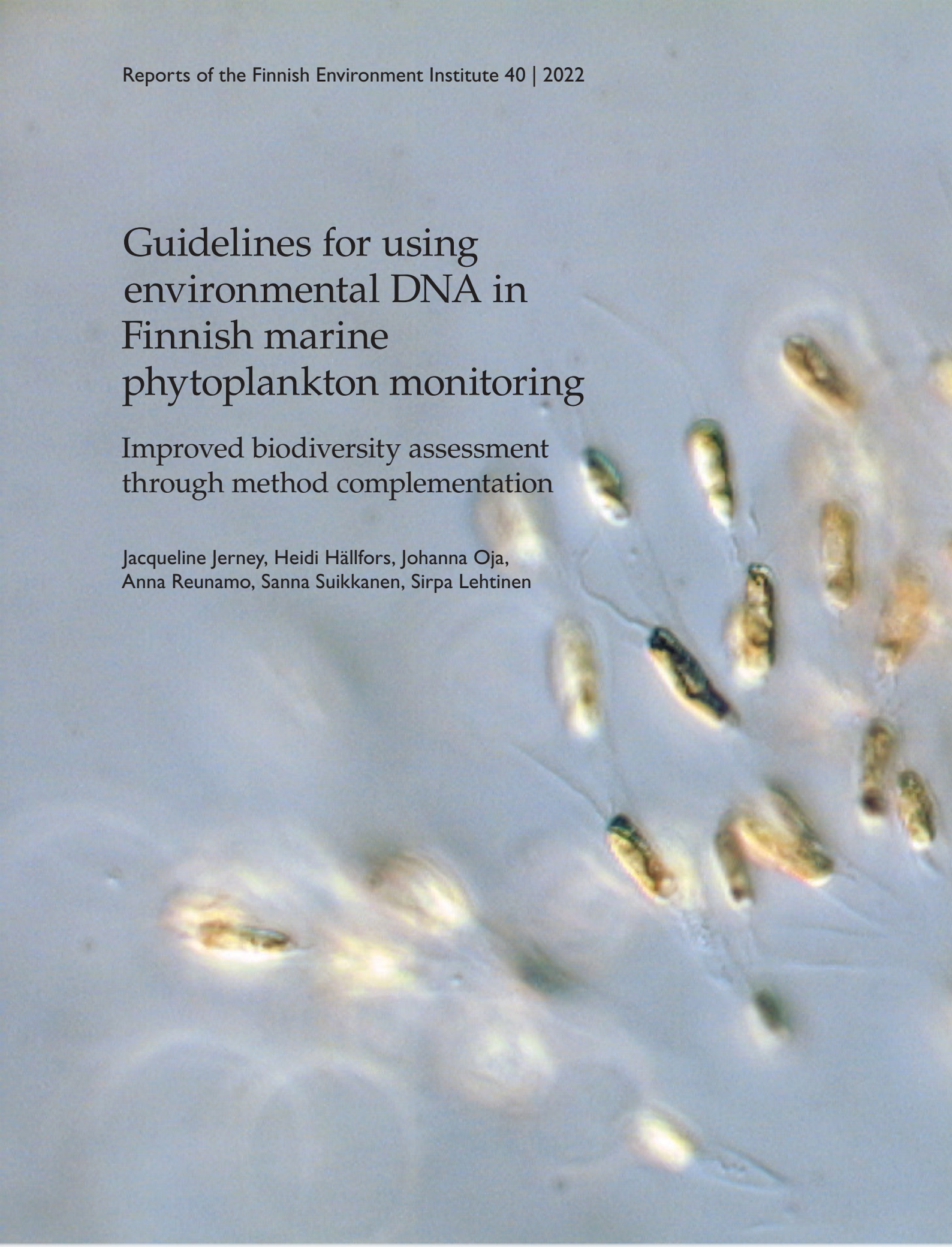
Reports of the Finnish Environment Institute 40 | 2022

Guidelines for using environmental DNA in Finnish marine phytoplankton monitoring

Improved biodiversity assessment through method complementation

Jacqueline Jerney, Heidi Hällfors, Johanna Oja,
Anna Reunamo, Sanna Suikkanen, Sirpa Lehtinen



Finnish Environment Institute

Reports of the Finnish Environment Institute 40 / 2022

Guidelines for using environmental DNA in Finnish marine phytoplankton monitoring

Improved biodiversity assessment through method complementation

Jacqueline Jerney, Heidi Hällfors, Johanna Oja, Anna Reunamo, Sanna Suikkanen, Sirpa Lehtinen

Reports of the Finnish Environment Institute 40 | 2022
Finnish Environment Institute
Marine research centre

Authors: Jacqueline Jerney, Heidi Hällfors, Johanna Oja,
Anna Reunamo, Sanna Suikkanen, Sirpa Lehtinen

Subject Editor: Riitta Autio

Financier: Finnish Ministry of Environment

Publisher and financier of the publication: Finnish Environment Institute (SYKE)
Latokartanonkaari 11, 00790 Helsinki, Finland, Phone +358 295 251 000, syke.fi

Layout: Jacqueline Jerney and Pirkko Väänänen
Cover photo: Seija Hällfors

The publication is available on the internet (pdf): syke.fi/publications | helda.helsinki.fi/syke

ISBN 978-952-11-5524-6 (PDF)
ISSN 1796-1726 (online)

Year of issue: 2022

Abstract

Guidelines for using environmental DNA in Finnish marine phytoplankton monitoring

This document is a methodological guide for using a genomic ecosystem survey technique (eDNA metabarcoding) to supplement conventional phytoplankton monitoring of the Finnish marine monitoring program. The guidelines describe the detection of eukaryotic and prokaryotic phytoplankton with 18S and 16S rDNA gene primers, using high-throughput sequencing. The document includes information on sampling, sample processing, molecular biological work, quality control, and bioinformatics so that the method can be applied in addition to standardized light microscopy. The guidelines are based on a first pilot project testing the integration of eDNA metabarcoding in Finnish marine phytoplankton monitoring and will be developed further, according to evolving genetic methods and international guidelines and standards. Suggestions on steps towards introducing eDNA methodology in phytoplankton monitoring are included in the guidelines. Using eDNA metabarcoding to complement standardized light microscopy advances conventional monitoring and research of phytoplankton communities to assess biodiversity and the status of the marine environment.

Keywords: metabarcoding, amplicon sequencing, biodiversity, method comparison, phytoplankton, monitoring, Baltic Sea

Tiivistelmä

Menetelmäohje DNA-viivakoodaustekniikan käyttöön meren kasviplanktonseurannassa

Tässä julkaisussa kuvataan menetelmäohje DNA-viivakoodeihin perustuvan eDNA (ympäristö-DNA) -metaviivakoodaustekniikan käytöstä Suomen merenhoidon seurantaohjelmaan kuuluvan kasviplankton-seurannan tukena. Aitotumaisen kasviplanktonin (mikrolevät) tunnistamiseen soveltuvan 18S rDNA -menetelmän lisäksi ohjeessa kuvataan esitumaisten syanobakteerien (sinilevät) analysointi 16S rDNA -menetelmällä käyttäen korkean käsittelytehon sekvensointia (high-throughput sequencing). Ohje sisältää yksityiskohtaiset tiedot eDNA-näytteiden keräämisestä, näytteiden käsittelystä, molekyylibiologisista työvaiheista, bioinformatiikka-analyyseistä ja laadunvarmennuksesta. Ohje perustuu ympäristö-ministeriön rahoittaman hankkeen tuloksiin. Ohjeistusta tullaan kehittämään jatkossa sitä mukaa, kun uusia tutkimustuloksia, kansainvälisiä ohjeita ja standardeja julkaistaan. Ohjeistus sisältää ehdotuksen etenemisvaiheista, jotka kannattaa huomioida, kun eDNA-menetelmää ryhdytään ottamaan mukaan kasviplanktonseurantaan. eDNA-menetelmää voidaan käyttää valomikroskopointimenetelmän ohella tarkentamaan ja tehostamaan kasviplanktonyhteisöjen biodiversiteetin seurantaa ja tutkimusta meriympäristön tilan arvioita varten.

Asiasanat: metaviivakoodaus, amplikonisekvenssi, biologinen monimuotoisuus, menetelmävertailu, kasviplankton, seuranta, Itämeri

Sammandrag

Metodanvisningar för användandet av eDNA-metabarcoding i monitoreringen av växtplankton i de finska havsområdena

I detta dokument beskrivs metodanvisningarna för användandet av en så kallad eDNA-metabarcodingteknik, för att komplettera den i Finlands havsförvaltningsplan ingående ordinarie växtplanktonmonitoreringen. Anvisningarna beskriver skönjandet av eukaryotiskt och prokaryotiskt växtplankton med 18S rDNA och 16S rDNA -metoderna och storskalig sekvensering (high-throughput sequencing). Anvisningarna innehåller detaljerade instruktioner för samlandet av eDNA-proven, provbehandlingen, de molekylärbiologiska arbetsskedena, bioinformatikanalyserna och kvalitetssäkringen. Dessa anvisningar baserar sig på resultaten av ett projekt finansierat av miljöministeriet i Finland. Anvisningarna kommer att utvecklas vidare i fortsättningen, allt efter som nya forskningsresultat, internationella anvisningar och standarder publiceras. Anvisningarna innehåller förslag på steg för att införa eDNA-metodiken vid växtplanktonmonitoreringen. Utnyttjandet av eDNA-metabarcodingtekniken för att komplettera den ordinarie växtplanktonmonitoreringen gör det möjligt att precisera och effektivera monitoreringen och forskandet av växtplanktonsamhällets biodiversitet och befrämjar således värderingen av biodiversiteten och havsmiljöns tillstånd.

Nyckelord: eDNA-metabarcoding, amplikonsekvensering, biodiversitet, metodjämförelse, växtplankton, monitorering, Östersjön

Preface

As primary producer, phytoplankton plays a key role in marine ecosystems, and the effects of eutrophication and climate change are often detectable earlier in the phytoplankton community than at higher trophic levels (Hällfors et al., 2013, Suikkanen et al., 2013, Lehtinen et al., 2016, Kuosa et al., 2017). Anthropogenic pressures on ecosystems, like eutrophication and climate change, are globally increasing, which leads to a rapid loss of biodiversity and may have far-reaching consequences for ecosystem functioning (IPBES, 2019). The functioning of ecosystems may be affected because biodiversity has a significant impact on the production, stability, and nutrient use efficiency of primary producer communities (Cardinale et al., 2012). Thus, monitoring the biodiversity of primary producers in marine habitats is of utmost importance in assessing their ecological status, as outlined by the European Union (EU) Water Framework Directive (European Union, 2000), the Marine Strategy Framework Directive (European Union, 2008) and the Biodiversity Strategy (European Union, 2020). In the Finnish marine monitoring program (Rantajärvi et al., 2020) and in the marine monitoring program of the Helsinki Commission (HELCOM), the standardized light microscopy method (CEN, 2015, HELCOM, 2021) is used to analyse phytoplankton samples, making results comparable both within the Baltic Sea and globally.

Light microscopy has been used for phytoplankton observations since the 18th century, the applied techniques have been improved over time and well-established standards are in use (CEN, 2015). However, by light microscopy, it is not possible to differentiate the smallest and most numerous organisms – single-celled picoplankton (HELCOM, 2021). Furthermore, rare and/or sparsely occurring species may remain undetected and accurate species identification of some organisms is not possible based on their morphology, especially when samples are preserved, as is routinely done in monitoring. Moreover, morpho-taxonomy based monitoring requires highly skilled professionals, is laborious, time-consuming, and therefore limited in temporal and spatial resolution, as well as affected by the variation in individual taxonomic expertise of the analysts (Leese, 2022).

Novel genomic tools can overcome many of the mentioned difficulties and could complement traditional bioassessment and biomonitoring. Currently, various approaches are independently developed in different institutions, thereby hindering harmonized routine application of genetic methods (Leese, 2022). The need to include genetic methods in monitoring programs is recognized nationally and internationally, but the implementation requires additional resources and expertise in molecular biology and bioinformatic tools. In Finland, the aim to develop genetic methods to support phytoplankton monitoring has been recorded in the Finnish Marine Monitoring Roadmap (Pitkänen et al., 2020) and in the Manual for marine monitoring in Finland 2020–2026 (Rantajärvi et al., 2020). At the EU level, progress is being made towards better knowledge of marine and aquatic ecosystems, and one of the goals is to consecutively sequence and publish 50% of the DNA of organisms living in these habitats by 2030 (Lamy et al., 2020).

Genetic methods have the potential to significantly improve monitoring and research of phytoplankton diversity. Species with similar morphological features may be distinguished and the size distribution of the species to be monitored may be extended the smallest organisms (e.g. picoplankton) unidentifiable by light microscopy. It may also be easier to detect alien and harmful species occurring in low numbers. Incorporation of genetic methods into routine monitoring would facilitate an expansion of the temporal and spatial resolution of phytoplankton data, which is crucial for monitoring phytoplankton communities with distinct seasonality and geographical variability. By combining genetic techniques and light microscopy-based species-specific abundance and biomass results, it will be possible to gain a better overview of phytoplankton species composition and diversity in the future.

Contents

1 Aim of the guidelines

The purpose of these guidelines is to provide all necessary information on sampling, sample processing, molecular biological work, quality control, and bioinformatics so that the eDNA metabarcoding method can be implemented in the Finnish marine phytoplankton monitoring program to complement the standardized light microscopy method. Implementing genetic methods in monitoring programs will improve biodiversity assessment and thereby the status assessment of marine ecosystems. Publishing national guidelines and protocols will promote the implementation and standardization of international DNA surveys in marine phytoplankton monitoring.

The presented guidelines are intended for practitioners and environmental authorities in marine management. They include a brief methodological introduction on eDNA metabarcoding and an overview on other genetic methods, discuss the advantages and disadvantages of different methods (**Section 2**), and provide a detailed description of all recommended work steps, from sampling to the processing of sequencing data (**Sections 3-4**). Core terms are explained in the **Section Abbreviations and lexicon**. The pilot study, on which these guidelines are based, is presented in the **Appendix**.

2 Background

2.1 eDNA metabarcoding

DNA-based methods for species detection have been used for decades to study the ecology of organisms from all kingdoms of life, inhabiting a broad range of habitats, and have revolutionized our ability to assess diversity (Bruce et al., 2021). Rapid technological advancements in the past decade have enabled researchers to use genetic methods with increasing frequency, lower effort and costs. So called environmental DNA (eDNA) can be extracted from environmental samples, such as water, air, or sediment without the actual target organism being present in the sample. Tissues, single cells or even fragments of genomic DNA that get distributed in the environment can be enough for successful detection of a target organism.

Automated identification of multiple species from a bulk sample containing entire organisms or from a single environmental sample containing degraded DNA can be described with an approach termed DNA metabarcoding (Taberlet et al., 2012). For DNA metabarcoding, general or universal polymerase chain reaction (PCR) primers are used on mixed DNA samples from any origin followed by so-called next-generation sequencing (NGS), also termed high-throughput sequencing (HTS), to determine the species composition of the sample (Ruppert et al., 2019). Concerning larger-sized organisms such as fish, eDNA metabarcoding has a great potential to be used in monitoring, because it offers several advantages, compared to traditional monitoring techniques: besides being a non-destructive method, it allows further automation, superior species detectability, detection without a priori knowledge of species, requires lower effort and can be implemented in areas where traditional surveys are impossible (Ruppert et al., 2019). Thus, eDNA metabarcoding is nowadays sufficiently advanced to replace traditional, more invasive monitoring methods of larger organisms (Keck et al., 2022). In contrast, DNA metabarcoding may continue to give complementary rather than identical estimates compared to conventional approaches for smaller organisms, like phytoplankton, zooplankton, macroinvertebrates, and microphytobenthos (Keck et al., 2022).

eDNA metabarcoding depends on the successful amplification of specific fractions of the genome. Small subunit ribosomal RNA (SSU rRNA) genes can be targeted with the help of specific primers, amplified by PCR and sequenced. Subsequently, the obtained sequences (“barcodes”) of an environmental sample are compared to a reference database for taxon identification. This procedure requires that taxa of interest have been correctly identified, isolated, sequenced and the sequences deposited in a reference database. For many larger organisms, comprehensive reference databases exist (Keck et al., 2022), they can be easily completed by sequencing new individuals, and the phylogeny of interest is usually less complex than for smaller organisms such as phytoplankton.

Although eDNA metabarcoding might be sufficiently advanced to replace traditional monitoring for certain larger organisms, many challenges still need to be overcome for smaller and more diverse groups, such as phytoplankton. Phytoplankton comprises several, not closely related, taxonomical groups, which belong to two kingdoms of life – eukaryotes (i.e. most phytoplankton groups) and prokaryotes (cyanobacteria) – and there is a lack of databases for universal primers, which can amplify eukaryotic and prokaryotic genes simultaneously. In addition, many taxa are difficult to culture and sequence as reference, which hinders reference database generation and a comparison of eDNA metabarcoding and light microscopy data. Other challenges include variable DNA extraction efficiencies between taxa due to different cell morphologies (e.g. robust cell coverings, such as the diatom silica frustule, lessens extraction efficiency) and the high copy number variation of ribosomal genes (e.g. Mäki et al., 2017), which constrains the quantitative interpretation of metabarcoding data. In addition, there is no consensus in the research community for one particular primer pair (Latz et al., 2022).

eDNA metabarcoding studies mostly utilize Illumina NGS sequencing technology, which is restricted to relatively short (a few hundred base pairs) sequences. Since the Illumina sequencing restricts the possible length of the barcode region, it also restricts the taxonomic resolution of the results. Closely related species regularly have identical barcode regions. Therefore, the taxonomy of amplicon sequences can often not be resolved beyond the genus level (Latz et al., 2022 and references therein).

Despite the above-mentioned challenges, eDNA metabarcoding is currently the most promising genetic approach for routine phytoplankton monitoring for several reasons. Many comparative studies between genetic and morphological methods are available (e.g. Kim et al., 2019, Bolaños et al., 2020, Salmaso et al., 2020, Malashenkov et al., 2021), sequencing costs are low, and well developed bioinformatic tools exist. In addition, metabarcoding has successfully been applied to study phytoplankton in various contexts, thereby advancing our knowledge about the group's biodiversity. For example, the expedition of the French *Tara Oceans Foundation* collected more than 300 samples to analyse the eukaryotic plankton community with eDNA metabarcoding globally and found unexpected diversity (de Vargas et al., 2015). In the Baltic Sea, Hu et al. (2016) used amplicon sequencing to study the diversity of prokaryotic and eukaryotic phytoplankton and Karlson et al. (2018) recommended to introduce phytoplankton metabarcoding in Swedish marine monitoring programs as a complement to other methods, based on a comparison of microscopy- and metabarcoding-based results. In addition, metabarcoding has already been used for monitoring harmful algal blooms (Sildever et al., 2019) and guidelines to monitor such blooms with metabarcoding in Chile have been developed (Yarimizu et al., 2020). Another example of existing guidelines is the recently published eDNA metabarcoding guidelines for monitoring Alpine waters, which contain detailed metabarcoding protocols for benthic diatoms, and for prokaryotic and eukaryotic phytoplankton (Elersek, 2021).

2.2 Other genetic methods

The quantitative interpretation of microeukaryote HTS data still has several limitations (Lamb et al., 2019), although approaches to overcome the gene copy number bias of eukaryotic phytoplankton are being investigated (e.g. Martin et al., 2022). To overcome these limitations, other genetic methods can be used to screen DNA samples for the presence of one or a few pre-defined species (Bruce et al., 2021). Such "targeted species detection" methods, like quantitative PCR (e.g. conventional, quantitative, digital and digital droplet PCR) use a taxon-specific primer set and species' presence is inferred based on successful amplification. In aquatic systems, single-species detection is often applied to track rare species or invasive species in the early stages of invasion (Taberlet et al., 2018). The advantages of such methods are their simplicity and potentially their sensitivity, whereas the disadvantages include the need for extensive and costly validation studies to ensure specificity of the assay, and the limited information obtained (Bruce et al., 2021).

While eDNA metabarcoding studies mostly utilize Illumina sequencing of relatively short sequences, which restricts the possible length of the barcode region and the taxonomic resolution, much longer sequences can be recovered with "third generation" or long-read sequencing, through platforms from Pacific Bioscience (PacBio) and Oxford Nanopore Technologies. Long-read sequencing allows sequencing of large fractions of the rRNA operon and potentially reconstruction of well-resolved phylogenetic trees, but the disadvantages include higher costs, technical limitations, and a lack of established pipelines (Latz et al., 2022 and references therein), which makes them currently less attractive for monitoring.

In contrast to the above-mentioned primer-based metabarcoding methods, primer-independent assays allow simultaneous assessment of entire communities by exploiting the taxonomic information contained in metagenomes (Obiol et al., 2020). Untargeted metagenomics uses shotgun sequencing of the genomic DNA isolated from an environmental sample and requires no PCR selection of a particular taxonomic group (Cordier et al., 2021). Recently, the photosynthetic gene *psbO*, which is universally

and exclusively present in photosynthetic prokaryotes and eukaryotes, mainly in one copy per genome, was targeted from metagenomes and found to be a promising new genetic marker for molecular-based evaluations of entire phytoplankton communities (Pierella Karlusich et al., 2022). Such a marker would allow simultaneous amplification and quantification of eukaryotes and prokaryotes.

Another interesting approach is metatranscriptomics, which also circumvents the PCR selection bias, but includes only transcribed RNA (mRNA, rRNA, supposedly functional) to depict the metabolic capabilities of the community, and the expressed genes at the moment of sampling (Cordier et al., 2021). Despite the advantages of novel genetic methods, they are still more expensive, and many challenges need to be overcome before they can be implemented in routine monitoring programs.

2.3 Standardization of eDNA methods

eDNA methods are not envisioned to replace conventional phytoplankton monitoring methods, although conventional monitoring approaches have several known limitations (see above, and e.g. Schaumburg et al., 2021, for an overview). The pros and cons of eDNA analysis, compared to traditional approaches, were summarized by Pawlowski et al. (2020), and among the most critical disadvantages are the lack of complete reference databases (or "barcode libraries") and standardization. The need for standard development and international coordination has been defined as one of the primary development areas in the "Roadmap for implementing environmental DNA (eDNA) and other molecular monitoring methods in Finland" (Norros et al., 2022). Standardized protocols for DNA-based algal monitoring exist so far only for benthic diatoms (CEN/TR 17245:2018) and for the management of diatom barcodes (CEN/TR 17244:2018). The use of diatom-eDNA based indices for water quality assessment is currently intensively studied and in Switzerland, a Swiss Molecular Diatom index is under development (Pawlowski et al., 2020).

2.4 Needs for biodiversity data

Accurate species data are needed since biodiversity is an essential community property that has been shown to drive fundamental ecosystem processes and regulate the stability of those processes both in terrestrial and aquatic ecosystems (Cardinale et al., 2012; Eriksson & Hillebrand 2020). Detailed biodiversity data are required to be able to develop an indicator that describes the diversity of phytoplankton and that can be used in marine management. Such a phytoplankton biodiversity indicator is currently lacking in e.g. the HELCOM area (the Baltic Sea).

Phytoplankton communities consist of numerous species with different functional characteristics. In addition to taxonomic diversity, also functional phytoplankton diversity needs to be considered more carefully in marine management in the future. Investigating functional diversity and structure relies currently on taxonomic data, since functional trait data are species-specific (Litchman & Klausmeier, 2008, Leruste et al., 2018, Graco-Roza et al., 2021). Because of this, it is important to develop monitoring methods that could lower the sensitivity of functional results to challenges in taxonomic identification. Thus, eDNA methods could also support the development of an indicator for functional phytoplankton diversity.

