Unravelling the phylogeny of a common intestinal protist: Intra-generic diversity of *Endolimax*

Emma Filtenborg Hocke[a], Mahwash Jamy[b], Fabien Burki[b,c], C. Graham Clark[d,1], Christen Rune Stensvold[a,1]

Affiliations: [a]Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark; [b]Department of Organismal Biology, Systematic Biology, Uppsala University, Uppsala, Sweden; [c]Science for Life Laboratory, Uppsala, Sweden; [d]Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

Running title: Intra-generic diversity of *Endolimax*

[1]Corresponding authors: C. Graham Clark, Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom. Tel.: +44 207 927 2351. Email: graham.clark@lshtm.ac.uk and Christen Rune Stensvold, Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Artillerivej 5, Copenhagen, Denmark. Tel.: +45 3268 8224. Email: run@ssi.dk

Abstract

Endolimax nana is a common endobiont of the human intestine, but members of the genus have also been reported in non-human hosts and in non-intestinal organs. Limited information is available regarding the genetic diversity of *Endolimax*, which is necessary to delineate species, host specificity and potential differences in clinical impact on the host. Here, we used cloning of PCR products followed by Sanger sequencing and next-generation PacBio Sequencing to obtain *Endolimax*-related nuclear ribosomal gene sequences and undertook a phylogenetic analysis to gain additional insight into the taxonomy of *Endolimax* and related organisms. The new sequences confirmed that *E. nana* forms a discrete clade within the Archamoebae and is related to *Endolimax piscium* and *Iodamoeba*. However, we identified substantial sequence divergence within *E. nana* and evidence for two distinct clades, which we propose to name *E. nana* ribosomal lineage 1 and *E. nana* ribosomal lineage 2. Both of the sequencing approaches applied in the study helped us to improve our understanding of genetic diversity across *Endolimax*, and it is likely that wider application of next-generation sequencing technologies will facilitate the generation of *Endolimax*-related DNA sequence data and help complete our understanding of its phylogenetic position and intrageneric diversity.

Key words: genetic diversity, intestinal parasite, host specificity, next-generation sequencing, ribosomal gene sequencing, taxonomy.

INTRODUCTION

Amoeboid protists of the genus *Endolimax* have been reported in faecal samples from humans, other mammals, reptiles, amphibians, fish, birds, and insects (Poulsen & Stensvold, 2016). So far, *Endolimax nana* is the only named species in humans (Constenla, Padrós, & Palenzuela, 2014) and is commonly reported in faecal samples (Fitzgerald & O'Farrell, 1954; Graczyk et al., 2005; Santos et al., 2014; Shah et al., 2012; Stauffer & Levine, 1974). *Endolimax nana* remains one of the most common but least studied species of the parasitic Archamoebae. In a recent review, Poulsen and Stensvold calculated weighted prevalence averages of 3.4% and 13.9% among 1,409,022 and 93,815 individuals with and without gut symptoms, respectively (Poulsen & Stensvold, 2016). While this protist has not been proven to cause pathology, *Endolimax* is often used as an indicator of faecal exposure and it is often observed in co-infections with organisms known to cause diarrhoea (Ignacio et al., 2017).

Endolimax has been found to exhibit extensive genetic diversity (Constenla et al., 2014; Poulsen & Stensvold, 2016; Stensvold et al., 2020). However, the phylogenetic relationships of *Endolimax* have changed over time as more information has become available. Initially, *Endolimax* was proposed as the closest relative to *Entamoeba*, based on limited taxon sampling (Silberman, Clark, Diamond, & Sogin, 1999). Most phylogenetic studies to date still use only this first complete sequence of an *Endolimax* small subunit (SSU) ribosomal RNA gene (rDNA) available in the NCBI Nucleotide Database, namely *E. nana* NIH:0591:1 (AF149916). Subsequently, *Endolimax* was proposed to be a sister taxon to the free-living protist genus *Mastigamoeba* (Cavalier-Smith, Chao, & Oates, 2004), while recent studies suggest *Endolimax* is a sister taxon to *Iodamoeba* (Stensvold, Lebbad, & Clark, 2012; Zadrobilkova, Walker, & Cepicka, 2015). Although *Endolimax* is now confirmed as a lineage within the Mastigamoebidae group B, the specific affinities of *Endolimax* species remain unclear, and wider sampling is needed to clarify the levels of intrageneric diversity.

