

Identifying group-specific primers for environmental *Heterolobosa* by high-throughput sequencing

Vincent Delafont¹  Anne Mercier,¹
Stéphanie Barrouilhet,² Marie-Laure Mollichella,¹
Pascaline Herbelin² and Yann Héchard^{1,*}

¹Université de Poitiers, Laboratoire Ecologie et Biologie des Interactions, CNRS UMR 7267, 1 rue Georges Bonnet, TSA 51106, 86073, Poitiers Cedex 9, France.

²EDF, Division Recherche et Développement, LNHE, 6 Quai Watier, 78401, Chatou, France.

Summary

Diversity of Heterolobosea (Excavata) in environments is poorly understood despite their ecological occurrence and health-associated risk, partly because this group tends to be under-covered by most universal eukaryotic primers used for sequencing. To overcome the limits of the traditional morphology-based biomonitoring, we constructed a primer database listing existing and newly designed specific primer pairs that have been evaluated for Heterolobosea 18S rRNA sequencing. *In silico* taxonomy performance against the current SILVA SSU database allowed the selection of primer pairs that were next evaluated on reference culture amoebal strains. Two primer pairs were retained for monitoring the diversity of Heterolobosea in freshwater environments, using high-throughput sequencing. Results showed that one of the newly designed primer pairs allowed species-level identification of most heterolobosean sequences. Such primer pair could enable informative, cultivation-free assays for characterizing heterolobosean populations in various environments.

Introduction

Heterolobosea is a distinctive lineage of amoebae, amoeboflagellates and flagellates within the eukaryotic

supergroup Excavata, encompassing about 150 species (Adl *et al.*, 2012; 2019; Pánek *et al.*, 2017). Almost all free-living heteroloboseans are largely soil and freshwater sediment aerobes, although the group also includes anaerobes, obligate halophiles and many thermophiles, revealing adaptation to various extreme environments (for review, see Pánek *et al.*, 2017). Most heteroloboseans adopt a three-phase asexual life cycle involving a trophozoite (amoeba) stage that can reversibly switch to either a flagellate or a cyst form. In exceptional cases, heteroloboseans can lack flagellate stage (e.g. *Vahlkampfia* spp.; Page, 1988) or trophozoite stage (e.g. *Stephanogon* spp.; Yubuki and Leander, 2008). The most studied heterolobosean members are from the *Naegleria* genus. As an example, *N. gruberi* has long been used as a cell biology model for studying flagellar apparatus development (Dingle and Fulton, 1966; Lee, 2010). *N. fowleri* is largely described as facultative human pathogen responsible for primary amoebic meningoencephalitis (Maciver *et al.*, 2020). While the disease is still very rare (a few hundred documented cases worldwide since 1968), the mortality is very high (>90%) (Gharpure *et al.*, 2020; Maciver *et al.*, 2020). *N. australiensis* was reported to be pathogenic for mice (De Jonckheere *et al.*, 1983), while *Naegleria* sp. was identified in brain infections of catfish species (Dykova and Lom, 2004). Strains of *Tetramitus ovis*, *Vahlkampfia* sp. and *Paravahlkampfia* sp. were isolated from keratitis patients, also suggesting pathogenicity (De Jonckheere and Brown, 2005; Visvesvara *et al.*, 2009). In addition, some heteroloboseans have been described as reservoir of various pathogenic bacteria such as *Legionella pneumophila* and *Mycobacterium* spp. (Newsome *et al.*, 1985; Ovrutsky *et al.*, 2013). Taken together, and taking into account potential effect of global warming on heteroloboseans, these studies support the importance of biomonitoring heteroloboseans in environment such as freshwaters used or exposed to various anthropogenic activities, in order to provide a better understanding of the Heterolobosea ecology, as well as the putative risks for human and animal health.

To this day, the diversity and ecology of Heterolobosea remain understudied and poorly described due to the limit of the traditional methods (Smirnov, 2003; Mrva and Garajová, 2018; Samba-Louaka *et al.*, 2019). As for other amoebal lineages, the biomonitoring techniques usually require a cultivation step, followed by either a morphological identification or DNA sequencing for

Received 15 September, 2021; revised 29 April, 2022; accepted 9 May, 2022.

For correspondence: E-mail yann.hechard@univ-poitiers.fr; vincent.delafont@univ-poitiers.fr; Tel. +33 (0)5 49 45 40 07; Fax +33 (0)5 49 45 40 07.

Microbial Biotechnology (2022) 15(9), 2476–2487
doi:10.1111/1751-7915.14080

Funding information

No funding information provided.

© 2022 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

identification. Indeed, the compulsory cultivation step has limitations, such as low sensitivity, lack of specificity, high response time and possibly an underestimation of populations due to the presence of viable but non-cultivable microorganisms.

High-throughput sequencing (HTS) of PCR-amplified loci, such as ribosomal RNA, has enabled the assessment of abundances, diversity and relatedness, of various microbial communities. While the HTS identification of microbial protists has significantly increased over the last decade (e.g. Oliverio *et al.*, 2020; Singer *et al.*, 2021), few studies are dedicated to the study of the Heterolobosea group. This is added to the fact that most universal eukaryotic primers do not provide a satisfactory coverage for this group, making it likely under-represented in global surveys (Aucher *et al.*, 2020; Vaulot *et al.*, 2021).

Only some primer pairs, mostly targeting part of the 18S rRNA gene, claimed to be specific to *Naegleria* and *Vahlkampfia* genera or to the *N. fowleri* species, were reported in the literature in relation to the Heterolobosea group (Sheehan *et al.*, 2003; Qvarnstrom *et al.*, 2006; Schild *et al.*, 2007; Le Calvez *et al.*, 2012; Scheikl *et al.*, 2014; Régoudis and Pélandakis, 2016; Taravaud *et al.*, 2018). For HTS of amplified loci, the primer selection, amplicon size and the amplification efficiency across the whole Heterolobosea taxonomic range are critical to obtain an accurate representation of such communities (Sze and Schloss, 2019). Therefore, it seems necessary to compare these existing primers and to search for new primers which will make it possible to precisely characterize environmental communities of Heterolobosea, using amplicon sequencing targeting part of the 18S rRNA gene.

In this study, we tested several primer pairs targeting Heterolobosea 18S rRNA gene sequences, consisting of existing and newly designed specific primer pairs. These primers were assessed *in silico* against an 18S rRNA gene database and experimentally by PCR amplification from DNA isolated for amoeba culture collection. Finally, two primer pairs were retained for assessing diversity of Heterolobosea in freshwater environments using HTS, via the Illumina MiSeq technology.

Results and discussion

In silico evaluation of previously and newly described primer pairs targeting Heterolobosea.

