Environmental DNA

WILEY

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

ORIGINAL ARTICLE OPEN ACCESS

V9 Hypervariable Region Metabarcoding Primers for Euglenozoa and Metamonada

Jiří Novák[1] | Sebastian Cristian Treitli[1,2] | Zoltán Füssy[1,3] | Kristína Záhonová[1,4,5,6] | Blanka Hamplová[7] | Štěpánka Hrdá[7] | Vladimír Hampl[1]

[1]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Vestec, Czech Republic | [2]Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany | [3]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic | [4]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic | [5]Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic | [6]Division of Infectious Diseases, Department of Medicine, University of Alberta, Edmonton, Alberta, Canada | [7]Core Facility – Genomics, Faculty of Science, Charles University, BIOCEV, Vestec, Czech Republic

Correspondence: Jiří Novák (novakjiri9@natur.cuni.cz) | Vladimír Hampl (vlada@natur.cuni.cz)

Received: 28 June 2024 | **Revised:** 19 September 2024 | **Accepted:** 23 September 2024

Funding: This work was supported by Grantová Agentura České Republiky, 23-07277S.

Keywords: 18S-V9 | Euglenozoa | freshwater diversity | metabarcoding | Metamonada

ABSTRACT

Short amplicon sequencing is a commonly used method to study the diversity of organisms in various habitats. The hypervariable regions of the small subunit rRNA gene (18S rDNA) are the most general barcodes for eukaryotes, which can provide detailed taxonomic information across a wide range of eukaryotic diversity. However, some organisms are often missed by universal primers, which have difficulty amplifying their barcodes. In this study, specific primers were designed for the amplification of the highly diverse 18S-V9 region of the Euglenozoa and Metamonada groups. The performance of the newly designed primers—V9Eug and V9Meta—was compared with the universal V9 primer on cultured communities derived from a range of freshwater environments of the Soos Natural Reserve and the Slavkov Forest in the Czech Republic. The V9Eug primer was more specific with Euglenozoa representing 91.8% of reads and 57.0% of OTUs, while the V9Meta primer showed lower specificity with only 48.4% of reads and 19.7% of OTUs assigned to Metamonada. Both the Euglenozoa and Metamonada primer pairs significantly improved recovery of their target groups compared to the universal V9 primer pair, detecting 2.7 and 1.8 times more OTUs, respectively. These results provide a more sensitive protocol for studying the diversity of these eukaryotic taxa.

1 | Introduction

Protists represent a broad paraphyletic grouping of unicellular eukaryotes encompassing all taxa outside the well-known groups of animals, plants, and fungi (Burki et al. 2020). The environmental DNA barcoding of protist communities has become increasingly popular, providing deeper insights into the functional ecology and diversity of various environments (Aguilar et al. 2016; Amaral-Zettler et al. 2009; de Vargas et al. 2015; Edgcomb et al. 2011; Ficetola and Taberlet 2023; Geisen et al. 2014, 2018; Hugerth et al. 2014; López-García et al. 2001; Pawlowski et al. 2012; Stoeck et al. 2010). In marine ecosystems, the importance of protists stems from their spatial dynamics and diverse trophic functions (Anderson and Harvey 2020; Berdjeb et al. 2018; Edgcomb 2016; Gutiérrez-Rodríguez et al. 2022; Trefault et al. 2021). Research on soil protist microbiota has recently gained attention (Bates et al. 2012; Geisen et al. 2014, 2018; Seppey et al. 2020) due to its importance for agriculture (Liao et al. 2024; Tong et al. 2024), and the estimates show that the diversity and richness of soil protists are higher than in marine or freshwater habitats (Singer et al. 2021). In comparison with soil and marine communities, the functional aspects of protist communities in freshwater ecosystems are less

---

studied (Cruaud et al. 2019; Karlicki et al. 2024), and the literature focuses more on the taxonomic composition (Debroas, Hugoni, and Domaizon 2015; Debroas et al. 2017; Jankowska et al. 2024; Metz et al. 2022; Oikonomou et al. 2015; Slapeta, Moreira, and Lopez-Garcia 2005; Taib et al. 2013).

The extensive variety of protists is reflected in the sequence diversity of their barcodes (Amaral-Zettler et al. 2009; de Vargas et al. 2015; Flegontova, Lukeš, and Horák 2023; Pawlowski et al. 2012; Schoenle et al. 2021). As a result, some protist lineages are not efficiently detected by universal eukaryotic primers and therefore missed in environmental DNA metabarcoding surveys. This is particularly true for two taxonomic groups of the former supergroup Excavata—Euglenozoa and Metamonada (Choi and Park 2020; Singer et al. 2021). Both groups display diverse morphologies and variable life strategies. Euglenozoa encompass microbial algae, free-living heterotrophs from various environments, (endo)commensals, and parasites of typically insect and vertebrate hosts (Bicudo and Menezes 2016; Butenko et al. 2021; Kostygov et al. 2021; Kubín, Juráň, and Kučera 2024). Metamonada comprises heterotrophic flagellates that are specialized for environments with low oxygen concentrations, such as animal guts or sediments (Adl et al. 2019). Despite this, the true diversity of these two groups is still not well captured in terms of genetic variability, as evidenced by the continuous discovery of new taxa on different taxonomic levels (Lax et al. 2021; Mazancová et al. 2023; Stairs et al. 2021; Yazaki et al. 2020; Záhonová et al. 2021).

The full-length 18S rDNA sequence has traditionally been the gold standard for diversity surveys. However, the next-generation sequencing (NGS) based on Illumina technology has become popular due to its ease and affordability, leading to the prevalent use of shorter barcodes for exploring diversity (Choi and Park 2020; Edgcomb et al. 2011; Singer et al. 2021; Stoeck et al. 2010). Currently, the most common approach involves amplifying the hypervariable regions V4 (amplicon size ~200–500 bps) and V9 (amplicon size ~100–200 bps) (Amaral-Zettler et al. 2009; Choi and Park 2020; Dunthorn et al. 2012, 2014; Hadziavdic et al. 2014; Korajkic et al. 2015; Stoeck et al. 2010) followed by massively parallel sequencing. There is ongoing discussion about which part of the 18S gene is better for barcoding eukaryotes. Some studies suggest that the V4 region yields more taxonomic units (Stoeck et al. 2010), especially for certain protist groups (Dunthorn et al. 2012; Łukomska-Kowalczyk et al. 2016). However, other studies claim the opposite, demonstrating that the V9 region can provide up to 20% more OTUs (Piredda et al. 2016; Tragin, Zingone, and Vaulot 2018) and outperform the V4 region in capturing rare taxa (Choi and Park 2020). Ultimately, researchers concur on the comparable performance of both regions (under certain conditions) and recommend using both regions simultaneously (Choi and Park 2020; Gaonkar and Campbell 2024; Piredda et al. 2016; Stoeck et al. 2010). Without argument, short amplicons provide robust and accurate output, which can be used to reveal ecological patterns and species richness in the studied areas. However, there are several challenges associated with these methods (Bradley, Pinto, and Guest 2016). One challenge is that the microbial diversity is underestimated due to poor primer coverage (Hong et al. 2009) or biased due to different amplification efficiencies (Polz and Cavanaugh 1998; Suzuki and Giovannoni 1996). Another issue is the short length of the barcode, which may not contain sufficient phylogenetic signal (Łukomska-Kowalczyk et al. 2016). These problems may lead to misinterpretation of the results. Newly developed long-read metabarcoding methods offer promising solutions to address these issues more robustly (Overgaard et al. 2023).

