote ecology. *Appl Environ Microbiol.* 75:5797-5808.

Chatton EPL. 1925. Pansporella perplexa, amoebien à spores protégées, parasite des daphnies. Réflexions sur la biologie et la phylogénie des protozoaires. *Ann Sci Nat Zool.* 8:5—85.

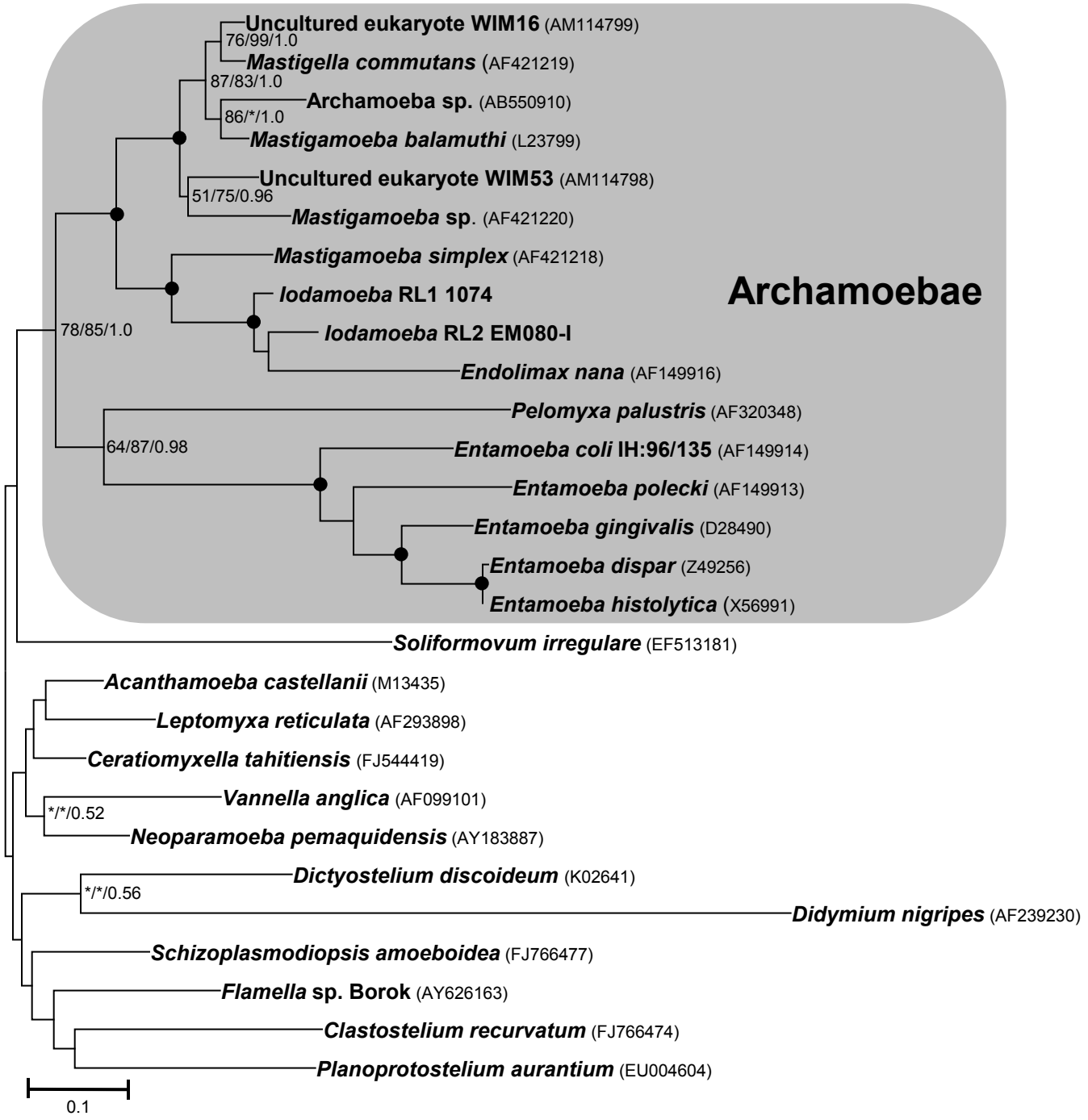
133 Clark CG, Kaffashian F, Tawari B, et al. (13 co-authors). 2006. New insights into the
134 phylogeny of Entamoeba species provided by analysis of four new small-subunit
135 rRNA genes. *Int J Syst Evol Microbiol.* 56:2235—2239.
136 Dobell C. 1919. The amoebae living in man. London: John Bale, Sons and Danieldsson Ltd.
137 Dutta GP. 1962. Cytochemistry of Iodamoeba butschlii. *Proc Nat Inst Sci India* 30:47—55.
138 Howells ME, Pruetz J, Gillespie TR. 2011. Patterns of gastro-intestinal parasites and
139 commensals as an index of population and ecosystem health: the case of sympatric
140 western chimpanzees (Pan troglodytes verus) and guinea baboons (Papio hamadryas
141 papio) at Fongoli, Senegal. *Am J Primatol.* 73:173—179.
142 Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees.
143 *Bioinformatics* 17:754—755.
144 Kessel JF. 1928. Intestinal protozoa of the domestic pig. *Am J Trop Med Hyg.* 8:481—501.
145 Lebbad M, Ankarklev J, Tellez A, Leiva B, Andersson JO, Svärd S. 2008. Dominance of
146 Giardia assemblage B in León, Nicaragua. *Acta Trop.* 106:44—53.
147 Levine ND. 1962. Protozoan parasites of domestic animals and of man. Minneapolis,
148 Minnesota: Burgess Publishing Co.
149 MacKinnon DL, Dibb MJ. 1938. Report on intestinal protozoa of some mammals in the
150 Zoological Gardens at Regent's Park. *Proc Zool Soc Lond. [Ser. B]* 108:323—345.
151 Nassonova E, Smirnov A, Fahrni J, Pawlowski J. 2010. Barcoding amoebae: comparison of
152 SSU, ITS and COI genes as tools for molecular identification of naked lobose
153 amoebae. *Protist* 161:102—115.
154 Ponce Gordo F, Herrera S, Castro AT, Garciá Durán B, Martínez Díaz RA. 2002. Parasites
155 from farmed ostriches (Struthio camelus) and rheas (Rhea americana) in Europe. *Vet*
156 *Parasitol.* 107:137—160.

