Ecology and Evolution

WILEY

RESEARCH ARTICLE OPEN ACCESS

Unveiling Protist Composition and Diversity Patterns With eDNA Metabarcoding: Comparing Short- and Long-Read Approaches

Dimitra-Ioli Skouroliakou[1,2] | Deborah W. E. Dupont[3,4] | Yelle Vandenboer[3] | Sofie D'Hont[2] | Koen Sabbe[2] | Isa Schön[3,5]

[1]IFREMER, DYNECO, Plouzané, France | [2]Laboratory of Protistology and Aquatic Ecology, Department of Biology, Ghent University, Ghent, Belgium | [3]Operational Directorate Natural Environment, Aquatic and Terrestrial Ecology, Freshwater Biology, Royal Belgian Institute of Natural Sciences, Brussels, Belgium | [4]Marine Biology Lab, Université Libre de Bruxelles (ULB), Brussels, Belgium | [5]Research Group Zoology, University of Hasselt, Diepenbeek, Belgium

Correspondence: Dimitra-Ioli Skouroliakou (dimitra.ioli.skouroliakou@ifremer.fr; dimitraioli.skouroliakou@ugent.be)

Received: 13 February 2026 | **Accepted:** 20 February 2026

ABSTRACT

Environmental DNA (eDNA) metabarcoding is a key tool in biodiversity monitoring due to its high-throughput, non-destructive nature. While short-read (SR) sequencing platforms such as Illumina Miseq have been routinely used in environmental monitoring, their limited read lengths (less than 600 bp) constrain the depth of taxonomic assignment, particularly for complex microbial eukaryotes like protists. Conversely, long-read (LR) sequencing technologies like Oxford Nanopore Technologies (ONT) offer promising alternatives but remain underutilized for studying protist communities. We conducted a comparative study of SR versus LR metabarcoding of protist communities along a coastal-offshore gradient in the Belgian part of the North Sea. Using amplicons targeting the V4 region (SR; 577 bp) and the V4–V5 region (LR; 745 bp) of the 18S rRNA gene, we compared diversity patterns, taxonomic assignment, and community composition between approaches. We observed general congruence in community composition at higher taxonomic levels, but under the applied workflows, LR metabarcoding yielded a greater depth of taxonomic annotation at lower taxonomic ranks. Notably, dinoflagellates were less overrepresented in LR data, and a unique detection of potential nuisance taxa (e.g., *Bellerochea*), and ecologically important genera such as haptophytes (e.g., *Gephyrocapsa*) was achieved. These results highlight the potential of LR metabarcoding to complement SR approaches by providing increased taxonomic annotation depth and ecological insights. Although both methods targeted only partial regions of the 18S rRNA gene, LR metabarcoding yielded a greater depth of taxonomic assignment under the applied workflows. As next-generation sequencing technologies continue to evolve, our research provides valuable insights for selecting optimal strategies in routine plankton monitoring and biodiversity assessment programs.

1 | Introduction

Over the past decade, eDNA metabarcoding has started to transform monitoring as a relatively fast and non-destructive method to characterize biodiversity and its changes worldwide (Ruppert et al. 2019). Environmental DNA (eDNA) refers to the genetic material that can be extracted from environmental samples of different matrices. It is acknowledged that eDNA contains a complex mixture of intracellular DNA, originating from living cells or potentially whole organisms, and extracellular DNA resulting from natural cell death and/or destruction of cell structure (Taberlet et al. 2012; Pawlowski et al. 2020;

Correction added on 19 April 2026, after first online publication: the affiliation of the authors, Sofie D'Hont and Koen Sabbe, has been updated in this version. Skouroliakou Dimitra-Ioli and Dupont Deborah W.E. contributed equally to this work.

Rodriguez-Ezpeleta et al. 2021). Amplicon sequencing of eDNA facilitates identification of the composition of natural assemblages without prior knowledge and also detects small-size organisms like pico- and nanoplankton, cryptic species with undistinguishable morphology, and rare species (López-García et al. 2001; Sunagawa et al. 2015; Burki et al. 2021; Duarte et al. 2023). It has been routinely used to study protist communities in various coastal systems (Lambert et al. 2019; Caracciolo et al. 2022; Longobardi et al. 2022; Skouroliakou et al. 2022). As its use becomes more widespread, eDNA is increasingly integrated into international monitoring initiatives aimed at evaluating the environmental status of marine ecosystems and predicting future changes, such as the European Marine Omics Biodiversity Observation Network (EMO BON) (Cardinale et al. 2012; Santi et al. 2023).

Protists form a significant part (over 60%) of marine biomass (Bar-On and Milo 2019) playing a key role in biogeochemical cycles and carbon sequestration and as primary producers in marine food webs (Falkowski et al. 1998; Field et al. 1998). Investigations of protists with eDNA have so far been based on short-read sequencing methods provided by Illumina technology (named short-read (SR) metabarcoding hereafter), targeting hypervariable parts of the 18S ribosomal region such as V4 (Pernice et al. 2013), often in conjunction with V5 or V9 (Amaral-Zettler et al. 2009). This approach has been able to identify diversity patterns of protist communities and their major drivers. For example, SR metabarcoding data revealed correlations of seasonal and/or spatial variation in protist diversity with environmental fluctuations (Caracciolo et al. 2022), interspecific interactions (Genitsaris et al. 2015), and water depth (Meziti et al. 2023). Protist species causing harmful algal blooms (HABs) have also been successfully detected with SR metabarcoding at low cell concentrations; an example is the HAB of *Lepidodinium chloroforum* that was identified outside of algae blooming periods at the coast of south Brittany (Roux et al. 2023).

Despite its widespread application, SR metabarcoding provides limited molecular information because commonly sequenced amplicons are typically short (approximately 300 to <600 base pairs). This restricted sequence length can lead to biased estimates of taxonomic composition, resulting in the over- or underrepresentation of certain taxa. For example, dinoflagellates, which possess numerous ribosomal gene copies per cell, are frequently overrepresented due to PCR and sequencing biases (Prokopowich et al. 2003; Wisecaver and Hackett 2011). In addition, SR metabarcoding often limits taxonomic identification to higher taxonomic ranks (Szoboszlay et al. 2023).

These limitations are particularly pronounced for protists, which comprise deeply divergent evolutionary lineages and exhibit extensive cryptic diversity as well as highly variable rates of rRNA gene evolution across taxa (Burki et al. 2021; Jamy et al. 2022). Accurate characterization of protist communities therefore requires sufficient phylogenetic signal to discriminate among closely related taxa and evolutionary lineages. Although SR metabarcoding has proven effective for detecting broad diversity patterns and ecological gradients in protist communities (Genitsaris et al. 2015; Caracciolo et al. 2022; Meziti et al. 2023), the limited length of commonly used markers constrains the depth of taxonomic assignment achievable for many groups—particularly those that are underrepresented or unevenly annotated in reference databases (Szoboszlay et al. 2023). In contrast, longer amplicons provide increased sequence context and, in principle, greater potential for improved taxonomic resolution and discrimination among closely related protist taxa (Jamy et al. 2022; Gaonkar and Campbell 2024; Chwalińska et al. 2025).

Long-read (LR) sequencing technologies such as Oxford Nanopore Technologies (ONT) have been developed in recent years. ONT LR metabarcoding has successfully characterized bacterial (Stoeck et al. 2024; van der Loos et al. 2021) and zooplankton communities (Semmouri et al. 2021) with eDNA. However, to our knowledge, only a few studies have compared protist diversity and composition between SR and LR metabarcoding approaches based on 18S rDNA (Jamy et al. 2022; Gaonkar and Campbell 2024; Chwalińska et al. 2025). Few other studies focused on specific protist groups such as Radiolaria (Sandin et al. 2022).

Long-read metabarcoding is advantageous in providing more DNA sequencing data from longer amplicons of the target DNA regions (more than 600 bp) and thus could potentially enable deeper taxonomic annotation and discrimination among closely related taxa. This could be particularly beneficial for taxonomically diverse groups like protists, where fine distinctions between closely related taxa are often required for accurate ecological assessments (Gaonkar and Campbell 2024). The most significant limitation of ONT, however, is its greater error rate as compared to Illumina sequencing based on base calling biases (Wang et al. 2021). Given the recent improvements in the ONT library kits and flow cell chemistry both reducing sequencing errors (Wang et al. 2021), it has now become possible to test whether LR metabarcoding using ONT can overcome the limitations of SR metabarcoding based on Illumina sequencing. As a result, differences between SR and LR outputs may reflect both biological signal and methodological artifacts. Evaluating how these approaches complement each other, and to what extent they can be integrated in comparative ecological analyses, remains an open question for protist metabarcoding.

The present study focuses on protist communities, comparing their taxonomic assignment, composition, and diversity across a coastal-offshore gradient between metabarcoding data generated with (SR) and (LR) sequencing. We expect congruence in diversity patterns based on higher taxonomic levels between both methods similarly to Gaonkar and Campbell (2024). Because LR metabarcoding generates longer sequence reads with increased phylogenetic information content, we hypothesize that LR metabarcoding will provide greater depth of taxonomic assignment for protist communities under the applied workflows, potentially revealing a higher number of taxa at lower taxonomic ranks compared to SR approaches (Jamy et al. 2022; Gaonkar and Campbell 2024; Chwalińska et al. 2025). In the context of the increasing integration of eDNA metabarcoding into studies of biodiversity, ecology and also monitoring, such comparative studies are urgently required to inform about methodological choices. Our work contributes practical insights into the strengths and limitations of commonly used sequencing strategies for the characterization of marine protist communities.

2 | Material and Methods

2.1 | Study Area

The Belgian Part of the North Sea (BPNS; 3454 km^2) is an epicontinental shallow area (maximum 35 m deep) with a coastal length at ca. 66 km (Mortelmans et al. 2019). The BPNS is affected by Atlantic waters through the English Channel and by freshwater inputs from the Scheldt and Meuse rivers (Aubert et al. 2022 and references therein). It is macrotidal and well-mixed without seasonal stratification (Blauw et al. 2012). The BPNS experienced eutrophication (1950s–1980s), de-eutrophication (1980s–2000s), and an increase in water temperature since the 1970s (Beaugrand 2004). This area is furthermore heavily impacted by the introduction of non-indigenous species, industrial and agricultural pollution, overfishing and trawling, offshore wind farming and heavy shipping traffic (Emeis et al. 2015). The protist communities of the BPNS have been extensively studied in the past with a wide range of molecular and classic microscopic and imaging techniques (Breton et al. 2006; Nohe et al. 2020; Aubert et al. 2022; Perneel et al. 2024), making it an ideal area for the current study.