Obtaining *Endolimax* SSU rDNA sequences can be challenging due to the absence of cultured material and the fact that the SSU rDNA of *E. nana* is relatively long (~2.5 kbp) (Silberman et al., 1999). PCR using general eukaryotic primers preferentially amplifies any shorter and more abundant SSU rDNA from co-infecting/co-colonising organisms present in the intestine. This is often *Blastocystis* sp., which is frequently observed in *Endolimax*-positive samples, as its SSU rDNA are around 700 bp shorter than those of *Endolimax*. Even when specific amplification is successful, the PCR product can be difficult to sequence due to high intra-genome variation among

the ribosomal gene copies (Poulsen & Stensvold, 2016). This makes direct Sanger sequencing of PCR products problematic and unable to clarify genetic diversity.

The present study used two different approaches to address *Endolimax* diversity, namely 1) cloning of PCR products from single faecal DNA samples followed by Sanger sequencing, as previously done with *Iodamoeba* (Stensvold et al., 2012), and 2) the use of next-generation PacBio sequencing to sequence single molecules. Both of these approaches helped improve our understanding of genetic diversity within this genus, and the latter method also provided evidence of a novel archamoebid related to *Endolimax* and *Iodamoeba*.

RESULTS

Endolimax SSU rDNA sequences obtained by Sanger sequencing

Endolimax SSU rDNA sequences were obtained by PCR amplification from two faecal DNA samples (H80028 and EN18) by combining a previously described forward primer and a new reverse primer (Table 1). The amplicons generated consensus sequences with a length of ~1,750 bp, which is equivalent to about 65%–70% of the complete SSU rDNA of *Endolimax*. The new consensus sequences were aligned with the two almost full-length *Endolimax nana* SSU rDNA sequences already present in the NCBI GenBank database (AF149916 and LC230015), and they showed between 84% and 98% identity in the region of overlap (1,791 bp).

SSU rDNA sequences generated using PacBio sequencing

Four *Endolimax*-related SSU rDNA sequences were obtained by PCR amplification from pooled wastewater DNA samples (SW01–SW04; DNA sequence length, ~1,580 bp). When aligned with the two almost full-length *Endolimax nana* SSU rDNA sequences in the NCBI GenBank database, three sequences showed between 85% and 98% identity in the region of alignment overlap (1,819 bp). In contrast, the fourth sequence (SW04) showed much less similarity to the other new sequences and the database sequences (61%–63% identity).

Genetic distances

An initial multiple sequence alignment was produced that included all the newly generated sequences, the sequences previously deposited in GenBank representing *E. nana*, *Endolimax* sp., *Endolimax piscium* from fish, and both ribosomal lineages (RL) of *Iodamoeba* (Stensvold et al., 2012). Only the region covered by the PacBio sequences was included, and regions of ambiguous alignment were excluded, leaving 967 aligned positions. Pairwise distances were calculated (Table 2). The values obtained clearly indicated two clusters of *E. nana* and *Endolimax* sp. sequences, consisting of NIH:0591:1, SW01 and SW02; and SW03, TDP-2, H80028, and EN18, respectively.

Phylogeny

Maximum likelihood and Bayesian phylogenetic analyses confirmed the inferences from the genetic distances and each analysis recovered the same topology (Figure 1). A single clade containing all *E. nana* and *Endolimax* sp. sequences, with two strongly supported subclades, was recovered with maximal support. However, monophyly of the genus *Endolimax* was not supported, as the *E. nana* clade did not cluster with *E. piscium*. A clade containing *Iodamoeba* and SW04 (Bootstrap, BP = 77, Posterior probability, PP = 0.96) was quite well supported, as was the grouping of all *Endolimax*, *Iodamoeba* and the SW04 sequences (BP = 82, PP = 1.0).

DISCUSSION

Amplification of *Endolimax nana* SSU rDNA has proven problematic when using broad-specificity SSU rDNA primers. This is in part due to the size of the gene – it is among the longest eukaryotic SSU rDNAs – and in part due to the frequency by which *E. nana* is found in mixed infections involving other human parasites. The gene length means that even when the relative numbers of *E. nana* in a sample are comparable, its SSU rDNA amplification will be less efficient, with shorter SSU rDNA sequences like those of *Blastocystis* being amplified preferentially. These observations mean that any investigation of the *Endolimax* DNA in a sample will require the use of specific primers rather than general primers. However, using this approach assumes we know enough about genetic diversity in these organisms to design primers that will amplify DNA from all relevant sample types. We believe that some of the primers used in this study (e.g., IO_LIMAX_F and IO_LIMAX_R) will amplify all *Endolimax* and related SSU rDNA, but inevitably there may be some organisms of interest that do not amplify due to sequence divergence in the primer locations.

