Evaluation of different primers of the 18S rRNA gene to profile amoeba communities in environmental samples

Article *in* Water Biology and Security · June 2022

DOI: 10.1016/j.watbs.2022.100057

CITATIONS
9

READS
253

5 authors, including:

Xiafei Zheng
Sun Yat-Sen University
34 PUBLICATIONS **946** CITATIONS
SEE PROFILE

Zhili He
Southern Marine Science and Technology Guangdong Laboratory (Zhuhai)
647 PUBLICATIONS **30,475** CITATIONS
SEE PROFILE

Qingyun Yan
Southern Maine Science and Engineering Guangdong Laboratory (Zhuhai)
151 PUBLICATIONS **5,391** CITATIONS
SEE PROFILE

Longfei Shu
Sun Yat-Sen University
78 PUBLICATIONS **1,966** CITATIONS
SEE PROFILE

Water Biology and Security 1 (2022) 100057

Contents lists available at ScienceDirect

Water Biology and Security

journal homepage: www.keaipublishing.com/en/journals/water-biology-and-security

Evaluation of different primers of the 18S rRNA gene to profile amoeba communities in environmental samples

Xiafei Zheng [a,b], Zhili He [a], Cheng Wang [a], Qingyun Yan [a,**], Longfei Shu [a,*]

[a] Environmental Microbiomics Research Center, School of Environmental Science and Engineering, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Guangzhou, 510006, China
[b] Ninghai Institute of Mariculture Breeding and Seed Industry, Zhejiang Wanli University, Ningbo, 315100, China

ARTICLE INFO

Keywords:
Amoeba community
Diversity
Drinking water safety
Metagenomics
18S rRNA gene
Protist

ABSTRACT

Amoeboid protists, an assemblage of organisms belonging to different phylogenetic lineages, have drawn increasing attention due to their crucial ecological roles in various environments and their potential health risks. Currently, 18S rRNA gene sequencing is widely applied for the detection of amoebae. However, it is not clear which is the best primer pair for 18S rRNA gene amplification in amoebae. This study compared the four most commonly used primer pairs for revealing the diversity, composition, core species, and community assembly processes of amoebae in water and sediments. We found that the choice of primers artificially influences the detection of community composition of amoebae. We also found that short-read fragments may lead to mismatches in taxonomy and were not suitable for phylogenetic analyses. In contrast, full-length primers could detect the highest number of amoeba lineages and annotate 80% of reads belonging to amoebae to known species. However, full-length primers did not detect as many amoeba species as V4 primers. Moreover, we showed that beta diversity and community assembly determination were largely unaffected by primer choice, but different primers could influence our interpretations of the ecological process underlying stochasticity and determinism. This study indicates that full-length read sequencing and V4 region Illumina sequencing are suitable for profiling amoeba diversity in the environment.

1. Introduction

Amoebae (amoeboid protists) are widespread in water, soil and the air. They are characterized by their ability to produce pseudopods and feed by phagocytosis. They play significant roles in nutrient cycling and energy flow in aquatic and terrestrial ecosystems (Anderson, 2018; Samba-Louaka et al., 2019; Zhang et al., 2022; Yu et al., 2022). Some amoebae are pathogenic and can cause human diseases such as amoebiasis, keratitis and granulomatous encephalitis (Shirley et al., 2018), while *Naegleria fowleri* (also known as the "brain-eating amoeba") can infect and destroy the human brain, resulting in a fatality rate over 97% in the USA (Bartrand et al., 2014; Cooper et al., 2019; Visvesvara et al., 2007). Non-pathogenic amoebae are also of concern to human health because they have complex interactions with bacteria, fungi, and viruses (Balczun and Scheid, 2017; Shu et al., 2018a, b; Shu et al., 2021). For example, *Acanthamoeba* spp., *Naegleria* sp., *Hartmannella* sp. and other amoeboid protists can serve as environmental reservoirs for pathogens, such as *Legionella pneumophila*, Mycobacteria, Chlamydiae, and viruses (Balczun and Scheid, 2017; Samba-Louaka et al., 2019; Shi et al., 2021). Furthermore, ingestion by amoebae enables some amoeba-resisting bacteria (ARBs) to evolve the ability to infect humans (Balczun and Scheid, 2017; Strassmann and Shu, 2017). Therefore, it is crucial to understand the diversity and distribution of amoebae in the environment.

The presence of amoebae in man-made water systems is an emerging health risk (Shi et al., 2021). Amoebae have been isolated from cooling towers (Tsao et al., 2019), drinking water networks (Delafont et al., 2016; Miller et al., 2018), and tap water (Delafont et al., 2013). These findings suggest that amoebae are widespread in drinking water systems. Like *Cryptosporidium* and *Giardia*, some amoebae are resistant to commonly used disinfectants such as chlorine (Dupuy et al., 2011; Thomas and Ashbolt, 2011; He et al., 2021, 2022). For instance, *Acanthamoeba* cysts can survive exposure to chlorine at 100 mg L^{-1} for 10 min or 50 mg L^{-1} for 18 h (Storey et al., 2004).

** Corresponding author.
* Corresponding author.,
E-mail addresses: yanqy6@mail.sysu.edu.cn (Q. Yan), shulf@mail.sysu.edu.cn (L. Shu).

https://doi.org/10.1016/j.watbs.2022.100057
Received 18 January 2022; Received in revised form 13 May 2022; Accepted 9 June 2022
Available online 14 June 2022

Accurate identification of amoebae is the key to successfully assessing their potential health risks. Traditionally, most studies relied on a culture-based approach, which is often laborious and not always successful (Samba-Louaka et al., 2019). DNA barcoding metagenomic techniques, such as the sequencing of 18S rRNA gene amplicons, have greatly improved our ability to detect amoebae (Samba-Louaka et al., 2019). Several studies have utilized such an approach to determine the diversity of amoebae (Buse et al., 2014; Delafont et al., 2013, 2016, 2019; Wang et al., 2020). Also, using amplicon pyrosequencing, Delafont et al. investigated the amoeba community in drinking water networks and found that *Vermamoeba vermiformis* was the most abundant species, followed by *Acanthamoeba* (Delafont et al., 2013, 2016). Another study found that more amoebae were detected in hot water than in cold water, in which *V. vermiformis* was also the most abundant species (Buse et al., 2014). In addition, determining how amoebae biodiversity and community composition are generated and maintained has received increasing attention in recent years. Previous studies indicated that both stochastic and deterministic processes drive the amoeba community assembly in subtropical freshwater reservoirs, but the former was the dominant mechanism (Ren et al., 2018; Wang et al., 2020). Therefore, how to use 18S rRNA gene metabarcoding to precisely quantify amoeba diversity and its ecological drivers is important for water security.

The successful application of 18S rRNA gene metabarcoding to detect amoeboid protists belonging to different phylogenetic lineages is constrained by the lack of a universal primer pair. Current studies utilize different sets of primers targeting diverse regions (e.g., V4 and V9) of the 18S rRNA gene (Countway et al., 2005; Ren et al., 2018; Stoeck et al., 2010; Thomas et al., 2006). No study, however, has compared or evaluated the detection efficiency of different primer pairs for amoebae. One recent study used qPCR and sequencing of 18S rRNA gene amplicons to detect pathogens in drinking water systems, and it was found that although both approaches could detect *Acanthamoeba*, *N. fowleri* was only detected by qPCR, highlighting the vital role of the choice of primers (Garner et al., 2018).

In this study, we aimed to identify a suitable primer pair for analyzing amoeba diversity in the environment by addressing two key questions: (1) how different primers influence the coverage and detection of amoebae; and (2) how the amplification primers influence the evaluation of amoeba community composition and assembly. We compared different 18S rRNA gene amplification primers to evaluate their accuracy and efficiency both in identifying core amoebae and potentially pathogenic amoebae from water and sediment samples, and in determining the phylogenetic relationships of these amoebae. The results provided a guideline for primers choosing in analyzing amoeboid protists from the environmental samples.

2. Materials and methods

2.1. Sampling sites and collection procedures

Water and sediment samples were collected from six aquatic ponds located at the Pearl River Estuary (22°35′5.26″N, 113°37′56.23″E) on May 31, 2018. These ponds are mainly used for culturing grass carp and have a salinity of 1–2‰. On the sampling date, the temperature was 32–33 °C, the pH was around 8, and the dissolved oxygen concentration ranged from 5 to 7 mg L^{-1}. Three water and sediment samples were collected from each pond, resulting in a total of 36 samples (18 each for water and sediment). All the samples were kept in a portable fridge and then transferred to the laboratory within 1 h. About 500 mL of water sample was filtered with 0.22 μm polycarbonate filters (47 mm diameter, Whatman, Maidstone, UK). The filters and sediment were stored at −80 °C for further DNA extraction.