2.2 | Sample Collection

A total of 79 samples were collected aboard the *R.V. Belgica* during 15 campaigns at a monthly basis during the regular monitoring program organized by RBINS (Table S1). Three stations were sampled from February 2022 to May 2023; the coastal station MOW1 (51°21.50′ N, 3°07.50′ E), the transitional station WO5 (51°25.00′ N, 2°48.50′ E), and the offshore station WO8 (51°27.61′ N, 2°20.91′ E) situated 87 km west of the coastal station (Figure S1). No samples were collected in May, November 2022, and January 2023 due to rough weather conditions and/or the boat's unavailability. At each station, sea water was collected with Niskin bottles at 1 m depth (subsurface) and 1 m above the seabed (epibenthic). Seawater ranging in volume from 0.1 to 1 L was filtered on board with a low-pressure vacuum pump through 0.45 μm membrane filters (47 mm, Merck-Millipore), depending on the quantity of suspended matter in the water (i.e., until clogging of the filter occurred). All filters were stored at −80°C onboard the vessel, after snap-freezing in liquid nitrogen, until DNA extractions took place in the dedicated eDNA laboratory at the RBINS. Amplification and sequencing procedures used SR (MiSeq, Illumina) and LR (GridIon, Oxford Nanopore Technologies) sequencing approaches (Figure S2). Here, 66 subsurface and 13 epibenthic water samples were analyzed together with five field and three laboratory negative controls to assess potential contamination.

2.3 | DNA Extraction

DNA was extracted from half of the filters following an adapted protocol of the DNeasy PowerLyzer Microbial kit (Qiagen, Germany). Modifications were made to maximize the yield, such as increasing the rotation speed of the PowerLyzer (30 Hz for 3 min), reducing the volume of eluents to 35 μL, and repeating elutions three times.

2.4 | Primer Selection and In Silico Tests

One of the most widely used eukaryote-specific primer pairs in SR metabarcoding studies are the forward primer TAReuk454FWD1 (CCAGCASCYGCGGTAATTCC) and the reverse primer TAReukREV3 (ACTTTCGTTCTTGATYRA) (Stoeck et al. 2010), which together amplify 577 bp of the V4 region of the 18S rRNA gene. We used the same primers for LR metabarcoding as Semmouri et al. (2021), who successfully characterized marine zooplankton communities of the North Sea in high resolution. The 18S primers F-566 (CAGCAGCCGCGGTAATTCC) and R-1200 (CCCGTGTTGAGTCAAATTAAGC) (Hadziavdic et al. 2014) amplify the V4 and V5 regions of the 18S rRNA gene with a total length of 745 bp (Figure S3). To assess how well these primers would amplify protists from a reference database, *in silico* tests were performed first using the PR2 Primer Database v.2 (Vaulot et al. 2022), which is curated to the species level for protists. Primers were matched with all protist sequences available in the database, allowing to identify mismatches and evaluate amplification efficiency. The SR primer pair TAReuk454FWD1—TAReukREV3 (Stoeck et al. 2010) detected in silico 83.2% of all eukaryotic barcodes in the PR2 database; and the LR primer pair F-566—R-1200 (Hadziavdic et al. 2014) 90.45% (Figures S4 and S5).

2.5 | Short-Read Amplification and Library Preparation

Forward and reverse primers with overhang and adapter sequences of 38 bp were used, which bind to Illumina indexes and sequencing adapters according to the standardized Illumina protocol. Samples were amplified in 25 μL reactions using the Kapa HotStart ReadyMix DNA polymerase (Roche Sequencing Store) with the following PCR settings: 95°C for 3 min, 30–35 cycles (depending on the amount of DNA) including 95°C for 30 s, 58°C–52°C for 30 s, 72°C for 1 min, and a final elongation step of 5 min at 72°C. All PCRs were conducted in replicates. Successful amplification was confirmed on 1% agarose gels. PCR products were purified with the AMPure XP bead-based reagent purification kit (Beckman Coulter Life Sciences). After the first amplification with overhangs had been performed, dual indexes barcodes (Nextera XT index kit v2 Set A, B, C) were ligated to the adapter sequences allowing multiplexing of PCR products from different samples. For the barcoding PCRs, the KAPA Hotstart ReadyMix DNA polymerase was also used in final volumes of 50 μL. PCR settings were: 95°C for 3 min, 8 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final elongation step of 5 min at 72°C. PCR products from each sample were pooled and purified using the AMPure XP bead-based reagent purification kit. DNA concentrations after barcoding and purification were measured with a Qubit dsDNA BR assay (ThermoFisher) and adjusted to 45 nM.

2.6 | Short-Read Sequencing and Bioinformatic Analysis

Pooled and purified amplicons were sequenced in paired-end mode on an Illumina MiSeq 2 × 300 platform (Genewiz, Germany GmbH, Leipzig). Quality filtering of reads,

identification of amplicon sequencing variants (ASVs), and taxonomic affiliations were conducted with the R-package *DADA2* v.1.26.0 (Callahan et al. 2016). Due to the poor quality of reverse sequences (Figure S6A), only the forward reads were retained for subsequent bioinformatic analysis (Figure S6B). A total of approximately 17,000,000 high quality forward reads were obtained, while only 8000 reads were recovered when using paired end reads (forward and reverse) (Figure S7). The final dataset with forward reads only included 87 samples (79 field samples, five field controls, and three lab controls). Forward reads were trimmed at position 280, primers were removed (*TrimLeft*) and reads with ambiguous nucleotides or with a maximum number of expected errors (*maxEE*) exceeding 2 were filtered out using the function *filterAndTrim()*. Chimeric sequences were identified and removed using DADA2's consensus-based chimera removal approach (*BimeraDenovo, pooled method*). To assign taxonomy to ASVs, the default RDP naive Bayesian classifier method was used with PR2 v5.0.0. *(assignTaxonomy)*, including a minimum bootstrap confidence threshold (minBoot = 50). Under this setting, taxonomic assignments were truncated when bootstrap support fell below the threshold. At the end of the analyses, a total of 12,627,934 eukaryotic reads and 17,543 ASVs were obtained.

2.7 | Long-Read Amplification and Library Preparation

Samples were amplified with the LR primer pair F-566—R-1200 in 25 μL reaction volumes using the KAPA HotStart ReadyMix DNA polymerase (Roche Sequencing) and the following PCR protocol: 94°C for 50 s, 25–30 cycles including 94°C for 50 s, 63°C for 50 s, 72°C for 1 min, and a final elongation step of 10 min at 72°C. Successful amplification was confirmed by 1% agarose gel electrophoresis with subsequent staining with Midori Green. Amplicons from single PCR reactions were purified with AMPure XP (Beckman Coulter Life Sciences) before LR sequencing at OHMX.bio (Ghent, Belgium). The libraries were prepared with the Native Barcoding kit 96 V14 (SQK-NBD114.96, ONT) and the manufacturer's protocol. Uniquely barcoded DNA amplicons were pooled, purified (0.4× AMPureXP beads), adapter-ligated, and repurified. Libraries were quantified (Qubit dsDNA HS assay, ThermoFisher) before adding Sequencing Buffer (SB) and Loading Beads for sequencing.

2.8 | Long-Read Sequencing and Bioinformatic Analysis

Sequencing was performed on a GridION R10.4 flow cell, preliminarily primed with Flow Cell Flush (FCF), Bovine Serum Albumin (BSA), and Flow Cell Tether (FCT) as priming mix. Twenty femtomoles of pooled libraries were loaded three times through the SpotON sample port and sequenced for 72 h with MinKNOW high-accuracy base calling; altogether, 22.5 million raw LRs were generated. The quality of the raw sequencing reads was controlled with PycoQC v.2.5.2, in which sequences were basecalled, demultiplexed with Guppy v6.4.8 (Nanopore) and low-quality LR data were removed with the > PHRED9 parameter. Then, the quality of demultiplexed FASTQ files was visualized with *NanoPlot* v1.46.2 (De Coster and Rademakers 2023). Based on the primer pair, the expected LR amplicon size was approximately 750 bp. LRs were filtered using a length range of 600–1000 bp to retain target amplicons, while allowing for expected variation in amplicon length. The resulting read-length distribution corresponded to the expected amplicon size, with a median read length of 762 bp (Figure S9). Therefore, LRs ranging between 600 and 1000 bp were kept with the *chopper* tool v0.12.05 and with read quality scores exceeding the median score value of the entire LR dataset (i.e., q = 14; Figure S8) (De Coster and Rademakers 2023). Profiling was done with EMU v3.5.0 (Expectation–Maximization algorithm) with the *keep-counts* command, and the PR2 v5.0.0_emu reference database. Filtered long reads were aligned to the PR2 reference sequences using *minimap2* v2.24, as implemented within the EMU pipeline. This pipeline is designed to estimate relative abundances at the species-level in two steps: first, it aligns reads to a reference database, and second, it applies an expectation–maximization-based error correction. Specifically, it assigns relative abundances by computing the likelihood that each read comes from each species using probabilistic alignment and iterating to refine initial estimates. This method improves the accuracy of community profiles at the genus and species levels, particularly with error-prone reads (Curry et al. 2022). Our LR approach generated a total of 13,898,821 long reads and identified 1752 species.

2.9 | Taxonomic Filtering in Protist Metabarcoding Datasets

Samples from the negative controls were excluded from the downstream analysis of both SR and LR metabarcoding datasets. For the SR metabarcoding dataset, all amplicon sequence variants (ASVs) were pooled to the species level using the *tax_glom* function *in phyloseq*; (McMurdie and Holmes 2013) to facilitate comparisons with the LR dataset. Given that the focus of this study was on protists, the following taxonomic groups were retained: Amoebozoa, Archaeplastida, CRUMs, Excavata, Haptista, TSAR, and Cryptista kingdoms. Within the Obazoa kingdom, the Breviatea division and the Ancyromonadida, Nibbleridia, and Apusomonada supergroups were included, as well as the genus *Tunicaraptor*. In the Opisthokonta domain, the Choanoflagellata division was considered (Figure S10). Sequences associated to the domains Bacteria, Fungi, Metazoans, the Ichthyosporea, the Streptophyta supergroup, as well as sequences corresponding to chloroplasts, mitochondria and any unclassified taxa at the domain level were excluded to ensure accurate analyses of protist data. Additionally, unassigned taxa at the species level were also removed. In total, 7,438,382 reads were identified as protists and classified as 1039 species based on the SR metabarcoding dataset, while 7,623,617 reads were identified as protists in 1180 species from the LR metabarcoding dataset. The relative abundance of protist taxa was calculated for each dataset including the taxonomic groups described above.

All bioinformatic analyses were conducted on a local server to increase computational efficiency.

