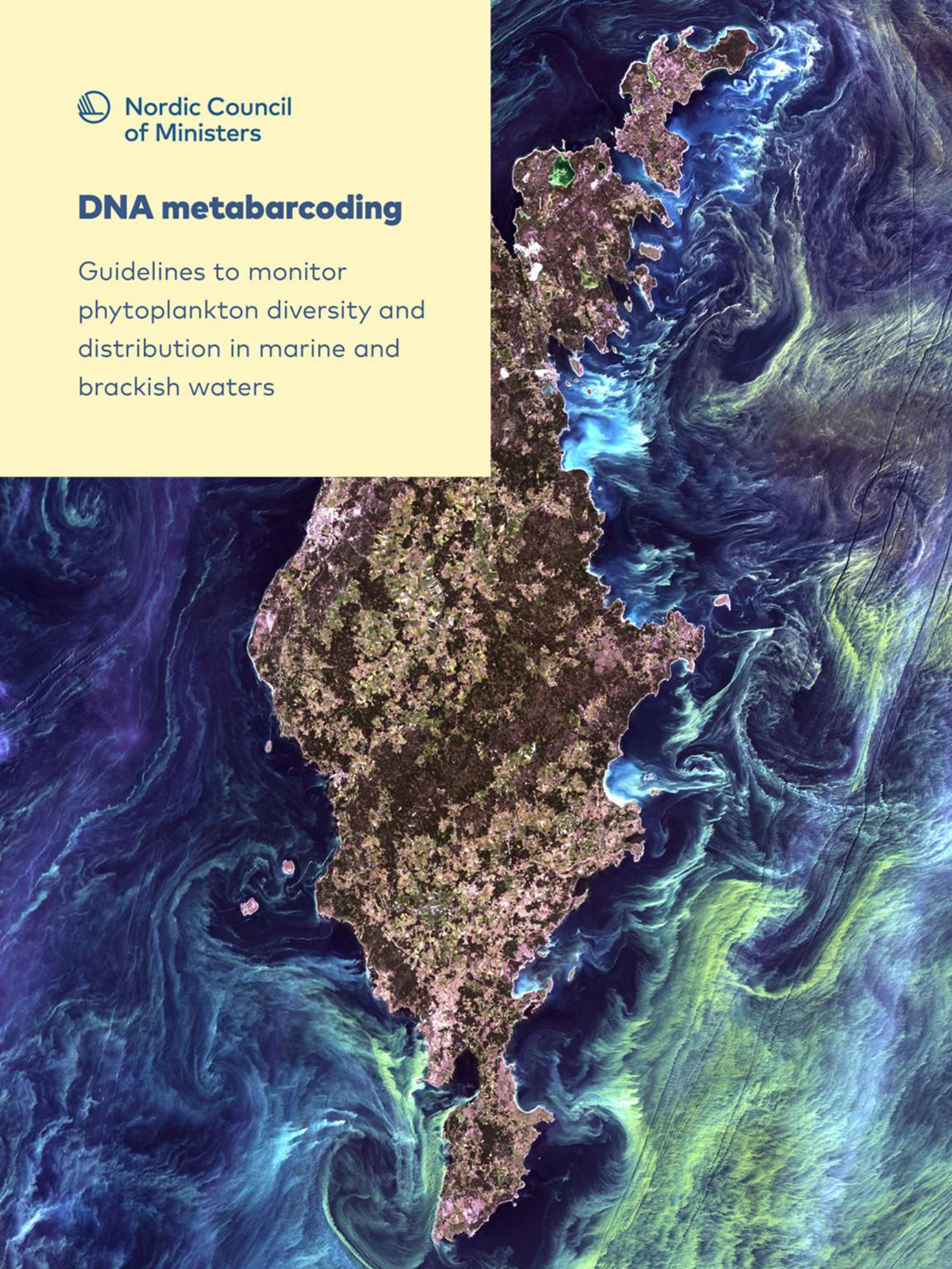


DNA metabarcoding

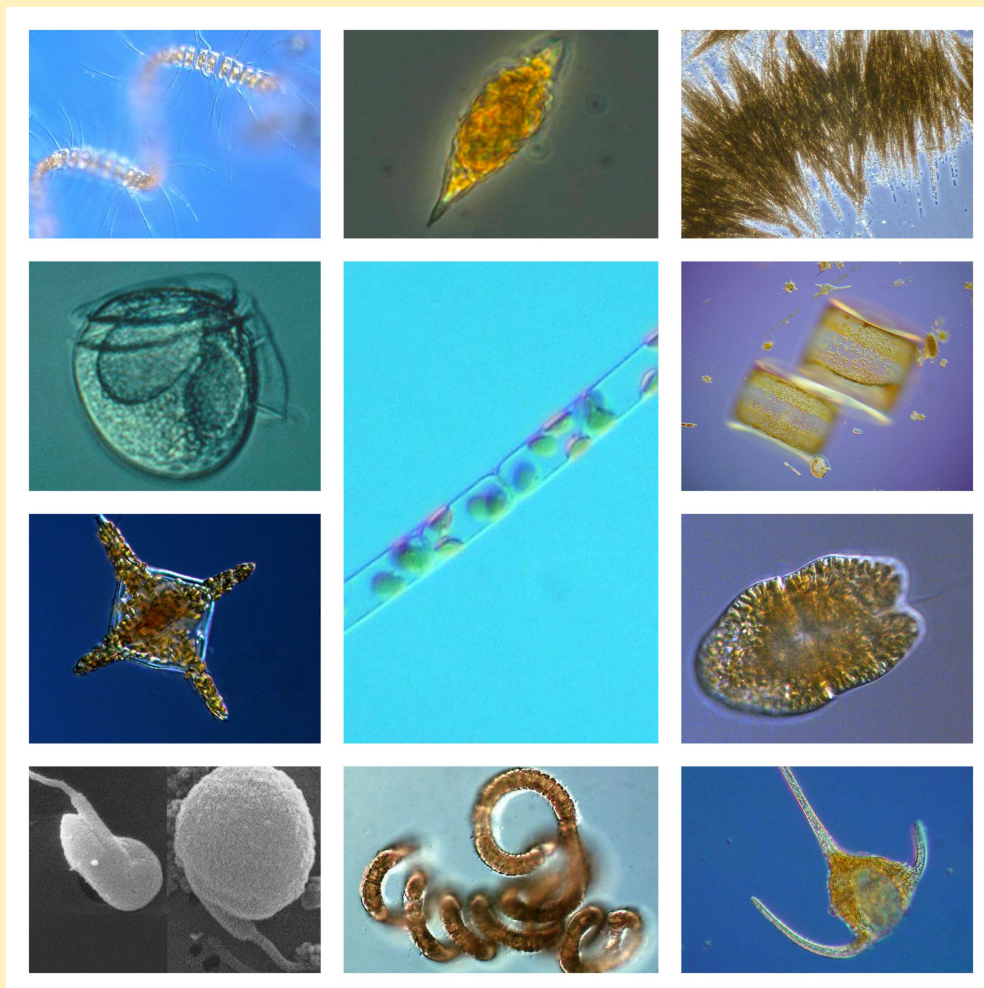
Guidelines to monitor
phytoplankton diversity and
distribution in marine and
brackish waters



Authors and affiliations

Jacqueline Jerney¹, Heidi Hällfors¹, Hans Jakobsen², Iveta Jurgensone³, Bengt Karlson⁴, Anke Kremp⁵, Sirpa Lehtinen¹, Markus Majaneva⁶, Kristian Meissner¹, Veera Norros¹, Sirje Sildever⁷, Sanna Suikkanen¹, Karolin Teeveer⁸

1. Finnish Environment Institute (Syke), Finland
2. Danish Centre for Environment and Energy, Aarhus University (AU), Denmark
3. Latvian Institute of Aquatic Ecology (LHEI), Latvia
4. Swedish Meteorological and Hydrological Institute (SMHI), Sweden
5. Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Germany
6. Norwegian Institute for Nature Research (NINA), Norway
7. Tallinn University of Technology (TalTech), Estonia
8. University of Tartu (UT), Estonia



Photographies in the collage by Seija Hällfors, Sirpa Lehtinen, Bengt Karlson and Ann-Turi Skjevik. The photographs have different scales.

Contents

Authors and affiliations	2
Foreword	5
Summary	7
Executive summary	8
1. Aims and scope	10
1.1 General aims	10
1.2 Scope and definitions	11
2. Background	12
2.1 General introduction	12
2.2 Application to EU directives	14
2.3 Pilot studies and guidelines for eDNA phytoplankton monitoring	15
2.4 A typical eDNA analysis workflow	16
3. Sampling and sample processing	17
3.1 Sampling	17
3.2 Good working practices for minimizing the risk of contamination	17
3.3 Filtration	18
3.3.1 Filtration systems and membranes	18
3.3.2 Filtration volume	19
3.4 Sample preservation	19
3.5 Replicates and filtration blanks	20
4. Molecular work – wet lab	23
4.1 DNA extraction	23
4.2 PCR amplification	25
4.2.1 Primers	25
4.2.2 Polymerases	27
4.2.3 Inhibitor removal	28
4.3 High-throughput amplicon sequencing (HTS)	28
4.4 Quality control	29
4.4.1 Negative controls	29
4.4.2 Positive controls	29
4.4.3 Tagging system controls	29
5. Bioinformatics - dry lab	31
5.1 Sequence quality filtering and clustering	31
5.2 Assigning taxonomy	32
5.3 Reference databases	33
6. Data storage and management	36
7. Future perspectives and recommendations	38

7.1 Standardization	38
7.2 Sequence reference databases	38
7.3 Phytoplankton groups to target with eDNA-metabarcoding	39
7.4 Long-read sequencing	39
8. Cost estimates	40
9. Conclusions and future perspectives	43
Acknowledgements	44
References	45
Appendix	62
About this publication	67

This publication is also available online in a web-accessible version at <https://pub.norden.org/temanord2023-505>.

Foreword

Over the past decades, the rapid advancement of molecular tools has improved our understanding of species diversity, ecology and evolution and their role in ecosystems. Despite their great potential, molecular tools have mostly been used in individual scientific studies and are currently not part of harmonized phytoplankton monitoring programs in Europe.

The guidelines in this document are the result of the project, *Implementing DNA-based methods in Northern European marine phytoplankton monitoring* (ID-NEP), funded by the Nordic Council of Ministers in 2022. ID-NEP aimed at harmonizing the implementation of DNA-based methods in marine (including brackish water) phytoplankton monitoring in Northern Europe and the Baltic Sea area to avoid unnecessary pluralism of national methods and to improve the comparability of results. The project also strengthened collaboration and knowledge exchange between participating institutions and generally benefits scientists and authorities involved in phytoplankton monitoring. ID-NEP aimed to contribute to the establishment of European standards for DNA-based phytoplankton monitoring, which will produce comparable, high-quality biodiversity data to be used in marine and brackish water environmental assessment.

The guidelines for DNA-based identification of prokaryotic and eukaryotic phytoplankton are designed to support existing joint monitoring and environmental protection frameworks such as The Baltic Marine Environment Protection Commission (HELCOM) and The Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR).

Traditional marine phytoplankton monitoring focuses on planktonic primary producing (photoautotrophic and mixotrophic) eukaryotic microalgae and prokaryotic cyanobacteria, but heterotrophic dinoflagellates and other heterotrophic flagellates are included to enable studies with a wider taxonomic scope. In addition, mixotrophic ciliates such as *Mesodinium rubrum* are included since they contribute to primary production because of endosymbiotic cryptophytes. Single-celled <2 µm-sized picoplankton is usually not included, since the conventional method (light microscopy of Lugol's-preserved samples) is not suitable for distinguishing single-celled autotrophic picoplankton from single-celled heterotrophic bacteria. In addition to phytoplankton (as defined above), eDNA metabarcoding can provide information on e.g. picoplankton, microzooplankton, heterotrophic bacteria and fungi present in the phytoplankton samples. Since the guidelines have been optimized for phytoplankton, they need to be carefully evaluated and eventually adapted for monitoring other groups of organisms (e.g. regarding the choice of primers and databases). eDNA metabarcoding can complement traditional monitoring and improve our knowledge about taxa not included in current monitoring. Since DNA-based methods are developing rapidly the guidelines presented here should be regularly updated and developed further in the future, to account for advances in DNA-based methods and harmonize with other international guidelines and standards.

The partner institutes involved in producing these guidelines are the Finnish Environment Institute (Syke), Finland; Aarhus University (AU), Denmark; Latvian Institute of Aquatic Ecology (LIAE), Latvia; Swedish Meteorological and Hydrological Institute (SMHI), Sweden; Norwegian Institute for Nature Research (NINA), Norway; Tallinn University of Technology (TalTech), Estonia; University of Tartu (UT), Estonia and Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Germany.

Summary

This is a technical guide for monitoring the taxonomic diversity of phytoplankton using environmental DNA metabarcoding, together with conventional phytoplankton monitoring techniques in marine and brackish waters. The guidelines focus on the detection of eukaryotic and prokaryotic phytoplankton with 18S and 16S rRNA primers, using high-throughput sequencing. Information and recommendations on sampling, sample processing, molecular biological work, quality control, bioinformatics, data storage and management and cost estimates are included so that the method can be used to complement standardized light microscopy. Furthermore, the guidelines contain information from scientific literature and a discussion on future perspectives, including reference databases and standardization. The guidelines were developed within the project, *Implementing DNA-based methods in Northern European marine phytoplankton monitoring (ID-NEP)*, supported by the Nordic Council of Ministers in 2022. These guidelines serve as a *status quo* to accommodate future advances in molecular methods, international guidelines and standards. Using eDNA metabarcoding to complement standardized light microscopy advances conventional monitoring and research on phytoplankton communities to assess biodiversity and the state of the marine environment.

Executive summary

Using DNA-based methods for monitoring can fundamentally transform environmental monitoring by providing new opportunities and overcoming the shortcomings of traditional methods. However, many aspects still require discussion and agreement amongst experts and stakeholders, before eDNA metabarcoding can be implemented complementary to traditional monitoring techniques. This technical guide for eDNA-based monitoring of phytoplankton in Northern European coastal waters and the Baltic Sea presents the results of the project *Implementing DNA-based methods in Northern European marine phytoplankton monitoring* (ID-NEP). ID-NEP was funded by the Nordic Council of Ministers to harmonize DNA-based methods among Northern European countries and Baltic Sea states and develop guidance on how to complement traditional microscopy-based phytoplankton analyses with DNA-based approaches.