The use of single-molecule long-read technology (PacBio sequencing) has been shown here to overcome some of the limitations of previous approaches. Cloning of PCR products prior to sequencing does avoid the issue of sequence variation between gene copies that makes direct sequencing of PCR products problematic. However, that approach is labour intensive and slow in comparison with sequencing of single molecules. The latter also allows the detection of multiple organisms present in different numbers in the same sample, which would again be difficult using other approaches. The number of reads obtained for SW01–SW04 varied over five-fold, from several hundreds to over 2000, but the number of reads adds to confidence that these sequences are not artefacts, such as sequence chimeras.

The new sequences generated in this study confirm that *E. nana* forms a discrete clade within the Archamoebae and is related to *Endolimax piscium* and *Iodamoeba.* However, within *E. nana* there is substantial sequence divergence and an indication that there may be at least two clades, perhaps as many as the four indicated in our recent study of Swedish wastewater samples (Stensvold et al., 2020). The two *E. nana* clades identified in the present study correspond to two of the clades identified previously (Stensvold et al., 2020), represented by NIH:0591:1 and H80028. We propose to call these clades *E. nana* RL1 and RL2, respectively (Figure 1). We moreover propose that the sequences deposited in GenBank as "*Endolimax* sp." (H80028 and TDP-2) should be considered *E. nana* since they are all closely related.

There are five sequences from the TDP-2 sample in the NCBI Database, all of which were obtained using a plasmid cloning procedure (Yoshida et al., 2019). Although not acknowledged by the authors, based on the analyses in the present study, the TDP-2 sequences do represent *E. nana.* Moreover, the TDP-2 sequences are from a pig and possibly the first DNA-based evidence of *Endolimax* in a pig host. For *Iodamoeba*, it is also known that one of the two ribosomal lineages identified to date is able to colonise pigs. Hence both *Endolimax* and *Iodamoeba* can colonise pigs, but it remains to be confirmed whether multiple RLs from each genus can colonise pigs.

Endolimax nana was found in human faecal samples but also in wastewater samples (Stensvold et al., 2020). The latter is not surprising. However, also isolated from wastewater was a related sequence (SW04) that did not cluster with *E. nana* or *E. piscium* but with *Iodamoeba*. We suspect that the organism from which this sequence derives has a non-human host, but in the absence of information about the organism's source and morphology, it is not even possible to assign this organism to a genus. While SW04 appears to be related to *Iodamoeba*, there is a sequence in

GenBank that shows greater similarity. The sequence, KU658872, was found in an anaerobic reactor sample from Luxembourg. Unfortunately, it is only 293 bp in length; however, SW04 and KU658872 exhibit 96% identity over the first 181 bp of SW04 and 84% identity over the full length of KU658872.

The phylogenetic relationships depicted in Figure 1 are consistent with those obtained by others (Zadrobilkova et al., 2015). As in previous analyses, the relationships among *E. nana*, *E. piscium*, and *Iodamoeba* are poorly resolved, although together they clearly form a clade within the Mastigamoebidae Group B (Ptáčková et al., 2013; Zadrobilkova et al., 2015). With the addition of SW04 as an additional distinct lineage within this clade, but with no information about its morphology, the question of the appropriate genus or genera for these organisms is also unresolved. Interestingly, in the early twentieth century *Iodamoeba* was sometimes assigned to the genus *Endolimax* (as *Endolimax williamsi*) (reviewed in Taliaferro and Becker, 1922). However, the cysts of *Iodamoeba* are morphologically quite distinct to those of *Endolimax* and this is widely used in microscopic diagnosis, which led to separation of the two genera. Other species of *Endolimax* from various hosts have been described over the years (Poulsen & Stensvold, 2016) but, to our knowledge, no DNA sequences are available for these, apart from *E. piscium*. However, from Figure 1 it is unclear whether *E. piscium* and *E. nana* should be considered congeneric and so the taxonomy of *Endolimax* will likely need to be revisited in the future. If sequences from other *Endolimax* and related species become available they may allow better resolution in this part of the Amoebozoan phylogenetic tree and, indeed, may well give us a very different picture of relationships between the lineages discussed above.