2.5 Current monitoring supplemented by eDNA metabarcoding

Finnish marine phytoplankton monitoring follows the guidelines described in the Manual for marine monitoring in Finland 2020–2026 (Rantajärvi et al., 2020) and the Manual for Marine Monitoring in the COMBINE program of HELCOM (HELCOM, 2021). In brief, phytoplankton samples from offshore areas are taken as integrated water samples from 0-10 m depth with a tube sampler (Niskin or

HydroBios sampler). The samples are immediately preserved with acidic Lugol's solution and stored in a dark refrigerator. Following the international CEN standard (CEN, 2015), the Utermöhl (1958) inverted light microscopy method is used to analyse the species composition and their abundances in the samples. The observed phytoplankton taxa are determined as accurately as possible, and the specimens are furthermore categorized into size classes (Olenina et al., 2006 and its annually updated annex), to facilitate the most accurate biovolume determination possible. The Finnish marine phytoplankton monitoring focuses on primary producing (photoautotrophic and mixotrophic) taxa, but heterotrophic dinoflagellates and other heterotrophic flagellates are included to enable wider taxonomic studies. Single-celled <2 µm-sized picoplankton is not included, since the method (light microscopy of Lugol's-preserved samples) is not suitable for separating single-celled autotrophic picoplankton from single-celled heterotrophic bacteria, and the settling of such small particles in the settling chambers is uncertain. The size-specific taxonomic counting results are converted to biomass (µg L^{-1}, wet weight), assuming a density of 1 kg L^{-1}. Data are stored in the open Hertta database of the Finnish Environment Institute (syke.fi/avoindata, in Finnish), and are openly available also through the database of the European Marine Observation and Data Network (EMODnet, emodnet-biology.eu).

Implementing the tested eDNA method has great potential in supplementing the current light microscopy method, since (1) species with similar morphological features could potentially be distinguished from each other genetically, and (2) the size distribution of monitored species could be expanded to cover small-sized organisms that cannot be enumerated or identified taxonomically by light microscopy (<2 µm-sized single-celled picoplankton). Furthermore, (3) eDNA methods potentially facilitate the detection of alien and harmful species that occur in numbers too low to be detected by routine light microscopy. Introducing eDNA metabarcoding to complement the standardized light microscopy method would advance conventional monitoring and research of phytoplankton communities to assess biodiversity and the state of the marine environment. Although amplicon sequencing has successfully been used in studies on Baltic Sea eukaryotic phytoplankton (e.g. Hu et al., 2016, Bennke et al., 2018, Filker et al., 2019), it has not yet been utilized in routine phytoplankton monitoring. Thus, there is an obvious need to test the usability of an eDNA method as a part of monitoring. The guidelines described in this document are the result of the first pilot project (**Appendix**) testing the integration of eDNA metabarcoding in Finnish marine phytoplankton monitoring. In the future, the guidelines will be developed further, according to evolving genetic methods and international guidelines and standards.

3 Protocol

This protocol is a practical guide for utilizing eDNA metabarcoding in the Finnish marine phytoplankton monitoring. All the necessary information from sampling to bioinformatics is provided in detail in the following sections.

3.1 Preparations before sampling

Sampling stations, sampling depths, and required water volumes need to be discussed and agreed with the cruise leader several months before the sampling cruise. To guarantee smooth sampling, all required equipment should be gathered well in advance, checked for functionality, cleaned if necessary and packed in stackable, clean boxes with lids before the cruise. The following preparations should be done before packing the material listed in **Table 1**:

Preparations on land

- All reusable labware should be cleaned in the lab dish washer using a combination of acidic and alkaline detergents (e.g. Labwash premium acidrinse C and Labwash premium alkamatic (VWR) containing citric acid, and active chlorine). Alternatively, equipment can be rinsed with sodium hypochlorite (5-10% bleach) and ddH_2O thereafter (Pawlowski et al., 2020).
- Reusable filtration units should be disassembled, the separated components washed in the lab-dishwasher and the dry, assembled filtration units autoclaved.
- Functionality of the vacuum pump and Bunsen burner should be checked.

Preparations onboard the research vessel

- Lab benches and laminar flow hood should be cleaned and decontaminated before setting up the filtration system.
- Permission from the captain is required to use the Bunsen burner in or next to the laminar flow hood and the cruise leader needs to be informed.
- Before each station, the cruise leader should be reminded about the required sampling volume for each depth.
- Sampling flasks should be labelled with respective depths.
- Protective shoes, glasses, lab coat, and gloves must be worn.

Table 1. Required material for phytoplankton monitoring sampling (eDNA and microscopy)

Purpose	Item	Amount	Comment
General	Single-use disposable gloves	4 pairs of gloves per station	2 for preparing the integrated sample, 2 for working in the laminar flow hood
	Protective glasses and shoes		
	Lab coat		
	Labelling tape		
	Pencil		
	Waterproof pen		
	Scissors		
	Lab book		
	Parafilm		
Sampling for microscopy	500 mL sampling flasks	One for each depth	
	Rack for flasks		
	2L flask		To prepare an integrated sample
	200 mL measuring cylinder		
	300 mL clear-glass bottles	1 per station	
	Acidic Lugol (0.5%)	1 ml / microscopy sample	
	1 mL sterile Pasteur pipettes	1 package	
	Rack or basket to store samples in fridge		Not needed, if samples can be placed otherwise safely in fridge
Sampling for eDNA	Nalgene filtration units	1 per station	Pack 3-5 additional units for filtration blanks and as back-up
	Silicon corks and tubing		To connect vacuum pump
	Sterile Mixed Cellulose Ester Membrane (MCE) filters	2 per station	0.2 µm pore Size, 47mm diameter (Whatman)
	2 mL cryovials	2 per station	
	Rack for cryovials		
	Metal forceps	3x	
	20 mL glass beaker		For 96% EtOH
	Bunsen burner or EtOH burner		Matches for EtOH burner
	Additional butane gas cartridge		Refill for Bunsen burner
	Sodium hypochlorite (5-10% bleach)		
	EtOH 70% and 96%	ca. 0.5-1 L of both per 100 samples	Cleaning, prepare fresh dilution in advance
	EtOH spray bottle	2x	
	Milli-Q spray bottle	2x	
	Cryo-box		To store samples at -80°C
	Temperature protection glove		For -80°C freezer
	Electric vacuum or peristaltic pump		
	Collection flask and hose to connect pump (autoclaved)		
	Hand vacuum pump		As back-up
	5L collection flask with rubber cork		To collect filtrate

3.2 Sampling and sample processing

Sampling onboard the research vessel should follow the phytoplankton monitoring guidelines (HELCOM, 2021), to allow a comparison between genetic data and microscopy data. Due to implementation of eDNA sampling, cleaning procedures need to be more careful compared to routine phytoplankton sampling for microscopy analyses.

Workspace: clean lab bench on the research vessel

1. **Rinsing**: Sampling flasks should be rinsed thoroughly three times with water from the respective depths (between each rinsing step remaining water drops should be shaken out of the flasks) before taking the actual sample.
2. **Sampling**: Approximately 500 mL of seawater should be collected in rinsed and labelled sampling flasks from each of the five depths (1 m, 2.5 m, 5 m, 7.5 m and 10 m).
3. **Cleaning**: Before preparing the integrated sample, all sampling flasks and the work bench should be dried and cleaned with 70% EtOH wearing a new pair of gloves.
4. **Integrated sample**: Equal amounts of sea water from each depth (400 ml) are pooled into a 2 L flask using a clean measuring cylinder (Figure 1). To guarantee a homogeneous distribution of cells, sampling flasks of each depth need to be mixed by inverting gently five times before measuring the volume.

Figure 1. Water samples from 5 different depths (1 m, 2.5 m, 5 m, 7.5 m and 10 m) and required equipment to prepare the integrated sample (0-10 m). Photo: Jacqueline Jerney.

3.2.1 Acidic Lugol's preservation of microscopy samples

Phytoplankton samples for the light microscopy analysis are taken from the mixed integrated water sample (0-10 m). Phytoplankton samples are preserved with acidic Lugol's solution immediately after pouring the integrated sample into a 300 mL glass flask. One phytoplankton microscopy sample per monitoring station needs to be prepared.

Workspace: clean lab bench on the research vessel

1. **Preservation**: The integrated sample should be mixed by inverting gently five times before transferring 300 mL to a clean clear-glass flask. 1 mL of acidic Lugol's solution (0.5 % v/v (HELCOM, 2021)) should be added immediately using a plastic Pasteur pipette, wearing safety equipment (lab coat, gloves, glasses).
2. **Storage**: Lugol's-preserved samples should be stored in the dark at 4°C until further processing (HELCOM, 2021).
3. **Cleaning**: Sampling flasks for each depth, the integrated sample flask and the measuring cylinder need to be rinsed three times with RO-water between the stations and stored bottom-up for drying.

3.2.2 eDNA filtration

The integrated live water sample (0-10 m) used for preparing the light microscopy sample, is also used to prepare the eDNA sample. Filtration of the eDNA sample is done in a clean laminar flow hood on the research vessel. The filtered eDNA samples are stored immediately at -80°C. Two eDNA samples per sampling station need to be filtered (one as back-up). For quality control, one station should be chosen during each cruise to filter three to five independent eDNA samples, and filtration blanks should be prepared as explained in **Section 3.4**.

Workspace: laminar flow hood on the research vessel

1. **Preparations**: Turn on the laminar flow hood and wait for 5 minutes (warm-up). Take a new pair of gloves, clean the gloves, equipment and the working surface with 70% EtOH and place the needed sterile / clean equipment (filtration unit, forceps, Bunsen burner, rack, cryovials, filters, labeling pen) in the laminar flow hood (note: the Bunsen burner should be placed outside the laminar if there is enough space). Clean the integrated sample flask with 70% EtOH, place it in the laminar flow hood and wait for 2 minutes (until air is purged). Rinse the filtration unit with sample water (first mix it)!
2. **Sterilize forceps**: Use 96% EtOH and the Bunsen burner to sterilize 2 forceps - let them cool down prior to filter handling.
3. **Sterile filter**: Using the forceps place a sterile Mixed Cellulose Ester Membrane (MCE) filter carefully in the reusable filtration unit without otherwise touching it – wet with Milli-Q.
4. **Filtration**: After mixing (by inverting the flask five times), pour 500 mL of the integrated sample into the filtration unit and turn on the vacuum (**Figure 2**).
5. Make sure the entire sample is filtered, so that no water remains between the filter and the upper part of the filtration unit.
6. Carefully open the filtration unit, fold the filter twice so that the upper side with the sample is inside the filter and transfer it to the first vial of the extraction kit (or to a sterile 2 ml cryovial / Eppendorf tube).
7. **Freeze** immediately at -80°C until further processing.
8. **Repeat** steps 3. to 7. for the second filter to be stored as back-up

Figure 2. Filtration set-up in the laminar flow hood onboard the research vessel. If there is sufficient space next to the laminar flow hood, the Bunsen burner should be placed outside of the laminar. Photo: Jacqueline Jerney.

3.3 DNA extraction, PCR amplification and sequencing

After the sampling cruise, DNA extraction, PCR amplification and sequencing are conducted in a dedicated DNA laboratory on land. The process also includes quality control, as described in **Section 3.4**. The filtered eDNA samples need to be stored at -80°C until processing them and must be kept at -20°C during the transport to the laboratory on land.

Workspace: dedicated DNA laboratory on land

Preparations

- Clean the laminar flow hood (Table 2) with 70% EtOH and turn on the UV-light for 20 min.
- Get samples (filters) from -80°C freezer and store them temporarily at -20°C, process a maximum of 8 filters at a time.
- Collect or set up required material listed in Table 2.
- Turn on the heating block.
- Label tubes.
- Always use gloves and clean your gloved hands with 70% EtOH.
- Turn on the laminar flow hood, clean equipment (racks, bottles, etc.) with 70% EtOH and place all the required material in the laminar flow hood.
- After introducing equipment, wait for a few minutes until the laminar air flow is re-established.
- Clean the forceps and scissors in a 50 mL beaker filled with 96 % EtOH, sterilize them with the Bunsen burner and let them cool down.

Table 2. Material required for DNA extraction.

Purpose	Item	Amount	Comment
DNA extraction	Commercial DNA extraction kit: DNeasy PowerSoil Pro Kit (Qiagen)	1	Choose the kit depending on the number of stations (50 or 250 preparations)
	Laminar flow hood	1	
	Single-use disposable gloves	2 per sample	Should resist EtOH, amount variable
	Bunsen burner	1	Check if there is enough refill gas
	Metal scissors	1	To cut filters
	Forceps	2	To handle filters
	Racks	2	For 1.5 mL centrifuge tubes
	Bead-beater	1	FastPrep-24 5G, MP Biomedicals Centrifuge
	Glass beaker 50 mL	1	
	EtOH 70% and 96%	ca. 0.5-1 L of both per 100 samples	Cleaning, prepare fresh dilution in advance
	Pipettes and tips	1 tip per sample	Volumes: 1 mL and 100-200 µL
	Table-top centrifuge	1	e.g. Eppendorf 5424
	Heating block	1	
	Molecular biology grade water	50 µL per sample	
	Empty flask	1	To collect extraction waste
	Spray bottle	1	EtOH for cleaning
	Paper towels	variable	Cleaning

3.3.1 DNA extraction

Follow the protocol of the kit manufacturer with the following modifications:

Workspace: dedicated DNA laboratory on land

1. Using the sterile forceps take the thawing filter, unfold it slightly and cut it aseptically into small pieces (ca. 4 mm^2) using the sterile scissors above or inside the bead tube, so that all filter particles are collected in the bead tube.
2. Add 800 µL of the first solution (CD1) and incubate at 65°C for 10 min.
3. Make sure that both the lids and the tubes are labelled. Homogenize samples with FastPrep-24 5G using the following settings: 6 m/s, 45 s, matrix B (for 0.1 mm silica beads), sample volume 1 mL, two cycles, 10 s wait time in between cycles. **Make sure that the rack is properly balanced.**
4. Elute DNA using 50 µL molecular biology grade water.
5. Mix the final DNA sample by pipetting, transfer 4 µL to a PCR tube for DNA measurement and freeze the sample immediately at -20°C.
6. Transfer the samples to -80°C after the DNA measurements were carried out.

3.3.2 DNA quantification and dilution

DNA quantification is done by measuring DNA concentration/absorbance spectra using NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham MA) by choosing "Nucleic acid, dsDNA". Record the results. In addition, measure the DNA concentration with a/the Qubit fluorometer (Invitrogen, Carlsbad CA) using dsDNA HS Assay Kits. Use Qubit for quantifying DNA concentrations, and Nanodrop for assessing DNA purity.

If primer amplification, library preparation and sequencing is carried out by an external service provider (e.g. the Finnish Institute for Molecular Medicine – FIMM), the concentration of all DNA samples should be harmonized, i.e. DNA concentration and volume need to be similar in all samples. This may require dilution of samples. A concentration of at least 5 ng μL^{-1} and a volume of 20 µL is required per primer and sample for sequencing, but a higher DNA concentrations (> 10 ng μL^{-1}) is recommended.

3.4 Quality control

For the purposes of quality assurance and error flagging, replicates and negative controls should be included for the eDNA samples, according to the protocols below. In addition, including positive controls is highly recommended.

3.4.1 Replicates

Choose one station during each cruise to prepare three to five independent eDNA filters from the integrated water sample (0-10 m) and process them according to the protocol (**Section 3.2.2, Section 3.3**).

3.4.2 Negative controls

Negative controls include filtration blanks and extraction kit controls. Filtration blanks are prepared on the research vessel before filtering the first eDNA sample and after filtering the last eDNA sample (**Section 3.2.2**). Extraction kit controls are processed during the DNA extraction in the laboratory on land (**Section 3.3**).

- Filtration blank
  Filtration blanks are used to check if the entire wet-lab workflow from filtration to sequencing is free of contamination. Before and after a sampling campaign (e.g. before filtering the first sample and after filtering the last sample) filter an equivalent volume of Milli-Q to your samples (e.g. 500 mL) and process the filter the same way as all other samples (Section 3.3).
- Extraction kit control
  To rule out any contamination during the DNA extraction, carry out one extraction without adding a filter, i.e. process an empty extraction tube the same way as all other samples (Section 3.3).

3.4.3 Positive controls

Preparing positive controls is recommended. Positive controls are prepared in the laboratory on land.

- **Mock community**
  Including an artificial phytoplankton community of known composition (mock community) to evaluate the laboratory steps and to adjust bioinformatic pipelines, is highly recommended (Bradley et al., 2016; Santoferrara, 2019). The mock community is used as a positive control to check if expected taxa are correctly identified and if the DNA extraction is sensitive enough. For this purpose, 5-10 taxa from e.g. the FINMARI Culture Collection / SYKE and Tvärminne Zoological Station of the University of Helsinki (FINMARI CC), with known sequences should be chosen and transferred to fresh medium to obtain cultures in exponential growth. Next, the

cell numbers of the exponentially growing cultures need to be assessed. Each culture should be diluted (if necessary) so that the final cell density in the mock community is 10,000 cells L^{-1} for each separate taxon. 500 mL of this mock community should be filtered **(Section 3.2.2)** and processed **(Section 3.3)** as all other samples.

- **Single-species cultures**
  Additionally, single-species cultures of interest (e.g. harmful, or invasive species) with known sequences can be filtered **(Section 3.2.2)** and processed **(Section 3.3)** as all other samples as positive controls. Several cell concentrations can be included to get an estimate of the detection limit.

3.5 Primers for High Throughput Sequencing (HTS)

The choice of HTS primers is one of the most important steps for phytoplankton eDNA metabarcoding, because the sequences obtained by using a specific primer need to be part of a reference databases for successful taxa identification. Primers targeting the 16S and 18S rRNA V4 gene region were used in many pro- and eukaryotic phytoplankton metabarcoding studies (e.g. Hu et al., 2016, Needham & Fuhrman, 2016, Djurhuus et al., 2017, Filker et al., 2019, Nowinski et al,.2019, Yarimizu et al., 2020) and they are well represented in databases like Silva (Quast et al., 2013) and PR^2 (Guillou et al., 2013).

For **eukaryotes**, PCR amplification of the 18S rRNA genes is performed by targeting a ~380-bp fragment of the 18S rRNA gene variable region V4 using the specific primer set:

- Forward: TAReuk454FWD1, 5'-CCAGCASCYGCGGTAATTCC -3' (Stoeck et al., 2010)
- Reverse: V4r, 5'-ACTTTCGTTCTTGAT -3' (Bradley et al., 2016)

For **prokaryotes**, PCR amplification of the 16S rRNA genes is performed by targeting a ~390 bp fragment of the 16S rRNA gene variable region V4 using the specific primer set:

- Forward: 515F, 5′-GTGYCAGCMGCCGCGGTAA-3′ (Parada et al., 2016)
- Reverse: 806R, 5′-GGACTACNVGGGTWTCTAAT-3′ (Apprill et al., 2015)

For PCR and sequencing, samples can be sent for example to the sequencing laboratory of the FIMM Technology Centre, University of Helsinki, where the above-mentioned primers have been tested already. Information on the number of samples, the multiplexing scheme and the sequencing depth should be communicated when requesting an offer from the sequencing facility.

3.6 Bioinformatics

Bioinformatics requires specific expertise; therefore this section is aimed at molecular biology experts who are familiar with the basics of bioinformatics. 16S and 18S rRNA gene HTS reads are analysed using standardized bioinformatic pipelines and the required software and examples of scripts for processing the sequences are provided below. The described protocol focuses on quality filtering and taxonomic assignment but does not cover the downstream data analysis and statistics, which needs to be carried out subsequently (e.g. calculating alpha- and beta diversity, rarefaction curves etc.).

3.6.1 Preparations

A prerequisite for receiving and working with the sequencing data is an active user account at the Finnish IT Center for Science (CSC). Instructions on how to get started with CSC services are available at the CSC webpages (https://research.csc.fi/accounts-and-projects). To access CSC servers with a Windows computer

an SSH client (e.g. MobaXterm) needs to be installed. Once a CSC account is available and the user has opened a new project, a suitable directory structure should be created within the disc area "scratch" of the respective project on CSC server Puhti. An example is given below. In the following sections comments or batch job settings are highlighted as green and script names as blue text. **Note**: "project_number" needs to be replaced with the CSC project code throughout the scripts and the most recent versions of the CSC instructions, databases and software should checked and the code adjusted accordingly.

```
# move to respective scratch directory and create new directory structure
cd /scratch/project_number # replace "number" with your project code
mkdir rawdata
reference_data \
Info \ # to store the metadata file
18Sv4 \ # qiime2 output files of eukaryotes
16Sv4 \ # qiime2 output files for prokaryotes
```

Sequencing data (FASTQ files) can be transferred from the FIMM server to CSC server Puhti via Funet FileSender by using the following commands:

```
# change to respective scratch directory
cd /scratch/project_number/rawdata

# downloading sequencing data (use the link from the sequencing facility)
wget -O amplicon.zip https://filesender.funet.fi/download.link
```

After unzipping the files, file permissions should be changed to prevent accidental overwriting of the raw data.

3.6.2 Sequencing quality

The sequencing facility usually provides FastQC (Babraham Bioinformatics) or MultiQC (Ewels et al., 2016) reports together with the actual sequencing data. The .html files can be explored with a web browser to assess the sequencing quality. Some basic instructions can be found here: www.bioinformatics.babraham.ac.uk/projects/fastqc/. Alternatively, raw sequences can be checked with FastQC on CSC server Puhti. Samples which were not sequenced correctly and thus have a too low number of reads (e.g. less than 1000 reads), should be excluded from further analysis.

3.6.3 Batch jobs

CSC uses a batch job system to execute computing tasks on supercomputers (docs.csc.fi/computing/running/getting-started/). To analyse data and use the supercomputers efficiently, it is recommended to submit batch jobs to the SLURM batch job manager. A batch job is a scheduled program that is assigned to run on a computer without further user interaction. Instructions on how to write batch job scripts and submit them to the server Puhti can be found in the CSC docs at docs.csc.fi/computing/running/creating-job-scripts-puhti/. Before working on the entire dataset, settings should be tested on a subset of samples by submitting test scripts. SLURM batch job settings need to be adjusted for each dataset.

3.6.4 QIIME2 workflow for eukaryotes

To identify amplicon sequencing variants (ASVs) the pipeline **DADA2** (Callahan et al., 2016), implemented in **QIIME2** (Bolyen et al., 2019) should be used on the CSC server Puhti. This option allows fast and simple data processing and upscaling. Alternatively, DADA2 can be used within the R software (R Core Team, 2021,R-project.org), which is not covered in this document, but instructions can be found in a metabarcoding protocol published by Salmaso et al. (2021a) and on the DADA2 webpage (benjjneb.github.io/dada2). DADA2 and QIIME2 are a part of the bioconda environment provided by CSC and the most recent version can be loaded by the user without installing the software, as described below. The most basic analysis can be carried out using the three scripts below. To submit a script to the SLURM batch job system, the following commands can be used:

```
Sbatch 01_qiime2_data_import_18sv4.sh
Sbatch 02_qiime2_dada2_denoise_18sv4.sh
Sbatch 03_qiime2_vsearch_taxonomy_classify_18sv4.sh
```

The first step is to import paired-end sequencing data to QIIME2, by adapting the following batch job script: "`01_qiime2_data_import_18sv4.sh`".

```
#!/bin/bash

#SBATCH --job-name=qiime2_import_18sv4
#SBATCH --account=Project_number
#SBATCH --time=00:15:00
#SBATCH --ntasks=1
#SBATCH --nodes=1
#SBATCH --output=output-%x-%j.txt
#SBATCH --error=errors-%x-%j.txt
#SBATCH --cpus-per-task=2
#SBATCH --mem=2G
#SBATCH --partition=small
#SBATCH --gres=nvme:100

# Load QIIME2 following instructions at https://docs.csc.fi/apps/qiime/
module load qiime2
export TMPDIR="$LOCAL_SCRATCH"

# Import data. To use this method, the file names should look like this:
# 03-LL7-n5-18S-V4_S27_L001_R1_001.fastq.gz/
# 03-LL7-n5-18S-V4_S27_L001_R2_001.fastq.gz

  qiime tools import \
  --type SampleData[PairedEndSequencesWithQuality] \
  --input-path /scratch/project_number/rawdata/18sv4 \
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \
  --output-path /scratch/project_number/18sv4/18sv4-demux-paired-end.qza
```

Before running the next script, a tab separated **metadata file** needs to be prepared. A detailed description on how to format the file is available in the QIIME2 documentation and an example is given

below. Note: The metadata file should not be prepared on a Windows system, since hidden characters remain in the file when copy + pasting text, which will cause errors when reading the metadata file.