Primer selection is critical to obtain an accurate representation of Heterolobosea communities in an environmental sample. Only four primer pairs were collected from the literature based on their reported ability to specifically amplify Heterolobosea spp. and *Naegleria* spp. rRNA 18S sequences, in particular (Table 1,

Table S1). To increase our primer database, ten primer pairs were newly designed using PrimerDesign_M from a subset of rRNA 18S sequences from Heterolobosea class available in the current SILVA SSU 132 database (Table S2). These new primer pairs cover different variable regions of the Heterolobosea 18S rRNA gene sequence (Fig. S1). The performance of each previously and newly designed primer pairs was evaluated using TestPrime by running an *in silico* PCR on the SILVA database. Coverages expressed as a percentage, for taxonomic group in all the taxonomies detailed by SILVA, were reported in Table 1. The most appropriate primer pairs for our HTS application associate the highest coverage of Heterolobosea with the lowest amplification among other taxa and with an amplicon size ranging from 300 to 550 bp. From this *in silico* taxonomic coverage of the primer pairs against the SILVA SSU, Nae1/Nae2, Vahl730f_C/Test6R, Vahl730f_C/R1200, HF_592/HR_982, HF_1198/HR_1680 and HF_1202/HR_1680 were selected for experimental evaluations. For these six selected primer pairs, the coverage for the Eukaryota domain was low (< 3.4%) as well as the coverage for Amoebozoa and Opisthokonta classes (< 1.5% and < 5.9%, respectively). In addition to a suitable coverage of the Heterolobosea class (> 73%), the newly designed primer pairs also had the highest coverage for *Naegleria* genus representing 100% of the 21 *Naegleria* sequences in the SILVA SSU 132 database that were expected to be amplified (Table 1).

Because microbial eukaryotes are highly diverse, many studies concluded that no universal primer pairs can be designed to target all taxa with the same efficiency, while avoiding a resort to excessively degenerated primers. In a recent study, it was underlined that Excavata, including the Heterolobosea class, constitute a clade that is most often discriminated, in comparison with other groups (Vaulot *et al.*, 2021). An alternative approach to these universal primers is the application of group-specific primers, allowing an optimal compromise between coverage and specificity of the primers. Nevertheless, our *in silico* analysis highlighted that the primer pairs described so far in the literature were not very efficient to target the Heterolobosea sequences present in the SILVA database since the coverages range from 24 to 70%. In addition, the Vahl_560_F/Vahl_730_R primer pair, giving the best results (70% of coverage) produced an amplicon which size (approx. 170 base pairs; Le Calvez *et al.*, 2012), is too short and not discriminant enough for HTS applications and taxonomic identification. In fact, these primers were mostly designed to perform qPCR, requiring short amplicons. On the contrary, several primer pairs newly designed in this study cover more than 70% of the Heterolobosea sequences present in the SILVA database and were compatible in length

Table 1. *In silico* taxonomic coverage, expressed as a percentage, of the previously and newly described primer pairs against the SILVA SSU 132.

Primer pairs	Amplicon length (bp)	Euk	Amo	Dis	Ac	Tub	Vr	Exc	Het	Ng	Opi	Nu
Previously described												
Nae1/Nae2	400							1.5	40.0	100.0		
Nae3-For/Nae3-Rev	164							0.1	2.0	6.0		
NaegIF192/NaegIR344	153							0.8	24.3	60.0		
Vahl_560_F/ Vahl_730_R	170							2.0	70.0	66.0		
Newly described												
Vahl730f_C / Test6R ^a	330		0.1	0.3				2.0	73.0	100.0		
Vahl730f_C / R1200 ^a	470	0.1	0.1	0.3				2.0	73.0	100.0		
HF_308 / HR_648	340	0.6	3.4	3.1				1.2	35.1	80.0		
HF_592 / HR_982	390	0.8	1.1	0.6		4.2		5.5	75.0	94.7	0.4	
HF_413 / HR_1114	701	11.0	12.3	3.2	0.7	36.3		1.6	36.6	83.3	4.1	
HF_628 / HR_1114	486	10.8	15.8	7.5	1.4	37.9		10.3	40.8	100.0	4.3	
HF_628 / HR_982 ^a	354	12.7	12.0	3.2		1.9		5.0	79.6	100.0	17.6	
HF_1245 / HR_1680 ^a	435	3.4	1.5	0.6				3.4	85.7	100.0	6.0	
HF_1198 / HR_1680 ^a	482	3.4	1.5	0.6				3.4	85.7	100.0	5.9	
HF_1202 / HR_1680 ^a	478	0.1						3.3	85.7	100.0		

Aca, *Acanthamoeba*; Amo, Amoebozoa; Dis, Discosea; Euk, Eukaryota; Exc, Excavata; Het, Heterolobosea; Ng, *Naegleria*; Nu, *Nucleia*; Opi, Opisthokonta; Tub, Tubulinea; Vr, *Vermamoeba*.

The colour code indicates the coverage of primers, ranging from blue (low coverage) to red (full coverage).

a. Primers selected for PCR conducted on DNA from various amoeba isolates.

with HTS. Moreover, they seem quite specific because they cover less than 5% of the other tested taxa.

PCR evaluation of the *in silico*-selected primer pairs on isolates of amoebae

In order to confirm the *in silico* specificity of the six selected primer pairs, PCR experiments were conducted on DNA from various amoeba isolates, belonging to the Discosea, Tubulinea, Heterolobosea and Nuclemycea classes (Table S3). The results indicate that the Nae1/Nae2 primer pair failed to amplify DNA from all the Heterolobosea isolates like those from the *Vahlkampfia* and *Willaertia* genera (Fig. 1). Again, this primer pair was previously designed to amplify *Naegleria* and it succeeded to do so. Three other primer pairs were able to amplify DNA from all the Heterolobosea isolates but produced non-specific amplification from isolates that do not belong to Heterolobosea. Thus, only two primer pairs (i) HF_1198/HR_1680 and (ii) HF_1202/HR_1680 gave amplification of a single PCR product at the expected size, without amplifying any of the isolates from non-target amoebal classes. Although we are aware that this evaluation does not guarantee that those primers will not amplify any DNA other than Heterolobosea in the environment, they seem specific enough to be tested on environmental samples. The amplicon size produced by using those two primer pairs is of 5pprox. 480 bases pairs (bp). This sequence size is likely long enough to be taxonomically resolutive, while allowing a sufficient overlap between forward and reverse sequences when using Illumina MiSeq 2x300 bp chemistry, thus ensuring to reconstruct the complete amplicon.

Both HF_1198/HR_1680 and HF_1202/HR_1680 primer pairs target the same 18S rRNA region and differ by a shift of 4 bp on the forward primer. To investigate whether this commonly amplified region could provide sufficient taxonomic resolution for genus and/or species-level identification, a distance matrix based on theoretical amplicons yielded by HF_1202/HR_1680 was computed (Fig. 2A). This analysis highlighted that most heterolobosean genera could be clearly identified and separated from each other based on amplicon sequence. Furthermore, species-level identification could be achieved for the majority of *Naegleria* and *Tetramitus* species (Fig. 2B and C). Of note, the analysed amplicon allows to formally identify and discriminate *Naegleria fowleri* sequences, from other *Naegleria* species; a divergence of 4 bp was identified with the closest relative species, *N. lovaniensis* (Fig. 2B).