METHODS

PCR, TA cloning procedure, and Sanger sequencing

Genomic DNA from two stool samples (H80028 and EN18) was used. These had previously been identified as positive for *Endolimax* by microscopy and/or PCR and Sanger sequencing methods. *Endolimax* SSU rDNA was amplified using genus-specific primers (Table 1). The PCR used Extract-N-Amp PCR ReadyMix (Sigma-Aldrich, Søborg, Denmark). Cycling conditions consisted of initial denaturation (3 min at 94 °C) followed by 35 amplification cycles (1 min at 94 °C, 1 min at 55 °C, and 2 min at 72 °C) followed by a final extension (5 min at 72 °C). An amplicon of ~1,750

bp was purified from 1.5% low melting point agarose gels using QIAquick PCR purification kit (Qiagen Inc., Valencia, California, USA). Amplicons were cloned into the pCR™2.1 Vector and transformed into One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen, Portland, Oregon, USA). The presence of the insert in transformants was confirmed by PCR with *Endolimax*-specific primers (Table 1). One clone from each sample was sequenced *in house* and bidirectionally using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fischer, Waltham, Massachusetts, USA) with the use of specific and general eukaryotic primers (Table 3). Sequences were assembled and edited using the Staden Package (Staden, Beal, & Bonfield, 2000).

PCR and PacBio sequencing

Genomic DNAs from 10 wastewater samples used in a recent study (Stensvold et al., 2020) were amplified using the *Endolimax-/Iodamoeba*-specific primers listed in Table 4 and Extract-N-Amp PCR ReadyMix. Cycling conditions consisted of an initial denaturation (3 min at 94 °C) and 35 amplification cycles (94 °C for 1 min, 54 °C for 1 min, 72 °C for 2 min) followed by final extension (5 min at 72 °C). PCR amplicons were pooled and sequenced by PacBio on a Sequel I SMRT cell. PacBio sequences were processed using a pipeline modified from a previous study (Jamy et al., 2020). Briefly, circular consensus sequences (CCS) were generated from raw reads using pbccs v3.4.0 (https://github.com/PacificBiosciences/unanimity/blob/develop/doc/PBCCS.md) with the following settings: minLength=10, maxLength=21000, minPasses=3, minPredictedAccuracy=0.99. This resulted in a fastq file containing 251665 CCS. A fasta file was generated using the fastq.info (pacbio=T) option in mothur v1.39.5. PCR artefacts such as incomplete amplicons and sequencing errors such as long homopolymers runs were then filtered out using the trim.seqs command in mothur using the following settings: qwindowsize=50 and qwindowaverage=30 (to trim CCS with a stretch of low quality nucleotides), maxhomop=9 (to discard CCS with a homopolymers run of more than 9 nucleotides), and minLength=900, maxLength=4000 (to discard non-specific and incomplete amplicons). The remaining non-specific PCR amplicons were filtered out using Barrnap v0.7 (--reject 0.3, --kingdom euk) (https://github.com/tseemann/barrnap). Only CCS containing the SSU rDNA were retained. An in-house script was used to detect sequences represented by the reverse strand; these were subsequently reverse-complemented, so that all sequences were in the same direction. The sequences were then dereplicated before performing *de novo* chimera detection. The curated sequences were then clustered at 97% identity using vsearch

v2.3.4 (--cluster_fast --id 0.97) to yield 6152 operational taxonomic units (OTUs). OTU sequences were used as queries against the NCBI nt database using blastn with default parameters and were found to cover a range of diversity in addition to *Endolimax*. Relevant OTU representatives were extracted if the best BLAST hit was *Endolimax nana* (187 OTUs) or *Endolimax piscium* (48 OTUs). For phylogenetic analysis, OTUs observed with fewer than 350 sequences were excluded.

Genetic distances

Pairwise distances between newly obtained sequences and those already in GenBank databases were calculated using MEGAX (Kumar et al. 2018) following sequence alignment using MUSCLE (Edgar, 2004) as implemented in MEGAX.

