



Comparison of DNA metabarcoding and microscopy in analysing planktonic protists from the European Arctic

Agata Weydmann-Zwolicka¹ · Anna Maria Dąbrowska² · Monika Mioduchowska^{1,3} · Adrian Zwolicki⁴

Received: 28 July 2023 / Revised: 31 January 2024 / Accepted: 23 March 2024
© The Author(s) 2024

Abstract

In the era of climate change-related restructuring of planktonic protist communities, it is especially important to identify possible shifts in their taxonomic composition. While traditional microscopy-based morphological classification is time-consuming and requires experienced taxonomists, metabarcoding seems to substantially accelerate the determination of taxonomic composition. In this study, based on samples collected in summer 2019 from the West Spitsbergen Current, we analysed planktonic protists using both methods. Metabarcoding, based on high-throughput sequencing of the V4 region of the 18S rRNA gene, resulted in a much higher number of operational taxonomic units (OTUs) and sample diversity than microscopy, although the resolution of taxonomic identification ranged from species to phyla. Most morphology-based identification was performed at the species or genus level, additionally allowing us to include information about dominants and size fractions. The highest proportion of 45% shared taxa by both methods was recorded at the class level. The composition of dominant protists differed between the approaches, with most similarities being observed in Bacillariophyceae, for which two genera, *Thalassiosira* and *Eucampia*, were found to be the most abundant with both methods. For Dinophyceae, the most abundant representatives identified by microscopy were *Gymnodinium* spp., *Prorocentrum minimum* and *Gonyaulax gracilis*, while in the metabarcoding approach, most dinoflagellates were identified to the class level only. Given the different levels of accuracy of taxonomic determinations and possible biases in results connected to the chosen methodology, we advocate using an integrative taxonomic approach for the classification of planktonic protists based on the combination of microscopy and molecular methods.

Keywords Phytoplankton · Integrative taxonomy · Next-generation sequencing · High-throughput sequencing · Spitsbergen · Fram Strait · Atlantification

Communicated by E. Fileman

This article is a contribution to the Topical Collection *Discovering the water column: integrated taxonomic approaches for measuring marine pelagic biodiversity*.

✉ Agata Weydmann-Zwolicka
agata.weydmann@ug.edu.pl

- ¹ Laboratory of Plankton Biology, Department of Marine Biology and Biotechnology, Faculty of Oceanography and Geography, University of Gdańsk, Gdynia, Poland
- ² Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland
- ³ Department of Evolutionary Genetics and Biosystematics, Faculty of Biology, University of Gdańsk, Gdańsk, Poland
- ⁴ Department of Vertebrate Ecology and Zoology, Faculty of Biology, University of Gdańsk, Gdańsk, Poland

Introduction

The present-day Arctic marine ecosystem is dominated by short and relatively simple trophic webs, in which sympagic and planktonic autotrophs constitute the primary group of producers (Bluhm et al. 2015). The main energy source for higher trophic levels thus comes from the two subsequent blooms, the first ice-algal bloom appearing by the end of winter and the second bloom of planktonic protists once the sea ice melts (Leu et al. 2015). However, recent climate change, due to temperature rise, reduction in sea-ice cover and related phenomena, influences both blooms and thus primary production (Lalande et al. 2019). Due to increased amounts of heat and freshwater received by the upper Arctic Ocean, mixing processes are not efficient enough to deliver nutrients from depth to the surface for phytoplankton growth, which results in a documented shift towards

the smallest-size phytoplankton (pico- and nano, up to 20 μm) and other (mostly flagellated) protists (Li et al. 2009). Additionally, the observable proceeding inflow of warmer, more saline Atlantic waters northwards (Walczowski et al. 2012; Polyakov et al. 2017), known as Atlantification, brings boreal expatriates of all trophic levels to this polar environment (e.g. Hegseth and Sundfjord 2008; Weydmann et al. 2014) thus boosting local biodiversity, but simultaneously posing a threat to naturally occurring communities, among others by competing for resources and trophic niches.

The area influenced by the West Spitsbergen Current, which is the main source of Atlantic waters and boreal organisms in the Arctic, is especially vulnerable to the processes related to Atlantification, including the planktonic protist community. The reductions in diatom biomass and production, with simultaneous increases in both nano- and microplanktonic taxa, such as dinoflagellates and ciliates, have been frequently reported from West Spitsbergen fjords and adjacent waters (e.g. Piwosz et al. 2009; Mayzaud et al. 2013), especially in warmer years (Kubiszyn et al. 2014). Additionally, in the open waters of the Fram Strait, located between Greenland and Spitsbergen, a possible climate-driven shift in dominance from mostly microplanktonic diatoms to the haptophyte *Phaeocystis pouchetii* and other flagellated or colony-forming species was observed, along with higher levels of chlorophyll *a* concentrations (Nöthig et al. 2015).

In light of climate change-related restructuring of planktonic protist communities, including phytoplankton, and the following inevitable changes in benthic and zooplankton communities, it is especially important to continuously monitor and identify possible shifts in their qualitative and quantitative compositions. Ideally, such evaluation of an ecosystem's current state should be based on a relatively fast, cheap, and large-scale taxonomic community assessment. However, until recent years, especially in regions that are nearly logistically inaccessible, such as the European Arctic, such monitoring has been temporarily and spatially limited due to the specific nature of traditional light microscopy examinations, including, *inter alia*, time-consuming analyses based on morphometric cell data, which require many years of identification experience. Considering these limitations, recently developed metabarcoding, a genomic method where all organisms are examined at the same time, seems to have the potential to revolutionize and speed up environmental monitoring. Comprehensive genetic-based studies on marine phytoplankton, or planktonic protists, are still not common, and it has been shown that operational taxonomic unit (OTU) diversity is greater than that described with microscopy-based taxonomic methods (e.g. Piredda et al. 2017; Gran-StadniczeŃko et al. 2019; Santi et al. 2021).