#SampleID	Station	Treatment	Volume
02-LL7-n3-18S-V4	LL7	N	300
03-LL7-n5-18S-V4	LL7	N	500
04-LL7-n8-18S-V4	LL7	N	800
05-LL7-s5-18S-V4	LL7	S	500
06-LL7-s8-18S-V4	LL7	S	800
07-LL7-s3-18S-V4	LL7	S	300
08-LL3-n3-18S-V4	LL3	N	300
09-LL3-n5-18S-V4	LL3	N	500
10-LL3-n8-18S-V4	LL3	N	800
13-LAN-n3-18S-V4	LAN	N	300
14-LAN-n5-18S-V4	LAN	N	500
15-LAN-n8-18S-V4	LAN	N	800
18-L17-n3-18S-V4	L17	N	300
20-L17-n8-18S-V4	L17	N	800
21-L17-n5-18S-V4	L17	N	500
22-F64-n3-18S-V4	F64	N	300
24-F64-n8-18S-V4	F64	N	800
25-F64-n5-18S-V4	F64	N	500
26-US5-n3-18S-V4	US5	N	300
28-US5-n8-18S-V4	US5	N	800
29-US5-n5-18S-V4	US5	N	500
30-NTC-n5-18S-V4	NTC	N	500
31-MS9-n3-18S-V4	MS9	N	300
31-MS9-n3-18S-V4	MS9	N	300
35-MS9-n5-18S-V4	MS9	N	500
37-LL7-L3-18S-V4	LL7	L	300
39-LL3-L3-18S-V4	LL3	L	300
41-LAN-L3-18S-V4	LAN	L	300
44-L17-L3-18S-V4	L17	L	300
46-F64-L3-18S-V4	F64	L	300
47-US5-L3-18S-V4	US5	L	300
49-MS9-L3-18S-V4	MS9	L	300
51-NTC-L3-18S-V4	NTC	L	300
52-MOC-N3-18S-V4	MOC	N	300
53-MOC-L3-18S-V4	MOC	L	300
54-blank-00-18S-V4	BLA	NA	0

As a next step, short reads need to be removed using the software cutadapt (Martin, 2011), implemented in QIIME2 followed by denoising as described in "`02_qiime2_dada2_denoise_18sv4.sh`". Using DADA2 in QIIME2 allows filtering out noisy sequences, correcting errors in marginal sequences, removing chimeric sequences, removing singletons, joining denoised paired end reads, and then dereplicating those sequences, so additional filtering should not be necessary (docs.qiime2.org/2022.2/tutorials/overview/#taxonomy-classification-and-taxonomic-analyses). Note: truncating settings (--p-trunc-len-f , --p-trunc-len-r) need to be adjusted for each dataset.

```
#!/bin/bash

#SBATCH --job-name=qiime2_denoise_18sv4
#SBATCH --account=Project_number
#SBATCH --time=01:00:00
#SBATCH --ntasks=1
#SBATCH --nodes=1
#SBATCH --output=output_%j_%x.txt
#SBATCH --error=errors_%j_%x.txt
#SBATCH --cpus-per-task=4
#SBATCH --mem=8G
#SBATCH --partition=small
#SBATCH --gres=nvme:100

# The cpu and memory requirements are per the csc qiime denoise example
# Load QIIME2 as per instructions at https://docs.csc.fi/apps/qiime/
module load qiime2
export TMPDIR="$LOCAL_SCRATCH"

# define top directory (store it as variable) to shorten paths
top=/scratch/project_number/your_path/ # replace with your specific path

# remove short reads
qiime cutadapt trim-paired \
  --verbose \
  --i-demultiplexed-sequences $top/18sv4-demux-paired-end.qza \
  --p-front-f ^CCAGCASCYGCGGTAATTCC \
  --p-front-r ^ACTTTCGTTCTTGAT \
  --p-overlap 12 \
  --p-minimum-length 100 \
  --o-trimmed-sequences $top/18sv4-demux-paired-end-trimmed.qza

# create visualization
qiime demux summarize \
  --i-data $top/18sv4-demux-paired-end-trimmed.qza \
  --o-visualization $top/18sv4-trimmed_reads.qzv

# download test-18sv4-trimmed_reads.qzv to your local computer
# drag and drop it to https://view.qiime2.org/
# check Interactive Quality Plot

# denoise
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs $top/18sv4-demux-paired-end-trimmed.qza \
  --p-trim-left-f 0 \
  --p-trim-left-r 0 \
  --p-trunc-len-f 214 \
  --p-trunc-len-r 182 \
  --p-chimera-method consensus \
  --p-n-threads $SLURM_CPUS_PER_TASK \
```

```
  --o-table $top/18sv4-dada2-table.qza \
  --o-representative-sequences $top/18sv4-dada2-rep-seqs.qza \
  --o-denoising-stats $top/18sv4-dada2-denoising-stats.qza \

# Summarize a feature table - check how many sequences dada2 removed
qiime feature-table summarize \
--i-table $top/18sv4-dada2-table.qza \
--m-sample-metadata-file /scratch/project_number/info/18sv4-metadata.tsv\
--o-visualization $top/18sv4-dada2-table.qzv

# Summarize representative sequences
qiime feature-table tabulate-seqs \
  --i-data $top/18sv4-dada2-rep-seqs.qza \
  --o-visualization $top/18sv4-dada2-rep-seqs.qzv

# Summarize denoising statistics
qiime metadata tabulate \
  --m-input-file $top/18sv4-dada2-denoising-stats.qza \
  --o-visualization $top/18sv4-dada2-denoising-stats.qzv
```

Check and evaluate the created output before continuing. Denoising settings need to be adjusted, depending on each individual dataset after inspecting the Interactive Quality Plot of the 18sv4-trimmed_reads.qzv file with the help of the qiime2view webpage. Furthermore, filtration blanks and negative controls must be checked for possible contamination. If blanks and controls are clean, they can be omitted from further analysis.

As the last step, run the third script "03_qiime2_vsearch_taxonomy_classify_18sv4.sh", which uses the VSEARCH (Rognes et al., 2016) plugin and the database PR2 to classify the new sequences.

```
#!/bin/bash

#SBATCH --job-name=qiime2_vsearch_tax_class_18sv4
#SBATCH --account=Project_number
#SBATCH --time=20:00:00
#SBATCH --ntasks=1
#SBATCH --nodes=1
#SBATCH --output=output_%x_%j.txt
#SBATCH --error=errors_%x_%j.txt
#SBATCH --cpus-per-task=24
#SBATCH --mem=8G
#SBATCH --partition=small
#SBATCH --gres=nvme:100


#Load QIIME2 as per instructions at https://docs.csc.fi/apps/qiime/
module load qiime2
export TMPDIR="$LOCAL_SCRATCH"

# define top directory
top=/scratch/project_number/
```

```
# Download most recent pr2database version
cd $top/reference_data
wget
https://github.com/pr2database/pr2database/releases/download/v4.14.0/pr2_
version_4.14.0_SSU_dada2.fasta.gz
gunzip pr2_version_4.14.0_SSU_dada2.fasta.gz

# change into the sample directory
cd $top/18sv4

# import reference data
qiime tools import \
  --type 'FeatureData[Sequence]' \
  --input-path $top/reference_data/pr2_version_4.14.0_SSU_mothur.fasta \
  --output-path $top/reference_data/pr2_4.14.0_SSU_ref-seqs.qza

qiime tools import \
  --type 'FeatureData[Taxonomy]' \
  --input-format HeaderlessTSVTaxonomyFormat \
  --input-path $top/reference_data/pr2_version_4.14.0_SSU_mothur.tax \
  --output-path $top/reference_data/pr2_4.14.0_SSU_ref-taxonomy.qza

# run classifier
qiime feature-classifier classify-consensus-vsearch \
  --i-query 18sv4-dada2-rep-seqs.qza \
  --i-reference-reads ../reference_data/pr2_4.14.0_SSU_ref-seqs.qza \
  --i-reference-taxonomy ../reference_data/pr2_4.14.0_SSU_ref-
taxonomy.qza \
  --p-threads $SLURM_CPUS_PER_TASK \
  --o-classification 18sv4-vsearch-classification.qza

# prepare visualization (list of sequences and associated taxa)
qiime metadata tabulate \
  --m-input-file 18sv4-vsearch-classification.qza \
  --o-visualization 18sv4-vsearch-classification.qzv

# generate bar plot
qiime taxa barplot \
  --i-table 18sv4-dada2-table.qza \
  --i-taxonomy 18sv4-vsearch-classification.qza \
  --m-metadata-file ../info/18sv4-metadata.tsv \
  --o-visualization 18sv4-bar-plots.qzv
```

The visualizations (e.g. bar plots, qzv. files) can be downloaded to a local computer, viewed at view.qiime2.org with a web browser and the associated taxa list downloaded for the downstream analysis.

The described workflow produces results for all eukaryotic protists, including heterotrophic ciliates and fungi. The following command can be used to filter the ASV table to contain only certain groups of

interest, like Dinophyceae, Chlorophyta, Cryptophyta, Haptophyta and Ochrophyta (including Bacillariophyceae).

```
# filtering eukaryotic algae
qiime taxa filter-table \
  --i-table 18sv4-dada2-table.qza \
  --i-taxonomy 18sv4-vsearch-classification-bold.qza \
  --p-include Dinophyceae,Chlorophyta,Cryptophyta,Haptophyta,Ochrophyta \
  --o-filtered-table ./algae/algae_final_table.qza \
```

3.6.5 QIIME2 workflow for cyanobacteria

The analysis of 16S rRNA gene sequence data is performed as described in 18S rRNA gene pipeline, with the following exceptions:

- When removing short reads with cutadapt command, remember to use 16S rRNA gene primer sequences as parameters:

  --p-front-f ^GTGYCAGCMGCCGCGGTAA \

  --p-front-r ^GGACTACNVGGGTWTCTAAT \

- For classifying the sequences, download the Naive Bayes classifier pre-trained on Silva 138 99% OTUs from 515F/806R region of sequences from QIIME2 website data.qiime2.org/2022.2/common/silva-138-99-515-806-nb-classifier.qza
- For classification, use the following batch job:

```
#!/bin/bash

#SBATCH --job-name= tax_class_16sv4
#SBATCH --account=project_number
#SBATCH --time=30:00:00
#SBATCH --ntasks=1
#SBATCH --nodes=1
#SBATCH --output=output_%x_%j.txt
#SBATCH --error=errors_%x_%j.txt
#SBATCH --cpus-per-task=8
#SBATCH --mem=64G
#SBATCH --partition=small
#SBATCH --gres=nvme:100

module load qiime2
export TMPDIR="$LOCAL_SCRATCH"

# change into the sample directory
cd /scratch/project_number/16sv4

# download the pre-trained classifier
wget https://data.qiime2.org/2022.2/common/silva-138-99-515-806-nb-
classifier.qza

# run classifier
```

```
qiime feature-classifier classify-sklearn \
  --i-classifier silva-138-99-515-806-nb-classifier.qza \
  --i-reads rep-seqs.qza \
  --p-n-jobs $SLURM_CPUS_PER_TASK \
  --o-classification 16sv4-taxonomy.qza

# prepare visualization (list of sequences and associated taxa)
qiime metadata tabulate \
  --m-input-file 16sv4-taxonomy.qza \
  --o-visualization 16sv4-taxonomy.qzv

# generate bar plot
qiime taxa barplot \
  --i-table 16sv4-table.qza \
  --i-taxonomy 16sv4-taxonomy.qza \
  --m-metadata-file sample-metadata.tsv \
  --o-visualization 16sv4-bar-plots.qzv
```

As a result of the 16S rRNA gene sequence analysis you will get information of the entire bacterial community of the sample. With the following command you can select only cyanobacteria for further analysis, such as creating bar plots, as described earlier.

```
# filtering prokaryotic algae
qiime taxa filter-table \
  --i-table sample-frequency-filtered-table.qza \
  --i-taxonomy taxonomy.qza \
  --p-include cyanobacteria \
  --o-filtered-table cyano-table.qza
```

4 Recommendations and future perspectives

4.1 Integrating eDNA metabarcoding into phytoplankton monitoring

The combination of eDNA metabarcoding and light microscopy can substantially improve phytoplankton community assessment (Ruppert et al., 2019, Huo et al., 2020, Pereira et al., 2021). Nevertheless, further development of eDNA metabarcoding is required to produce more accurate taxon lists, as has been suggested previously (Hanžek et al., 2021). Especially the reference databases for sequences need to be extended. An improved integrated methodology will allow more accurate assessment of the phytoplankton biodiversity. This is expected to improve assessing the ecological status of the Baltic Sea, which is crucial to understanding the impact of anthropogenic activities on ecosystem stability and functioning. The presented guidelines are based on the pilot study described in the **Appendix**. The guidelines will be developed further in the future, according to evolving genetic methods and international guidelines or standards.

We recommend integrating eDNA metabarcoding into routine phytoplankton monitoring, to complement the current morphological light microscopy-based approach. Preferably, the introduction of eDNA methodology should include the following steps: 1) pilot study combining light microscopy and eDNA metabarcoding to test the presented guidelines in different seasons, 2) isolating, culturing and sequencing phytoplankton species to improve reference databases, 3) improving and evaluating the bioinformatic pipeline using phylogenetic placement 4) assessing long-term data management solutions, 5) investigating primer-independent and taxonomy-free environmental genomics strategies, 6) outlining how to fine-tune biodiversity assessment by combining light microscopy, eDNA metabarcoding and other methods such as imaging and flow cytometry, and 7) standardizing eDNA workflows, according to international developments. We recommend that all these steps should include institutional, national, and international collaboration, so that the work phases, costs, and knowledge could be shared among the collaborators.

4.2 Filtration and sample processing

We recommend reusable filtration units and MCE filters for phytoplankton eDNA monitoring because they offer several advantages compared to closed filters (systems in which the filter membrane is enclosed within an outer housing): better DNA recovery, easier handling of the filter if mechanical cell disruption is part of the protocol, less plastic waste, and a lower price. MCE filters might also yield more consistent community composition than polyethersulfone (PS) filters, as suggested for metazoan communities (Majaneva et al., 2018). Although some authors suggested using closed filters for eDNA sampling from aquatic environments (Pawlowski et al., 2020, Yarimizu et al., 2020, Elersek, 2021), the presented guidelines advocate open filters, because they allow easier mechanical disruption of cells, which is needed for DNA extraction from e.g. diatoms (Bruce et al., 2021). Open filters carry a greater risk of contamination, compared to closed filters, therefore filter handling must be carried out with the utmost care (e.g. under sterile conditions in the laminar flow hood onboard the ship and on land) and the use of negative controls is necessary. One optional improvement for future eDNA work is the use of a sodium hypochlorite (NaOCl, 5-10% bleach,) to decontaminate all equipment for filtration and DNA extraction.

In these guidelines we suggest filtering 500 mL (of 0-10 m integrated water sample, **Section 3.2.2**) when sampling for the northern Baltic Sea spring bloom to obtain enough DNA, without the risk of reduced DNA extraction efficiency due to inhibiting substances. Before incorporating the described protocol into routine monitoring programs, further inhibition testing should be carried out. A common reason for false negative results is PCR inhibition by chemicals and compounds in the sample that

interact with the PCR and either reduce its efficiency or cause it to fail completely, even when target DNA is present (Bruce et al., 2021). Complex humic substances are a well-known inhibitor of PCR in water samples (Stoeckle et al., 2017) and due to the high load of humic substances to the Baltic Sea, it is important to test the extracted DNA for inhibition to avoid false negative results (Bruce et al., 2021).

We recommend DNA concentrations of >10 ng μL^{-1} for monitoring to improve the detection of rare and sparsely occurring species. In addition, we recommend using non-preserved samples for DNA extraction. Mäki et al. (2017) found that acidic Lugol's solution resulted in equal DNA yields and PCR performance but affected the community profile of a mock community, i.e. the dominance of some taxa decreased or increased. Thus, acidic Lugol's preservation may lead to biased community composition if the sequences of some taxa are overrepresented, compared to rare sequences. In addition, acidic Lugol's preservation reduced the DNA yield for some samples in the pilot study (**Appendix**) for unknown reasons. Since the effect of Lugol's solution is still ambiguous, untreated samples should be used for DNA extraction and sequencing.

4.3 Recommended quality control

The inherent sensitivity of eDNA methods to contamination calls for increased rigor in quality assurance and control measures to prevent and detect contamination. It is highly recommended to use sequence replicates to allow for error flagging and removal, and controls (e.g. randomized negative controls at field and lab steps) to monitor contamination and cross-contamination (e.g. Santoferrara, 2019 and references therein, Pawlowski et al., 2020). Therefore, we suggest including three to five independent samples from at least one station per cruise to be used as sequence replicates.

Negative controls should be included as described in the protocol (**Section 3.4**). Positive controls are recommended to be included using a mock community of species with known sequences (**Section 3.4**). Species selected for the mock community should not change their morphology too much in culture over time (as e.g. certain centric diatoms do), so that they can still be identified and quantified by light microscopy, which is required for the positive control. It is recommended to check the candidate mock community species by light microscopy before establishing the mock community.

4.4 DNA measurement

DNA concentration should be measured with two methods: Qubit fluorometer (Invitrogen, Carlsbad CA) using dsDNA HS Assay Kits to get accurate DNA quantification and Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham MA) or other spectrophotometer to assess DNA quality, as outlined by Bruce et al. (2021, and references therein):

- The Qubit fluorometer uses an intercalating dye that binds directly to DNA and thus measures the quantity of double or single stranded DNA precisely even at very low concentrations, although it provides no estimate of DNA purity. Users should be aware that readings from the Qubit are affected by temperature, so care needs to be taken to ensure consistency in this regard.
- The Nanodrop spectrophotometer usually gives an inflated estimate of DNA concentration because it measures even single nucleotides. However, the 260/280 ratio given by the Nanodrop can be useful for indicating RNA and protein contamination (values between 1.8 and 2.0 denote pure DNA; below 1.8 indicates phenol, salt, protein, or polysaccharide contamination).

4.5 Bioinformatics

The suggested pipeline for eukaryotes is a basic first step and does not yield the best possible result at lower taxonomic (i.e. more specific) levels. More complex pipelines for phytoplankton, applying e.g.,

phylogenetic placement (Czech et al., 2022), are still in development and need to be thoroughly evaluated before they can be reliably implemented in phytoplankton monitoring. Annotation by using an evolutionary placement algorithm (EPA) rather than taxonomic assignment with VSEARCH can improve the taxonomic resolution considerably, as has been demonstrated for freshwater phytoplankton communities (Huo et al., 2020). An alternative pipeline using an in-house database and phylogenetic inference was tested for diatoms and dinoflagellates during the pilot study (**Appendix,** unpublished data), and further testing of the pipeline with all major phytoplankton groups should be carried out. In addition, the used primers should be evaluated with the newly assembled database metaPR, consisting of processed 18S rRNA (mostly V4) metabarcodes that are annotated with the PR^2 reference sequence database (Vaulot et al., 2022) and alternative primers considered.

For 16S rRNA gene analysis the suggested pipeline is a comprehensive tool for analysing and visualizing data, but alternative primers amplifying a longer region should be considered to gain higher taxonomic accuracy. For example, the V3-V4 primers 341F-805R have been used in Baltic Sea (Herlemann et al., 2011).

In addition to QIIME2, an open-source software Mothur (Schloss et al., 2009) is a suitable tool for microbial bioinformatics. To simplify data visualization and downstream data analysis in the future, the R software implementation of DADA2 could be adapted from the continuously updated version available at the DADA2 web page (benjjneb.github.io/dada2, Callahan et al., 2016) for 18S and 16S rRNA genes. This has been already demonstrated by Salmaso et al. (2021b) for freshwater phytoplankton.

4.6 Alternative strategies and methods

Usually one of the following general implementation strategies is applied for monitoring of ecosystems with genomic methods (Cordier et al., 2021): (a) taxonomy-based analyses focused on identification of known bioindicators or described taxa; (b) *de novo* bioindicator analyses; (c) structural community metrics including inferred ecological networks; and (d) functional community metrics (metagenomics or metatranscriptomics). The strategy applied in these guidelines is a taxonomy-based analysis. The challenge in using this strategy is that it heavily relies on the completeness of reference databases, and that adding new missing taxa involves expensive and laborious cell isolation, culturing, and sequencing. If a taxon of interest has not been sequenced and is not part of a reference database yet, then taxonomic assignment of environmental samples is less accurate, i.e. a taxon of interest is assigned to a less specific taxonomical level. For certain purposes, e.g. for developing bioindicators or assessing functional diversity, alternative implementation strategies should also be considered for biomonitoring. Such alternative strategies could be e.g. the utilization of genetic results to assess biodiversity even though taxonomic identities are not known (OTU or ASV approach) or utilizing information on functions that are present in the community even though it is not clear which taxa are behind the functions. Although there is still room for improvement of the taxonomy-based strategy to better bridge the current gaps between taxonomy-dependent molecular and morphology-based methods, it is a promising approach to improve biomonitoring and develop bioindicators (Cordier et al., 2021).

Besides improving the presented guidelines, new promising developments should be considered as alternatives for complementing monitoring in the future. For example, primer-free shotgun metagenomics, which allows identification and quantification of prokaryotes and eukaryotes simultaneously without primer-bias (Pierella Karlusich et al., 2022) seems to be a promising approach. For prokaryotes, shotgun metagenomics has become the standard in microbial ecology studies due to increases in sequencing throughput at lower costs and continued improvements in bioinformatic analysis pipelines (Grieb et al., 2020). Alternatively, the three-domain universal primer (515Y/926R) might be a good choice for monitoring, since it can resolve community composition for 16S and 18S rRNA genes in a single PCR reaction with quantifiable and manageable biases (Yeh et al., 2021).

Acknowledgements

The pilot study and the compilation of the presented methodological guidelines were funded by the Finnish Ministry of Environment (project VN/9072/2021-YM-1, Geneettiset menetelmät kasviplanktonseurantaan, GeMeKa). We thank the SYKE Seed Money project "A pilot for implementing environmental DNA (eDNA) based methods into environmental and biomonitoring", led by researcher Tiina Laamanen, for the collaboration in collecting the samples and analysing the bioinformatics of cyanobacterial eDNA samples. PhD Henna Savela is thanked for the expert advice. We thank the Baltic Sea and Nordic countries phytoplankton eDNA expert network, launched within the framework of the GeMeKa project, and planned to continue beyond, for constructive discussions in our online meetings. SYKE and the Finnish Marine Research Infrastructure FINMARI are thanked for providing the facilities onboard research vessel Aranda and at the Marine Research Laboratory.

Abbreviations and lexicon

Modified and amended from Pawlowski et al. (2020).

16S rRNA	16S ribosomal RNA. The 16S rRNA gene is used for phylogenetic studies of prokaryotes as it is highly conserved between different species of bacteria and archaea.
18S rRNA	18S ribosomal RNA. Sequence data from 18S rRNA genes is used for phylogenetic studies of eukaryotes as its slow evolutionary rate makes it suitable to reconstruct ancient divergences.
Amplicon	A piece of DNA or RNA that is the source and/or product of amplification or replication events. It can be formed artificially, using various methods including polymerase chain reactions (PCR) or naturally through gene duplication.
ASV	Amplicon sequence variant. Individual DNA sequences produced by high-throughput amplicon sequencing after the removal of spurious sequences generated during PCR amplification and sequencing. ASVs are always the same (constant) and can be compared between different data sets.
Barcode	A barcode sequence is a short nucleotide sequence from a standard genetic locus for use in species identification.
Bulk sample	A sample consisting of whole organisms and their fragments originating from the environment and collected manually.
Chimera	Genomic artefact created during PCR amplification by combining DNA fragments of different origins.
Contamination	Presence of extraneous DNA, which does not originate from the sample.
CSC	Finnish IT Center for Science.
Digital PCR (dPCR)	A PCR technique in which a sample is partitioned into thousands of subsamples (droplets in digital droplet PCR). A PCR reaction occurs within each subsample and successful amplification is detected by fluorescence.
DNA extraction	A laboratory process of chemical and physical steps to release and purify DNA from cells or other material.
eDNA	Environmental DNA, a complex mixture of genomic DNA from many different organisms found in an environmental sample.
FINMARI CC	FINMARI Culture Collection/SYKE Marine Research Centre and Tvärminne Zoological Station of the University of Helsinki.
Genetic marker	A genomic DNA region (e.g. fragment of a gene, or region of the 18S rRNA gene), which allows to identify species within a particular taxonomic group.
HELCOM	Helsinki Commission. Baltic Marine Environment Protection Commission.
HTS	High-throughput sequencing: a method producing millions of DNA sequences through massively parallel sequencing technologies, also known as next-generation sequencing (NGS).

Library A collection of DNA fragments prepared for high-throughput sequencing. Each DNA fragment is flanked with specific adapters to both ends.