Phylogenetic analysis

Sequences were aligned with mafft-qinsi (Katoh and Standley 2013) and then trimmed with trimal (-gt 0.1, -st 0.001) to remove the sites with most gaps (Capella-Gutiérrez et al. 2009). The final alignment contained 21 *Endolimax* and related Mastigamoebidae A and B taxa and 2,067 positions.

Phylogenetic relationships were inferred using two different approaches: Maximum Likelihood (ML) and Bayesian Inference (BI). ML analyses were carried out in raxml-ng (Kozlov et al. 2019) using the GTR+Gamma model. The topology with the best likelihood score out of 20 ML searches was selected and support was evaluated with 100 bootstrap replicates (until bootstrap convergence). BI was carried out in MrBayes v3.2.6 (Huelsenbeck and Ronquist 2001) under the GTR+Gamma model, with two Markov chains run for 4,000,000 generations and sampled every 1000 generations. The average standard deviation of split frequencies was < 0.01, which indicated that the Markov chains reached convergence. Consensus tree and posterior probabilities were calculated using 3000 trees after discarding the first 1000 trees as burn-in.

Data deposition

DNA sequences for samples SW01, SW02, SW03, EN18 and SW04 were submitted to the NCBI GenBank Database with the accession numbers OK483220, OK483221, OK483222, OK483223 and OK483224, respectively; the sequence from sample H80028 was submitted under the accession

number MN556101. The raw PacBio data are available in the European Nucleotide Archive under accession PRJEB48208.

Acknowledgements

We thank our colleagues Marianne Lebbad, Jessica Beser, and Anette Hansen from the Public Health Agency of Sweden for sharing DNA from wastewater samples with us. We also thank Prof Staffan Svärd, Department of Cell and Molecular Biology, Microbiology, Uppsala University, Uppsala, Sweden, for facilitating the PacBio work. Part of this work in FB's laboratory was supported by a grant from Science for Life Laboratory and a grant from the Swedish Research Council Vetenskapsrådet (2017-04563).

264 **Table 1.** Primers used for amplification of *Endolimax* SSU rDNA in the two clinical samples
265 (H80028 and EN18).

Primer Name	**Primer Sequence (5'-3')**	**Reference** 266
Limax_2F	GGAGCAATTGGAATGAAAGCAAG	Poulsen & Stensvold, 267 2016
Limax_2R (2018)	GAACCTTAATATCTAGAGGAAGGAG	Present study 268 269

Table 2. Pairwise distances among *Endolimax* and related sequences.