Rapid climate change and Atlantification of the Arctic are predicted to cause shifts in planktonic protist communities

(Kubiszyn et al. 2014; Dąbrowska et al. 2020) and have a major impact on primary production (Lalande et al. 2019; Ardyna and Arrigo 2020), also influence Arctic marine and terrestrial ecosystem functions (Findlay et al. 2015; Stempniewicz et al. 2021; Weydmann-Zwolicka et al. 2021). Therefore, it is especially important to improve tools allowing for fast and proper species identification and assessment of different levels of biodiversity that would allow for early detection of possible planktonic biodiversity shifts. Therefore, the aim of this study was to test whether metabarcoding based on high-throughput sequencing of the V4 region of the 18S rRNA gene can be an alternative to traditional light microscopy-based morphological classification of planktonic protists, which is time-consuming and requires specialistic knowledge. The use and possibilities offered by both approaches were tested on samples collected from the Fram Strait along the West Spitsbergen Current, which is the main route of boreal organisms and heat inflow to the Arctic Ocean. Therefore, our study also contributes to the application of novel methods for a better understanding of the protist plankton community structure and dynamics in the era of Atlantification of the Arctic Ocean.

Materials and methods

Study area

As the study area, we selected the Fram Strait, which is mainly influenced by the West Spitsbergen Current (WSC, Fig. 1). The WSC, which is a continuation of the Norwegian Atlantic Current, transports Atlantic waters and associated biota northwards through the Norwegian and Greenland Seas, and thus influences the hydrological properties and planktonic communities of the Arctic Ocean. There are two branches of the WSC eastern and western, which most differ in physical water properties (Walczowski et al. 2012), taxonomic composition of planktonic communities (Kubiszyn et al. 2014; Weydmann et al. 2014; Dąbrowska et al. 2020), and key species population structure (Weydmann et al. 2018; Mańko et al. 2022); their distribution patterns are mostly driven by the bottom topography. The west coast of Spitsbergen, the largest island of Svalbard Archipelago, is additionally influenced by the coastal South Cape Current, which carries cold, less saline Arctic-type water from the northeast Barents Sea to the West Spitsbergen Shelf (Fig. 1; Cottier et al. 2005; Walczowski et al. 2012).

Sampling

Ten sampling stations were located at hydrological transects: H (73°30'N), K (75°N), N (76°30'N), and EB (78°50'N), perpendicular to the main direction of the

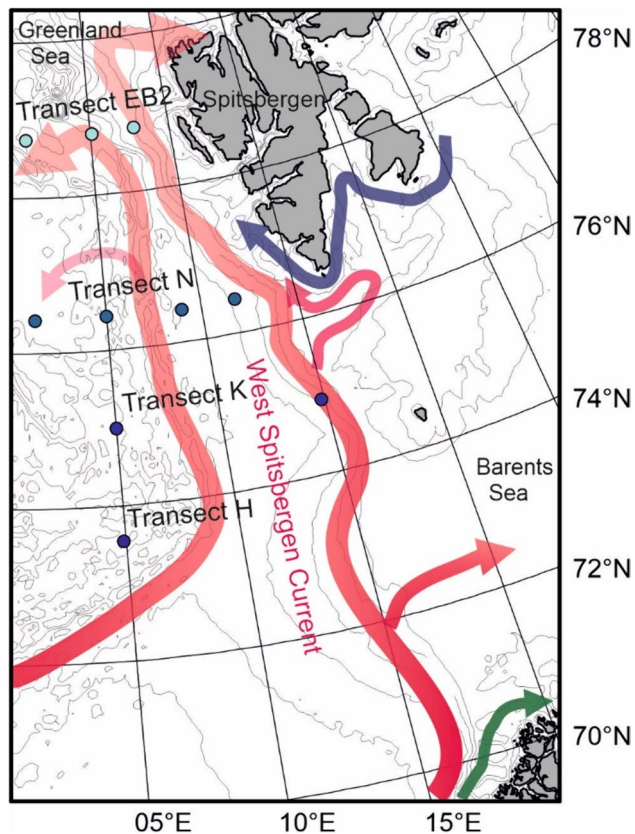


Fig. 1 Location of sampling stations in the transects along the West Spitsbergen Current, with a schematic representation of the oceanic circulation pattern (West Spitsbergen Current—red, South Cape Current—blue, Norwegian Coastal Current—green)

Atlantic Water inflow and spanning both WSC branches (Fig. 1; Walczowski et al. 2012), which are annually sampled by the Institute of Oceanology, Polish Academy of Sciences (IO PAN). Sampling was performed at the end of June and beginning of July 2019 during the routine Arctic cruises of RV *Oceania*, belonging to IO PAN.

Samples for the microscopic and metabarcoding analyses of planktonic protists (pico-, nano-, and microplankton) were collected from a depth close to the surface with the use of a 5 l Niskin bottle. Subsamples of 200 ml for the determination of planktonic protist taxonomic composition using traditional microscopy were immediately fixed with acidic Lugol's solution and, after 24 h, with glutaraldehyde. The fixatives were added to final concentrations of 2 and 1%, respectively, and the samples were kept in the dark until processed in the laboratory. The remaining water samples were filtered through disposable Whatman® Grade 4 Qualitative Filter Papers with 20 µm pore sizes and a diameter of 42.5 mm, which were kept in -80 °C until further processing for DNA extraction.

Microscopy

The samples were qualitatively and quantitatively analysed using protocols described by Utermöhl (Utermöhl 1958) and modified. A 10–50-ml subsample was placed in a counting chamber for 24 h and then the protists were counted under an inverted microscope equipped with phase and interference contrast (Nikon Eclipse TE-300). To count the microplanktonic cells (>20 µm), the entire chamber surface was scanned at ×100 magnification. The nanoplanktonic cells (3–20 µm) were counted at ×400 magnification by moving the field of view along the length of three transverse transects. The taxa were identified to the lowest possible taxonomic level according to the systematics used in the Protist Ribosomal Reference (PR2 v4.13.0, <https://pr2-database.org/>) sequence database.

DNA extraction, PCR amplification, and Illumina MiSeq sequencing

Molecular laboratory procedures were conducted under laminar flow conditions using sterile equipment to avoid sample cross-contamination. DNA extraction from planktonic protists collected by filtration was performed using the commercial DNeasy PowerWater Kit (Qiagen, Germany), with some modifications to the manufacturer's protocol. In the first step, to lyse phytoplankton samples more effectively, incubation was carried out overnight at 37 °C and for 2 h at 50 °C with shaking at 500 rpm in an Eppendorf ThermoMixer©. Subsequent steps were performed according to the manufacturer's protocol. The concentration of purified DNA was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, USA). The extracted genomic DNA was then stored at -20 °C for further analysis.