Overall, the set of analyses conducted *in silico* and *in vitro* suggest that two pairs of primer newly designed in this study provided an amplicon enabling a sensitive identification, down to genus and sometimes species level of heterolobosean representatives. By focusing on *Naegleria* genus, we were able to differentiate most species by variations of one or several nucleotidic positions within the amplicon. However, no difference could be observed among sequences from *N. gruberi*, *N. clarki* and *N. italica*. Interestingly, *N. fowleri* sequences could be clearly distinguished from their closest relative that is *N. lovaniensis*, by a 4-bp difference. Thus, the primer pair newly designed in this study would be proven useful for characterizing the presence of pathogenic *Naegleria* in environmental samples. By focusing on *Tetramitus*, the second most diversified genus among Heterolobosea, we

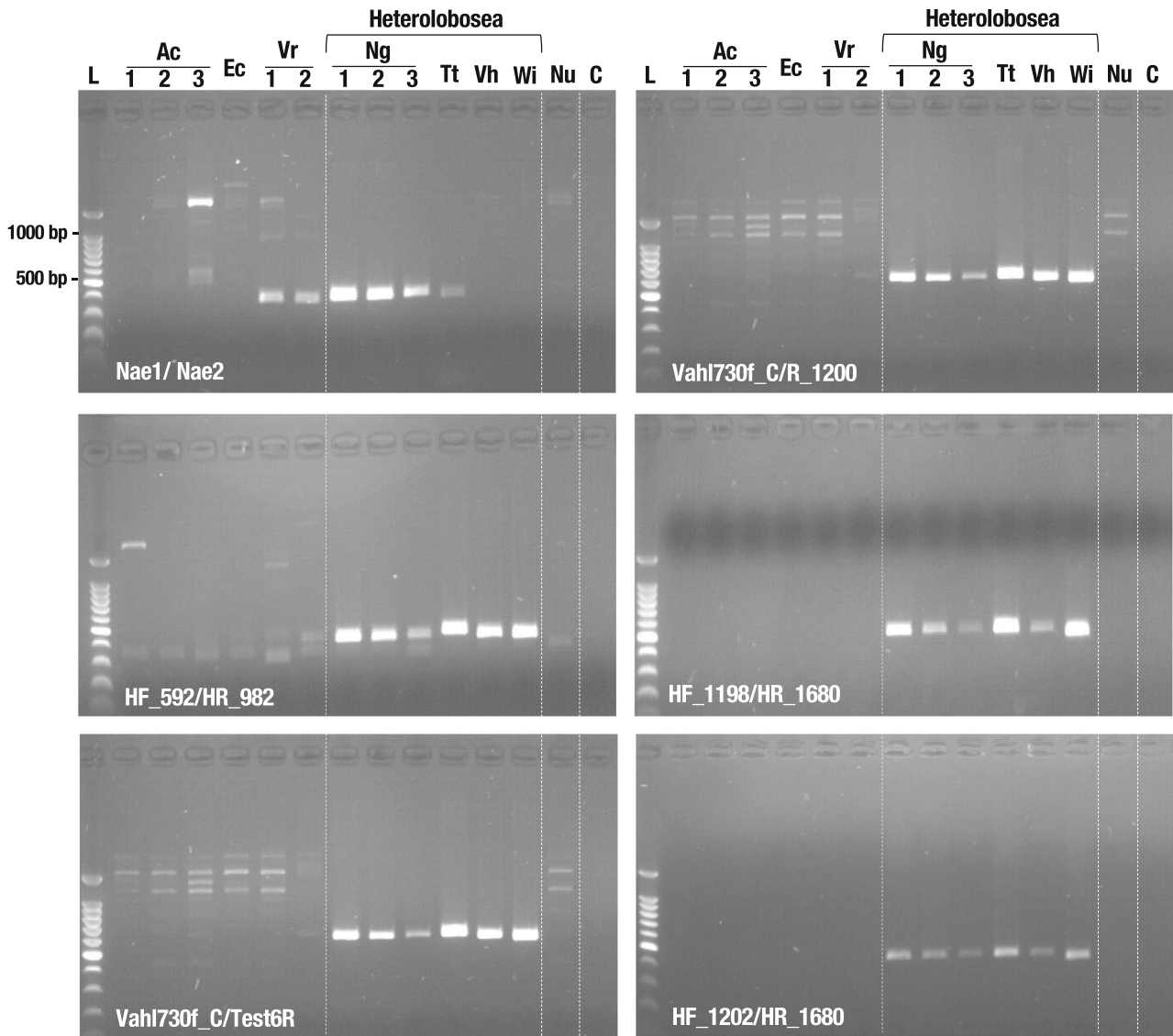


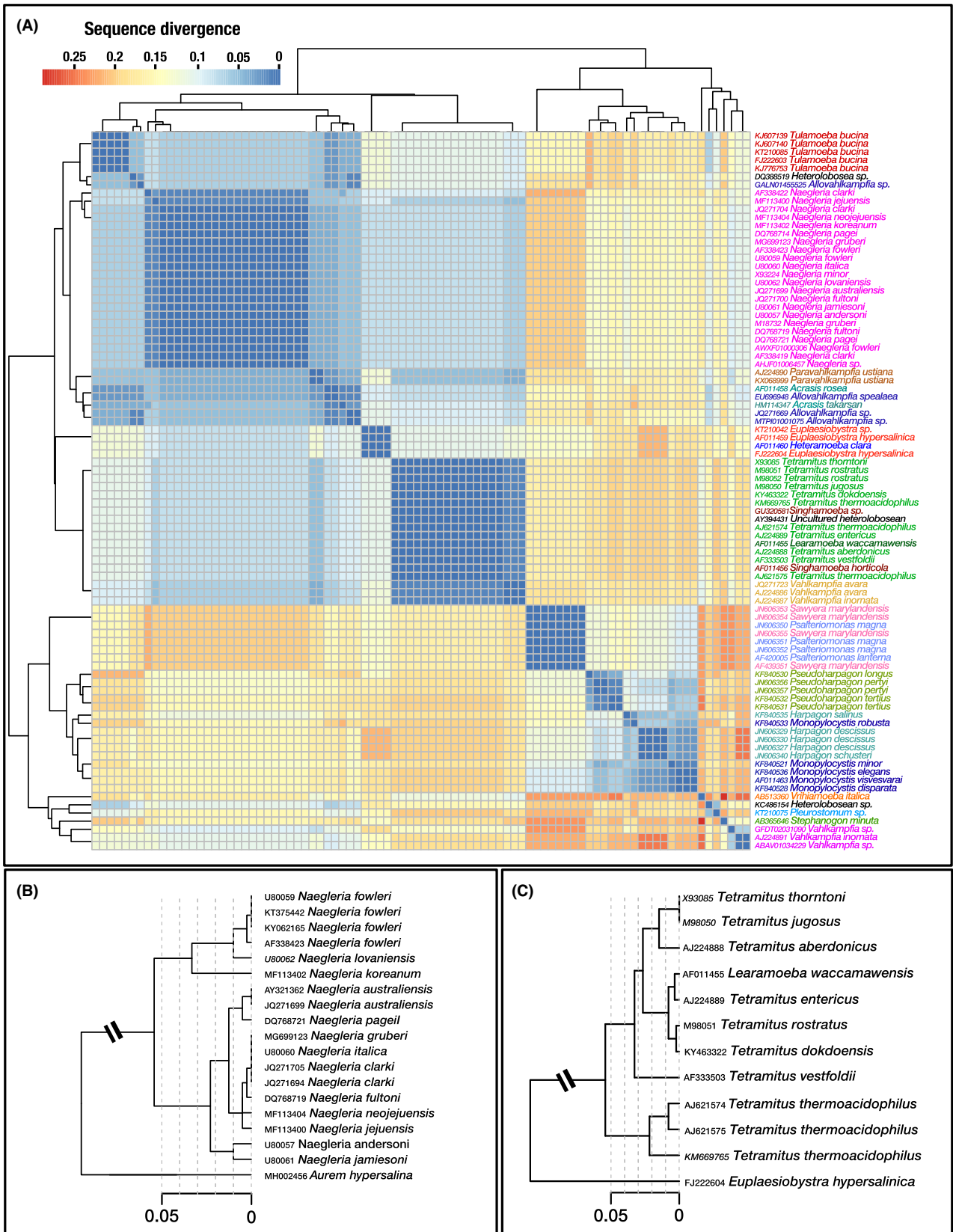
Fig. 1. Specificity of selected primer pairs for detection of the Heterolobosea. Agarose gel electrophoreses (2%) of PCR products using 6 different primer pairs for the environmental isolates of amoebae mentioned in Table S3. Ac: *Acanthamoeba*, Ec: *Echinamoeba*, Vr: *Vermamoeba*, Ng: *Naegleria*, Tt: *Tetramitus*, Vh: *Vahlkampfia*, Wi: *Willaertia*, Nu: *Nuclearia*, C: control PCR reaction with no template DNA. The numbers on the left correspond to the molecular weight (bp) of the BenchTop 100 bp DNA ladder (L). Results from PCR experiment conducted at 56°C.