2.2. DNA isolation, PCR amplification and sequencing

Environmental DNA from filters and sediments were extracted by PowerWater DNA Isolation Kit and PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, USA), respectively. For sediments, we combined the freeze-grind DNA extraction method with the MoBio PowerSoil kit as described in http://www.ou.edu/ieg/tools/protocols. The quantity and quality of DNA were determined using a NanoDrop Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). We used four primer pairs to amplify the eukaryotic small-subunit 18S rRNA gene (Table S1). In brief, full-length primers Euk-A (5′ AACCTGGTTGATCCTGCCAGT 3′) and Euk-B (5′ GATCCTTCTGCAGGTTCACCTAC 3′) were used to target the full length of the eukaryotic small-subunit 18S rRNA gene (Full-length, ~2000–2540 bp) (Countway et al., 2005). Amoeba-specific primers Ami6F1 (5′ CCAGCTCCAATAGCGTATATT 3′) and Ami9R (5′ GTTGAGTCGAATTAAGCCGC 3′) were used to target the amoebae (Amoeba-specific, ~600–700 bp) (Delafont et al., 2016). The primers TAReuk454FWD1 (5′-CCAGCA(G/C)C(C/T)GCGGTAATTCC-3′) and TAReukREV3 (5′-ACTTTCGTTCTTGAT(C/T)(A/G)A-3′) were used to target the V4 region of the eukaryotic 18S rRNA gene (V4, ~420 bp) (Stoeck et al., 2010). The primers 1380F (5′-CCCTGCCHTTTGTACACAC-3′) and 1510R (5′-CCTTCYGCAGGTTCACCTAC-3′) were used to target the V9 region of the eukaryotic 18S rRNA gene (V9, ~130 bp) (Ren et al., 2018).

The PCR program for full-length amplification was composed of an initial denaturation step at 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 2 min; and a final extension at 72 °C for 5 min. Amplifications by amoeba-specific primers were carried out with an initial denaturation step at 95 °C for 5 min, 30 cycles of denaturation at 95 °C for 30 s, primer annealing at 50 °C for 30 s, elongation at 72 °C for 1 min and a final elongation step at 72 °C for 5 min. The V4 and V9 amplifications comprised an initial activation step at 95 °C for 5 min, followed by 30 three-step cycles consisting of 94 °C for 30 s, 47 °C for 45 s and 72 °C for 1 min, and a final extension at 72 °C for 5 min. The PCR products by full-length and amoeba-specific primers were sequenced on the PacBio Sequel II System (Pacific Biosciences of California, Inc.). The PCR products of V4 and V9 amplifications were sequenced on the Illumina HiSeq2500 platform (Illumina, Inc., San Diego, CA, USA). The original sequence data have been deposited in the NCBI Sequence Read Archive (SRA) database under the BioProject number PRJNA667969.

2.3. Bioinformatic analysis

The raw reads were mapped to sample barcodes, and both forward and reverse primers were trimmed to allow one mismatch by Cutadapt (1.16) (Martin, 2011). For the sequences of V4 and V9 amplifications, paired-end reads of sufficient length (>200 bases) were combined with at least 10 bp overlap by FLASH program (Magoc and Salzberg, 2011). The sequences generated from PacBio did not need to combine. For all four kinds of amplification sequences, poorly overlapped and low-quality sequences, such as those with a quality score <20 with a window size of 5, were removed by Btrim program (Kong, 2011). Any sequences with an ambiguous base were discarded. Then, UPARSE was used to remove chimeras and classify the sequences into operational taxonomic units (OTUs) at the identity of 97%; singletons were discarded (Edgar, 2013). A single representative sequence for each OTU was aligned using PyNAST (Caporaso et al., 2010). The ribosomal database project (RDP) classifier (Wang et al., 2007) was used to assign 18S rDNA sequences to protist taxonomy according to the Protist Ribosomal Reference (PR^2) database (Guillou et al., 2013). We used the default parameters in RDP classifier except the assignment confidence cutoff set as 0.5. Finally, the full-length, amoeba-specific, V4 and V9 primer datasets retained 113729, 277812, 1776254 and 2597671 reads, respectively.

Amoebae were selected from the total microbial eukaryotic community according to the phylogenetic lineages containing amoeboid protists (Adl et al., 2019; Goldstein, 2003; Ren et al., 2018; Samba-Louaka et al., 2019). Finally, the full-length, amoeba-specific, V4 and V9 primer datasets retained 13875, 21637, and 95652, 41676 reads, respectively. All samples were rarefied to an equal sequencing depth when calculating the alpha diversity of the amoeba communities. Core microbiome refers to the

species that were detected in a high fraction of all the samples above a given abundance threshold. In this study, the core amoebae were defined with a threshold of 20% prevalence in all samples and above the relative abundance of 0.01% in any given sample. We further compared the special core amoebae in the V9 dataset with the full-length dataset using the Basic Local Alignment Search Tool (BLAST) with perc_identity as 90, and the highest-scoring results were selected. The sequences were aligned by the PyNAST method using the align_seqs.py script in QIIME 1. The aligned sequences were then used to build the phylogenetic trees by fasttree method (Price et al., 2009) using make_phylogeny.py script in Qiime 1. These trees were visualized with MicrobiomeAnalyst (Chong et al., 2020).

2.4. Community assembly

The Sloan neutral community model (NCM) was used to assess the contributions of neutral processes in the amoeba community (Sloan et al., 2006). In the neutral community model, the relative abundance of taxa was governed by random dispersal, birth and death. The neutral model was fitted using the R codes from Burns et al. (2016). The OTUs that fall within the 95% confidence intervals of the neutral model's best fit are considered neutrally distributed. The OTUs that are above the 95% confidence interval (above prediction) are likely positively selected by the environment or have increased dispersal ability relative to other microbes. The OTUs that fall below the 95% confidence interval (below prediction) are selected against by the environment or have limited dispersal compared to others.

Null model analysis was further carried out to quantify the deterministic and stochastic processes of community assembly using the framework described by Stegen et al. (2013). In the null model, β-Nearest Taxon Index (βNTI) and Bray–Curtis-based Raup-Crick (RC_{bray}), were used to quantify the impacts of major ecological processes in governing the community assembly. If $|\beta NTI| > 2$, the main driver of community assembly is determinism. If $\beta NTI > 2$ indicates more phylogenetic turnover than expected (i.e. heterogeneous selection) whereas $\beta NTI < -2$ indicates less phylogenetic turnover than expected (i.e. homogeneous selection). If $|\beta NTI| < 2$, the main driver of community assembly is stochasticity. The relative contribution of homogenizing dispersal is estimated by the percentage of pairwise comparisons with $|\beta NTI| < 2$ and $RC_{bray} < -0.95$. The relative contribution of dispersal limitation is estimated by the percentage of pairwise comparisons with $|\beta NTI| < 2$ and $RC_{bray} > 0.95$. The percentage of pairwise comparisons with $|\beta NTI| < 2$ and $|RCbray| < 0.95$ are used to estimate the influence of other undominated processes, which include weak selection, weak dispersal, diversification, and drift (Zhou and Ning, 2017).

2.5. Statistical analyses

The alpha diversity of amoebae determined with different primers was compared by Kruskal-Wallis test, and multiple testing was corrected by controlling the false discovery rate (FDR) with a cutoff of 0.05 (Benjamini et al., 2006). The principal component analysis (PCA) was performed based on Bray-Curtis distance and plotted with the *factoextra* package (https://cran.r-project.org/package=factoextra). Permutational multivariant analysis of variance (PERMANOVA) was applied to examine the differences in the composition of the amoeba communities between water and sediment using the vegan package (Dixon, 2003). The similarity of amoeba community composition was performed based on Bray-Curtis distance at the phylum and species levels. Kruskal-Wallis test and multiple testing corrections were conducted in GraphPad Prism 7.00. PCA and PERMANOVA were performed in R program (3.4.1) (R Core Team, 2010).

3. Results

3.1. Overall comparison of taxonomy annotation among different primers

Amoebozoa, Heterolobosea, Vampyrellida, Euglyphida and Rhizaria were the highest-rank taxonomic clades recovered, and we found the full-length primer pair had good coverage of these clades (Table 1). However, the full-length primer pair failed to detect the genera *Rhogostoma* or *Nuclearia*. In contrast, the amoeba-specific and V4 primer pairs used in this study failed to detect Foraminifera and Heterolobosea, respectively (Table 1). Most sequences amplified by full-length, amoeba-specific and V4 primers could identify taxa to species level, but only half of the sequences could identify taxa to species level using the V9 primer pair (Fig. S1). The sequences identified to species level by the full-length and amoeba-specific primers were about 10% and 8%, respectively, of the total sequences in all microbial eukaryotic communities, but only 6% and 1% for the V4 and V9 primers, respectively (Fig. S1).

To quantify the difference of primers in the taxonomy annotation, we compared the numbers of taxa at each taxonomic rank. The results showed that V4 primers detected the least supergroups but the most species (Table 2). The amoeba-specific primer pair detected the least number of amoeba species (Table 2).