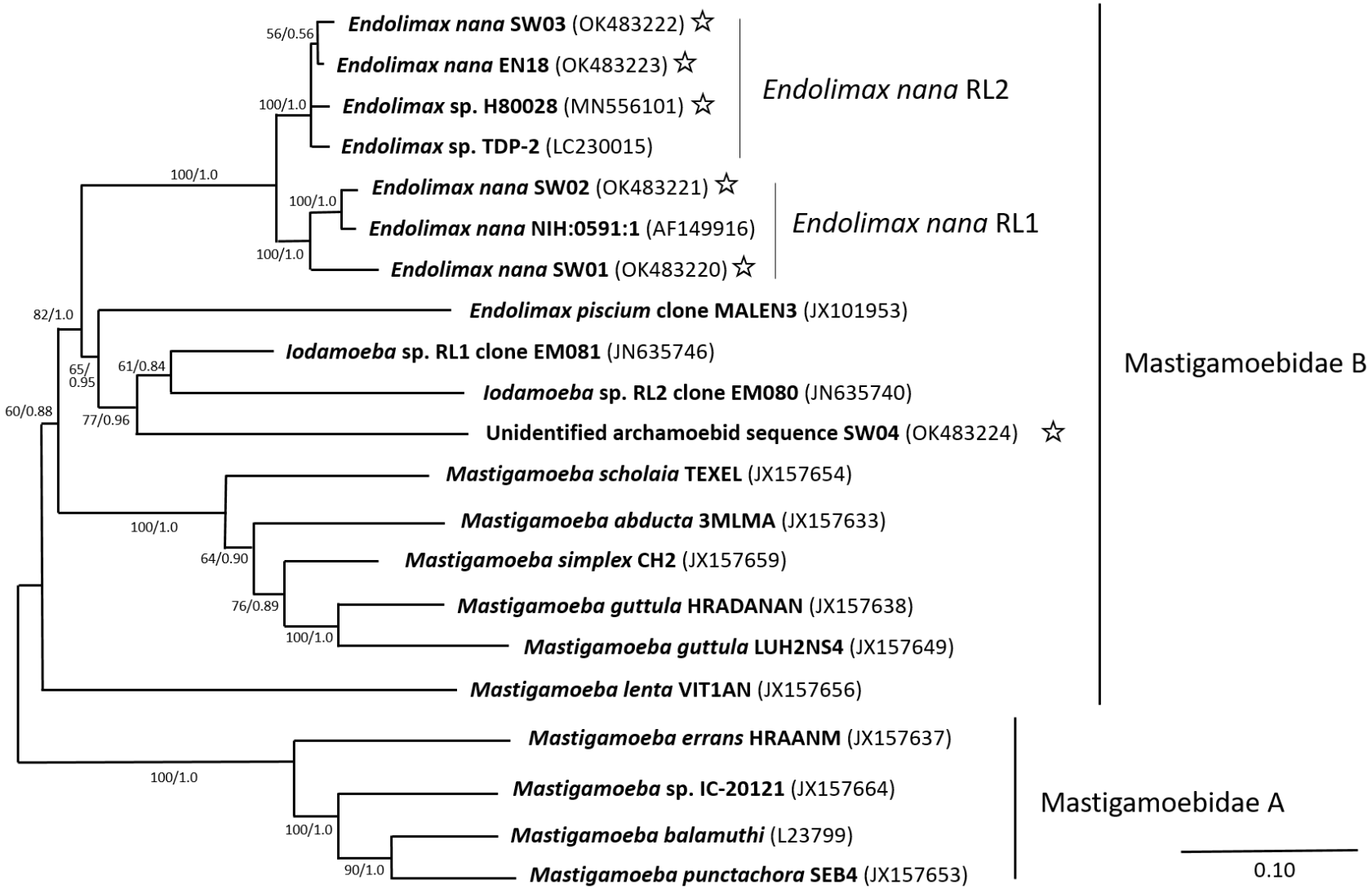
	Iodamoeba sp. RL1	*Iodamoeba* sp. RL2	*Endolimax piscium*	*Endolimax nana* NIH:0591:1	*Endolimax* sp. TDP-2	*Endolimax* sp. H80028	*Endolimax nana* EN18	*Endolimax nana* SW01	*Endolimax nana* SW02	*Endolimax nana* SW03
Iodamoeba sp. RL1										
Iodamoeba sp. RL2	0.123									
Endolimax piscium	0.232	0.232								
Endolimax nana NIH:0591:1	0.236	0.228	0.259							
Endolimax sp. TDP-2	0.253	0.238	0.272	0.045						
Endolimax sp. H80028	0.240	0.231	0.266	0.048	0.010					
Endolimax nana EN18	0.265	0.249	0.285	0.056	0.017	0.017				
Endolimax nana SW01	0.248	0.236	0.268	0.032	0.067	0.064	0.078			
Endolimax nana SW02	0.240	0.232	0.267	0.005	0.043	0.045	0.053	0.035		
Endolimax nana SW03	0.250	0.235	0.269	0.045	0.007	0.007	0.015	0.067	0.043	
Unidentified archamoebid sequence SW04	0.173	0.196	0.211	0.260	0.266	0.252	0.278	0.259	0.260	0.263

Table 3. Primers used for Sanger sequencing of *Endolimax nana* SSU rDNA.

Primer Name	Primer Sequence (5'-3')	Reference
Limax_2F	GGAGCAATTGGAATGAAAGCAAG	Poulsen & Stensvold, 2016
Endoligenus F	GTGGAATGCTTTCGCTCTC	Poulsen & Stensvold, 2016
Limax_2R	GTCGTAGTCTCAACCATAAACG	Poulsen & Stensvold, 2016
1055F	GTGGTGCATGGCCGT	Stensvold et al., 2011
1055R	ACGGCCATGCACCAC	Stensvold et al., 2011
Limax_2R (2018)	GAACCTTAATATCTAGAGGAAGGAG	Present study

Table 4. Primers used for amplification of *Endolimax* SSU rDNA in 10 pooled genomic DNAs extracted from wastewater samples.