showed that amplicon sequences also allowed for species-level identification of most species. *T. thomtoni* could not be distinguished from *T. jugosus*. Overall, our analyses suggest that fine identification, likely at the species level, of most heterolobosean can be theoretically achieved using the primer pairs HF_1198/HR_1680 and HF_1202/HR_1680.

Experimental evaluation of two newly described primer pairs on environmental samples

Based on *in silico* and *in vitro* analyses, the two newly designed primers HF_1198/HR_1680 and HF_1202/HR_1680 were further tested by using them for HTS experiments. For this, total DNA isolated from river water

Fig. 2. Sequence dissimilarity based on amplicons yielded by primer pair HF_1202/HR_1680, allows for a clear identification of heterolobosean genera (A). Each genus noted on the right side is colour-coded. B) Dendrogram based on DNA distance matrix, depicting sequence divergence among the amplicon yielded by primer pair HF_1202/HR_1680 for *Naegleria* (B) and *Tetramitus* (C) genera. Scale bar represents the percentage of nucleotidic divergence, with each dotted line representing 1% divergence.



and sediment samples were used as templates for PCR and sequencing, using an Illumina MiSeq platform. With each primer pair, a negative template control was performed and added to the sequencing run. In total, 2'387'524 raw sequences were generated (Table 2, Table S4). After the various filtering steps, 363'887 high-quality, eukaryotic sequences were retained for further analyses (mean of $45'486 \pm 15'736$ sequences per sample, excluding controls which comprised 0 or 6 sequences). Rarefaction analyses indicated that diversity contained within all samples was well covered and that the number of sequences generated for this study is largely sufficient (Fig. S2). On average, 241 ± 26 amplicons sequence variants (ASV) were identified per sample. Taxonomic identification of ASV highlighted that among those, 25 ± 11 ASVs were affiliated to the Heterolobosea group. These heterolobosean ASVs represented 26 to 66% of identified eukaryotic sequences in all environmental samples (Table 2; Fig. 3). Heterolobosean sequences thus consistently represented an important fraction of generated sequences, for all samples. By comparing both primer pairs, $29.6 \pm 4.1\%$ of total sequences were affiliated to Heterolobosea using the HF_1198/HR_1680 primer pair, compared to $41.6 \pm 16.6\%$ with the HF_1202/HR_1680 primer pair.

The diversity within Heterolobosea, as identified using a standard taxonomic pipeline implemented in QIIME2 (vsearch, coupled with SILVA database, see methods), was further explored at the genus level (Fig. 4A). Using the conventional pipeline, three genera could be identified, namely *Naegleria*, *Neovahlkampfia* and *Tetramitus*. *Naegleria* and *Tetramitus* accounted for the majority of sequences in all samples, while *Neovahlkampfia* was only identified in one sample and only with the primer pair HF_1202/HR_1680. Also, not all sequences could be identified down to the genus level, hence the presence of unidentified Heterolobosea and/or Tetramitia in 6 out of 8 samples analysed. In most samples, *Naegleria* sequences represented at least 75% of all sequences. Only two samples presented a majority (around 70%) of *Tetramitus* sequences. The tendency is that *Tetramitus* was more represented in sediments than in water. Similar results regarding Heterolobosea relative abundances and diversity were obtained when using the Protist Reference database PR² (v4.14, Fig. S3; Guillou *et al.*, 2013).

Because *in silico* analyses of amplicon taxonomic resolution suggested the possibility to identify ASV at the species level (Fig. 2), we tentatively re-classified all sequences using a manually curated database of 90 high-quality Heterolobosea sequences (Table S5). Identifications were performed using the BLAST algorithm, and species-level identification was achieved when coverage and identities provided unambiguous results.

Identities threshold were defined in light of previously computed distance matrix, manually inspected and confirmed for each ASV. Such an approach considerably improved the taxonomic identification of ASV in all samples (Fig. 4B). In total, eleven species were formally identified in the samples. While no difference could be observed regarding trends of genera abundances, numerous ASV could be identified down to species level, especially among *Naegleria* and *Tetramitus* genera, not all ASV could reliably be assigned to defined species, mostly due to a lack of identity or coverage. Overall, the global profiles of heterolobosean communities identified in all samples were comparable from one primer pair to another. Only minor differences emerged, such as the detection of *Vahlkampfia* sp. Sequences in Rhône water and of *Naegleria pagei* in Rhône sediments only observed using HF_1202/HR_1680. These differences are in line with the observed number of ASV and with diversity indexes, which were always slightly higher, though not significantly (Wilcoxon rank sum test with continuity correction, $W = 5$, $P = 0.4678$) for samples generated using HF_1202/HR_1680 primer pair (Table 2).

The characterization of amoebae belonging to the Heterolobosea group heavily relies on cultivation methods. The development of a primer pair, enabling the molecular analysis of heterolobosean diversity, gives access to the uncultivable fraction within this taxonomic group. Designing priming specifically covering this group was also motivated by the fact that heteroloboseans are frequently under-covered by other, more universal primers used for describing eukaryotic diversity (see primers referenced in Vaultot *et al.*, 2021). Our study thus aims to provide a new tool for better characterizing Heterolobosea diversity, without cultivation steps.

The presented approach for taxonomic identification relies on manual inspection of ASV identities compared with a curated set of reference sequences. The use of ASV is essential in completing such analyses; this approach enables a nucleotide-level comparison with reference sequences (Callahan *et al.*, 2017). A clear benefit of this approach is the enhancement of species-level identification of heterolobosean, as compared to other methods. Most notably, the set of primers tested in this study, coupled with the above-mentioned taxonomic identification method, theoretically allows the identification of the numerous heterolobosean species, as well as the unambiguous detection of *Naegleria fowleri* from high-throughput amplicons sequences. The presented work thus provides a method and a new set of primers to identify heterolobosean diversity with superior precision, as compared to other known methods.

Even though relative abundances were reproducible between the two primer pairs, such values should be

Table 2. Numeric and taxonomic features of sequences generated by high-throughput sequencing, per sample.