157 Ray HN, Banik DC. 1964. Occurence of Iodamoeba sp. in an Indian camel. *Bull Calcutta Sch*
158 *Trop Med.* 12:170.
159 Rodenhuis Y. 1919. Bijdrage tot de kennis von die joodcysten en hunne amoeben. *Geneesk*
160 *Tijdschr v Nederl-Indië* 59:813.
161 Sano M, Kino H, de Guzman TS, Ishii AI, Kino J, Tanaka T, Tsuruta M. 1980. Studies on the
162 examination of imported laboratory monkey, Macaca fascicularis for E. histolytica
163 and other intestinal parasites. *Int J Zoonoses* 7:34—39.
164 Silberman JD, Clark CG, Sogin ML. 1996. Dientamoeba fragilis shares a recent common
165 evolutionary history with the trichomonads. *Mol Biochem Parasitol.* 76:311—314.
166 Stensvold CR, Lebbad M, Victory EL, Verweij JJ, Tannich E, Alfellani M, Legarraga P,
167 Clark CG. 2011. Increased sampling reveals novel lineages of Entamoeba:
168 consequences of genetic diversity and host specificity for taxonomy and molecular
169 detection. *Protist* 162:525—541.
170 Taliaferro WH, Becker ER. 1922. The human intestinal amoeba, Iodamoeba williamsi, and
171 its cysts (iodine cysts). *Am J Hyg.* 2:188—207.
172 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular
173 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
174 maximum parsimony methods. *Mol Biol Evol.* (In press).
175 Wenyon CM. 1926. Protozoology, vol. 1. New York: William Wood and Company.
176 Zaman V. 1972. Ultrastructure of Iodamoeba butschlii cysts. *Ann Trop Med Parasitol.*
177 66:251—253.
178

