Parasite Epidemiology and Control 9 (2020) e00131

Contents lists available at ScienceDirect

Parasite Epidemiology and Control

journal homepage: www.elsevier.com/locate/parepi

Differentiation of *Blastocystis* and parasitic archamoebids encountered in untreated wastewater samples by amplicon-based next-generation sequencing

Christen Rune Stensvold [a,*], Marianne Lebbad [b], Anette Hansen [b], Jessica Beser [b], Salem Belkessa [a,c,d], Lee O'Brien Andersen [a], C. Graham Clark [e]

[a] Laboratory of Parasitology, Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Artillerivej 5, DK–2300 Copenhagen S, Denmark
[b] Department of Microbiology, Public Health Agency of Sweden, SE-171 82 Solna, Sweden
[c] Department of Biochemistry and Microbiology, Faculty of Biological and Agronomic Sciences, Mouloud Mammeri University of Tizi Ouzou, 15000 Tizi Ouzou, Algeria
[d] Department of Natural and Life Sciences, Faculty of Exact Sciences and Natural and Life Sciences, Mohamed Khider University of Biskra, 07000 Biskra, Algeria
[e] Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

ARTICLE INFO

Article history:
Received 20 September 2019
Received in revised form 6 November 2019
Accepted 16 December 2019
Available online xxxx

Keywords:
Sewage
Topic:
Wastewater
Scandinavia
Topic:
Genetic diversity
Parasite
Amoebae
Ameba
Topic:
Archamoebae
Stramenopiles
NGS
Survey
Phylogeny
Ribosomal lineage

ABSTRACT

Background: Application of next-generation sequencing (NGS) to genomic DNA extracted from sewage offers a unique and cost-effective opportunity to study the genetic diversity of intestinal parasites. In this study, we used amplicon-based NGS to reveal and differentiate several common luminal intestinal parasitic protists, specifically *Entamoeba*, *Endolimax*, *Iodamoeba*, and *Blastocystis*, in sewage samples from Swedish treatment plants.

Materials and methods: Influent sewage samples were subject to gradient centrifugation, DNA extraction and PCR-based amplification using three primer pairs designed for amplification of eukaryotic nuclear 18S ribosomal DNA. PCR products were sequenced using ILLUMINA® technology, and resulting sequences were annotated to species and subtype level using the in-house BION software, sequence clustering, and phylogenetic analysis.

Results: A total of 26 samples from eight treatment plants in central/southern Sweden were analysed. *Blastocystis* sp. and *Entamoeba moshkovskii* were detected in all samples, and most samples ($n = 20$) were positive for *Entamoeba coli*. Moreover, we detected *Entamoeba histolytica*, *Entamoeba dispar*, *Entamoeba hartmanni*, *Endolimax nana*, and *Iodamoeba bütschlii* in 1, 11, 4, 10, and 7 samples, respectively. The level of genetic divergence observed within *E. nana* and *E. moshkovskii* was 20.2% and 7.7%, respectively, across the ~400-bp region studied, and two clades of *E. moshkovskii* were found. As expected, *Blastocystis* sp. subtypes 1–4 were present in almost all samples; however, ST8 was present in 10 samples and was the only subtype not commonly found in humans that was present in multiple samples.

Conclusions: *Entamoeba* and *Blastocystis* were identified as universal members of the "sewage microbiome". *Blastocystis* sp. ST8, which has been rarely reported in humans, was a very common finding, indicating that a hitherto unidentified but common host of ST8 contributed to the sewage influent. The study also provided substantial new insight into the intra-generic diversity of *Entamoeba* and *Endolimax*.

* Corresponding author.
E-mail address: run@ssi.dk. (C.R. Stensvold).

https://doi.org/10.1016/j.parepi.2019.e00131

1. Introduction

Over the past few decades, molecular methods have enabled us to optimise detection and differentiation of intestinal parasites in human faecal material (Verweij and Stensvold, 2014; Stensvold et al., 2011a; van Lieshout and Verweij, 2010; Stensvold and Nielsen, 2012). Studies of the genetic diversity within these parasites have been increasing due to the recognition that differences in genetic make-up may reflect variable clinical/public health significance of these parasites, including host specificity and clinical impact of infection (Stensvold et al., 2011a; Stensvold, 2019; Stensvold et al., 2009a; Verweij et al., 2001).

For common luminal intestinal parasitic protists (CLIPPs) such as *Blastocystis* sp. and the archamoebids, reports of novel ribosomal lineages are emerging regularly (Stensvold, 2019; Verweij et al., 2001; Jacob et al., 2016; Elsheikha et al., 2018). CLIPPS, a term recently coined to address all single-celled intestinal parasites usually confined to the intestinal lumen only (Stensvold, 2019), are transmitted faecal-orally and have environmentally resistant stages that help them survive outside the host. Some of these lineages may represent species complexes. In this context, a species complex is a cluster of nuclear ribosomal RNA gene sequences that exhibit substantial levels of genetic divergence, on a scale usually observed between species in that particular genus. Examples include *Entamoeba coli*, which is known to have two lineages with 18S ribosomal RNA gene sequences differing by up to at least 13% (Stensvold et al., 2011b), and *Iodamoeba bütschlii*, which is known also to comprise at least two ribosomal lineages with sequences that differ by about 30% (Stensvold et al., 2012a). Preliminary data indicate a similar situation is present in *Endolimax nana* (Poulsen and Stensvold, 2016).