Identifier	Origin	Sample Type	Primer pair	Eukaryotic sequences	ASV number	Heterolobosean sequences	Heterolobosean ASV	Heterolobosean sequences (%)	Shannon H (Heterolobosea)	Overall mean genetic distance
11-11	26.24	1.60	1198_S36	Vienne		Freshwater	HF_1198-HR_1680	32 878	265	8626
14		0.62								
11-36.50	1.81	0.67	1202_S19			HF_1202-HR_1680	25 810	216	9421	18
13-27.56	1.99	0.59	1198_S7	Sediment		HF_1198-HR_1680	66 964	225	18 457	40
13-0.60			1202_S29	HF_1202-HR_1680	64 830	236	20 171	44	31.11	2.13
22-14	35.59	0.91	1198_S3	Rhône		Freshwater	HF_1198-HR_1680	51 644	292	18 379
22-66.07	1.00	0.64	1202_S25			HF_1202-HR_1680	28 288	222	18 691	23
25-29.01	3.04	0.32	1198_S13	Sediment		HF_1198-HR_1680	49 474	231	14 354	24
25-0.59			1202_S35	HF_1202-HR_1680	43 999	244	14 257	25	32.40	3.05

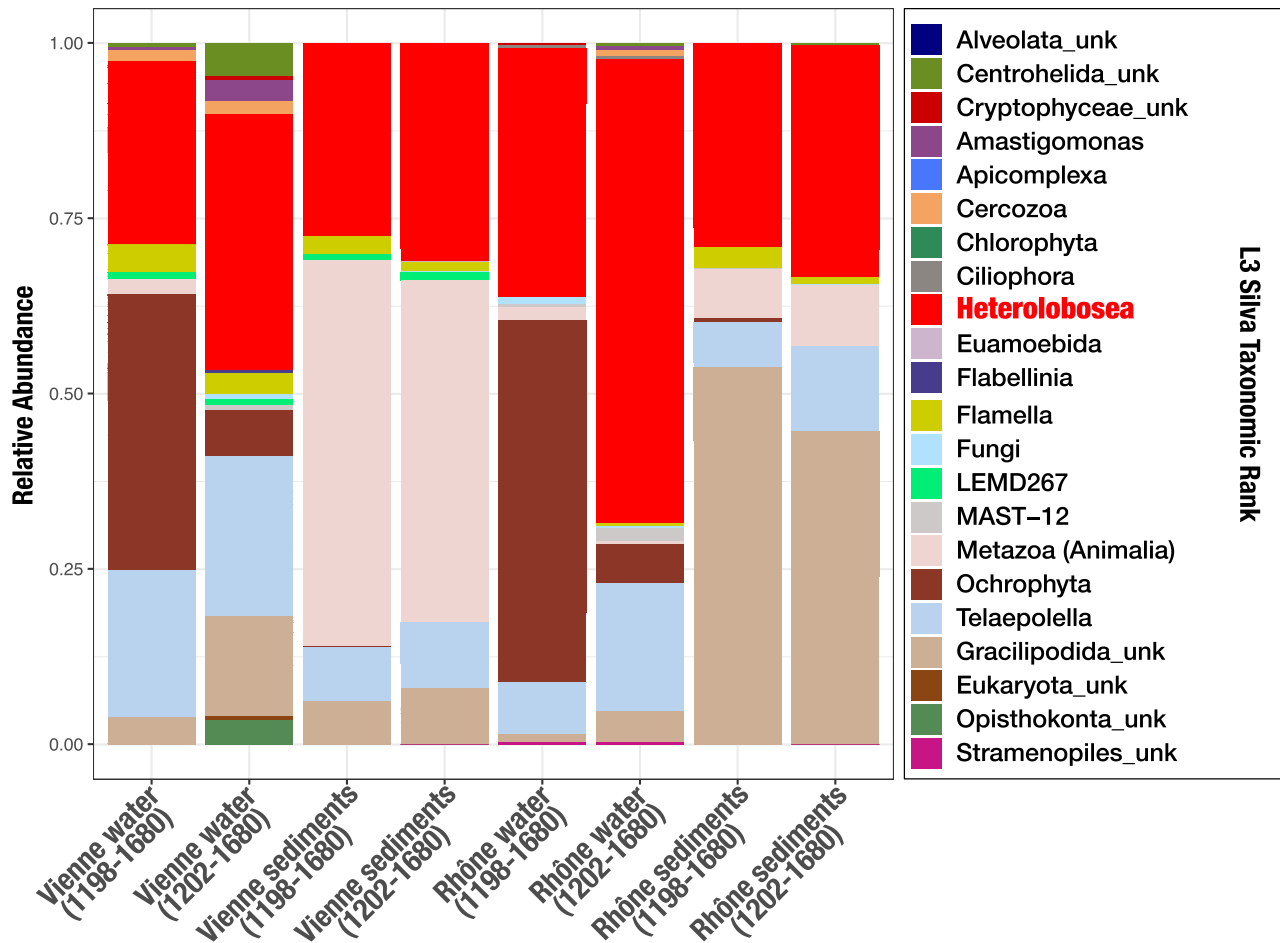


Fig. 3. Taxonomic assignments of amplicons yielded by primer pairs HF_1198/HR_1680 and HF_1202/HR_1680, from 4 environmental samples. Relative abundances are shown for each sample, and sequences affiliated to the Heterolobosea clade are highlighted in red.

interpreted with caution. Indeed, 18S rRNA gene copy numbers can vary a lot within protists cells, both because of interspecific and intraspecific variations (Lavrinenko *et al.*, 2021). Moreover, a peculiarity of *Naegleria* species lies in the fact that the rRNA cistron is exclusively located on an intranuclear plasmid, found in approx. 4000 copies per cell (Clark and Cross, 1987). Early research suggested this feature might be conserved within other heterolobosean such as *Tetramitus* and *Vahlkampfia* genera, though it has not been recently investigated (Clark and Cross, 1988). Thus, large variation in 18S rRNA gene copy number across heterolobosean species could be observed, prompting caution during interpretation of relative abundance data.

Conclusion

In conclusion, we designed a new PCR primer pair dedicated to Heterolobosea identification by high-throughput sequencing. We designed and tested several primer

pairs *in silico*, on pure culture and on environmental samples. The experiments led to the selection of one primer pair, HF_1202/HR_1680, showing the highest specificity and allowing to detect the highest diversity among Heterolobosea. Interestingly, the amplicon sequences were discriminant enough to distinguish most Heterolobosea at the species level. This primer pair allows, for the first time, a global description of the Heterolobosea community in environmental samples. In these samples, *Naegleria* (in majority), *Tetramitus*, *Vahlkampfia*, *Neovahlkampfia* and *Learamoeba* were detected. This primer pair should be further tested on samples coming from various environments in order to confirm its specificity and universality towards the Heterolobosea group.