In this study, we describe primer pairs for the 18S-V9 region that show improved detection efficiency for Euglenozoa and Metamonada. These primers were tested on cultures derived from various freshwater samples (such as oligotrophic stream, swamp sediment, eutrophic pond, and mofettas) and recovered 263 and 124 OTUs from the targeted taxonomic groups, respectively, suggesting improved detection of these organisms. By using a combination of the universal and newly designed primer pairs, we detected 1597 OTUs, representing most of the main eukaryotic supergroups. This indicates a high protist diversity in the tested freshwater environments.

2 | Material and Methods

2.1 | Sample Collection and DNA Extraction

A total of 62 samples were collected from various locations in western Czech Republic in August 2021: Pond Amerika (50°06′40.3″ N 12°19′34.0″ E; L1), the Soos National Nature Reserve (50°08′48.0″ N 12°24′10.6″ E; L2), Kladský Pond (50°01′36.5″ N 12°40′32.9″ E; L3), Vlček Wetland (50°01′58.1″ N 12°43′59.4″ E; L4), and the Smraďoch Nature Reserve (50°00′47.1″ N 12°43′05.2″ E; L4) (Figure 1; Table S1). Selected locations varied in the type of aquatic environment. The pond Amerika (L1) is a fish breeding pond and from this location, we collected material from the pond itself and nearby puddles. The Soos National Nature Reserve (L2) encompassed a pond with a few inflows and streams but also peat bogs and many mofettas (ejections of dry carbon dioxide). Kladský Pond (L3) encompassed a pond with a few inflows, streams, and a swamp. We decided to merge samples (L4) from the Vlček wetland and Smraďoch Nature Reserve because these were small locations. Vlček wetland was composed of systems of puddles, and Smraďoch was just a big mofetta. In general, from each aquatic type, we collected water and sediment samples. Low biomass samples were collected using a plankton net (20 μm mesh size) and transferred into 50 mL tubes, while high biomass samples were collected directly into sterile 50 mL tubes. To increase the abundance of anoxic and microaerophilic organisms, the samples were subcultured and cultivated under microaerophilic conditions in 10 mL of Sonneborn's *Paramecium* medium ATCC 802 (Sonneborn 1950) for 2 months (4 passages) at room temperature with or without illumination, resulting in a total of 70 samples. For DNA extraction, four pools reflecting the sampling locations were prepared by pooling 1 mL culture from each corresponding sample (Table S1). The DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol.

2.2 | Primer Design

The primers for the V9 region of 18S rDNA were designed using the ARB software (arb-7.0) (Ludwig et al. 2004) with the SILVA SSU database version r138.1 RefNR collection

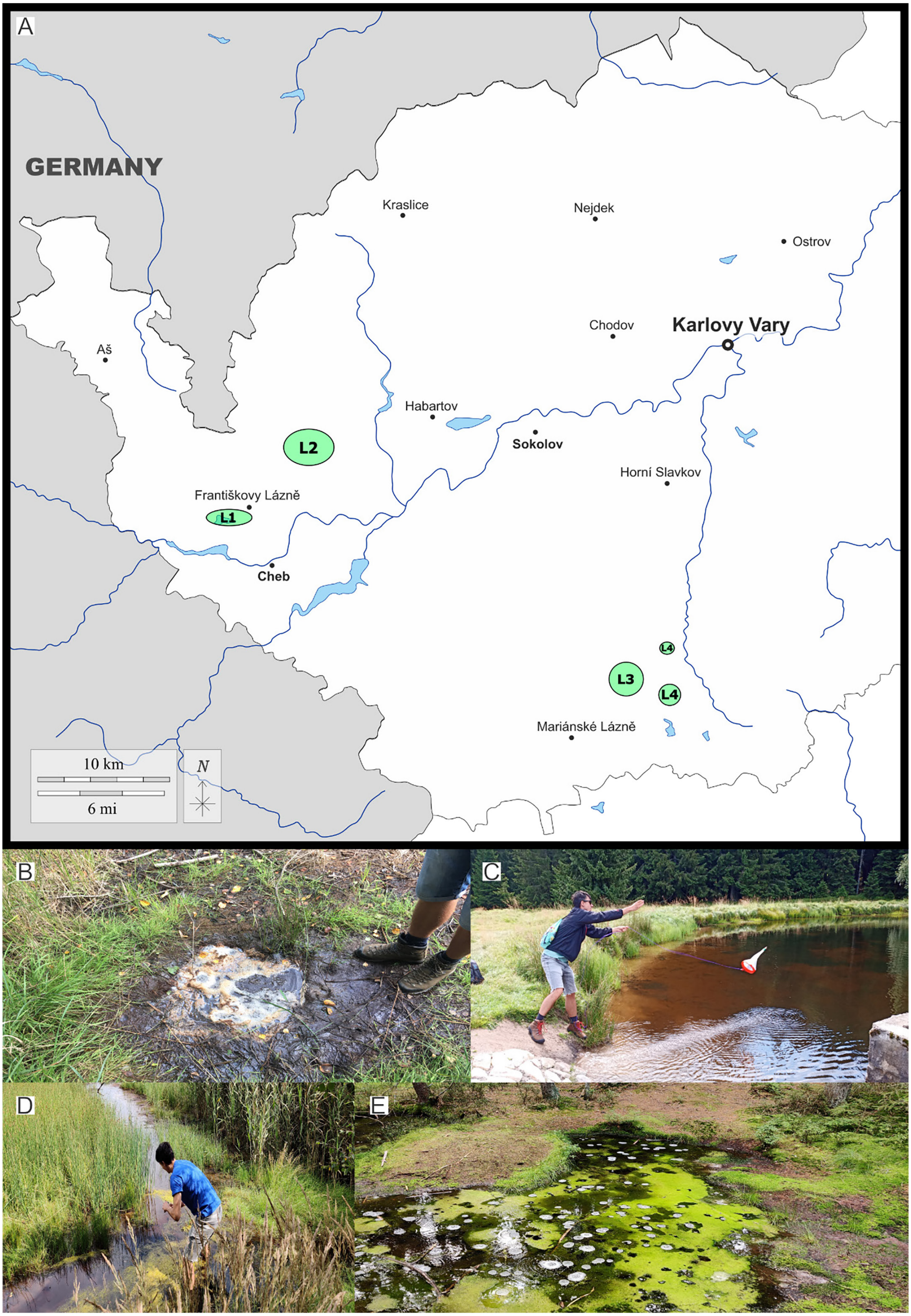

FIGURE 1 | Freshwater sampling sites. (A) Map with the sampling locations: L1—Pond Amerika, L2—Soos National Nature Reserve, L3—Kladský Pond, L4—Vlček wetland (top circle) and Nature Reserve Smraďoch (bottom circle). The photographs show examples of sampled locations, (B) mofetta in the Soos National Nature Reserve, (C) freshwater Kladský pond, (D) freshwater stream in the Soos National Nature Reserve, (E) mofetta in the Nature reserve Smraďoch.

(Quast et al. 2012). Manual adjustments were made to improve specificity for targeted groups using alignments from the Protist Ribosomal Reference (PR2) database v4.14.0 (Guillou et al. 2013). Two group-optimized reverse primers, V9Eug (5′-ACCTTGTTACGACTTTTGC-3′) and V9Meta (5′-GTTACGACTTCTCSTTCCT-3′), were designed by shifting the binding sites upstream relative to the general primer 1510R (5′-CCTTCYGCAGGTTCACCTAC-3′) (Amaral-Zettler et al. 2009), hereafter V9Gen (Figure 2). This shift was intended to increase recovery of the target groups due to increased complementarity with the respective sites in their sequences while increasing the number of mismatches with other taxa. The V9 forward primer (5′-TTTGTACACACCGCCC-3′), hereafter V9F, is a universal primer published as r7 (Milyutina et al. 2001). To verify the coverage of the newly designed primers, their performance was evaluated using SILVA-ARB TestPrime 1.0 (Klindworth et al. 2013) with a maximum of 1 allowed mismatch and a minimum of 5 base-long 0-mismatch zone at the 3′ end of either primer. The primer constructs were synthesized and HPLC purified at Biomers.net (Ulm, Germany). The expected amplicon length was approximately 100–200 bp.