Untreated sewage samples taken from sewage treatment plants represent pools of human excreta, and are valuable resources to explore the geographical distribution of and genetic diversity within CLIPPs. Application of next-generation sequencing (NGS) to DNA from sewage offers a unique and cost-effective opportunity to study the genetic diversity of CLIPPs, and, in the long run, may enable efficient and standardised surveillance of parasitic organisms.

We recently developed a NGS-based platform for detection and differentiation of prokaryotic and eukaryotic nuclear ribosomal DNA, which has already been used in a few studies (Krogsgaard et al., 2018; Ring et al., 2017; Lear et al., 2019). In the present study, we used this method to reveal and differentiate taxa within several CLIPP genera in samples from Swedish sewage treatment plants.

2. Materials and methods

2.1. Samples and DNA extraction

Genomic DNAs were available from samples collected from eight different wastewater treatment plants (also referred to as sampling sites in the following) in southern Sweden from May 2014–December 2014. These wastewater samples were the same as those used in a previous limited study (Stensvold et al., 2018), with an additional five samples added.

Each wastewater sample was shaken to ensure a homogenous sample, and 200 mL was filtered into a 200-mL Falcon bottle through four layers of gauze and centrifuged at 1700 ×*g* (3,000 rpm) for 10 min. The supernatants were decanted, leaving approximately 10 mL of each sample, which was transferred to a 10-mL centrifuge tube and centrifuged at 1700 ×*g* for 10 min, after which the supernatants were discarded. The sediments were washed again with Milli-Q water for 10 min at 1700 ×*g* and supernatants discarded. The resulting sediments were processed by a sucrose gradient to enrich these for spores and (oo)cysts, and DNA was extracted from separated cysts as previously described (Lebbad et al., 2008).

2.2. PCR and sequencing

18S rRNA gene segments were amplified by PCR using a total of three primer sets as previously described (Krogsgaard et al., 2018; Ring et al., 2017; Lear et al., 2019); G3F1/G3R1 (5′-GCCAGCAGCCGCGGTAATTC-3′/5′-ACATTCTTGGCAAATGCTTTCGCAG-3′), G4F3/G4R3 (5′-CAGCCGCGGTAATTCCAGCTC-3′/5′-GGTGGTGCCCTTCCGTCAAT-3′) and G6F1/G6R1 (5′-TGGAGGGCAAGTCT GGTGCC-3′/5′-ACGGTATCTGATCGTCTTCGATCCC-3′). G3 and G6 primers target the hyper-variable regions V3-V4 of the 18S rRNA gene, and G4 primers target V3-V5. 18S rDNA was amplified in a 25 μL volume, using the REDExtract-N-Amp PCR ReadyMix (Sigma-Aldrich, St Louis, MO, USA) with 0.4 μM of each primer and 2 μL of template. PCR cycling conditions included initial denaturation at 95 °C for 3 min, 20 cycles of 95 °C for 1 min, 60 °C for 1 min, and 72 °C for 30 s, and a final elongation at 72 °C for 4 min. PCR products were prepared for sequencing by a second PCR (adaptor PCR), using the same PCR protocol as described above. This PCR attached an adaptor A, an index barcode, and a forward sequencing primer site to the 5′ end of the amplicons and an adaptor B, an index barcode, and a reverse sequencing primer site (RSP) to the 3′ end of the amplicons. DNA was quantified using the Quant-ITTM dsDNA High Sensitive Assay Kit (Thermo Fisher Scientific), and PCR2 products were pooled in equimolar amounts across samples. Undesirable DNA amplicons were removed from the pooled amplicon library by Agencourt AMPure XP bead (Beckman Coulter) purification. The resulting AMPure beads-purified pooled amplicon library was diluted to its final concentration of 11.5 pM DNA in 0.001 N NaOH and used for sequencing on the Illumina MiSeq desktop sequencer (Illumina Inc., San Diego, CA 29122, USA). The library was sequenced with the 500 rxn MiSeq Reagent Kit V2 in a 2 × 250 nt setup (Illumina Inc., San Diego, CA 29122, USA).

2.3. Sequence-based identification of parasite taxa

Data were analysed with BION (http://box.com/bion), a recently developed pipeline that accepts raw sequence data and includes steps for de-multiplexing, primer-extraction, sampling, sequence- and quality-based trimming and filtering, de-replication, clustering, chimera-checking, reference data similarities and taxonomic mapping and formatting. Non-overlapping paired reads are allowed for analysis. The filtered abundance matrix was analysed in R using the phyloseq (v.1.26.0) and ggplot (v.3.1.0) packages.

Sequences identified as being from Archamoebae and *Blastocystis* by BION were extracted from the html browser of BION for each of the samples and managed individually. The fasta-formatted sequences were clustered using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/; CLUSTAL O(1.2.4) multiple sequence alignment). Sequence alignment results were viewed with 'Jalview', and the consensus sequence of each cluster was analysed and confirmed to genus level by e.g., BLAST.

For *E. nana* and *Entamoeba moshkovskii* data, consensus sequences so obtained were aligned with reference sequences using the multiple sequence alignment option 'Muscle' in MEGA 6.0 (Tamura et al., 2013). Alignments were edited manually so that ambiguously aligned bases were deleted and the edited alignment was used for phylogenetic analysis using the Neighbor-Joining algorithm as implemented in MEGA 6.0.