This document presents precise guidelines on technical procedures used for eDNA-based prokaryotic and eukaryotic phytoplankton monitoring in Nordic waters and the Baltic Sea, covering all steps from sampling, sample processing and sequencing procedures to data analysis, interpretation and improvement of databases. The guidelines support implementing a DNA-based approach in parallel with microscopy-based analyses, which allows comprehensive comparison of results obtained from both methods and supports existing joint monitoring and environmental protection frameworks.

DNA-based monitoring can improve biodiversity and status assessment of marine and brackish waters, but it is crucial to have an internationally standardized method to improve large-scale biodiversity assessment. Currently, there is no common phytoplankton biodiversity indicator for European marine and brackish waters. eDNA metabarcoding could kick start the development of such an indicator, which is urgently needed for improved assessment of biodiversity and the status of marine and brackish water bodies.

The technology and methodology described in this publication offer major advantages compared to classic approaches, such as the remarkably improved ability to identify and monitor a very broad range of organisms. The guidelines give advice and an overview of important aspects to consider for the successful implementation of eDNA metabarcoding of phytoplankton. An expansion of the methodology to other groups of organisms, not covered by traditional monitoring, should be considered. Another advantage of eDNA metabarcoding is the limited effort required for upscaling, once common, standardized protocols are in use. Ideally, the guidelines presented herein will be developed further and integrated into e.g., the HELCOM framework and standardized through either CEN or ISO. Doing so will produce comparable, standardized data and allow more cost-efficient eDNA-based monitoring.

One of the biggest short-term challenges for eDNA metabarcoding implementation into routine monitoring is the required improvement of existing reference databases. This publication provides an overview and recommendations on which databases to

use and how to manage sequencing data. The improvement of reference databases requires a coordinated effort and will be expedited if many European states contribute.

In summary, the benefits of using eDNA-based methods for monitoring far outweigh the challenges. We are confident that future marine phytoplankton monitoring will include environmental DNA survey techniques. We hope that these guidelines will facilitate better understanding and harmonization of eDNA-based methods and speed their implementation into routine monitoring programs.

1. Aims and scope

1.1 General aims

The guidelines presented here provide detailed descriptions of procedures necessary for the implementation of eDNA metabarcoding-based phytoplankton monitoring in European seas. Here, we put forward the use of eDNA metabarcoding in combination with – and potentially cost-efficient replacement of – light microscopy in monitoring and research of phytoplankton communities. By introducing standardized protocols for eDNA metabarcoding, which will be updated regularly to accommodate rapid advances in eDNA technologies, this initiative aims to bring forward a harmonized phytoplankton monitoring approach that will enable a more precise assessment of phytoplankton biodiversity and the state of the marine and brackish water environment. The ultimate goal of these guidelines is to initiate the harmonization of eDNA based approaches and respective assessment methods and eventually contribute to the improvement of the status of aquatic environments in Europe.

The resulting data can be used to

- Assess marine phytoplankton biodiversity coherently in Northern Europe
- Study the distribution of phytoplankton in time and space – and produce comprehensive biogeography and phenology data for national surveys
- Conduct a status assessment required by the various EU directives across national borders ([Section 2.2](#))
- Study long-term changes of phytoplankton to understand anthropogenic impacts (e.g., climate change, environmental pollution)
- Assess the impacts of local anthropogenic activities (e.g., power plants, fish farms, etc.)
- Assess the establishment and spreading of non-indigenous organisms or harmful algal blooms (HABs)

1.2 Scope and definitions

Terms and abbreviations used in these guidelines are summarized in the glossary at the end of this document (see [Appendix A2](#)).

- **Taxonomic scope**

The taxonomic scope of the guidelines presented here includes mainly planktonic primary producing (photoautotrophic and mixotrophic) eukaryotic microalgae and prokaryotic cyanobacteria. Heterotrophic dinoflagellates and other heterotrophic flagellates are also included to enable studies with a wider taxonomic scope. In addition, mixotrophic ciliates such as *Mesodinium rubrum* are included since they contribute to primary production because of endosymbiotic cryptophytes. Although the focus of these guidelines is on phytoplankton, the obtained data can also be used to monitor other groups of organisms such as heterotrophic bacteria, microzooplankton or fungi, since eDNA metabarcoding provides additional information on those groups.

- **Ecosystem scope**

The target habitats of the guidelines are marine and brackish pelagic waters, but most principles also apply to freshwater and benthic habitats. For shallow and benthic habitats, the sampling method needs to be adapted and additional primers should be considered (e.g., for benthic diatoms, see [section 4.2.1](#).)

- **Methodological scope**

The approach targeted here is eDNA metabarcoding. eDNA comprises free or particle-bound DNA found in an environment and metabarcoding is an approach to identify multiple species in a complex sample (e.g., environmental or bulk sample) based on high-throughput amplicon sequencing (HTS). Although the focus is on eDNA metabarcoding, alternative methods are discussed, and many principles of this method are applicable to other molecular methods.

- **Geographic scope**

The geographical areas in focus are the Baltic Sea, the North Sea, the Norwegian Sea and adjacent sea areas. The guidelines are therefore applicable to all HELCOM and OSPAR areas.

2. Background

2.1 General introduction

Phytoplankton plays an important role in biogeochemical cycles and forms the basis of the marine food web. Plankton monitoring is used for many different purposes, such as examining ecological status, long-term changes in diversity and biomass, the effects of stressors such as eutrophication or acidification, the presence of non-indigenous species (NIS) or harmful phytoplankton species. The monitoring of phytoplankton has traditionally been performed using a light microscopy method which has remained more or less unaltered for more than six decades (Utermöhl, 1958). This method identifies and quantifies phytoplankton specimens from acidic Lugol's preserved samples using an inverted light microscope. The current monitoring guidelines of the Baltic Sea (HELCOM, 2021) recommend biomass of individual species to be estimated based on cell volume (Olenina et al., 2006) and subsequent conversion to carbon (Menden-Deuer & Lessard, 2000). The method is time consuming and requires extensive taxonomic expertise. The method has produced valuable decadal time series of qualitative and quantitative phytoplankton data in the Baltic Sea, dating back approximately 50 years. However, despite its inherent power of visualizing and confirming species, this method is limited in that it cannot always reliably resolve subtle morphological differences delineating species, and thus may not reflect true diversity of phytoplankton communities. In addition, microscopic analyses are time consuming and even strenuous on the analysts. Other limitations are artefacts due to fixation (Culverhouse et al., 2014) and problems to identify particularly small taxa, due to technical limitations of microscopy. Thus, many taxa cannot be identified to species level which compromises the assessment of biodiversity. Finally, disagreements, and misinterpretation, as well as inexperience by taxonomists (Culverhouse et al., 2003) can flaw the final assessment (Jakobsen et al., 2015).

Species identification based on a short section of DNA (DNA barcode), which requires the use of specific DNA markers (e.g. the 16S rRNA gene for prokaryotic organisms, and the 18S rRNA gene for eukaryotic organisms) provides a complement to light microscopy analyses to study marine phytoplankton and other microorganisms. Once DNA barcodes are used for automated identification of multiple species from a single sample (or from a single bulk sample containing entire organisms) the approach can be described as DNA metabarcoding (Taberlet et al., 2012, see [Appendix A2 Glossary](#) for a more detailed explanation of the terminology). DNA metabarcoding produces a large number (typically tens or hundreds of thousands) of DNA barcode sequences from a single water sample by HTS, enabling the determination of the phytoplankton community composition in the sample. The cost of DNA metabarcoding has been reduced markedly during the past decade by the development of HTS technology, and this trend is predicted to continue (Prof. Florian Leese 2021, personal communication). While the costs are highly dependent on the specifics of the analysis, generally the cost per sample is now lower for DNA metabarcoding compared to microscopy when a sufficiently large number of samples is analyzed, and workflows have been optimized.

However, DNA barcoding, too, has its limitations and further method harmonization is needed to produce comparable results. In addition, technology is evolving rapidly,

while the reference databases that control the identification (taxonomic determination) of the sequences are still continuously being improved. Sequencing of species from the relevant sea areas and reporting of sequences helps to expand international databases, e.g. the Protist Reference Database 2 (PR², Guillou et al., 2013), which is important for ensuring high quality results from gene-based environmental monitoring. By storing barcode sequences with the highest possible genetic resolution, i.e. exact amplicon sequence variants (ASVs), data can be re-annotated in the future as the reference databases are updated. For some groups of organisms, the 18S rRNA gene sequences have too low variability to resolve species. In these cases, sequencing of longer parts of the genome is required, e.g. including ITS and 28S rRNA gene region.

DNA metabarcoding can provide an estimate of species diversity and relative abundances (occurrence) but does currently not provide absolute abundance, biomass or cell number data, as different species have different numbers of marker genes in their genomes. The ability of DNA metabarcoding to accurately estimate species diversity critically depends on the taxonomic resolution of the marker, and sufficiently stringent bioinformatic criteria for removing false diversity. Absolute abundance in terms of gene copies from a taxon, cannot be inferred from metabarcoding, due to the PCR amplification step. Relative read abundance may in some cases correspond to relative abundance of biomass or biovolume, and often the most abundant taxa in terms of biomass will also have high relative read abundance. However, the relative read abundance of a taxon or species is also a function of taxon-specific copy number of the marker gene. Thus, a taxon with high copy number may be overrepresented in the reads, compared to in the biomass or biovolume. For instance dinoflagellates and ciliates are known to have very high copy number of the ribosomal operon and appear very abundant relative to other groups with lower copy number (Elianne Egge, personal communication).

A brief overview on the advantages and challenges for implementing eDNA metabarcoding into routine phytoplankton monitoring is given in **Table 1** and more extensive overviews on DNA-based methods for biodiversity assessment in aquatic environments have been written by Pawlowski et al. (2020) and Bruce et al. (2021).

While quantitative aspects of molecular phytoplankton monitoring methods are under development, some facets of phytoplankton biodiversity, such as cell size distribution and some phenotypic traits, are impossible to capture by DNA metabarcoding. DNA metabarcoding should therefore be complemented by traditional microscopy-based analyses as long as reliable and applicable molecular quantification is not possible. Another aspect is the compatibility of time series established by the different methods and data types. Microscopy-based time series can reach back decades and even centuries and statistical methods must be developed to ensure compatibility. Sufficiently extensive data sets directly comparing traditional and DNA-based methods should be collected before replacement of traditional microscopy by eDNA based phytoplankton monitoring methods can be considered. Nevertheless, complementation of eDNA metabarcoding with new imaging techniques like high-frequency imaging flow cytometry, which can provide community composition information at frequencies impossible to attain with traditional microscopy (e.g. Kraft et al., 2021) has a great potential to overcome current limitations of traditional phytoplankton methods. Although metabarcoding of single taxonomic groups (e.g. diatoms) seems to be an effective operational tool for quality assessment of rivers (Apothéloz-Perret-Gentil

et al., 2021), increased taxonomic knowledge could be gained by method complementation to develop biodiversity indicators based on community composition in the future.

Table 1. Summary of the main advantages and challenges for using eDNA metabarcoding in phytoplankton monitoring.

	Topic	Comments
Advantages	Easy to scale up	With skilled technicians / robots, hundreds of samples can be processed "simultaneously", also bioinformatics can largely be automatized.
	Cost efficient	Once a harmonized method has been taken into use and samples are collected and sequenced together by several countries
	Detection of rare biosphere	The volume of the sample processed is much larger than in traditional monitoring, thus the rare biosphere can be detected better.
	Comparable across laboratories	If standardized and harmonized methods are used, human error will have a lower impact on the identification of taxa.
Challenges	Reference libraries	Coverage must be improved.
	Erroneous reference sequences	Taxonomic and genetic-expert-annotated reference libraries are a bottle neck.
	Lack of standards	Harmonization is difficult because of rapid technological development.
	Lack of universal primers	Can be circumvented by multiple primers. However, cost per sample increases linearly with every new primer pair.

2.2 Application to EU directives

The EU Water Framework Directive (WFD; Directive 2000/60/EU, European Union, 2000) includes composition, abundance and biomass of phytoplankton as one of the biological quality elements in coastal waters, transitional waters and lakes (see sections 1.1.2, 1.1.3 and 1.1.4 of the WFD). eDNA metabarcoding adds information on the species composition of phytoplankton and can overcome some limitations of traditional monitoring methods, therefore an integration of both methods is advisable.

The Marine Strategy Framework Directive (MSFD) set out a list of 11 descriptors of environmental status for which good status must be achieved (Directive 2008/56/EC, European Union, 2008). Phytoplankton is included in descriptors 1, 2, 3, and 5 of the MSFD, which was first established in 2008 (2008/56/EC) (European Union, 2008) and updated in 2017 (2017/848/EU) (European Union 2017a,b).

The MSFD descriptor 1 (D1C6) includes monitoring the habitat's biotic structure and its functions. Molecular data can provide additional community information such as the habitat's typical species composition, a rough estimate of biomass-based relative species abundances (depending on the method), the presence of particularly sensitive or vulnerable species or species providing essential functions. These parameters may be linked to human stressors and thus additional indicators for Good Environmental Status (GES) could be developed.

The MFSD descriptor 2 (D2C1, D2C2 and D2C3) focuses on various aspects of

introduced NIS. Identification of phytoplankton using light microscopy is inherently challenging for rare and cryptogenic species and eDNA analysis as a supplementary tool in monitoring was recommended to detect NIS (Tsiamis et al., 2021). Molecular tools may provide a more precise detection of the arrival of new species (D2C1) and more timely observation of NIS turning into invasive species (D2C2 and D2C3). Moreover, molecular tools can be used to efficiently detect harmful species already when they are present in low numbers, which is relevant for the MFSD descriptor 4 (marine food webs) and descriptor 5 (eutrophication, including harmful algal blooms (e.g. cyanobacteria) in the water column).