Polymerase chain reaction (PCR) amplification of the V4 region of the 18S rRNA gene (approximately 470 bp in length) was performed using the following primers (adapters are underlined): forward 5'-CCAGCASYCGGTAATTC C-3' and reverse 5'- ACTTTCGTTCTTGATYRATGA-3' (Piredda et al. 2017; Tragin et al. 2017), with Illumina adapters. All PCRs were performed in a 20 µl volume containing 0.8×JumpStart Taq ReadyMix (1 U of JumpStart Taq DNA polymerase, 4 mM Tris-HCl (pH 8.3), 20 mM KCl, 0.6 mM MgCl₂, 0.08 mM dNTPs; Sigma-Aldrich, Germany), 0.4 µM of forward and reverse primers and approximately 10 ng of DNA. In turn, the PCR conditions consisted of an initial denaturation step of 5 min at 95 °C followed by 40 cycles of 30 s at 95 °C, 45 s at 50 °C, 90 s at 72 °C, and a final elongation step of 10 min at 72 °C. The reactions were performed in a BiometraTProfessional thermocycler. PCR products were run on a 1% agarose gel in TBE (Tris-borate-EDTA) buffer and visualized by SimplySafe (EURx,

Poland) staining under UV light using the GelDoc-It2 310 UVP imaging system.

The PCR products were sent to the Genomed Laboratory (Poland; <http://www.genomed.pl/>) for Illumina MiSeq 18S rRNA sequencing with paired-end (PE) technology. All obtained PCR products passed the final quality control (QC) check using MultiQC tools (Ewels et al. 2016). The company used MiSeq Reporter (MSR) v2.6 (<https://www.illumina.com/systems/sequencing-platforms/miseq/products-services/miseq-reporter.html>) to perform high-throughput sequencing (HTS) of the V4 region of the 18S rRNA gene.

Bioinformatic analysis of raw reads

The sequenced data were analysed using QIIME 1.9.1 (Caporaso et al. 2010) according to a standard bioinformatics pipeline. The following bioinformatic steps were applied: (i) removing adapter sequences using the cutadapt programme (Martin 2011); (ii) performing a read quality analysis and removing low quality sequences (quality < 20) using the cutadapt programme; (iii) joining paired sequences, with a minimum overlap length of 10 nucleotides, using the fastq-join algorithm (Aronesty 2011); (iv) removing chimeric sequences using the usearch algorithm (Edgar 2010); (v) clustering of 18S rRNA amplicon sequences into OTUs using a hybrid method (open-reference OTU), with 97% similarity based on the selected Protist Ribosomal Reference (PR2 v4.13.0, <https://pr2-database.org/>) reference sequence database using the uclust algorithm (Edgar 2010). Subsequently, sequences without matches in the database were clustered using the de novo method; and (vi) after completing the clustering, taxonomy assignment was performed with the use of the same algorithm.

Data analysis

Within many morphologically recognized taxa, especially those identified to genus or order, specimen sizes were also noted, although all size fractions within a respective

taxonomic level were summed as one taxon (genus or order, respectively) for the purpose of data analysis.

To compare planktonic protist diversity between the methods, the Shannon diversity index (H' , \log_e) was calculated for each sample. Afterwards, the pairs of indices obtained for each method within a given sampling station were tested using a paired *t*-test. The calculation of the Shannon diversity index was performed in PRIMER 6.1.15 (Clarke and Gorley 2006), and other analyses and graphical presentation of the results were performed in the R 4.1.0 environment (R Core Team 2021). The following packages were used: 'tidyverse' (Wickham et al. 2019), 'ggpubr' (Kassambara 2020), and 'eulerr' (Larsson 2018) for visualization and 'vegan' (Oksanen et al. 2019) for cumulative curve calculations. All of the above analyses were performed on presence-absence transformed data. To describe the composition of the planktonic protist community and their dominance structure obtained by metabarcoding and microscopy, we presented the relative abundances of the most numerous species/genera grouped at the class level, and the relative abundances of the most numerous taxa, which were summed for all samples within each method.

Results

Using microscopy, 106 taxa were identified in all 10 samples (Fig. 2a), with the taxonomic level spanning from species to classes, and with 75 taxa identified down to a species level. Using DNA metabarcoding and after removing nonmarine organisms, 244 OTUs were distinguished based on (Fig. 2a), with taxonomic precision ranging from the species to phylum levels, including 116 species. This striking difference in the number of taxa identified by both methods is clearly seen in Fig. 2, which shows rarefaction curves plotted for the cumulative number of OTUs obtained from DNA metabarcoding and microscopy.

Consequently, Shannon diversity indices calculated for the samples collected at the same stations, differed

Fig. 2 Comparison of DNA metabarcoding (blue) and microscopy (pink) results. **a** Rarefaction curves plotted for the cumulative number of taxa. **b** Shannon diversity indices calculated for both methods for the samples collected at the same stations

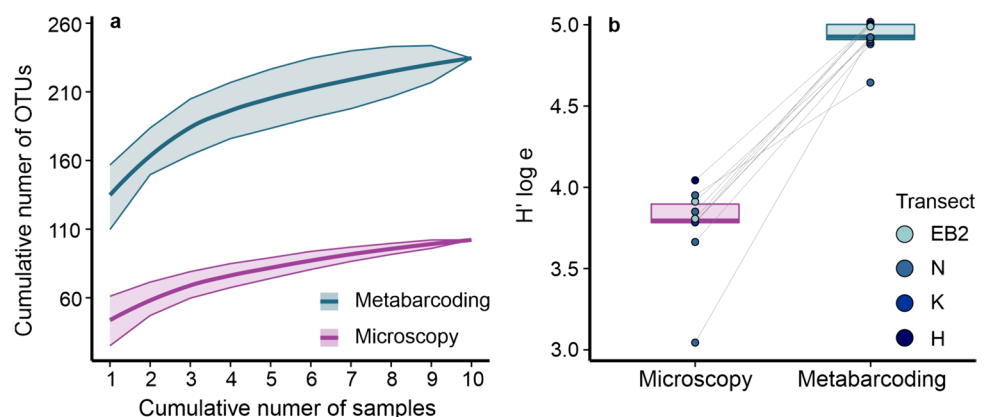


Fig. 3 Venn diagrams indicating the proportions of classes (a), orders (b), families (c) or genera (d) that resulted from DNA metabarcoding (blue) and microscopy (pink). The yellow intersection area indicates the proportion of shared taxa obtained by both methods