FIGURE LEGEND

Fig.1. Phylogenetic position of *Iodamoeba*. The analysis used 1,430 unambiguously aligned positions from 14 archamoebae, 2 *Iodamoeba* and a broad selection of 12 non-archamoeba amoebozoan sequences. Alignments were generated using MEGA 5 (Tamura et al., 2011) and the inbuilt MUSCLE alignment algorithm then edited. Phylogenetic analyses used three different approaches: distance-based analysis (MEGA 5) used the Neighbor-Joining algorithm and the Maximum Composite Likelihood model, while Bayesian (MrBayes 3.1.2; Huelsenbeck and Roquist 2001) and maximum likelihood (MEGA 5) analyses both used the General Time Reversible (GTR) model of nucleotide substitution with four categories of among-site rate variation and the proportion of invariant sites, selected as best using ModelTest (MEGA5). Statistical support for distance and maximum likelihood trees was evaluated using bootstrapping (1,000 replicates). Bayesian analysis used four Markov chain Monte Carlo (MCMC) strands and 5,000,000 generations, with trees sampled every 100 generations. In the Bayesian analysis the final average standard deviation of split frequencies was less than 0.01. A consensus tree was produced after excluding an initial burn-in of 25% of the samples, as recommended. The Maximum Likelihood tree is shown. Bootstrap values and posterior probabilities from the three types of phylogenetic analyses are shown in the following order: Maximum Likelihood/Distance/Bayesian. Nodes where both bootstrap values are >95% and the posterior probability is >0.95 are indicated by black circles. Bootstrap values of <50 or posterior probabilities of < 0.50 are indicated by an asterisk and where all three analyses show these low support values the node is not labelled.

Fig. 1.


Uncultured eukaryote WIM16 (AM114799)
76/99/1.0
Mastigella commutans (AF421219)
87/83/1.0
Archamoeba sp. (AB550910)
86/*/1.0
Mastigamoeba balamuthi (L23799)
Uncultured eukaryote WIM53 (AM114798)
51/75/0.96
Mastigamoeba sp. (AF421220)
Mastigamoeba simplex (AF421218)
Iodamoeba RL1 1074
Iodamoeba RL2 EM080-I
78/85/1.0
Endolimax nana (AF149916)
Archamoebae
Pelomyxa palustris (AF320348)
64/87/0.98
Entamoeba coli IH:96/135 (AF149914)
Entamoeba polecki (AF149913)
Entamoeba gingivalis (D28490)
Entamoeba dispar (Z49256)
Entamoeba histolytica (X56991)
Soliformovum irregulare (EF513181)
Acanthamoeba castellanii (M13435)
Leptomyxa reticulata (AF293898)
Ceratiomyxella tahitiensis (FJ544419)
Vannella anglica (AF099101)
//0.52
Neoparamoeba pemaquidensis (AY183887)
Dictyostelium discoideum (K02641)
//0.56
Didymium nigripes (AF239230)
Schizoplasmodiopsis amoeboidea (FJ766477)
Flamella sp. Borok (AY626163)
Clastostelium recurvatum (FJ766474)
Planoprotostelium aurantium (EU004604)
0.1

201 Table 1. Samples used and sequences produced for phylogenetic analyses.

DNA Sample ID	Source of DNA	Host	Geographical info (Travel/Origin)	Sequence ID	Sequence Length (bp)	GenBank Accession No.[c]	Sequence from PCR Products /Clone	*Iodamoeba* ribosomal lineage (RL)
EM081	Cysts	*Homo sapiens*	Thailand	EM081-6	1,752	JN635745	Clone	1
				EM081-3.1	1,961	JN635746	Clone	1
1074	Faeces	*Homo sapiens*	NA[b]	1074	2,376	JN635741	PCR	1
82	Faeces	*Homo sapiens*	NA	82	2,193	JN635742	PCR	1
28	Faeces	*Homo sapiens*	NA	28	509	JN635743	PCR	1
215	Faeces	*M. fascicularis*	NA	215	150	NA	PCR	1
EM080	Cysts	*Homo sapiens*	Cuba	EM080-I	2,215	JN635740	Clone	2
				EM080-A-H	252-257	JN635747-51	Clone	2
Mabel	Culture[a]	*Sus scrofa*	UK	Mabel	1,190	JN635744	Clone	2

202 [a]Clark et al. 2006.

203 [b]NA: Not available

204 [c]GenBank accession number for sequence 215 was not available since the length of the sequence was < 200 bp.

Supplementary Table 1. Primers used in the study for PCR amplification and sequencing.