Four consensus sequences for *E. nana* and two for *E. moshkovskii* were deposited in GenBank (MN508053–MN508056 and MN498050–MN498051, respectively).

For *Blastocystis* data, consensus sequences were queried by the BLAST algorithm in the NCBI Database and annotated to subtype level based on the level of similarity to reference sequences.

3. Results

The number of archamoebid and *Blastocystis* sequences produced by the G3 primers was negligible compared with those produced by the G4 and G6 primers, and so only G4 and G6 sequence outputs were used. G6 sequences were typically shorter than the G4 sequences, and in contrast to most G4 sequences did not have gap in the middle when assembled.

Archamoebid sequences were identified in all samples (Table 1). *Entamoeba moshkovskii* was present in all samples, while *E. coli* was present in 20 (77%) and *Entamoeba dispar*, *Entamoeba hartmanni*, and *Entamoeba histolytica* were detected in 11, 4, and 1 of the 26 samples, respectively. Finally, *E. nana* and *I. bütschlii* were found in 10 and 7 samples, respectively.

Table 1
Distribution of species and subtypes of Archamoebae and *Blastocystis*, in the 26 wastewater samples sourced from seven different Swedish water treatment plants (*E. histolytica* = *Entamoeba histolytica*; *E. dispar* = *Entamoeba dispar*; *E. coli* ST1 = *Entamoeba coli* ST1; *E. coli* ST2 = *Entamoeba coli* ST2; *E. moshkovskii* Clade 1 = *Entamoeba moshkovskii* Clade 1; *E. moshkovskii* Clade 2 = *Entamoeba moshkovskii* Clade 2; *E. nana* = *Endolimax nana*; *I. bütschlii* = *Iodamoeba bütschlii*). See text for further details.

Sample ID	Sampling site	*E. histolytica*	*E. dispar*	*E. hartmanni*	*E. coli* ST1	*E. coli* ST2	*E. moshkovskii* Clade 1	*E. moshkovskii* Clade 2	*E. nana*	*I. bütschlii* RL1	*Blastocystis*					
											ST1	ST2	ST3	ST4	ST8	ST10
A46	1	ND	ND	ND	+	ND	ND	+	ND	ND	+	+	+	+	+	ND
A47	2	ND	ND	ND	ND	ND	+	+	ND	ND	+	+	+	+	+	ND
A48	3	ND	ND	ND	+	ND	+	+	+	ND	+	+	+	+	ND	ND
A52	4	ND	ND	ND	+	ND	+	+	+	ND	+	+	+	+	ND	ND
A53	5	ND	ND	ND	+	ND	ND	+	+	ND	+	+	+	+	ND	ND
A55	5	ND	ND	ND	ND	ND	ND	+	+	ND	+	+	+	+	ND	ND
A56	6	ND	+	ND	+	ND	ND	+	+	ND	+	+	+	+	+	ND
A59	6	ND	+	ND	ND	+	ND	+	ND	ND	+	+	+	+	+	ND
A63	4	ND	ND	ND	+	ND	ND	+	+	ND	+	+	ND	+	ND	ND
A64	6	ND	+	+	+	ND	ND	+	ND	+	+	+	+	+	+	ND
A65	4	ND	ND	ND	+	ND	+	+	ND	ND	+	+	+	+	ND	ND
A67	5	ND	ND	ND	+	+	ND	+	ND	+	+	+	+	+	+	ND
A69	6	ND	+	ND	+	ND	ND	+	ND	ND	+	+	+	+	ND	ND
A70	4	ND	ND	ND	ND	ND	ND	+	ND	ND	ND	+	+	+	ND	ND
A72	5	ND	+	ND	+	ND	ND	+	+	+	+	+	+	+	ND	ND
A77	4	ND	ND	+	+	+	+	+	ND	ND	+	+	+	+	ND	ND
A79	5	ND	+	+	+	ND	ND	+	ND	ND	+	+	+	+	+	ND
A81	1	ND	+	ND	+	+	ND	+	ND	+	+	+	+	+	+	ND
A82	2	+	+	ND	+	ND	ND	+	+	+	+	+	+	+	ND	ND
A83	7	ND	ND	ND	+	ND	+	+	ND	ND	+	+	+	+	ND	ND
A88	1	ND	ND	ND	ND	ND	+	+	ND	ND	+	+	+	+	ND	ND
A89	2	ND	+	ND	+	ND	+	+	ND	+	+	+	+	+	ND	ND
A95	7	ND	ND	ND	ND	ND	+	+	ND	ND	+	+	+	+	ND	+
A96	8	ND	ND	ND	ND	ND	+	+	ND	ND	+	+	+	+	ND	ND
A100	8	ND	+	+	+	+	+	+	+	+	+	+	+	+	+	ND
A101	4	ND	+	ND	+	+	+	+	+	ND	+	+	+	+	+	ND