MCE filter Mixed Cellulose Ester Membrane filter.

Metabarcode The sequences resulting from metabarcoding and produced by high-throughput amplicon sequencing.

Metabarcoding An approach to identify multiple species in a complex sample (e. g., eDNA or bulk sample) based on high-throughput amplicon sequencing.

Mock community A defined mixture of microbial cells created *in vitro* to simulate the composition of a microbiome sample, or the nucleic acid isolated therefrom.

mRNA Messenger ribonucleic acid, a type of single-stranded RNA involved in protein synthesis.

Multiplexing An approach consisting of simultaneous PCR amplification of different markers or sequencing of different samples to optimize the molecular workflow.

Negative controls Measures that allow tracking potential contaminations during field sampling, DNA extraction, and PCR.

Operational taxonomic unit (OTU)

A cluster of sequences grouped by similarity that are considered as a proxy for molecular species. Nowadays more commonly used in prokaryote research, while eukaryote research is moving toward using ASVs. OTUs cannot be compared between different data sets, unlike ASVs.

PCR inhibitor A factor preventing or limiting amplification of DNA during PCR through interaction with the DNA template, polymerase enzyme or other cofactors used in the PCR.

Polymerase An enzyme that synthesizes DNA molecules during PCR by replicating an existing DNA sequence.

Polymerase chain reaction (PCR)

A process to generate copies of a particular fragment of DNA with the help of DNA polymerase.

Positive control A group in an experiment that receives a treatment with a known result, and therefore should show a particular change during the experiment.

Primer A short single-stranded piece of DNA utilized for DNA replication during PCR. Usually two primers (also called a primer pair) that flank the region to be replicated by polymerase are used.

PS filter Polyethersulfone filter.

Quantitative PCR (qPCR)

An approach that allows quantification of DNA products during polymerase chain reactions based on fluorescence intensity. The fluorescent signal can be produced either by a non-specific dye binding to double stranded DNA or by a specific probe. The fluorescent signal increases with the accumulation of DNA and is then quantified against the signal produced by a known amount of DNA.

Reads	A common term used for DNA sequences generated during high-throughput sequencing.
Reference database	A collection of DNA sequences individually linked to morphologically identified specimens, which are ideally stored in museum collections. Reference databases serve to taxonomically assign DNA sequences retrieved from eDNA or bulk samples.
Replicate	Repeated DNA sampling or PCR amplification to estimate the variability associated with the method and control the consistency of obtained results.
rRNA	Ribosomal ribonucleic acid, the RNA component of ribosomes, the molecular machines that catalyze protein synthesis.
Shotgun sequencing	Laboratory technique for determining the DNA sequence of an organism's genome. The method involves randomly breaking up the genome into small DNA fragments that are sequenced individually. A computer program looks for overlaps in the DNA sequences, using them to reassemble the fragments in their correct order to reconstitute the genome.
SSU rRNA	Small subunit ribosomal RNA.
Taxonomic assignment	The taxonomic identification of DNA sequences based on reference databases.

References

Andersson, A.F., Riemann, L., Bertilsson, S. 2010. Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. ISME J. 4: 171–181. https://doi.org/10.1038/ismej.2009.108

Apprill, A., McNally, S., Parsons, R., Weber, L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat. Microb. Ecol. 75: 129–137. https://doi.org/10.3354/ame01753

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., Stamatakis, A. 2019. EPA-ng: Massively parallel evolutionary placement of genetic sequences. Syst. Biol. 68: 365–369. https://doi.org/10.1093/sysbio/syy054

Bennke, C.M., Pollehne, F., Müller, A., Hansen, R., Kreikemeyer, B., Labrenz, M. 2018. The distribution of phytoplankton in the Baltic Sea assessed by a prokaryotic 16S rRNA gene primer system. J. Plankton Res. 40: 244–254. https://doi.org/10.1093/plankt/fby008

Bolaños, L.M., Karp-Boss, L., Choi, C.J., Worden, A.Z., Graff, J.R., Haëntjens, N., Chase, A.P., Della Penna, A., Gaube, P., Morison, F., et al. 2020. Small phytoplankton dominate western North Atlantic biomass. ISME J. 14: 1663–1674. https://doi.org/10.1038/s41396-020-0636-0

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37: 852–857. https://doi.org/10.1038/s41587-019-0209-9

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

Bruce, K., Blackman, R., Bourlat, S.J., Hellström, A.M., Bakker, J., Bista, I., Bohmann, K., Bouchez, A., Brys, R., Clark, K., et al. 2021. A practical guide to DNA-based methods for biodiversity assessment. Advanced Books 1: e68634. https://doi.org/10.3897/ab.e68634

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13: 581–583. https://doi.org/10.1038/nmeth.3869

Callahan, B.J., Grinevich, D, Thakur, S., Balamotis, M.A., Yehezkel, T.B. 2021. Ultra-accurate microbial amplicon sequencing with synthetic long reads. Microbiome 9: 130. https://doi.org/10.1186/s40168-021-01072-3

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., et al. 2012. Biodiversity loss and its impact on humanity. Nature 486: 59–67. https://doi.org/10.1038/nature11148

Celepli, N., Sundh, J., Ekman, M., Dupont, C.L., Yooseph, S., Bergman, B., Ininbergs, K. 2017. Meta-omic analyses of Baltic Sea cyanobacteria: diversity, community structure and salt acclimation. Environ. Microbiol. 19: 673–686. https://doi.org/10.1111/1462-2920.13592

CEN 2015. DIN EN 16695 Water quality – guidance on the estimation of phytoplankton biovolume: English version EN 16695:2015.

Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A., Chariton, A., Creer, S., Frühe, L., Keck, F., et al. 2021. Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Mol. Ecol. 30: 2937–2958. https://doi.org/10.1111/mec.15472

Czech, L., Stamatakis, A., Dunthorn, M., Barbera, P. 2022. Metagenomic analysis using phylogenetic placement—A review of the first decade. Front. Bioinform. 2: 871393. https://doi.org/10.3389/fbinf.2022.871393

Djurhuus, A., Port, J., Closek, C.J., Yamahara, K.M., Romero-Maraccini, O., Walz, K.R., Goldsmith, D.B., Michisaki, R., Breitbart, M., Boehm, A.B., et al. 2017. Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. Front. Mar. Sci. 4: 314. https://doi.org/10.3389/fmars.2017.00314

Eikrem, W., Throndsen, J. 1990. The ultrastructure of *Bathycoccus* gen. nov. and *B. prasinos* sp. nov., a non-motile picoplanktonic alga (Chlorophyta, Prasinophyceae) from the Mediterranean and Atlantic. Phycologia 29: 344–350. https://doi.org/10.2216/i0031-8884-29-3-344.1

Elersek, T. (ed.) 2021. Technical guidelines for eDNA monitoring in Alpine waters for stakeholders and end-users. National Institute of Biology. https://www.alpine-space.org/projects/eco-alpswater/en/project-results/publications

Eriksson, B.K., Hillebrand, H. 2020. Rapid reorganization of global biodiversity. Science 366: 308-309. https://doi.org/10.1126/science.aaz4520.

European Union 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Communities L 327, 22/12/2000. p. 1–73.

European Union 2008. Directive 2008/56/EC of the European Parliament and of the Council of June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Official Journal of the European Union L 164, 25.6.2008. p. 19–40.

European Union 2020. Document 52020DC0380. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. EU Biodiversity Strategy for 2030. Bringing nature back into our lives. COM/2020/380. Brussels, 20.5.2020.

Ewels, P., Magnusson, M., Lundin, S., Käller, M. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Filker, S., Kühner, S., Heckwolf, M., Dierking, J., Stoeck, T. 2019. A fundamental difference between macrobiota and microbial eukaryotes: Protistan plankton has a species maximum in the freshwater-marine transition zone of the Baltic Sea. Environ. Microbiol. 21: 603–617. https://doi.org/10.1111/1462-2920.14502

Gong, W., Marchetti, A. 2019. Estimation of 18S gene copy number in marine eukaryotic plankton using a next-generation sequencing approach. Front. Mar. Sci. 6: 219. https://www.frontiersin.org/articles/10.3389/fmars.2019.00219

Graco-Roza, C., Soininen, J., Corrêa, G., Pacheco, F.S., Miranda, M., Domingos, P., Marinho, M.M. 2021. Functional rather than taxonomic diversity reveals changes in the phytoplankton community of a large dammed river. Ecol. Indic. 121: 107048. https://doi.org/10.1016/j.ecolind.2020.107048

Grieb, A., Bowers, R.M., Oggerin, M., Goudeau, D., Lee, J., Malmstrom, R.R., Woyke, T., Fuchs, B.M. 2020. A pipeline for targeted metagenomics of environmental bacteria. Microbiome 8: 21. https://doi.org/10.1186/s40168-020-0790-7

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

Guiry, M.D., Guiry, G.M. 2022. AlgaeBase. - World-wide electronic publication, National University of Ireland, Galway. https://www.algaebase.org. Accessed 15.8.2022, 30.8.2022, 2.9.2022, 5.9.2022.

Hällfors, G. 2004. Checklist of Baltic Sea phytoplankton species (including some heterotrophic protistan groups). HELCOM Baltic Sea Environment Proceedings 95: 1–208. https://www.helcom.fi/wp-content/uploads/2019/10/BSEP95.pdf

Hällfors, H. 2013. Studies on dinoflagellates in the northern Baltic Sea. Ph.D. thesis, Faculty of Biological and Environmental Sciences, University of Helsinki. Walter and Andrée de Nottbeck Foundation Scientific Reports 39. 71 pp. + 4 papers. https://helda.helsinki.fi/handle/10138/38315

Hällfors, G., Hällfors, S. 1992. The Tvärminne collection of algal cultures. Tvärminne Studies 5: 15-17.

Hällfors, H., Backer, H., Leppänen, J.-M., Hällfors, S., Hällfors, G., Kuosa, H. 2013. The northern Baltic Sea phytoplankton communities in 1903–1911 and 1993–2005: A comparison of historical and modern species data. Hydrobiologia 707: 109–133. https://doi.org/10.1007/s10750-012-1414-4

Hanžek, N., Gligora Udovič, M., Kajan, K., Borics, G., Várbíró, G., Stoeck, T., Žutinić, P., Orlić, S., Stanković, I. 2021. Assessing ecological status in karstic lakes through the integration of phytoplankton functional groups, morphological approach and environmental DNA metabarcoding. Ecol. Indic. 131: 108166. https://doi.org/10.1016/j.ecolind.2021.108166

Hasle, G.R., Syvertsen E.E. 1997. Marine diatoms. In: Tomas, C.R. (ed.), Identifying marine phytoplankton, 5-385. Academic Press, Harcourt Brace & Company, San Diego, New York etc. 858 pp.

Head, M.J., Harland, R., Matthiessen, J. 2001. Cold marine indicators of the late Quaternary: the new dinoflagellate cyst genus *Islandinium* and related morphotypes. J. Quaternary Sci. 16: 621–636. https://doi.org/10.1002/jqs.657

HELCOM 2021. Guidelines for monitoring of phytoplankton species composition, abundance and biomass. Updated November 2021. 22 pages. Available online (25.8.2022): https://helcom.fi/wp-content/uploads/2020/01/HELCOM-Guidelines-for-monitoring-of-phytoplankton-species-composition-abundance-and-biomass.pdf

Herlemann, D., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., Andersson, A.F. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 5: 1571–1579. https://doi.org/10.1038/ismej.2011.41

Hu, Y.O.O., Karlson, B., Charvet, S., Andersson, A.F. 2016. Diversity of pico- to mesoplankton along the 2000 km salinity gradient of the Baltic Sea. Front. Microbiol. 7: 679. https://doi.org/10.3389/fmicb.2016.00679

Huo, S., Li, X., Xi, B., Zhang, H., Ma, C., He, Z. 2020. Combining morphological and metabarcoding approaches reveals the freshwater eukaryotic phytoplankton community. Environmental Sciences Europe 32: 37. https://doi.org/10.1186/s12302-020-00321-w

IPBES 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Brondizio, E.S., Settele, J., Díaz, S., Ngo, H.T. (eds.), 63. IPBES Secretariat, Bonn, Germany. 1148 pages. https://doi.org/10.5281/zenodo.3831673

Jaanus, A. 2011. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Ph.D. thesis, Dissertationes Biologicae Universitatis Tartuensis 198. 56 pp. + 5 papers.

Karlson, B., Mohlin, M., Hu, Y.O.O., Andersson, A.F. 2018. Miljöövervakning av växtplankton i Kattegatt och Östersjön med rDNA-barcoding och mikroskopi. Havs-och vattenmyndighetens rapport 2018: 22. 63 pp. https://www.havochvatten.se/data-kartor-och-rapporter/rapporter-och-andra-publikationer/publikationer/2018-07-16-miljoovervakning-av-vaxtplankton-i-kattegatt-och-ostersjon-med-rdna-barcoding-och-mikroskopi.html

Keck, F., Blackman, R.C., Bossart, R., Brantschen, J., Couton, M., Hürlemann, S., Kirschner, D., Locher, N., Zhang, H., Altermatt, F. 2022. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. Mol. Ecol. 31: 1820–1835. https://doi.org/10.1111/mec.16364

Kim, D.-K., Park, K., Jo, H., Kwak, I.-S. 2019. Comparison of water sampling between environmental DNA metabarcoding and conventional microscopic identification: A case study in Gwangyang Bay, South Korea. Applied Sciences 9: 3272. https://doi.org/10.3390/app9163272

Komárek, J., Anagnostidis, K. 1999. Cyanoprocaryota 1. Teil: Chroococcales. −Süßwasserflora von Mitteleuropa 19/1:I−VI, 1−548. [In English]

Kooistra, W.H.C.F., Sarno, S., Balzano, S., Haifeng, H., Andersen, R.A., Zingone, A. 2008. Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). Protist 159: 177–193. https://doi.org/10.1016/j.protis.2007.09.004

Kuosa, H. 1991. Picoplanktonic algae in the northern Baltic Sea: seasonal dynamics and flagellate grazing. Mar. Ecol. Prog. Ser. 73: 269–276.

Kuosa, H., Fleming-Lehtinen, V., Lehtinen, S., Lehtiniemi, M., Nygård, H., Raateoja, M., Raitaniemi, J., Tuimala, J., Uusitalo, L., Suikkanen, S. 2017. A retrospective view of the development of the Gulf of Bothnia ecosystem. J. Mar. Syst. 167: 78–92. https://doi.org/10.1016/j.jmarsys.2016.11.020

Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G., Taylor, M.I. 2019. How quantitative is metabarcoding: A meta-analytical approach. Mol. Ecol. 28: 420–430. https://doi.org/10.1111/mec.14920

Lamy, P., Citores, A., Deidun, A., Evans, L., Galgani, F., Heffernan, P., Karageorgis, A., Kauppi, L., Manakovski, D., Meissner, G., et al. 2020. European Commission Independent Expert Report. Proposed mission, Mission Starfish 2030, Restore our Ocean and Waters. Report of the Mission Board Healthy Oceans, Seas, Coastal and Inland waters.

Latz, M.A.C., Grujcic, V., Brugel, S., Lycken, J., John, U., Karlson, B., Andersson, A., Andersson, A.F. 2022. Short- and long-read metabarcoding of the eukaryotic rRNA operon: Evaluation of primers and comparison to shotgun metagenomics sequencing. Mol. Ecol. Resour. 22: 2304–2318. https://doi.org/10.1111/1755-0998.13623

Laza-Martinez, A., Arluzea, J., Miguel I., Orive, E. 2012: Morphological and molecular characterization of *Teleaulax gracilis* sp. nov. and *T. minuta* sp. nov. (Cryptophyceae). Phycologia 51: 649-661. https://doi.org/10.2216/11-044.1

Leese, F. 2022. DNAqua-Net. http://dnaqua.net/about/. Accessed 8 September 2022.

Lehtinen, S., Suikkanen, S., Hällfors, H., Kauppila, P., Lehtiniemi, M., Tuimala, J., Uusitalo, L., Kuosa, H. 2016. Approach for supporting food web assessments with multi-decadal phytoplankton community analyses — case Baltic Sea. Front. Mar. Sci. 3: 220. https://doi.org/10.3389/fmars.2016.00220

Leruste, A., Villéger, S., Malet, N., DeWit, R., Bec, B. 2018. Complementarity of the multidimensional functional and the taxonomic approaches to study phytoplankton communities in three Mediterranean coastal lagoons of different trophic status. Hydrobiologia 815: 207–227. https://doi.org/10.1007/s10750-018-3565-4

Lindh, M.V., Sjöstedt, J., Andersson, A.F., Baltar, F., Hugerth, L.W., Lundin, D., Muthusamy, S., Legrand, C., Pinhassi, J. 2015. Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. Environ. Microbiol. 17: 2459–2476. https://doi.org/10.1111/1462-2920.12720

Lipsewers, T. 2020. Spring bloom dynamics in the Baltic Sea: From the environment to macroelements and microbial interactions. Ph.D. thesis, Faculty of Biological and Environmental Sciences, University of Helsinki. 68 pp. + 3 papers. https://helda.helsinki.fi/handle/10138/316698

Litchman, E., Klausmeier, C.A. 2008. Trait-based community ecology of phytoplankton. Annu. Rev. Ecol. Evol. Syst. 39: 615–639. https://doi.org/10.1146/annurev.ecolsys.39.110707.173549

Majaneva, M., Rintala, J.-M., Piisilä, M., Fewer, D.P. Blomster, J. 2012. Comparison of wintertime eukaryotic community from sea ice and open water in the Baltic Sea, based on sequencing of the 18S rRNA gene. Polar Biol. 35: 875-889. https://doi.org/10.1007/s00300-011-1132-9

Majaneva, M., Diserud, O.H., Eagle, S.H.C., Boström, E., Hajibabaei, M., Ekrem, T. 2018. Environmental DNA filtration techniques affect recovered biodiversity. Sci. Rep. 8: 4682. https://doi.org/10.1038/s41598-018-23052-8

Mäki, A., Salmi, P., Mikkonen, A., Kremp, A., Tiirola, M. 2017. Sample preservation, DNA or RNA extraction and data analysis for high-throughput phytoplankton community sequencing. Front. Microbiol. 8: 1848. https://doi.org/10.3389/fmicb.2017.01848

Malashenkov, D.V., Dashkova, V., Zhakupova, K., Vorobjev, I.A., Barteneva, N.S. 2021. Comparative analysis of freshwater phytoplankton communities in two lakes of Burabay National Park using morphological and molecular approaches. Sci. Rep. 11: 16130. https://doi.org/10.1038/s41598-021-95223-z

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17: 10–12. https://doi.org/10.14806/ej.17.1.200

Martin, J.L., Santi, I., Pitta, P., John, U., Gypens, N. 2022. Towards quantitative metabarcoding of eukaryotic plankton: An approach to improve 18S rRNA gene copy number bias. Metabarcoding and Metagenomics 6: e85794. https://doi.org/10.3897/mbmg.6.85794

Montresor, M., Procaccini, G. Stoecker, D. 1999. *Polarella glacialis*, gen. nov., sp. nov. (Dinophyceae): Suessiaceae are still alive! - J. Phycol. 35: 186-197. https://doi.org/10.1046/j.1529-8817.1999.3510186.x

Montresor, M., Lovejoy, C., Orsini, L., Procaccini, G. Roy, S. 2003. Bipolar distribution of the cyst-forming dinoflagellate *Polarella glacialis*. Polar Biol. 26: 186–194. https://doi.org/10.1007/s00300-002-0473-9

Needham, D.A., Fuhrman, J.A. 2016. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat. Microbiol. 1: 16005. https://doi.org/10.1038/nmicrobiol.2016.5

Nowinski, B., Smith, C.B., Thomas, C.M., Esson, K., Marin, R. III., Preston, C.M., Birch, J.M., Scholin, C.M., Huntemann, M., Clum, A., et al. 2019. Microbial metagenomes and metatranscriptomes during a coastal phytoplankton bloom. Sci. Data 6: 129. https://doi.org/10.1038/s41597-019-0132-4

Norros, V., Laamanen, T., Meissner, K., Iso-Touru, T., Kahilainen, A., Lehtinen, S., Lohtander-Buckbee, K., Nygård, H., Pennanen, T., Ruohonen-Lehto, M., et al. 2022. Roadmap for implementing environmental DNA (eDNA) and other molecular monitoring methods in Finland – Vision and action plan for 2022–2025. Reports of the Finnish Environment Institute 20/2022. 71 pp. https://helda.helsinki.fi/handle/10138/342992

Obiol, A., Giner, C.R., Sánchez, P., Duarte, C.M., Acinas, S.G., Massana, R. 2020. A metagenomic assessment of microbial eukaryotic diversity in the global ocean. Mol. Ecol. Resources 20: 718–731. https://doi.org/10.1111/1755-0998.13147

Olenina, I., Hajdu, S., Edler, L., Andersson, A., Wasmund, N., Busch, S., Göbel, J., Gromisz, S., Huseby, S., Huttunen, M. et al. 2006. Biovolumes and size-classes of phytoplankton in the Baltic Sea. HELCOM Baltic Sea Environment Proceedings 106: 1-142. The annually updated annex is available at DOME (Marine Environment) (ices.dk) (choose DOWNLOAD PEG BIOVOLUME at the right side of the page).

Parada, A.E., Needham, D.M., Fuhrman, J.A. 2016. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol. 18: 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pawlowski, J., Apothéloz-Perret-Gentil, L., Mächler, E., Altermatt, F. 2020. Environmental DNA applications in biomonitoring and bioassessment of aquatic ecosystems. Guidelines. Federal Office for the Environment, Bern. Environmental Studies. no. 2010: 71 pp. https://doi.org/10.5167/uzh-187800

Pereira, C.L., Gilbert, M.T.P., Araújo, M.B., Matias, M.G. 2021. Fine-tuning biodiversity assessments: A framework to pair eDNA metabarcoding and morphological approaches. Methods Ecol. Evol 12: 2397–2409. https://doi.org/10.1111/2041-210X.13718

Pierella Karlusich, J.J., Pelletier, E., Zinger, L., Lombard, F., Zingone, A., Colin, S., Gasol, J.M., Dorrell, R.G., Henry, N., Scalco, E., et al. 2022. A robust approach to estimate relative phytoplankton cell abundances from metagenomes. Mol. Ecol. Resources (in press). https://doi.org/10.1111/1755-0998.13592

Pitkänen, H., Raateoja, M., Kankaanpää, P., Uusitalo, L., Heiskanen, A.-S., Kettunen, J., Kankaanpää, H., Korpinen, S. (eds.) 2020. Meriseurannan tiekartta – SYKEn ylläpitämien ja koordinoimien meren tilaseurantojen nykytila ja kehittäminen (Roadmap for marine monitoring – current state and development of marine monitoring maintained and coordinated by SYKE). In Finnish. Suomen ympäristökeskuksen raportteja 26/2020. https://helda.helsinki.fi/handle/10138/320467

Pospelova, V. Head, M.J. 2002. *Islandinium brevispinosum* sp. nov. (Dinoflagellata), a new organic-walled dinoflagellate cyst from modern estuarine sediments of New England (USA). J. Phycol. 38: 593–601. https://doi.org/10.1046/j.1529-8817.2002.01206.x

Potvin, E., Rochon, A., Lovejoy, C. 2013. Cyst–theca relationship of the arctic dinoflagellate cyst *Islandinium minutum* (Dinophyceae) and phylogenetic position based on SSU rDNA and LSU rDNA. J. Phycol. 49: 848–866. https://doi.org/10.1111/jpy.12089

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Research, Volume 41, Issue D1, 1, Pages D590–D596, https://doi.org/10.1093/nar/gks1219

R Core Team 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rantajärvi, E., Pitkänen, H., Korpinen, S., Nurmi, M., Ekebom, J., Liljaniemi, P., Cederberg, T., Suomela, J., Paavilainen, P. Lahtinen, T. (eds.) 2020. Seurantakäsikirja Suomen merenhoitosuunnitelman seurantaohjelmaan vuosille 2020–2026 (Manual for marine monitoring in Finland 2020–2026). In Finnish. Suomen ympäristökeskuksen raportteja 47/2020. 219 pp. https://helda.helsinki.fi/handle/10138/323600.