Primer Name	Primer Sequence (5'-3')	Reference
IO_LIMAX_F	CTGCCAGTAGTCATATGCTTGTG	Present Study
IO_LIMAX_R	GAGACTACGACGGTATCTGATCG	Present Study


Endolimax nana SW03 (OK483222) ☆
Endolimax nana EN18 (OK483223) ☆
Endolimax sp. H80028 (MN556101) ☆
Endolimax sp. TDP-2 (LC230015)
Endolimax nana SW02 (OK483221) ☆
Endolimax nana NIH:0591:1 (AF149916)
Endolimax nana SW01 (OK483220) ☆
Endolimax nana RL2
Endolimax nana RL1
Endolimax piscium clone MALEN3 (JX101953)
Iodamoeba sp. RL1 clone EM081 (JN635746)
Iodamoeba sp. RL2 clone EM080 (JN635740)
Unidentified archamoebid sequence SW04 (OK483224) ☆
Mastigamoeba scholaia TEXEL (JX157654)
Mastigamoeba abducta 3MLMA (JX157633)
Mastigamoeba simplex CH2 (JX157659)
Mastigamoeba guttula HRADANAN (JX157638)
Mastigamoeba guttula LUH2NS4 (JX157649)
Mastigamoeba lenta VIT1AN (JX157656)
Mastigamoeba errans HRAANM (JX157637)
Mastigamoeba sp. IC-20121 (JX157664)
Mastigamoeba balamuthi (L23799)
Mastigamoeba punctachora SEB4 (JX157653)
Mastigamoebidae B
Mastigamoebidae A
56/0.56
100/1.0
100/1.0
100/1.0
100/1.0
82/1.0
65/ 0.95
61/0.84
77/0.96
60/0.88
100/1.0
64/0.90
76/0.89
100/1.0
100/1.0
100/1.0
90/1.0
0.10

283

Figure 1. Maximum likelihood phylogeny of *Endolimax* and relatives, reconstructed from an SSU rDNA alignment consisting of 21 taxa and 2067 positions. Maximum likelihood bootstrap values and Bayesian posterior probabilities are shown in that order on each bipartition. GenBank accession numbers are indicated in brackets. Sequences generated in this study are indicated with a star; sequences from sewage have the prefix SW.

290 **References:**

291 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment
292 trimming in large-scale phylogenetic analyses. Bioinformatics 25**(15):**1972–1973. doi:
293 10.1093/bioinformatics/btp348.

294 Cavalier-Smith T, Chao EEY, Oates B (2004) Molecular phylogeny of Amoebozoa and the
295 evolutionary significance of the unikont *Phalansterium*. Eur J Protistol **40:** 21–48
296 doi:10.1016/j.ejop.2003.10.001

297 Constenla M, Padrós F, Palenzuela O (2014) *Endolimax piscium* sp. nov. (Amoebozoa), causative
298 agent of systemic granulomatous disease of cultured sole, *Solea senegalensis* Kaup. J Fish Dis **37:**
299 229–240. doi:10.1111/jfd.12097

300 Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput.
301 Nucleic Acids Res **32:**1792–1797. doi:10.1093/nar/gkh340

302 Fitzgerald O, O'Farrell TT (1954) Chronic diarrhoea due to *Endolimax nana* infestation. Ir J Med
303 Sci **346:**467–468. doi:10.1007/BF02952047

304 Graczyk TK, Shiff CK, Tamang L, Munsaka F, Beitin AM, Moss WJ (2005) The association of
305 *Blastocystis hominis* and *Endolimax nana* with diarrheal stools in Zambian school-age children.
306 Parasitol Res **98:**38–43. doi:10.1007/s00436-005-0003-0

307 Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees.
308 Bioinformatics **17:**754–755. doi: 10.1093/bioinformatics/17.8.754

309 Ignacio CF, Silva MECD, Handam NB, Alencar MFL, Sotero-Martins A, Barata MML, Moraes
310 AHA Neto (2017) Socioenvironmental conditions and intestinal parasitic infections in Brazilian
311 urban slums: a cross-sectional study. Rev Inst Med Trop Sao Paulo **59:**e56. doi: 10.1590/S1678-
312 9946201759056.

313 Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D,
314 Burki F (2020) Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and
315 taxonomically resolve environmental diversity. Mol Ecol Resour **20:**429–443. doi:10.1111/1755-
316 0998.13117

317 Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:
318 Improvements in performance and usability. Mol Biol Evol **30:**772–780. doi:
319 10.1093/molbev/mst010.