2.3 | Amplicon PCR and Sequencing

Three amplicon libraries were generated from each DNA pool using the V9F primer in combination with the V9Gen, V9Meta, or V9Eug primer using the same strategy as described in (Caporaso et al. 2012), resulting in a total of 12 amplicon libraries (Table S2). Amplifications were carried out in 50 μL reactions using PrimeSTAR Max DNA Polymerase Premix (Takara Bio, Kusatsu, Shiga, Japan), with a final primer concentration 0.4 μM and ~10 ng of genomic DNA. The thermal protocol consisted of an initial step at 98°C for 3 min, followed by 10 cycles of 98°C for 10 s, 60°C for 20 s and 72°C for 30 s, then 25 cycles of 98°C for 10 s, 57°C for 20 s, and 72°C for 30 s, and a final extension at 72°C for 10 min. The correct fragment sizes were confirmed using 2.0% agarose gel electrophoresis and extracted from the gel using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany) followed by a second purification using magnetic beads based on the BOMB protocol (Oberacker et al. 2019). DNA concentration and size distribution were assessed using the Quantus fluorometer (Promega, Madison, WI, United States) and Agilent 2100 Bioanalyzer (Agilent, La Jolla, CA, United States), respectively. The amplicons were pooled and sequenced with a 2×150 paired-end configuration on an Illumina MiSeq at the Biocev Core Facility (OMICS-Genomics Biocev, Vestec, Czech Republic).

2.4 | Read Processing, OTU Clustering, and Taxonomic Annotation

The raw sequencing reads were processed using a swarm-based pipeline (https://github.com/frederic-mahe/swarm/wiki/Fred's-metabarcoding-pipeline; Mahé et al. 2014). Briefly, overlapping reads were merged using VSEARCH v.2.4.3 (Rognes et al. 2016) with the parameter --fastq_minovlen set to 50. After merging, forward and reverse primers were removed using cutadapt v.1.8.3 (Martin 2011). Reads shorter than 80 bp in length and reads containing ambiguous bases were discarded. Amplicons were further dereplicated and clustered using Swarm v.3.1.4 (Mahé et al. 2014, 2021) with the parameter $d = 1$ and fastidious option on. Chimera detection was performed by UCHIME (Edgar et al. 2011).

The taxonomic assignment was carried out using the "sequence taxonomic assignment by massive pairwise alignments"

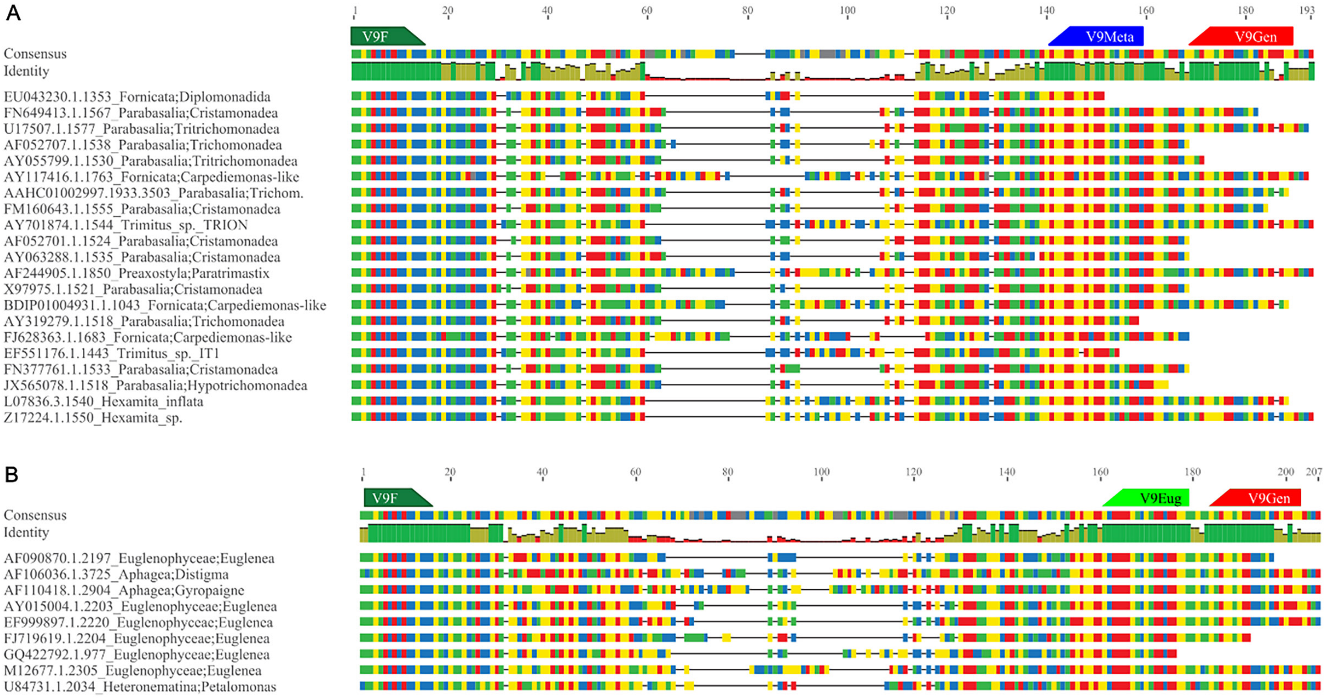

FIGURE 2 | Positions of primer pairs. Positions of primers are indicated on alignments of V9 region 18S rDNA of (A) Metamonada and (B) Euglenida species. The common forward primer is highlighted in dark green color and the universal primer (V9Gen) is highlighted in red. Newly designed specific primers are highlighted in blue (V9Meta) and light green (V9Eug).

(stampa) pipeline (Mahé 2016) based on VSEARCH. PR2 v.4.14.0 served as the reference database with the taxonomic system of eight levels, consistent with the taxonomic description from (Adl et al. 2019). For stampa, the sequences in the reference database were trimmed, and only those containing the V9 region corresponding to our primer sets were used. The V9_DeepSea database (Schoenle et al. 2021) was also included to enrich our reference dataset. The taxonomic assignment was retained if the OTUs were assigned with at least 50.0% identity to the reference sequence. For the detection of potential prokaryotic OTUs, the sequences were subjected to blastn v.2.11.0+ (Altschul et al. 1990) searches against the NCBI nt database (download date: 7. 2. 2021). Any OTU that had a prokaryotic match from the NCBI nt database at an identity level of 80.0% or higher, was classified as prokaryotic. As the last step of taxonomic assignment, the sequences were manually verified and corrected using the best hit of blastn search against the NCBI nt database.

Since each sample was sequenced three times with different reverse primers, many OTUs were found to be identical in overlapping sequence but differed in length, leading to an overestimation of OTU numbers. To address this, OTUs were clustered using CD-HIT-EST (Li and Godzik 2006) with the parameter $c = 1$. Finally, residual filtering included removing all chimeric, low-quality (minimal expected error rate/sequence length > 0.01), singleton (OTUs formed by a single sequence within a sample), bacterial, and unassigned OTUs.