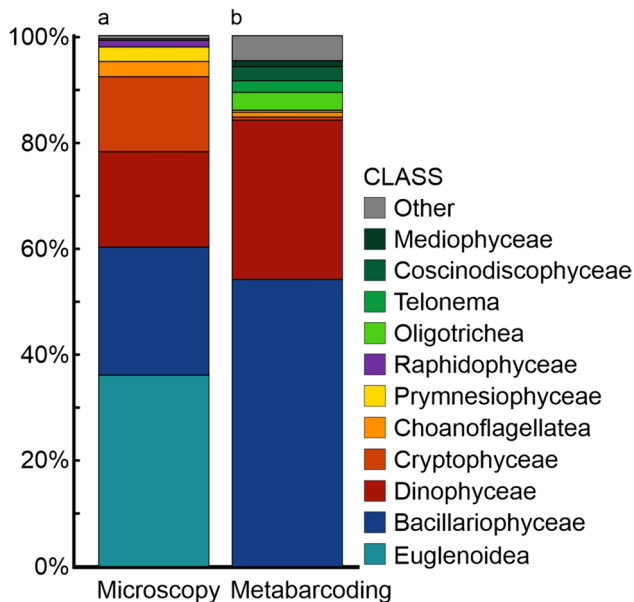
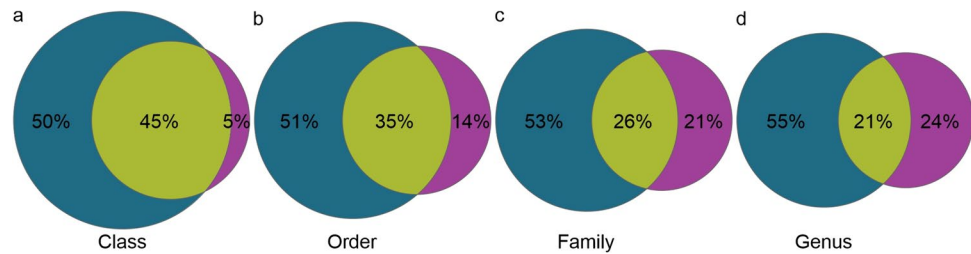


Fig. 4 Relative abundances (%) of the most numerous classes identified by microscopy (a) and metabarcoding (b). The category ‘Other’ includes classes which abundance was lower than 1%

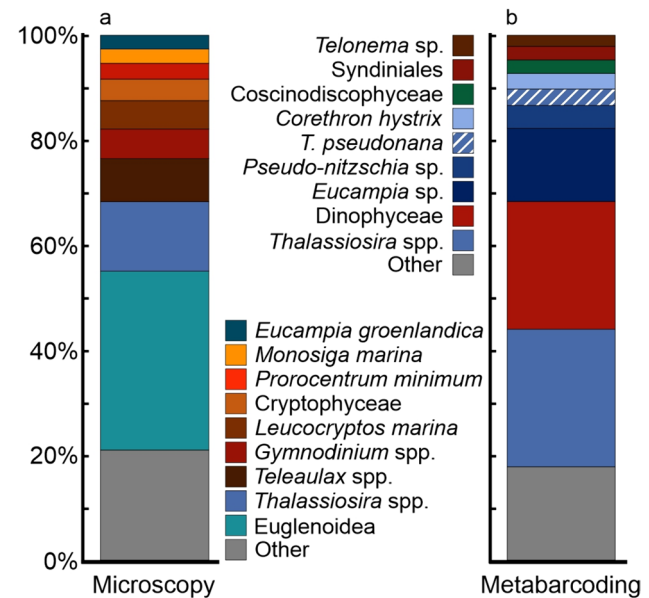


Fig. 5 Relative abundances (%) of the most numerous taxa (summed for all samples) identified by microscopy (a) and metabarcoding (b)

significantly between the methods (paired t -test = 12.17, $p < 0.001$, $n = 10$, with large effect size = 3.85), and in each case, the index was much higher with the metabarcoding approach (Fig. 2b).

The number and proportion of taxa identified, and shared, by both methods changed with taxonomic resolution (Fig. 3). At the genus level, 143 taxa were identified, 108 by metabarcoding and 65 by microscopy, and only 30 genera (21%) were shared by both methods. The proportion of shared taxa increased at the family level to 26% (30 out of 115 families), followed by the order level, at which 84 taxa were identified, with a shared fraction of 35% (29 orders). At the class level, the shared proportion of taxa reached 45% (17 out of 38 classes), with 5% (2 classes) recorded by microscopy only, and 50% (19 classes) identified only by metabarcoding (Fig. 3a).

The composition of the planktonic protist community greatly varied between the metabarcoding and morphology-based approaches (Fig. 4), even at the class level, which had the highest proportion of 45% shared taxa (Fig. 3a). Based

on the number of the lowest taxonomic units such as genera/species obtained by microscopy, Euglenoidea was the most abundant class constituting 36.1% of the planktonic protist community (Fig. 4), followed by Bacillariophyceae (24%), Dinophyceae (18%), and Cryptophyceae (14%). In comparison, Bacillariophyceae, which comprised 54% genera/species, was the most abundant class identified by metabarcoding, followed by Dinophyceae (30%), and Oligotrichea (3%). Out of the most observed classes, Bolidophyceae, Labyrinthulea, Mamiellophyceae, and Phyllopharyngea were detected by metabarcoding only. Notably, a protozoan class Euglenoidea was identified only by microscopy.

In terms of numbers, the composition of dominant protistan taxa differed between the identification methods (Fig. 5). According to microscopy, the most abundant taxon was the protozoan class Euglenoidea (34%), followed by the diatom *Thalassiosira* spp. (13%), a cryptophyte *Teleaulax* spp. (8%), and dinophycean *Gymnodinium* spp. (6%). In the case of metabarcoding the most numerous OTUs belonged to Bacillariophyceae *Thalassiosira* spp. (26%), and *Eucampia* spp. (14%), as well as Dinophyceae as a class (24%).