Some phytoplankton species produce phycotoxins that may accumulate in filter feeders such as oysters and mussels or in animals of higher trophic levels, such as in eider ducks and fish. Since the phycotoxins in human food pose a health risk, there are strict regulations on monitoring of phycotoxin producing phytoplankton in areas where bivalve molluscs are harvested for human consumption (European Union, 2019). Light microscopy (Utermöhl, 1958) and fluorescence microscopy with staining of dinoflagellate theca (Andersen, 2010) are the recommended methods according to the EU guideline (Serret et al., 2019). However, given the lack of specific characteristics to distinguish many phycotoxin producing phytoplankton by light microscopy, molecular tools to identify problematic species in shellfish harvesting areas would be a solid complementation to the EU guideline by Serret et al. (2019).

2.3 Pilot studies and guidelines for eDNA phytoplankton monitoring

First steps towards implementation of eDNA metabarcoding have been taken, but currently eDNA metabarcoding is not used for harmonized routine monitoring of marine phytoplankton in Northern Europe to the best of our knowledge. Several pilot projects and studies have been carried out in recent years, e.g. Hu et al. (2016) used amplicon sequencing to study the diversity of prokaryotic and eukaryotic phytoplankton along a salinity gradient in the Baltic Sea. Karlson et al. (2018) recommended to introduce phytoplankton metabarcoding in Swedish marine monitoring programs as a complement to other methods and a project testing eDNA metabarcoding alongside traditional monitoring is still ongoing in Sweden. Initiatives on the Faroe Islands include a technical pilot study, investigating appropriate primers for amplicon sequencing (Ása Jacobsen, personal communication) and an ongoing whole year sampling study using the complementary approaches eDNA, microscopy and flowcam analysis (Fiskaaling, n.d.). Furthermore, guidelines for integrating eDNA metabarcoding in Finnish marine phytoplankton monitoring were published recently (Jerney et al., 2022).

Outside Europe, eDNA metabarcoding has been used e.g. to monitor HABs in the Sea of Okhotsk, Japan (Sildever et al., 2019), and a protocol for HAB monitoring using metabarcoding has been developed in Chile (Yarimizu et al., 2020). Another example of existing phytoplankton eDNA monitoring guidelines are the metabarcoding protocols for benthic diatoms, prokaryotic and eukaryotic freshwater phytoplankton developed for Alpine waters (Elersek, 2021). An overview on some published guidelines and protocols containing instructions for phytoplankton eDNA metabarcoding is given in the [Appendix Table A1](#).

2.4 A typical eDNA analysis workflow

The standard procedure for eDNA metabarcoding involves several steps, including sampling, processing of eDNA samples and DNA extraction to obtain DNA sequences of taxa present in those samples. These steps are depicted in **Figure 1** and can be summarized as follows:

1. Water samples are collected from the environment, followed by filtration in the field or on a research vessel. Cells, organelles, and free DNA molecules are captured on filters, which are then stored at -20°C or -80°C.
2. Extraction of DNA from samples takes place in a dedicated laboratory on land. A single optimized DNA extraction protocol is recommended to achieve good extraction efficiency.
3. Extracted DNA is purified, and specific regions of the 16S and/or 18S rRNA gene are amplified by polymerase chain reaction (PCR)
4. The PCR products are sent for HTS.
5. Sequence data are processed in a standardized workflow through a so-called bioinformatics pipeline, including several quality filtering steps.
6. The result are exact sequencing variants or ASVs.
7. These are annotated by means of matching against a specific reference database. The resulting data can be analyzed and e.g. the community composition assessed. Data should be stored at a national data host and made available to international data repositories.

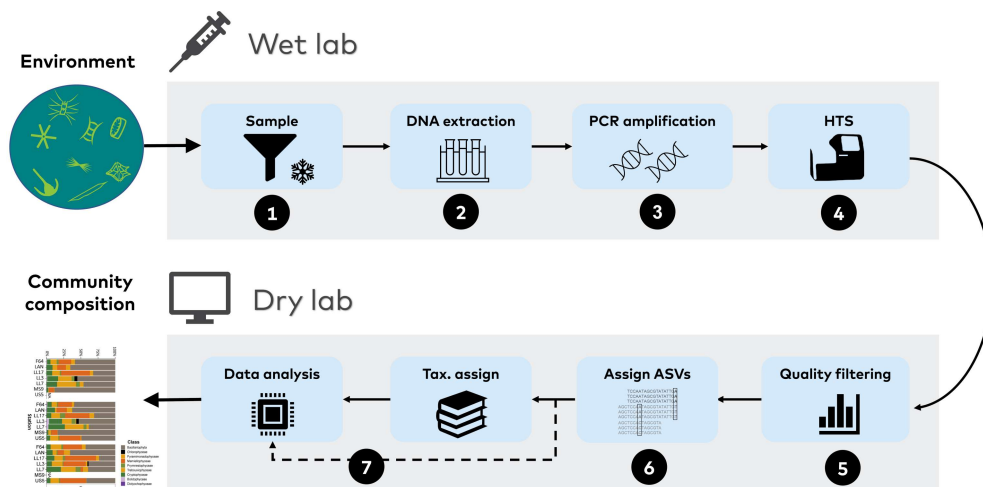


Figure 1. Description of eDNA analysis workflow. Steps 1–4 need to be carried out in a designated DNA laboratory (wet lab) and steps 5–7 are part of the bioinformatic processing (dry lab). Collection of organisms and filtration (1), DNA extraction (2), purification and PCR amplification (3) and sequencing (4), quality filtration (5), identification of ASVs (6), taxonomic assignment using reference libraries and downstream data analysis (7), adapted from Pawlowski et al. (2018).

3. Sampling and sample processing

The following section of the guidelines has been optimized for offshore sampling in productive sea areas on research vessels which are equipped with laboratory infrastructure. Thus, depending on the national monitoring programs, the available infrastructure, the targeted ecosystem and expected phytoplankton biomass, various settings need to be adjusted, or alternative methods used. A summary of recommendations is given in **Box 1** at the end of this section and further guidance can be found in the [Appendix Table A1](#), which lists examples of guidelines and protocols for aquatic eDNA metabarcoding.

3.1 Sampling

In general, sampling should be aligned with a national monitoring program to allow comparability of the genetic results with the microscopy-based results, which are attained from the conventional phytoplankton monitoring according to specific guidelines (e.g., for the Baltic Sea, see HELCOM, 2021 and for the North East Atlantic see OSPAR, 2016). Since phytoplankton shows a substantial seasonal variation (e.g. G. Hällfors et al., 1981; Gromisz & Witek, 2001; Wasmund & Siegel, 2008; Jaanus et al., 2011; Jurgensone et al., 2011), sampling needs to cover the entire growing season, which e.g., in parts of the Baltic Sea extends over the entire year.

For quantitative studies in the open sea, the minimum requirement is to take an integrated sample from 0–10 m depth using a hose (Lindahl, 1986) or by pooling equal amounts of water collected from fixed depths between 0 and 10 m using a water sampler (see Majaneva et al., 2009 for examples). The recommended sampling depths are 0–1 m, 2.5 m, 5 m, 7.5 m and 10 m. The integrated sample should be thoroughly, but gently (to avoid breaking of fragile cells) mixed in a bucket or similar container. In coastal areas, sampling depends on the water depth and local environmental conditions and should be modified accordingly, e.g. a sample from 0–1 m or an integrated sample (0–10 m) could be collected (HELCOM, 2021).

3.2 Good working practices for minimizing the risk of contamination

Methods for the detection of eDNA are optimized to discover small traces of DNA and therefore these techniques are extremely susceptible to contamination. To minimize the risk of contamination, steps need to be taken at the different stages of the sampling work. All material and equipment that is in contact with the eDNA sample must be either of single-use or cleaned to be DNA-free. The sampling equipment should always be handled with new gloves at each sampling station, and, if several field methods are used, always start with collecting the eDNA samples. This not only hinders contamination by human DNA but will also reduce cross-contamination between sampling sites.

All reusable material or equipment need to be cleaned between sampling sites. All reusable labware needs to be cleaned using acidic and alkaline detergents or rinsed with sodium hypochlorite (5–10%) and distilled water (Pawłowski et al., 2020).

Commercial products, such as 'DNA away' or 'DNA-Exitus Plus' are decontamination solutions that are safe to use in the field but are more expensive. Rinsing with ethanol or water is not sufficient to decontaminate; rinsing may reduce cross-contamination by dilution but does not break down DNA. However, it is recommended to rinse the cleaned equipment with ddH₂O after cleaning to remove any bleach or detergent before the collection of new eDNA samples (Pawłowski et al., 2020).

3.3 Filtration

Careful consideration of methods is important because the choice of filter material can create false negative detections and underestimate taxonomic richness from environmental samples. Water samples should be filtered immediately after collection, onboard the sampling vessel, ideally in a laminar flow hood in a designated laboratory. If filtration cannot be done immediately or at least within one hour, the collected samples must be kept near *in situ* temperatures or cooled to 4 °C until filtration. If samples are not filtered within one hour the time and storage conditions must be recorded. It is better to filter directly in the field and freeze the filters immediately to reduce the risk of cross-contamination and keep the community composition of the sample as undisturbed as possible. Filtration in the laboratory on land may allow large volumes to be filtered but is only feasible when the laboratory facilities can be reached within a short time, to avoid degradation of microbial eDNA or other changes in the community. Filtering in the laboratory on land also requires greater precaution, as all samples are handled at the same place and additional actions are needed to minimize cross-contamination (Pawłowski et al., 2020).

3.3.1 Filtration systems and membranes

Several filtration methods have been utilized, but no single best method has been identified (Pawłowski et al., 2020). The filtration can be done using a hand-operated, peristaltic, or vacuum pump, but importantly the filtration time must be below 60 minutes and filtration pressure should not exceed 200mm Hg / 267mbar / 27kPa (Andersson et al., 2022).

If the filtration is done in a dedicated laboratory on a research vessel, reusable open filtration units and individually packed sterile 0.22 µm pore size filter membranes (47 mm diameter) are recommended because they offer several advantages compared to closed filters (i.e. systems in which the filter membrane is enclosed within an outer housing): better DNA recovery, easier handling of the filter for mechanical cell disruption, less plastic waste, and a lower price. Djurhuus et al. (2017) compared 0.2 µm filters for environmental DNA biodiversity assessments of seawater samples across multiple trophic levels and concluded that filter membranes composed of Nitrocellulose (NC), Polycarbonate Track Etch (PCTE), Polyethersulfone (PES), and Polyvinylidene Difluoride (PVDF) can be used to obtain statistically comparable results across multiple trophic levels. Mixed Cellulose Ester (MCE) filters might yield

a more consistent community composition than PES filters, as suggested for metazoan communities (Majaneva et al., 2018). Furthermore, binding affinities of various filter membranes were different for prokaryote eDNA fragments with the polycarbonate (PC) binding the least and mixed cellulose acetate and cellulose nitrate binding the most (Liang & Keeley, 2013). Thus, MCE, PES, PCTE, PVDF, or NC filter membranes are a suitable choice.

Although closed filters were recommended for eDNA sampling from aquatic environments (Pawlowski et al., 2020; Yarimizu et al., 2020), we advocate open filters, because they allow easier mechanical disruption of cells, which is needed for improved DNA extraction from e.g. diatoms (Bruce et al., 2021). Compared to closed filters, open filters carry a greater risk of contamination, therefore filter handling must be carried out with utmost care and the use of negative controls is necessary (see [Section 4.4](#)).

3.3.2 Filtration volume

Generally, the more water is sampled, the more likely it is to detect a species. However, the amount of possible PCR inhibitors also increases with the volume of water filtered, and filtering larger volumes is logistically challenging. The sample volume filtered depends on the type of environment (Pawlowski et al., 2020) and the time of sampling. More biomass and DNA can be expected e.g. in coastal waters or during a phytoplankton bloom in spring or summer, compared to the open sea or at times of low productivity. During a phytoplankton bloom, filtering 500 mL of seawater (of a 0–10 m integrated water sample) is recommended to obtain enough DNA, without the risk of reduced extraction efficiency due to inhibiting substances. This recommendation is based on existing protocols and guidelines developed for the Baltic Sea (Andersson et al., 2022; Jerney et al., 2022). Thus, the filtration volume can be higher if low phytoplankton biomass is expected.

3.4 Sample preservation

eDNA can degrade relatively quickly in water samples or on filters due to microbial activity, therefore the filtered samples should be frozen immediately on the ship at -20°C, if the DNA is extracted within a month after sampling, or at -80°C if stored for several months before extraction. Freezing samples is also recommended because it aids mechanical disruption of rigid cells, but several freeze-thawing cycles should be avoided before DNA extraction (e.g. during the transport from a research vessel to the laboratory on land), since they can reduce the DNA quality. After the sampling DNA extraction should be carried out as soon as possible.

If immediate freezing of the filter is not possible, buffer solutions can be used, but they are more expensive and may contain hazardous substances. An overview of properties of commonly used preservative solutions with practical considerations is provided by Bruce et al. (2021).

Water samples for routine phytoplankton monitoring by light microscopy are usually immediately preserved with acidic Lugol's solution and stored in a dark refrigerator. DNA extraction from Lugol's-preserved samples is possible, but not recommended, since the effect of Lugol's solution on the phytoplankton community profiles is so far still ambiguous (Mäki et al., 2017; Jerney et al., 2022). Mäki et al. (2017) found that

acidic Lugol's solution resulted in equal DNA yields and PCR performance but affected the community profiles of a mock community, i.e., the dominance of some taxa decreased or increased. Thus, more research is needed to assess the effect of Lugol's solution on eDNA metabarcoding-derived phytoplankton community profiles before it can be used in routine monitoring.

3.5 Replicates and filtration blanks

The sensitivity of eDNA methods to contamination requires special attention on 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 for sampling and laboratory steps) to monitor contamination and cross-contamination (e.g. Santoferrara, 2019 and references therein; Pawlowski et al., 2020). In general, the level of replication should be aligned with the monitoring strategy, but at least at one station during each sampling campaign three to five independent water samples should be taken from each depth to prepare replicate integrated samples (i.e., biological replicates). In addition, technical replicates should be included annually by sequencing the same integrated sample (one biological replicate) several times to improve error-flagging. Furthermore, inter-laboratory calibration test should be planned and carried out every third year between laboratories following these guidelines to evaluate the wet lab workflow and check for contamination.