To gain a better understanding of the OTUs taxonomy and to correct potentially erroneously assigned OTUs, the obtained OTUs were mapped on a comprehensive and robust reference tree. The reference tree was built using 11,685 full-length 18S rDNA sequences, representing all major protist groups, extracted from the PR2 database, aligned using MAFFT v.7.516 (--large --globalpair) (Yamada, Tomii, and Katoh 2016) and trimmed with trimAl v.1.2rev59 (--automated1) (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009). A reference phylogenetic tree was calculated using IQ-TREE v.2.2.0 (Minh et al. 2020) with the GTR+I+G model. Ultrafast bootstrap was performed with --B 1000 and --nm 5000 to ensure tree convergence. The final OTUs from the respective primer sets were added to the aligned full-length 18S rDNA dataset using MAFFT (--addfragments --multipair). The resulting alignment was trimmed with trimAl (--gt 0.01), and incorrectly aligned V9 fragments were discarded. The evolutionary placement algorithm (EPA) option of RAxML v.8.2.12 (Stamatakis 2014) was used for the phylogenetic placement estimation with the GTRCATI model. Placement signals were accumulated using gappa (--threshold 0.8) (Czech, Barbera, and Stamatakis 2020). EPA trees for OTUs for each primer set were visualized by iTOL (Letunic and Bork 2021), not considering read abundance.

Further data analysis and visualization was conducted using R v.4.2.1 (R Core Team 2023) with ggplot2 (Wickham 2016) and ggVenn (Gao et al. 2024) packages.

3 | Results

The *in silico* evaluation of the primers showed that the newly designed primer set for Metamonada (V9F + V9Meta) had a broad coverage, matching 74.6% of the species in the database. Additionally, the coverage was also broad for non-Metamonada taxa, such as Archaeplastida, Amoebozoa, Cryptista, Haptista, and SAR with a coverage > 64.5%. The primer set designed for Euglenozoa (V9F + V9Eug) had a 60.0% estimated coverage, and the probability of detecting non-Euglenozoa species was very low, at 18.0% for *Telonema*, and close to zero for other taxonomic groups. Both primer pairs had higher coverage for the targeted groups compared to the universal primer pair (V9F + V9Gen), with 31.0% for Metamonada species and 37.6% for Euglenozoa species.

The sequencing run produced 8,484,133 paired reads, which after processing yielded 5,833,209 merged reads (in average 486,100 reads per sample). These included 1,854,818 reads generated using Euglenozoa-optimized, 1,827,730 reads generated using Metamonada-optimized, and 2,150,661 reads generated using universal primer pairs. To simplify comparisons, OTU clustering was performed using the reads from all three primer combinations, resulting in 7462 OTUs. Out of these, 3062 OTUs (317,153 reads) were potentially prokaryotic (0.81%, 0.03%, and 14.02% of reads from V9Eug, V9Meta, and V9Gen primer pairs, respectively). During the OTU clustering step, the pipeline assumes that all amplicons were obtained with the same primer pair and have the same length. Reads generated by different primers, having different lengths but identical in sequence, are thus clustered to different OTUs. To remove this artificial redundancy, we clustered the resulting OTUs using CD-HIT at 100% identity. After this step, potentially chimeric OTUs, OTUs with prokaryotic hits, low-quality OTUs, and singletons were removed, leaving a set of 1794 OTUs (Table S3). The initial taxonomic assignment provided by stampa was verified by phylogenetic placement using RAxML-EPA (Figure 3A–C) and homology searches against the NCBI nt database (Table S3). After the correction and removal of the incorrectly annotated prokaryotic (11 OTUs) and unassigned (186 OTUs) taxa, the final number of supergroup assigned eukaryotic OTUs dropped to 1597.

The eukaryotic reads and OTUs were classified into supergroups and further divided based on the primer pair used to detect them. OTUs containing reads from multiple primer pairs, were assigned to more primer groups. In total, 1199 OTUs were detected using the general primer pair V9Gen (1,780,783 reads), while 402 OTUs (1,832,344 reads) and 513 OTUs (1,737,391 reads) were detected using the group-optimized primer pairs V9Eug and V9Meta, respectively (Figure 4A,B). Initially, our focus was on the performance of each primer pair in detecting the two groups of interest, Euglenozoa and Metamonada. A total of 263 OTUs belonging to Euglenozoa were detected, with 51 amplified by both V9Gen and V9Eug primers, 33 amplified exclusively by V9Gen, and 178 amplified exclusively by V9Eug primers (Figure 5A). Only one Euglenozoa OTU (OTU1623) was detected using the V9Meta primer but was not included in the comparison described above. Similarly, of the total 124 OTUs belonging to Metamonada, 34 were amplified by both V9Gen and V9Meta primers, 22 were specifically amplified by V9Gen, and 68 specifically by V9Meta (Figure 5B). When comparing the specificity of each primer set, we found that V9Eug showed high specificity, with 1,682,471 reads (91.8% of reads) and 229 OTUs (57.0% of OTUs) assigned to Euglenozoa. From

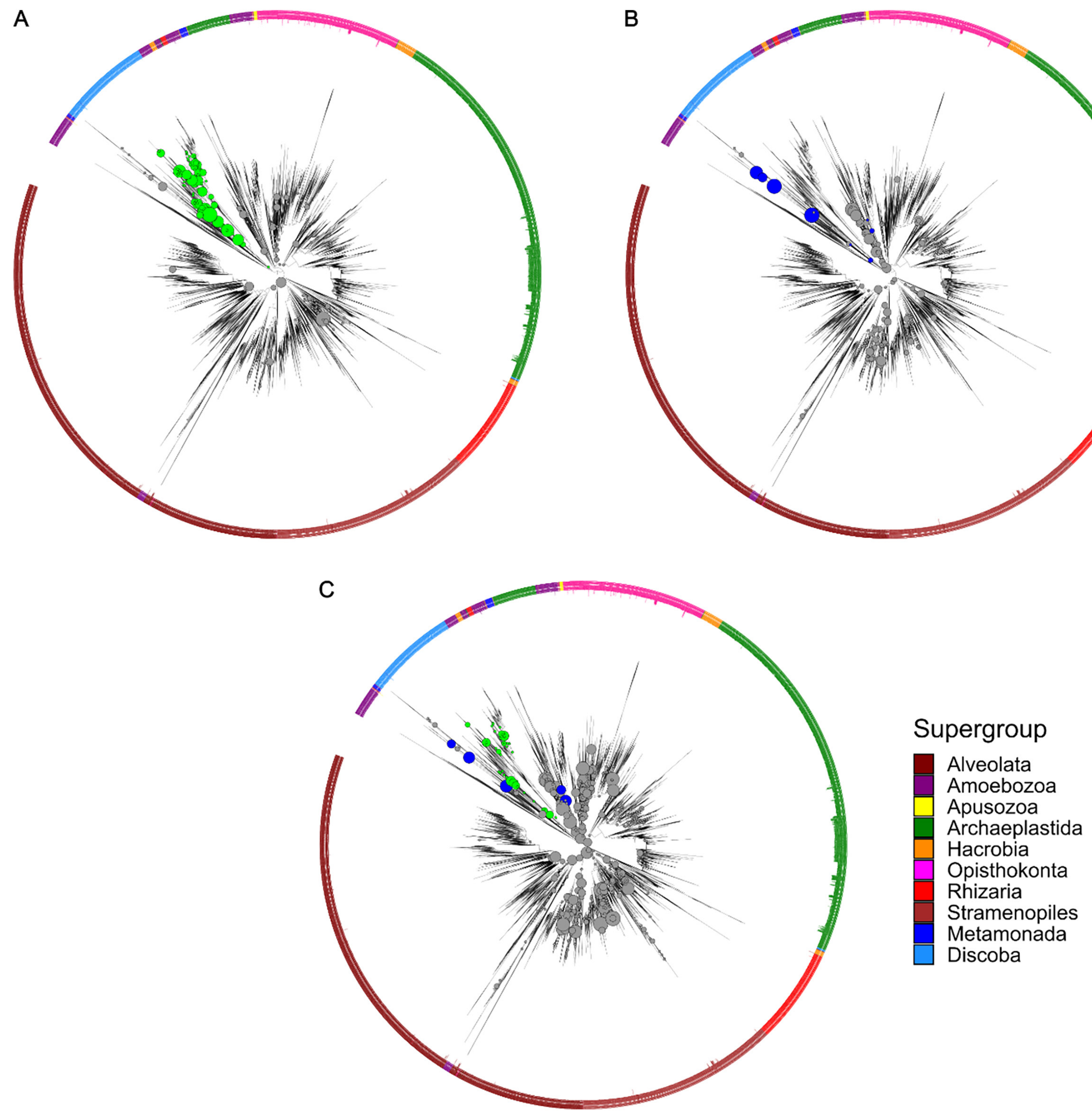

FIGURE 3 | EPA-RAxML placement of V9 barcodes. The position of barcodes obtained with (A) Euglenozoa-optimized primer pair, (B) Metamonada-optimized primer pair, and (C) general primer pair is shown on the phylogenetic tree constructed with full-length 18S rDNA sequences. Gray, green and blue circles represent the position of non-targeted OTUs, Euglenozoa OTUs and Metamonada OTUs, respectively. The color code on the edge of the trees represents respective protist supergroups.