3.2. Diversity and composition of amoebae revealed by different primers

The OTU number, Shannon diversity, and Simpson evenness varied among the four primer pairs in both water and sediment samples (Fig. 1a). In sediment, the alpha diversity was higher in V9 and V4 datasets than in amoeba-specific and full-length primer datasets. In water, alpha diversity had no apparent trend (Fig. 1a). Also, all the primer pairs could successfully separate water and sediment samples (Fig. 2, PERMANOVA, $P < 0.001$). At the phylum level, the V9 primer pair detected the most phyla but recovered the lowest relative abundance of amoebae in the aquatic microbial eukaryotic community (Fig. 1b). In the sediment, full-length and V9 primers detected a high relative abundance of Conosa compared to amoeba-specific and V4 primers, and V9 detected a small percentage of Cercozoa. The amoebae compositions detected using amoeba-specific and V4 primers were similar in both water and sediment (Fig. 1c). At the species level, the amoebae identified by the V9 primer were quite different to those identified by other primer pairs (Fig. S2). Leptophryidae X sp. was detected as a dominant species in the full-length, amoeba-specific and V4 datasets, but was recovered as a rare species in the V9 dataset in both water and sediment (Fig. S2). In contrast, *Korotnevella* sp. and *Mastigamoeba* sp. were detected as the most abundant species in water and sediment, respectively, by the V9 primer pair, but were absent or rare in other primer datasets (Fig. S2). *Theratromyxa weberi* and *Kelleromyxa fimicola* were detected by the full-length primer pair as dominant species in water and sediment, respectively, but were absent in the other primer datasets (Fig. S2). Altogether, the alpha diversity indexes of the amoebae were the highest in the V9 and V4 primer datasets, but the community compositions of amoeba at the phylum and species levels were quite different from those in other primer datasets.

3.3. Core amoebae and potential pathogenic amoebae

We compared the resolution of the four primers in detecting amoeba species. The results showed that the core amoebae detected by the four primers differed (Fig. 3). Nine species were detected as core amoebae by all of the four primer pairs. The most dominant amoeba identified by full-length, amoeba-specific, and V4 primers was *Leptophryidae* X sp., whereas the most dominant species identified by the V9 primer was *Mastigamoeba* sp. (Fig. 3). We found all the core amoebae in the amoeba-specific or V4 datasets could also be found in the other primer datasets. The V9 dataset had some particular core amoebae which were absent in the other datasets. Therefore, we blasted the sequences of those particular core amoebae in the V9 dataset against the full-length dataset. We found that most of these species, except *Acanthamoeba castellanii,* could be matched with the full-length 18S rRNA gene region (Table 3). Surprisingly, some of the annotations from V9 primers were incorrect (Table 3). For instance, the sequences of *Korotnevella* sp., *Platyreta germanica*, or

Table 1
The primer pair influence the detection of the amoeba community.

Taxonomy	PR[2]			
	Full-length	Amoeba-specific	V4	V9
• Amoebozoa	✓	✓	✓	✓
• Excavata	✓	✓	×	✓
Heterolobosea (C)				
• Opisthokonta	×	×	×	✓
Nucleariа (G)				
Fonticula (G)	×	×	×	×
• Rhizaria	✓	×	✓	✓
Foraminifera (P)				
Vampyrellida (F)	✓	✓	✓	✓
Euglyphida (O)	✓	✓	✓	✓
Trivalvulariida (O)	×	×	×	×
Rhogostoma (G)	×	×	✓	×
Pseudodifflugia (G)	✓	×	×	×
Frenzelina (G)	×	×	×	×
Diaphoropodon (G)	×	×	×	×
Gymnophrys (G)	×	×	×	×
Kraken (G)	×	×	×	×
Reticulomyxa (G)	×	×	×	×

P, phylum; C, Class; O, Order; F, Family; G, Genus. √, detected; ×, not detected.

Table 2
Number of taxa at each taxonomic rank of amoebae identified using different primers.

Primer	Supergroup	Phylum	Class	Order	Family	Genus	Species
Full-length	3	6	13	21	35	46	53
Amoeba-specific	3	6	13	19	37	42	46
V4	2	6	14	22	50	65	77
V9	4	7	17	27	45	47	59

Tetramitus entericus were classified into different phyla or genera in V9 and full-length datasets (Table 3).

As pathogenic amoebae are increasingly considered an emerging health risk, we also compared the ability of these four primer pairs to detect potential pathogenic amoebae. The four primer pairs identified five potential pathogenic species, but no single primer pair could detect all of these species (Table 4). Only *Entamoeba moshkovskii* could be detected by all the four primer pairs. The V9 primer pair could detect four species, while other primer pairs could detect three species (Table 4).

3.4. Phylogenetic relationships among detected amoebae

The phylogenetic relationships among amoebae identified by the primers were compared. The results showed that the model and method of phylogenetic analysis could not completely separate the amoebae into different clades according to their taxonomy at phylum level (Fig. 4). However, the branches in the phylogenetic tree based on the V9 dataset were significantly more interlaced than those in trees based on other datasets (Fig. 4d). The phylogenetic trees based on sediment samples contained more diverse branches than those based on water samples. In the amoeba-specific and V4 phylogenetic trees, the UI13E03-lineage X in amoebozoa X was close to Lobosa (Fig. 4c). We also found that the V9 phylogenetic tree had several unclassified branches that had no close relationships with known taxa (Fig. 4d).

3.5. Primer choice affects the observed pattern of amoeba community assembly

The amoeba communities in water and sediment detected by each of the four primer pairs fit well with the Sloan neutral community model (Fig. 5). The fit of the model for the sediment communities was more conserved than that for the water communities. More than 90% of OTUs in water and sediment could be predicted as neutrally distributed with all the primer pairs (Fig. S3). The higher the estimated migration rate (m) represented, the less dispersal limitation. The relatively low estimated migration rate in the sediment based on the V9 dataset indicated that the contribution of dispersal limitation increased (Fig. 5). The null model further showed that stochastic processes mainly governed the amoeba populations in both water and sediment (Fig. 6). Consistent with the neutral model results, we found that the amoebae in water samples with full-length primers had a relatively high heterogeneous selection, and the amoeba communities in sediment samples detected with V9 primers had a relatively high percentage of dispersal limitation (Fig. 6). Altogether, all the primer pairs indicated stochastic processes mainly governed the amoebae, but the details of the ecological processes were subject to different biases according to the primer pair.

4. Discussion

We compared the four common primer pairs to reveal amoeba community diversity, composition and assembly, core amoeba species, potentially pathogenic amoebae, and phylogenetic relationships of amoebae. This study showed that the choice of 18S rRNA gene primers influenced the taxonomic coverage and species identity of the amoebae. Short-read amplicon sequencing primers, especially the V9, had biases in revealing the amoeba community profile and composition. In our study, no single primer pair could efficiently detect all the potentially pathogenic amoebae. Full-length 18S rRNA gene amplicon sequencing provides advantages in taxonomy coverage and elucidating phylogenetic relationships compared to short-read sequencing. There was no apparent difference between stochasticity and determinism as drivers of community assembly, but the underlying ecological processes were subject to bias according to the primer pair used.

4.1. Full-length primers provide better taxonomic coverage and community profiles, and more robust phylogenetic relationships than other primers

We found amoeba-specific and V4 primers had a shortcoming in detecting the Foraminifera and Heterolobosea, respectively. Our results were consistent with the previous studies that reported Heterolobosea