Ratnasingham, S., Hebert, P.D.N. 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Mol. Ecol. Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F. 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ 4: e2584. https://doi.org/10.7717/peerj.2584

Ruppert, K.M., Kline, R.J., Rahman, M.S. 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecol. Conservation 17: e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Salazar, V.W., Tschoeke, D.A., Swings, J., Cosenza, C.A., Mattoso, M., Thompson, C.C., Thompson, F.L. 2020. A new genomic taxonomy system for the Synechococcus collective. Environ. Microbiol. 22: 4557–4570. https://doi.org/10.1111/1462-2920.15173

Salmaso, N., Boscaini, A., Pindo, M. 2020. Unraveling the diversity of eukaryotic microplankton in a large and deep perialpine lake using a high throughput sequencing approach. Front. Microbiol. 11: 789. https://www.frontiersin.org/articles/10.3389/fmicb.2020.00789

Salmaso, N., Riccioni, G., Pindo, M., Kurmayer, R., Vasselon, V., Domaizon, I. 2021a. Metabarcoding protocol – Analysis of protists using the 18S rRNA gene and a DADA2 pipeline. Protocol prepared as part of the Interreg Alpine Space project Eco-AlpsWater (ASP569) - Innovative Ecological Assessment and Water Management Strategy for the Protection of Ecosystem Services in Alpine Lakes and Rivers. https://doi.org/10.5281/zenodo.5233527

Salmaso, N., Riccioni, G., Pindo, M., Vasselon, V., Domaizon, I., Kurmayer, R. 2021b. Metabarcoding protocol – Analysis of Bacteria (including Cyanobacteria) using the 16S rRNA gene and a DADA2 pipeline. Protocol prepared as part of the Interreg Alpine Space project Eco-AlpsWater (ASP569) - Innovative Ecological Assessment and Water Management Strategy for the Protection of Ecosystem Services in Alpine Lakes and Rivers. https://doi.org/10.5281/zenodo.5232772

Santoferrara, L.F. 2019. Current practice in plankton metabarcoding: Optimization and error management. J. Plankton Res. 41: 571–582. https://doi.org/10.1093/plankt/fbz041

Schaumburg, J., Mischke, U., Salmoso, N., Rund, H., Capelli, C., Wanzenpöck, J., Domaizon, I. 2021. Recommendations for the inclusion of innovative monitoring approaches in water quality assessment and management (WFD/WPO) (Work Package WPT4; Eco-AlpsWater). https://www.alpine-space.org/projects/eco-alpswater/deliverables-final/d-t.4.2.2_recommendations-inclusion-in-management_wfd_wpo_14102021.pdf

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. 2009. Introducing mothur: Open-Source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75: 7537–7541. https://doi.org/10.1128/AEM.01541-09

Seenivasan, R., Sausen, N., Medlin, L.K., Melkonian, M. 2013. *Picomonas judraskeda* gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as 'picobiliphytes'. PLOS ONE. 8: e59565. https://doi.org/10.1371/journal.pone.0059565

Shalchian-Tabrizi, K., Eikrem, W., Klaveness, D., Vaulot, D., Minge, M.A., Le Gall, F., Romari, K., Throndsen, J., Botnen, A., Massana, R., et al. 2006: Telonemia, a new protist phylum with affinity to chromist lineages. Proc. R. Soc. B 273: 1833–1842. https://doi.org/10.1098/rspb.2006.3515

Sildever, S., Kawakami, Y., Kanno, N., Kasai, H., Shiomoto, A., Katakura, S., Nagai, S. 2019. Toxic HAB species from the Sea of Okhotsk detected by a metagenetic approach, seasonality and environmental drivers. Harmful Algae 87: 101631. https://doi.org/10.1016/j.hal.2019.101631

Somervuo, P., Yu, C.W., Xu, C.C.Y., Ji, Y., Hultman, J., Wirta, H., Ovaskainen, O. 2017. Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. Methods Ecol. Evol. 8: 398–407. https://doi.org/10.1111/2041-210X.12721

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., Richards, T.A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. 19: 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Stoeckle, B.C., Beggel, S., Cerwenka, A.F., Motivans, E., Kuehn, R., Geist, J. 2017. A systematic approach to evaluate the influence of environmental conditions on eDNA detection success in aquatic ecosystems. PLOS ONE 12: e0189119. https://doi.org/10.1371/journal.pone.0189119

Subirana, L., Péquin, B., Michely, S., Escande, M.-L., Meilland, J., Derelle, E., Marin, B., Piganeau, G., Desdevises, Y., Moreau, H., Grimsley, N.H. 2013. Morphology, genome plasticity and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). Protist 164: 643–659. https://doi.org/10.1016/j.protis.2013.06.002

Suikkanen, S., Pulina, S., Engström-Öst, J., Lehtiniemi, M., Lehtinen, S., Brutemark, A. 2013. Climate change and eutrophication induced shifts in northern summer plankton communities. PLOS ONE 8: e66475. https://doi.org/10.1371/journal.pone.0066475

Sundström, A.M., Kremp, A., Tammilehto, A., Tuimala, J., Larsson, U. 2010. Detection of the bloom-forming cold-water dinoflagellate *Biecheleria baltica* in the Baltic Sea using LSU rRNA probes. Aquat. Microb. Ecol. 61: 129–140. https://doi.org/10.3354/ame01442

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21: 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taberlet, P., Bonin, A., Zinger, L., Coissac, E. 2018. Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Tamm, M., Laas, P., Freiberg, R., Nõges, P., Nõges, T. 2018. Parallel assessment of marine autotrophic picoplankton using flow cytometry and chemotaxonomy. Sci. Total Environ. 625: 185–193. https://doi.org/10.1016/j.scitotenv.2017.12.234

Utermöhl, H. 1958. Zur Vervollkommung der quantitativen phytoplankton-methodik. Mitt. Int. Ver. Limnol. 9: 38.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., et al. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science 348: 1261605. https://doi.org/10.1126/science.1261605

Vaulot, D., Sim, C.W.H., Ong, D., Teo, B., Biwer, C., Jamy, M., Lopes dos Santos, A. 2022. metaPR^2: A database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists. Mol. Ecol. Resources (in press). https://doi.org/10.1111/1755-0998.13674

Wasmund, N., Göbel, J., v. Bodungen, B. 2008. Changes in the phytoplankton community of Kiel Bight (Baltic Sea) during the last century. J. Mar. Syst. 73: 300–322. https://doi.org/10.1016/j.jmarsys.2006.09.009

WoRMS Editorial Board 2022. World Register of Marine Species. Available from https://www.marinespecies.org at VLIZ. Accessed 15.8.2022, 30.8.2022, 2.9.2022, 5.9.2022.

Yarimizu, K., Fujiyoshi, S., Kawai, M., Norambuena-Subiabre, L., Cascales, E.-K., Rilling, J.-I., Vilugrón, J., Cameron, H., Vergara, K., Morón-López, J., et al. 2020. Protocols for monitoring harmful algal blooms for sustainable aquaculture and coastal fisheries in Chile. Int. J. Env. Res. Pub. He. 17: 7642. https://doi.org/10.3390/ijerph17207642

Yarimizu, K., Sildever, S., Hamamoto, Y., Tazawa, S., Oikawa, H., Yamaguchi, H., Basti, L., Mardones, J.I., Paredes-Mella, J., & Nagai, S. 2021. Development of an absolute quantification method for ribosomal RNA gene copy numbers per eukaryotic single cell by digital PCR. Harmful Algae 103: 102008. https://doi.org/10.1016/j.hal.2021.102008

Yeh, Y.-C., McNichol, J., Needham, D.M., Fichot, E.B., Berdjeb, L., Fuhrman, J.A. 2021. Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities, and estimation of sequencing bias against 18S. Environ. Microbiol. 23: 3240–3250. https://doi.org/10.1111/1462-2920.15553

Young, J., Geisen, M., Cros, L., Kleijne, A., Sprengel, C., Probert, I., Østergaard, J. 2003. A guide to extant coccolithophore taxonomy. Journal of Nannoplankton Research Special Issue 1. 125 pp.

Appendix

Pilot study

This appendix describes the pilot study on which the protocol and instructions in these guidelines are based. During the two-year pilot project, the integration of eDNA metabarcoding in Finnish marine phytoplankton monitoring was tested and the additional information gained by eDNA metabarcoding assessed. The outcome was used to decide on the materials and methods suggested in the provided protocol (**Section 3**).

A 1 Sampling and sample processing

The integration of the eDNA metabarcoding technique in the Finnish marine phytoplankton monitoring was tested utilizing monitoring samples taken during the HELCOM COMBINE cruise in April 2021 on research vessel (R/V) Aranda. In total, seven stations were sampled: two stations in the Bothnian Sea (US5B and MS9), one station in the Åland Sea (F64), one station in the northern Baltic Proper (LL17), two stations in the western Gulf of Finland (Längden (LAN) and LL7), and one station in the eastern Gulf of Finland (LL3A) (**Figure A1**). The study utilized SYKE's marine research infrastructure, which is a part of the national FINMARI RI consortium.

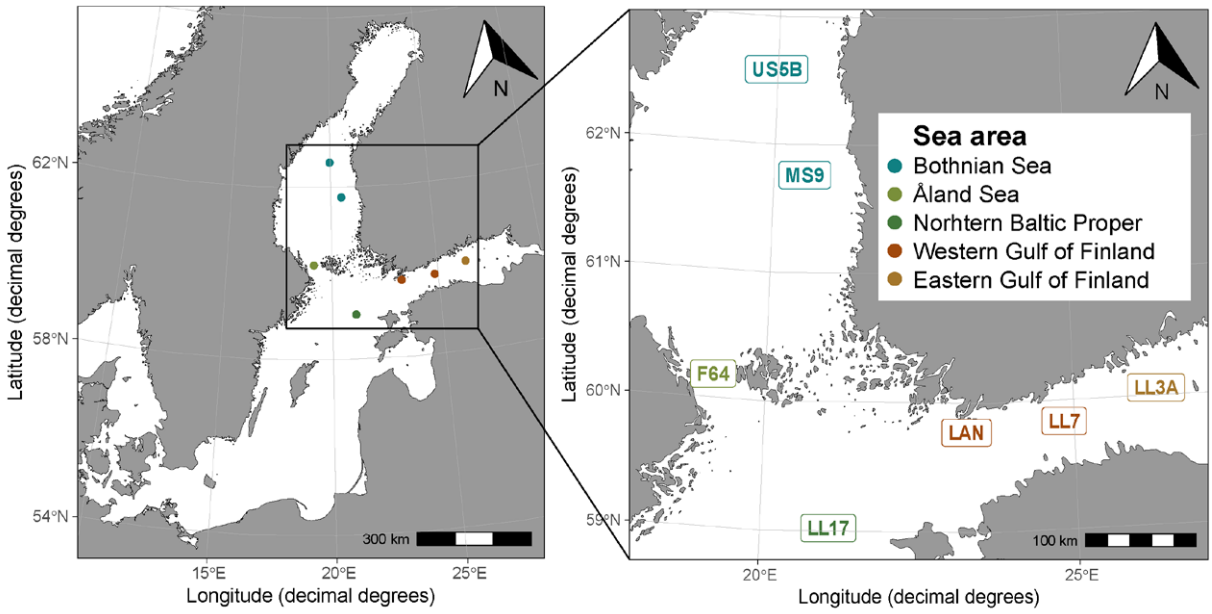


Figure A1. Maps showing the location of the sampling stations in the Baltic Sea.

Samples for both phytoplankton microscopy and eDNA analyses were taken as integrated water samples (0-10 m depth) according to the monitoring guidelines (HELCOM, 2021), with tube samplers (Niskin or HydroBios). Samples were processed as described in **Section 3.2**. After preparing the integrated sample, three main processing options were compared for the eDNA samples (**Figure A2**):

a) filtration of the untreated sample with a disposable filtration unit (Steritop, Merck) with 0.2 µm pore size polyethersulfone (PS) membranes to reduce the risk of contamination,
b) filtration of the untreated sample using reusable filtration units (Nalgene) with sterile 0.2 µm MCE filters and
c) preservation with acidic Lugol's solution (0.5%), followed by filtration with the reusable filtration units.

In addition, three different volumes (300, 500 and 800 mL) were compared in option **b)** to assess the effect of the filtration volume on the DNA concentration. DNA extraction and sequencing were carried out as described in **Section 3.2** to facilitate the processing of as many samples as possible, i.e. also those with low DNA yields.

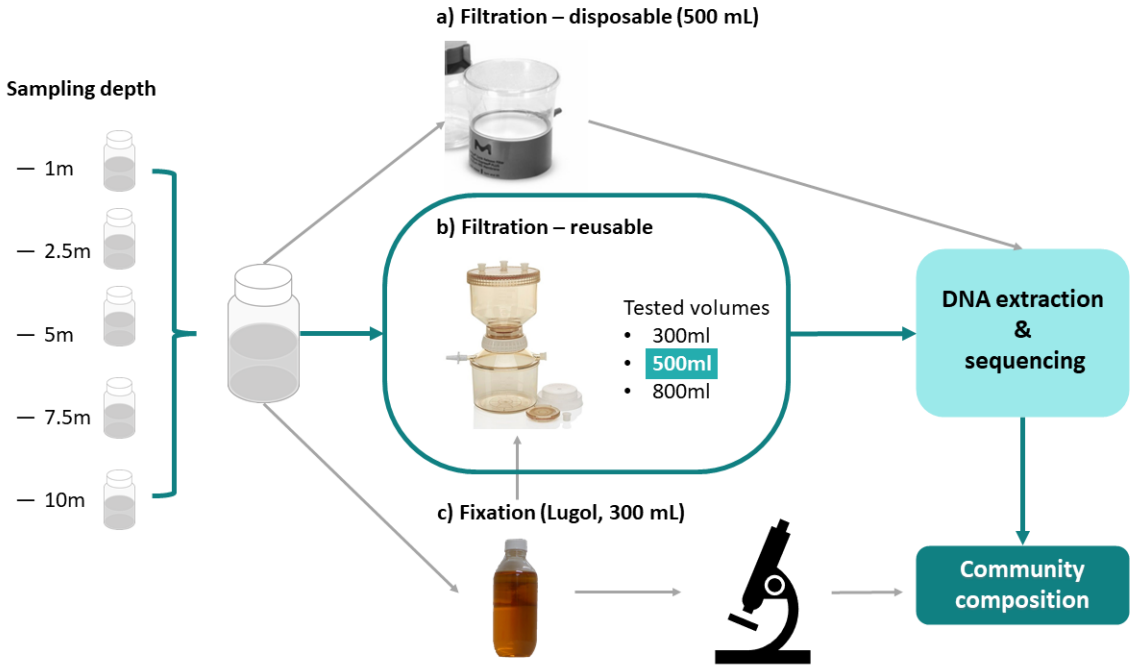


Figure A2. Schematic overview of the sample processing. Turquoise arrows and boxes highlight the recommended approach for monitoring. For option a) the filter needs to be detached from the disposable funnel with a disposable scalpel.

A 2 Positive and negative controls

Positive and negative controls were used in the pilot study as described in **Section 3.4** of the guidelines.

A 2.1 Negative control preparation

Preparation of the negative controls during the sampling cruise is described in the protocol (**Section 3.4**). Although the amount of DNA extracted from the filtration blanks and the kit controls was below the detection limit both negative controls were sequenced together with all other samples.

A 2.2 Positive control preparations

For positive controls, both a single-species culture and an artificial phytoplankton community of known composition (mock community) were used. As the single-species positive control, a clonal culture of the potentially harmful bloom-forming dinoflagellate species *Alexandrium ostenfeldii* was used (strain AOB203, isolated in Föglö, Åland, in 2015, by Jacqueline Jerney). Clonal cultures for the mock culture were obtained from the FINMARI CC (**Table A1**). The cell number in each clonal culture was roughly assessed by light microscopy using a 1 mL Sedgewick Rafter counting chamber, and required dilutions were made before preparing the final single-species culture and the final mock culture. At least 400 cells were counted for each taxon when assessing the cell number.

The mock culture was prepared by selecting several species that are expected to occur in the field sample, and which have been sequenced and the sequences deposited in the Barcode of Life Data

System (BOLD, Ratnasingham & Hebert, 2007). Cultures were added either in low (~10^3 cells L^{-1}) or high (~10^6 cells L^{-1}) concentration to one litre of filtered seawater (FSW, salinity of 6) to get an estimate for the sensitivity of the tested protocol. Afterwards, 300 mL of the untreated and acidic Lugol's preserved mock sample were filtered with MCE filters and processed as all the other samples for sequencing.

Table A 1. FINMARI CC phytoplankton strains, which were used for the positive control mock community. The strain isolation sites are in the western Gulf of Finland, unless otherwise indicated.

FINMARI CC ID	Taxon	Isolation site	Isolation year	Isolated by
CWTV-1401	*Chaetoceros wighamii*	Tvärminne pier	2014	P. Hakanen
PCTV0906	*Peridiniella catenata*	Tvärminne Storfjärden	2009	A. Kremp
C14 07	*Skeletonema marinoi*	Bothnian Sea	2011	C. Sjöqvist
TBLL7-1302	*Thalassiosira baltica*	Station LL7	2013	A. Kremp
TV335	*Phaeodactylum tricornutum*	Tvärminne rockpool	≤1992	G. Hällfors & S. Hällfors (1)
LL7 SH1-01	*Apocalathium malmogiense*	Station LL7	2015	A. Kremp
GCTV-C1	*Gymnodinium corollarium*	Station BY29, Northern Baltic Proper	2005	A. Kremp
BBTV-1401	*Biecheleria baltica*	Tvärminne pier	2014	P. Hakanen
Crypto08-B1	*Rhodomonas marina*	Tvärminne Storfjärden	2008	A. Kremp
HATV-1401	*Heterocapsa arctica subsp. frigida*	Tvärminne Storfjärden	2014	P. Hakanen
DTTV-1401	*Diatoma tenuis*	Tvärminne Storfjärden	2014	P. Hakanen

(1) Hällfors, G., Hällfors, S. 1992, The Tvärminne collection of algal cultures. Tvärminne Studies 5: 15-17.

A 3 Microscopical analyses

A 3.1 Monitoring samples

The light microscopy analysis of the phytoplankton monitoring samples was carried out according to the HELCOM monitoring guidelines (HELCOM, 2021). 10, 25 or 50 mL of the integrated 0-10 m water sample was settled, depending on cell density. For details regarding microscopy, see e.g. Lehtinen et al. (2016). Photoautotrophic and mixotrophic planktonic microalgae, cyanobacteria, the mixotrophic ciliate *Mesodinium rubrum*, and heterotrophic dinoflagellates as well as other heterotrophic flagellates were enumerated. Single-celled <2 µm-sized picoplankton was not included, since the method (light microscopy of Lugol's-preserved samples) is not suitable for separating single-celled photoautotrophic picoplankton from single-celled heterotrophic bacteria, and the settling of such small particles in the settling chambers is uncertain. The observed taxa were determined to the most detailed possible taxonomical level, and the specimens were furthermore categorized into size classes (Olenina et al., 2006, and its annually updated annex). All specimens cannot be identified to species level. For these, the size classes enable further information separating specimens into different groups based on their size, shape, and trophy. For all specimens, including taxa that are identified to species level, information on size classes facilitates as accurate biovolume determination as possible.

A 3.2 Microscopy of positive controls

The light microscopical analysis of the mock community (positive control) was carried out according to the HELCOM monitoring guidelines (HELCOM, 2021). For the mock community also epifluorescence microscopy was used, since three important cold-water dinoflagellates, namely *Apocalathium malmogiense, Biecheleria baltica*, and *Gymnodinium corollarium*, cannot reliably be identified to species level due to their high morphological similarity in Lugol's-preserved samples when observed with light microscopy. In phytoplankton monitoring, these three taxa are usually enumerated together as the *Apocalathium* complex (syn. *Scrippsiella/Biecheleria/Gymnodinium* complex *sensu* Sundström et al., 2010; also termed *Scrippsiella* complex; Jaanus 2011), but the three species can be separated and quantified by inverted epifluorescence microscopy. To quantify the contribution of the three taxa, 10 mL of the mock community was preserved with neutral Lugol's solution 0.5% (v/v) and concentrated for 48h according to Utermöhl (1958). Cells were stained with freshly prepared 0.1% solution of Calcofluor white MR2 (Fluorescence Brightener 28, Sigma-Aldrich) as described by Lipsewers (2020) and enumerated with an inverted epifluorescence microscope (Leica DMI 3000 B, Leica DFC 490 camera) at 800x magnification. In total, 670 cells were enumerated.

A 4 Sequencing and data analysis

Primers were chosen as described in **Section 3.5** and PCR tests for a subset of samples were carried out before the actual library preparation and sequencing, to confirm successful primer amplification. Amplicon library preparation and sequencing on the Illumina MiSeq platform (300 bp paired-end reads, 100k sequencing depth), were carried out at the FIMM Technology Centre, University of Helsinki.

For eukaryotes, amplicon sequence variants (ASVs) were produced as described in **Section 3.6.4.** Briefly, the raw reads were trimmed with cutadapt (Martin, 2011) to find and remove sequencing primers (--p-front-f ^CCAGCASCYGCGGTAATTCC, --p-front-r ^ACTTTCGTTCTTGAT) and filter out reads that are shorter than 100bp (--p-minimum-length 100), requiring an overlap of 12 base pairs between forward and reverse read (--p-overlap 12) before denoising (--p-trunc-len-f 214, --p-trunc-len-r 182) and removal of chimeras. No abundance cutoffs were applied. Taxonomy classification was carried out using the VSEARCH consensus method with default settings (Rognes et al., 2016). The Protist Ribosomal Reference database (PR2 v.4.14.0) was used as a reference database (Guillou et al., 2013). Relevant in-house sequences from the FINMARI CC were used to extend the database to include Baltic Sea species not included in the standard release of the database. One sequence of each of the following taxa were added: *Heterocapsa triquetra*, *Peridiniella catenata*, *Apocalathium malmogiense*, *Biecheleria baltica* and *Gymnodinium corollarium.* Sequence analysis for cyanobacteria was performed as described in **Section 3.6.5**. Mitochondria and chloroplasts were removed from the data.

For comparing the effect of different treatments (i.e. Lugol preservation and filtering different volumes) only the phytoplankton groups Dinophyceae, Chlorophyta, Cryptophyta, Haptophyta and Ochrophyta were considered and filtered in QIIME2 based on the assigned taxonomy, as described in **Section 3.6.4**. Microsoft Excel and packages "reshape", "ggplot2") in R version 4.1.1 (R Core Team, 2021) were used for visualizing the data.

A 5 Results and discussion

A 5.1 DNA extraction yields

We compared the DNA yield obtained by using two different filtration systems / filters at one station (LL7) (**Figure A3**). For all tested volumes, the total DNA yield was higher when using MCE filters, compared to polyethersulfone (PS) filters. In addition, MCE filters are more practical, because they can be easily transferred from the filtration unit to a reaction vial for DNA extraction.

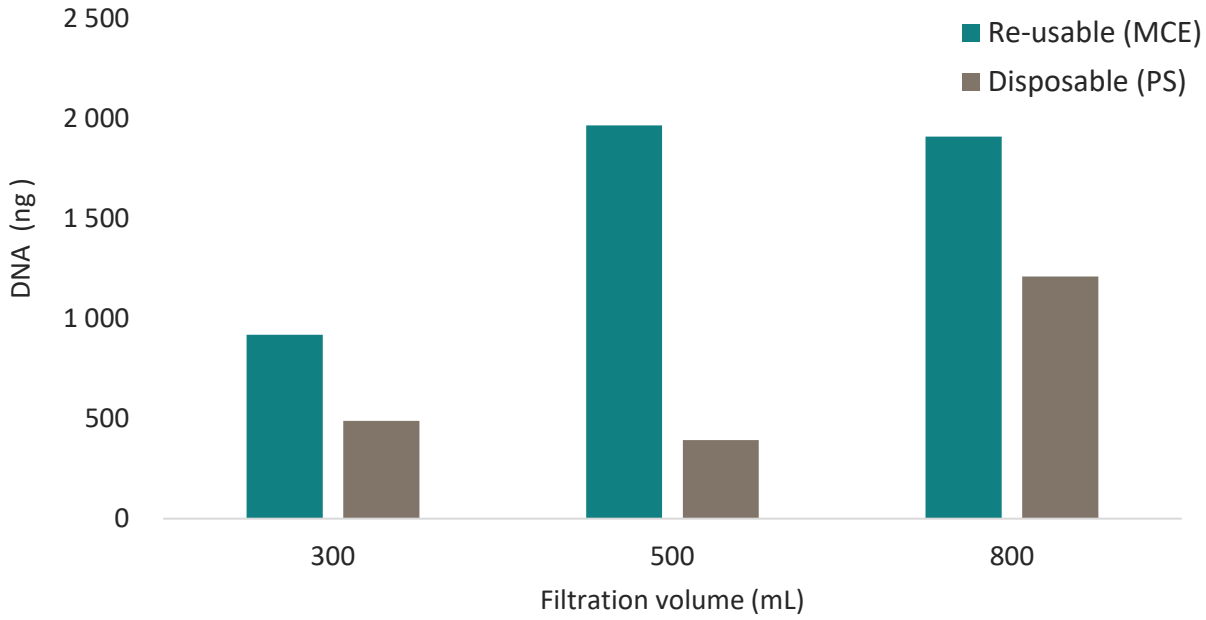


Figure A 3. DNA yield of different filtration systems and volumes compared for station LL7. MCE filters were used in reusable filtration units and PS filters are a part of disposable filtration units.