Several filtration blanks should be included in each sequencing run to check if the entire workflow from filtration to sequencing is free of contamination. During a sampling campaign (e.g. before filtering the first sample and after filtering the last sample) the same volume of Milli-Q (e.g. 500 mL) should be filtered and processed as all other samples.

Box 1. Recommendations for sampling and sample processing

Sampling

- Should be aligned with a national monitoring program to allow comparability of the genetic results with the microscopy-based results
- Integrated sample from 0–10 m depth using a hose or by pooling equal amounts of water collected from fixed depths between 0 and 10 m using a water sampler
- The recommended sampling depths are 0–1 m, 2.5 m, 5 m, 7.5 m and 10 m

Good working practices to reduce the risk of contamination

- All material and equipment that is in contact with the eDNA sample must be either single-use equipment or cleaned to be DNA-free
- All reusable material or equipment must be cleaned between sampling sites using acidic and alkaline detergents or rinsed with sodium hypochlorite (5–10%) and distilled water

Filtration

- Immediately after collection or at least within one hour, ideally in a designated laboratory onboard the sampling vessel
- If the time until filtration exceeds one hour, the collected samples should be kept near *in situ* temperatures or cooled to 4°C until filtration; temperature and storage time must be recorded
- Filtration time must be below 60 minutes and filtration pressure should not exceed 200mm Hg / 267mbar / 27kPa
- Reusable, open filtration units and sterile packed filters with a pore size of 0.22 µm and a diameter of 47 mm
- Filtering 500 mL of seawater is recommended to obtain enough DNA during a phytoplankton bloom
- The filtration volume can be increased if low phytoplankton biomass is expected and filtration time and pressure remain within the suggested limits

Sample preservation

- Filters should be frozen immediately on the ship at -20°C, if the DNA is extracted within a month after sampling, or at -80°C if stored for several months before extraction
- DNA extraction should be carried out as soon as possible

Replicates and filtration blanks

- Three to five independent integrated samples (i.e. biological replicates) should be filtered at one station during each sampling campaign
- Technical replicates should be included annually by sequencing one biological sample several times
- An inter-laboratory calibration test should be carried out every third year between laboratories that agreed to follow these guidelines
- Several filtration blanks should be included for each sampling campaign

4. Molecular work – wet lab

All laboratory work must be done following strict protocols and practices, to avoid erroneous results due to contamination. Molecular work must be carried out in a laboratory environment specifically dedicated to eDNA work. The laboratory must be equipped with infrastructure and devices devoted to the specific working steps and needs a predefined cleaning procedure. A general guidance for the molecular laboratory workflow for freshwater eDNA samples has been published by Pawlowski et al. (2020). The same principles apply to marine eDNA samples and are summarized below, followed by an overview of the general eDNA metabarcoding workflow, which comprises three main sample processing steps. A summary of recommendations for molecular work is provided in **Box 2** at the end of this section.

Best practices

- Use of dedicated working areas (especially a post-PCR room)
- Unidirectional workflow (DNA extraction room → PCR room → post-PCR room)
- Single-use material and protective clothing (lab coats and shoes dedicated to each room, gloves, hairnets, face masks, etc.)

General workflow

1. DNA extraction

All DNA molecules contained in an environmental sample (DNA of living cells and their organelles, extracellular DNA) are isolated ([Section 4.1](#)).

2. Polymerase Chain Reaction (PCR) amplification

Copies of a targeted gene region are produced in a series of enzymatic reactions by using specific primers ([Section 4.2.1](#)) and polymerases ([Section 4.2.2](#)).

3. High-throughput Sequencing (HTS)

The PCR-amplified products serve as template for massively parallel DNA sequencing producing millions of sequences ([Section 4.3](#)).

4.1 DNA extraction

Various protocols for DNA extraction exist, and they usually follow the same basic steps. In a first step cells are lysed, ideally including mechanical cell disruption (like bead-beating). Mechanical cell disruption is highly recommended, since cell walls of some eukaryotes, such as diatoms and dinoflagellates, may be resilient to harsh lytic incubation (Yuan et al., 2015). In a second step DNA is separated from the other cell components (organic and inorganic), including potential inhibitors, and in a third step the DNA is purified (Pawlowski et al., 2020).

Various methods have been used for the extraction of eDNA from aquatic environments for metabarcoding of phytoplankton (**Figure 2**). Based on a literature survey consisting of 67 unique peer-reviewed studies ([Appendix Table A2](#)), the most

common approach is the use of commercial kits such as Qiagen's DNeasy PowerWater, Blood & Tissue, and Plant kits. Other methods, such as Chelex, CTAB, phenol-chloroform, and custom extraction protocols are also used (Pawlowski et al., 2020), but less frequently (**Figure 2**). The studies comparing different extraction methods have found significant differences in the amount and quality of DNA extracted (Mäki et al., 2017; Vasselon et al., 2017; Deiner et al., 2018; Berry et al., 2019; Muñoz-Colmenero et al., 2021) and in some cases in the community composition (Djurhuus et al., 2017; but see Vasselon et al., 2017). Thus, as for the sampling, it is important to standardize the DNA extraction, to allow the use of the same PCR parameters at the amplification stage and to allow comparison of data between different data sets.

To reduce the use of hazardous chemicals (e.g. phenol and chloroform) and allow comparison between different samples, extraction by commercial kits has been suggested (Djurhuus et al., 2017). Commercial DNA extraction kits were used for testing the integration of eDNA metabarcoding in phytoplankton monitoring (Karlson et al., 2018; Jerney et al., 2022) and seem to be a simple and reliable option. Nevertheless, before eDNA metabarcoding is implemented into large-scale phytoplankton monitoring, open-source DNA extraction protocols should be developed. This would ensure independence from commercial products with unknown composition, which are sometimes only available for a limited time, and allow long-term data consistency. Ideally, the chosen extraction method is time and cost-efficient, open-source, results in a high DNA yield for all relevant taxa, does not contain hazardous substances and creates a minimum amount of waste.

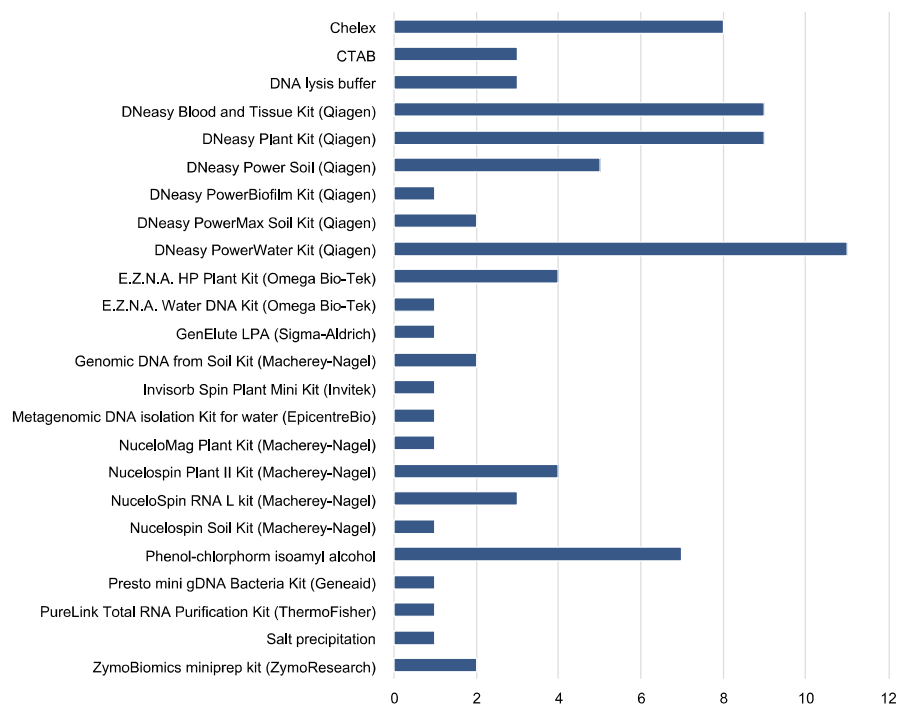


Figure 2. Overview of the DNA extraction methods (in alphabetical order) based on 67 peer-reviewed published papers focusing on environmental DNA from aquatic environments, targeting phytoplankton and other organisms (references given in the [Appendix Table A2](#)). X-axis shows the number of studies using a particular kit. Y-axis shows various DNA extraction methods.

4.2 PCR amplification

The amount of target DNA present in an environmental sample is usually relatively low. Therefore, the DNA barcode region of interest needs to be multiplied by PCR before sequencing. During the PCR, the number of amplified DNA fragments, also called amplicons, increases exponentially, producing enough material for sequencing (Pawlowski et al., 2020). How well the PCR amplifies target DNA depends on many factors, like the choice of primers ([Section 4.2.1](#)) and the polymerase ([Section 4.2.2](#)) or the presence of inhibitors ([Section 4.2.3](#)).

4.2.1 Primers

The detection of species in eDNA samples strongly depends on the choice of PCR primers. For assessing the community composition of phytoplankton, primers should ideally be generic enough to amplify all taxa of interest, but at the same time specific enough to distinguish species. Choosing a single primer that fulfills both requirements is challenging, since phytoplankton comprises several, distant taxonomical groups belonging to two kingdoms of life – eukaryotes (i.e. most phytoplankton groups) and prokaryotes (cyanobacteria). Of the two main types of barcoding markers, protein-coding genes (e.g. ribulose-bisphosphate carboxylase, *rbcL*) can ensure very good identification at the species level and seem to be a promising option for benthic freshwater diatoms (Pawlowski et al., 2018; Apothéloz-Perret-Gentil et al., 2021). The second type of markers are ribosomal genes (e.g. 18S, 16S, 12S, ITS, 23S, 28S), which consist of an assortment of variable and conserved regions, allowing the synthesis of both highly specific primers and more universal primers. Primers of the latter type are commonly used for eukaryotic and prokaryotic phytoplankton (e.g. de Vargas et al., 2015; Hu et al., 2016; Kim et al., 2019; Bolaños et al., 2020; Salmaso et al., 2020; Malashenkov et al., 2021).

The diversity detected by the eDNA metabarcoding approach can be influenced by several factors ranging from sampling to bioinformatic analysis (Lindahl et al., 2013). One such factor is the gene and region targeted, exemplified by the differences in species diversity detected by the different markers (Bucklin et al., 2016; Tanabe et al., 2016; Gran-Stadniczeňko et al., 2017; Smith et al., 2017; Günther et al., 2018; Forster et al., 2019; Sildever et al., 2019). Thus, the usage of multiple markers and primer pairs to detect higher diversity (Smith et al., 2017; Stat et al., 2017; Alberdi et al., 2018; Zhang et al., 2018; Choi & Park, 2020) and for describing new species (Ott et al., 2022) has been suggested.

Different regions of the rRNA 18S gene have been targeted in a number of studies investigating eukaryotic diversity using the eDNA metabarcoding and HTS approach (Monchy et al., 2012; Massana et al., 2015; Hu et al., 2016; Nagai et al., 2016; Tanabe et al., 2016; Majaneva et al., 2017; Moreno-Pino et al., 2018; Bucklin et al., 2019) and there are several universal 18S primers for eukaryotes suitable for HTS (e.g. Amaral-Zettler et al., 2009; Stoeck et al., 2010; Hadziavdic et al., 2014; Hugerth et al., 2014; Tanabe et al., 2016), but there is no agreement on the best choice.

Based on a literature survey of 67 peer-reviewed papers on metabarcoding in aquatic environments, primers targeting different regions and genes of ribosomal RNA have been commonly used for targeting phytoplankton (**Figure 3**, [Appendix Table A2](#)). Among those, the majority of the unique primer pairs detected target the V4 region of the 18S rRNA gene (**Figure 3**), while the universal eukaryotic primers

that target the 18S rRNA gene V9 region (Amaral-Zettler et al., 2009) were used in the highest number of studies (29% of the 38 studies utilizing the 18S rRNA gene). At the same time, there are only a few sets of primers targeting the 28S rRNA gene, 23S rRNA gene, internal transcribed spacer (ITS), *rbcL* gene, and cytochrome c oxidase subunit 1 (COI). In the case of the 16S rRNA gene, most of the unique primers target the V3-V4 or V4 regions. In a few studies, group-specific primers were also used for targeting diatoms, dinoflagellates or cyanobacteria. Since no primer pair can target all protists equally well, an informed primer choice is necessary, which can be supported by using the R-based web application pr2-primers (Vaulot et al., 2022) and a primer evaluation study carried out by Latz et al. (2021).

Improved detection of species can be achieved by utilizing multiple markers (Stat et al., 2017; Sawaya et al., 2019; Choi & Park, 2020; Sildever et al., 2021) and could comprise a set of lineage specific primers, as suggested by Pawlowski et al. (2012). Yet, since sequencing costs increase for each additional set of primers, it might still be too costly to use a lineage-specific multi primer approach.

To detect prokaryotic and eukaryotic organisms it is recommended to use at least two primer pairs, which have a high coverage in databases. We recommend using 18S V4 primers for eukaryotes, since they have been widely used to study eukaryotes, comparative studies between light microscopy exist and they are relatively well covered in databases, although euglenophytes and haptophytes are currently underrepresented in databases. We recommend using either primers by Balzano et al. (2015) which were found to be a good compromise by Latz et al. (2022), who carried out an evaluation of primers, or the primer pair by Stoeck et al. (2010) which is most commonly used, according to Vaulot et al. (2022). Unfortunately, neither recommended primer pair amplifies the group Excavata well, and thus only around 70% of the Eukaryota sequences which are part of the reference database PR² would be amplified (tested using the option "Test your primer set" of the web application pr2-primers (<https://app.pr2-primers.org/pr2-primers/>) by Vaulot et al., 2022). Furthermore, for prokaryotes we recommend using 16S V3-4 rRNA gene primers, which cover a longer region and will render higher taxonomic accuracy. The suggested primers have been extensively used, e.g. to study phytoplankton (Bennke et al., 2018) and bacteria (Bunse et al., 2016) in the Baltic Sea, were judged to be the most promising in an evaluation by Klindworth et al. (2013) and are recommended in Illumina's 16S library protocol (Illumina, 2020). An extended discussion on primers is included in the [Appendix A1](#).