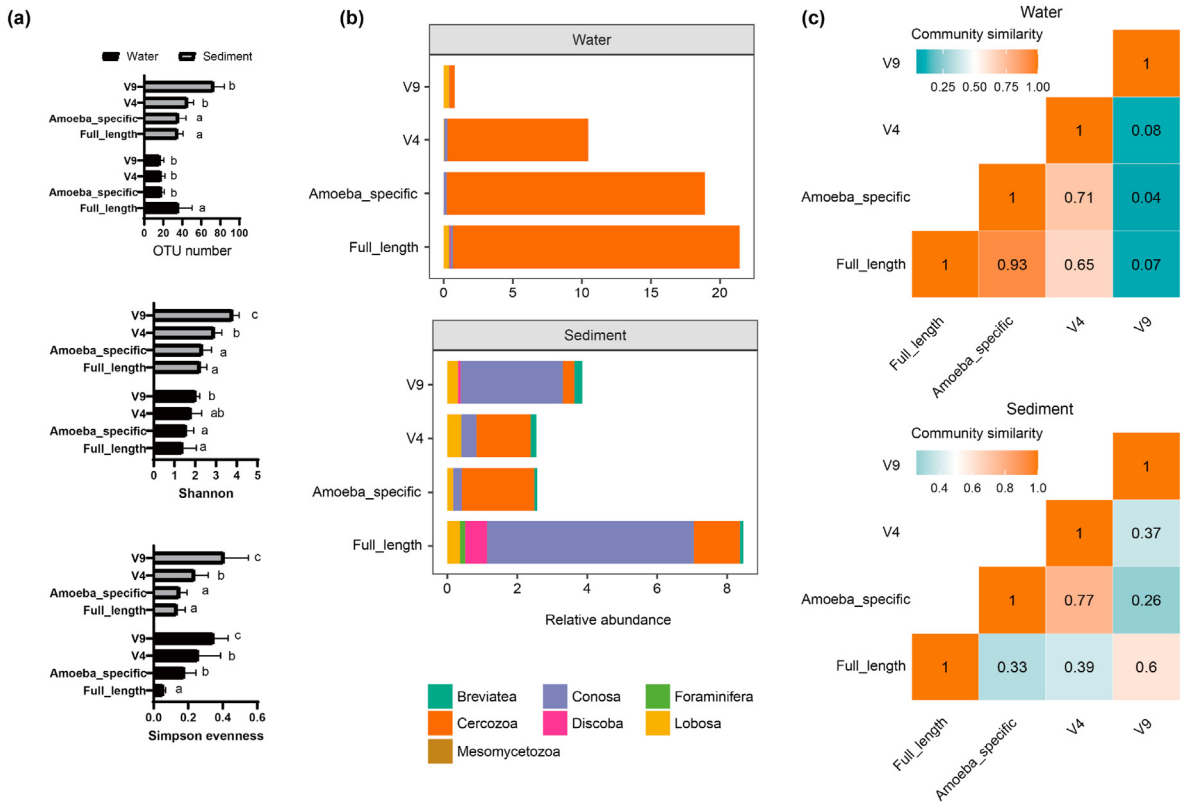

Fig. 1. Alpha diversity and community composition of amoeba communities revealed by different primers. Different letters indicate statistical differences at 0.05 significance level (Kruskal-Wallis test). (**a**) Alpha diversity in water and sediment samples. (**b**) Amoeba community at the phylum level. (**c**) Community similarity at the phylum level.

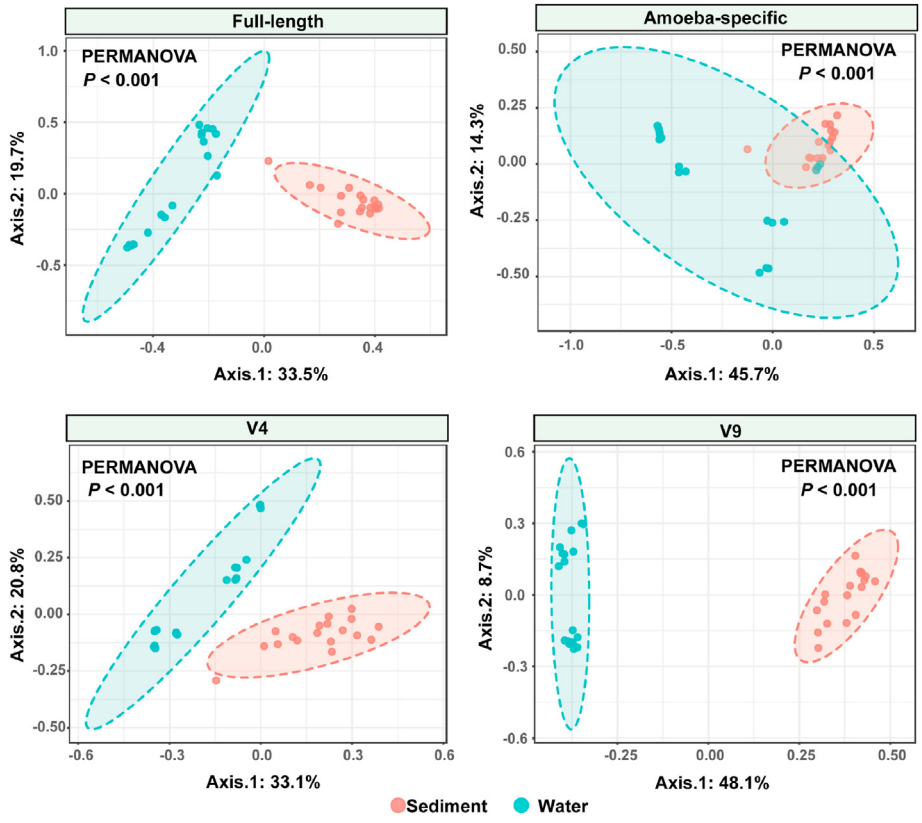

Fig. 2. Principal component analysis (PCA) on amoebae communities revealed by different primers. The dashed lines represent ellipses with 95% confidence.

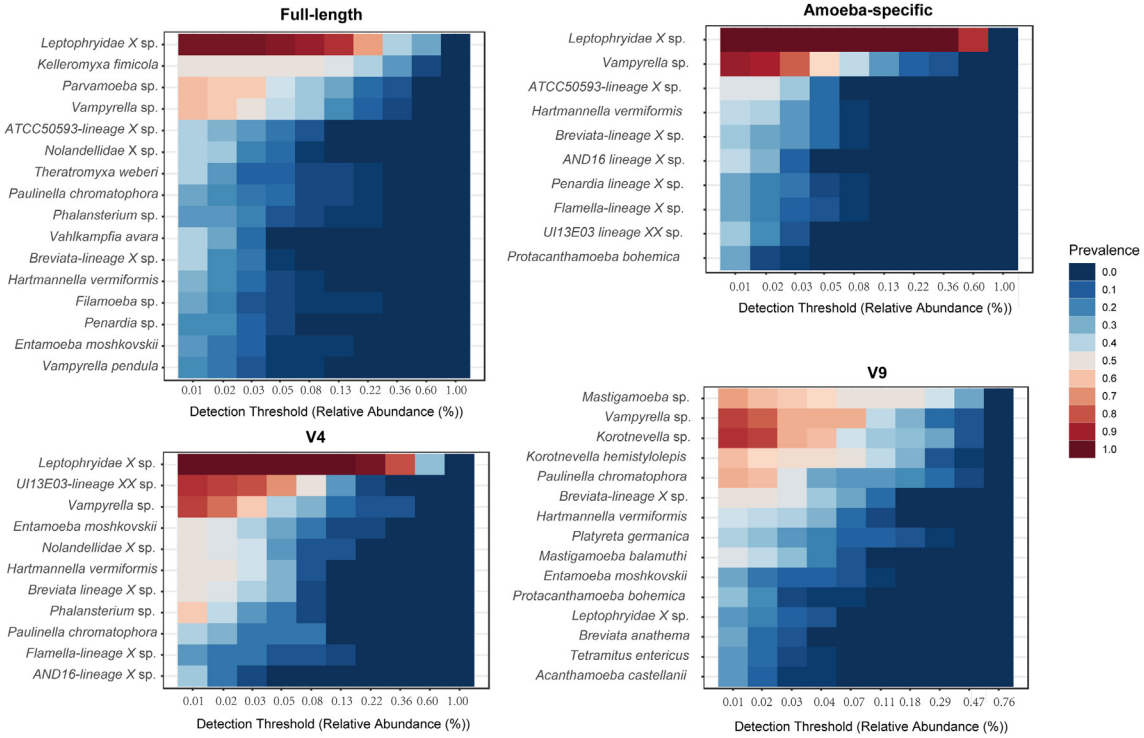

Fig. 3. Core-microbiome in amoeba communities revealed by different primers. These amoebae were detected in a high fraction of all the samples above a given abundance threshold.

(phylum Discoba) is frequently absent in the microbial eukaryotic community when V4 primer pair is used (Choi and Park, 2020; Maritz et al., 2019; Salmaso et al., 2020). The full-length and V9 primers could detect the majority of amoebae, but the species classification ratio (the proportion of sequences that are assigned to a species) of the V9 dataset was lower than that of the other primer datasets. It can be explained that the PR^2 reference database includes only 29.4% of full-length sequences, 63.7% of sequences including the V4 region, and 12.1% of sequences including the V9 region (Guillou et al., 2013). Therefore, we concluded that the full-length primer pair could detect and identify amoebae in the environment more comprehensively than short-read amplicon sequences.

When analyzing the eukaryotic community, the V9 region usually provides about 20% more OTUs than the V4 region (Choi and Park, 2020; Tragin et al., 2018). However, our results indicated that V9 primers detected more OTUs but fewer amoebae species than V4 primers. Also, the underestimation of amoeboid representatives of Cercozoa by V9 led to its amoeba community structure being quite different from others in both water and sediment samples. In previous studies, Cercozoa also had a relatively low abundance (Ren et al., 2018) but their relative abundance did not differ significantly between V4 and V9 datasets (Stoeck et al., 2010). We found that V9 had a biased amplification of *Korotnevella* sp. (Lobosa) and *Mastigamoeba* sp. (Conosa), and may therefore also have a biased amplification of Cercozoa. Moreover, full-length primers could detect *Theratromyxa weberi* and *Kelleromyxa fimicola*, which were absent in the other primer datasets. The community composition profiles of amoebae in sediment were more similar among the four primers than those for amoebae in water. As the amoeba diversity was higher in sediment than in water, this indicates that the application of short-read amplicons in environments with low amoeba diversity, such as drink water systems, might be more susceptible to producing high variations in the amoeba community diversity and composition. Overall, full-length reads provide better amoeba community profiles than short ones, especially the V9 region of the 18S rRNA gene.