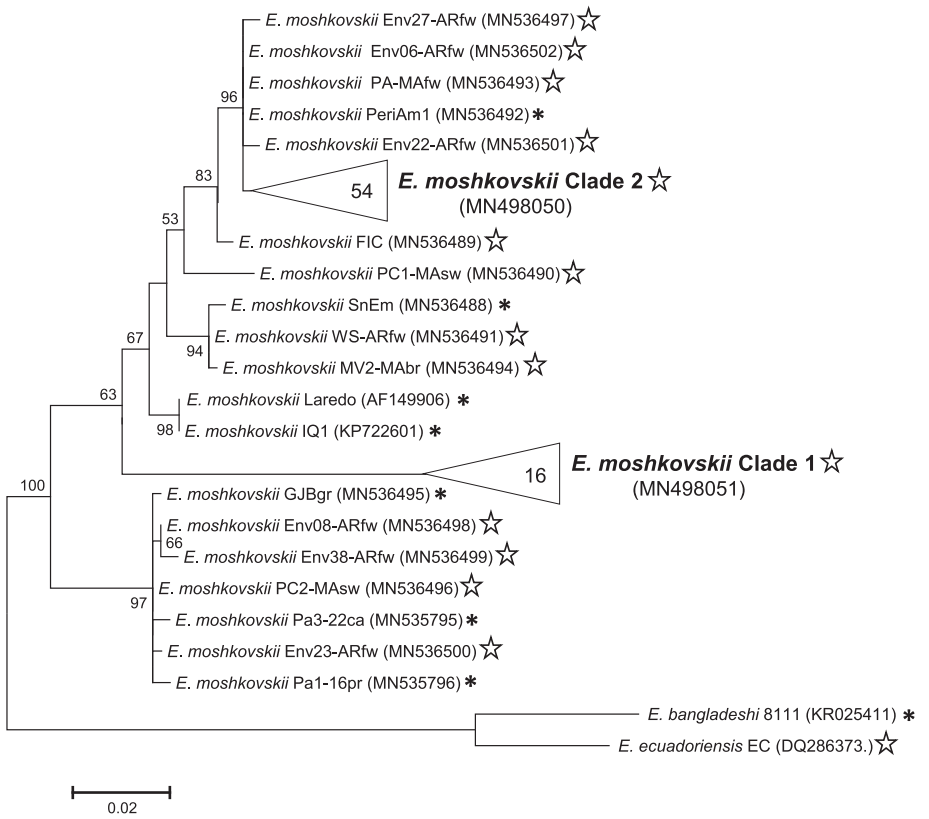

Fig. 1. Phylogenetic analysis of *Entamoeba moshkovskii* sequences. This Neighbor–joining tree was produced using the Tajima-Nei Model, as selected by Model Test implemented in MEGA 6.0 (Tamura et al., 2013). Bootstrap proportions are shown where a node is supported by over 50% of the 1000 replicates. *E. moshkovskii* sequences identified by stars have an environmental origin while those identified by asterisks are from animal hosts, including humans.

Phylogenetic analysis using *E. moshkovskii* sequences obtained by the G4 and G6 primer pairs resulted in the formation of two clades, Clade 1 and Clade 2 (Fig. 1). Clade 1 was found in 12 samples (46%), while Clade 2 was present in all samples. *Entamoeba moshkovskii* clade consensus sequences differed by 6.7%, indicating substantial intra-species diversity.

Both subtypes of *E. coli* acknowledged to date (Stensvold et al., 2011b) were found. *Entamoeba coli* ST1 was most common and found in 19/26 samples (73%), while *E. coli* ST2 was found in six samples (23%). *Entamoeba coli* ST1 was found at all sampling sites and *E. coli* ST2 at 5/8 sampling sites.

Entamoeba dispar was found at 6/8 sampling sites, and at one of these, it was observed in each of the four samples taken. In contrast, *E. histolytica* was found at only one of the treatment plants and only in one sample.

Extensive genetic diversity has been identified previously in both *Endolimax* and *Iodamoeba*. In the present study, only one of the two acknowledged ribosomal lineages of *Iodamoeba* was detected, namely RL1. With regard to *Endolimax*, phylogenetic analysis resulted in the recovery of no fewer than four clades; in sample A56, all four clades were found. Half of the sequences fell into a clade that included an as yet unpublished *Endolimax* sequence from a human (H80028).

Blastocystis subtypes 1–4 were found in all but two samples: in sample A63, no *Blastocystis* ST3 was found, while in A70, no *Blastocystis* ST1 was found. *Blastocystis* ST8 was found at 6/8 sampling sites, and in a total of 10/26 samples (38%). Finally, one sample was positive for *Blastocystis* ST10.

For quite a few samples, a genus or species (complex) was represented by multiple distinct lineages (e.g., four different sequences of *Endolimax* from a single sample mentioned above).