320 Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019) RAxML-NG: A fast, scalable and
321 user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics **35:**4453–4455.
322 doi: 10.1093/bioinformatics/btz305.

323 Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics
324 analysis across computing platforms. Mol Biol Evol **35:**1547–1549. doi:10.1093/molbev/msy096

325 Poulsen CS, Stensvold CR (2016) Systematic review on *Endolimax nana*: A less well studied
326 intestinal ameba. Trop Parasitol **6:**8–29. doi:10.4103/2229-5070.175077

327 Ptáčková E, Kostygov AY, Chistyakova LV, Falteisek L, Frolov AO, Patterson DJ, Walker G,
328 Cepicka I (2013) Evolution of Archamoebae: Morphological and molecular evidence for pelobionts
329 including *Rhizomastix*, *Entamoeba*, *Iodamoeba*, and *Endolimax*. Protist **164:**380–410.
330 doi:10.1016/j.protis.2012.11.005

331 Santos CB, Araújo KC, Jardim-Botelho A, Santos MB, Rodrigues A, Dolabella SS, Gurgel RQ
332 (2014) Diarrhea incidence and intestinal infections among rotavirus vaccinated infants from a poor
333 area in Brazil: A spatial analysis. BMC Public Health **14:**399. doi:10.1186/1471-2458-14-399

334 Shah M, Tan CB, Rajan D, Ahmed S, Subramani K, Rizvon K, Mustacchia P (2012) *Blastocystis*
335 *hominis* and *Endolimax nana* co-infection resulting in chronic diarrhea in an immunocompetent
336 male. Case Rep Gastroenterol **6:**358–364. doi:10.1159/000339205

337 Silberman JD, Clark CG, Diamond LS, Sogin ML (1999) Phylogeny of the genera *Entamoeba* and
338 *Endolimax* as deduced from small-subunit ribosomal RNA sequences. Mol Biol Evol **16:**1740–
339 1751.

340 Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. Methods Mol Biol **132:**115–
341 130. doi:10.1385/1-59259-192-2:115

342 Stauffer JQ, Levine WL (1974) Chronic diarrhea related to *Endolimax nana*: Response to treatment
343 with metronidazole. Am J Dig Dis **19:**59–63. doi:10.1007/BF01073354

344 Stensvold CR, Lebbad M, Clark CG (2012) Last of the human protists: The phylogeny and genetic
345 diversity of *Iodamoeba*. Mol Biol Evol **29:**39–42. doi: 10.1093/molbev/msr238.

346 Stensvold CR, Lebbad M, Hansen A, Beser J, Belkessa S, O'Brien Andersen L, Clark CG (2020)
347 Differentiation of *Blastocystis* and parasitic archamoebids encountered in untreated wastewater
348 samples by amplicon-based next-generation sequencing. Parasite Epidemiol Control **9:**e00131.
349 doi:10.1016/j.parepi.2019.e00131

350 Stensvold CR, Lebbad M, Victory EL, Verweij JJ, Tannich E, Alfellani M, Legarraga P, Clark CG
351 (2011) Increased sampling reveals novel lineages of *Entamoeba*: Consequences of genetic diversity
352 and host specificity for taxonomy and molecular detection. Protist **162:**525–541. doi:
353 10.1016/j.protis.2010.11.002.

354 Taliaferro WH, Becker ER (1922) The human intestinal amoeba, *Iodamoeba williamsi*, and its cysts
355 (Iodine cysts). Am J Epidemiol **2:**188–207.

356 Yoshida N, Kobayashi S, Suzuki J, Azuma Y, Kobayashi-Ogata N, Kartikasari DP, Yanagawa Y,
357 Iwata S (2019) Growth-promoting effects of the hydrogen-sulfide compounds produced by
358 *Desulfovibrio desulfuricans* subsp. desulfuricans co-cultured with *Escherichia coli* (DH5alpha) on
359 the growth of *Entamoeba* and *Endolimax* species isolates from swine. Biosci Trends, **13**:402–410.
360 doi:10.5582/bst.2019.01233

361 Zadrobílková E, Walker G, Čepička I (2015) Morphological and molecular evidence support a close
362 relationship between the free-living archamoebae *Mastigella* and *Pelomyxa*. Protist **166:**14–41.
363 doi:10.1016/j.protis.2014.11.003

364