2.10 | Statistical Analysis

Visualization of data and statistical analyses were performed in R version 4.1.0. (R Core Team 2021) using the *phyloseq* (McMurdie and Holmes 2013), the *ggplot2*, and the *vegan* packages (Wickham 2016). As eDNA metabarcoding data are compositional and influenced by sequencing depth, read counts were normalized across samples prior to alpha diversity analysis. For this purpose, samples within each dataset were rarefied to the lowest sequencing depth observed (10,131 reads for both datasets). Rarefaction is commonly applied in metabarcoding studies to reduce biases associated with uneven sequencing effort (Schloss 2024), and unequal sampling effort (Bruce et al. 2021; Ramond et al. 2021). To avoid any further bias by unequal sample size, Random Under-Sampling (RUS) was applied using the base R sample function (Dittman et al. 2014). After rarefaction and resampling, both SR and LR datasets contained 76 samples. Alpha diversity indexes (richness, Shannon, Gini-Simpson (1-D), and Chao1) were calculated to describe and compare diversity of protist communities. The Kruskal–Wallis test was used to evaluate if protist alpha diversity significantly differed among stations, and the Wilcoxon test was applied to test for differences in alpha diversity between metabarcoding methods. To further describe protist patterns, three thresholds for rare taxa were defined for each dataset (SR and LR datasets, respectively): (i) species were considered as abundant if they were present in more than 0.1% of the average relative read abundance, (ii) rare species were present in less than 0.001% of the average relative read abundance, and (iii) intermediate rare species contributed to 0.001 and 0.1% of the average read relative abundance (Logares et al. 2014). NMDS ordination plots with a centered log-ratio (CLR) transformation were constructed to evaluate whether protist communities were structured according to the coastal-offshore gradient; NMDS plots used Bray-Curtis similarities based on a non-transformed number of rarefied reads for each dataset. Then, PERMANOVA analyses were used to investigate what proportion of the variance in community composition was explained by the location (Anderson 2017).

3 | Results

3.1 | Numbers of Sequencing Read and Taxonomic Assignment of SR and LR Metabarcoding

Potential contamination was assessed first from negative field and lab controls. Protist sequences detected in the field samples were genuine and not the result of contamination, given that field samples clearly contained much higher frequencies of protist taxa (98% for SR and 85% for LR metabarcoding) than the negative controls (Figure S11).

A total of 17 million eukaryotic reads were generated using SR Illumina sequencing, whereas LR Nanopore sequencing produced 22 million reads. After removing primers and filtering, approximately 14 million reads were retained for the LR, and 13 million for the SR dataset for the 79 field samples. Both sequencing platforms detected a similar number of reads being assigned to protists. SR produced a total of 7,438,382 reads identifying 1039 species, and the LR metabarcoding dataset resulted in 7,623,617 protist reads, with 1180 identified species. In terms of relative abundance of total reads (%) for each dataset, protists contributed more than 60% to read relative abundance in both datasets, with SR exhibiting a 10% higher relative abundance of reads as compared to LR (Figure S12). The taxonomic assignment of protists varied greatly between the two sequencing methods. The LR metabarcoding dataset contained only 1.4% of reads being unassigned to protists at the species level and 0% at the genus level. In contrast, the SR metabarcoding dataset had a 10 times higher proportion of unassigned reads, with 12% being unassigned at the genus level and 15.2% at the species level (Figure S13).

3.2 | Protist Composition Across Taxonomic Levels

With both metabarcoding techniques, TSAR (Telonemia, Stramenopiles, Alveolata, and Rhizaria) was the most prevalent kingdom, being identified as 87.4% of relative read abundances in the LR and 81.7% in the SR dataset (Figure 1A), respectively. The second most dominant kingdom, Haptista, contributed 4.2% of relative read abundances in the LR and 7.2% in the SR dataset. At the supergroup level, Stramenopiles and Alveolata were the most dominant groups. Stramenopiles dominated the LR dataset with 43.5%, as compared to 33.2% in the SR dataset (Figure 1B). In contrast, Alveolata was 38.4% more abundant in the SR dataset as compared to 35.9% in the LR dataset. This pattern was similar at lower taxonomic levels (Figure 1C–H). Specifically, Gyrista was the most abundant division in LR (40.2%), while in SR, Gyrista (29.5%) and Dinoflagellata (28.5%) were nearly equally represented (Figure 1C). At the class level (Figure 1D), Mediophyceae prevailed in the LR (24.5%) and Dinophyceae in the SR sequencing reads (19.2%). Spirotricheae were present in both datasets in comparable read relative abundances (LR: 9.5%, SR: 8.4%) while Syndiniales, Prymnesiophyceae, and Filosa-Thecofilosea were slightly more abundant in SR. The order Thalassiosirales dominated in LR (12%) and Gymnodiniales in SR (10.4%) datasets, respectively (Figure 1E). Similarly, at the family level, Thalassiosiraceae (7.8%) and Gymnodiniaceae (8.6%) prevailed in the LR and SR results, respectively (Figure 1F). The genus *Thalassiosira* was more prevalent in the LR dataset (7.5%), while *Gyrodinium* was more abundant in the SR dataset (Figure 2G). It is worth noting that the SR dataset did not detect the same dominant diatom genera as LR: *Chaetoceros* (5.6%) was most frequent in SR, while *Thalassiosira* (4.1%) was dominant in LR. The species *Thalassiosira curviseriata* was dominant in LR (4%), while *Phaeocystis globosa* (4.7%) and the heterotrophic dinoflagellate *Gyrodinium fusiforme* (1.8%) were most abundant in SR (Figure 1H).

3.3 | Comparing the Identification of Unique Taxa Between SR and LR Metabarcoding

Venn diagrams showed a high congruence between the two metabarcoding approaches in the detection of common taxa at higher taxonomic ranks, with progressively fewer shared and more unique taxa at lower taxonomic levels (Figure 2A–H). Both SR and LR datasets identified the same kingdoms (Figures 1A and 2A), and at the supergroup and division levels, approximately 90% of detected taxa were shared between the two approaches (Figure 2B,C). SR detected three unique

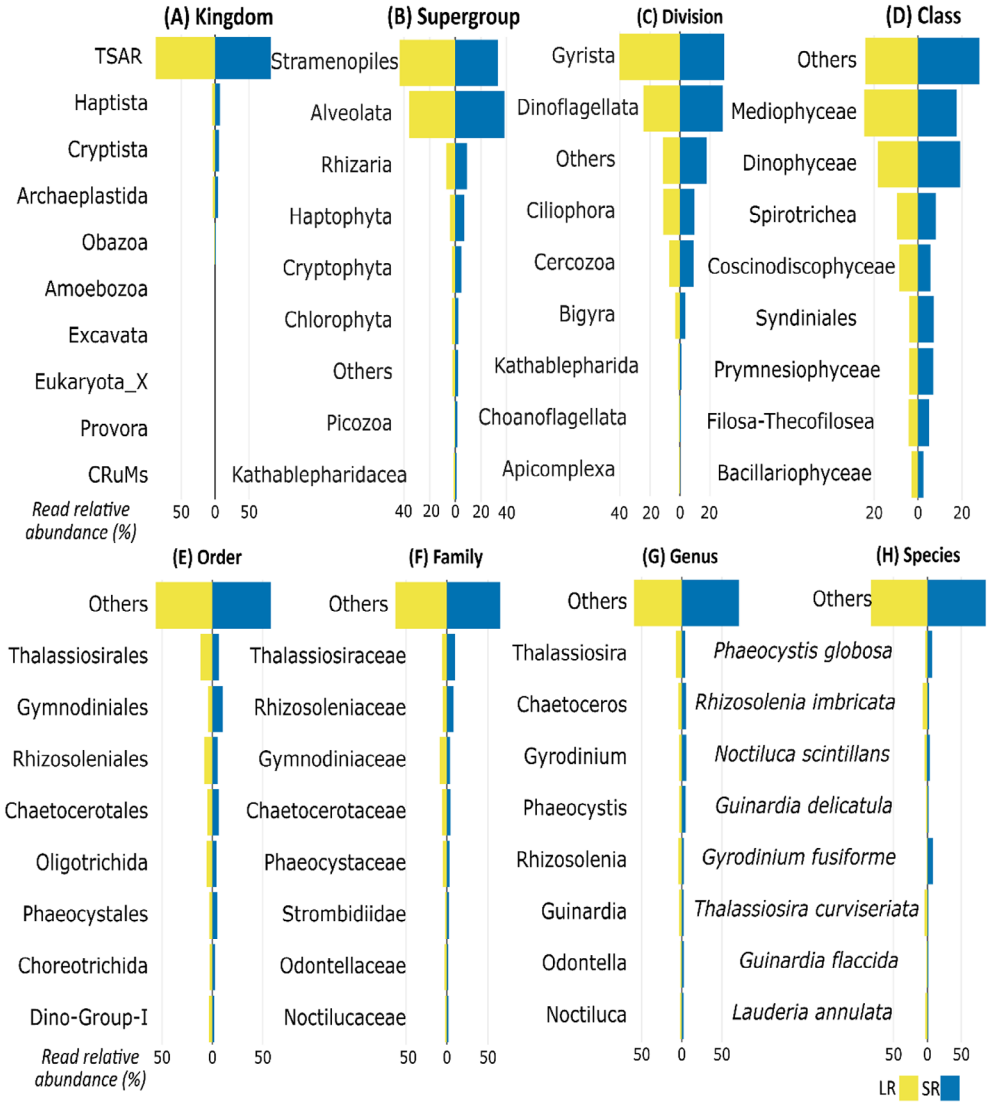

FIGURE 1 | Dual-horizontal histograms of protist compositions based on read relative abundance (%) obtained through LR (yellow) and SR metabarcoding (blue) across the following taxonomic ranks: (A) Kingdom, (B) Supergroup, (C) Division, (D) Class, (E) Order, (F) Family, (G) Genus, and (H) Species, in 79 samples. In total, 7,438,382 reads were identified as protists and classified as 1039 species based on the SR metabarcoding dataset, while 7,623,617 reads were identified as protists in 1180 species from the LR metabarcoding dataset.

supergroups (Figure 2B) (Amoebozoa_X, Mantamonadidae, and Rhodelphidia). The difference between the two approaches became more pronounced at the order and family levels, where 108 families (18%) were only identified in the LR dataset, whereas 82 families (14%) were unique to the SR dataset (Figure 2E,F). At genus and species levels, the percentage of common taxa further declined, with less than 60% and 50% of taxa being shared between the two methods (Figure 2G,H). Overall, the LR dataset showed a higher proportion of unique taxa from the family to species levels (Figure 2). The intersection part of the Venn diagram, illustrating the shared taxa between the two datasets, was dominated by "abundant" taxa with a relative read abundance of more than 0.1%. (Table S2, Figure 2). In contrast, uniquely detected taxa were rare or "intermediate rare" taxa (Table S2). Only a few unique genera were abundant (see Table S3). Eleven abundant genera were detected exclusively with LR, while three were unique to SR (Table S3). For example, only LR detected the autotrophic protists *Gephyrocapsa* (ex. *Emiliania*, Haptophyta) and *Bellerochea* (Bacillariophyceae), while the flagellates *Torodinium* and Telonemia-Group-2 were only identified with SR (Table S3). When comparing the distribution of identified rare taxa at various taxonomic levels between the two datasets, the SR dataset consisted mostly of abundant taxa, while the LR dataset contained more rare taxa (Table S2).