these, 845,355 reads (46.1%) in 173 OTUs (43.0%) were assigned to Euglenida, and 837,116 reads (45.7%) in 56 OTUs (13.9%) were assigned to Kinetoplastea. The specificity of the V9Meta primer was lower, with 841,739 reads (48.4%) in 101 OTUs (19.7%) assigned to Metamonada, and the rest ascribed to other lineages.

Out of a total count of eukaryotic 1783 OTUs, 1597 OTUs were assigned to a supergroup, while 186 were only classified as "Eukaryota" without further description. The 263 OTUs of Euglenozoa were assigned to the genera *Astasia*, *Bodo*, *Cryptoglena*, *Cyclidiopsis*, *Dimastigella*, *Discoplastis*, *Distigma*, *Entosiphon*, *Euglena*, *Gyropaigne*, *Ichthyobodo*, *Menoidium*, *Monomorphina*, *Neobodo*, *Neometanema*, *Notosolenus*, *Parabodo*, *Parmidium*, *Petalomonas*, *Phacus*, *Procryptobia*, *Rhabdomonas*, *Rhynchobodo*, *Rhynchomonas*, and *Trachelomonas*; the 124 OTUs of Metamonada were assigned to genera *Chilomastix*, *Enteromonas*, *Hexamita*, *Paratrimastix*, *Trepomonas*, and *Trimitus*. However, some of the taxa assignments (e.g., parasitic genera) are likely incorrect due to the low phylogenetic resolution provided by the sequenced region (Łukomska-Kowalczyk et al. 2016). Among the nontargeted protist groups, we detected Alveolata (270 OTUs), Amoebozoa (230 OTUs), Apusozoa (1 OTU), Archaeplastida (103 OTUs), Discoba (1 unassigned and 15 Heterolobosea OTUs), Hacrobia (32 OTUs), Opisthokonta (267 OTUs), Rhizaria (67 OTUs), and Stramenopiles (224 OTUs). More details are available in Table S3.

4 | Discussion

Metamonada and Euglenozoa encompass protist species with highly divergent 18S rDNA sequences (Kolisko et al. 2020; Łukomska-Kowalczyk et al. 2016). This makes it challenging to capture them using universal metabarcoding primers targeting the hypervariable regions of 18S rDNA, leading to an underestimation of their actual diversity. In our study, we tested newly designed primer pairs optimized for amplifying the 18S-V9 rDNA region of the two protist groups. As expected, both newly designed primer pairs showed better coverage for the targeted groups than the universal primer pair. The *in silico* evaluation of the primer pair optimized for Metamonada (V9Meta) indicated reasonable coverage (74.6%) but low specificity for the group. This primer pair detected almost twice as many (102) Metamonada OTUs compared to the general primer pair (56). Both universal and Metamonada-optimized primer sets detected the same number of protist genera. For the Euglenozoa-optimized primer pair (V9Eug), the *in silico* coverage estimation was only 60.0%, but it showed high specificity for Euglenozoa. The results of the metabarcoding survey strongly supported the high specificity, as almost all obtained reads and the majority of OTUs were assigned to Euglenozoa. V9Eug recovered almost three-times more (229) Euglenozoa OTUs compared to the general primer pair (84). The universal primer pair detected only 19 out of the total 25 Euglenozoa genera detected in the samples, therefore, the specific primer brings considerable improvement in covering Euglenozoa taxa.

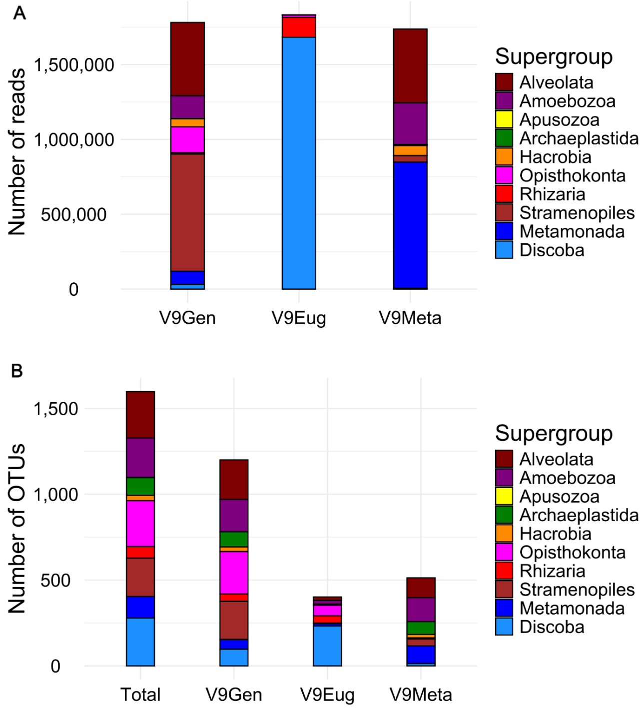

FIGURE 4 | The coverage of the eukaryotic supergroups. (A) The number of eukaryotic reads assigned to eukaryotic supergroups per tested primer pair. The graph illustrates the high specificity of Euglenozoa-optimized primer with 91.8% of reads being assigned as Euglenozoa taxa. The Metamonada-optimized primer shows moderate specificity, with 48.4% of reads being assigned as Metamonada taxa. (B) The number of annotated eukaryotic OTUs assigned to eukaryotic supergroups per tested primer pair. The majority of OTUs retrieved by the Euglenozoa-optimized primer set assigned as Euglenozoa (57.0%) indicate considerable level of specificity. On the other hand, only 19.7% of OTUs retrieved by the Metamonada-optimized primer set were assigned to Metamonada displaying a poor specificity but high coverage for Metamonada species. The "Total" column represents the number of all observed unique OTUs with all primer pairs together.

In some species, however, the optimized primers did not detect the entire variability within the sequences captured by the universal primer. For example, in the case of *Bodo saltans*, V9Eug and V9Gen detected congruently 15 OTUs (4 shared, 11 unique for both) in 815,422 and 23,986 reads, respectively. One reason for the missing OTU is the variable positions localized at 3′ end of the V9 region. As the optimized primers are shifted toward 5′ end, they do not cover them. Another reason might be errors introduced during the PCR amplification by one or the other primer pair producing artificial, primer pair-specific OTUs. However, the probability of this happening is very low because of the clustering during the OTU formation.