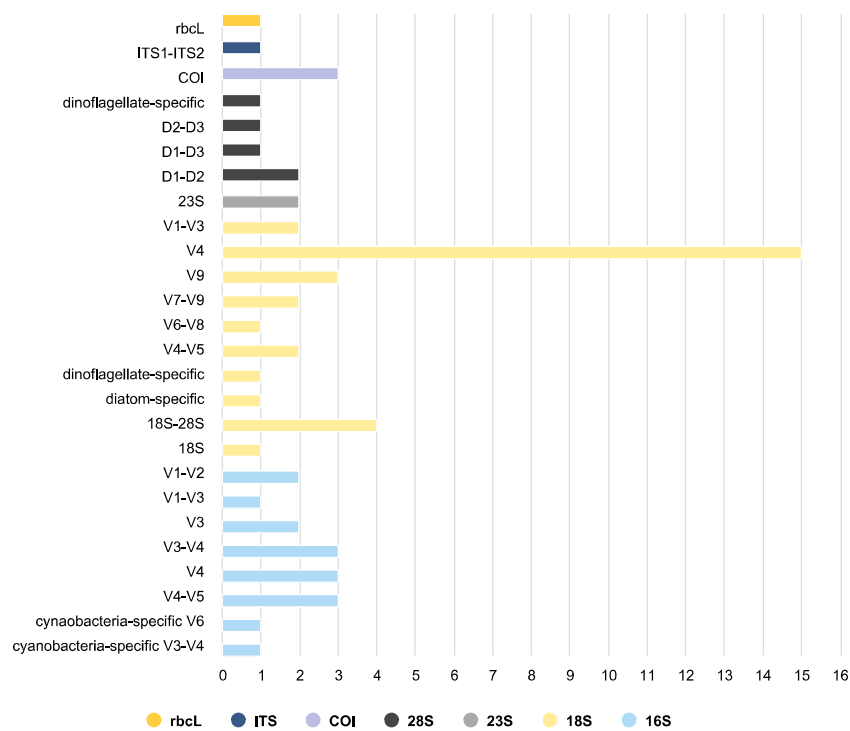


Figure 3. Overview of the number of unique primer pairs targeting a specific gene and region based on a literature survey of 67 peer-reviewed studies (references given in the [Appendix Table A2](#)).

4.2.2 Polymerases

Although PCR amplification is a very useful method to increase the signal of metabarcodes, it is also the main source of technical errors during the metabarcoding workflow (Berney et al., 2004; Aird et al., 2011). Thus, PCR amplification can have a notable influence on the diversity and the relative sequence abundances of eDNA samples (Haas et al., 2011; Brandariz-Fontes et al., 2015; Kelly, et al. 2019). In an inter-laboratory experiment the choice of the polymerase had a consistently significant effect on the variability in metabarcoding data (Zaiko et al., 2021) and it can influence HTS data in terms of PCR-related errors (Quail et al., 2012; Gohl et al., 2016), chimera formation (Lahr & Katz, 2009), species occurrence and relative sequence abundances (Haas et al., 2011; Brandariz-Fontes et al., 2015; Oliver et al., 2015; Nichols et al., 2018; Kawato et al., 2021), community composition and quality of HTS data (Sze & Schloss, 2019). Nagai et al. (2022) compared PCR product error profiles of 14 different PCR kits using a eukaryotic DNA mock community and found that kits containing specific polymerases (KOD Plus Ver.2 at 56°C, KOD plus Neo, and HotStart Taq DNA polymerase at the annealing temperature of 65°C) produced better results associated with sequence bias and species identification.

In general, the usage of high-fidelity DNA polymerase has been suggested for reducing erroneous sequences obtained by HTS (Sze & Schloss, 2019), although other factors (e.g. diversity and PCR conditions) can further influence the results. We recommend using a polymerase that has been tested together with the chosen primer pair and using it over extended time periods to ensure long-term data consistency.

4.2.3 Inhibitor removal

False negative PCR results are often a consequence of inhibition by chemicals and compounds present in the sample that interact with the PCR. Thereby they can reduce the PCR efficiency or cause it to fail completely, even when target DNA is present (Bruce et al., 2021). Well-known PCR inhibitors in water samples are complex humic substances (Stoeckle et al., 2017). It is important to test the extracted DNA for inhibition to avoid false negative results (Bruce et al., 2021), especially when high amounts of humic substances are expected, as in the Baltic Sea. If DNA samples contain too many impurities that inhibit PCR amplification the samples can be pretreated using chemicals, specific kits, or equipment (Tedersoo et al., 2022), but ideally the additional processing steps are reduced to a minimum to improve comparability between pretreated and untreated samples.

4.3 High-throughput amplicon sequencing (HTS)

During the past two decades, the evolution of technical advances of DNA sequencing has been rapid. The low-throughput base-by-base Sanger sequencing (Sanger et al., 1977) has evolved to high-throughput of second-generation short-read and third generation long-read sequencing technologies (Heather & Chain, 2016). Currently, the second-generation short-read sequencing technologies (i.e., Illumina MiSeq and NovaSeq) are most used in DNA metabarcoding studies. Of the second-generation technologies, the most cost-effective and suitable for detecting different species from communities is Illumina NovaSeq, which allows including up to 700–800 samples into one sequencing run and still results in millions of reads per sample. Illumina NovaSeq also outperforms Illumina MiSeq in detecting more DNA sequence diversity within samples even at exactly the same number of reads per sample (Singer et al., 2019). There is, however, a trade-off in the length of sequences obtained by NovaSeq and MiSeq, the former having 2x250 bp and the latter 2x300 bp sequencing lengths. Recently, Illumina also released a new kit for their NextSeq instrument, allowing 2x300 bp sequencing, which gives significantly more data per cost compared to MiSeq (Anders Andersson, personal communication). Thus, NextSeq can be considered as alternative to MiSeq if the new kit allowing sequencing 2x300 bp is used. A comparison between the most commonly used sequencing platforms is available through the Illumina webpage (Illumina, n.d.). The small difference of 50bp becomes critical when choosing primers – for example, the recommended 16S V3-4 rRNA gene primers for prokaryotes (Herlemann et al., 2011) are amplifying a fragment just too long for NovaSeq but suitable to be sequenced with MiSeq (Latz et al., 2022). In the future, this issue may be circumvented since the quality of third-generation long-read sequencing technologies has increased substantially (Callahan et al., 2021), will gain momentum and may replace second-generation short-read technologies. The definite benefit of long reads is the increased phylogenetic signal available for species identification. For the time being, we recommend using Illumina NovaSeq, MiSeq or NextSeq but urge for researching the possible benefits of changing to third-generation long-read sequencing technologies for increased phylogenetic resolution.

4.4 Quality control

4.4.1 Negative controls

In addition to the filtration blank ([Section 3.5](#)) one empty extraction tube (i.e., without a sample) should be processed as control, identically to regular samples, according to the used protocol to check for contamination during the DNA extraction.

4.4.2 Positive controls

An artificial phytoplankton community of known composition (mock community) can be used as positive control. Including a mock community to evaluate laboratory steps and adjust bioinformatic pipelines is highly recommended (Santoferrara, 2019). For this purpose, 5–10 taxa with known sequences (acquired e.g. from a culture collection) should be cultured to reach exponential growth and sufficiently high cell numbers. The mock community should include prokaryotes and eukaryotes as well as taxa representing different phytoplankton groups, that are easy to distinguish by light microscopy. The cell density of each culture should be assessed by light microscopy and an aliquot of each culture mixed to yield a cell density of ~100,000 cells L⁻¹ in the mock community. The mock community should be filtered and processed as all other samples. Additionally, single cultures of interest (e.g. harmful algae or invasive species) can be filtered and sequenced as positive controls in different cell concentrations to get an estimate of the detection limit.

In addition, positive controls consisting of known DNA samples, or even synthetic sequences reflecting the attributes of the targeted products should be sequenced along the biological samples (Sepulveda et al., 2020).

4.4.3 Tagging system controls

Due to the large number of samples usually involved in DNA metabarcoding studies, it is advisable to implement a tagging system, where each sample displays a unique combination of forward and reverse tags (Taberlet et al., 2018). By including unused tag combinations in the experimental design it is possible to detect and filter out chimeras (Taberlet et al., 2018).

Box 2. Recommendations for molecular work – wet lab

DNA extraction

- Commercial extraction kits should be used which
 - Are available for at least 5 years continuously
 - Include mechanical cell disruption
 - Remove inhibitors
 - Contain a reduced amount of hazardous substances
 - Produce as little waste as possible

Primers

- Eukaryotes: 18S rRNA gene region V4
 - V4_Balzano_F 5'-CCAGCASCYGC GGTAATTCC-3' (Stoeck et al., 2010)
 - V4_Balzano_R 5'-ACTTTCGTTCTTGATYRR-3' (Balzano et al., 2015, adapted from Stoeck et al., 2010)
- Prokaryotes: 16S rRNA gene region V3-4
 - Forward: 341F, 5'-CCTACGGGNGGCWGCAG -3' (Herlemann et al., 2011)
 - Reverse: 805R, 5'-GACTACHVGGGTATCTAATCC -3' (Herlemann et al., 2011)

Polymerases

- High-fidelity DNA polymerase
- Using a polymerase that has been tested together with the chosen primer pair
- Using the same polymerase over extended time periods for data consistency

Inhibitor removal

- Inhibition testing when a high load of humic substances is expected
- Use commercial extraction kits which are optimized for inhibitor removal

High-throughput sequencing

- Illumina NovaSeq, MiSeq or NextSeq paired-end sequencing
- Sequencing depth of 100.000 reads/sample

Quality control

The following controls should be included and processed as all other samples

- Negative control
 - Filtration blank
 - Kit control
- Positive control
 - Mock community
 - Single culture in two different concentrations (10^3 and 10^6 cells L⁻¹)

5. Bioinformatics - dry lab

The bioinformatic workflow comprises four main steps (**Figure 4**) and the most important recommendations are summarized in **Box 3**.

1. **Quality-filtering** – Amplicon sequences of low quality and/or with ambiguous bases are removed. The paired-end sequences are merged into a contiguous full-length sequence.
2. **Detecting ASVs** – Individual variants of sequences are defined by algorithms using complex error reduction models and potential chimeras are removed; this step can be performed together with step 1.
3. **Taxonomic assignment** – ASVs are compared to a reference database and assigned to taxa depending on their sequence similarity or other criteria.
4. **Data analysis** – The list of ASVs is used to determine the taxonomic composition of each sample and their relation to environmental variables.

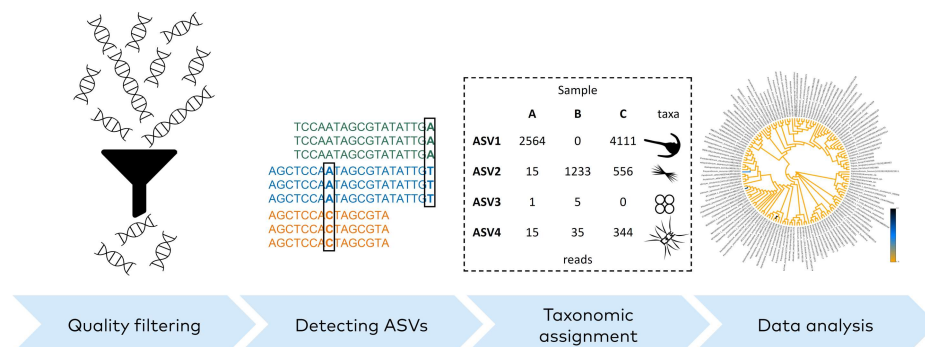


Figure 4. Workflow of HTS data analysis adapted from Pawlowski et al. (2020).

5.1 Sequence quality filtering and clustering

The first step of sequence handling and curation is to determine the global quality of the sequencing run. If the overall quality is not sufficient, a resequencing of samples should be considered. The software FastQC (Babraham Bioinformatics) or MultiQC (Ewels et al., 2016) can be used to visualize and evaluate the quality.

As second step the raw sequences can be processed with widely used pipelines, which usually integrate various software components and algorithms to separate the sequences of different samples (demultiplexing), do quality filtering, identify either ASVs or cluster similar sequences into OTUs and assign taxonomy to the sequences ([Section 5.2](#)). Performing quality control is also referred to as “denoising” and usually includes the removal of short reads, primers, chimeras, and paired-end read joining. In the past years various bioinformatic tools have been developed to

handle metabarcoding data and identify ASVs or OTUs, but due to improvements in reusability, reproducibility, and comprehensiveness (Callahan et al., 2017) we recommend using ASVs instead of the formerly used OTUs.

The pipeline DADA2 (Callahan et al., 2016) is widely used to identify ASVs, either as an R integration, which simplifies downstream data processing and analysis, or as part of the software QIIME2 (Bolyen et al., 2019). In addition to QIIME2, Mothur (Schloss et al., 2009) is frequently used for microbial bioinformatics. Further commonly used algorithms are vsearch (Rognes et al., 2016) and obitools (Boyer et al., 2016).

5.2 Assigning taxonomy

Hleap et al. (2021) categorized taxonomic assignment approaches into four strategies: sequence similarity, sequence composition, phylogenetic, and probabilistic techniques. The sequence similarity methods directly compare obtained reads to reference sequences using global or local alignments, i.e., BLAST (Altschul et al., 1997). The sequence composition methods build a model that links nucleotide frequency patterns extracted from obtained reads to specific taxonomic groups, a widely used example being the RPD classifier (Wang et al., 2007) that uses a naïve Bayesian method to do this. Phylogenetic methods reconstruct a phylogenetic tree from the 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). Probabilistic techniques employ a probabilistic framework to assess the probability of correctly placing the obtained reads to taxonomic levels. An example is PROTAX (Somervuo et al., 2016) which is based on a multinomial regression model using any kind of read similarity measures or outputs of classifiers as predictors.