Primer name[a]	Primer sequence (5'--3')	Primer specificity[b]	Primer position[c]	DNA sample sequenced
RD5	ATCTGGTTGATCCTGCCAGT	E	1-20	EM080-I, EM081-3.1, 1074, 82
RD3	ATCCTTCCGCAGGTTCACCTAC	E	2,194-2,215	EM080-I, EM081-6, 1074
AEMH5.2[d]	TCTAAGGAAGGCAGCAGGC	E	581-599	EM081-6, EM081-3.1, MABEL
AEMH3.1[d]	AAGGGCATCACGGACCTGTT	E	1835-1854	EM081-3.1, MABEL
528F	GCGGTAATTCCAGCTC	E	745-760	EM080-I, 1074
528R	GAGCTGGAATTACCGC	E	745-760	EM080-I
1200F	CAGGTCTGTGATGCCC	E	1837-1852	EM080-I
IODAGENUS1580F	ATCGAGTGAGTGTATGGGCTTC	G	1468-1489	1074
IODAGENUS_F	GGGGTGGTTTATATTTCATAGCG	G	1199-1222	EM080-A, EM080-B, EM080-C, EM080-E, EM080-H, 215
IODAGENUS_R	TCTCTCTAGGTGCTGGAGGAGTC	G	1443-1465	EM080-I, EM081-6, EM081-3.1
IODAGENUS2300R	CCGAAGCCCATACACTCATTC	G	1471-1491	1074, 28
IODAGENUS650F	GTAGTGACGACAAATACCGATG	G	629-650	EM080-I, 82, 28

IODAGENUS780R	CCGCAACAGCTTTAGTATACACTC	G	764-787	EM080-I, 82
IODAMOEBA100F	AAGGATAACCCTGTTAATTGTAGAG	G	141-165	EM080-I, 82
IODAMOEBA2080R	CCCCAGCTTGATGAACATTAC	G	1930-1950	EM080-I
IODAMOEBA1610R	CAGCCTTGCGACCATACTC	G	1468-1486	EM080-A, EM080-B, EM080-C, EM080-E, EM080-H, EM080-I, 1074
IODAGENUS1230F	AATTGGGGTGGTTTATATTTCATAGC	G	1172-1197	1074, 82, 215
IODAGENUS2220R	CAAATCCAACATTTTCACCG	G	2074-2093	82
IODAM1200R	ATGCACTACCCACAGCACAC	O	**1161-1182**	EM081-3.1
IODAM1450R	TACACCCTGTGTTACCAGTGTG	O	**911-932**	1074
IODAM1400R	GTCTGCAGCGATTGTTTCTATTC	O	**1463-1485**	1074, 82
IODAM1500R	CAAAACATCACATAAATGTTCTGCC	O	**873-897**	1074, EM081-3.1
IODAM520R	CACACACAAGTGCGCACTG	O	**1149-1166**	EM081-6, 1074
IODAM450F	GAAGATATGTCTCGTGGGTGC	O	**1066-1086**	EM081-6
IODclone3.1_1200F	TGAGCGTCACAACAGTGGC	O	**1244-1262**	EM081-3.1
IODAM580R	CGTGGTCAATATGCATAGTTTATTATAGAC	O	**1208-1237**	EM081-6, EM081-3.1, 1074
IODAM600F	CACAACCAGTGCTTAGGAATAGAC	O	**1250-1273**	EM081-6, 1074

IODAM500F	GTGTCTAGTTGCAGTGCGC	O	**1138-1156**	EM081-6, 1074
T7	TAATACGACTCACTATAGGG	GSP	NA	EM081-6, EM081-3.1
SP6	ATTTAGGTGACACTATAG	GSP	NA	EM081-6, EM081-3.1

[a] Primer name does not necessarily reflect position of primer

[b] E = Broad-specificity, eukaryotic primer; G = *Iodamoeba* genus-specific primer (with a maximum of 2-3 mismatches in primer and/or conserved 3'-end; O = lineage-/strain-specific (RL1) ; GSP = General sequencing primer (cloning).

[c] Primer position relative to EM080-I (non-bold) and 1074 (bold); NA = Not applicable.

[d] part of AEMH5/3 pool (Clark et al., 2006).