The Euglenozoa-optimized primer pair also seems to show low sensitivity for the heterotrophic euglenids (Lax and Simpson 2020), which are expected to be present in this type of environment but are rare among the reads. In our case, we detected a relatively low number of OTUs from the *Entosiphon*, *Petalomonas*, *Notosolenus*, and *Neometanema* genera. However, the diversity of heterotrophic euglenids in freshwater bodies is expected to be much broader (Kubín, Juráň, and Kučera 2024). This discrepancy could be due to several factors, such as their absence in the samples, their extinction during cultivation, primer bias, and bias of the reference database. Primer bias

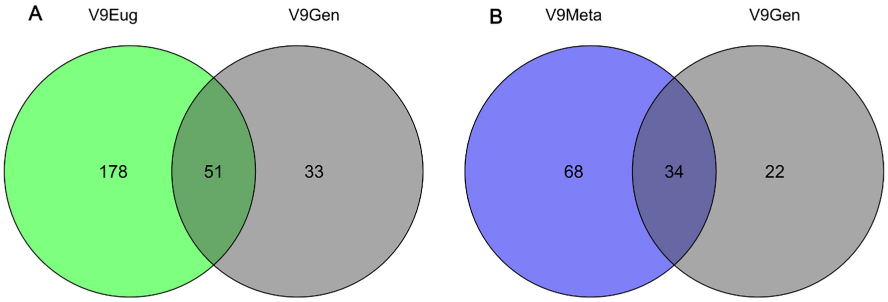

FIGURE 5 | Venn diagram comparison of commonly and exclusively detected OTUs obtained with each tested primer set pair. (A) Euglenozoa OTUs detected using Euglenozoa-optimized (V9Eug) and universal (V9Gen) primer pairs. (B) Metamonada OTUs detected using Metamonada-optimized (V9Meta) and universal (V9Gen) primer pairs.

is unlikely as the V9Eug primers show absolute or high complementarity to many sequences of heterotrophic euglenids in the PR2 database. Similarly, bias in the reference database is improbable as the PR2 contains many representatives, and the annotations were further verified by Blast against the GenBank database, which is even richer. Our light microscopy observations of fresh samples confirmed the presence of a variety of heterotrophic euglenids, while their frequency in the cultures seemed to decrease. This leads us to conclude that the absence of reads from heterotrophic euglenids is most likely caused by their extinction during the cultivation phase and not by the metabarcoding procedure. Nevertheless, the primers should be further tested on samples rich in heterotrophic euglenids to verify their sensitivity to this group.

To allow a fair comparison of the primer sets, we decided to keep singletons present in a primer set if the exact same OTU sequence is recovered with higher support using the other primer sets. Singletons play an important role as they may represent truly rare taxa or sequencing errors, and thus may affect the results (Schloss 2020). Our results support the idea that rare taxa can be missed by excluding singletons. For example, from the 51 Euglenozoa OTUs detected using both general and specific primer sets, 11 of these are represented by only one read in the general primer set and would not be detected if the environment was sampled using only this primer set, as they would be discarded as singletons. This shows that the specific primer set can help to enrich for low abundance taxa that would otherwise be missed.

The use of the general eukaryotic primer pair provided insight into the diversity of eukaryotes in the sampled biotopes and allowed comparisons with previous studies. Despite the inclusion of a cultivation step, our results mostly align with the previously described composition of microbial communities in freshwater environments (Cruaud et al. 2019; Debroas, Hugoni, and Domaizon 2015; Debroas et al. 2017; Korajkic et al. 2015; Singer et al. 2021; Slapeta, Moreira, and Lopez-Garcia 2005; Taib et al. 2013). In our cultures, the majority of protist diversity is represented by Alveolata and Stramenopiles. A high number of OTUs was assigned to Amoebozoa, which is unusual as these organisms are rare in freshwater environments (Oikonomou et al. 2015; Singer et al. 2021). We attribute this to the admixture of sediments during sampling and the inclusion of the cultivation step. Conversely, we captured a relatively low number of Hacrobia, and Archaeplastida, most likely because of the cultivation without oxygen and illumination. This type of cultivation might be especially problematic for Hacrobia as they are very abundant during the warm seasons of the year (Cruaud et al. 2019). We attributed a relatively low number of Rhizaria OTUs to their mostly soil-based natural habitat (Singer et al. 2021), which was not our primary focus, and to the possible influence of the season in which we collected samples (Karlicki et al. 2024). Additionally, we captured a high number of Fungi, but their taxonomic assignment was problematic due to the low molecular resolution provided by our marker (Tedersoo et al. 2022).

In conclusion, this study presents two new promising group-optimized 18S-V9 primer pairs for two highly morphologically and molecularly diverse protist groups, Metamonada and Euglenozoa. These primer pairs exhibit minimal recovery of prokaryotes and variable recovery of other eukaryotic groups. They are particularly effective in targeting the specified groups significantly improving their recovery in metabarcoding surveys compared to universal primers. Regardless of the improvement, primers might still be enhanced in future. In the case of V9Meta primer, an improvement might be necessary to achieve higher sensitivity and specificity. For the V9Eug primer, testing its sensitivity for heterotrophic euglenids is desirable.

Author Contributions

S.C.T., Z.F., and V.H. designed the study, Z.F. designed new primers, J.N. conducted the wet lab experiments and cultivations, S.C.T., B.H., and Š.H. designed the sequencing strategy, J.N., S.C.T., and K.Z. conducted the bioinformatical analyses, and J.N. wrote the manuscript with input from all authors. All authors approved the final version of the manuscript.

Acknowledgments

We would like to express our gratitude to Zdeněk Soukup for leading us through Soos National Park and to Justýna Zítek, Priscilla Peña-Diaz, and Yi-Kai Fang for assisting us in collecting samples.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data analyzed in this study are deposited under the BioProject ID PRJNA1051831. Alignment for the reference phylogenetic tree, computed tree and OTU placements are available on Zenodo: 10.5281/zenodo.13771462.

References

Adl, S. M., D. Bass, C. E. Lane, et al. 2019. “Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes.” *Journal of Eukaryotic Microbiology* 66: 4–119.

Aguilar, M., E. Richardson, B. Tan, et al. 2016. “Next-Generation Sequencing Assessment of Eukaryotic Diversity in Oil Sands Tailings Ponds Sediments and Surface Water.” *Journal of Eukaryotic Microbiology* 63: 732–743.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215: 403–410.

Amaral-Zettler, L. A., E. A. McCliment, H. W. Ducklow, and S. M. Huse. 2009. “A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes.” *PLoS One* 4: e6372.

Anderson, S. R., and E. L. Harvey. 2020. “Temporal Variability and Ecological Interactions of Parasitic Marine Syndiniales in Coastal Protist Communities.” *mSphere* 5: e00209.

Bates, S. T., J. C. Clemente, G. E. Flores, et al. 2012. “Global Biogeography of Highly Diverse Protistan Communities in Soil.” *ISME Journal* 7: 652–659.

Berdjeb, L., A. Parada, D. M. Needham, and J. A. Fuhrman. 2018. “Short-Term Dynamics and Interactions of Marine Protist Communities During the Spring–Summer Transition.” *ISME Journal* 12: 1907–1917.

Bicudo, C. E. M., and M. Menezes. 2016. “Phylogeny and Classification of Euglenophyceae: A Brief Review.” *Frontiers in Ecology and Evolution* 4: 17.

Bradley, I. M., A. J. Pinto, and J. S. Guest. 2016. "Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities." *Applied and Environmental Microbiology* 82: 5878–5891.

Burki, F., A. J. Roger, M. W. Brown, and A. G. B. Simpson. 2020. "The New Tree of Eukaryotes." *Trends in Ecology & Evolution* 35: 43–55.

Butenko, A., M. Hammond, M. C. Field, M. L. Ginger, V. Yurchenko, and J. Lukeš. 2021. "Reductionist Pathways for Parasitism in Euglenozoans? Expanded Datasets Provide New Insights." *Trends in Parasitology* 37: 100–116.

Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25: 1972–1973.

Caporaso, J. G., C. L. Lauber, W. A. Walters, et al. 2012. "Ultra-High-Throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms." *ISME Journal* 6: 1621–1624.