4. Discussion

Surveillance of parasites in raw and treated wastewater should employ robust, cost-effective, and precise molecular tools for detection and differentiation. Nevertheless, large-scale analysis of parasites in wastewater samples is still a costly and laborious activity. Studies relying on PCR followed by cloning and Sanger sequencing are also time consuming and expensive. Metagenomic analysis of microbial communities in faecal material, including sewage, has been undertaken previously, but most frequently with a view to surveying antimicrobial resistance genes rather than exploring taxonomic diversity. Moreover, most studies have focused on bacteria, and only a few studies have addressed eukaryotic populations in wastewater (Berglund et al., 2017; Li et al., 2012; Kitajima et al., 2014; Zahedi et al., 2019), typically using real-time PCR for targeted detection especially of *Giardia* and

Cryptosporidium. Finally, compared with bacteria, reference data for some parasitic eukaryotic organisms such as the Archamoebae are limited (Stensvold, 2019). Nevertheless, recent studies of the genetic diversity within CLIPPs have led to the recognition of dozens of species and ribosomal lineages within *Blastocystis*, *Entamoeba*, *Iodamoeba*, and *Endolimax*. For instance, at least 17 ribosomal lineages of *Blastocystis* (the so-called subtypes, which are arguably species) have been identified in mammals and birds (Alfellani et al., 2013a). In addition, what are known as *E. coli*, *I. bütschlii*, and *E. nana* based on morphology actually reflect species complexes rather than single species (Stensvold et al., 2011b; Stensvold et al., 2012a). For example, *E. coli* ST1 (AF149915) and *E. coli* ST2 (AF149914) only share 86.7% identity across the entire 18S rRNA gene.

Unsurprisingly, parasites identified in the sewage samples analysed in the present study largely reflected parasites found in human stool. After the introduction of DNA-based methods to detect and differentiate *Blastocystis*, the estimated prevalence of this parasite increased tremendously (Stensvold et al., 2012b; El Safadi et al., 2014; Poulsen et al., 2016; Oliveira-Arbex et al., 2018). Numerous surveys of the subtypes colonising humans have been undertaken, and it was recently estimated that more than 95% of all human *Blastocystis* carriage is attributed to subtypes 1–4 (Alfellani et al., 2013b), all four of which were found in almost all our sewage samples. While subtypes 1–4 are the *Blastocystis* subtypes commonly seen in humans, these subtypes have also been found in non-human hosts, and so the question remains to what extent the *Blastocystis* detected in these samples might originate from hosts other than humans. Recent 18S rRNA gene allele analyses have detected cryptic host specificity in these subtypes (Alfellani et al., 2013c; Vaisusuk et al., 2018), such that, e.g., ST3 strains isolated from humans are largely distinct from ST3 strains isolated from non-human primates. Since data for the same gene region are not produced by the methods used in this study, unfortunately it is not possible to speculate on the host origin of these subtypes. However, since the prevalence of parasites commonly found in livestock and wild animals, such as *Blastocystis* ST10, was low (only found in 1/26 samples), we believe that the *Blastocystis* ST1–ST4 identified here are mostly of human origin.

Given the distribution of parasite taxa found in this study and the broad specificity of the primers used, it is likely that only the most abundant species/genera were detected. We know from our previous study (Stensvold et al., 2018) that three of the samples included in the present study were positive for *Entamoeba polecki* using single-round conventional PCR; this species, however, did not show up in any reads from the present study. This circumstance also implies that since *Blastocystis* ST8 was detected so frequently, it must be relatively abundant in the sewage in which it was found. However, there are only single reports of ST8 in Sweden (Forsell et al., 2017) and Denmark (Stensvold et al., 2008), in imported human cases. Due to the fact that *Blastocystis* ST8 has been so rarely reported in humans, it was a surprise to detect this particular subtype in so many of the samples. This raises the question of whether ST8 in the sewage is a result of human excretion or as a result of excretion from non-human hosts, such as rodents, birds or arthropods. The latter groups have had limited sampling, so there is still limited knowledge of the subtypes colonising them. To date, ST8 has only been identified with regularity in some New World monkeys (Alfellani et al., 2013c). There are only sporadic reports of ST8 in humans (Forsell et al., 2017; Stensvold et al., 2008; Barbosa et al., 2018; Mattiucci et al., 2016; Stensvold et al., 2009b; Seguí et al., 2018), non-human primates (Alfellani et al., 2013c; Helenbrook et al., 2015; Scicluna et al., 2006), birds (Abe et al., 2003), other hosts (Noradilah et al., 2017a) and water (Noradilah et al., 2017b). Although ST8 has been reported once in bird droppings (Abe et al., 2003) the subtypes usually reported in birds are ST6 (Jacob et al., 2016) and ST7 (Stensvold et al., 2009b; Greige et al., 2018; Deng et al., 2019). As these subtypes were not detected in this study, a limited contribution of birds to the sewage samples is suggested. Similarly, *Blastocystis* ST10, a subtype commonly found in ruminants, particularly in cattle (Stensvold et al., 2009b; Greige et al., 2019; Zhu et al., 2017), is the only strictly 'non-human subtype' found in the material. It was found in only one of the samples, suggesting that the contribution of stool from larger animals to the sewage was limited. The host of the *Blastocystis* ST8 in our samples remains unknown at present.

The finding of *E. coli* ST2 in so many of the samples may provide another useful clue to contamination of the sewage by non-human faeces. *Entamoeba coli* has traditionally been regarded as a parasite of humans. Although there are no reports of *E. coli* in rats, a DNA sequence in the NCBI Database from a rat attributed to *Entamoeba muris* (FN396613) is in fact *E. coli* ST2 (Jacob et al., 2016). This suggests that rats can be hosts of *E. coli* ST2. However, other non-primate hosts are also known to host *E. coli* ST2, namely the rhea (FN396614) and chinchilla (FR686439), suggesting a wider and incompletely known host range. Octonucleate-cyst producing amoebae have been reported in birds – *E. gallinarum* in chickens, ducks and geese, for example. However, although it is tempting to speculate that rodents and birds may be common hosts of *E. coli* ST2, no DNA sequence data are available for *E. gallinarum* and until the genetic diversity and host specificity of this *Entamoeba* group is better mapped, the host origins of these protists in our samples remain obscure. With regard to *E. coli* ST1, DNA sequences belonging to this clade to date belong exclusively to humans and non-human primates.