3.4 | Protist Diversity and Community Structure Across a Coastal-Offshore Gradient

The diversity indexes Richness, Shannon and Gini-Simpson (1-D) showed a slightly decreasing trend from the coastal to the offshore station for both datasets, with no significant differences among locations with SR (Figure 3A–C; Tables S4–S7). In contrast, in the LR dataset, significant differences among locations were observed for all diversity indexes (i.e., Richness, Shannon, Gini-Simpson, Chao1; Tables S6 and S7). In contrast to Shannon and Simpson indexes, the Chao index reached higher mean values with LR (Figure 4D). Protist

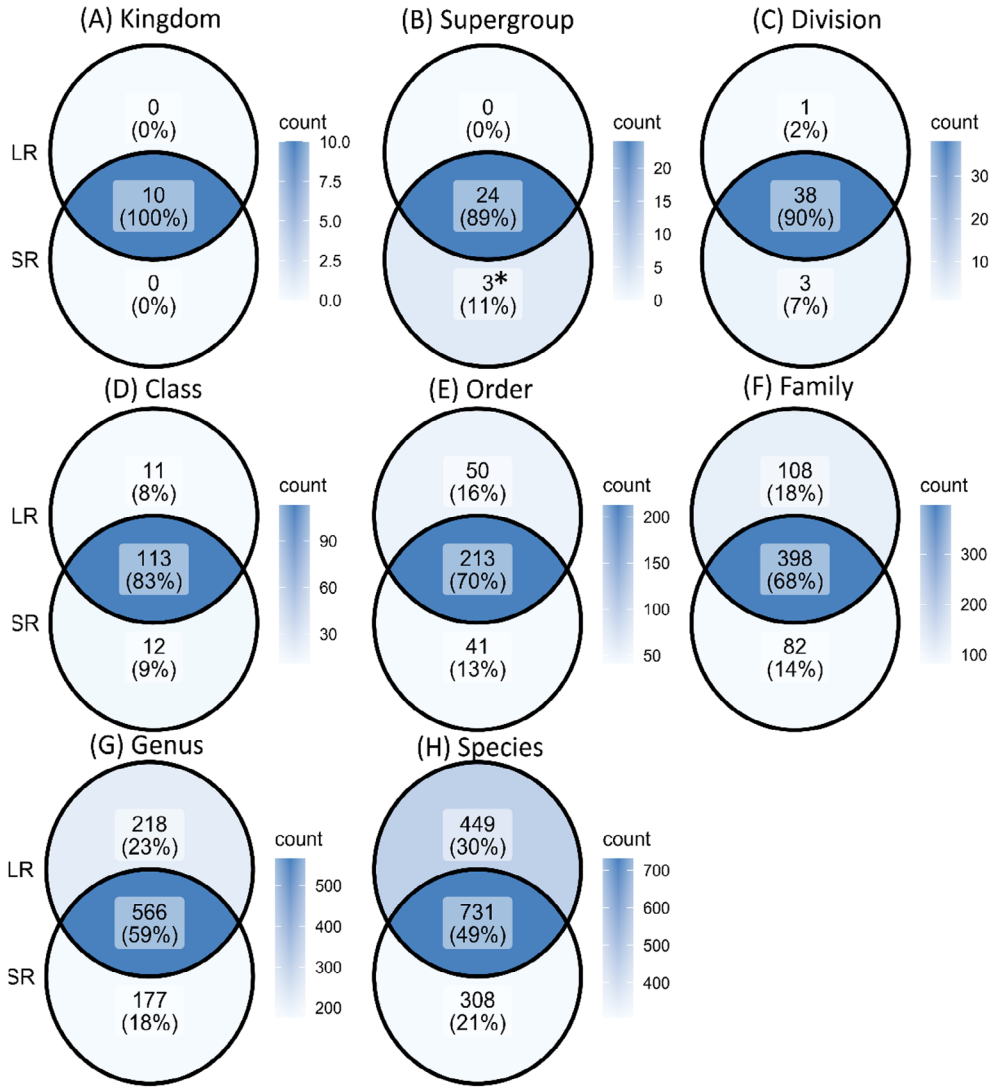

FIGURE 2 | Venn diagrams illustrating the absolute number of protist taxa and their relative abundance (%) as detected with LR, and SR metabarcoding approaches across the following taxonomic ranks: (A) Kingdom, (B) Supergroup (the asterisk indicate the unique Supergroup detected in SR: Amoebozoa_X, Mantamonadidae, and Rhodelphidia), (C) Division, (D) Class, (E) Order, (F) Family, (G) Genus, and (H) Species in 79 samples. The intersecting area indicates the common taxa identified by both methods. In total, 7,438,382 reads were identified as protists and classified as 1039 species based on the SR metabarcoding dataset, while 7,623,617 reads were identified as protists in 1180 species from the LR metabarcoding dataset.

communities displayed the highest observed richness in the LR dataset (e.g., 455) at the coastal station MOW1 (Figure 3A, Table S6). The rarefaction curves showed that a lower sampling size is sufficient for LR to capture the same protist diversity as with SR (Figure 3E). A comparison between the mean ranks with the Wilcoxon test revealed significant differences between the two metabarcoding approaches in Shannon, Gini-Simpson and Chao1 alpha diversity indexes but not in richness (Table S7), although mean richness was slightly lower with LR (Table S6). In addition, alpha diversity indexes (except Chao1) of the LR dataset showed higher variability than those of the SR dataset (Figure 4A–E).

NMDS and PERMANOVA analyses indicated that protist communities differed significantly among stations ($p < 0.001$ for both datasets); however, only 7% and 9% of the total variation was explained by location in SR and LR, respectively (Figure 4). The coastal-offshore gradient was somewhat more noticeable in the NMDS of the LR dataset. When comparing spatial patterns of the composition of protist communities at the order level, both methods detected the same orders along the coastal-offshore gradient, albeit in different proportions (Figure 5). Both methods showed that the diatom order Mediophyceae dominated the coastal station, while at the transitional and offshore stations, Dinophyceae were more abundant, again with both SR and LR.

4 | Discussion

4.1 | Sequencing Yield

Regarding overall sequencing yields, both SR and LR metabarcoding approaches were equally suitable, with LR producing somewhat higher overall numbers of sequencing reads than SR (22 million vs. 17 million raw reads, and 14 million as compared

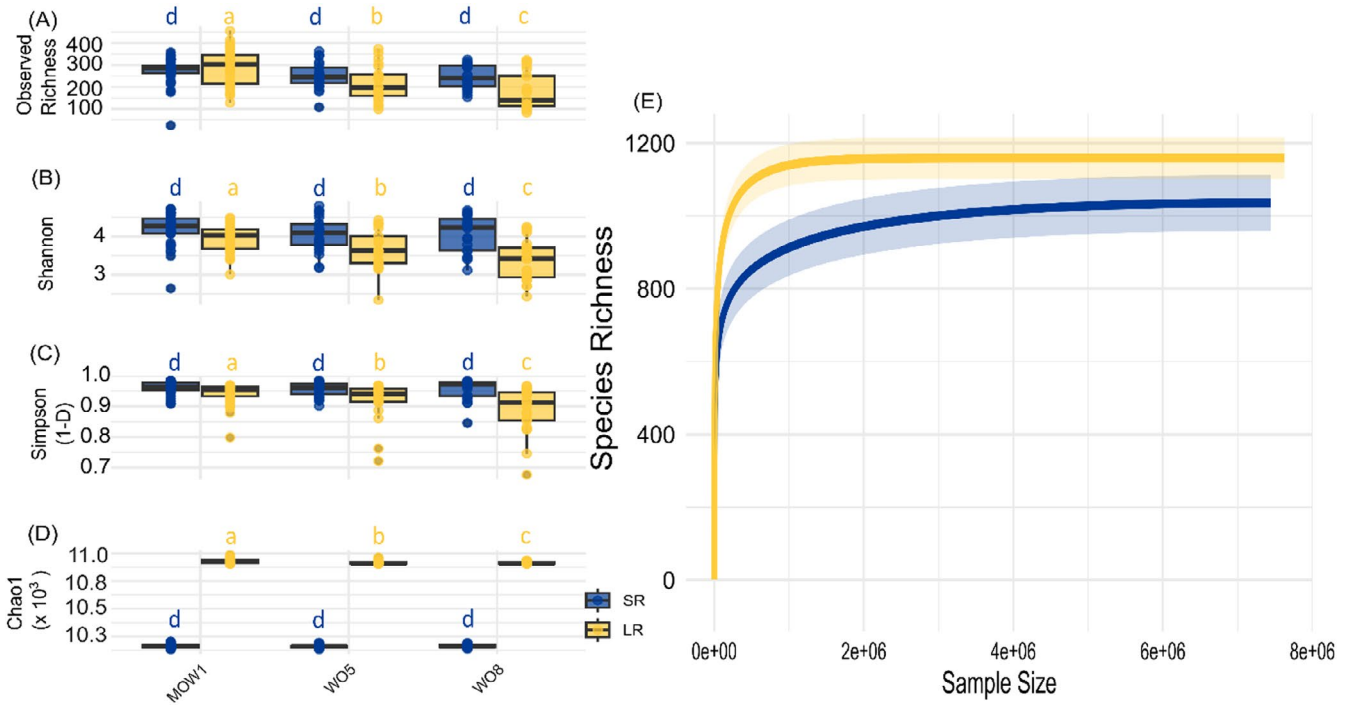

FIGURE 3 | Comparison of protist species α-diversity indexes in: (A) observed species richness, (B) Shannon index, (C) Gini-Simpson (1-D), (D) Chao1 index; for short-read (SR, blue) and long-read metabarcoding approaches (LR, yellow) and three different sampling locations (MOW1; WO5; WO8), (E) Rarefaction curves of protist species richness obtained through short-read (SR, blue) and long-read metabarcoding (LR, yellow) approaches with increasing read counts (i.e., sequencing depth). Significant differences between geographic locations (with Kruskal–Wallis tests) are indicated by different small letters (a, b, c), the absence of significance is indicated with the small letter d. The SR and LR datasets were rarefied at 10,131 reads per sample. Each dataset contained a total of 769,956 protist rarified reads in 76 samples. A total of 1039 and 1180 species were detected in SR and LR respectively.