Experimental procedures

Primer design and *in silico* primer evaluation

Specific primer pairs target the 18S rRNA gene sequence relative to Heterolobosea, were collected from

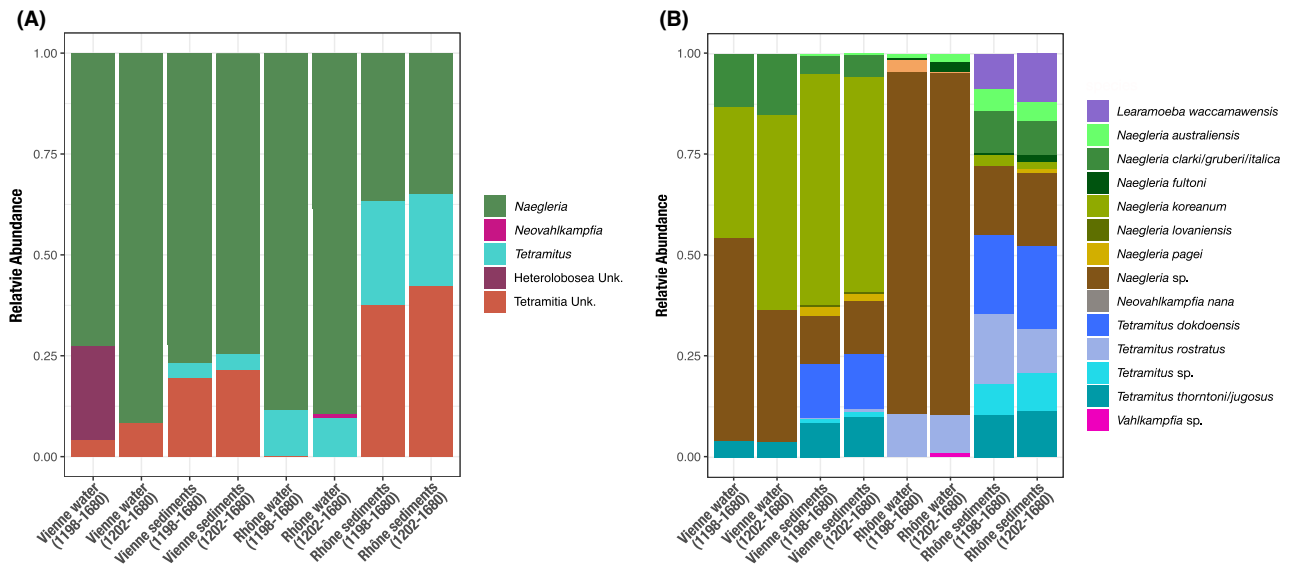


Fig. 4. Taxonomic assignments of heterolobosean amplicons yielded by primer pairs HF_1198/HR_1680 and HF_1202/HR_1680, from 4 environmental samples. The conventional pipeline using VSEARCH and SILVA allows genus-level taxonomic identification (A), while the improved pipeline described in this study allows for species-level identification of most amplicons (B).

the literature. In addition, a reference 18S rRNA sequences database was created with sequences from 17 Heterolobosea, available from the current SILVA SSU 132 database (<https://www.arb-silva.de/documentation/release-132/>; Table S2). Only 18S rRNA sequences with length longer than 1400 nt were considered for alignment using Muscle (Edgar, 2004). From an optimal sequence alignment, PrimerDesign_M (Yoon and Leitner, 2015) was performed to determine sites of conservation across the Heterolobosea sequences and identify primer pairs. The Flex fragment overlap option was used, as well as the following settings: an expected read length of 280–470 nucleotides (nt), no barcode tag, a minimum primer length of 18 nt, a maximum primer length of 25 nt, a detection limit at 5%, a complexity limit at 32 and a maximum difference in T_m at 3°C. In silico taxonomic coverage of the previously and newly designed primer pairs against the SILVA SSU 132 was performed using TestPrime, allowing for a maximum of 2 mismatches (www.arb-silva.de/search/testprime/, accessed in February 2019).

For evaluating the taxonomic resolution of amplicons yielded by selected primer pairs, a distance matrix was computed with sequence listed in Table S5. Sequences were aligned using MUSCLE (Edgar, 2004), and sequence dissimilarity was computed, using kimura parameter, in R environment and ape package (Paradis *et al.*, 2004; R Development Core Team, 2013). Heatmap was generated using gplot package (Ward, 2008), while dendrogram was generated using vegan (Oksanen *et al.*, 2013).

DNA extraction, PCR amplification and sequencing

DNA was extracted from 13 purified isolates of amoebae deposited in our laboratory culture collection of environmental free-living amoebae (FLA) using the Nucleospin Tissue kit according to manufacturer's instructions. Isolates were selected from the Discosea, Tubulinea, Heterolobosea and Nucleomyxa classes (Table S3).

Water ($n = 2$) and sediment ($n = 2$) samples were collected between June and October 2018 from the Vienne and Rhône rivers, France. Water samples (1 l) were filtered through a 3 µm nitrocellulosic membrane (Millipore) within 24 h after sampling. Total DNA was extracted directly from membranes (i.e. without classic enrichment step) using the DNeasy PowerWater DNA and from 500 mg (wet weight) of sediment using the DNeasy PowerSoil Pro DNA Kit, with the QIAcube workstation according to Qiagen's instructions. The integrity of the extracted DNA was observed with 0.8% agarose gel electrophoresis. The amount of extracted DNA was quantified using a NanoDrop™ One spectrophotometer (Thermo Scientific™). DNA from FLA isolates and from environmental samples were diluted to a standard concentration of 5 ng µl⁻¹, stored at -20°C until use for molecular applications.

All PCR reactions were carried in a total volume of 50 µl PCR mix with 0.5 µM of each forward and reverse primer (Table S1), 1X PCR mix buffer with 200 mM of each dNTP and 1 U of Q5 high fidelity DNA polymerase (NEB) and 25 ng of template DNA, using a vapo.protect™ Mastercycler (Eppendorf). DNA from FLA isolates

and from environmental samples was amplified under the following conditions: an initial denaturation at 98°C for 1 min, followed by 30 cycles at 98°C for 10 s, 56°C for 30 s and 72°C for 30 s and a final extension performed at 72°C for 2 min. The amplified products were checked by 2% agarose gel electrophoresis. For sequencing, triplicate PCR products for each environmental sample were pooled, purified using the PCR clean-up Kit with the QIAcube connect workstation according to the manufacturer's instructions (Qiagen) and quantified using a Qubit 2 fluorimeter (Invitrogen). Libraries were prepared using the NEBNext® Ultra II DNA library Prep Kit for Illumina (New England Biolabs). Sequencing was performed on the Illumina MiSeq sequencing platform with the V3 Illumina chemistry (2 × 300 cycles) at the ICM Institute (Paris, France; <https://icm-institute.org/fr/>) according to standard protocols.

Sequence data processing

Raw fastq files were recovered and processed using QIIME2 software package. First, data were imported ('qiime tools import' command) into the software package and overall quality data were visualized ('qiime demux summarize' command). Quality filtering, denoising and merging of paired reads were performed using DADA2 with standard parameters, as implemented in QIIME2 ('qiime dada2 denoise-paired' command; Callahan *et al.*, 2016). Reads were trimmed on their 5' ends in order to remove primer sequences and truncated on their 3' ends based on quality data (median quality < Q30). Thus, forward reads were truncated at 280 nt and reverse reads at 250 nt. The amplicon sequence variants (ASV) resulting from this step were then aligned using mafft (Kato, 2002), and a midpoint-rooted phylogeny was computed using FastTree (Price *et al.*, 2010). Alpha diversity parameters (Chao1, Faith phylogenetic diversity, Shannon indices, Observed OTUs, Goods coverage) were computed ('qiime diversity alpha-rarefaction' command) and visualized using QIIME 2 View online interface (Table 2, Fig. S2). Estimates of the mean evolutionary diversity indices for the entire population were calculated based on ASV sequence alignments, for each sample, using MEGA 11 (Tamura *et al.*, 2021).