The composition of dominant protistan genera and species belonging to the two most numerous classes, Bacillariophyceae and Dinophyceae (Fig. 4), identified by both methods differed. Most similarities were observed in the case of Bacillariophyceae, for which, according to microscopy, the most common were *Eucampia groenlandica*, which were present at every station (mean abundance 1.57×10^7 ind. m^{-3}), and *Thalassiosira* spp. (mean abundance 8.0×10^7 ind. m^{-3}); while in the case of metabarcoding, it was *Thalassiosira* spp., *Eucampia* spp., and *Pseudo-nitzschia* spp. as shown in Fig. 5. Based on the morphological approach, the most numerous representatives of Dinophyceae were *Gymnodinium* spp., *Prorocentrum minimum* (recently *Prorocentrum cordatum*) and *Gonyaulax gracilis*, while with metabarcoding most taxa belonging to this class were identified to the class level only (Fig. 5), and the most abundant taxa at the two lowest taxonomic levels were *Biecheleria* sp. and *Pelagodinium beii*.

Discussion

The use of metabarcoding resulted in a much higher number of identified taxa and sample H' diversity than the traditional microscopy-based approach, which was similar to the findings of other studies on marine communities, including planktonic protists (e.g. Piredda et al. 2017; Gran-Stadniczeňko et al. 2019; Santi et al. 2021), seagrass meadows (Coward et al. 2015), periphyton (Groendahl et al. 2017), or zooplankton (Lindeque et al. 2013). This could certainly be due to the taxonomic resolution achieved by the two methods: for metabarcoding it ranged from species to phyla, while most morphology-based identification was performed at the species or genus levels. Additionally, metabarcoding data included a large portion of nonmarine organisms, and some that should not have been present in the Arctic, probably constituting a genetic (DNA) remnant of species living in remote areas, brought with ocean currents to the study area.

The results obtained by metabarcoding are influenced by different factors (Mioduchowska et al. 2022), including the filtration step and sample volume, a molecular marker, or markers, applied, and chosen PCR primers, sequence similarity threshold, and reference database. The higher number of OTUs obtained with metabarcoding may have resulted from different sample volumes, as in microscopy only 200 ml was analysed, allowing fewer species to be observed, compared to the molecular method, for which almost 5 l of sea water were filtered. However, the 20 μ m pore size filter used might have also influenced the metabarcoding results, causing bias towards the microplanktonic fraction, while the nanoplanktonic cells could also be identified with microscopy.

As a molecular marker, we chose the hypervariable region V4 of the 18S rRNA gene, which is widely used as a metabarcoding marker, and was proven to give good results in terms of taxonomic resolution for different marine algal communities, including phytoplankton (Piredda et al. 2017; Tragin et al. 2017; Gran-Stadniczeňko et al. 2019; MacNeil et al. 2022) and periphyton (Groendahl et al. 2017). Additionally, the use of the V4 region for studying environmental microbial diversity has been shown to be superior to other 18S rRNA hypervariable regions, as the entire length (ca. 400 nt) is easily obtainable with the current paired-end high-throughput sequencing technologies (Mahé et al. 2015). However, different taxa are not equally amplified with some PCR primers, e.g. the V4 18S rRNA gene is not efficient in amplifying Euglenophyta (Gran-Stadniczeňko et al. 2019), while some taxa may have high copy numbers of a target gene, such as putatively high rRNA gene copy number in dinoflagellates (Prokopowich et al. 2003). Both of the above observations might have also affected our results, especially the lack of members of the protozoan class Euglenoidea obtained by metabarcoding, with their simultaneous high abundance identified by microscopy and a high number of OTUs belonging to Dinophyceae. Due to possible bias in the metabarcoding results obtained from one marker, some studies recommend a multimarker approach that can reveal a high diversity of marine eukaryotes (e.g. Coward et al. 2015; Leray and Knowlton 2016; MacNeil et al. 2022).

Additionally, the use of a sequence similarity threshold and its level may result in obtaining a lower resolution of taxonomic data. In the case of the V4 region of the 18S rRNA gene, sequences obtained by metabarcoding of planktonic protists had been clustered in OTUs at similarities from 95% (Piredda et al. 2017) to 98% (Gran-Stadniczeňko et al. 2019). We decided to apply operational taxonomic units at a 97% sequence similarity threshold, as it has been successfully used in previous eukaryotic plankton studies (e.g. Liu et al. 2017, 2019; Xue et al. 2018; Mo et al. 2021). The selection of a reference database can also largely affect metabarcoding results. Out of various databases allowing for taxonomic assignments for small ribosomal subunit RNA sequences, including Greengenes (DeSantis et al. 2006), Ribosomal Database Project (RDP-II, Cole et al. 2005), SILVA (Pruesse et al. 2007), and Protist Ribosomal Reference (PR2, Guillou et al. 2013), we decided to choose the latter. The chosen PR2 database provides unique access to eukaryotic small ribosomal subunit RNA, as well as DNA sequences, with curated taxonomy. Moreover, the PR2 database consists mainly of protistan sequences, which were in the area of interest, although sequences originating from eukaryotic organelles, land plants, macrosporidic fungi, and metazoans are also included (e.g. Liu et al. 2019; Mo et al. 2021). However, such discrepancies leave some doubts about possible misidentified taxa in the reference sequence

databases and, to some extent, question the reliability of such data.

Generally, the planktonic protist community structure dominated by two classes, Bacillariophyceae and Dinophyceae, obtained in our study by both methods is in line with previous research conducted during summers in the area influenced by the West Spitsbergen Current (Piwoż et al. 2009; Kubiszyn et al. 2014; Dąbrowska et al. 2020). However, the lower the taxonomic level of taxon identification, the higher the inconsistency and more differences between the methods could be observed. Interestingly, the exception to the above observation is the identification of dominant diatoms, which gave similar results for both methods, with the representatives of two genera, *Thalassiosira* and *Eucampia* (although identified to a genus or species level, or both, by either method), being the most abundant, which may indicate a recent spring bloom or melting of these taxa from sea ice transported from the north and/or the Barents Sea. This, however, changes in regard to Dinophyceae; the use of morphology-based microscopy allowed us to identify the most abundant dinoflagellate species or genera, such as *Gymnodinium* spp., *Prorocentrum minimum*, and *Gonyaulax gracilis*, while in the case of metabarcoding, most Dinophyceae had been identified to the class level only. Interestingly, similar observations were made by Santi et al. (2021), who compared DNA metabarcoding, based on the V9 of the 18S rRNA gene, and microscopy methodologies for the study of the marine plankton community from the Eastern Mediterranean and found that the two methodologies did not vary significantly for Bacillariophyta, although they did for Dinoflagellata and Ciliophora.