Until now, the 18S rRNA gene is the most widely used marker for single-gene phylogenetic analyses of amoebae (Samba-Louaka et al., 2019). A previous study showed that V4 and V9 region sequences could not adequately resolve phylogenetic relationships at the highest-rank taxonomic group level (Choi and Park, 2020). Our study revealed that the molecular phylogenetic trees with all primers form paraphyletic or polyphyletic groups at the phylum level. This suggests that the evolution of amoebae is complicated, and the current classification method may not fully reflect their phylogenetic relationships. Generally, the V9 region is too short of revealing their phylogenetic relationships (Geisen et al., 2019). In our study, the branches from the same family did not cluster together in the V9 tree, which is consistent with the findings of Choi and Park (2020). Moreover, the Vampyrellidae were located far away from Leptophryidae and Paulinellidae in the full-length primer tree, but these families clustered together in the amoeba-specific and V4 trees. In addition, it is possible to employ a phylogenetic approach to classify sequences taxonomically and obtain a robust evolutionary framework of environmental samples (Jamy et al., 2020). In our study, NAMAKO-1-lineage XX had a close relationship with Vampyrellidae, and UI13E03-lineage X was closed to Parvamoebidae in the full-length, amoeba-specific and V4 trees. It helps to understand the ecological role of these amoebae without their exactly taxonomical information. We can also deduce a large number of unclassified amoebae to *Kelleromyxa fimicola* (order: Stemonitales-Physarales) using the phylogenetic relationships constructed by full-length sequences of 18S rRNA gene. It can thus be concluded that short amplicon sequences, which were mainly conducted by second-generation sequencing platforms, had a shortcoming in revealing phylogenetic relationships of amoebae.

4.2. *Short read sequences are defective in the species identification*

Precise species identification of amoebae from environmental samples is essential to evaluating their health risk. The core amoebae identified in the amoeba-specific and V4 datasets could also be detected in the full-length and V9 datasets, indicating that all the primers could detect the majority of dominant amoebae. However, the composition of the most dominant amoebae in the full-length, amoeba-specific and V4 datasets differed from that in the V9 dataset. For example, the most

Table 3
Mismatched core amoebae identified by the V9 primers.
For those species identified by the V9 primers that were matched with a different genus in the full-length dataset (indicated by the grey background), the full rank taxonomy is presented.

V9	Full-length	Identities	Gaps	Expect
Mastigamoeba sp. (OTU_11058)	Mastigamoeba sp. (OTU_10387)	144/151 (95%)	0/151 (0%)	1e-63
Mastigamoeba sp. (OTU_167)	Mastigamoeba balamuthi (OTU_16281)	143/151 (95%)	0/151 (0%)	5e-62
Amoebozoa Lobosa Discosea-Flabellinia Dactylopodida Paramoebidae *Korotnevella* *Korotnevella* sp. (OTU_10505)	Alveolata Ciliophora Ciliophora X Ciliophora XX Ciliophora XXX *Mesodinium* *Mesodinium pulex* (OTU_4454)	139/144 (97%)	0/144 (0%)	3e-63
Amoebozoa Lobosa Discosea-Flabellinia Dactylopodida Paramoebidae *Korotnevella* *Korotnevella* sp. (OTU_260)	Alveolata Ciliophora Ciliophora X Ciliophora XX Ciliophora XXX *Mesodinium* *Mesodinium pulex* (OTU_4454)	144/144 (100%)	0/144 (0%)	2e-71
Korotnevella hemistylolepis (OTU_4123)	*Korotnevella hemistylolepis* (OTU_2279)	142/144 (99%)	0/144 (0%)	3e-68
Korotnevella hemistylolepis (OTU_428)	*Korotnevella hemistylolepis* (OTU_12132)	144/144 (100%)	0/144 (0%)	2e-71
Mastigamoeba balamuthi (OTU_528)	*Mastigamoeba balamuthi* (OTU_16281)	139/152 (91%)	2/152 (1%)	1e-53
Rhizaria Cercozoa Endomyxa Vampyrellida Leptophryidae *Platyreta* *Platyreta germanica* (OTU_295)	Rhizaria Cercozoa Endomyxa Vampyrellida Leptophryidae *Theratromyxa* *Theratromyxa weberi* (OTU_16885)	140/140 (100%)	0/140 (0%)	2e-69
Rhizaria Cercozoa Endomyxa Vampyrellida Leptophryidae *Platyreta* *Platyreta germanica* (OTU_295)	Alveolata Ciliophora Spirotrichea Tintinnida TIN 03 *Codonella* TIN 03 *Codonella cratera* (OTU_4686)	140/140 (100%)	0/140 (0%)	2e-69
Breviata anathema (OTU_6795)	*Breviata anathema* (OTU_2570)	133/137 (97%)	0/137 (0%)	5e-61
Excavata Discoba Heterolobosea Heterolobosea X Vahlkampfiidae *Tetramitus* *Tetramitus entericus* (OTU_852)	Excavata Discoba Heterolobosea Heterolobosea X Vahlkampfiidae *Learamoeba* *Learamoeba waccamawensis* (OTU_987)	124/124 (100%)	0/124 (0%)	2e-60

For those species identified by the V9 primers that were matched with a different genus in the full-length dataset (indicated by the grey background), the full rank taxonomy is presented.

dominant amoeba *Mastigamoeba* sp. (Conosa) in the V9 dataset was not recovered as a core amoeba in other primer datasets. These results are consistent with the finding that V9 had an unbalanced bias between Amoebozoa (Lobosa and Conosa) and Rhizaria (Cercozoa) supergroups. Furthermore, some species were unique to the V9 dataset, so the variations of relative abundance could result from the bias of primer preferences. First, it can be explained that some of the amoebae uniquely detected in the V9 dataset were the results of misclassifications with

Table 4
Detection of common pathogenic amoebae using different primers.

Pathogenic amoebae	Full-length	Amoeba-specific	V4	V9
Naegleria fowleri	-	-	-	-
Naegleria australiensis	+	+	-	+
Entamoeba histolytica	-	-	-	-
Entamoeba moshkovskii	+	+	+	+
Entamoeba gingivalis	-	-	-	-
Entamoeba invadens	-	-	-	-
Balamuthia mandrillaris	-	-	+	+
Vermamoeba vermiformis	-	-	+	-
Acanthamoeba	+	-	-	+
Sappinia	-	-	-	-

"+" indicates the amoeba was detected; "–" indicates the amoeba was not detected.

short sequences of 16S rRNA gene amplicons (Lam et al., 2020). The other reason may be the low representation of V9 region sequences in the PR^2 reference database (Guillou et al., 2013) which may compromise the accuracy of taxonomic classification. However, some species were unique to the full-length dataset, suggesting that the commonly used short sequences of 18S rRNA gene amplicons have a shortcoming in species identification. Our study further showed that the profiles of pathogenic amoebae were inconsistent among the four primer datasets, indicating that their detection is highly dependent on the primer choice. Short-read amplicon sequencing is the most inexpensive taxonomic profiling method, which is also the best-suited method for probing rare community variants with the great sequence depth per sample provided (Callahan et al., 2019). Therefore, we recommend using full-length and V4 primers in combination for the health risk evaluation of environmental samples, since this should detect most pathogenic amoebae.

4.3. Minor influence of primers choice on the beta diversity and community assembly

In a previous 16S rRNA gene sequencing study, Tremblay et al. (2015) found that beta diversity metrics were robust to both primer and sequencing platform biases. Moreover, Bradley et al. (2016) indicated that both the V4 and the V8–V9 regions could differentiate microbial eukaryotic communities in freshwater, coastal, and wastewater samples. In this study, we found that the beta diversity of amoeba communities in water and sediment samples was not affected by primer choice. Besides beta diversity, there is increasing interest in evaluating the contributions of deterministic and stochastic processes in microbial community assembly (Zhou and Ning, 2017). According to previous studies, stochastic processes are more critical than deterministic processes in driving the amoebae community structure in aquatic environments (Ren et al., 2018; Wang et al., 2020). All the primers indicated that stochastic processes were more critical in the amoeba community assembly based on the neutral and null model analyses in this study.