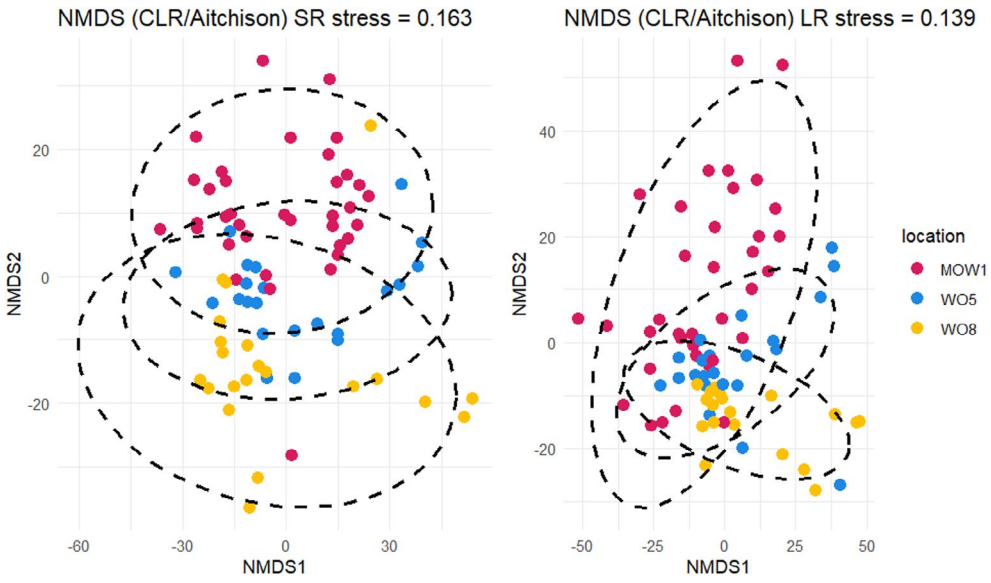

FIGURE 4 | Ordination with multidimensional scaling (NMDS) and CLR transformation of protist communities for short-read (SR, left) and long-read metabarcoding (LR, right) methods and across geographic sampling locations (MOW1, pink; WO5, blue; WO8, yellow, 79 samples in total). Each dot represents one individual sample, and each color represents a sampling station. In total, 7,438,382 reads were identified as protists and classified as 1039 species based on the SR metabarcoding dataset, while 7,623,617 reads were identified as protists in 1180 species from the LR metabarcoding dataset.

to 13 million filtered reads). If only sequencing reads being assigned to protists (~7.6 million vs. ~7.4 million) and their species richness (1180 vs. 1039) were considered, again, both approaches performed alike and showed a dominance of protists. A slightly greater relative abundance of protists (> 10%) was observed with SR than with LR (Figure S12).

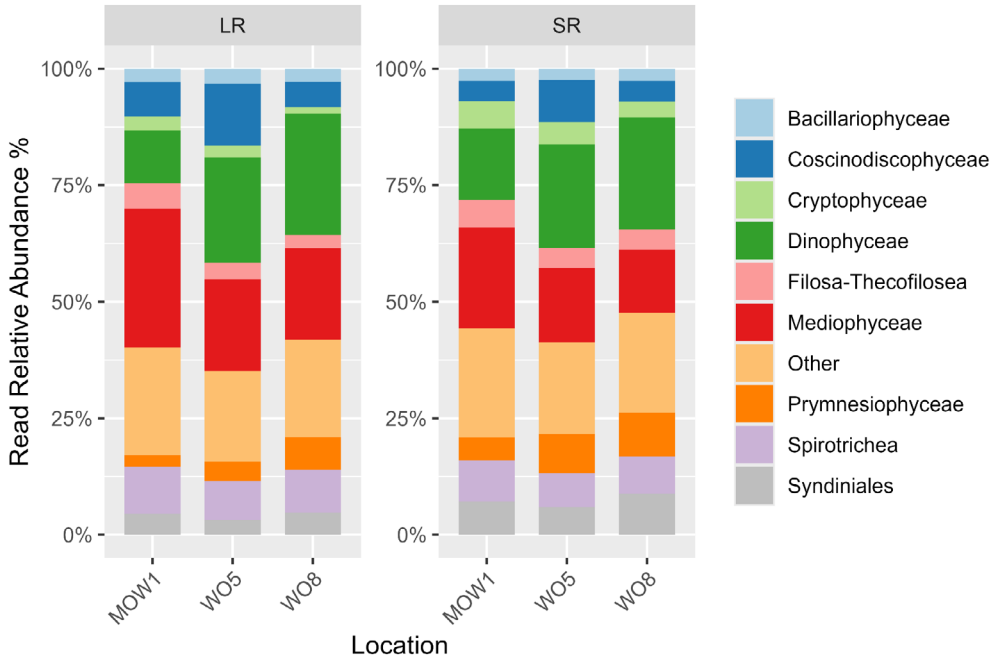

FIGURE 5 | Relative read abundance of protist reads for short-read (SR, right) and long-read metabarcoding (LR, left) techniques and the three sampling locations (MOW1; WO5; WO8) in 79 samples. In total, 7,438,382 reads were identified as protists and classified as 1039 species based on the SR metabarcoding dataset, while 7,623,617 reads were identified as protists in 1180 species from the LR metabarcoding dataset.

4.2 | Taxonomic Assignment

A higher proportion of unassigned taxa was observed with the SR method (Figure S13), matching the metabarcoding results of protist communities along the coast of Texas (Gaonkar and Campbell 2024). Furthermore, taxa identified by both methods predominately belonged to higher taxonomic levels (Figure 2), which is consistent with findings by Gaonkar and Campbell (2024). In the latter study, sequencing reads covering the entire 18S rDNA region provided more accurate taxonomic assignments with fewer unassigned reads when compared to results based on the V4 (380bp) and V8–V9 regions (330 bp). Although our study amplified only a part of the 18S rDNA region using both methods, LR still provided more reliable taxonomic identifications than SR (i.e., SR: 15.2% vs. LR: 1.2% unassigned species). This suggests that the longer 18S rDNA fragment being analyzed by the LR approach in our study was more effective in capturing a broad range of taxa. Differences in taxonomic assignment between SR and LR datasets may partly be attributed to different bioinformatic pipelines including different classifier methods (Mugnai et al. 2023; Hleap et al. 2021). Expectation–maximization algorithms such as EMU can distinguish closely related taxa or assign unknown sequences to their nearest available reference (Curry et al. 2022). This probabilistic framework is designed to account for sequencing errors inherent to LR sequencing data. Unlike the naïve Bayesian classifier used for SR data with DADA2, EMU does not apply a fixed bootstrap confidence cutoff but instead estimates relative abundances based on probabilistic alignment to reference sequences. Consequently, EMU may yield deeper taxonomic assignments with the same reference database, even when sequence identity is lower. As EMU tends to assign reads to the nearest available reference even with limited similarity, whereas the naïve Bayesian classifier that was used for SR analyses truncates assignments when confidence is low, a higher proportion of species-level assignments in LR data may reflect algorithmic permissiveness. In our study, the use of different classifiers followed commonly applied, technology-specific workflows for SR and LR metabarcoding.

In addition, it is worth noting that while recent improvements in Oxford Nanopore library kits and flow cell chemistry have substantially reduced chimeras and sequencing error rates (Wang et al. 2021), PCR chimeras, and systematic indel and homopolymer errors might remain potential problems in LR amplicon datasets and may contribute to inflated richness or spurious species-level assignments. The EMU pipeline used in this study is optimized for error-prone reads, yet it does not include an explicit chimera detection or removal step, Consequently, undetected chimeras may have contributed to elevated richness or deeper taxonomic assignments in the LR dataset, and results at the species level should therefore be interpreted with caution. Reference based chimera detection tools such “uchime2_ref” have been applied to higher quality long reads produced by PacBio or Nanopore using unique molecular identifiers during library preparation before sequencing (Karst et al. 2021). However, these approaches might remove genuine biological sequences or fail to detect chimeras when sequencing error rates are high (Stock et al. 2025). Finally, custom pipelines such as CONCOMPRA have been recently applied to the analyses of prokaryotic communities, but to our knowledge these approaches have not yet been implemented in eukaryotic LR metabarcoding (Stock et al. 2025).

Differences in the detection of protist taxa between the two methods were particularly evident at finer (lower) taxonomic levels (Figure 2). Both datasets recovered most of the abundant

taxa, while unshared taxa were typically rare (Table S3). Notably, more rare taxa were detected in the LR metabarcoding dataset (Table S2). This is consistent with previous findings that ONT long sequencing was better suited for detecting rare taxa in a microbiome study (Szoboszlay et al. 2023). To account for the potentially higher detection of rare taxa in our study, we applied an abundance threshold. Without this control, the probabilistic model used by the EMU algorithm in LR analyses might artificially inflate the number of rare species (Curry et al. 2022).

4.3 | Composition Differences and Identification of Specific Taxa

We expected congruence between the SR and LR metabarcoding approaches in the relative read abundance at high taxonomic levels (Figure 1), in line with the findings by Gaonkar and Campbell (2024). However, at lower taxonomic levels, notable differences in protist relative abundance were observed between the two methods. Specifically, Dinophyceae (dinoflagellates), and closely related groups (e.g., Syndiniales), were more abundant in the SR dataset, while the LR dataset showed a prevalence of Bacillariophyceae (diatoms, Figure 1). Our results thus confirm one of the known limitations of SR metabarcoding; the over-presentation of dinoflagellates due to their high rDNA copy numbers (Prokopowich et al. 2003; Georges et al. 2014; Bradley et al. 2016 and reference therein; Santi et al. 2021; Yeh et al. 2021). This limitation is especially relevant in diatom-dominated marine systems such as the BPNS (Nohe et al. 2020). In contrast, the LR approach appeared to mitigate this bias, probably due to several possible reasons, besides the differences in bioinformatic pipelines mentioned in the previous section: (1) Technological differences: SR methods rely on bridge amplification in Illumina sequencing, which might introduce amplification bias. In contrast, LR sequencing uses ligation and direct sequencing, which may reduce such bias (Mikheyev and Tin 2014) and help correct the over-representation of dinoflagellates. (2) Completeness of reference database: The dinoflagellate reference database remains limited at the genus/species level, potentially further contributing to biased classifications with SR methods (Gaonkar and Campbell 2024; Mordret et al. 2023). Conversely, diatom reference databases are more complete, enhancing the LR's ability to accurately detect and classify them (Gaonkar and Campbell 2024). An alternative approach to address database limitations would be taxonomy-free assignments, such as de novo assembly. While promising and already applied to microbial profiling (Stock et al. 2025) these approaches currently require metagenomic or long read sequencing data. (3) Intracellular haplotype diversity: Certain dinoflagellates, such as *Tripos* sp., exhibit high single-cell haplotype diversity, where a single cell may contain multiple haplotypes (Huang et al. 2024). In our study, *Tripos* was a dominant genus (ca. 0.1% in relative abundance) in both datasets. Such intracellular variation can affect the clustering of short reads, leading to inflated diversity at lower taxonomic levels by misclassifying intraspecific variation, as separate OTUs (Operational Taxonomic Units) (Stoeck et al. 2024; Huang et al. 2024). We minimized this bias by using non-clustering techniques such as amplicon sequence variants (ASVs) for SR and EMUs for LR. (4) Primer bias: Primers targeting specific parts of the 18S rDNA gene often have narrow taxonomic coverage, which can affect detection efficiency (Vaulot et al. 2022) and could have further contributed to discrepancies between the two methods.