Choi, J., and J. S. Park. 2020. "Comparative Analyses of the V4 and V9 Regions of 18S rDNA for the Extant Eukaryotic Community Using the Illumina Platform." *Scientific Reports* 10: 6519.

Cruaud, P., A. Vigneron, M.-S. Fradette, et al. 2019. "Annual Protist Community Dynamics in a Freshwater Ecosystem Undergoing Contrasted Climatic Conditions: The Saint-Charles River (Canada)." *Frontiers in Microbiology* 10: 2359.

Czech, L., P. Barbera, and A. Stamatakis. 2020. "Genesis and Gappa: Processing, Analyzing and Visualizing Phylogenetic (Placement) Data." *Bioinformatics* 36: 3263–3265.

de Vargas, C., S. Audic, N. Henry, et al. 2015. "Eukaryotic Plankton Diversity in the Sunlit Ocean." *Science* 348: 1261605.

Debroas, D., I. Domaizon, J.-F. Humbert, et al. 2017. "Overview of Freshwater Microbial Eukaryotes Diversity: A First Analysis of Publicly Available Metabarcoding Data." *FEMS Microbiology Ecology* 93: fix023.

Debroas, D., M. Hugoni, and I. Domaizon. 2015. "Evidence for an Active Rare Biosphere Within Freshwater Protists Community." *Molecular Ecology* 24: 1236–1247.

Dunthorn, M., J. Klier, J. Bunge, and T. Stoeck. 2012. "Comparing the Hyper-Variable V4 and V9 Regions of the Small Subunit rDNA for Assessment of Ciliate Environmental Diversity." *Journal of Eukaryotic Microbiology* 59: 185–187.

Dunthorn, M., J. Otto, S. A. Berger, et al. 2014. "Placing Environmental Next-Generation Sequencing Amplicons From Microbial Eukaryotes Into a Phylogenetic Context." *Molecular Biology and Evolution* 31: 993–1009.

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. "UCHIME Improves Sensitivity and Speed of Chimera Detection." *Bioinformatics* 27: 2194–2200.

Edgcomb, V. 2016. "Marine Protist Associations and Environmental Impacts Across Trophic Levels in the Twilight Zone and Below." *Current Opinion in Microbiology* 31: 169–175.

Edgcomb, V., W. Orsi, J. Bunge, et al. 2011. "Protistan Microbial Observatory in the Cariaco Basin, *Caribbean* I. Pyrosequencing vs Sanger Insights Into Species Richness." *ISME Journal* 5: 1344–1356.

Ficetola, G. F., and P. Taberlet. 2023. "Towards Exhaustive Community Ecology via DNA Metabarcoding." *Molecular Ecology* 32: 6320–6329.

Flegontova, O., J. Lukeš, and A. Horák. 2023. "Intragenomic Diversity of the V9 Hypervariable Domain in Eukaryotes Has Little Effect on Metabarcoding." *iScience* 26: 107291.

Gao, C., C. Chen, T. Akyol, et al. 2024. "ggVennDiagram: Intuitive Venn Diagram Software Extended." *iMeta* 3: e177.

Gaonkar, C. C., and L. Campbell. 2024. "A Full-Length 18S Ribosomal DNA Metabarcoding Approach for Determining Protist Community Diversity Using Nanopore Sequencing." *Ecology and Evolution* 14: e11232.

Geisen, S., C. Bandow, J. Römbke, and M. Bonkowski. 2014. "Soil Water Availability Strongly Alters the Community Composition of Soil Protists." *Pedobiologia* 57: 205–213.

Geisen, S., E. A. D. Mitchell, S. Adl, et al. 2018. "Soil Protists: A Fertile Frontier in Soil Biology Research." *FEMS Microbiology Reviews* 42: 293–323.

Guillou, L., D. Bachar, S. Audic, et al. 2013. "The Protist Ribosomal Reference Database (PR2): A Catalog of Unicellular Eukaryote Small Sub-Unit rRNA Sequences With Curated Taxonomy." *Nucleic Acids Research* 41: D597–D604.

Gutiérrez-Rodríguez, A., A. L. dos Santos, K. Safi, et al. 2022. "Planktonic Protist Diversity Across Contrasting Subtropical and Subantarctic Waters of the Southwest Pacific." *Progress in Oceanography* 206: 102809.

Hadziavdic, K., K. Lekang, A. Lanzen, I. Jonassen, E. M. Thompson, and C. Troedsson. 2014. "Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers." *PLoS One* 9: e87624.

Hong, S., J. Bunge, C. Leslin, S. Jeon, and S. S. Epstein. 2009. "Polymerase Chain Reaction Primers Miss Half of rRNA Microbial Diversity." *ISME Journal* 3: 1365–1373.

Hugerth, L. W., E. E. L. Muller, Y. O. O. Hu, et al. 2014. "Systematic Design of 18S rRNA Gene Primers for Determining Eukaryotic Diversity in Microbial Consortia." *PLoS One* 9: e95567.

Jankowska, K., M. Łukomska-Kowalczyk, R. Milanowski, A. Fells, and B. Zakryś. 2024. "Biodiversity of Autotrophic Euglenids Based on the Group Specific DNA Metabarcoding Approach." *Protist* 175: 126024.

Karlicki, M., A. Bednarska, P. Hałakuc, K. Maciszewski, and A. Karnkowska. 2024. "Spatio-Temporal Changes of Small Protist and Free-Living Bacterial Communities in a Temperate Dimictic Lake: Insights From Metabarcoding and Machine Learning." *bioRxiv* 100: 5328.

Klindworth, A., E. Pruesse, T. Schweer, et al. 2013. "Evaluation of General 16S Ribosomal RNA Gene PCR Primers for Classical and Next-Generation Sequencing-Based Diversity Studies." *Nucleic Acids Research* 41: e1.

Kolisko, M., O. Flegontova, A. Karnkowska, et al. 2020. "EukRef-excavates: seven curated SSU ribosomal RNA gene databases." *Database* 2020: baaa080.

Korajkic, A., L. W. Parfrey, B. R. McMinn, et al. 2015. "Changes in Bacterial and Eukaryotic Communities During Sewage Decomposition in Mississippi River Water." *Water Research* 69: 30–39.

Kostygov, A. Y., A. Karnkowska, J. Votýpka, et al. 2021. "Euglenozoa: Taxonomy, Diversity and Ecology, Symbioses and Viruses." *Open Biology* 11: 200407.

Kubín, J., J. Juráň, and J. Kučera. 2024. "Checklist of Colourless Euglenoids of The Czech Republic, With Several Taxonomic Additions." *Protist* 175: 126045.

Lax, G., M. Kolisko, Y. Eglit, et al. 2021. "Multigene Phylogenetics of Euglenids Based on Single-Cell Transcriptomics of Diverse Phagotrophs." *Molecular Phylogenetics and Evolution* 159: 107088.

Lax, G., and A. G. B. Simpson. 2020. "The Molecular Diversity of Phagotrophic Euglenids Examined Using Single-Cell Methods." *Protist* 171: 125757.

Letunic, I., and P. Bork. 2021. "Interactive Tree of Life (iTOL) V5: An Online Tool for Phylogenetic Tree Display and Annotation." *Nucleic Acids Research* 49: W293–W296.

Li, W., and A. Godzik. 2006. "Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences." *Bioinformatics* 22: 1658–1659.

Liao, H., X. Hao, Y. Li, et al. 2024. "Protists Regulate Microbially Mediated Organic Carbon Turnover in Soil Aggregates." *Global Change Biology* 30: e17102.

López-García, P., F. Rodríguez-Valera, C. Pedrós-Alió, and D. Moreira. 2001. "Unexpected Diversity of Small Eukaryotes in Deep-Sea Antarctic Plankton." *Nature* 409: 603–607.