Taxonomic assignment of obtained good-quality reads to higher-level ranks (phylum, class, order) is relatively straightforward in any of the approaches mentioned above since the used marker genes (here, 16S and 18S rRNA genes) have enough variability to assign reads reliably. Assigning reads to lower-level ranks (family, genus, species) has remained more difficult. This is because (1) there is limited variability in the short reads, (2) the reference databases are incomplete, with missing taxa and limited within-taxon sampling, and (3) widely applied methods for taxonomic assignment lack a proper assessment of identification reliability (Somervuo et al., 2017). To tackle the first hurdle, it is recommended to shift to third-generation long-read sequencing technologies, which can produce reads covering the whole 16S and 18S rRNA genes, enabling more accurate lower-level rank identification of the target organisms (Callahan et al., 2021). For the second issue, the remedy is to populate reference databases with the missing species and to add more specimens per species. The third problem can be improved by including methods that assess reliability of taxonomic assignment.

It is often assumed that using more complex methods to assign taxonomy always yields better results, but Hleap et al. (2021) demonstrated that sequence similarity and sequence composition methods outperformed phylogenetic and probabilistic methods tested in their study. They highlighted that curation of reference databases and tuning the assignment parameters are of paramount importance for all taxonomic assignment methods. Further, probabilistic and phylogenetic methods will

be very informative if carefully curated databases are available since they offer extra information, i.e., in the form of probability of finding a given species if it is known to be present in the study area (Axtner et al., 2019), and evolutionary history (Czech et al., 2022).

For monitoring phytoplankton in the HELCOM and OSPAR areas we suggest using sequence similarity and sequence composition methods since no well-curated reference databases for probabilistic methods or reference trees for phylogenetic methods exist yet. The main problem is missing species in databases. One option to get a reliability assessment of the obtained taxonomic assignments is to follow a protocol presented in Åström et al. (2022). First, they suggest combining results from different assignment approaches in order to identify possible misidentifications. This can be done by comparing results of sequence composition methods with sequence similarity results; if sequence similarity (BLAST) against sequence composition reference database (RDP classifier database) is low, the read might be mislabeled. Second, they suggest rating the assignments by giving them a confidence rating – high, moderate or low confidence of correct species assignment. Assignment gets a high confidence if it is assigned to species level, if >80% of the study area's species in the genus are in the database, and if the marker distinguishes at species level within the genus. Assignment confidence is moderate if the marker distinguishes at species level within the genus, but it is not assigned at species level, or if <80% of the study area's species in the genus are in the database. Species assignment is of low confidence if the marker does not distinguish at species level within the genus. Another option is to use sequence similarity methods that employ the lowest common ancestor (LCA) strategy to assign taxonomy. LCA-based approaches use several BLAST hits to determine the taxonomic level into which the assignment is placed (Huson et al., 2016) possibly revealing presence of species at higher taxonomic levels when a genus or species identification is not possible due to missing reference sequences. In the case of mislabeled references, the LCA method places the obtained read on a higher taxonomic rank where the BLAST matches converge.

5.3 Reference databases

Accurate taxonomic classification of obtained reads is critical to make DNA-based monitoring methods relevant for management. The taxonomic assignment methods have some impact on the accuracy (Hleap et al., 2021) but the coverage and quality of the used reference database/library are of utmost importance (Ekrem et al., 2007). The most comprehensive database is the NCBI GenBank (Benson et al., 2013) that is built through submissions from individual researchers. It is not restricted to any given gene sequence, and it offers a suite of BLAST programs to perform sequence similarity searches against it. Being an all-inclusive database without external taxonomic curation, the GenBank has a share of mislabeled sequences. However, it has been shown that identifications in the GenBank are surprisingly accurate (Leray et al., 2019). Nevertheless, a sequence gap analysis carried out by Salmaso et al. (2022) showed that, in the identity range of 99–100%, approximately 30% (plankton) and 60% (biofilm) of freshwater prokaryotic and eukaryotic microalgae sequences did not find any close counterpart in the NCBI GenBank.

Similarly, a taxonomic gap analysis carried out by the same authors showed that approximately 50% of the cyanobacterial and eukaryotic microalgal species identified by light microscopy were not represented in NCBI GenBank.

To counteract the mislabeled references in the GenBank, several expert-curated and/or gene-specific databases have been created. In the phytoplankton context, the most relevant ones are SILVA (Quast et al., 2013), PR² (Guillou et al., 2013) and GTDB (Parks et al., 2022). SILVA provides comprehensive, quality-checked and updated datasets of aligned 16S, 18S, 23S and 28S rRNA gene sequences for Archaea, Bacteria and Eukarya. The PR² database provides carefully curated and annotated 18S rRNA gene sequences of mainly single-celled eukaryotes, but also sequences from multicellular animals, fungi and plants as well as a limited set of 16S sequences from plastids and bacteria. The PR² database also includes comprehensive metadata fields. Both the SILVA and PR² databases have been translated for and are available for most of the taxonomic assignment approaches, making their usage in automated processes easy. Unlike SILVA and PR², GTDB is not gene-specific but is based on complete prokaryotic genomes (bacteria and archaea), sourced from the NCBI Assembly database. It can accommodate isolate genomes and the tens of thousands of metagenome-assembled genomes (MAGs) now being obtained also from environmental samples and provides a framework for standardized prokaryotic taxonomy.

An interesting and promising addition to the expert-curated reference libraries is the *µgreen-db* (Djemiel et al., 2020), which is a database for the 23S rRNA gene of both eukaryotic plastids and cyanobacteria. This marker gene would allow usage of a single primer pair (Sherwood & Presting, 2007) and a reference library to identify both prokaryotic and eukaryotic phytoplankton species. However, applying it in phytoplankton monitoring suffers greatly from the low coverage of the reference library. Another somewhat phytoplankton-relevant reference library is the *Diat.barcode* (Rimet et al., 2019), which is a curated library for the chloroplast gene *rbcL* but it contains mainly benthic diatoms. To make the reference libraries better and more usable for phytoplankton monitoring, we strongly encourage including as many different marker genes as possible in the future barcoding efforts. This will of course increase the cost of such work (see [Section 8](#)) but it will inevitably come with huge benefits in the future in the form of better coverage and usability of all the different reference libraries.

We recommend using a publicly available, comprehensive and curated database (SILVA, PR² or GTDB) in order to make the data and processes towards the final results as accessible and open as possible.

Box 3. Recommendations for bioinformatics

Sequence quality filtering and clustering

- Amplicon sequencing variants (ASVs) should be used
- The pipeline DADA2 is recommended as R implementation

Assigning taxonomy

- Sequence similarity (i.e. BLAST), and
- Sequence composition methods (e.g. RPD classifier)

Reference databases

- Publicly available, comprehensive and curated database (SILVA, PR² or GTDB)

6. Data storage and management

Across the globe, organizations conducting environmental monitoring are becoming increasingly committed to the principles of open data and open science, according to which both raw data and processed results should be distributed as freely as possible unless there are valid grounds for their restriction (e.g. legal issues or concern for the persistence of threatened species). Moreover, it is widely agreed that ideally monitoring data should be FAIR – Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016). To promote the wide usability and impact of monitoring data in the society, we recommend adopting these aims also for eDNA-based phytoplankton monitoring.

It is a commonly followed good practice to deposit raw sequence data in an international database, primarily one of the centers of International Nucleotide Sequence Database Collaboration (INSDC), which is a long-standing foundational initiative that operates between several other databases (DNA data bank of Japan, DDBJ, European Nucleotide Archive, ENA and the National Center for Biotechnology Information, NCBI). However, to make the DNA-based observations findable and accessible outside the expert community, and to enable their integration with traditional monitoring data, it is also important to mainstream their inclusion in biodiversity databases. Apart from some national platforms (e.g. the Atlas of Living Australia), both the Global Biodiversity Information Facility (GBIF) and the Ocean Biodiversity Information System (OBIS) now accommodate DNA-based observations (Andersson et al., 2020; Nilsson et al., 2022). These observations can be uploaded using the existing Darwin core data standard for biodiversity observations, thus making publication of DNA-derived occurrence data compatible with GBIF practices. DNA-based observations can be included at any taxonomic level, recognizing that their present taxonomic resolution is limited by current identification methods and reference databases. As long as the DNA sequence (ASV) is included in the metadata of each observation, it will be possible to update the identification in the future as the methodology and reference libraries improve. We recommend publishing raw sequence data in INSDC and quality-controlled observation data in GBIF with links between these databases. Alternatively, data should be stored at the European Marine Observation and Data Network (EMODnet) or the National Oceanographic Data Centres (NODC), which are structural elements of the International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission of UNESCO, or similar that already handle phytoplankton data.

In addition to standard metadata related to phytoplankton samples, eDNA metadata should include all relevant information about wet and dry lab procedures, as summarized in **Box 4**. An example data format named "PlanktonBarcoding" is available from the Swedish Oceanographic Data Centre (see example for year 2013 at: <http://sharkweb.smhi.se>).

Box 4. Recommendations for data management

FAIR principles

- eDNA monitoring data should be FAIR (Findable, Accessible, Interoperable and Reusable, <https://www.go-fair.org/>)

Publishing

- Raw sequence data should be published in INSDC or submitted to ENA
- Quality-controlled observation data should be published in GBIF with links between these databases

Metadata

- Detailed metadata should be provided, including information about
 - Replicate
 - Number of replicates
 - Filtered volume
 - Type of filter
 - DNA extraction kit
 - Target gene (e.g. 16S rDNA, 18S rDNA)
 - Primer (forward, reverse, reference)
 - Amplicon Sequence Variant id (identifiable in raw data)
 - Sequence
 - Sequencing method and instrument
 - Number of reads
 - Relative abundance (reads for this ASV divided by the total number of reads for all ASVs)
 - Annotation confidence
 - Spike in (if applicable)
 - Bio-informatics pipeline (version and date)
 - Reference databases (version and date)
 - Latest re-analyses (date)

7. Future perspectives and recommendations

7.1 Standardization

To assure the uptake of the described recommended procedures, the creation and use of internationally agreed upon standards is necessary. Already existing structures in the European standardization organization CEN (i.e. working groups in its Technical Committee (TC) 230) and the international standardization organization ISO (its TC 147) should be preferred endpoints to process the recommendations described in [Section 3](#) and especially [Sections 4](#) "Molecular work – wet lab" and [5](#) "Bioinformatics - dry lab". In addition, minimum criteria should be defined also for metadata to improve comparability and use of data for large geographical assessments ([Section 6](#)).

If such steps towards standardization are not taken, there is a serious risk that individual nations will develop independent procedures and in doing so, reduce the comparability of results on a transnational scale. As the field of molecular methods is fast evolving, focus should be put on defining minimum criteria that methods and results must fulfill to attain high levels of result accuracy and constancy. Defining only the minimum requirements will not prevent innovation in method development but will assure interoperability and backward comparability despite probable advances in methodology.

Great care should be given to assure inclusivity with regards to stakeholders and individual countries. Inclusivity in involvement during the process of establishing minimum criteria to be standardized will better account for the great variety in the gradient of technical and budget constraints across many European countries. Thus, open platforms and online or hybrid meetings are to be preferred to purely physical in-person ones. Lastly it will be important that central international entities (organizations working under the EU commission or UN) will be involved in co-creatively setting the goals of the standardization process to assure their speedy uptake. This is not to say that such entities should be pursuing the task of standardization itself, which should be left to the dedicated standardization bodies of ISO and CEN.

7.2 Sequence reference databases

For the accurate molecular identification of phytoplankton species, the completeness and accuracy of sequence reference databases is crucial. Reference sequences of interest, e.g. of ribosomal RNA genes, relevant for a certain geographic area need to be part of the reference database for ASVs to be reliably assigned to taxa. Thus, it is recommended to check if the public sequence reference databases contain taxa of interest for a particular geographic area, e.g. for the Baltic Sea, the "Checklist of Baltic Sea Phytoplankton Species" (G. Hällfors, 2004) can be used and compared to the reference database. If they are not present, or only few sequences

(maybe even from geographically remote locations) are part of the sequence reference database, full-length sequences of the target organisms (as outlined in [Section 7.4](#)) isolated from the area of interest should be used to populate reference databases. Isolating and sequencing organisms from the geographic area and environment of interest can improve taxonomic assignment. It is also possible to use strains from culture collections in the region, in northern Europe, like the FINMARI CC (FINMARI Culture Collection/SYKE Marine Research Centre and Tvärminne Zoological Station), NORCCA (The Norwegian Culture Collection of Algae), or GUMACC (Algal Bank at the University of Gothenburg).

7.3 Phytoplankton groups to target with eDNA-metabarcoding

The implementation of eDNA-based methods into traditional phytoplankton monitoring would be most beneficial for members of phytoplankton groups that are difficult or impossible to identify by light microscopy of Lugol's preserved samples. Several common phytoplankton groups which can occur in high abundance are, e.g. the cyanobacterial orders Chroococcales, Oscillatoriales, and Synechococcales, the dinoflagellate order Gymnodiniales (*sensu lato*), cryptophytes (Cryptophyta) and haptophytes (Haptophyta). The diversity of those groups remains largely hidden using routine monitoring methods. Importantly, many potentially harmful species (e.g. S. Hällfors, 2007; Karlson et al., 2021) and potential invasive non-indigenous species (e.g. HELCOM, 2015) cannot be identified in conventional phytoplankton monitoring. Therefore, these groups should be prioritized when improving reference databases to gain the most benefits from eDNA-based phytoplankton monitoring.

7.4 Long-read sequencing

Improved detection of species can be achieved by long-read sequencing (third-generation sequencing) covering the full length of the gene or genes of interest, e.g. ribosomal RNA genes (Schlaeppli et al., 2016; Singer et al., 2016; Heeger et al., 2018; Tedersoo et al., 2018). As an example, the amplification of a full-length 18S rRNA gene yielded more unique sequences compared to sequencing only the V4 region (Latz et al., 2022). However, usage of longer sequences still has some limitations, e.g. the cost per base pair is high, technical issues (variability in the amplicon lengths, potentially leading to the overrepresentation of shorter amplicons; Tedersoo et al., 2018), increased potential for chimera formation by amplifying long fragments, no publicly available bioinformatic pipeline with the denoising of long-read sequences, and a lack of long reads in the public databases (Latz et al., 2022). Those limitations need to be overcome before long-read sequencing can be used for phytoplankton monitoring. Thus, we recommend obtaining full-length sequences of the genes of interest from the target organisms using Sanger sequencing to populate reference databases, but not for the actual monitoring.