Interestingly, while the SR approach identified a greater number of abundant species, the LR method detected a higher number of unique, abundant taxa (Table S2). This difference is particularly important, as only the LR approach identified key taxa that have been widely reported in the BPNS using conventional methods. Two notable autotrophic protists are the colonial diatom *Bellerochea* and the calcifying haptophyte *Gephyrocapsa* (Table S3). *Bellerochea* blooms have been recently intensified in the BPNS, likely due to temperature-driven declines in copepod populations (Mortelmans et al. 2024). The future intensification of these blooms is alarming as they might reduce oxygen levels in the water column, potentially threatening higher trophic levels, including larval fish (Mortelmans et al. 2024). *Gephyrocapsa*, on the other hand, is ecologically important due to its substantial production of biomass and calcium carbonate. It contributes to carbon dioxide uptake and emits dimethyl sulfide, a gas with climate-cooling properties (Paasche 2001). This haptophyte forms extensive blooms that can span thousands of kilometers in summer and autumn (Raitsos et al. 2006). These blooms can occur locally in the North Sea (Holligan et al. 1993), or they can be further transported via the inflow of Atlantic waters (Head et al. 1998). To date, most research on this haptophyte has relied on satellite data to characterize its spatial distribution (Ladd et al. 2018; Terrats et al. 2020), while studies on its phenology using eDNA approaches remain very limited (Neri et al. 2025). Future studies on *Gephyrocapsa* in the BPNS should integrate eDNA and satellite technologies to better resolve the spatial and temporal dynamics of this haptophyte at local and regional scales.

4.4 | Coastal-Offshore Gradient Influencing Protist Diversity and Community Structure

Although both SR and LR methods revealed similar trends in protist diversity patterns, community structure, and composition at higher taxonomic levels along the coastal-offshore gradient (Figures 3–5, and Table S1), notable differences emerged in the estimated diversity metrics between the two methods (Figures 3 and 5). Given the increased sequence context provided by LR metabarcoding, we initially expected LR to yield deeper taxonomic annotation and potentially identify a higher number of taxa under the applied workflow compared to SR. Discrepancies in diversity estimates can at least partially be explained by the higher number of rare taxa identified by LR, as mentioned in the previous section.

Because the Shannon diversity index is sensitive to the presence of rare taxa (Roswell et al. 2021), the prominence of such rare taxa in the LR dataset most likely influenced this index (Figure 3B; Table S2). Similarly, the Gini-Simpson index, which emphasizes taxon dominance, was lower in the LR dataset, reflecting its detection of fewer dominant taxa compared to SR (Figure 3C; Table S2; Roswell et al. 2021). In contrast, the Chao1 index (Figure 3D), which estimates species richness while correcting for rare taxa, may provide a more reliable estimate of protist richness in the LR dataset (Chao et al. 2006, 2020). Chao1's mathematical framework reduces bias resulting

from underrepresentation of rare species, making it particularly suited for use with high-resolution sequencing data (Figure 3D).

Similar inconsistencies in diversity estimates between SR and LR methods have been reported in microbiome research. For instance, Szoboszlay et al. (2023) found that SR data tended to overestimate bacterial diversity in fecal samples, likely due to sequencing noise. These findings emphasize the importance of selecting appropriate diversity indexes based both on the metabarcoding method used and the nature of the data.

Phylogenetic-based metrics such as the Faith's Phylogenetic diversity index (PD; Faith 1992) and UniFrac (Lozupone and Knight 2005) may offer more accurate insights for metabarcoding data by incorporating evolutionary relationships. However, the primary aim of the current study was not to evaluate all available diversity indexes, but rather to provide a first comparison of the most commonly used indexes within the context of eDNA-based protist community analyses.

In contrast to the diversity indexes, protist community structure differed significantly across sampling stations, although geographic location accounted for less than 10% of the variability for both methods (Figure 4). Nonetheless, the differences in protist communities along the coastal-offshore gradient were more pronounced in the LR dataset. This pattern may be attributed to the greater depth of taxonomic assignment and the detection of additional low-abundance taxa under the LR metabarcoding workflow, which likely captured subtle shifts in community composition. At higher taxonomic levels, both methods provided consistently detected a dominance of Mediophyceae (diatoms) at the coastal station, while Dinophyceae (dinoflagellates) prevailed at the transient and offshore stations (Figure 5). The dominance of diatoms at the coastal station MOW1 can probably be explained by riverine inputs from the River Schelde (Aubert et al. 2022).

4.5 | Recommendations

Given the inherent limitations of each method, the choice between SR and LR metabarcoding for marine protist monitoring studies depends heavily on the goals of the study. SR metabarcoding remains a reliable and cost-effective tool for detecting abundant taxa and providing broad snapshots of community diversity. Its use in monitoring for almost a decade (Teeling et al. 2016; Lambert et al. 2019; Caracciolo et al. 2022; Skouroliakou et al. 2024) has enable consistent data generation over time. One of its key advantages is its established use.

As highlighted by Ducklow et al. (2009), long-term oceanographic surveys are crucial for capturing episodic events and assessing the impacts of climate change, processes that often unfold over extended timescales. Recent initiatives, such as the European Marine Omics Biodiversity Observation Network (EMO BON), represent promising initiatives to standardize and sustain SR-based eDNA monitoring across European waters (Santi et al. 2023).

Despite its advantages, SR metabarcoding has well-known limitations, including amplification bias, limited depth of taxonomic assignment, and the overrepresentation of dinoflagellates due to high rDNA copy numbers. In contrast, LR metabarcoding (particularly using ONT) is emerging as a powerful complement to SR approaches. We would recommend additional verification of species assignments with LR approaches in future studies with other analyses like BLAST searches, alignments or phylogenetics, and suggest to also use mock communities or positive controls but these are beyond the scope of the present study. Recent advances in ONT accuracy and throughput are addressing previous limitations related to error rates, making LR sequencing more accessible and suitable for marine studies (Wang et al. 2021). The application of LR metabarcoding in monitoring initiatives is promising and offers several distinct advantages: Unlike SR, LR metabarcoding might provide increased sequence context and, under certain workflows, offer deeper taxonomic assignment that may facilitate differentiation among closely related taxa; it may reduce amplification bias, resulting in more balanced community profiles; and it can enhance the detection of low-abundance taxa, thereby increasing sensitivity to subtle shifts in biodiversity. Last but not least, LR techniques quantify key phytoplankton genera more accurately, including diatoms and haptophytes such as *Bellerochea* and *Gephyrocapsa*, with pronounced harmful effects in the BPNS and other parts of the North Sea.

5 | Conclusions

Molecular methods for the long-term monitoring of marine ecosystems continue to evolve alongside advances in sequencing technologies. Although still underutilized in protist research, LR sequencing offers significant potential to address limitations of traditional SR metabarcoding. In the current study, we directly compared protist diversity, taxa detection, composition, and community structure using LR and SR metabarcoding approaches. While both metabarcoding methods showed consistent patterns in community composition at higher taxonomic ranks and similar trends along the coastal-offshore gradient in the BPNS, data from LR revealed less overrepresentation of dinoflagellates and a greater capacity to detect certain diatoms with harmful effects and haptophytes in general. These results underscore the value of integrating LR sequencing into future marine eDNA monitoring frameworks, particularly for tracking harmful and/or nuisance taxa and assessing fine-scale biodiversity dynamics across spatial gradients.

Author Contributions

Dimitra-Ioli Skouroliakou: conceptualization (equal), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), software (lead), validation (lead), visualization (equal), writing – original draft (lead), writing – review and editing (lead). **Deborah W. E. Dupont:** data curation (equal), formal analysis (equal), investigation (equal), methodology (equal), software (equal), validation (equal), visualization (equal), writing – original draft (equal), writing – review and editing (equal). **Yelle Vandenboer:** data curation (equal), investigation (equal), software (equal), writing – review and editing (equal). **Sofie D'Hont:** methodology (equal), validation (equal), writing – review and editing (equal). **Koen Sabbe:** conceptualization (equal), funding acquisition (equal), investigation (equal), project administration (equal), resources (equal), supervision (equal), validation (equal), writing – review and editing (equal). **Isa Schön:** conceptualization (equal), funding acquisition (equal), investigation (equal), project administration (equal),

resources (equal), supervision (equal), validation (equal), writing – review and editing (equal).

Acknowledgments

We acknowledge funding from the European Maritime, Fisheries and Aquaculture Fund (EMFAF) for the ZeroImpact project. We also thank Sophie Derycke for coordinating ZeroImpact and the Freshwater Biology team at the Royal Belgian Institute of Natural Sciences (RBINS) for their involvement in various sampling campaigns. This research was supported by the Belgian Science Policy (BELSPO) within the BRAIN-be program (BG-PART, contract number B2/202/P1/BG-PART), PiNS, contract number RV/21/PiNS, and EMBRC Belgium—FWO international research infrastructure I001621N and I000825N (infrastructural funding for PAE). We thank Marie Cours, Jeroen Venderickx, the crew of RV Belgica, and all other participants in the field campaigns for their valuable technical support. We acknowledge Willem Stock for his useful insights on bioinformatic analyses. Finally, we also thank the reviewers and editor for their comments and suggestions that helped to improve our manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Bioinformatic workflow for SR and LR analyses, processed input tables and scripts for all figures are available to the Zenodo repository: https://doi.org/10.5281/zenodo.18366431. Raw sequencing SR and LR data are available under the Sequence Read Archive of NCBI under BioProject number PRJNA1282445 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1282445).

References

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, no. 7: e6372. https://doi.org/10.1371/journal.pone.0006372.

Anderson, M. J. 2017. "Permutational Multivariate Analysis of Variance (PERMANOVA)." In *Wiley StatsRef: Statistics Reference Online*, 1–15. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat07841.

Aubert, A., O. Beauchard, R. de Blok, et al. 2022. "From Bacteria to Zooplankton: An Integrative Approach Revealing Regional Spatial Patterns During the Spring Phytoplankton Bloom in the Southern Bight of the North Sea." *Frontiers in Marine Science* 9. https://doi.org/10.3389/fmars.2022.863996.

Bar-On, Y. M., and R. Milo. 2019. "The Biomass Composition of the Oceans: A Blueprint of Our Blue Planet." *Cell* 179, no. 7: 1451–1454. https://doi.org/10.1016/j.cell.2019.11.018.