Another clear difference in the composition of communities determined by both methods was the lack of some important taxa identified by either method. We observed a clear dominance of mixotrophic Euglenoidea (most likely *Eutreptia* or *Eutreptiella*) in microscopic examinations, but it was absent in the metabarcoding results. First, we cannot exclude that, due to the shape of these cells and their plasticity, they were lost at the filtration stage prior to DNA extraction. Second, although abundant, especially in freshwater ecosystems, it is one of the least researched groups of microorganisms and is known to be underrepresented in environmental sequencing surveys (Łukomska-Kowalczyk et al. 2016). In the case of metabarcoding of the V4 18S rRNA gene, the lack of Euglenoidea can be explained by the poor amplification of this marker (Gran-Stadniczeńko et al. 2019). On the other hand, morphology-based microscopy failed to identify the representatives of picoplanktonic Mamiellophyceae and Bolidophyceae, which mostly belong to small-sized species (Vaulot et al. 2008) and thus are difficult to detect with microscopy. It is worth noting that the failure of one method in detecting taxa that are dominant in the other method, may bias the results of H' diversity. This is

evident in Fig. 2 where the N-transect sample had lower H' diversity than the remaining ones according to microscopy, but an average diversity based on the results of metabarcoding. Low H' in microscopy was probably caused by the lack of Euglenoidea in this given sample, which was generally dominant in the morphology-based results and absent in metabarcoding.

Taking into account our results and previous studies on marine planktonic protistan communities that compared metabarcoding and morphology-based microscopy (e.g. Piredda et al. 2017; Gran-Stadniczeńko et al. 2019; Santi et al. 2021), among the main disadvantages of metabarcoding are possible biases connected to the chosen methodology, as discussed above, and the lack of available molecular references for some microeukaryotic taxa that are difficult to identify through microscopy or cultivation. The advantages of metabarcoding include less time needed to analyse the set of samples, no need for long-term training in laboratory work, and a larger amount of water possible to analyse as a single sample, which allows the sampling of more specimens and possibly more taxa (Piredda et al. 2017; Gran-Stadniczeńko et al. 2019). Metabarcoding may also provide information about cryptic diversity and species or developmental stages that are impossible to identify based on morphological features; at the same time, these are among the main disadvantages of microscopy. Morphology-based taxonomy is also time-consuming and requires training, and the results of analysis are highly correlated with taxonomic expert experience. On the other hand, a well-trained and experienced taxonomist can not only qualitatively and quantitatively analyse samples but also provide information about the sizes of planktonic prokaryotes, identified mode of nutrition, cell condition, abundance, and biomass (Santi et al. 2021).

In summary, the low resolution of metabarcoding-based taxonomic identification of planktonic protists belonging to dominant taxa, such as Dinophyceae, and the absence of whole groups, such as Euglenophyceae that were not detected by metabarcoding, while Bolidophyceae, Labyrinthulea, and Mamiellophyceae were not identified by microscopy, is an important argument for the combined use of both methods. Such bias in the dominant taxa influences the conclusions of ecological studies and precludes any comparison of protistan community structure between areas, seasons, or years, which would hinder the improvement of the general understanding of changes in global biodiversity patterns and possible ecosystem responses to climate change or Atlantification. Therefore, we postulate that the combination of methods based on modern molecular biology techniques, such as metabarcoding, and traditional morphology-based taxonomy, could be the answer to the above-described limitations of both approaches. A similar approach has been postulated, among others, from other studies on aquatic

microbial eukaryotes (Gran-Stadniczeńko et al. 2019; Santi et al. 2021), seagrass communities (Coward et al. 2015), and microalgal periphyton communities (Groendahl et al. 2017).

Given the consensus on the need to combine molecular biology techniques with morphology-based taxonomic approaches, the question remains as to how to complement these two methods in future studies. For example, metabarcoding could be widely used for selecting regions, in which the processes connected to climate change, intensified inflow of Atlantic water masses or exposure to potential anthropogenic pressure are especially significant. If any of the above conditions are identified, samples from this area should be analysed with the use of microscopy. In this way, a combined approach based on integrative taxonomy that would allow for a relatively fast and cheap qualitative assessment of planktonic protist biodiversity with the use of high-throughput sequencing, coupled with morphology-based taxonomy for samples from selected areas or seasons, could be implemented in regions as remote as the European Arctic.

Conclusions

- The use of metabarcoding of the V4 region of the 18S rRNA gene resulted in a much higher number of OTUs and diversity than morphology-based microscopy. The lower the taxonomic level of taxon identification, the higher the inconsistency and more differences between metabarcoding and microscopy could be observed.
- The composition of dominant protistan genera and species belonging to the two most numerous classes identified by both methods, Dinophyceae and Bacillariophyceae, differed between the classes, with most similarities noticed in the case of Bacillariophyceae.
- According to both methods, the representatives of two diatom genera, *Thalassiosira* and *Eucampia*, were the most abundant in the study area. In the case of dinoflagellates; the use of microscopy allowed us to indicate the most abundant species, or genera, *Gymnodinium* spp., *Prorocentrum minimum*, and *Gonyaulax gracilis*, while in the metabarcoding approach, most Dinophyceae were identified to the class level only.
- The largest difference between the methods, and one of the most important arguments for needing both, was in the groups that were not at all identified by one of them, e.g. Bolidophyceae, Labyrinthulea, and Mamiellophyceae were not identified by microscopy, while Euglenophyceae was not identified by metabarcoding.
- Given the different speeds of obtaining results, the different levels of accuracy of taxonomic determinations, possible biases in results connected to the chosen methodology, and various types of information provided by these two approaches, which is especially important in the era

of rapid climate change and Atlantification of the Arctic, we advocate using an integrative taxonomic approach for the classification of planktonic protists based on the combination of morphological features and molecular methods.