Taxonomic classification of ASV was performed using vsearch and the SILVA release 132 database ('qiime feature-classifier classify-consensus-vsearch' command; Pruesse *et al.*, 2007; Rognes *et al.*, 2016). ASV DNA sequences, as well as a table grouping ASV counts per sample and full taxonomic paths, were then exported ('qiime tools export' command). In further steps, only ASVs affiliated to the 'Discicristata' taxonomic group were considered. In order to retain only heterolobosean sequences, sequences from Euglenozoa were filtered

out, as this group constitutes a distinctive class from the Heterolobosea within the Discicristata eukaryotic clade. To improve taxonomic classification down to genus and or species level, a collection of 18S rRNA sequences from Heterolobosea representative was constituted (Table S5), covering the 18S rRNA region amplified in this study. Heterolobosean ASVs were queried against this custom database using the BLASTn algorithm (Altschul *et al.*, 1990). Output was manually curated and ASVs were assigned to defined genera and species when unambiguous results were obtained, using threshold defined by DNA distance matrices (Fig. 3). Results were visualized using ggplot2 in R environment (Wickham, 2009).

Raw sequences were deposited on the Sequence Read Archive and are freely available under project number PRJNA725645.

Acknowledgements

We thank Nicolas Bech and Elie Morin for their guidance and assistance, for computing genetic diversity analyses. This work benefited from equipment and services from the iGenSeq core facility, at ICM.

Conflict of interest

None declared.

References

- Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A. *et al.* (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol* **66**: 4–119. doi:10.1111/jeu.12691.
- Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S. *et al.* (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**: 429–493. doi:10.1111/j.1550-7408.2012.00644.x.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2.
- Aucher, W., Delafont, V., Ponlaitiac, E., Alafaci, A., Agogue, H., Leboulanger, C. *et al.* (2020) Morphology and ecology of two new amoebae, isolated from a Thalassohaline Lake, Dziani Dzaha. *Protist* **171**: 125770. doi:10.1016/j.protis.2020.125770.
- Callahan, B.J., McMurdie, P.J., and Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* **11**: 2639–2643. doi:10.1038/ismej.2017.119.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583. doi:10.1038/nmeth.3869.
- Clark, C.G., and Cross, G.A. (1987) rRNA genes of *Naegleria gruberi* are carried exclusively on a 14-kilobase-pair

- plasmid. *Mol Cell Biol* **7**: 3027–3031. doi:[10.1128/MCB.7.9.3027](https://doi.org/10.1128/MCB.7.9.3027).
- Clark, C.G., and Cross, G.A. (1988) Circular ribosomal RNA genes are a general feature of schizopyrenid amoebae. *J Protozool* **35**: 326–329.
- De Jonckheere, J.F., Aerts, M., and Martinez, A.J. (1983) *Naegleria australiensis*: experimental meningoencephalitis in mice. *Trans R Soc Trop Med Hyg* **77**: 712–716. doi:[10.1016/0035-9203\(83\)90212-2](https://doi.org/10.1016/0035-9203(83)90212-2).
- De Jonckheere, J.F., and Brown, S. (2005) Isolation of a vahlkampfiid amoeba from a contact lens: *Tetramitus ovis* (Schmidt, 1913) n. comb. *Eur J Protistol* **41**: 93–97. doi:[10.1016/j.ejop.2004.11.006](https://doi.org/10.1016/j.ejop.2004.11.006).
- Dingle, A.D., and Fulton, C. (1966) Development of the flagellar apparatus of *Naegleria*. *J Cell Biol* **31**: 43–54. doi:[10.1083/jcb.31.1.43](https://doi.org/10.1083/jcb.31.1.43).
- Dykova, I., and Lom, J. (2004) Advances in the knowledge of amphizoic amoebae infecting fish. *Folia Parasitol* **51**: 81–97. doi:[10.14411/fp.2004.014](https://doi.org/10.14411/fp.2004.014).
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797. doi:[10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340).
- Gharpure, R., Bliton, J., Goodman, A., Ali, I.K.M., Yoder, J., and Cope, J.R. (2020) Epidemiology and clinical characteristics of primary amebic meningoencephalitis caused by *Naegleria fowleri*: a global review. *Clin Infect Dis* **73**: e19–e27. doi:[10.1093/cid/ciaa520](https://doi.org/10.1093/cid/ciaa520).
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bitner, L. et al. (2013) The protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small subunit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41**: D597–D604. doi:[10.1093/nar/gks1160](https://doi.org/10.1093/nar/gks1160).
- Katoh, K. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066. doi:[10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436).
- Lavrinenko, A., Jernfors, T., Koskimäki, J.J., Pirttilä, A.M., and Watts, P.C. (2021) Does intraspecific variation in rDNA copy number affect analysis of microbial communities? *Trends Microbiol* **29**: 19–27. doi:[10.1016/j.tim.2020.05.019](https://doi.org/10.1016/j.tim.2020.05.019).
- Le Calvez, T., Trouilhé, M.-C., Humeau, P., Moletta-Denat, M., Frère, J., and Héchard, Y. (2012) Detection of free-living amoebae by using multiplex quantitative PCR. *Mol Cell Probes* **26**: 116–120. doi:[10.1016/j.mcp.2012.03.003](https://doi.org/10.1016/j.mcp.2012.03.003).
- Lee, J.H. (2010) De novo formation of basal bodies during cellular differentiation of *Naegleria gruberi*: Progress and hypotheses. *Semin Cell Dev Biol* **21**: 156–162. doi:[10.1016/j.semdb.2009.12.009](https://doi.org/10.1016/j.semdb.2009.12.009).
- Maciver, S.K., Piñero, J.E., and Lorenzo-Morales, J. (2020) Is *Naegleria fowleri* an emerging parasite? *Trends Parasitol* **36**: 19–28. doi:[10.1016/j.pt.2019.10.008](https://doi.org/10.1016/j.pt.2019.10.008).
- Mrva, M., and Garajová, M. (2018) The efficiency of cultivation Media in Recovering Naked Lobose Amoebae from freshwater environments. *Zool Stud* **9**: 1–7. doi:[10.6620/ZS.2018.57-09](https://doi.org/10.6620/ZS.2018.57-09).
- Newsome, A.L., Baker, R.L., Miller, R.D., and Arnold, R.R. (1985) Interactions between *Naegleria fowleri* and *Legionella pneumophila*. *Infect Immun* **50**: 449–452. doi:[10.1128/iai.50.2.449-452.1985](https://doi.org/10.1128/iai.50.2.449-452.1985).
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2013) Package 'vegan.' *R Packag.* ver. 2.0–8, 254.
- Oliverio, A.M., Geisen, S., Delgado-Baquerizo, M., Maestre, F.T., Turner, B.L., and Fierer, N. (2020) The global-scale distributions of soil protists and their contributions to belowground systems. *Sci Adv* **6**: eaax8787. doi:[10.1126/sciadv.aax8787](https://doi.org/10.1126/sciadv.aax8787).
- Ovrutsky, A.R., Chan, E.D., Kartalija, M., Bai, X., Jackson, M., Gibbs, S. et al. (2013) Cooccurrence of free-living amoebae and nontuberculous mycobacteria in hospital water networks, and preferential growth of *Mycobacterium avium* in *Acanthamoeba lenticulata*. *Appl Environ Microbiol* **79**: 3185–3192. doi:[10.1128/AEM.03823-12](https://doi.org/10.1128/AEM.03823-12).
- Page, F. C. (1988). *A new key to freshwater and soil gymnamoebae: with instructions for culture* [WWW document]. URL https://books.google.fr/books/about/A_New_Key_to_Freshwater_and_Soil_Gymnamo.html?id=9DIJAQAAMAAJ&pgis=1.
- Pánek, T., Simpson, A.G.B., Brown, M.W., and Dyer, B.D. (2017) *Heterolobosea*. Cham, Switzerland: Springer.
- Paradis, E., Claude, J., and Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290. doi:[10.1093/bioinformatics/btg412](https://doi.org/10.1093/bioinformatics/btg412).
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490. doi:[10.1371/journal.pone.0009490](https://doi.org/10.1371/journal.pone.0009490).
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196. doi:[10.1093/nar/gkm864](https://doi.org/10.1093/nar/gkm864).
- Qvarnstrom, Y., Visvesvara, G.S., Sriram, R., and Da Silva, A.J. (2006) Multiplex real-time PCR assay for simultaneous detection of *acanthamoeba* spp., *Balamuthia mandrillaris*, and *naegleria fowleri*. *J Clin Microbiol* **44**: 3589–3595. doi:[10.1128/JCM.00875-06](https://doi.org/10.1128/JCM.00875-06).
- R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Régoudis, E., and Pélandakis, M. (2016) Detection of the free living amoeba *Naegleria fowleri* by using conventional and real-time PCR based on a single copy DNA sequence. *Exp Parasitol* **161**: 35–39. doi:[10.1016/j.exppara.2015.12.007](https://doi.org/10.1016/j.exppara.2015.12.007).
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584. doi:[10.7717/peerj.2584](https://doi.org/10.7717/peerj.2584).
- Samba-Louaka, A., Delafont, V., Rodier, M.-H., Cateau, E., and Héchard, Y. (2019) Free-living amoebae and squatters in the wild: ecological and molecular features. *FEMS Microbiol Rev* **43**: 415–434. doi:[10.1093/femsre/fuz011](https://doi.org/10.1093/femsre/fuz011).
- Scheikl, U., Sommer, R., Kirschner, A., Rameder, A., Schrammel, B., Zweimüller, I. et al. (2014) Free-living amoebae (FLA) co-occurring with legionellae in industrial waters. *Eur J Protistol* **50**: 422–429. doi:[10.1016/j.ejop.2014.04.002](https://doi.org/10.1016/j.ejop.2014.04.002).
- Schild, M., Gianinazzi, C., Gottstein, B., and Müller, N. (2007) PCR-based diagnosis of *Naegleria* sp. infection in formalin-fixed and paraffin-embedded brain sections. *J Clin Microbiol* **45**: 564–567. doi:[10.1128/JCM.01357-06](https://doi.org/10.1128/JCM.01357-06).
- Sheehan, K.B., Fagg, J.A., Ferris, M.J., and Henson, J.M. (2003) PCR detection and analysis of the free-living