Beaugrand, G. 2004. "The North Sea Regime Shift: Evidence, Causes, Mechanisms and Consequences." *Progress in Oceanography* 60, no. 2: 245–262. https://doi.org/10.1016/j.pocean.2004.02.018.

Blauw, A. N., E. Benincà, R. W. P. M. Laane, N. Greenwood, and J. Huisman. 2012. "Dancing With the Tides: Fluctuations of Coastal Phytoplankton Orchestrated by Different Oscillatory Modes of the Tidal Cycle." *PLoS One* 7, no. 11: e49319. https://doi.org/10.1371/journal.pone.0049319.

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, no. 19: 5878–5891. https://doi.org/10.1128/AEM.01630-16.

Breton, E., V. Rousseau, J.-Y. Parent, J. Ozer, and C. Lancelot. 2006. "Hydroclimatic Modulation of Diatom/Phaeocystis Blooms in Nutrient-Enriched Belgian Coastal Waters (North Sea)." *Limnology and Oceanography* 51, no. 3: 1401–1409. https://doi.org/10.4319/lo.2006.51.3.1401.

Bruce, K., R. Blackman, S. J. Bourlat, et al. 2021. *A Practical Guide to DNA-Based Methods for Biodiversity Assessment*. Advanced Books. https://doi.org/10.3897/ab.e68634.

Burki, F., M. M. Sandin, and M. Jamy. 2021. "Diversity and Ecology of Protists Revealed by Metabarcoding." *Current Biology* 31, no. 19: R1267–R1280. https://doi.org/10.1016/j.cub.2021.07.066.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. "DADA2: High-Resolution Sample Inference From Illumina Amplicon Data." *Nature Methods* 13: 581–583. https://doi.org/10.1038/nmeth.3869.

Caracciolo, M., F. Rigaut-Jalabert, S. Romac, et al. 2022. "Seasonal Dynamics of Marine Protist Communities in Tidally Mixed Coastal Waters." *Molecular Ecology* 31, no. 14: 3761–3783. https://doi.org/10.1111/mec.16539.

Cardinale, B. J., J. E. Duffy, A. Gonzalez, et al. 2012. "Biodiversity Loss and Its Impact on Humanity." *Nature* 486, no. 7401: 59–67. https://doi.org/10.1038/nature11148.

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2006. "Abundance-Based Similarity Indices and Their Estimation When There Are Unseen Species in Samples." *Biometrics* 62, no. 2: 361–371. https://doi.org/10.1111/j.1541-0420.2005.00489.x.

Chao, A., Y. Kubota, D. Zelený, et al. 2020. "Quantifying Sample Completeness and Comparing Diversities Among Assemblages." *Ecological Research* 35, no. 2: 292–314. https://doi.org/10.1111/1440-1703.12102.

Chwalińska, M., M. Karlicki, S. Romac, F. Not, and A. Karnkowska. 2025. "From Short to Long Reads: Enhanced Protist Diversity Profiling via Nanopore Metabarcoding." *Metabarcoding and Metagenomics* 9: e163750.

Curry, K. D., Q. Wang, M. G. Nute, et al. 2022. "Emu: Species-Level Microbial Community Profiling of Full-Length 16S rRNA Oxford Nanopore Sequencing Data." *Nature Methods* 19, no. 7: 845–853. https://doi.org/10.1038/s41592-022-01520-4.

De Coster, W., and R. Rademakers. 2023. "NanoPack2: Population-Scale Evaluation of Long-Read Sequencing Data." *Bioinformatics* 39, no. 5: btad311. https://doi.org/10.1093/bioinformatics/btad311.

Dittman, D. J., T. M. Khoshgoftaar, and A. Napolitano. 2014. "Selecting the Appropriate Data Sampling Approach for Imbalanced and High-Dimensional Bioinformatics Datasets." In *2014 IEEE International Conference on Bioinformatics and Bioengineering*, 304–310. https://doi.org/10.1109/BIBE.2014.61.

Duarte, S., L. Simões, and F. O. Costa. 2023. "Current Status and Topical Issues on the Use of eDNA-Based Targeted Detection of Rare Animal Species." *Science of the Total Environment* 904: 166675. https://doi.org/10.1016/j.scitotenv.2023.166675.

Ducklow, H. W., S. C. Doney, and D. K. Steinberg. 2009. "Contributions of Long-Term Research and Time-Series Observations to Marine Ecology and Biogeochemistry." *Annual Review of Marine Science* 1: 279–302. https://doi.org/10.1146/annurev.marine.010908.163801.

Emeis, K.-C., J. van Beusekom, U. Callies, et al. 2015. "The North Sea — A Shelf Sea in the Anthropocene." *Journal of Marine Systems* 141: 18–33. https://doi.org/10.1016/j.jmarsys.2014.03.012.

Faith, D. P. 1992. "Conservation Evaluation and Phylogenetic Diversity." *Biological Conservation* 61, no. 1: 1–10. https://doi.org/10.1016/0006-3207(92)91201-3.

Falkowski, P. G., R. T. Barber, and V. Smetacek. 1998. "Biogeochemical Controls and Feedbacks on Ocean Primary Production." *Science* 281, no. 5374: 200–206. https://doi.org/10.1126/science.281.5374.200.

Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. "Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components." *Science* 281, no. 5374: 237–240. https://doi.org/10.1126/science.281.5374.237.

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, no. 4: e11232. https://doi.org/10.1002/ece3.11232.

Genitsaris, S., S. Monchy, E. Viscogliosi, T. Sime-Ngando, S. Ferreira, and U. Christaki. 2015. "Seasonal Variations of Marine Protist Community Structure Based on Taxon-Specific Traits Using the Eastern English Channel as a Model Coastal System." *FEMS Microbiology Ecology* 91, no. 5: fiv034. https://doi.org/10.1093/femsec/fiv034.

Georges, C., S. Monchy, S. Genitsaris, and U. Christaki. 2014. "Protist Community Composition During Early Phytoplankton Blooms in the Naturally Iron-Fertilized Kerguelen Area (Southern Ocean)." *Biogeosciences* 11, no. 20: 5847–5863. https://doi.org/10.5194/bg-11-5847-2014.

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, no. 2: e87624. https://doi.org/10.1371/journal.pone.0087624.

Head, R. N., D. W. Crawford, J. K. Egge, et al. 1998. "The Hydrography and Biology of a Bloom of the Coccolithophorid *Emiliania huxleyi* in the Northern North Sea." *Journal of Sea Research* 39, no. 3: 255–266. https://doi.org/10.1016/S1385-1101(97)00060-9.

Hleap, J. S., J. E. Littlefair, D. Steinke, P. D. Hebert, and M. E. Cristescu. 2021. "Assessment of Current Taxonomic Assignment Strategies for Metabarcoding Eukaryotes." *Molecular Ecology Resources* 21: 2190–2203. https://doi.org/10.1111/1755-0998.13407.

Holligan, P. M., S. B. Groom, and D. S. Harbour. 1993. "What Controls the Distribution of the Coccolithophore, *Emiliania huxleyi*, in the North Sea?" *Fisheries Oceanography* 2, no. 3–4: 175–183. https://doi.org/10.1111/j.1365-2419.1993.tb00133.x.

Huang, X., Y. Li, H. Du, and N. Chen. 2024. "Comparative Assessment of the Intragenomic Variations of Dinoflagellate Tripos Species Through Single-Cell Sequencing." *Marine Pollution Bulletin* 206: 116690. https://doi.org/10.1016/j.marpolbul.2024.116690.

Jamy, M., C. Biwer, D. Vaulot, et al. 2022. "Global Patterns and Rates of Habitat Transitions Across the Eukaryotic Tree of Life." *Nature Ecology & Evolution* 6: 1458–1470. https://www.nature.com/articles/s41559-022-01838-4.

Karst, S. M., R. M. Ziels, R. H. Kirkegaard, et al. 2021. "High-Accuracy Long-Read Amplicon Sequences Using Unique Molecular Identifiers With Nanopore or PacBio Sequencing." *Nature Methods* 18: 165–169. https://doi.org/10.1038/s41592-020-01041-y.

Ladd, C., L. B. Eisner, S. A. Salo, C. W. Mordy, and M. D. Iglesias-Rodriguez. 2018. "Spatial and Temporal Variability of Coccolithophore Blooms in the Eastern Bering Sea." *Journal of Geophysical Research: Oceans* 123: 9119–9136. https://doi.org/10.1029/2018JC014302.

Lambert, S., M. Tragin, J.-C. Lozano, et al. 2019. "Rhythmicity of Coastal Marine Picoeukaryotes, Bacteria and Archaea Despite Irregular Environmental Perturbations." *ISME Journal* 13, no. 2: 388–401. https://doi.org/10.1038/s41396-018-0281-z.

Logares, R., S. Audic, D. Bass, et al. 2014. "Patterns of Rare and Abundant Marine Microbial Eukaryotes." *Current Biology* 24, no. 8: 813–821. https://doi.org/10.1016/j.cub.2014.02.050.

Longobardi, L., L. Dubroca, F. Margiotta, D. Sarno, and A. Zingone. 2022. "Photoperiod-Driven Rhythms Reveal Multi-Decadal Stability of Phytoplankton Communities in a Highly Fluctuating Coastal Environment." *Scientific Reports* 12, no. 1: 3908. https://doi.org/10.1038/s41598-022-07009-6.

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, no. 6820: 603–607. https://doi.org/10.1038/35054537.

Lozupone, C., and R. Knight. 2005. "UniFrac: A New Phylogenetic Method for Comparing Microbial Communities." *Applied and Environmental Microbiology* 71, no. 12: 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005.

McMurdie, P. J., and S. Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PLoS One* 8, no. 4: e61217. https://doi.org/10.1371/journal.pone.0061217.

Meziti, A., E. Smeti, D. Daniilides, S. Spatharis, G. Tsirtsis, and K. A. Kormas. 2023. "Increased Contribution of Parasites in Microbial Eukaryotic Communities of Different Aegean Sea Coastal Systems." *PeerJ* 11: e16655. https://doi.org/10.7717/peerj.16655.

Mikheyev, A. S., and M. M. Y. Tin. 2014. "A First Look at the Oxford Nanopore MinION Sequencer." *Molecular Ecology Resources* 14, no. 6: 1097–1102. https://doi.org/10.1111/1755-0998.12324.

Mordret, S., R. Piredda, G. Zampicinini, et al. 2023. "Metabarcoding Reveals Marked Seasonality and a Distinctive Winter Assemblage of Dinoflagellates at a Coastal LTER Site in the Gulf of Naples." *Marine Ecology* 44, no. 3: e12758. https://doi.org/10.1111/maec.12758.

Mortelmans, J., J. Goossens, L. Amadei Martínez, K. Deneudt, A. Cattrijsse, and F. Hernandez. 2019. "LifeWatch Observatory Data: Zooplankton Observations in the Belgian Part of the North Sea." *Geoscience Data Journal* 6: 76–84. https://doi.org/10.1002/gdj3.68.