In the present study, it was found that the specific ecological processes of community assembly in water and sediment samples revealed by full-length and V9 primers differed from other primers in null models. The neutral model was constructed based only on OTU information, i.e., without taxonomic or phylogenetic information. However, the null

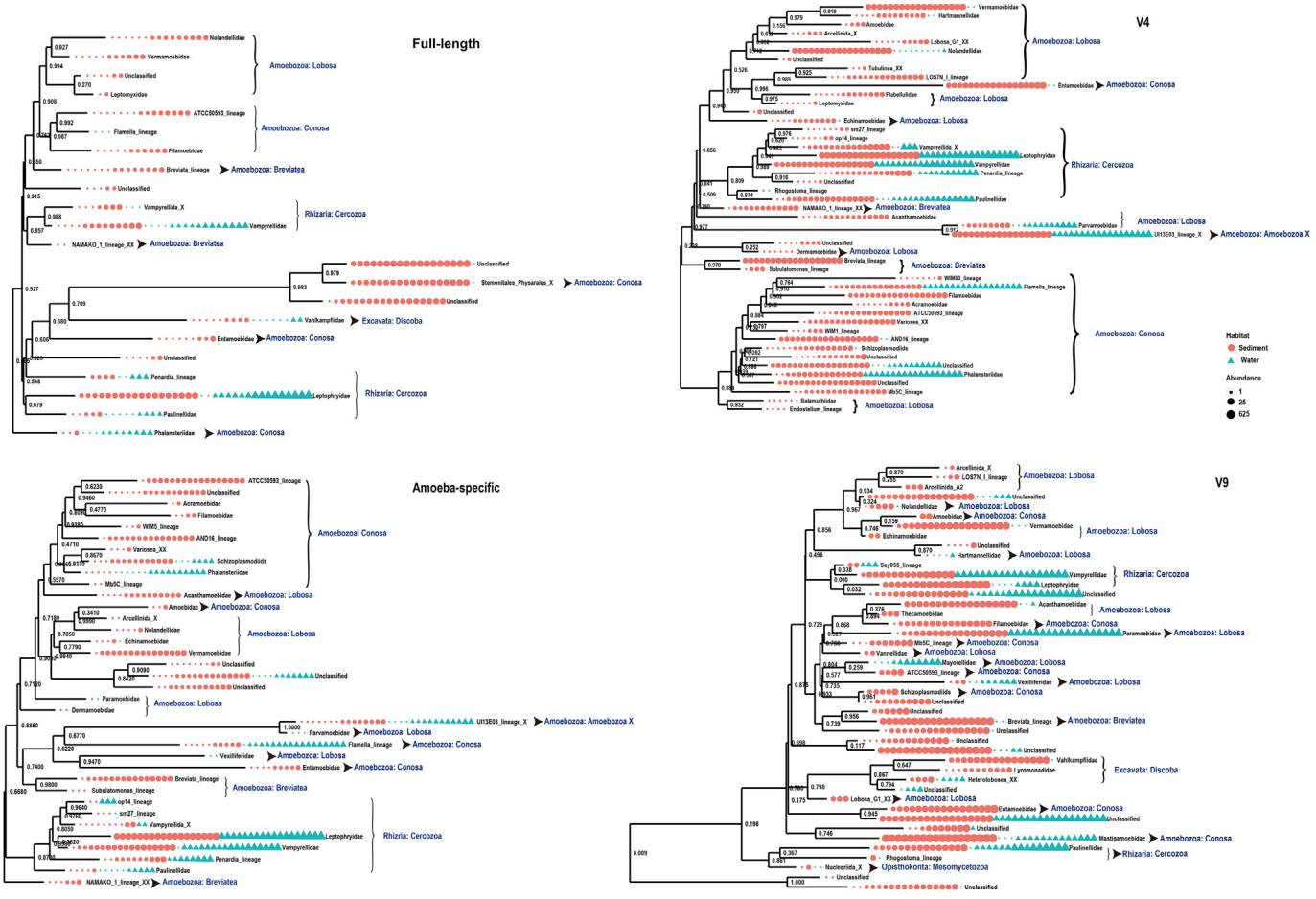

Fig. 4. Phylogenetic tree of amoebae at family level. The circles and triangles represent the sediment and water samples, respectively. The size of the circles and triangles is proportional to the number of sequence reads.

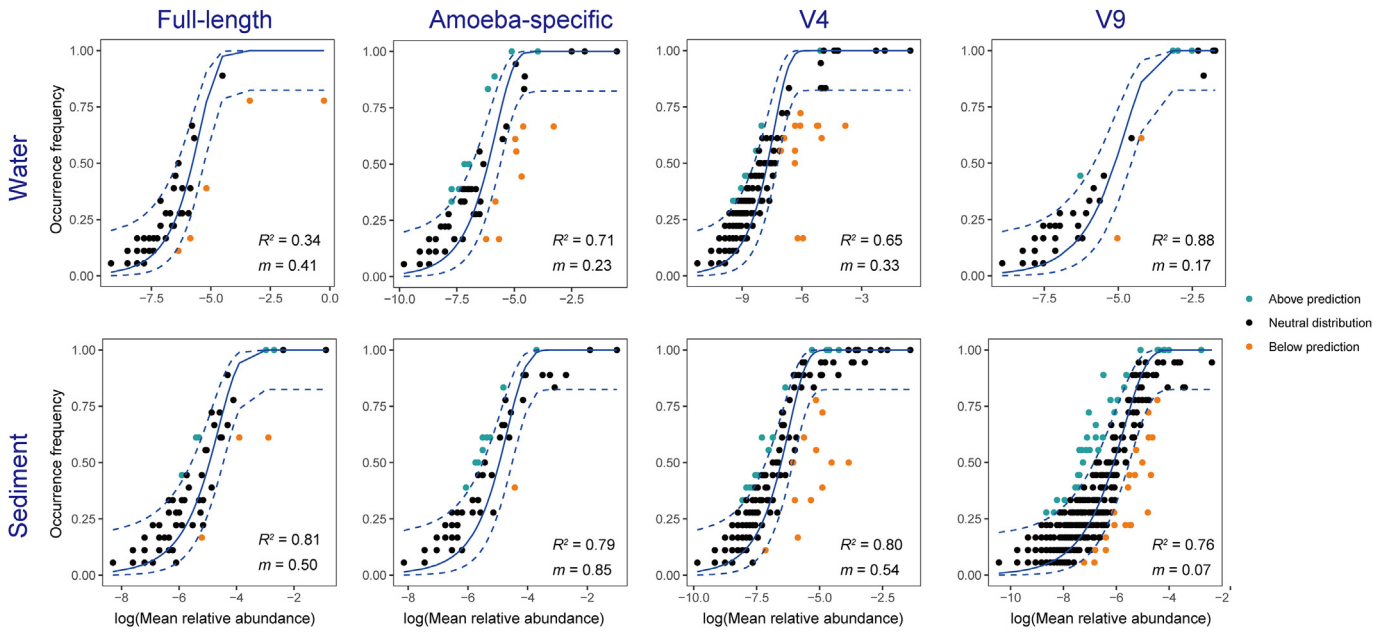

Fig. 5. Fit of the neutral models for the amoeba communities. The OTUs that occurred more frequently than predicted by the model are shown in blue, while those that occurred less frequently than predicted are shown in orange. Blue dashed lines represent 95% confidence intervals around the model prediction and those OTUs that fall within the confidence intervals are considered to be neutrally distributed. The goodness of fit of the neutral model (R^2) and the estimated migration rate (m) are also presented.

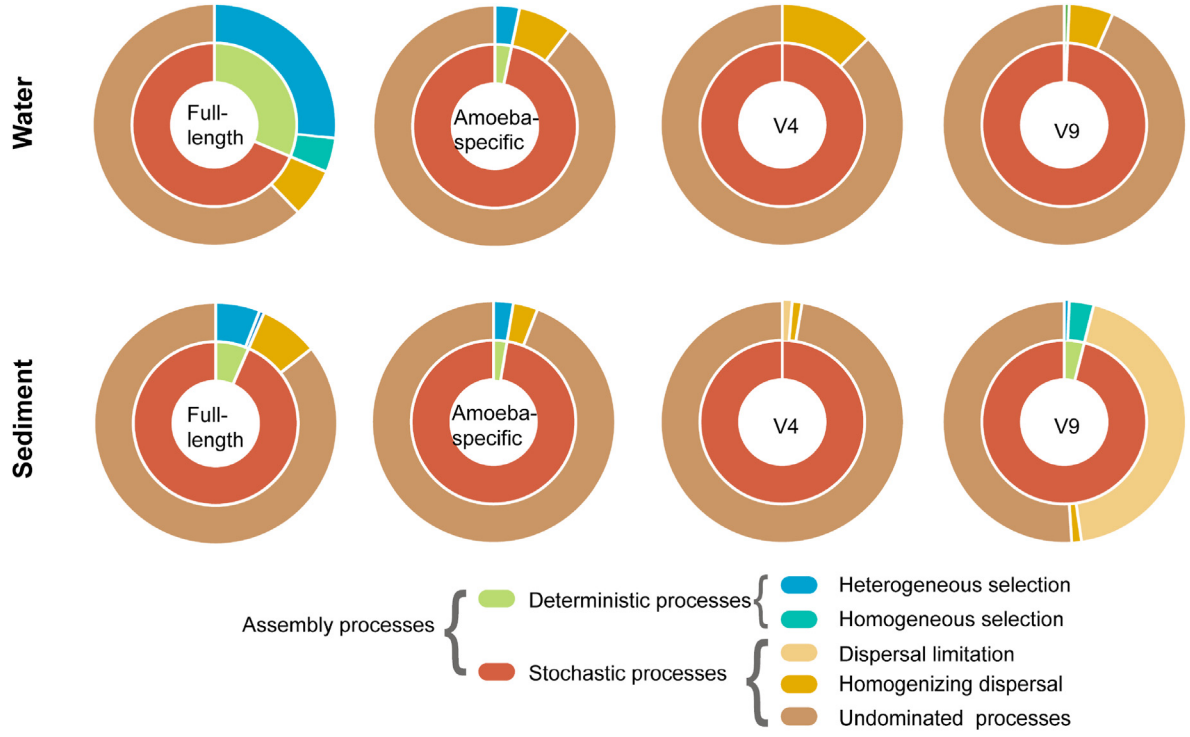

Fig. 6. Quantification of the ecological processes that drive the assembly of amoeba communities. The inner circle quantifies the percentages of deterministic and stochastic processes. The outer circle quantifies the percentages of heterogeneous selection, homogeneous selection, dispersal limitation, homogenizing dispersal, and undominated processes.