8. Cost estimates

Conducting ecosystem monitoring comes with costs for e.g. sampling, sample processing and reporting. The optimal monitoring program design is a trade-off between maximal sample coverage and minimal cost. DNA-based methods in phytoplankton monitoring will bring benefits (see above sections) and presenting realistic cost estimates facilitates inclusion of the methods in routine national monitoring programs.

In their cost-efficiency report, Fu et al. (2021a) identified three steps in eDNA-based surveys – sampling, filtering, and DNA extraction – which are universal for such surveys independent of target organisms. Then, they reviewed literature and compiled an exhaustive list of supplies that are used in eDNA-based surveys, assessed the cost of each supply, and compared the costs of hypothetical fish surveys based on eDNA and conventional methods. However, they reported only costs related to buckets (sampling), filters (filtration), DNA extraction kits (DNA extraction) and some labor costs. This information is not sufficient for gaining a holistic view of costs required for establishing international DNA-based phytoplankton monitoring. Fu et al. (2021a) estimate that their hypothetical fish eDNA-survey would require 30–42% of the cost of conventional survey, using expert personnel. The main cause for the difference in the costs are labor needs, the eDNA survey requiring fewer hours than a survey based on conventional methodology in fish monitoring. In the case of phytoplankton monitoring, however, a fraction of the DNA-based monitoring costs (e.g., sample processing and data analysis) would be added to conventional monitoring since these are not compensatory but complementary, meaning that in an ideal monitoring, both conventional and DNA-based monitoring would be included. Thus, expensive ship time, for example, would be a shared cost, and sampling and preparation of an integrated sample would be shared work time.

Another recent cost estimate report (Magnussen & Navrud, 2021) considered not only collecting monitoring data using DNA-based methods but also benefits, pre-requisites and improvements of such work. Their method was based on asking experts conducting different DNA-based projects for costs related to their work. According to their estimates, the average costs associated with collecting, securing, and making available reference data, i.e., building and maintaining DNA reference libraries, are approximately 1100 € per species. They estimated that the already existing reference library for Norwegian species has cost over 46 million euros, and to include all remaining organisms reported to occur in Norway would still cost approximately 26 million euros. Magnussen and Navrud (2021) highlight that this is an average and the cost of each individual species will vary a lot depending on how difficult it is to obtain a sample of the species; the most expensive species are the ones in aquatic environments, particularly marine deep-water species that can cost up to 15 000 € to be included in a reference library. Collecting and making reference data available is a one-off cost (investment), but they estimate that storing reference material, both as physical and digital objects, costs approximately 8 € per year and per species in addition to the investment. Put into a phytoplankton perspective and taking the Baltic Sea as an example, ca. 1630 eukaryotic

phytoplankton species have been microscopically observed in the Baltic Sea (G. Hällfors, 2004), and based on a recent search (Henna Savela, personal communication), 573 (35%) of these species have at least one reference sequence in the PR² or SILVA database. Thus, adding all the currently known Baltic Sea phytoplankton species to reference databases would cost over 1 million €, according to the calculations of Magnussen and Navrud (2021). In addition to the currently known species, it is likely that the real biodiversity of phytoplankton in this specific sea area is higher, and the number of species will increase with increasing scientific knowledge.

Collection of DNA reference material gives a number of benefits, in the form of increased knowledge about species and ecosystems. This also gives increased non-use values and increased option values because collecting and securing DNA references gives the opportunity to use the information in the future. Moreover, reference databases provide the opportunity to analyze and use the collected information in environmental monitoring and management, and this is where the greatest benefits of building and maintaining DNA reference libraries lie for management. According to Magnussen and Navrud (2021), monitoring can in many cases not only be more cost-effective, but eDNA samples can also provide more information and knowledge about species, populations and ecosystems. This in turn provides an opportunity for better management based on more in-depth knowledge of the nature to be managed. Furthermore, these new methods can contribute to innovation and "spin-offs", which can be interesting for environmental management, but also for a wider range of applications. Magnussen and Navrud (2021) conclude that the greatest benefit-cost impact of using DNA-based methods for monitoring is where collection costs can be reduced by simpler sampling, and where relatively many species that are time-consuming to determine morphologically but easy to identify by DNA sequences are present. The latter point applies to phytoplankton. As an extreme example, Magnussen and Navrud (2021) show that the DNA-based monitoring of flying insects in Norway costs approximately 2 million euros per year while the same monitoring based on conventional methodology would cost approximately 48 million euros per year.

Here, we have compiled a table (**Table 2**) of working hours related to conventional (light microscopy) and DNA-based phytoplankton monitoring. We have used our expert knowledge on time used in each step and estimated the total working hours per sample for conducting monitoring using both methods and assuming that 100–1000 samples are taken. The working hours for sampling are the same for the conventional and the DNA-based method since the same sampling methodology should be used. Filtration causes additional costs for the DNA-based method, requiring specific filtering equipment, working time, and a clean workspace. Phytoplankton sampling for microscopy is very simple and does not require much equipment or working time. Costs of using the research vessel are not included in the calculations, since both microscopy and eDNA samples for phytoplankton can be taken during the same sampling cruises along with all other biological, chemical, and physical samples for routine monitoring.

Identifying taxa with eDNA metabarcoding requires more working steps compared to the conventional method, which only requires fixation, settling and microscopy. However, increasing the number of samples decreases the cost per sample of the DNA-based survey while it does not decrease the cost of the conventional survey. This is one of the definite benefits of DNA-based methods and should be taken into

consideration in management where large amounts of samples from large areas are needed for a comprehensive overview of the status of the environment. As an example, all phytoplankton samples across the HELCOM and OSPAR areas could be processed together using coherent sampling and sample processing, reducing the overall cost markedly since the cost of laboratory (sequencing costs) and computer work (bioinformatic filtering and taxonomic assignment) per sample decreases as the number of samples increases.

In addition to labor costs, filtration, DNA extraction and PCR amplification require specific equipment, chemicals and consumables. Laboratory facilities and equipment are an expensive investment but usually available already within institutes or through collaboration. A rough estimate for chemicals and consumables to be used for 100 samples is 4000–5000 €.

Table 2. Estimates of working hours for conventional (light microscopy) and DNA-based monitoring of phytoplankton after sampling and sample preparation (i.e. preparing an integrated sample, which is estimated to take ca. 0.20h / sample, can be carried out onboard the research vessel and done simultaneously for light microscopy and DNA-based monitoring). Costs for sequencing of the eDNA samples are also shown. Costs for using the research vessel for sampling cruises and for other equipment are not included in the calculations.

Process	Conventional monitoring		DNA-based monitoring	
	Subprocess	Working hours	Subprocess	Working hours/Cost
Sample processing	Preservation	ca. 0.10 h/sample (onboard research vessel)	Filtration	0.25 h/sample (depending on biomass)
			DNA extraction	0.12 h/sample
Identification	Sample preparation and microscopy	ca. 8 h/sample	PCR amplification*	0.05 h/sample
			Sequencing**	15–20 €/sample
			Bioinformatic filtering*	0.03–0.05 h/sample
			Assigning taxonomy*	0.04–0.2 h/sample
Total		ca. 8–9 h/sample		0.59–0.87 h/sample + sequencing cost

*The more samples included the less time is spent per sample.

**Usually as a purchased service.

9. Conclusions and future perspectives

Using DNA-based methods can fundamentally transform environmental monitoring, given the new opportunities that arise, and the potential to overcome the shortcomings of traditional methods. However, there are still many aspects that require careful consideration and discussion amongst experts and stakeholders, before eDNA metabarcoding can be fully implemented as a complementary approach to microscopy. It is crucial to agree on details that will affect the big picture and can help improve biodiversity assessment. Currently, there is no common phytoplankton biodiversity indicator for European marine and brackish waters. eDNA metabarcoding could aid in the development of such an indicator, which is urgently needed for improved assessment of the biodiversity and status of marine and brackish water bodies.

The technology and methodology described here offer major advantages over classical approaches, the most remarkable of which is the ability to identify and monitor a very broad range of organisms. The guidelines provide an overview and guidance on important aspects needed for the successful implementation of eDNA metabarcoding for phytoplankton monitoring. An expansion to other groups of organisms, not covered by traditional monitoring, should be considered in the future as this would provide valuable insight into complex interactions. For example, the inclusion of nano- and picoplankton, microzooplankton as well as associated fungal or apicomplexan parasites which form associations with phytoplankton would be important.

Another advantage of eDNA metabarcoding is the low effort required for upscaling once a common protocol has been taken into use. Thus, ideally, the guidelines presented herein will be developed further and integrated into e.g. the HELCOM framework and standardized under CEN or ISO. Thereby comparable standardized data can be produced, and eDNA-based monitoring will be more cost-efficient. At the same time, phytoplankton monitoring with eDNA metabarcoding should be aligned with existing monitoring guidelines which use light microscopy, to enable comparison of results obtained from both methods.

One of the biggest challenges for implementing eDNA metabarcoding into routine phytoplankton monitoring is the improvement of reference databases. This publication provides an overview and recommendations on which databases to use and how to manage sequencing data. Further improvement of reference databases requires coordinated effort and can be achieved much faster and easier if several European states contribute.

Overall, the benefits of using DNA-based methods for monitoring are convincing and the new opportunities outweigh the challenges. Thus, we are confident that future phytoplankton biomonitoring or bioassessment will include environmental DNA survey techniques. We hope that these guidelines will facilitate a better understanding and harmonization of eDNA-based methods and will contribute to their implementation into routine monitoring programs.

Acknowledgements

Markus Lindh, Anders F. Andersson, Elianne Egge and Silke Van den Wyngaert are acknowledged for their helpful comments and advice on the first draft of the manuscript. The participants of the workshop organized to discuss the first draft of the guidelines are thanked for their critical questions, comments, and feedback. Furthermore, the authors wish to thank the following colleagues for advice given during personal communication: Henna Savela, Florian Leese, Ása Jakobsen. Danielle Bansfield is acknowledged for proofreading parts of the manuscript. The ID-NEP project was funded by the Nordic Council of Ministers (Project 220002).

References

- Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C. & Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.*, 12, R18. DOI: 10.1186/gb-2011-12-2-r18
- Alberdi, A., Aizpurua, O., Gilbert, M.T.P. & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol. Evol.*, 9, 134–147. DOI: 10.1111/2041-210X.12849
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402. DOI: 10.1093/nar/25.17.3389
- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. & Huse, S.M. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA Genes. *PLoS ONE*, 4, e6372. DOI: 10.1371/journal.pone.0006372
- Andersen, P. (2010). Filtering – calcofluor staining – quantitative epifluorescence microscopy for phytoplankton analysis. In: Karlson, B., Cusack, C. & Bresnan, E. (eds.), *Microscopic and molecular methods for quantitative phytoplankton analysis*. IOC Manuals and Guides 55, UNESCO, Paris, p. 31–36. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000187824>
- Andersson, A.F., Bissett, A., Finstad A.G., Fossøy, F., Grosjean, M., Hope, M., Jeppesen, T.S., Køljalg, U., Lundin, D., Nilsson, R.N., et al. (2020). *Publishing DNA-derived data through biodiversity data platforms*. v1.0 Copenhagen: GBIF Secretariat. 47 p. DOI: 10.35035/doc-vf1a-nr22.
- Andersson, A., Karlson, B., Andersson, A.F., Brugel, S., Latz, M., Lycken, J., Hedblom, M., Torstensson, A. & Lindh, M. (2022). *DNA extraction protocol for DNA-metabarcoding of marine phytoplankton using Zymobiomics DNA minprep kit (Zymo Research; D4300)*. Protocols.io. Retrieved from [dx.doi.org/10.17504/protocols.io.bucjnsun](https://doi.org/10.17504/protocols.io.bucjnsun)
- Apothéloz-Perret-Gentil, L., Bouchez, A., Cordier, T., Cordonier, A., Guéguen, J., Rimet, F., Vasselon, V. & Pawlowski, J. (2021). Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. *Mol. Ecol.*, 30, 2959–2968. DOI: 10.1111/mec.15646
- 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. DOI: 10.3354/ame01753
- Åström, J., Birkemoe, T., Dahle, S., Davey, M., Ekrem, T., Endrestøl, A., Fossøy, F., Hanssen, O., Laugsand, A., Staverløkk, A., et al. (2022). *Insektovervåking på Østlandet og i Trøndelag. Rapport fra feltsesong 2021*. NINA Rapport 2070. Norsk institutt for naturforskning. 119 p. Retrieved from <https://hdl.handle.net/11250/2976316>
- Axtner, J., Crampton-Platt, A., Horig, L.A., Mohamed, A., Xu, C.C.Y., Yu, D.W. &