Another common finding was *E. moshkovskii*, which is considered largely free-living but with the potential to infect humans, cattle and turtles (Jacob et al., 2016). In 1961, Levine considered *E. moshkovskii* "not as a parasite of animals, but of the municipal digestive tract" and as a parasite that "occurs in sewage" (Levine, 1961). Our findings strongly support the idea of *E. moshkovskii* as a "sewage parasite". Nevertheless, there were two major variants detected in this study. One of the variants (Clade 2, 55 sequences) was a consistent finding across all samples. It shares high similarity with a sequence from a Canadian sewage *E. moshkovskii* (FIC). The other clade (Clade 1, 16 sequences) was present in almost half of the samples. It has only limited similarity to previously identified *E. moshkovskii* sequences. A third clade, which includes the most widely studied *E. moshkovskii* strain (Laredo) (Wilson et al., 2019) was not found in our sewage samples, and neither was a fourth clade comprising mostly environmental isolates.

We detected vast genetic variation within *Endolimax*, with no fewer than four clades detected across the eight *Endolimax*-positive samples included in Fig. 2. Although insight into the genetic diversity within *Endolimax* is still limited, these data appear

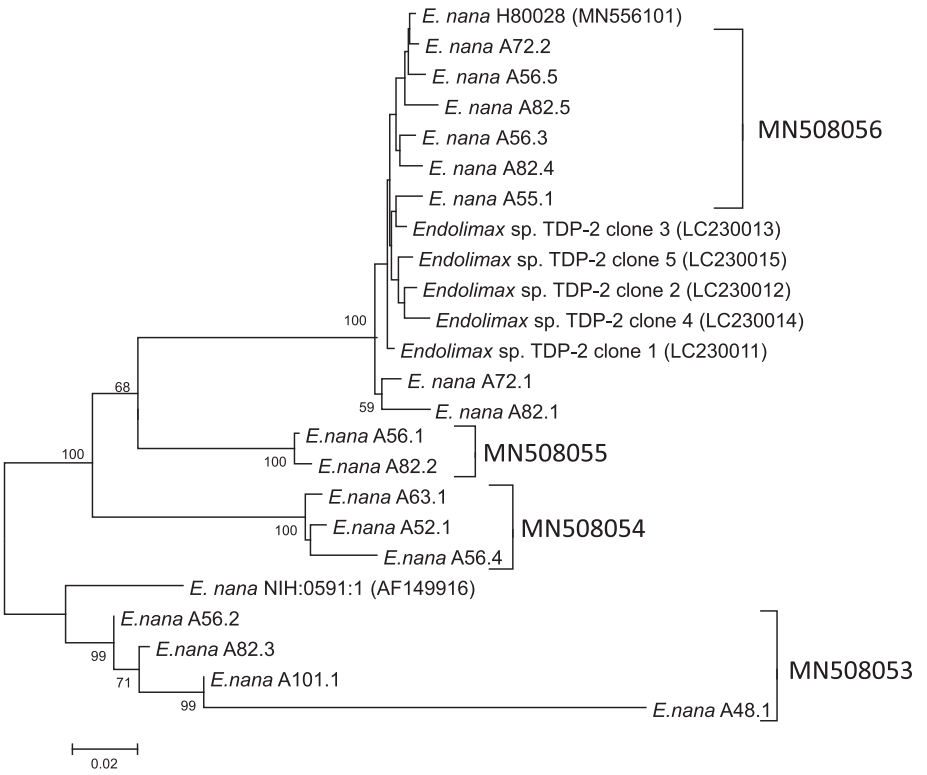

Fig. 2. Phylogenetic analysis of *Endolimax* sequences obtained by the G6 primers. A total of 17 sequences for 8 samples were aligned with two reference sequences. This Neighbor–joining tree was produced using the Tajima-Nei Model with rate variation modelled with a Gamma distribution (shape parameter = 0.42), as selected by Model Test implemented in MEGA 6.0. Bootstrap proportions are shown where a node is supported by over 50% of the 1000 replicates.

to confirm that, like *I. bütschlii*, parasites referred to as *E. nana* based on morphology actually reflect a species complex. This knowledge of course has consequences for studies seeking to investigate the role of *Endolimax* in health and disease, including the epidemiology of the parasite.

In this study, DNA was extracted only after isolation of cysts by gradient centrifugation. This means, that the DNA used may not yield the same results as DNA extracted directly from the sewage material. Importantly, the amount of DNA from trophozoites present in the sample material would be very limited. This might explain why we did not detect *Dientamoeba fragilis*, which is a parasite that is very common in Sweden and the rest of Scandinavia. Whether *D. fragilis* produces cysts is a matter of current debate (Clark et al., 2014; Munasinghe et al., 2013), but if such cysts existed, DNA from such cysts would likely be present in the extracted DNA. However, we also know that the sensitivity of our platform in terms of detecting flagellate DNA is low, so *Giardia* and *Dientamoeba* are detected only when tested samples are highly positive as evidenced by e.g. real-time PCR (unpublished observations).