Acknowledgements We would like to thank Dr hab. Anna Iglukowska from the Department of Evolutionary Genetics and Biosystematics, University of Gdańsk, for her support in collecting samples. We would also like to thank the Reviewers for their thorough feedback on the manuscript, which greatly helped us to improve it.

Funding This study was financed by the project HIDEA – Hidden diversity of plankton in the European Arctic, no. 2017/27/B/NZ8/01056 from the National Science Centre, Poland, granted to Agata Weydmann-Zwolicka.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval No animal testing was performed during this study.

Sampling and field studies All necessary permits for sampling and observational field studies have been obtained by the authors from the competent authorities and are mentioned in the ‘Acknowledgements,’ if applicable.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contribution statement AWZ took the lead in writing the manuscript, developed the conception of the research, contributed to data interpretation, and acquired funds. AMD participated in the sampling, analysed the samples using morphology-based microscopy, and contributed to microscopy-metabarcoding database processing and data interpretation. MM analysed the samples using metabarcoding and wrote respective parts of the manuscript. AZ analysed the data, wrote parts of the manuscript, and mostly contributed to data interpretation. All authors edited and reviewed the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ardyna M, Arrigo KR (2020) Phytoplankton dynamics in a changing Arctic Ocean. *Nat Clim Chang* 10:892–903. <https://doi.org/10.1038/s41558-020-0905-y>

- Aronesty E (2011) ea-utils: “Command-line tools for processing biological sequencing data”. <https://github.com/ExpressionAnalysis/ea-utils>
- Bluhm BA, Gebruk AV, Gradinger R et al (2015) Arctic marine biodiversity: an update of species richness and examples of biodiversity change. *Oceanogr* 24:232–248. <https://doi.org/10.5670/oceanogr.2011.75>
- Callahan BJ, McMurdie PJ, Rosen MJ et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>
- Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336. <https://doi.org/10.1038/nmeth.f.303>
- Clarke K, Gorley RN (2006) Primer v6: User Manual/Tutorial. PRIMER-E Ltd, Plymouth, UK
- Cole JR, Chai B, Farris RJ et al (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* 33:294–296. <https://doi.org/10.1093/nar/gki038>
- Cottier FR, Tverberg V, Inall ME et al (2005) Water mass modification in an Arctic fjord through cross-shelf exchange: the seasonal hydrography of Kongsfjord, Svalbard. *J Geophysical Res* 110. <https://doi.org/10.1029/2004JC002757>
- Cowart DA, Pinheiro M, Mouchel O et al (2015) Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE* 10(2):e0117562. <https://doi.org/10.1371/journal.pone.0117562>
- Dąbrowska AM, Wiktor JM Jr, Merchel M, Wiktor JM (2020) Planktonic protists of the Eastern Nordic Seas and the Fram Strait: spatial changes related to hydrography during early summer. *Front Mar Sci* 7:557. <https://doi.org/10.3389/fmars.2020.00557>
- DeSantis TZ, Hugenholtz P, Larsen N et al (2006) Greengenes, a chimera-checked 16s rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072. <https://doi.org/10.1128/AEM.03006-05>
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinform* 26(19):2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinform* 32:3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Findlay HS, Gibson G, Kędra M et al (2015) Responses in Arctic marine carbon cycle processes: conceptual scenarios and implications for ecosystem function. *Polar Res* 34:24252. <https://doi.org/10.3402/polar.v34.24252>
- Gran-Stadniczeńko S, Egge E, Hostyeva V et al (2019) Protist diversity and seasonal dynamics in Skagerrak plankton communities as revealed by metabarcoding and microscopy. *J Eukaryot Microbiol* 66:494–513. <https://doi.org/10.1111/jeu.12700>
- Groendahl S, Kahlert M, Fink P (2017) The best of both worlds: a combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. *PLoS ONE* 12(2):e0172808. <https://doi.org/10.1371/journal.pone.0172808>
- Guillou L, Bachar D, Audic S et al (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 41:D597–D604. <https://doi.org/10.1093/nar/gks1160>
- Hegseth EN, Sundfjord A (2008) Intrusion and blooming of Atlantic phytoplankton species in the high Arctic. *J Mar Syst* 74:108–119. <https://doi.org/10.1016/j.jmarsys.2007.11.011>
- Kassambara A (2020) ggpubr: “ggplot2” based publication ready plots. R package version 0.4.0, 438
- Kubiszyn AM, Piewosz K, Wiktor JM Jr, Wiktor JM (2014) The effect of inter-annual Atlantic water inflow variability on the planktonic protist community structure in the West Spitsbergen waters during the summer. *J Plankton Res* 36(5):1190–1203. <https://doi.org/10.1093/plankt/fbu044>
- Lalande C, Nöthig EM, Fortier L (2019) Algal export in the Arctic Ocean in times of global warming. *Geophys Res Lett* 46:5959–5967. <https://doi.org/10.1029/2019GL083167>
- Larsson J (2018) eulerr: area-proportional Euler diagrams with ellipses
- Leray M, Knowlton N (2016) Censusing marine eukaryotic diversity in the twenty-first century. *Phil Trans R Soc B* 371:20150331. <https://doi.org/10.1098/rstb.2015.0331>
- Leu E, Mundy CJ, Assmy P et al (2015) Arctic spring awakening – steering principles behind the phenology of vernal ice algal blooms. *Prog Oceanogr* 139:151–170. <https://doi.org/10.1016/j.pocean.2015.07.012>
- Li WKW, McLaughlin FA, Lovejoy C et al (2009) Smallest algae thrive as the Arctic Ocean freshens. *Science* 326:539. <https://doi.org/10.1126/science.1179798>
- Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next generation sequencing reveals the hidden diversity of zooplankton assemblages. *PLoS ONE* 8(11):e81327. <https://doi.org/10.1371/journal.pone.0081327>
- Liu L, Liu M, Wilkinson DM, Chen HH, Yu XQ, Yang J (2017) DNA metabarcoding reveals that 200- μ m-size-fractionated filtering is unable to discriminate between planktonic microbial and large eukaryotes. *Mol Ecol Resour* 17:991–1002. <https://doi.org/10.1111/1755-0998.12652>
- Liu L, Chen H, Liu M et al (2019) Response of the eukaryotic plankton community to the cyanobacterial biomass cycle over 6 years in two subtropical reservoirs. *The ISME Journal* 13:2196–2208. <https://doi.org/10.1038/s41396-019-0417-9>
- Łukomska-Kowalczyk M, Karnkowska A, Krupska M, Milanowski R, Zakryś B (2016) DNA barcoding in autotrophic euglenids: evaluation of COI and 18s rDNA. *J Phycol* 52:951–960. <https://doi.org/10.1111/jpy.12439>
- MacNeil L, Desai DK, Costa M et al (2022) Combining multi-marker metabarcoding and digital holography to describe eukaryotic plankton across the Newfoundland Shelf. *Sci Rep* 12:13078. <https://doi.org/10.1038/s41598-022-17313-9>
- Mahé F, Mayor J, Bunge J et al (2015) Comparing high-throughput platforms for sequencing the V4 region of SSU- rDNA in environmental microbial eukaryotic diversity surveys. *J Eukaryot Microbiol* 62:338–345. <https://doi.org/10.1111/jeu.12187>
- Mańko MK, Merchel M, Kwaśniewski S, Weydmann-Zwolicka A (2022) Atlantification alters the reproduction of jellyfish *Aglantha digitale* in the European Arctic. *Limnol Oceanogr* 67:1836–1849. <https://doi.org/10.1002/lno.12170>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>
- Mayzaud P, Boutoute M, Noyon M et al (2013) Lipid and fatty acids in naturally occurring particulate matter during spring and summer in a high arctic fjord (Kongsfjorden, Svalbard). *Mar Biol* 160:383–398. <https://doi.org/10.1007/s00227-012-2095-2>
- Mioduchowska M, Iglíkowska A, Jastrzębski JP, Kaczorowska AK, Kotlarska E, Trzebnay A, Weydmann-Zwolicka A (2022) Challenges of comparing marine microbiome community composition data provided by different commercial laboratories and classification databases. *Water* 14:3855. <https://doi.org/10.3390/w14233855>
- Mo Y, Peng F, Gao X et al (2021) Low shifts in salinity determined assembly processes and network stability of microeukaryotic plankton communities in a subtropical urban reservoir. *Microbiome* 9:128. <https://doi.org/10.1186/s40168-021-01079-w>
- Nöthig EM, Bracher A, Engel A et al (2015) Summertime plankton ecology in Fram Strait – a compilation of long-and short-term