Lugol's preservation led to a slightly increased DNA yield for samples from some stations (LL7, LL17 and US5B) compared to untreated samples. But, due to a lack of replicates it is unclear if the effect of Lugol's preservation was significant. In contrast, preservation with Lugol's solution resulted in a clearly reduced DNA yield, compared to untreated samples from stations LL3A, LAN, and F64 (**Figure A4**).

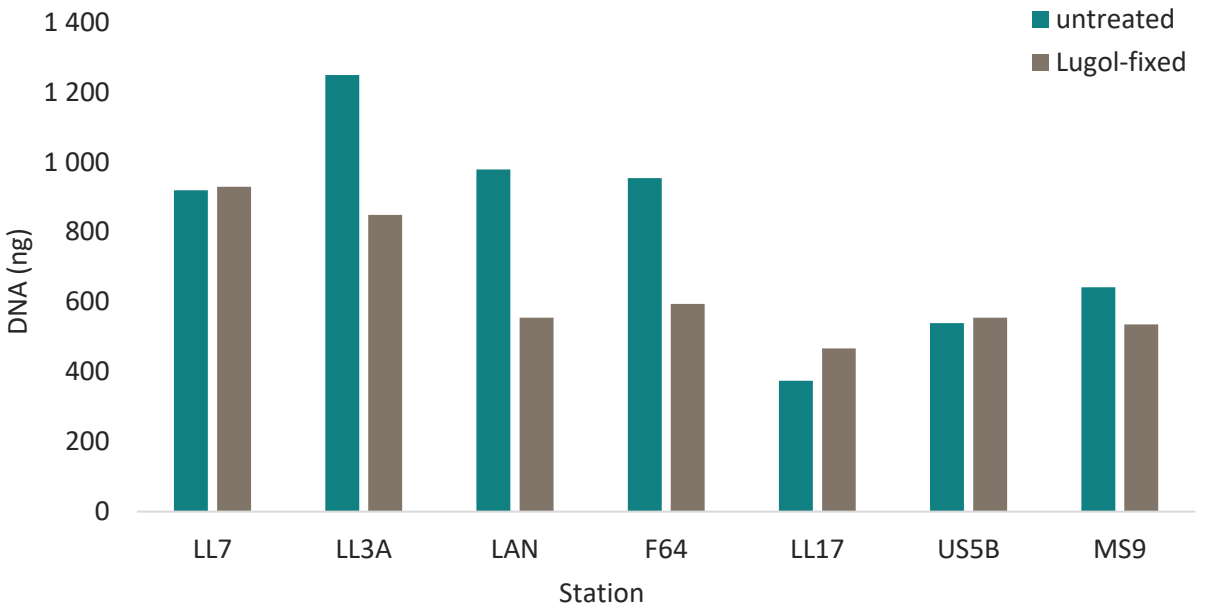


Figure A 4. DNA yield extracted from 300 mL of untreated or acidic Lugol's-preserved samples at all stations.

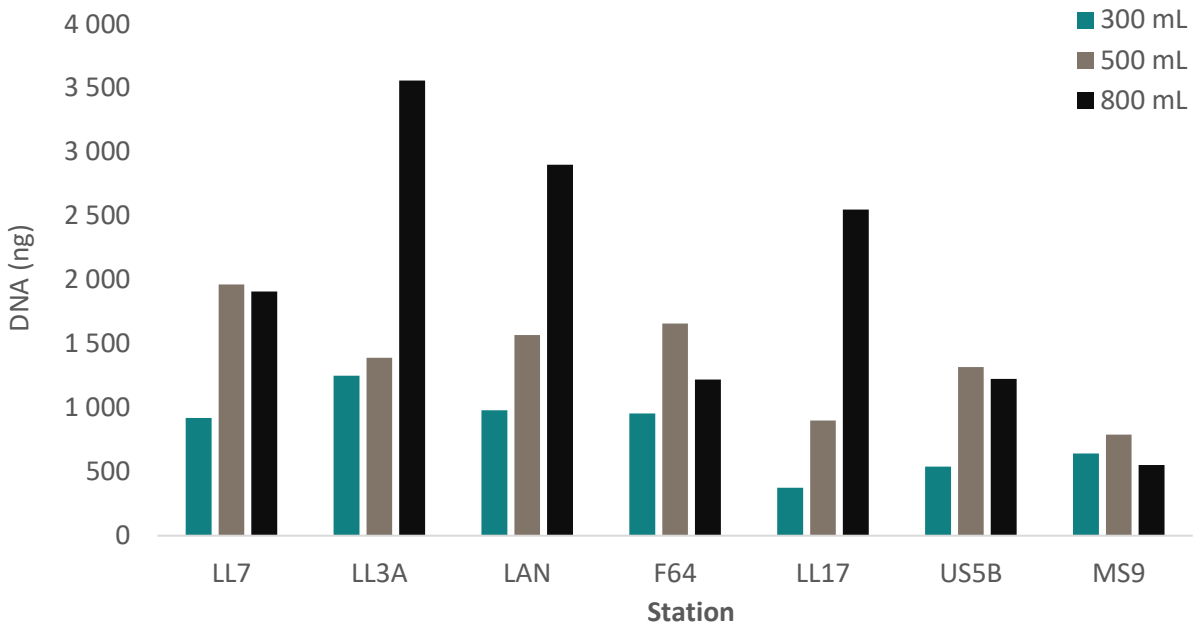


Figure A 5. DNA yields from different volumes of seawater, filtered with reusable filtration units and MCE filters.

When comparing different filtration volumes (300, 500 and 800 mL, **Figure A5**) we found no linear relationship between filtration volume and DNA yield. This could be explained by an increase of DNA inhibitors with larger filtration volumes. In the future, inhibitor testing should be carried out using biological and technical replicates.

A 5.2 Controls – sequencing and microscopy results

Sequencing of negative controls, i.e. filtration blanks and kit controls, resulted in less than 1000 reads, except for the acidic Lugol's preserved non-template control (300 mL), which had 1440 reads, when using the 18S V4 rRNA gene primer. After denoising, no ASVs remained in any of the negative controls, thus they were excluded from the downstream analysis. These results confirm that the suggested protocol sufficiently allows minimizing contamination of samples with eukaryotic sequences during the entire workflow.

The single-species positive control was successfully characterized as *Alexandrium ostenfeldii* (99.97%) and the remaining 0.03% of the ASVs were assigned to the class Dinophyceae. In the mock sample, only five of the taxa added with a high concentration (i.e. $>10^6$ cells L^{-1}) were detected by metabarcoding and the correct taxonomy was only assigned to two taxa at species level (*Apocalathium malmogiense* and *Skeletonema marinoi)*. Sparse taxa ($<10^6$ cells L^{-1}) were not detected. Taxonomic assignment of the mock community was clearly insufficient (see **Table A2**) and needs to be improved further in the future so that all taxa are assigned to species level.

Table A 2. Composition of the mock community, evaluated by light microscopy and eDNA metabarcoding (eDNA)

Taxon added	Abundance in mock	Light microscopy identification	eDNA identification	Light microscopy (cells L^{-1})	eDNA % contribution
Chaetoceros wighamii	high	*Chaetoceros wighamii* (also Centrales spp.?)	*Chaetoceros*	311 520	<1
Peridiniella catenata	high	*Peridiniella catenata*	Dinophyceae	1 446 552	30
Skeletonema marinoi	high	*Skeletonema marinoi* [1] (also Centrales spp.?)	*Skeletonema marinoi*	2 322 240	<1
Thalassiosira baltica	high	Centrales spp.		3 915 240	
Phaeodactylum tricornutum	high	*Phaeodactylum tricornutum*		1 840 540	
Apocalathium malmogiense	high	cf. *Apocalathium malmogiense*	*Apocalathium malmogiense*	1 774 696	68
Gymnodinium corollarium	high	*Apocalathium* complex	Gymnodiniales	753 344	1
Biecheleria baltica	low				
Rhodomonas marina	low	*Rhodomonas marina* [2]		14 158	
Heterocapsa arctica subsp. *frigida*	low	*Heterocapsa arctica* subsp. *frigida*		1 000	
Diatoma tenuis	low	*Diatoma tenuis*		3 552	

[1] The diatom formerly identified as *Skeletonema costatum* constitutes several species which cannot reliably be separated using light microscopy (see Kooistra et al., 2008 and references therein). Strains which subsequently have been investigated from the Baltic Sea have been identified as *S. marinoi* (Kooistra et al., 2008 and references therein; Wasmund et al., 2008), whereas investigating 18S rRNA gene diversity in the Baltic Sea, Majaneva et al. (2012) also found *S. grevillei.* In Baltic Sea phytoplankton monitoring, the *S. costatum* -like specimens are collectively determined as *S. marinoi.*

[2] In monitoring samples, *Rhodomonas marina* is not identifiable to species or even genus level; however, since the composition of the mock community was known, it was determined to species level.

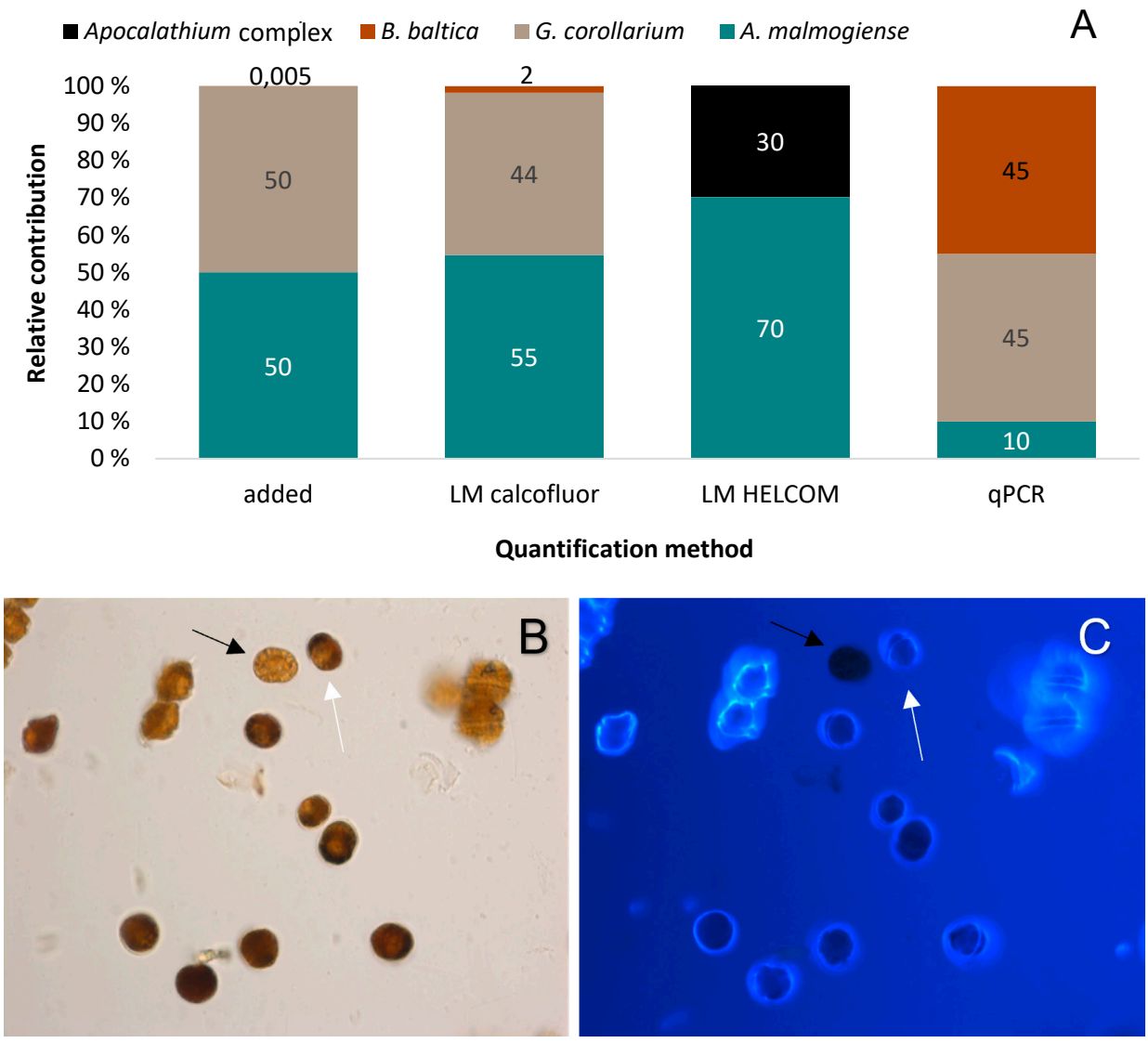


Figure A6. A) Evaluation of quantification methods used to distinguish the three taxa of the *Apocalathium* complex in the mock community. Relative contribution of single cultures before adding them to the mock community, based on light microscopy counted with the Sedgewick Rafter chamber (added); by standardized light microscopy with calcofluor stain (LM calcofluor) and without (LM HELCOM), the latter using an Utermöhl chamber. Typical composition of the three taxa during the spring bloom in the Baltic Sea (various sub-basins, years 2013-2015), based on qPCR (qPCR, Lipsewers et al., 2020). B) LM micrograph C) stained with calcofluor. In images B and C the white arrows point at *Apocalathium malmogiense* and the black arrows at *Gymnodinium corollarium*.

A 5.3 Sequencing results for photoautotrophic and mixotrophic microalgae

For the 18S rRNA gene V4 primer a total of 1,673,912 raw reads (avg. 45,241) were obtained for all samples (n = 37). Six samples (two non-template controls, one filtration blank, LL7-s8, US5-n3, MOC-n3) with less than 1500 sequences were removed from downstream processing, since no ASVs were assigned to them. On average, 43.56 % of the sequences per sample passed the denoising filter and 532 unique ASVs retained from 31 samples (including the mock and *A. ostenfeldii* culture), which had a total frequency of 470,379 ASVs. Taxonomic assignment of those ASVs resulted in 102 different taxa, including all treatments, as summarized in **Table S2**.

After filtering the dataset to include only the photoautotrophic and mixotrophic planktonic microalgae groups Dinophyceae, Chlorophyta, Cryptophyta, Haptophyta and Ochrophyta (including Bacillariophyceae) a total of 463,756 ASVs remained with an average length of 378.6 (SD 20.85) base pairs. Of those, 352 ASVs were unique.

The majority of ASVs were assigned to the class Dinophyceae at all stations, regardless of the filtration volume (**Figure A7**), indicating that there is a large undetected diversity within this group. The majority of ASVs assigned to Dinophyceae were not assigned to more detailed taxonomic levels, which is probably due to a lack of reference sequences but might be also related to artificial inflation of diversity, because many dinoflagellates are known to have a high gene copy number. In addition, the relatively short sequences obtained from 18S V4 primers, restrict the possible length of the barcode region and the taxonomic resolution. Thus, the presented results need to be interpreted with caution.

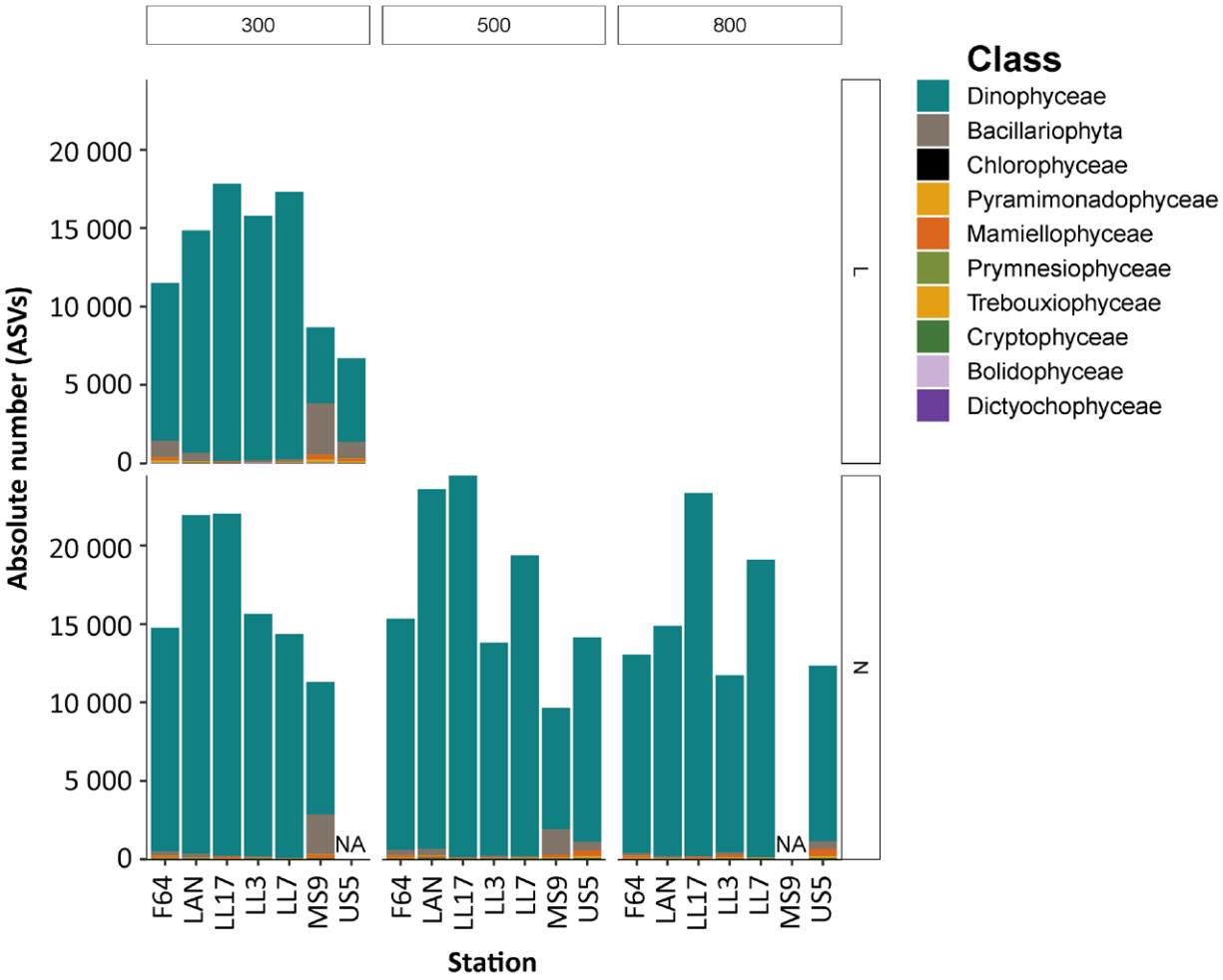


Figure A7. Absolute number of ASVs for photoautotrophic and mixotrophic eukaryotes obtained for different filtration volumes (300 mL, 500 mL and 800 mL – from left to right panel) at the seven sampling stations. Upper panel shows results for acidic Lugol's preserved samples (L). Lower panel shows results for untreated samples (N). All samples were filtered with MCE filters in reusable filtration units.

To illustrate the contribution of all other groups, besides Dinophyceae, they were plotted in relative numbers in **Figure A8**. The second largest number of ASVs was assigned to Bacillariophyta at most stations, except for station L17 (all filtration volumes) and USB5 when 800 mL of sample was filtered. Lugol's preservation had a weak, but noticeable effect on the detected community composition. For most stations more ASVs were assigned to Bacillariophyta in Lugol preserved samples, whereas more ASVs were assigned in Trebouxiophyceae and Cryptophyceae in untreated samples (**Figure A8**, 300

mL filtrations). In contrast, more ASVs were assigned to Dinophyceae in untreated samples at some stations (F64, LL17, LAN, MS9), compared to Lugol's preserved samples (**Figure A7**).

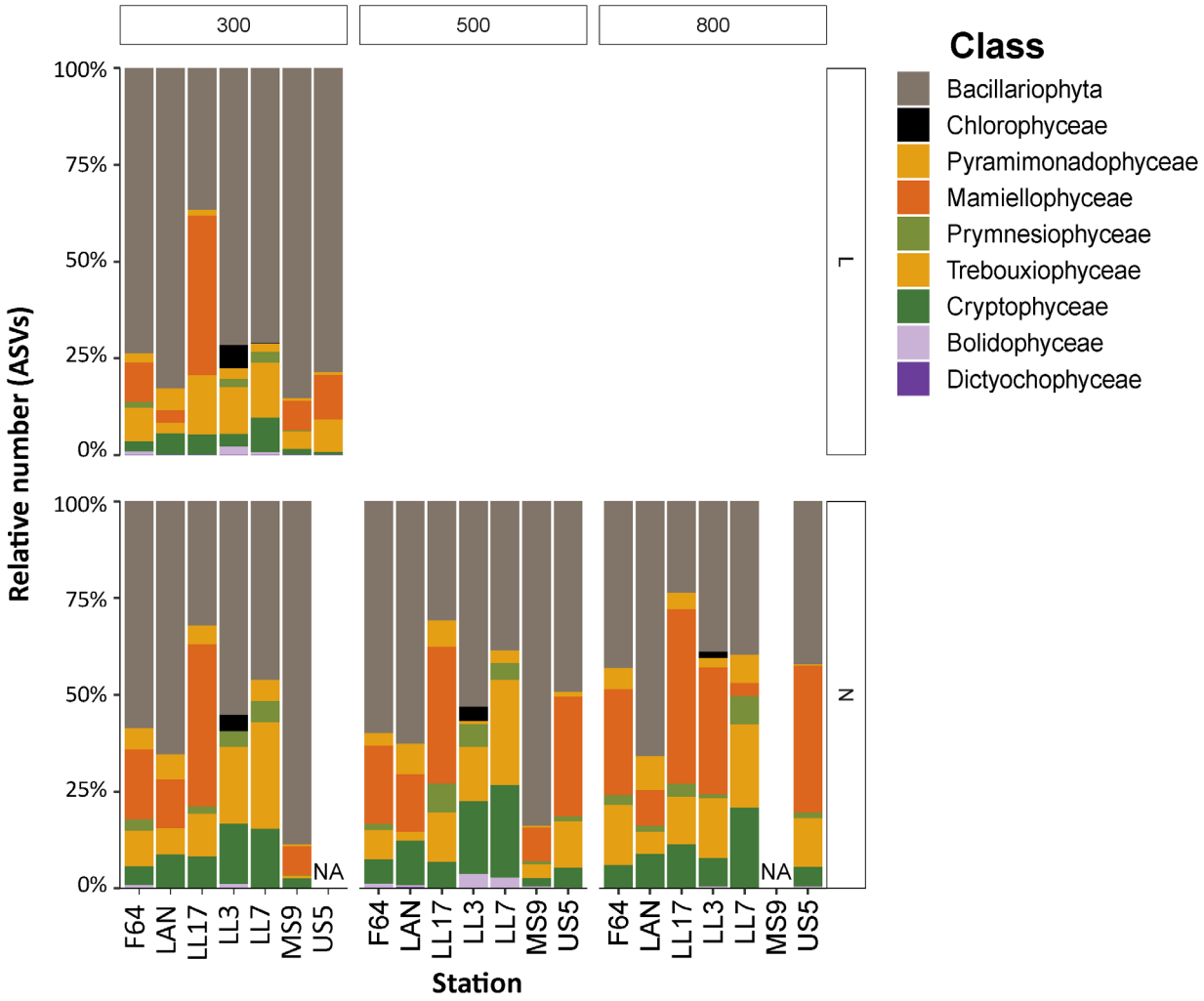


Figure A8. Relative number of ASVs for photoautotrophic and mixotrophic eukaryotes, excluding Dinophyceae, obtained for different filtration volumes (300 mL, 500 mL and 800 mL – panels from left to right) at the seven sampling stations. Upper panel shows acidic Lugol's preserved samples (L). Lower panel shows untreated samples (N). All samples were filtered with MCE filters in reusable filtration units.