Given the complexity of sewage material, the present dataset should not be seen as an exhaustive analysis of the *Blastocystis* and Archamoebae present in the samples, due to template competition for primers. Nevertheless, the present dataset provides insight into the diversity of *Blastocystis* and Archamoebae present in sewage material in Sweden, and may be useful to inform future studies on genetic diversity within the genera detected and also - at least to some extent - on host specificity. Moreover, the diversity captured by this dataset should reflect the species circulating in the population of middle and southern Sweden.

Acknowledgements

We thank the technical staff at Laboratory of Parasitology, Statens Serum Institut. We would also like to thank Dr. Jeff Silberman for making available his unpublished 18S rRNA sequences from *E. moshkovskii* before release in GenBank.

References

Abe, N., et al., 2003. Molecular characterization of *Blastocystis* isolates from birds by PCR with diagnostic primers and restriction fragment length polymorphism analysis of the small subunit ribosomal RNA gene. Parasitol. Res. 89 (5), 393–396.
Alfellani, M.A., et al., 2013a. Genetic diversity of *Blastocystis* in livestock and zoo animals. Protist 164 (4), 497–509.
Alfellani, M.A., et al., 2013b. Variable geographic distribution of *Blastocystis* subtypes and its potential implications. Acta Trop. 126 (1), 11–18.
Alfellani, M.A., et al., 2013c. Diversity and distribution of *Blastocystis* sp. subtypes in non-human primates. Parasitology 1–6.