- observations. *Polar Res* 34:23349. <https://doi.org/10.3402/polar.v34.23349>
- Oksanen J, Blanchet FG, Friendly M et al (2019) Package ‘vegan’. Community ecology package, version 2(9)
- Piredda R, Tomasino MP, D’Erchia AM et al (2017) Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean long term ecological research site. *FEMS Microbiol Ecol* 93:fiw200. <https://doi.org/10.1093/femsec/fiw200>
- Piwosz K, Walkusz W, Hapter R et al (2009) Comparison of productivity and phytoplankton in a warm (Kongsfjorden) and a cold (Hornsund) Spitsbergen fjord in mid-summer 2002. *Polar Biol* 32:549–559. <https://doi.org/10.1007/s00300-008-0549-2>
- Polyakov IV, Pnyushkov AV, Alkire MB et al (2017) Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. *Science* 356:285–291. <https://doi.org/10.1126/science.aai8204>
- Prokopowich CD, Gregory TR, Crease TJ (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46:48–50. <https://doi.org/10.1139/g02-103>
- Pruesse E, Quast C, Knittel K et al (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35:7188–7196. <https://doi.org/10.1093/nar/gkm864>
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 13 Oct 2023
- Santi I, Kasapidis P, Karakassis I, Pitta P (2021) A comparison of DNA metabarcoding and microscopy methodologies for the study of aquatic microbial eukaryotes. *Diversity* 13:180. <https://doi.org/10.3390/d13050180>
- Stempniewicz L, Weydmann-Zwolicka A, Strzelewicz A, Goc M, Głuchowska M, Kidawa D, Walczowski W, Węśławski JM, Zwolicki A (2021) Advection of Atlantic water masses influences seabird community foraging in a high-Arctic fjord. *Progr Oceanogr* 193:102549. <https://doi.org/10.1016/j.pocean.2021.102549>
- Tragin M, Zingone A, Vaultot D (2017) Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environ Microbiol* 20:506–520. <https://doi.org/10.1111/1462-2920.13952>
- Utermöhl H (1958) Zur vervollkommnung der quantitativen phytoplankton-methodik. *Mitt Int Ver Theor Angew Limnol* 9:1–38
- Vaultot D, Eikrem W, Viprey M, Moreau H (2008) The diversity of small eukaryotic phytoplankton (< or =3 microm) in marine ecosystems. *FEMS Microbiol Rev* 32:795–820. <https://doi.org/10.1111/j.1574-6976.2008.00121.x>
- Walczowski W, Piechura J, Goszczko I, Wieczorek P (2012) Changes of the Atlantic Water properties: an important factor of the European Arctic marine climate. *ICES J Mar Sci* 69:864–869. <https://doi.org/10.1093/icesjms/fss068>
- Weydmann A, Carstensen J, Goszczko I, Dmoch K, Olszewska A, Kwasniewski S (2014) Shift towards the dominance of boreal species in the Arctic: inter-annual and spatial zooplankton variability in the West Spitsbergen Current. *Mar Ecol Prog Ser* 501:41–52. <https://doi.org/10.3354/meps10694>
- Weydmann A, Walczowski W, Carstensen J, Kwasniewski S (2018) Warming of Subarctic waters accelerates development of a key marine zooplankton *Calanus finmarchicus*. *Global Change Biol* 24:172–183. <https://doi.org/10.1111/gcb.13864>
- Weydmann-Zwolicka A, Prątnicka P, Łacka M, Majaneva S, Cottier F, Berge J (2021) Zooplankton and sediment fluxes in two contrasting fjords reveal Atlantification of the Arctic. *Sci Total Environ* 773:145599. <https://doi.org/10.1016/j.scitotenv.2021.145599>
- Wickham H, Averick M, Bryan J et al (2019) Welcome to the Tidyverse. *J Open Source Softw* 4(43):1686
- Xue YY, Chen HH, Yang JR, Liu M, Huang BQ, Yang J (2018) Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. *ISME J* 12:2263–2277. <https://doi.org/10.1038/s41396-018-0159-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.