model construction needs both OTU information and phylogenetic trees (Jiao et al., 2020; Stegen et al., 2013). Therefore, if we used the neutral model to reveal community assembly mechanisms, we need to pay more attention to the OTU clustering method and sequence depth. If we further use the neutral model to quantify specific ecological processes (e.g., homogeneous selection, heterogeneous selection, homogenizing dispersal, and dispersal limitation) in determining the community assembly, we need phylogenetic trees in addition to OTU information, which may lead to a minor difference according to the choice of different primer pairs. Therefore, the interpretation of stochastic or deterministic

processes should require more caution and take account of other information, e.g., the relationship between the amoeba community and environmental parameters (Zhou and Ning, 2017).

4.4. Recommendation for primer choice in the study of amoebae

Full-length sequences by Pacbio sequencing can provide better taxonomic and phylogenetic information, but the sequencing depth is not as high as second-generation sequencing. We suggest combining second and third-generation sequencing technologies in case of the missed information from rare species. Hugerth et al. (2014) compared the primer pairs V4–V5, V5–V7, V7, and V7–V8, and found those targeting the V4 region performed best for profiling eukaryotic communities. Geisen et al. (2019) recently concluded that primers targeting the V4 region of 616–1132 performed best after comparing the widely used primers to study protist diversity. V9 primers may lead to a mismatch in taxonomy, and pyrosequencing is no longer commercially available for amoeba-specific primers. Furthermore, third-generation sequencing technology, e.g. PacBio Sequencing, is not cost-effective if it provides sequencing depth that is equal to second-generation sequencing for a given sample. Therefore, we suggest using full-length and V4 primers in combination in amplicon-based high-throughput sequencing to investigate amoeba diversity from environmental samples.

5. Conclusions

This study showed that full-length amplicon sequencing covers more phylogenetic lineages of amoeboid protists and provides reliable species identification. In contrast, short read amplicon sequences generated by second-generation sequencing give higher community diversity and have advantages in detecting rare amoebae species. As the PacBio sequencing of full-length 18S rRNA gene can not fully detect all the pathogenetic amoebae, we recommend using PacBio full-length read sequencing and V4 region Illumina sequencing in combination to identify potentially pathogenetic amoebae in the environment to reduce human health risks.

Author contributions

Xiafei Zheng: Conceptualization, Methodology, Visualization, Original draft preparation, Investigation. Zhili He: Writing- Reviewing and Editing. Cheng Wang: Writing- Reviewing and Editing. Qingyun Yan: Writing- Reviewing and Editing. Longfei Shu: Conceptualization, Writing- Reviewing and Editing.

Funding

This material is based upon work supported by the National Natural Science Foundation of China (31970384, 41907021, 21806044, 92051120, 31802350), the Fundamental Research Funds for the Central Universities Sun Yat-sen University (22lgqb22, 19lgzd28), the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311021006), the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (SML2021SP203), the Guangdong Natural Resources Department Contract (GDNRC[2021]62), and Guangzhou Basic and Applied Basic Research Foundation (202102020257).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.watbs.2022.100057.

References

Adl, S.M., Bass, D., Lane, C.E., Lukes, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., et al., 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J. Eukaryot. Microbiol. 66, 4–119. https://doi.org/10.1111/jeu.12691.

Anderson, O.R., 2018. A half-century of research on free-living amoebae (1965-2017): review of biogeographic, ecological and physiological studies. Acta Protozool. 57, 1–28. https://doi.org/10.4467/16890027ap.18.001.8395.

Balczun, C., Scheid, P.L., 2017. Free-living amoebae as hosts for and vectors of intracellular microorganisms with public health significance. Viruses 9, 65. https://doi.org/10.3390/v9040065.

Bartrand, T.A., Causey, J.J., Clancy, J.L., 2014. Journal of the American water works association. J. Am. Water Works Assoc. 106, E418–E432. https://doi.org/10.5942/jawwa.2014.106.0140.

Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 491–507. https://doi.org/10.1093/biomet/93.3.491.

Bradley, I.M., Pinto, A.J., Guest, J.S., 2016. Design and evaluation of illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. Appl. Environ. Microbiol. 82, 5878–5891. https://doi.org/10.1128/aem.01630-16.

Burns, A.R., Stephens, W.Z., Stagaman, K., Wong, S., Rawls, J.F., Guillemin, K., Bohannan, B.J.M., 2016. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. ISME J. 10, 655–664. https://doi.org/10.1038/ismej.2015.142.

Buse, H.Y., Lu, J., Lu, X., Mou, X., Ashbolt, N.J., 2014. Microbial diversities (16S and 18S rRNA gene pyrosequencing) and environmental pathogens within drinking water biofilms grown on the common premise plumbing materials unplasticized polyvinylchloride and copper. FEMS Microbiol. Ecol. 88, 280–295. https://doi.org/10.1111/1574-6941.12294.

Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., McGill, S.K., Dougherty, M.K., 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res. 47, e103. https://doi.org/10.1093/nar/gkz569.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R., 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26, 266–267. https://doi.org/10.1093/bioinformatics/btp636.

Choi, J., Park, J.S., 2020. Comparative analyses of the V4 and V9 regions of 18S rDNA for the extant eukaryotic community using the Illumina platform. Sci. Rep. 10, 6519. https://doi.org/10.1038/s41598-020-63561-z.

Chong, J., Liu, P., Zhou, G., Xia, J., 2020. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat. Protoc. 15, 799–821. https://doi.org/10.1038/s41596-019-0264-1.

Cooper, A.M., Aouthmany, S., Shah, K., Rega, P.P., 2019. Killer amoebas: primary amoebic meningoencephalitis in a changing climate. Off. J. Am. Acad. Physician Assistants (JAAPA) 32. https://doi.org/10.1097/01.JAA.0000558238.99250.4a.

Countway, P.D., Gast, R.J., Savai, P., Caron, D.A., 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. J. Eukaryot. Microbiol. 52, 95–106. https://doi.org/10.1111/j.1550-7408.2005.05202006.x.

Delafont, V., Bouchon, D., Hechard, Y., Moulin, L., 2016. Environmental factors shaping cultured free-living amoebae and their associated bacterial community within drinking water network. Water Res. 100, 382–392. https://doi.org/10.1016/j.watres.2016.05.044.

Delafont, V., Brouke, A., Bouchon, D., Moulin, L., Héchard, Y., 2013. Microbiome of free-living amoebae isolated from drinking water. Water Res. 47, 6958–6965. https://doi.org/10.1016/j.watres.2013.07.047.

Delafont, V., Perrin, Y., Bouchon, D., Moulin, L., Héchard, Y., 2019. Targeted metagenomics of microbial diversity in free-living amoebae and water samples. In: Buchrieser, C., Hilbi, H. (Eds.), Legionella: Methods and Protocols. Springer New York, New York, NY, pp. 421–428.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

Dupuy, M., Mazoua, S., Berne, F., Bodet, C., Garrec, N., Herbelin, P., Ménard-Szczebara, F., Oberti, S., Rodier, M.-H., Soreau, S., et al., 2011. Efficiency of water disinfectants against *Legionella pneumophila* and *Acanthamoeba*. Water Res. 45, 1087–1094. https://doi.org/10.1016/j.watres.2010.10.025.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998. https://doi.org/10.1038/nmeth.2604.

Garner, E., McLain, J., Bowers, J., Engelthaler, D.M., Edwards, M.A., Pruden, A., 2018. Microbial ecology and water chemistry impact regrowth of opportunistic pathogens in full-scale reclaimed water distribution systems. Environ. Sci. Technol. 52, 9056–9068. https://doi.org/10.1021/acs.est.8b02818.

Geisen, S., Vaulot, D., Mahé, F., Lara, E., de Vargas, C., Bass, D., 2019. A User Guide to Environmental Protistology: Primers, Metabarcoding, Sequencing, and Analyses. bioRxiv, 850610. https://doi.org/10.1101/850610.