Barbosa, C.V., et al., 2018. Intestinal parasite infections in a rural community of Rio de Janeiro (Brazil): prevalence and genetic diversity of *Blastocystis* subtypes. PLoS One 13 (3), e0193860.
Berglund, B., et al., 2017. Occurrence and removal efficiency of parasitic protozoa in Swedish wastewater treatment plants. Sci. Total Environ. 598, 821–827.
Clark, C.G., et al., 2014. Transmission of *Dientamoeba fragilis*: pinworm or cysts? Trends Parasitol. 30 (3), 136–140.
Deng, L., et al., 2019. First report of *Blastocystis* in giant pandas, red pandas, and various bird species in Sichuan province, southwestern China. Int J Parasitol Parasites Wildl 9, 298–304.
El Safadi, D., et al., 2014. Children of Senegal River Basin show the highest prevalence of *Blastocystis* sp. ever observed worldwide. BMC Infect. Dis. 14, 164. https://doi.org/10.1186/1471-2334-14-164.
Elsheikha, H.M., et al., 2018. Novel *Entamoeba* findings in nonhuman primates. Trends Parasitol. 34 (4), 283–294.
Forsell, J., et al., 2017. The relation between *Blastocystis* and the intestinal microbiota in Swedish travellers. BMC Microbiol. 17 (1), 231.
Greige, S., et al., 2018. Prevalence and subtype distribution of *Blastocystis* sp. isolates from poultry in Lebanon and evidence of zoonotic potential. Parasit. Vectors 11 (1), 389. https://doi.org/10.1186/s13071-018-2975-5.
Greige, S., et al., 2019. First report on the prevalence and subtype distribution of *Blastocystis* sp. in dairy cattle in Lebanon and assessment of zoonotic transmission. Acta Trop. 194, 23–29.
Helenbrook, W.D., et al., 2015. Characterization of *Blastocystis* species infection in humans and mantled howler monkeys, *Alouatta palliata aequatorialis*, living in close proximity to one another. Parasitol. Res. 114 (7), 2517–2525.
Jacob, A.S., et al., 2016. Expanding the *Entamoeba* universe: new hosts yield novel ribosomal lineages. J. Eukaryot. Microbiol. 63 (1), 69–78.
Kitajima, M., et al., 2014. Occurrence of *Cryptosporidium*, *Giardia*, and *Cyclospora* in influent and effluent water at wastewater treatment plants in Arizona. Sci. Total Environ. 484, 129–136.
Krogsgaard, L.R., et al., 2018. Characteristics of the bacterial microbiome in association with common intestinal parasites in irritable bowel syndrome. Clin. Transl. Gastroenterol. 9 (6), 161.
Lear, R., et al., 2019. Tart cherry concentrate does not alter the gut microbiome, glycaemic control or systemic inflammation in a middle-aged population. Nutrients 11 (5). https://doi.org/10.3390/nu11051063 (pii: E1063).
Lebbad, M., et al., 2008. Dominance of *Giardia* assemblage B in León, Nicaragua. Acta Trop. 106 (1), 44–53.
Levine, N.D., 1961. Protozoan Parasites of Domestic Animals and of Man. Burgess Pub. Co., Minneapolis.
Li, N., et al., 2012. Molecular surveillance of *Cryptosporidium* spp., *Giardia duodenalis*, and *Enterocytozoon bieneusi* by genotyping and subtyping parasites in wastewater. PLoS Negl. Trop. Dis. 6 (9), e1809.
van Lieshout, L., Verweij, J.J., 2010. Newer diagnostic approaches to intestinal protozoa. Curr. Opin. Infect. Dis. 23 (5), 488–493.
Mattiucci, S., et al., 2016. Molecular epidemiology and genetic diversity of *Blastocystis* infection in humans in Italy. Epidemiol. Infect. 144 (3), 635–646.
Munasinghe, V.S., et al., 2013. Cyst formation and faecal-oral transmission of *Dientamoeba fragilis*—the missing link in the life cycle of an emerging pathogen. Int. J. Parasitol. 43 (11), 879–883.
Noradilah, S.A., et al., 2017a. Molecular epidemiology of *Blastocystis* sp in animals reared by the aborigines during wet and dry seasons in rural communities, Pahang, Malaysia. Southeast Asian J. Trop. Med. Public Health 48 (6), 1151–1160.
Noradilah, S.A., et al., 2017b. Molecular epidemiology of blastocystosis in Malaysia: does seasonal variation play an important role in determining the distribution and risk factors of *Blastocystis* subtype infections in the aboriginal community? Parasit. Vectors 10 (1), 360. https://doi.org/10.1186/s13071-017-2294-2.
Oliveira-Arbex, A.P., et al., 2018. *Blastocystis* genetic diversity among children of low-income daycare center in Southeastern Brazil. Infect. Genet. Evol. 57, 59–63.
Poulsen, C.S., Stensvold, C.R., 2016. Systematic review on *Endolimax nana*: a less well studied intestinal ameba. Trop Parasitol 6 (1), 8–29.
Poulsen, C.S., et al., 2016. Epidemiological aspects of *Blastocystis* colonization in children in Ilero, Nigeria. Am. J. Trop. Med. Hyg. 95 (1), 175–179.
Ring, H.C., et al., 2017. The follicular skin microbiome in patients with hidradenitis suppurativa and healthy controls. JAMA Dermatol. 153 (9), 897–905.
Scicluna, S.M., et al., 2006. DNA barcoding of *Blastocystis*. Protist 157 (1), 77–85.
Seguí, R., et al., 2018. Prevalence of intestinal parasites, with emphasis on the molecular epidemiology of *Giardia duodenalis* and *Blastocystis* sp., in the Paranaguá Bay, Brazil: a community survey. Parasit. Vectors 11 (1), 490.
Stensvold, C.R., 2019. Pinning down the role of common luminal intestinal parasitic protists in human health and disease - status and challenges. Parasitology 146 (6), 695–701.
Stensvold, C.R., Nielsen, H.V., 2012. Comparison of microscopy and PCR for detection of intestinal parasites in danish patients supports an incentive for molecular screening platforms. J. Clin. Microbiol. 50 (2), 540–541.
Stensvold, C.R., et al., 2008. Symptomatic infection with *Blastocystis* sp. subtype 8 successfully treated with trimethoprim-sulfamethoxazole. Ann. Trop. Med. Parasitol. 102 (3), 271–274.
Stensvold, C.R., et al., 2009a. Pursuing the clinical significance of *Blastocystis*—diagnostic limitations. Trends Parasitol. 25 (1), 23–29.
Stensvold, C.R., et al., 2009b. Subtype distribution of *Blastocystis* isolates from synanthropic and zoo animals and identification of a new subtype. Int. J. Parasitol. 39 (4), 473–479.
Stensvold, C.R., et al., 2011a. The impact of genetic diversity in protozoa on molecular diagnostics. Trends Parasitol. 27 (2), 53–58.
Stensvold, C.R., et al., 2011b. Increased sampling reveals novel lineages of *Entamoeba*: consequences of genetic diversity and host specificity for taxonomy and molecular detection. Protist 162 (3), 525–541.
Stensvold, C.R., et al., 2012a. Last of the human protists: the phylogeny and genetic diversity of *Iodamoeba*. Mol. Biol. Evol. 29 (1), 39–42.
Stensvold, C.R., et al., 2012b. Development and evaluation of a genus-specific, probe-based, internal process controlled real-time PCR assay for sensitive and specific detection of *Blastocystis*. J. Clin. Microbiol. 50 (6), 1847–1851.
Stensvold, C.R., et al., 2018. Evaluation of a PCR method for detection of *Entamoeba polecki*, with an overview of its molecular epidemiology. J. Clin. Microbiol. 56 (5). https://doi.org/10.1128/JCM.00154-18 (pii: e00154-18).
Tamura, K., et al., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30 (12), 2725–2729.
Vaisusuk, K., et al., 2018. *Blastocystis* subtypes detected in long-tailed macaques in Thailand-further evidence of cryptic host specificity. Acta Trop. 184, 78–82.
Verweij, J.J., Stensvold, C.R., 2014. Molecular testing for clinical diagnosis and epidemiological investigations of intestinal parasitic infections. Clin. Microbiol. Rev. 27 (2), 371–418.
Verweij, J.J., et al., 2001. Genetic variation among human isolates of uninucleated cyst-producing *Entamoeba* species. J. Clin. Microbiol. 39 (4), 1644–1646.
Wilson, I.W., et al., 2019. Genetic diversity and gene family expansions in members of the genus *Entamoeba*. Genome Biol. Evol. 11 (3), 688–705.
Zahedi, A., et al., 2019. Identification of eukaryotic microorganisms with 18S rRNA next-generation sequencing in wastewater treatment plants, with a more targeted NGS approach required for *Cryptosporidium* detection. Water Res. 158, 301–312.
Zhu, W., et al., 2017. First report of *Blastocystis* infections in cattle in China. Vet. Parasitol. 246, 38–42.