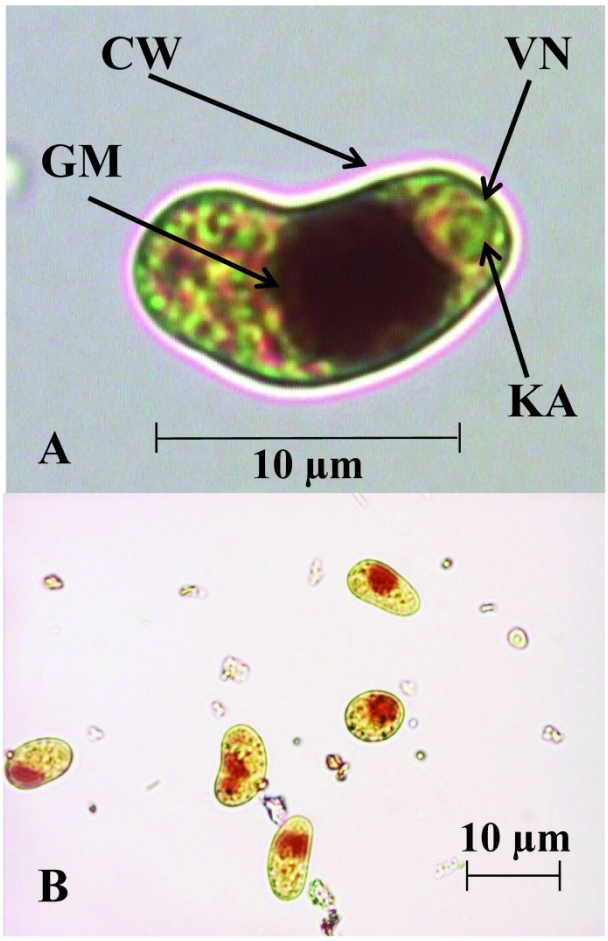
Supplementary figures

Supp. Fig. 1. *Iodamoeba* cysts observed by light microscopy of an iodine stained preparation. **A:** Single cyst showing morphological features. Size is indicated. CW = Cyst wall, VN = Vesicular nucleus, KA = Karyosome, GM = Glycogen mass. **B:** Cyst preparation of *Iodamoeba* EM080 showing the absence of other protist cysts.

Supp. Fig 2. Genetic diversity of *Iodamoeba*. *Iodamoeba* inter-sample phylogeny showing two ribosomal lineages and substantial intra-lineage diversity. A total of 383 unambiguously aligned positions in the region common to all sequences were used in the analysis. Maximum likelihood tree produced as in Fig. 1 is shown.

Supp. Fig. 3. Alignment of EM080 SSU-rDNA clones showing intra-sample genetic diversity. The sequence shown corresponds to positions 1,192—1,447 in EM081-I. * = identical base in all clones. The diversity detected consisted in two instances of differences in homopolymer length: at position 103 EM080-E has a homopolymer of four Gs, whereas the other clones have five Gs. At position 219, three clones have a homopolymer of three Ts while the others have two Ts. Also, a short region starting at position 182 exhibited single nucleotide polymorphisms (SNPs) and insertions/deletions (indels); in the same region EM080-C is clearly divergent from the other clones although it still belongs to the same ribosomal lineage (RL2).

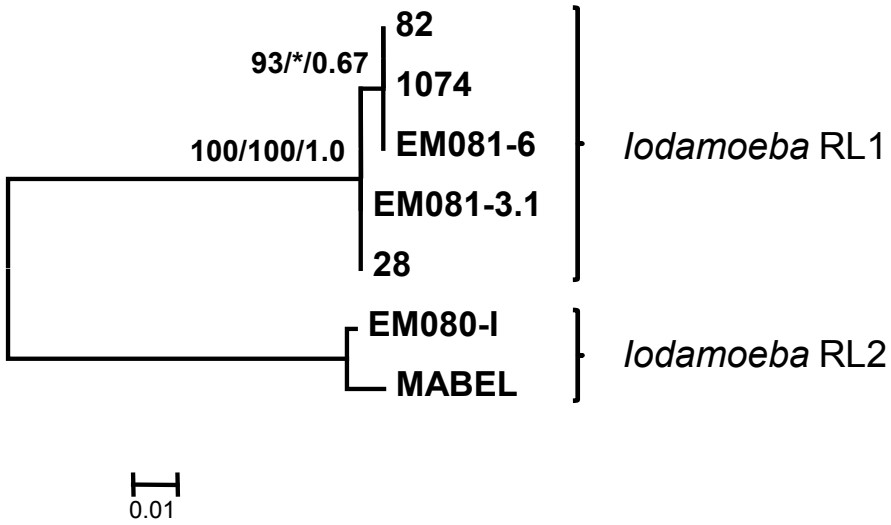
230 Supplementary Fig. 1.



231

232

Supplementary Fig. 2



Supplementary Fig. 3.