- amoeba *Naegleria* in hot springs in Yellowstone and Grand Teton National Parks. *Appl Environ Microbiol* **69**: 5914–5918. doi:[10.1128/AEM.69.10.5914](https://doi.org/10.1128/AEM.69.10.5914).
- Singer, D., Seppey, C.V.W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L. *et al.* (2021) Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ Int* **146**: 106262. doi:[10.1016/j.envint.2020.106262](https://doi.org/10.1016/j.envint.2020.106262).
- Smirnov, A.V. (2003) Optimizing methods of the recovery of gymna amoebae from environmental samples: a test of ten popular enrichment media, with some observations on the development of cultures. *Protist* **3**: 47–57.
- Sze, M.A., and Schloss, P.D. (2019) The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. *mSphere* **4**: e00163-19. doi:[10.1128/msphere.00163-19](https://doi.org/10.1128/msphere.00163-19).
- Tamura, K., Stecher, G., and Kumar, S. (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol* **38**: 3022–3027. doi:[10.1093/molbev/msab120](https://doi.org/10.1093/molbev/msab120).
- Taravaud, A., Ali, M., Lafosse, B., Nicolas, V., Féliers, C., Thibert, S. *et al.* (2018) Enrichment of free-living amoebae in biofilms developed at upper water levels in drinking water storage towers: an inter- and intra-seasonal study. *Sci Total Environ* **633**: 157–166. doi:[10.1016/j.scitotenv.2018.03.178](https://doi.org/10.1016/j.scitotenv.2018.03.178).
- Vaulot, D., Geisen, S., Mahé, F., and Bass, D. (2021). pr2-primers: an 18S rRNA primer database for protists. *bioRxiv*. doi:[10.1101/2021.01.04.425170](https://doi.org/10.1101/2021.01.04.425170).
- Visvesvara, G.S., Sriram, R., Qvarnstrom, Y., Bandyopadhyay, K., Da Silva, A.J., Pieniazek, N.J. *et al.* (2009) *Paravahlkampfia francinae* n. sp. masquerading as an agent of primary amoebic meningoencephalitis. *J Eukaryot Microbiol* **56**: 357–366. doi:[10.1111/j.1550-7408.2009.00410.x](https://doi.org/10.1111/j.1550-7408.2009.00410.x).
- Warnes, G. R. (2008) *gplots: Various R programming tools for plotting data v. 2.6.0* [WWW document]. URL <https://www.scienceopen.com/document?vid=0e5d8e31-1fe4-492f-a3d8-8cd71b2b8ad9>.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York, NY, USA: Springer-Verlag.
- Yoon, H., and Leitner, T. (2015) PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. *Bioinformatics* **31**: 1472–1474. doi:[10.1093/bioinformatics/btu832](https://doi.org/10.1093/bioinformatics/btu832).
- Yubuki, N., and Leander, B.S. (2008) Ultrastructure and molecular phylogeny of *Stephanopogon minuta*: an enigmatic microeukaryote from marine interstitial environments. *Eur J Protistol* **44**: 241–253. doi:[10.1016/j.ejop.2007.12.001](https://doi.org/10.1016/j.ejop.2007.12.001).

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Positions the multiple primers designed and used in this study, placed on an alignment of heterolobosean 18S rRNA gene sequences, generated using PrimerDesignM. The x-axis shows nucleotide position, and y-axes show entropy (light brown) and complexity (green). Primers are displayed as arrowheads, coloured according to the combinations used. *: F_628 was also used in combination with R_982.

Fig. S2: Rarefaction curves obtained for each sample, using Chao1, Goods coverage, Faith Phylogenetic diversity and the number of observed OTUs. The sampling depth was normalized at 20'000 sequences.

Fig. S3. Taxonomic assignments of amplicons yielded by primer pairs HF_1198/HR_1680 and HF_1202/HR_1680, from 4 environmental samples, using VSEARCH and the Protist Reference Database PR2 v4.14 (Guillou *et al.*, 2013). Taxonomic assignments are shown at level 4 (A), showing abundances of heterolobosean sequences, and at level 9, showing genus and species-level identification (B).

Table S1. Primer pairs to identify Heterolobosea communities by high-throughput sequencing

Table S2. 18S rRNA sequences of Heterolobosea from SILVA SSU 132 database used for the design of new primer pairs

Table S3. Isolates of amoebae to evaluate the specificity of the primer pairs

Table S4. Impact of the various quality control and filtering steps on the high-throughput sequencing dataset.

Table S5. 18S rRNA sequences of Heterolobosea from SILVA SSU 132 database used species identification.