Goldstein, S.T., 2003. Foraminifera: a biological overview. In: Modern Foraminifera. Springer Netherlands, Dordrecht, pp. 37–55.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., et al., 2013. The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41, D597–D604. https://doi.org/10.1093/nar/gks1160.

He, Z., Wang, L., Ge, Y., Zhang, S., Tian, Y., Yang, X., Shu, L., 2021. Both viable and inactivated amoeba spores protect their intracellular bacteria from drinking water

disinfection. J. Hazard Mater. 417, 126006. https://doi.org/10.1016/j.jhazmat.2021.126006.
He, Z., Zheng, N., Zhang, L., Tian, Y., Hu, Z., Shu, L., 2022. Efficient inactivation of intracellular bacteria in dormant amoeba spores by FeP. J. Hazard Mater. 425, 127996. https://doi.org/10.1016/j.jhazmat.2021.127996.
Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D., Wilmes, P., Andersson, A.F., 2014. Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. PLoS One 9. https://doi.org/10.1371/journal.pone.0095567.
Jamy, M., Foster, R., Barbera, P., Czech, L., Kozlov, A., Stamatakis, A., Bending, G., Hilton, S., Bass, D., Burki, F., 2020. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. Mol. Ecol. Resour. 20, 429–443. https://doi.org/10.1111/1755-0998.13117.
Jiao, S., Yang, Y., Xu, Y., Zhang, J., Lu, Y., 2020. Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. ISME J. 14, 202–216. https://doi.org/10.1038/s41396-019-0522-9.
Kong, Y., 2011. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics 98, 152–153. https://doi.org/10.1016/j.ygeno.2011.05.009.
Lam, T.Y.C., Mei, R., Wu, Z., Lee, P.K.H., Liu, W.T., Lee, P.H., 2020. Superior resolution characterisation of microbial diversity in anaerobic digesters using full-length 16S rRNA gene amplicon sequencing. Water Res. 178, 115815. https://doi.org/10.1016/j.watres.2020.115815.
Magoc, T., Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507.
Maritz, J.M., Ten Eyck, T.A., Elizabeth Alter, S., Carlton, J.M., 2019. Patterns of protist diversity associated with raw sewage in New York City. ISME J. 13, 2750–2763. https://doi.org/10.1038/s41396-019-0467-z.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17, 10–12. https://doi.org/10.14806/ej.17.1.200.
Miller, H.C., Wylie, J.T., Kaksonen, A.H., Sutton, D., Puzon, G.J., 2018. Competition between *Naegleria fowleri* and free living amoeba colonizing laboratory scale and operational drinking water distribution systems. Environ. Sci. Technol. 52, 2549–2557. https://doi.org/10.1021/acs.est.7b05717.
Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26, 1641–1650. https://doi.org/10.1093/molbev/msp077.
R Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from. http://www.R-project.org.
Ren, K.X., Xue, Y.Y., Ronn, R., Liu, L.M., Chen, H.H., Rensing, C., Yang, J., 2018. Dynamics and determinants of amoeba community, occurrence and abundance in subtropical reservoirs and rivers. Water Res. 146, 177–186. https://doi.org/10.1016/j.watres.2018.09.011.
Salmaso, N., Boscaini, A., Pindo, M., 2020. Unraveling the diversity of eukaryotic microplankton in a large and deep perialpine lake using a high throughput sequencing approach. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.00789.
Samba-Louaka, A., Delafont, V., Rodier, M.H., Cateau, E., Hechard, Y., 2019. Free-living amoebae and squatters in the wild: ecological and molecular features. FEMS Microbiol. Rev. 43, 415–434. https://doi.org/10.1093/femsre/fuz011.
Shirley, D.-A.T., Farr, L., Watanabe, K., Moonah, S., 2018. A Review of the global burden, new diagnostics, and current therapeutics for amebiasis. Open Forum Infect. Dis. 5. https://doi.org/10.1093/ofid/ofy161.
Shi, Y., Queller, D.C., Tian, Y., Zhang, S., Yan, Q., He, Z., He, Z., Wu, C., Wang, C., Shu, L., 2021. The ecology and evolution of amoeba-bacterium interactions. Appl. Environ. Microbiol. 87. https://doi.org/10.1128/aem.01866-20 e01866– e012022.
Shu, L., Brock, D.A., Geist, K.S., Miller, J.W., Queller, D.C., Strassmann, J.E., DiSalvo, S., 2018a. Symbiont location, host fitness, and possible coadaptation in a symbiosis between social amoebae and bacteria. Elife 7, e42660. https://doi.org/10.7554/eLife.42660.
Shu, L., Zhang, B., Queller, D.C., Strassmann, J.E., 2018b. Burkholderia bacteria use chemotaxis to find social amoeba Dictyostelium discoideum hosts. ISME J. 12, 1977–1993. https://doi.org/10.1038/s41396-018-0147-4.
Shu, L., He, Z., Guan, X., Yang, X., Tian, Y., Zhang, S., Wu, C., He, Z., Yan, Q., Wang, C., Shi, Y., 2021. A dormant amoeba species can selectively sense and predate on different soil bacteria. Funct. Ecol. 35, 1708–1721. https://doi.org/10.1111/1365-2435.13824.
Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., Curtis, T.P., 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ. Microbiol. 8, 732–740. https://doi.org/10.1111/j.1462-2920.2005.00956.x.
Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., Rockhold, M.L., Konopka, A., 2013. Quantifying community assembly processes and identifying features that impose them. ISME J. 7, 2069–2079. https://doi.org/10.1038/ismej.2013.93.
Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., Richards, T.A., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. 19, 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x.
Storey, M.V., Winiecka-Krusnell, J., Ashbolt, N.J., Stenstrom, T.A., 2004. The efficacy of heat and chlorine treatment against thermotolerant Acanthamoebae and Legionellae. Scand. J. Infect. Dis. 36, 656–662. https://doi.org/10.1080/00365540410020785.
Strassmann, J.E., Shu, L., 2017. Ancient bacteria–amoeba relationships and pathogenic animal bacteria. PLoS Biol. 15, e2002460. https://doi.org/10.1371/journal.pbio.2002460.
Thomas, J.M., Ashbolt, N.J., 2011. Do free-living amoebae in treated drinking water systems present an emerging health risk? Environ. Sci. Technol. 45, 860–869. https://doi.org/10.1021/es102876y.
Thomas, V., Herrera-Rimann, K., Blanc, D.S., Greub, G., 2006. Biodiversity of amoebae and amoeba-resisting bacteria in a hospital water network. Appl. Environ. Microbiol. 72, 2428–2438. https://doi.org/10.1128/aem.72.4.2428-2438.2006.
Tragin, M., Zingone, A., Vaulot, D., 2018. Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. Environ. Microbiol. 20, 506–520. https://doi.org/10.1111/1462-2920.13952.
Tremblay, J., Singh, K., Fern, A., Kirton, E.S., He, S.M., Woyke, T., Lee, J., Chen, F., Dangl, J.L., Tringe, S.G., 2015. Primer and platform effects on 16S rRNA tag sequencing. Front. Microbiol. 6. https://doi.org/10.3389/fmicb.2015.00771.
Tsao, H.-F., Scheikl, U., Herbold, C., Indra, A., Walochnik, J., Horn, M., 2019. The cooling tower water microbiota: seasonal dynamics and co-occurrence of bacterial and protist phylotypes. Water Res. 159, 464–479. https://doi.org/10.1016/j.watres.2019.04.028.
Visvesvara, G.S., Moura, H., Schuster, F.L., 2007. Pathogenic and opportunistic free-living amoebae: Acanthamoeba spp., Balamuthia mandrillaris, Naegleria fowleri, and Sappinia diploidea. FEMS Immunol. Med. Microbiol. 50, 1–26. https://doi.org/10.1111/j.1574-695X.2007.00232.x.
Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/Aem.00062-07.
Wang, W., Ren, K., Chen, H., Gao, X., Rønn, R., Yang, J., 2020. Seven-year dynamics of testate amoeba communities driven more by stochastic than deterministic processes in two subtropical reservoirs. Water Res. 185, 116232. https://doi.org/10.1016/j.watres.2020.116232.
Yu, H., He, Z., He, Z., Yan, Q., Shu, L., 2022. Soil amoebae affect iron and chromium reduction through preferential predation between two metal-reducing bacteria. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c08069.
Zhang, S., He, Z., Wu, C., Wang, Z., Mai, Y., Hu, R., Zhang, X., Huang, W., Tian, Y., Xia, D., Wang, C., Yan, Q., He, Z., Shu, L., 2022. Complex bilateral interactions determine the fate of polystyrene micro- and nanoplastics and soil protists: implications from a soil amoeba. Environ. Sci. Technol. 56, 4936. https://doi.org/10.1021/acs.est.1c06178.
Zhou, J.Z., Ning, D.L., 2017. Stochastic community assembly: does it matter in microbial ecology? Microbiol. Mol. Biol. Rev. 81. https://doi.org/10.1128/MMBR.00002-17 e00002-00017.