Ludwig, W., O. Strunk, R. Westram, et al. 2004. "ARB: A Software Environment for Sequence Data." *Nucleic Acids Research* 32: 1363–1371.

Łukomska-Kowalczyk, M., A. Karnkowska, M. Krupska, R. Milanowski, and B. Zakryś. 2016. "DNA Barcoding in Autotrophic Euglenids: Evaluation of COI and 18s rDNA." *Journal of Phycology* 52: 951–960.

Mahé, F. 2016. "Stampa: Sequence Assignment by Massive Pairwise Alignments." (Version 2.0.0) [Computer software]. https://github.com/frederic-mahe/stampa.

Mahé, F., L. Czech, A. Stamatakis, et al. 2021. "Swarm V3: Towards Tera-Scale Amplicon Clustering." *Bioinformatics* 38: 267–269.

Mahé, F., T. Rognes, C. Quince, C. de Vargas, and M. Dunthorn. 2014. "Swarm: Robust and Fast Clustering Method for Amplicon-Based Studies." *PeerJ* 2: e593.

Martin, M. 2011. "Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads." *EMBnet Journal* 17: 10–12.

Mazancová, E., E. Zadrobílková, N. Yubuki, and I. Čepička. 2023. "Phylogenetic and Morphological Diversity of Free-Living Diplomonads." *European Journal of Protistology* 91: 126024.

Metz, S., P. Huber, V. Accattatis, et al. 2022. "Freshwater Protists: Unveiling the Unexplored in a Large Floodplain System." *Environmental Microbiology* 24: 1731–1745.

Milyutina, I. A., V. V. Aleshin, K. A. Mikrjukov, O. S. Kedrova, and N. B. Petrov. 2001. "The Unusually Long Small Subunit Ribosomal RNA Gene Found in Amitochondriate Amoeboflagellate Pelomyxa Palustris: Its rRNA Predicted Secondary Structure and Phylogenetic Implication." *Gene* 272: 131–139.

Minh, B. Q., H. A. Schmidt, O. Chernomor, et al. 2020. "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era." *Molecular Biology and Evolution* 37: 1530–1534.

Oberacker, P., P. Stepper, D. M. Bond, et al. 2019. "Bio-On-Magnetic-Beads (BOMB): Open Platform for High-Throughput Nucleic Acid Extraction and Manipulation." *PLoS Biology* 17: e3000107.

Oikonomou, A., S. Filker, H. Breiner, and T. Stoeck. 2015. "Protistan Diversity in a Permanently Stratified Meromictic Lake (Lake Alatsee, SW Germany)." *Environmental Microbiology* 17: 2144–2157.

Overgaard, C. K., M. Jamy, S. Radutoiu, F. Burki, and M. K. D. Dueholm. 2023. "Benchmarking Long-Read Sequencing Strategies for Obtaining ASV-Resolved rRNA Operons From Environmental Microeukaryotes." *bioRxiv* 2023: 558928.

Pawlowski, J., S. Audic, S. Adl, et al. 2012. "CBOL Protist Working Group: Barcoding Eukaryotic Richness Beyond the Animal, Plant, and Fungal Kingdoms." *PLoS Biology* 10: e1001419.

Piredda, R., M. P. Tomasino, A. M. D'Erchia, et al. 2016. "Diversity and Temporal Patterns of Planktonic Protist Assemblages at a Mediterranean Long Term Ecological Research Site." *FEMS Microbiology Ecology* 93: fiw200.

Polz, M. F., and C. M. Cavanaugh. 1998. "Bias in Template-To-Product Ratios in Multitemplate PCR." *Applied and Environmental Microbiology* 64: 3724–3730.

Quast, C., E. Pruesse, P. Yilmaz, et al. 2012. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41: D590–D596.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. "VSEARCH: A Versatile Open Source Tool for Metagenomics." *PeerJ* 4: e2584.

Schloss, P. D. 2020. "Removal of Rare Amplicon Sequence Variants From 16S rRNA Gene Sequence Surveys Biases the Interpretation of Community Structure Data." *bioRxiv* 2020: 422279.

Schoenle, A., M. Hohlfeld, K. Hermanns, et al. 2021. "High and Specific Diversity of Protists in the Deep-Sea Basins Dominated by Diplonemids, Kinetoplastids, Ciliates and Foraminiferans." *Communications Biology* 4: 501.

Seppey, C. V. W., O. Broennimann, A. Buri, et al. 2020. "Soil Protist Diversity in the Swiss Western Alps Is Better Predicted by Topo-Climatic Than by Edaphic Variables." *Journal of Biogeography* 47: 866–878.

Singer, D., C. V. W. Seppey, G. Lentendu, et al. 2021. "Protist Taxonomic and Functional Diversity in Soil, Freshwater and Marine Ecosystems." *Environment International* 146: 106262.

Slapeta, J., D. Moreira, and P. Lopez-Garcia. 2005. "The Extent of Protist Diversity: Insights From Molecular Ecology of Freshwater Eukaryotes." *Proceedings of the Royal Society B: Biological Sciences* 272: 2073–2081.

Sonneborn, T. M. 1950. "Methods in the General Biology and Genetics of Paramecium Aurelia." *Journal of Experimental Zoology* 113: 87–147.

Stairs, C. W., P. Táborský, E. D. Salomaki, et al. 2021. "Anaeramoebae Are a Divergent Lineage of Eukaryotes That Shed Light on the Transition From Anaerobic Mitochondria to Hydrogenosomes." *Current Biology* 31: 5605–5612.

Stamatakis, A. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30: 1312–1313.

Stoeck, T., D. Bass, M. Nebel, et al. 2010. "Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water." *Molecular Ecology* 19: 21–31.

Suzuki, M. T., and S. J. Giovannoni. 1996. "Bias Caused by Template Annealing in the Amplification of Mixtures of 16S rRNA Genes by PCR." *Applied and Environmental Microbiology* 62: 625–630.

Taib, N., J.-F. Mangot, I. Domaizon, G. Bronner, and D. Debroas. 2013. "Phylogenetic Affiliation of SSU rRNA Genes Generated by Massively Parallel Sequencing: New Insights Into the Freshwater Protist Diversity." *PLoS One* 8: e58950.

Tedersoo, L., M. Bahram, L. Zinger, et al. 2022. "Best Practices in Metabarcoding of Fungi: From Experimental Design to Results." *Molecular Ecology* 31: 2769–2795.

Tong, Y., L. Wang, C. Wu, et al. 2024. "Promotion of the Symbiotic Interaction Between Phagotrophic Protists and Beneficial Bacteria, Mediated via a Balancing of Soil Nutrients, Reduces the Incidence of Watermelon Fusarium Wilt." *Pedosphere*.

Tragin, M., A. Zingone, and D. Vaulot. 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." *Environmental Microbiology* 20: 506–520.

Trefault, N., R. D. Iglesia, M. Moreno-Pino, et al. 2021. "Annual Phytoplankton Dynamics in Coastal Waters From Fildes Bay, Western Antarctic Peninsula." *Scientific Reports* 11: 1368.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Yamada, K. D., K. Tomii, and K. Katoh. 2016. "Application of the MAFFT Sequence Alignment Program to Large Data—Reexamination of the Usefulness of Chained Guide Trees." *Bioinformatics* 32: 3246–3251.

Yazaki, E., K. Kume, T. Shiratori, et al. 2020. “Barthelonids Represent a Deep-Branching Metamonad Clade With Mitochondrion-Related Organelles Predicted to Generate No ATP.” *Proceedings of the Royal Society B: Biological Sciences* 287: 20201538.

Záhonová, K., G. Lax, S. D. Sinha, et al. 2021. “Single-Cell Genomics Unveils a Canonical Origin of the Diverse Mitochondrial Genomes of Euglenozoans.” *BMC Biology* 19: 103.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.