A 5.4 Sequencing results for cyanobacteria

A total of 5,623,747 raw reads were obtained for 16S rRNA gene samples (n = 35). Three samples (two non-template controls and one filtration blank) with less than 1000 sequences were removed from downstream processing. After removing these samples, the average number of reads per sample was 175,209. On average 72% of the sequences per sample passed the denoising filter. The amount of cyanobacteria varied between 1% and 4% of the total bacterial community. *Cyanobium* PCC-6307 (Family Cyanobiaceae) was the most abundant genus of cyanobacteria in all samples (68.4-99.3%), except in Längden (LAN), where the proportion of *Aphanizomenon* MDT14a (family Nostocaceae) was 75.8% in one of the three replicates (Figure A9). Previous studies have shown that *Synechococcus* (Andersson et al., 2010, Lindh et al., 2015) and *Cyanobium* (Celepli et al., 2017) dominate in the Baltic Sea cyanobacterial community. In all these studies, the Silva database was used for classification.

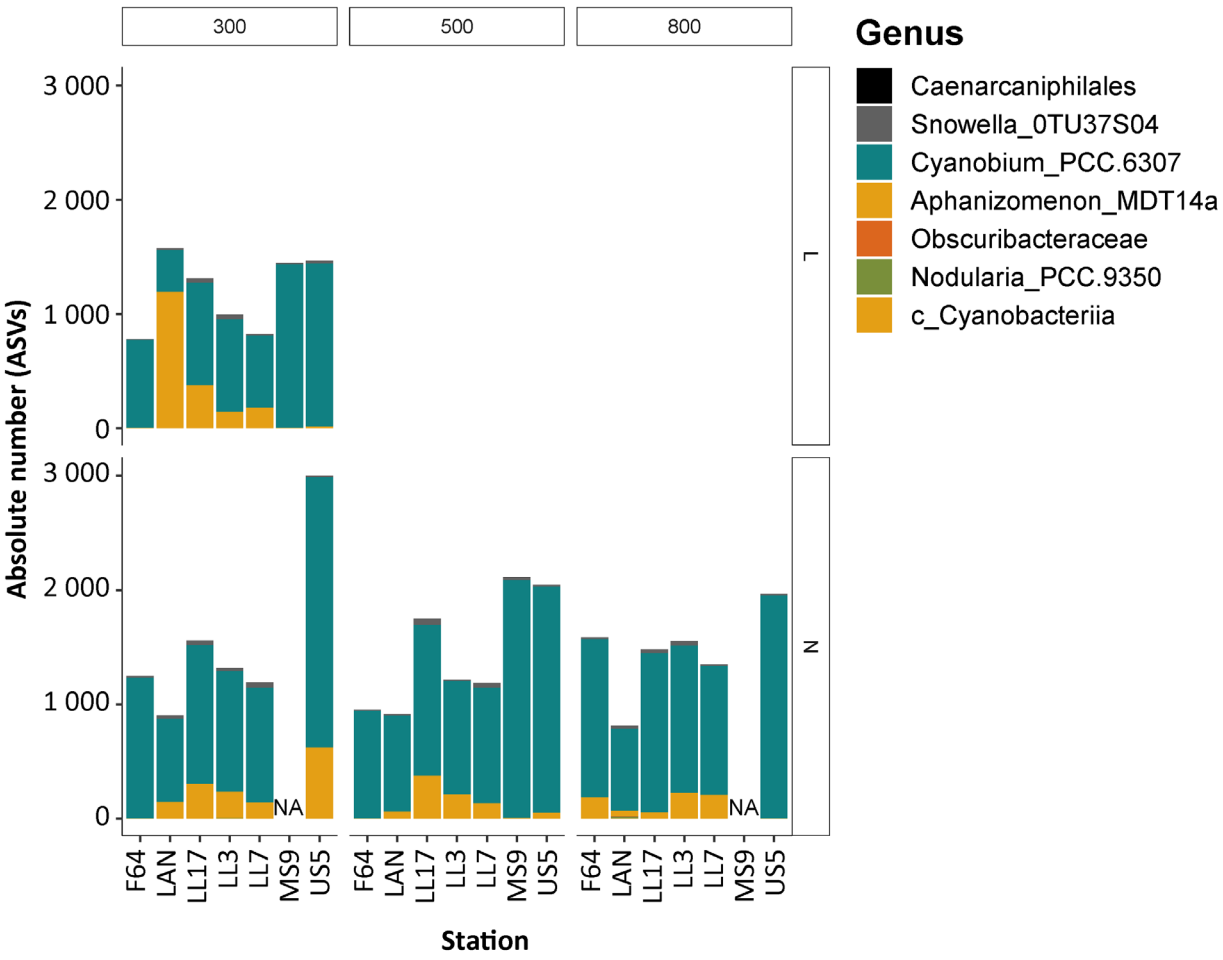


Figure A9. Absolute number of ASVs for cyanobacterial groups detected with 16S rRNA gene primers at genus level (except of c_Cyanobacteriia, which were summarised at class level) for different filtration volumes (300 mL, 500 mL and 800 mL – panels from left to right). Upper panel shows acidic Lugol's preserved samples (L). Lower panel shows untreated samples (N). All samples were filtered with MCE filters in reusable filtration units.

A 5.5 Comparison of microscopy and eDNA results: species composition

We compared the phytoplankton species composition results of the routine phytoplankton monitoring method and the eDNA metabarcoding method. The phytoplankton taxa observed in the light microscopical analysis and the eDNA analysis (ASVs, 500 mL sample, MCE filter, reusable filtration units) are presented in **Table S1**. Comparing and harmonizing the results and taxonomy of the two methods proved surprisingly challenging. A combination of the HELCOM Expert Group on Phytoplankton (EG PHYTO) taxon and biovolume list (Olenina et al., 2006, and its annually updated annex, version 2021), AlgaeBase (Guiry & Guiry, 2022) and WoRMS (WoRMS Editorial Board, 2022) were utilized as sources for the nomenclature and higher taxonomy.

Very few of the same taxa were found using both light microscopy and eDNA metabarcoding. Only two species, the common spring bloom diatom *Skeletonema marinoi* and the common bloom-forming cyanobacterium Aphanizomenon *flosaquae* were detected with both methods (**Table S1**). Contrary to expectations, common spring bloom and/or generalist species, such as the dinoflagellates *Peridiniella catenata*, *Heterocapsa rotundata*, *Apocalathium malmogiense*, *Biecheleria baltica* and *Gymnodinium corollarium* (the latter three constituting the so called *Apocalathium* complex) were not among the assigned eDNA taxa, although they were part of the used reference database (PR^2 plus in-house sequences). Most likely because only one sequence was included for each taxon (except of *Heterocapsa*

rotunda). The diatoms *Chaetoceros wighamii* and *Thalassiosira baltica* were not among the assigned eDNA taxa, since they were not part of the used reference database.

Also, the commonly occurring phototrophic ciliate Mesodinium rubrum was observed in the light microscopy analysis. Several ciliates were detected using eDNA (**Table S2**), however *M. rubrum* was not among them. *M. rubrum* is included in the data of primary producing phytoplankton in the Baltic Sea (HELCOM, 2021), since it is mixotrophic due to its cryptophyte endosymbiont.

Three genera, the haptophyte genus *Chrysochromulina*, the diatom genus *Chaetoceros* and the prasinophyte genus *Pyramimonas*, all commonly occurring taxa in the northern Baltic Sea, were detected using both methods. Also, two order level determinations, Cryptomonadales (a common group of cryptophytes) and Gymnodiniales (a common group of dinoflagellates) were determined using both methods. However, the results concerning *Chrysochromulina*, Cryptomonadales and Gymnodiniales are not quite straightforward to compare, since in the light microscopy analysis (out of practical reasons), these taxa potentially also include morphologically similar taxa belonging in other groups. Generally speaking, comparing more specific taxonomical levels is of greater value than comparing less specific taxonomic levels, since the taxa that are identified to a less specific level usually comprise a wide range of species with widely differing ecological preferences.

Some particularly interesting eDNA findings were made. *Teleaulax gracilis*, which is a fairly recently described cryptophyte (Laza-Martinez et al., 2012), was detected in all sea areas except the northern Baltic Proper (station L17). It has not been found in the Baltic Sea previously, based on microscopy (G. Hällfors, 2004; or subsequently, to the best of our knowledge). It differs from the two *Teleaulax* species occurring in the Baltic Sea (*T. acuta* and *T. amphioxeia;* G. Hällfors, 2004) by having two flagella of equal length (Laza-Martinez et al., 2012). Based on this feature it might be identifiable from monitoring samples; a cultured specimen would need to be preserved and scrutinized to evaluate this.

The marine diatom *Thalassiosira hispida* was detected in the Bothnian Sea (station MS9), even though it has not been found previously in the Baltic Sea (G. Hällfors, 2004; or subsequently, to the best of our knowledge). The distribution pattern of the species fits, since it occurs in northern cold and temperate regions (Hasle & Syvertsen, 1997). The species should be distinguishable from *T. baltica* and *T. levanderi,* the two in the northern Baltic Sea most widely distributed (G. Hällfors, 2004) and most commonly occurring *Thalassiosira* species, at least from girdle view based on cell shape, which in *T. hispida is* more or less octagonal (cf. Hasle & Syvertsen, 1997).

The dinoflagellate *Polarella glacialis* is an example of a species that is known to occur in the Baltic Sea, but which cannot be determined using conventional phytoplankton monitoring methods. It has previously been recorded from the Baltic Sea based on 18S rRNA gene (Majaneva et al., 2012). In addition, cysts of this species have been found in the Baltic Sea plankton community (H. Hällfors, 2013). In the present pilot study, it was observed in the Bothnian Sea (MS9), the northern Baltic Proper (LL17), and the western Gulf of Finland (Längden), but as expected, only using eDNA metabarcoding. In samples preserved with acidic Lugol's solution the vegetative stage of the dinoflagellate resembles a small nondescript *Gymnodinium* cell, and the species is primarily identifiable by its quite characteristic cyst (cf. Montresor et al., 1999, 2003). However, resting stages are not recorded within phytoplankton monitoring, and thus the species does not show up in the data.

While we here focus on the results from the treatment recommended in the guidelines (i.e. 500 mL untreated sample, MCE filters; **Section 4.2**), in other tested treatments two additional interesting taxa were detected. In the acidic Lugol's solution preserved, 300 mL MCE filtered sample from the Bothnian Sea (MS9), the dinoflagellate genus *Islandinium* was detected. It has previously not been found in the Baltic Sea based on microscopy (G. Hällfors, 2004; or subsequently, to the best of our knowledge). The genus includes at least euryhaline (5-10 psu, up to 35 psu) coldwater species (*I. minutum*; Head et al., 2001, Potvin et al., 2013) and species occupying specific environments with somewhat reduced salinities and elevated nutrient levels in the warm temperate zone (*I. brevispinosum*; Pospelova & Head, 2002). The genus *Islandinium* includes peridinioids without particularly characteristic gross-

morphological features (i.e. they have roundish cells lacking distinctive horns, spines, or flanges) and is probably too difficult to identify to even genus level in monitoring samples. The vegetative stages of all *Islandinium* cysts found in modern sediments are not known, but the genus is protoperidiniacean, indicating that a 'typical' protoperidinoid gross morphology (antapical horns and a pointed epitheca) of the vegetative stage is plausible also. It is thus possible that specimens of *Islandinium* are determined as Peridiniales or *Protoperidinium* spp. in monitoring samples.

The other noteworthy finding in a test treatment (i.e. in the 300 mL sample), is that in the western Gulf of Finland the family Noelaerhabdaceae was detected. The family comprises the coccolithophorid genera *Emiliania*, *Gephyrocapsa* and *Reticulofenestra* (Young et al., 2003). *Emiliania huxleyi* is a marine species, which has, based on microscopy, been recorded from the Kattegat and the Belt Sea area; an uncertain record is indicated for the Gulf of Finland (the observation is based on single scales only; G. Hällfors, 2004). In the northern Baltic Sea, the family has previously been detected from the Gulf of Bothnia based on 18S rRNA gene (Hu et al., 2016). Our finding constitutes the potential confirmation of noelaerhabdacean coccolithophorids also in the Gulf of Finland. In Noelaerhabdaceae the haptonema is vestigial (Young et al., 2003), and furthermore coccoliths (calcium carbonate plates) often dissolve in acidic Lugol's solution. Noelaerhabdaceans are therefore likely not identified as any kind of coccolithophorids or even haptophytes in monitoring samples, but merely as flagellates.

In addition to phytoplankton, the eDNA method also detected other organism groups, e.g. multicellular animals (Metazoa), ciliates (Ciliophora), and fungi (Fungi). All observed ASVs are listed in **Table S2**.

Since the volume of a sample processed for the eDNA metabarcoding is much larger than that for light microscopy samples (500 mL filtered for eDNA analysis, whereas 10-50 mL is settled for light microscopy and only a fraction of the volume is actually analysed), the lack of expected common and abundant taxa in the eDNA results is unlikely an artefact caused by sampling. Most likely, it is due to shortcomings of the reference library or the use of short reads. Assigning reads to lower-level ranks (family, genus, species) is more difficult because (1) there is limited variability in the short reads, (2) reference databases are incomplete (taxa are missing and there is limited within-taxon sampling i.e. variability represented) and the applied method for taxonomic assignment lacks a proper assessment of identification reliability (Somervuo et al., 2017). Thus, using third-generation long-read sequencing technologies, which can produce reads covering the whole 16S and 18S rRNA genes would allow more accurate and more specific identification of the target organisms (Callahan et al. 2021). Databases need to be improved by adding the missing species and adding more specimens per species. The third problem can be solved by including methods that assess the reliability of taxonomic assignment.

Another potential improvement could be achieved by fine-tuning the parameter settings of the pipeline used in our pilot study or by using phylogenetic placement methods. Phylogenetic methods reconstruct a phylogenetic tree based on obtained reads or place the obtained reads onto an existing reference tree using an evolutionary placement algorithm, for example EPA-ng (Barbera et al. 2019). Taxonomic assignment using a phylogenetic placement method was tested during this pilot study, but further validation of the method is necessary before it can be used for monitoring and no results are presented in this document.

A 5.6 Comparison of microscopy and eDNA results: biodiversity

We compared the numbers of phytoplankton taxa, size classes, and ASVs that were observed in the monitoring samples using the standardized light microscopy method (CEN, 2015, HELCOM, 2021) and in the 500 ml eDNA samples filtered using reusable filtration units. Results are shown for the five different sea areas sampled during the monitoring cruise (**Table A3**). Photoautotrophic and mixotrophic planktonic microalgae, cyanobacteria, heterotrophic dinoflagellates, and other heterotrophic flagellates, as well as the mixotrophic ciliate *Mesodinium rubrum* were included in this comparison. Single-celled <2 µm-sized picoplankton taxa are included only in the eDNA results since the light microscopy method is not suitable for analysing them (HELCOM, 2021). Due to the small number of samples, statistical analyses were not conducted.

Table A3. Number of phytoplankton taxa, size classes, and amplicon sequence variants (ASVs) observed in the monitoring samples by light microscopy and/or in the 500 ml eDNA samples filtered using MCE filters and reusable filtration units. Photoautotrophic and mixotrophic planktonic microalgae, cyanobacteria, heterotrophic dinoflagellates, and other heterotrophic flagellates, and the mixotrophic ciliate *Mesodinium rubrum* are included. Single-celled picoplankton taxa are included only in the eDNA results

	Phytoplankton taxa		Size classes	ASVs
	Light microscopy	eDNA	Light microscopy	eDNA
Bothnian Sea	34–38	26–30	90–105	11 605–16 089
Åland Sea	45	24	108	16 282
Northern Baltic Proper	33	19	70	26 042
Western Gulf of Finland	34–42	19-20	79–100	20 195–24 710
Eastern Gulf of Finland	39	19	101	14 873

Sampling stations are situated in the HELCOM sea areas as follows: stations US5B and MS9 in the Bothnian Sea, F64 in the Åland Sea, LL17 in the northern Baltic Proper, Längden (LAN) and LL7 in the western Gulf of Finland, and LL3A in the eastern Gulf of Finland (see **Figure A1**).

A 5.6.1 Number of observed taxa

The light microscopy method yielded the highest number of observed taxa in the Åland Sea and the eDNA method yielded the highest number of observed taxa in the Bothnian Sea (**Table A3**). Both methods yielded the lowest number of observed taxa in the northern Baltic Proper.

For all sea areas, the number of observed taxa was lower with the eDNA method compared to the light microscopy method. This result emphasises the importance of reference database development for the genetic sequences and testing and developing more suitable tools for the bioinformatic analyses.

A 5.6.2 Number of size classes

The light microscopy analysis facilitates assessment of phytoplankton size classes. The number of size classes was the highest in the sea areas where the number of observed taxa was also the highest, i.e, in the Åland Sea and in the Bothnian Sea (**Table A3)**. The number of size classes is usually higher than the number of observed taxa, which indicates that the biodiversity is higher than could be concluded based merely on the number of observed taxa. This is because the size classes separate specimens that are identified on less specific taxonomic levels into different groups based on their size, shape, and trophy. For example, the dinoflagellates that were identified as Gymnodiniales spp. (order level) were enumerated in a total of 10 different size classes in this data set. Size classes are also used for species level identifications, to facilitate as accurate biovolume determination as possible. Size classes on the species level may also, at least in some cases, potentially indicate intra-specific genetic variability, i.e. biodiversity.

A 5.6.3 Number of Amplicon Sequence Variants (ASVs)

ASV data are produced in the eDNA analysis. The number of ASVs was highest in the northern Baltic Proper and in the western Gulf of Finland (**Table A3**). As expected, the number of ASVs was not in line with the number of observed taxa. Even though the number of ASVs was the highest in the northern Baltic Proper, the number of observed taxa was the lowest in that area, based on both light microscopy and eDNA results. In addition, the number of observed size classes was the lowest in the northern Baltic Proper, based on the light microscopy method. Most of the ASVs in the eDNA samples were assigned to the dinoflagellate class Dinophyceae (18S rRNA gene primers, eukaryotes) and to the cyanobacteria genus *Cyanobium* (16S rRNA gene, prokaryotes).

In the case of dinoflagellates, the unspecific taxonomic assignment of many ASVs to the class Dinophyceae does not allow any further conclusions about their ecology or morphology (i.e., trophy, if

they were free-living or parasitic, or if they had a theca or not) as it is possible when using light microscopy. However, there may be a large diversity among the spring-bloom dinoflagellates in the northern Baltic Sea, even though their species level identities are currently not revealed by the eDNA analysis. Populating sequencing reference databases and testing and developing more suitable bioinformatic tools is needed to reveal more detail. On the other hand, it is known that dinoflagellates have a high gene copy number variation (e.g. Gong & Marchetti, 2019, Yarimizu et al., 2021), which likely inflated the number of ASVs assigned to dinoflagellates in the samples. In the monitoring samples, analysed quantitatively by light microscopy, dinoflagellates formed only <1 % of the total cell abundance (units L-1) and of the biomass (µg L-1). Based on the light microscopy analysis of the monitoring samples, the most abundant (>2 µm-sized) phytoplankton groups were diatoms and unidentified nanoflagellates, and diatoms dominated by biomass.

The second group to which most of the 16S rRNA gene-derived ASVs were assigned at genus level, was *Cyanobium*, which consists of picoplanktonic cyanobacteria from the order Synechococcales (Salazar et al., 2020). This information is interesting for phytoplankton monitoring, since <2 µm-sized single-celled picoplankton cannot be enumerated or identified by the currently used light microscopy method (HELCOM, 2021). This result highlights the need for method development to characterize *picoplankton*, which is a particularly important part of the primary producing phytoplankton in the northern Baltic Sea (Kuosa, 1991, Celepli et al., 2017, Tamm et al., 2018).

A 5.6.4 Discussion on biodiversity indicator development

Detailed biodiversity data are required to develop an indicator that describes the diversity of phytoplankton and that can be used in marine management. Such a phytoplankton biodiversity indicator is currently lacking in e.g. the HELCOM area (the Baltic Sea). In addition to taxonomic diversity, also functional phytoplankton diversity needs to be considered more carefully in marine management in the future.

For various types of environments and organism groups, biodiversity is most often assessed based on taxonomic data. eDNA data might be used in the future as one of the non-taxonomic biodiversity indicators as well. To develop such an indicator, eDNA results should be analysed together and compared with the environmental status and anthropogenic pressure data.

Taxonomic data offer a possibility to develop both taxonomic and functional biodiversity indicators. In the future, genetic information on phytoplankton functions, which are currently present in the community (e.g. nitrogen fixation, toxicity), may enable further development of functional biodiversity indicators. However, combining quantitative light microscopy data and eDNA data is currently the most promising approach to gain better information for both taxonomic and functional phytoplankton biodiversity indicators.

A 6 Required resources

The pilot study required working hours (human resources) for planning the eDNA sampling, surveying relevant literature, acquiring the equipment and materials, performing the sampling and sample processing onboard the research vessel and in the laboratory, growing and sampling the phytoplankton culture strains, performing the light and epifluorescence microscopy, conducting the bioinformatics analysis and the taxonomical analysis, and finally, the compiling, examining and reporting of the results. The most laborious individual work phase was bioinformatics analysis. Bioinformatics requires special expertise in molecular biology, pipelines, and analysing tools. Once an approved bioinformatic pipeline is taken into use, work time and costs will be strongly reduced, since upscaling is easy. However, like taxonomical light microscopical phytoplankton analysis, which requires extensive expertise and working time, also implementing molecular biology methods requires additional human resources with specific skills and working time.

The required working hours were the highest expense in the project, but resources were needed also for acquiring the equipment and materials, and for the sequencing itself. Samples were sent to another laboratory specialized in sequencing. The sequencing costs were the second highest expense after the working hours.

Supplementary information

Table S1. The phytoplankton taxa and amplicon sequence variants (ASVs) observed in the monitoring samples by light microscopy (LM), in the 500 ml eDNA samples filtered using the reusable filtration units (DNA), or by both methods (DNA / LM). Photoautotrophic and mixotrophic planktonic microalgae, cyanobacteria, heterotrophic dinoflagellates and other heterotrophic flagellates, and the mixotrophic ciliate *Mesodinium rubrum* are included. Single-celled picoplankton taxa are included only in the eDNA results. The higher taxonomy and nomenclature follow Olenina et al. (2006, and its annually updated annex, version 2021), a requisite in Baltic Sea phytoplankton monitoring. The taxa not found there follow AlgaeBase (Guiry & Guiry, 2022), WoRMS (WoRMS Editorial Board, 2022), or a combination of the first mentioned and either of the latter two, or the SILVA 138 database. The original taxonomic assignments by the eDNA databases are found in Tables S2 and S3. BS = Bothnian Sea (stations US5B and MS9), ÅS = Åland Sea (F64), NBP = northern Baltic Proper (LL17), WGF = western Gulf of Finland (Längden and LL7), EGF = eastern Gulf of Finland (LL3A).