Mortelmans, J., I. Semmouri, M. Perneel, et al. 2024. "Temperature-Induced Copepod Depletion and the Associated Wax of *Bellerochea* in Belgian Coastal Waters: Implications and Shifts in Plankton Dynamics." *Journal of Sea Research* 201: 102523. https://doi.org/10.1016/j.seares.2024.102523.

Mugnai, F., F. Costantini, A. Chenuil, M. Leduc, J. M. Gutierrez Ortega, and E. Meglécz. 2023. "Be Positive: Customized Reference Databases and New, Local Barcodes Balance False Taxonomic Assignments in Metabarcoding Studies." *PeerJ* 11: e14616. https://doi.org/10.7717/peerj.14616.

Neri, F., M. Ubaldi, S. Accoroni, et al. 2025. "Comparative Analysis of Phytoplankton Diversity Using Microscopy and Metabarcoding: Insights From an eLTER Station in the Northern Adriatic Sea." *Hydrobiologia* 852, no. 1: 169–183. https://doi.org/10.1007/s10750-024-05692-2.

Nohe, A., A. Goffin, L. Tyberghein, et al. 2020. "Marked Changes in Diatom and Dinoflagellate Biomass, Composition and Seasonality in the Belgian Part of the North Sea Between the 1970s and 2000s." *Science of the Total Environment* 716: 136316. https://doi.org/10.1016/j.scitotenv.2019.136316.

Paasche, E. 2001. "A Review of the Coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), With Particular Reference to Growth, Coccolith Formation, and Calcification-Photosynthesis Interactions." *Phycologia* 40, no. 6: 503–529. https://doi.org/10.2216/i0031-8884-40-6-503.1.

Pawlowski, J., L. Apothéloz-Perret-Gentil, and F. Altermatt. 2020. "Environmental DNA: What's Behind the Term? Clarifying the Terminology and Recommendations for Its Future Use in Biomonitoring." *Molecular Ecology* 29: 4258–4264. https://doi.org/10.1111/mec.15643.

Perneel, M., R. Lagaisse, J. Mortelmans, S. Maere, and P. I. Hablützel. 2024. "Seasonal Metabolic Dynamics of Microeukaryotic Plankton: A Year-Long Metatranscriptomic Study in a Temperate Sea." *MBio* 15, no. 8: e00383-24. https://doi.org/10.1128/mbio.00383-24.

Pernice, M. C., R. Logares, L. Guillou, and R. Massana. 2013. "General Patterns of Diversity in Major Marine Microeukaryote Lineages." *PLoS One* 8, no. 2: e57170. https://doi.org/10.1371/journal.pone.0057170.

Prokopowich, C. D., T. R. Gregory, and T. J. Crease. 2003. "The Correlation Between rDNA Copy Number and Genome Size in Eukaryotes." *Genome* 46, no. 1: 48–50. https://doi.org/10.1139/g02-103.

R Core Team. 2021. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raitsos, D., S. Lavender, Y. Pradhan, T. Tyrell, P. Reid, and M. Edwards. 2006. "Coccolithophore Bloom Size Variation in Response to the Regional Environment of the Subarctic North Atlantic." *Limnology and Oceanography* 51, no. 5: 2122–2130. https://doi.org/10.4319/lo.2006.51.5.2122.

Ramond, P., R. Siano, S. Schmitt, et al. 2021. "Phytoplankton Taxonomic and Functional Diversity Patterns Across a Coastal Tidal Front." *Scientific Reports* 11, no. 1: 2682. https://doi.org/10.1038/s41598-021-82071-0.

Rodriguez-Ezpeleta, N., O. Morissette, C. W. Bean, et al. 2021. "Trade-Offs Between Reducing Complex Terminology and Producing Accurate Interpretations From Environmental DNA: Comment on "Environmental DNA: What's Behind the Term?" by Pawlowski Et al., (2020)." *Molecular Ecology* 30: 4601–4605. https://doi.org/10.1111/mec.15942.

Roswell, M., J. Dushoff, and R. Winfree. 2021. "A Conceptual Guide to Measuring Species Diversity." *Oikos* 130, no. 3: 321–338. https://doi.org/10.1111/oik.07202.

Roux, P., M. Schapira, K. N. Mertens, et al. 2023. "When Phytoplankton Do Not Bloom: The Case of the Dinoflagellate *Lepidodinium chlorophorum* in Southern Brittany (France) Assessed by Environmental DNA." *Progress in Oceanography* 212: 102999. https://doi.org/10.1016/j.pocean.2023.102999.

Ruppert, K. M., R. J. Kline, and M. S. Rahman. 2019. "Past, Present, and Future Perspectives of Environmental DNA (eDNA) Metabarcoding: A Systematic Review in Methods, Monitoring, and Applications of Global eDNA." *Global Ecology and Conservation* 17: e00547. https://doi.org/10.1016/j.gecco.2019.e00547.

Sandin, M., S. Romac, and F. Not. 2022. "Intra-Genomic rRNA Gene Variability of Nassellaria and Spumellaria (Rhizaria, Radiolaria) Assessed by Sanger, MinION and Illumina Sequencing." *Environmental Microbiology* 24, no. 7: 2979–2993.

Santi, I., O. Beluche, M. Beraud, et al. 2023. "European Marine Omics Biodiversity Observation Network: A Strategic Outline for the Implementation of Omics Approaches in Ocean Observation." *Frontiers in Marine Science* 10. https://doi.org/10.3389/fmars.2023.1118120.

Santi, I., P. Kasapidis, I. Karakassis, and P. Pitta. 2021. "A Comparison of DNA Metabarcoding and Microscopy Methodologies for the Study of Aquatic Microbial Eukaryotes." *Diversity* 13, no. 5: 180. https://doi.org/10.3390/d13050180.

Schloss, P. D. 2024. "Rarefaction Is Currently the Best Approach to Control for Uneven Sequencing Effort in Amplicon Sequence Analyses." *mSphere* 9, no. 2: e00354-23. https://doi.org/10.1128/msphere.00354-23.

Semmouri, I., K. A. C. De Schamphelaere, S. Willemse, M. B. Vandegehuchte, C. R. Janssen, and J. Asselman. 2021. "Metabarcoding Reveals Hidden Species and Improves Identification of Marine Zooplankton Communities in the North Sea." *ICES Journal of Marine Science* 78, no. 9: 3411–3427. https://doi.org/10.1093/icesjms/fsaa256.

Skouroliakou, D.-I., E. Breton, and U. Christaki. 2024. "*Phaeocystis globosa* and Diatom Blooms Promote Distinct Bacterial Communities and Associations in a Coastal Ecosystem." *Environmental Microbiology Reports* 16, no. 4: e13313. https://doi.org/10.1111/1758-2229.13313.

Skouroliakou, D.-I., E. Breton, S. Irion, L. F. Artigas, and U. Christaki. 2022. "Stochastic and Deterministic Processes Regulate Phytoplankton Assemblages in a Temperate Coastal Ecosystem." *Microbiology Spectrum* 10, no. 6: e02427-22. https://doi.org/10.1128/spectrum.02427-22.

Stock, W., C. Rousseau, G. Dierickx, et al. 2025. "Breaking Free From References: A Consensus-Based Approach for Community Profiling With Long Amplicon Nanopore Data." *Briefings in Bioinformatics* 26, no. 1: bbae642. https://doi.org/10.1093/bib/bbae642.

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, no. s1: 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x.

Stoeck, T., S. N. Katzenmeier, H.-W. Breiner, and V. Rubel. 2024. "Nanopore Duplex Sequencing as an Alternative to Illumina MiSeq Sequencing for eDNA-Based Biomonitoring of Coastal Aquaculture Impacts." *Metabarcoding and Metagenomics* 8: e121817. https://doi.org/10.3897/mbmg.8.121817.

Sunagawa, S., L. P. Coelho, S. Chaffron, et al. 2015. "Structure and Function of the Global Ocean Microbiome." *Science* 348, no. 6237: 1261359. https://doi.org/10.1126/science.1261359.

Szoboszlay, M., L. Schramm, D. Pinzauti, J. Scerri, A. Sandionigi, and M. Biazzo. 2023. "Nanopore Is Preferable Over Illumina for 16S Amplicon Sequencing of the Gut Microbiota When Species-Level Taxonomic Classification, Accurate Estimation of Richness, or Focus on Rare Taxa Is Required." *Microorganisms* 11, no. 3. https://doi.org/10.3390/microorganisms11030804.

Taberlet, P., E. Coissac, M. Hajibabaei, and L. H. Rieseberg. 2012. "Environmental DNA." *Molecular Ecology* 21: 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05470.x.

Teeling, H., B. M. Fuchs, C. M. Bennke, et al. 2016. "Recurring Patterns in Bacterioplankton Dynamics During Coastal Spring Algae Blooms." *eLife* 5: e11888. https://doi.org/10.7554/eLife.11888.

Terrats, L., H. Claustre, M. Cornec, A. Mangin, and G. Neukermans. 2020. "Detection of Coccolithophore Blooms With BioGeoChemical-Argo Floats." *Geophysical Research Letters* 47: e2020GL090559. https://doi.org/10.1029/2020GL090559.

van der Loos, L. M., S. D'Hondt, A. Willems, and O. De Clerck. 2021. "Characterizing Algal Microbiomes Using Long-Read Nanopore Sequencing." *Algal Research* 59: 102456. https://www.sciencedirect.com/science/article/pii/S2211926421002757.

Vaulot, D., S. Geisen, F. Mahé, and D. Bass. 2022. "pr2-Primers: An 18S rRNA Primer Database for Protists." *Molecular Ecology Resources* 22, no. 1: 168–179. https://doi.org/10.1111/1755-0998.13465.

Wang, Y., Y. Zhao, A. Bollas, Y. Wang, and K. F. Au. 2021. "Nanopore Sequencing Technology, Bioinformatics and Applications." *Nature Biotechnology* 39, no. 11: 1348–1365. https://doi.org/10.1038/s41587-021-01108-x.

Wickham, H. 2016. *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4.

Wisecaver, J. H., and J. D. Hackett. 2011. "Dinoflagellate Genome Evolution." *Annual Review of Microbiology* 65: 369–387. https://doi.org/10.1146/annurev-micro-090110-102841.

Yeh, Y., J. McNichol, D. M. Needham, E. B. Fichot, L. Berdjeb, and J. A. Fuhrman. 2021. "Comprehensive Single-PCR 16S and 18S rRNA Community Analysis Validated With Mock Communities, and Estimation of Sequencing Bias Against 18S." *Environmental Microbiology* 23, no. 6: 3240–3250. https://doi.org/10.1111/1462-2920.15553.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix S1:** ece373218-sup-0001-supinfo.docx. **Appendix S2:** ece373218-sup-0002-Tables.xlsx.