Division	Class	Order	Taxon / ASV	BS	ÅS	NPB	WGF	EGF
Cyanophyta	Cyanophyceae	Chroococcales	Chroococcales	LM			LM	
Cyanophyta	Cyanophyceae	Synechococcales	*Merismopedia* spp.					LM
Cyanophyta	Cyanophyceae	Synechococcales	*Planktolyngbya* spp.		LM		LM	LM
Cyanophyta	Cyanophyceae	Synechococcales	*Snowella* (0TU37S04)	DNA	DNA	DNA	DNA	DNA
Cyanophyta	Cyanophyceae	Synechococcales	*Snowella* sp.				DNA	DNA
Cyanophyta	Cyanophyceae	Synechococcales	cf. *Snowella* spp.				LM	
Cyanophyta	Cyanophyceae	Synechococcales	cf. *Woronichinia* spp.				LM	
Cyanophyta	Cyanophyceae	Synechococcales	*Synechococcus rubescens* (1)	DNA	DNA	DNA	DNA	DNA
Cyanophyta	Cyanophyceae	Synechococcales	*Synechococcus* (uncultured) (1)	DNA	DNA	DNA	DNA	
Cyanophyta	Cyanophyceae	Synechococcales	*Synechococcus* (-like strain) (1)				DNA	DNA
Cyanophyta	Cyanophyceae	Synechococcales	*Cyanobium* (PCC-6307) (2)	DNA	DNA	DNA	DNA	DNA
Cyanophyta	Cyanophyceae	Oscillatoriales	Oscillatoriales	LM			LM	
Cyanophyta	Cyanophyceae	Nostocales	*Aphanizomenon flosaquae*	DNA / LM	DNA / LM	DNA / LM	DNA / LM	DNA / LM
Cyanophyta	Cyanophyceae		Cyanophyceae (Cyanobacteriia)			DNA		
Cyanobacteria	Vampirivibrionia	Caenarcaniphilales	Caenarcaniphilales (uncultured cyanobacterium) (3)	DNA				
Cryptophyta	Cryptophyceae	Pyrenomonadales	*Hemiselmis virescens*			LM	LM	LM
Cryptophyta	Cryptophyceae	Pyrenomonadales	*Plagioselmis prolonga*	LM	LM	LM	LM	LM
Cryptophyta	Cryptophyceae	Pyrenomonadales	*Teleaulax gracilis*	DNA	DNA		DNA	DNA
Cryptophyta	Cryptophyceae	Pyrenomonadales	*Teleaulax* spp.	LM	LM	LM	LM	LM
Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales (Cryptomonadales_XX_sp.)	DNA	DNA	DNA	DNA	DNA
Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales	LM				
Cryptophyta	Cryptophyceae		Cryptophyceae (Basal_Cryptophyceae-1_X_sp.)	DNA	DNA			
Dinophyta	Dinophyceae	Amphidiniales	*Amphidinium sphenoides*	LM	LM	LM	LM	
Dinophyta	Dinophyceae	Dinophysiales	*Dinophysis acuminata*	LM	LM			
Dinophyta	Dinophyceae	Dinophysiales	*Dinophysis* spp.		LM		LM	

Division	Class	Order	Taxon / ASV	BS	ÅS	NPB	WGF	EGF
Dinophyta	Dinophyceae	Gonyaulacales	*Peridiniella catenata*	LM	LM	LM	LM	LM
Dinophyta	Dinophyceae	Gymnodiniales	*Gymnodinium stellatum*				LM	
Dinophyta	Dinophyceae	Gymnodiniales	*Gyrodinium spirale*			LM		
Dinophyta	Dinophyceae	Gymnodiniales	cf. *Gyrodinium* spp.				LM	
Dinophyta	Dinophyceae	Gymnodiniales	Gymnodiniales	LM	LM	LM		
Dinophyta	Dinophyceae	Gymnodiniales	Gymnodiniales				DNA	
Dinophyta	Dinophyceae	Syndiniales	Syndiniales (Dino-Group-I-Clade-1_X_sp.) (4)	DNA	DNA	DNA	DNA	DNA
Dinophyta	Dinophyceae	Syndiniales	Syndiniales (Dino-Group-I-Clade-5_X_sp.) (4)			DNA		
Dinophyta	Dinophyceae	Syndiniales	Syndiniales (Dino-Group-II-Clade-1_X_sp.) (4)					DNA
Dinophyta	Dinophyceae	Syndiniales	Syndiniales (Dino-Group-II-Clade-10-and-11_X_sp.)(4)	DNA				
Dinophyta	Dinophyceae	Syndiniales	Syndiniales (Dino-Group-II-Clade-4_X_sp.) (4)	DNA		DNA		
Dinophyta	Dinophyceae	Syndiniales	Syndiniales (Dino-Group-III_XX_sp.) (4)	DNA				
Dinophyta	Dinophyceae	Peridiniales	*Heterocapsa arctica* subsp. *frigida*		LM		LM	LM
Dinophyta	Dinophyceae	Peridiniales	*Heterocapsa rotundata*	LM		LM	LM	
Dinophyta	Dinophyceae	Peridiniales	*Oblea rotunda* complex	LM				
Dinophyta	Dinophyceae	Peridiniales	*Protoperidinium bipes*	LM	LM		LM	
Dinophyta	Dinophyceae	Peridiniales	*Protoperidinium granii*	LM	LM		LM	LM
Dinophyta	Dinophyceae	Peridiniales	*Protoperidinium* spp.	LM		LM	LM	LM
Dinophyta	Dinophyceae	Peridiniales	Peridiniales	LM				
Dinophyta	Dinophyceae	Peridiniales	cf. Peridiniales		LM			
Dinophyta	Dinophyceae	Prorocentrales	*Prorocentrum* sp.	DNA		DNA		
Dinophyta	Dinophyceae	Suessiales	*Polarella glacialis*	DNA		DNA	DNA	
Dinophyta	Dinophyceae	Suessiales	Suessiales (Suessiales_XX_sp.) (5)	DNA			DNA	DNA
Dinophyta	Dinophyceae	Suessiales	Suessiales (5)				DNA	
Dinophyta	Dinophyceae	Thoracosphaerales	*Apocalathium* CPX	LM	LM	LM	LM	LM
Dinophyta	Dinophyceae	Tovelliales	*Katodinium glaucum*	LM		LM		
Dinophyta	Dinophyceae		Dinophyceae (Dinophyceae_XXX_sp.)	DNA	DNA	DNA	DNA	DNA
Dinophyta	Dinophyceae		Dinophyceae	DNA	DNA	DNA	DNA	DNA
Dinophyta			Dinophyta (Dinoflagellata)	DNA	DNA			
Haptophyta	Prymnesiophyceae	Prymnesiales	*Chrysochromulina* spp.	LM	LM			
Haptophyta	Prymnesiophyceae	Prymnesiales	*Chrysochromulina* sp.	DNA	DNA	DNA	DNA	DNA
Haptophyta	Prymnesiophyceae	Prymnesiales	cf. *Chrysochromulina* spp.			LM		LM
Haptophyta	Prymnesiophyceae	Prymnesiales	*Prymnesium*	DNA				
Chrysophyta	Chrysophyceae	Chromulinales	*Dinobryon* spp.		LM	LM	LM	
Chrysophyta	Chrysophyceae	Pedinellales	*Pseudopedinella thomsenii*	LM	LM	LM	LM	LM
Chrysophyta	Chrysophyceae	Pedinellales	*Pseudopedinella* spp.	LM	LM	LM	LM	LM
Chrysophyta	Chrysophyceae	Pedinellales	Pedinellales (Pedinellales_X_sp.)				DNA	

Division	Class	Order	Taxon / ASV	BS	ÅS	NPB	WGF	EGF
Ochrophyta	Chrysophyceae	Parmales	Parmales (6)	DNA				
Chrysophyta	Diatomophyceae	Eupodiscales	*Actinocyclus octonarius*	LM	LM			LM
Chrysophyta	Diatomophyceae	Eupodiscales	*Attheya* spp.	LM	LM		LM	
Chrysophyta	Diatomophyceae	Eupodiscales	*Chaetoceros* cf. *danicus*		LM		LM	
Chrysophyta	Diatomophyceae	Eupodiscales	*Chaetoceros* cf. *tenuissimus*			LM		
Chrysophyta	Diatomophyceae	Eupodiscales	*Chaetoceros wighamii*	LM	LM		LM	LM
Chrysophyta	Diatomophyceae	Eupodiscales	*Chaetoceros* spp.	LM	LM		LM	LM
Chrysophyta	Diatomophyceae	Eupodiscales	*Chaetoceros* sp.		DNA			
Chrysophyta	Diatomophyceae	Eupodiscales	*Chaetoceros*	DNA	DNA		DNA	DNA
Chrysophyta	Diatomophyceae	Eupodiscales	*Melosira arctica*	LM	LM			LM
Chrysophyta	Diatomophyceae	Eupodiscales	*Skeletonema marinoi*	LM	LM	LM	DNA / LM	LM
Chrysophyta	Diatomophyceae	Eupodiscales	*Thalassiosira baltica*	LM	LM	LM	LM	LM
Chrysophyta	Diatomophyceae	Eupodiscales	*Thalassiosira hispida*	DNA				
Chrysophyta	Diatomophyceae	Eupodiscales	*Thalassiosira levanderi*	LM	LM		LM	LM
Chrysophyta	Diatomophyceae	Eupodiscales	Diatomophyceae (Polar-centric-Mediophyceae)	DNA	DNA			
Chrysophyta	Diatomophyceae	Eupodiscales	Diatomophyceae (Radial-centric-basal-Coscinodiscophyceae_X_sp.)	DNA	DNA			
Chrysophyta	Diatomophyceae	Eupodiscales	Centrales spp.		LM		LM	LM
Chrysophyta	Diatomophyceae	Bacillariales	*Cylindrotheca closterium*	LM	LM			LM
Chrysophyta	Diatomophyceae	Bacillariales	*Nitzschia frigida*	LM				LM
Chrysophyta	Diatomophyceae	Bacillariales	*Pauliella taeniata*	LM		LM	LM	LM
Chrysophyta	Diatomophyceae	Bacillariales	Diatomophyceae (Raphid-pennate)	DNA		DNA	DNA	DNA
Chrysophyta	Diatomophyceae	Bacillariales	Pennales spp.		LM			
Chrysophyta	Diatomophyceae		Diatomophyceae (Bacillariophyta_X)	DNA	DNA		DNA	DNA
Euglenophyta	Euglenophyceae	Euglenales	*Eutreptiella* spp.	LM	LM		LM	LM
Chlorophyta	Chlorophyceae	Chlamydomonadales	Chlamydomonadales (Chlamydomonadales_X)					DNA
Chlorophyta	Chlorophyceae	Sphaeropleales	*Monoraphidium contortum*	LM	LM	LM	LM	LM
Chlorophyta	Chlorophyceae		cf. Chlorophyceae					LM
Chlorophyta	Mamiellophyceae	Mamiellales	*Bathycoccus prasinos* (7)	DNA	DNA	DNA	DNA	
Chlorophyta	Mamiellophyceae	Mamiellales	*Mantoniella*	DNA	DNA			
Chlorophyta	Mamiellophyceae	Mamiellales	*Ostreococcus mediterraneus* (8)	DNA		DNA		
Chlorophyta	Pyramimonadophyceae	Pyramimonadales	*Pyramimonas* spp.	LM	LM	LM	LM	
Chlorophyta	Pyramimonadophyceae	Pyramimonadales	*Pyramimonas*		DNA			DNA
Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pyramimonadales	DNA	DNA	DNA	DNA	
Chlorophyta	Trebouxiophyceae	Chlorellales	*Oocystis* spp.	LM				
Chlorophyta	Trebouxiophyceae	Chlorellales	cf. *Oocystis* spp.		LM			

Division	Class	Order	Taxon / ASV	BS	ÅS	NPB	WGF	EGF
Chlorophyta	Ulvophyceae	Ulothrichales	*Binuclearia lauterbornii*			LM		
Chlorophyta			Chlorophyta (9)	DNA	DNA	DNA	DNA	DNA
Choanozoa	Choanoflagellatea	Acanthoecida	*Diaphanoeca* spp.	LM	LM		LM	LM
Choanozoa	Choanoflagellatea	Acanthoecida	Acanthoecida (Stephanoecidae_Group_D_X_sp.)	DNA		DNA	DNA	DNA
Choanozoa	Choanoflagellatea	Acanthoecida	Acanthoecida (Stephanoecidae_Group_H_X_sp.)	DNA				
Choanozoa	Choanoflagellatea	Craspedida	*Calliacantha natans*		LM	LM	LM	LM
Choanozoa	Choanoflagellatea	Craspedida	*Calliacantha* spp.	LM				
Choanozoa	Choanoflagellatea	Craspedida	cf. *Calliacantha* spp.	LM	LM	LM	LM	LM
Choanozoa	Choanoflagellatea		Choanoflagellida	DNA			DNA	
Protozoa Incertae sedis	Ebriophyceae	Ebriales	*Ebria tripartita*	LM	LM	LM	LM	LM
Incertae sedis			*Katablepharis ovalis*			LM	LM	
Incertae sedis			*Katablepharis remigera*		LM	LM	LM	LM
Incertae sedis			*Katablepharis* spp.	LM	LM	LM	LM	LM
Incertae sedis			*Telonema subtile*		LM		LM	LM
Telonemia			Telonemia (Telonemia-Group-1_X_sp.) (10)	DNA	DNA		DNA	
Telonemia			Telonemia (Telonemia-Group-2_X_sp.) (10)	DNA	DNA			
Picozoa			Picozoa (11)	DNA	DNA		DNA	DNA
Ciliophora	Litostomatea	Cyclotrichiida	*Mesodinium rubrum*	LM	LM	LM	LM	LM
Others			Flagellates	LM	LM	LM	LM	LM
Others			Unicell	LM	LM	LM	LM	LM

(1) The genus *Synechococcus* comprises species which occur as single cells or grouped into irregular clusters or sometimes into short pseudofilamentous formations, but not in mucilaginous colonies (Komárek & Anagnostidis, 1999). Their cell size varies; the picoplanktonic solitary forms would not be recorded in light microscopy monitoring samples.

(2) Members of the genus *Cyanobium* occur as a solitary form and have the cell size ca 1-2 (4) x 1(-3) µm), i.e. they are in the borderline picoplankton-nanoplankton size range. They may or may not be recorded in phytoplankton monitoring; the larger specimens may be among the cells recorded as 2-3 µm Unicells.

(3) The ASV Caenarcaniphilales (uncultured cyanobacterium) higher taxonomy according to the SILVA 138 database.

(4) The Syndiniales is an order of parasitic dinoflagellates. The free-living life stage would possibly be enumerated in monitoring samples as *Gymnodinium,* Gymnodiniales, or Dinophyceae. The endoparasitic stage would not be recorded.

(5) Dinoflagellates in the order Suessiales have *Gymnodinium*-like cells. Suessiales comprises among other species *Biecheleria baltica*, a common Baltic Sea spring bloom dinoflagellate, determined in monitoring samples as the *Apocalathium* complex.

(6) Members of the order Parmales are in the picoplankton-nanoplankton size range; the order includes both flagellated and coccoid forms. They are likely identified as Flagellates or Unicells in monitoring samples.

(7) *Bathycoccus prasinos* has the cell size 1.5-2.5 x 1-2 µm (Eikrem & Throndsen, 1990), i.e. is in the borderline picoplankton-nanoplankton size range. It may or may not be recorded in phytoplankton monitoring; the larger specimens may be among the cells recorded as 2-3 µm Unicells.

(8) *Ostreococcus mediterraneus* has minute cells, 1.27 ± 0.5 µm (Subirana et al. 2013), i.e. is in the picoplankton size range and not recorded in phytoplankton monitoring. However, as the exact size of such small cells is difficult to measure in LM, some specimens may be among the cells recorded as 2-3 µm Unicells.

(9) Chlorophyta may also be (one or several) macrophyte(s) and/or littoral/benthic species, i.e. non-phytoplankton. However, since we also in phytoplankton monitoring in some cases are unable to determine specimens to a very specific level, and as the samples were collected in the open sea (rather than in coastal areas), we opted to keep this result in the table.

(10) Telonemia is a phylum described fairly recently by Shalchian-Tabrizi et al. (2006), currently comprising only the genus *Telonema*.

(11) Picozoa is a phylum comprising heterotrophic, marine protists of pico-nanoplanktonic size described fairly recently by Seenivasan et al. (2013), currently comprising the species *Picomonas judraskeda* (2.6–3.8 x 2–2.5 µm), which is in the nanoplankton size range and is thus likely recorded in phytoplankton monitoring as 2-3 µm or 3-5 µm Flagellates.

Table S 2. Taxonomic assignments for eukaryotes, including all treatments and samples, based on 18S V4 rDNA gene primers, VSEARCH classification using the PR^2 database plus relevant in-house sequences

Kingdom	Supergroup	Division	Class	Order	Family	Genus	Species
Eukaryota	Alveolata	Ciliophora	CONTH_7	CONTH_7_X	CONTH_7_XX	CONTH_7_XXX	CONTH_7_XXX_sp.
Eukaryota	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
Eukaryota	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	CONThreeP_XXX	CONThreeP_XXX_sp.
Eukaryota	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	Urotrichidae	Urotricha	Urotricha_sp.
Eukaryota	Alveolata	Ciliophora	Litostomatea	Haptoria_4	Didiniidae	Didiniidae_X	Didiniidae_X_sp.
Eukaryota	Alveolata	Ciliophora	Litostomatea	Haptoria_5	Pleurostomatida	Pleurostomatida_X	Pleurostomatida_X_sp.
Eukaryota	Alveolata	Ciliophora	Litostomatea	Haptoria_6	Lacrymariidae	Lacrymariidae_X	Lacrymariidae_X_sp.
Eukaryota	Alveolata	Ciliophora					
Eukaryota	Alveolata	Ciliophora	Oligohymenophorea	Peritrichia_2	Vaginicolidae		
Eukaryota	Alveolata	Ciliophora	Phyllopharyngea	Cyrtophoria_1	PHYLL_4	PHYLL_4_X	PHYLL_4_X_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Leegaardiellidae_B	Leegaardiella	Leegaardiella_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Choreotrichida			
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_A	Rimostrombidium_A	Rimostrombidium_A_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_C	Strobilidiidae_C_X	Strobilidiidae_C_X_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_I	Pelagostrobilidium	Pelagostrobilidium_neptuni
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_I	Pelagostrobilidium	Pelagostrobilidium_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea				
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida			
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae		
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae	Strombidiidae_X	Strombidiidae_X_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_L	Strombidiidae_L_X	Strombidiidae_L_X_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_M		
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_M	Strombidiidae_M_X	Strombidiidae_M_X_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_R		
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_R	Strombidium_R	Strombidium_R_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida_B	Strombidiida_B_X	Strombidiida_B_XX	Strombidiida_B_XX_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Strombidiida_G	Strombidiida_G_X	Strombidiida_G_XX	Strombidiida_G_XX_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Tintinnida			
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Tintinnida	TIN_03	TIN_03_X	TIN_03_X_sp.
Eukaryota	Alveolata	Ciliophora	Spirotrichea	Tintinnida	Tintinnidiidae	Tintinnidium	Tintinnidium_sp.
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Tovelliaceae	Woloszynskia	Woloszynskia_halophila
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Dinophysiales			
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Goniodomataceae	Alexandrium	Alexandrium_ostenfeldii
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales			

Kingdom	Supergroup	Division	Class	Order	Family	Genus	Species
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Warnowiaceae		
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae				
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Amphidiniopsidaceae	Islandinium	
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Amphidiniopsidaceae		
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_sp.
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Suessiales			
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Polarella	Polarella_glacialis
Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiales_X	Suessiales_XX	Suessiales_XX_sp.
Eukaryota	Alveolata	Dinoflagellata					
Eukaryota	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
Eukaryota	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-5	Dino-Group-I-Clade-5_X	Dino-Group-I-Clade-5_X_sp.
Eukaryota	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
Eukaryota	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
Eukaryota	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
Eukaryota	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II			
Eukaryota	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-III	Dino-Group-III_X	Dino-Group-III_XX	Dino-Group-III_XX_sp.
Eukaryota	Alveolata						
Eukaryota	Archaeplastida	Chlorophyta	Chlorophyceae	Chlamydomonadales	Chlamydomonadales_X		
Eukaryota	Archaeplastida	Chlorophyta	Chlorophyceae				
Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	Bathycoccus_prasinos
Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Ostreococcus	Ostreococcus_mediterraneus
Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Mantoniella	
Eukaryota	Archaeplastida	Chlorophyta					
Eukaryota	Archaeplastida	Chlorophyta	Pyramimonadales	Pyramimonadales_X	Pyramimonadales_XX		
Eukaryota	Archaeplastida	Chlorophyta	Pyramimonadales	Pyramimonadales_X	Pyramimonadales_XX	Pyramimonas	
Eukaryota	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Cryptomonadales_XX	Cryptomonadales_XX_sp.
Eukaryota	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Teleaulax	
Eukaryota	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Teleaulax	Teleaulax_gracilis
Eukaryota	Hacrobia	Cryptophyta	Cryptophyceae	Cryptophyceae_X	Basal_Cryptophyceae-1	Basal_Cryptophyceae-1_X	Basal_Cryptophyceae-1_X_sp.
Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Isochrysidales	Noelaerhabdaceae		
Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae				
Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	
Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.

Kingdom	Supergroup	Division	Class	Order	Family	Genus	Species
Eukaryota	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX		
Eukaryota	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-1		
Eukaryota	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-1	Telonemia-Group-1_X	Telonemia-Group-1_X_sp.
Eukaryota	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
Eukaryota							
Eukaryota	Opisthokonta	Choanoflagellida	Choanoflagellatea	Acanthoecida	Stephanoecidae_Group_D	Stephanoecidae_Group_D_X	Stephanoecidae_Group_D_X_sp.
Eukaryota	Opisthokonta	Choanoflagellida	Choanoflagellatea	Acanthoecida	Stephanoecidae_Group_H	Stephanoecidae_Group_H_X	Stephanoecidae_Group_H_X_sp.
Eukaryota	Opisthokonta	Choanoflagellida					
Eukaryota	Opisthokonta	Fungi	Cryptomycota	Cryptomycotina	Cryptomycotina_X	Cryptomycotina_XX	Cryptomycotina_XX_sp.
Eukaryota	Opisthokonta	Fungi					
Eukaryota	Opisthokonta	Metazoa	Arthropoda	Crustacea	Maxillopoda	Acartia	Acartia_bifilosa
Eukaryota	Opisthokonta	Metazoa	Arthropoda	Crustacea	Maxillopoda	Maxillopoda_X	Maxillopoda_X_sp.
Eukaryota	Opisthokonta	Metazoa	Arthropoda	Crustacea	Maxillopoda		
Eukaryota	Opisthokonta	Metazoa	Rotifera	Rotifera_X	Rotifera_XX		
Eukaryota	Opisthokonta						
Eukaryota	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
Eukaryota	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X			
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_sp.
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae		
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Skeletonema	Skeletonema_marinoi
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_hispida
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Radial-centric-basal-Coscinodiscophyceae_X	Radial-centric-basal-Coscinodiscophyceae_X_sp.
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Cylindrotheca	Cylindrotheca_closterium
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate		
Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Raphid-pennate_X	Raphid-pennate_X_sp.

Kingdom	Supergroup	Division	Class	Order	Family	Genus	Species
Eukaryota	Stramenopiles	Ochrophyta	Bolidophyceae	Parmales			
Eukaryota	Stramenopiles	Ochrophyta	Dictyochophyceae	Dictyochophyceae_X	Pedinellales	Pedinellales_X	Pedinellales_X_sp.
Eukaryota	Stramenopiles	Ochrophyta					
Eukaryota	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
Eukaryota	Stramenopiles	Pseudofungi					
Eukaryota	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Thraustochytriaceae_X	Thraustochytriaceae_X_sp.

Table S3. Taxonomic assignments for cyanobacteria, including all treatments and samples, based on 16S V4 rDNA primers, through aligning to the SILVA 138 database.

Domain	Phylum	Class	Order	Family	Genus	Species
Bacteria	Cyanobacteria	Cyanobacteriia	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	
Bacteria	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Aphanizomenon MDT14a	Aphanizomenon flos-aquae
Bacteria	Cyanobacteria	Cyanobacteriia	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	Synechococcus rubescens
Bacteria	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Microcystaceae	Snowella 0TU37S04	
Bacteria	Cyanobacteria	Cyanobacteriia	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured Synechococcus
Bacteria	Cyanobacteria	Cyanobacteriia	Synechococcales	Cyanobiaceae	Cyanobium PCC-6307	Synechococcus-like_str.
Bacteria	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Microcystaceae	Snowella 0TU37S04	
Bacteria	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nodularia_PCC-9350	
Bacteria	Cyanobacteria	Cyanobacteriia				
Bacteria	Cyanobacteria	Vampirivibrionia	Caenarcaniphilales	Caenarcaniphilales		
Bacteria	Cyanobacteria	Vampirivibrionia	Obscuribacterales	Obscuribacteraceae		
Bacteria	Cyanobacteria	Vampirivibrionia	Caenarcaniphilales	Caenarcaniphilales		

Guidelines for using environmental DNA in Finnish marine phytoplankton monitoring
Improved biodiversity assessment through method complementation

Finnish Environment Institute

SYKE

ISBN 978-952-11-5524-6 (PDF)
ISSN 1796-1726 (online)