- Wilting, A. (2019). An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies. *GigaScience*, 8, giz029. DOI: 10.1093/gigascience/giz029
- Bakker, J., Wangenstein, O.S., Baillie, C., Buddo, D., Chapman, D.D., Gallagher, A.J., Guttridge, T.L., Hertler, H. & Mariani, S. (2019). Biodiversity assessment of tropical shelf eukaryotic communities via pelagic eDNA metabarcoding. *Ecol. Evol.*, 9, 14341–14355. DOI: 10.1002/ece3.5871
- Balzano, S., Abs, E. & Leterme, S.C. (2015). Protist diversity along a salinity gradient in a coastal lagoon. *Aquat. Microb. Ecol.*, 74, 263–277. DOI: 10.3354/ame01740
- 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. DOI: 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. DOI: 10.1093/plankt/fby008
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2013). *Nucleic Acids Res.*, 41(D1), D36–D42. DOI: 10.1093/nar/gks1195
- Berney, C., Fahrni, J. & Pawlowski, J. (2004). How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol.*, 2, 13. DOI: 10.1186/1741-7007-2-13
- Berry, T.E., Saunders, B.J., Coghlan, M.L., Stat, M., Jarman, S., Richardson, A.J., Davies, C.H., Berry, O., Harvey, E.S. & Bunce, M. (2019). Marine environmental DNA biomonitoring reveals seasonal patterns in biodiversity and identifies ecosystem responses to anomalous climatic events. *PLoS Genet.*, 15, e1007943. DOI: 10.1371/journal.pgen.1007943
- 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. DOI: 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. DOI: 10.1038/s41587-019-0209-9
- Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.*, 16, 176–182. DOI: 10.1111/1755-0998.12428
- Bradley, I.M., Pinto, A.J. & Guest, J.S. (2016). Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. *Appl. Environ. Microbiol.*, 82, 5878–5891. DOI: 10.1128/AEM.01630-16
- Brandariz-Fontes, C., Camacho-Sanchez, M., Vila, C., Vega-Pla, J.L., Rico, C. & Leonard, J.A. (2015). Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. *Sci. Rep.*, 5, 8056. DOI: 10.1038/srep08056

- Bråte, J., Logares, R., Berney, C., Ree, D.K., Klaveness, D., Jakobsen, K.S. & Shalchian-Tabrizi, K. (2010). Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME J.*, *4*, 1144–1153. DOI: 10.1038/ismej.2010.39
- 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. DOI: 10.3897/ab.e68634
- Bucklin, A., Lindeque, P.K., Rodriguez-Ezpeleta, N., Albaina, A. & Lehtiniemi, M. (2016). Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *J. Plankton Res.*, *38*, 393–400. DOI: 10.1093/plankt/fbw023
- Bucklin, A., Yeh, H.D., Questel, J.M., Richardson, D.E., Reese, B., Copley, N.J. & Wiebe, P.H. (2019). Time-series metabarcoding analysis of zooplankton diversity of the NW Atlantic continental shelf. *ICES J. Mar. Sci.*, *76*, 1162–1176. DOI: 10.1093/icesjms/fsz021
- Bunse, C., Bertos-Fortis, M., Sassenhagen, I., Sildever, S., Sjöqvist, C., Godhe, A., Gross, S., Kremp, A., Lips, I., Lundholm, N., et al. (2016). Spatio-temporal interdependence of bacteria and phytoplankton during a Baltic Sea spring bloom. *Front. Microbiol.*, *7*, 517. DOI: 10.3389/fmicb.2016.00517
- 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. DOI: 10.1038/nmeth.3869
- Callahan, B.J., McMurdie, P.J. & Holmes, S.P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.*, *11*, 2639–2643. DOI: 10.1038/ismej.2017.119
- 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. DOI: 10.1186/s40168-021-01072-3
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N. & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*, *108*(S1), 4516–4522. DOI: 10.1073/pnas.1000080107.
- Chai, Z.Y., He, Z.L., Deng, Y.Y., Yang, Y.F. & Tang, Y.Z. (2018). Cultivation of seaweed *Gracilaria lemaneiformis* enhanced biodiversity in a eukaryotic plankton community as revealed via metagenomic analyses. *Mol. Ecol.*, *27*, 1081–1093. DOI: 10.1111/mec.14496
- Choi, J. & Park, J.S. (2020). Comparative analyses of the V4 and V9 regions of 18S rDNA for the extant eukaryotic community using the Illumina platform. *Sci. Rep.*, *10*, 6519. DOI: 10.1038/s41598-020-63561-z
- Choi, C. J., Jimenez, V., Needham, D. M., Poirier, C., Bachy, C., Alexander, H., Wilken, S., Chavez F. P., Sudek, S., Giovannoni, S. J., Worden, A. Z. (2020). Seasonal and Geographical Transitions in Eukaryotic Phytoplankton Community Structure in the Atlantic and Pacific Oceans. *Front. Microbio.*, *11*, 542372. DOI:10.3389/fmicb.2020.542372
- Comeau, A.M., Li, W.K.W., Tremblay, J.-É., Carmack, E.C. & Lovejoy, C. (2011). Arctic

- Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE*, 6, e27492. DOI: 10.1371/journal.pone.0027492
- Cordier, T., Frontalini, F., Cermakova, K., Apothéoz-Perret-Gentil, L., Treglia, M., Scantamburlo, E., Bonamin, V. & Pawlowski, J. (2019). Multi-marker eDNA metabarcoding survey to assess the environmental impact of three offshore gas platforms in the North Adriatic Sea (Italy). *Mar. Environ. Res.*, 146, 24–34. DOI: 10.1016/j.marenvres.2018.12.009
- Culverhouse, P.F., Williams, R., Reguera, B., Herry, V. & Gonzalez-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Mar. Ecol. Prog. Ser.*, 247, 17–25. DOI: 10.3354/meps247017
- Culverhouse, P.F., Macleod, N., Williams, R., Benfield, M.C., Lopes, R.M. & Picheral, M. (2014). An empirical assessment of the consistency of taxonomic identifications. *Mar. Biol. Res.*, 10, 73–84. DOI: 10.1080/17451000.2013.810762
- Czech, L., Stamatakis, A., Dunthorn, M. & Barbera, P. (2022). Metagenomic analysis using phylogenetic placement—A review of the first decade. *Front. Bioinform.*, 2, 871393. DOI: 10.3389/fbinf.2022.871393
- 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. DOI: 10.1126/science.1261605
- Deiner, K., Lopez, J., Bourne, S., Holman, L.E., Seymour, M., Grey, E.K., Lacoursiere-Roussel, A., Li, Y., Renshaw, M.A., Pfrender, M.E., et al. (2018). Optimising the detection of marine taxonomic richness using environmental DNA metabarcoding: the effects of filter material, pore size and extraction method. *Metabarcoding and Metagenomics*, 2, e28963. DOI: 10.3897/mbmg.2.28963
- Djemiel, C., Plassard, D., Terrat, S., Crouzet, O., Sauze, J., Mondy, S., Nowak, V., Wingate, L., Ogée, J. & Maron, P.A. (2020). µgreen-db: a reference database for the 23S rRNA gene of eukaryotic plastids and cyanobacteria. *Sci. Rep.*, 10, 5915. DOI: 10.1038/s41598-020-62555-1
- 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. & Chavez, F.P. (2017). Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. *Front. Mar. Sci.*, 4, 314. DOI: 10.3389/fmars.2017.00314
- Djurhuus, A., Closek, C.J., Kelly, R.P., Pitz, K.J., Michisaki, R.P., Starks, H.A., Walz, K.R., Andruszkiewicz, E.A., Olesin, E., Hubbard, K., et al. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nat. Commun.*, 11, 254. DOI: 10.1038/s41467-019-14105-1
- Ekrem, T., Willassen, E. & Stur, E. (2007). A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Mol. Phylogenet. Evol.*, 43, 530–542. DOI: 10.1016/j.ympev.2006.11.021
- Elsersek, T. (Ed.). (2021). *Technical guidelines for eDNA monitoring in Alpine waters for stakeholders and end-users*. National Institute of Biology. Retrieved from <https://www.alpine-space.eu/project/eco-alpswater/>
- Elferink, S., Neuhaus, S., Wohlrab, S., Toebe, K., Voß, D., Gottschling, M., Lundholm, N., Krock, B., Koch, B.P., Zielinski, O., et al. (2017). Molecular diversity patterns among

various phytoplankton size-fractions in West Greenland in late summer. *Deep-Sea Res. I*, 121, 54–69. DOI: 10.1016/j.dsr.2016.11.002

Elwood, H.J., Olsen, G.J. & Sogin, M.L. (1985). The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. *Mol. Biol. Evol.*, 2, 399–410. DOI: 10.1093/oxfordjournals.molbev.a040362

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, 1–72.

European Union (2008). Directive 2008/56/EC of the European Parliament and of the Council of 17 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, 19–40.

European Union (2017a). Commission Directive (EU) 2017/845 of 17 May 2017 amending Directive 2008/56/EC of the European Parliament and of the Council as regards the indicative lists of elements to be taken into account for the preparation of marine strategies. *Official Journal of the European Union*, L 125, 27–33.

European Union (2017b). Commission Decision (EU) 2017/848 of 17 May 2017 laying down criteria and methodological standards on good environmental status of marine waters and specifications and standardised methods for monitoring and assessment, and repealing Decision 2010/477/EU. *Official Journal of the European Union*, L 125, 43–74.

European Union (2019). Commission Implementing Regulation (EU) 2019/627 of 15 March 2019 laying down uniform practical arrangements for the performance of official controls on products of animal origin intended for human consumption in accordance with Regulation (EU) 2017/625 of the European Parliament and of the Council and amending Commission Regulation (EC) No 2074/2005 as regards official controls. *Official Journal of the European Union*, L 131, 51–100.

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. DOI: 10.1093/bioinformatics/btw354

Ficetola, G.F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P. & Pompanon, F. (2010). An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, 11, 434. DOI: 10.1186/1471-2164-11-434

Fiskaaling (n.d.). *Integrated approaches for phytoplankton dynamics*. Retrieved (12 Dec 2022) from <https://fiskaaling.fo/en/departments/biotech/current-research-projects/integrated-approaches-for-phytoplankton-dynamics/>

Fischer, M.A., Güllert, S., Neulinger, S.C., Streit, W.R. & Schmitz, R.A. (2016). Evaluation of 16S rRNA gene primer pairs for monitoring microbial community structures showed high reproducibility within and low comparability between datasets generated with multiple archaeal and bacterial primer pairs. *Front. Microbiol.*, 7, 1297. DOI: 10.3389/fmicb.2016.01297

Forster, D., Filker, S., Kochems, R., Breiner, H.W., Cordier, T., Pawlowski, J. & Stoeck, T. (2019). A comparison of different ciliate metabarcode genes as bioindicators for environmental impact assessments of salmon aquaculture. *J. Eukaryot. Microbiol.*, 66, 294–308. DOI: 10.1111/jeu.12670

- Fu, M., Hemery, L. & Sather, N. (2021a). *Cost efficiency of environmental DNA as compared to conventional methods for biodiversity monitoring purposes at marine energy sites*. Pacific Northwest National Laboratory PNNL-32310. Prepared for the U.S. Department of Energy under Contract DE-AC05-76RL01830. 38 p. Retrieved from https://tethys.pnnl.gov/sites/default/files/publications/Fu_et_al_2021.pdf
- Fu, Z., Piumsomboon, A., Punnarak, P., Uttayarnmanee, P., Leaw, C.P., Lim, P.T., Wang, A. & Gu, H. (2021b). Diversity and distribution of harmful microalgae in the Gulf of Thailand assessed by DNA metabarcoding. *Harmful Algae*, 106, 102063. DOI: 10.1016/j.hal.2021.102063
- Gohl, D.M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T.J., Clayton, J.B., Johnson, T.J., Hunter, R., et al. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.*, 34, 942–949. DOI: 10.1038/nbt.3601
- Gran-Stadniczeňko, S., Šupraha, L., Egge, E.D. & Edvardsen, B. (2017). Haptophyte diversity and vertical distribution explored by 18S and 28S ribosomal RNA gene metabarcoding and scanning electron microscopy. *J. Eukaryot. Microbiol.*, 64, 514–532. DOI: 10.1111/jeu.12388
- Gromisz, S. & Witek, Z. (2001). Main phytoplankton assemblages in the Gulf of Gdansk and the Pomeranian Bay from 1994 to 1997. *Bulletin of the Sea Fisheries Institute*, 2, 31–51. Retrieved from <https://mir.gdynia.pl/pliki/osrodek/biuletyn/biulet2-01.pdf#page=30>
- Grzebyk, D., Audic, S., Lasserre, B., Abadie, E., de Vargas, C. & Bec, B. (2017). Insights into the harmful algal flora in northwestern Mediterranean coastal lagoons revealed by pyrosequencing metabarcodes of the 28S rRNA gene. *Harmful Algae*, 68, 1–16. DOI: 10.1016/j.hal.2017.06.003
- 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 (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.*, 41(D1), D597–D604. DOI: 10.1093/nar/gks1160
- Günther, B., Knebelsberger, T., Neumann, H., Laakmann, S. & Martínez Arbizu, P. (2018). Metabarcoding of marine environmental DNA based on mitochondrial and nuclear genes. *Sci. Rep.*, 8, 14822. DOI: 10.1038/s41598-018-32917-x
- Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaq, D., Highlander, S.K., Sodergren, E., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.*, 21, 494–504. DOI: 10.1101/gr.112730.110
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M. & Troedsson, C. (2014). Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS ONE* 9: e87624. DOI: 10.1371/journal.pone.0087624
- Hällfors, G., Niemi, Å, Ackefors, H., Lassig, J. & Leppäkoski, E. (1981). Biological Oceanography. In: Voipio, A. (ed.), *The Baltic Sea. Elsevier Oceanography Series*, 30, 219–274. DOI: 10.1016/S0422-9894(08)70142-9
- Hällfors, G. (2004). Checklist of Baltic Sea phytoplankton species (including some heterotrophic protistan groups). *HELCOM Baltic Sea Environment Proceedings*, 95, 208 p. Retrieved from <https://www.helcom.fi/wp-content/uploads/2019/10/>

Hällfors, S. (ed.) (2007). Annex 6: Potentially harmful phytoplankton species of the Baltic Sea. – In: ICES, Report of the ICES-IOC-SCOR Working Group on GEOHAB Implementation in the Baltic (WGGIB), 7–9 May 2007, Helsinki, Finland, 21–28. *ICES CM 2007/BCC:05*. 51 pp.

Hamsher, S.E., Evans, K.M., Mann, D.G., Pouličková, A. & Saunders, G.W. (2011). Barcoding diatoms: Exploring alternatives to COI-5P. *Protist*, 162, 405–422. DOI: 10.1016/j.protis.2010.09.005

Hassouna, N., Mithot, B. & Bachellerie, J.P. (1984). The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucleic Acids Res.*, 12, 3563–3583. DOI: 10.1093/nar/12.8.3563

Heather, J.M. & Chain, B. (2016). The sequence of sequencers: the history of sequencing DNA. *Genomics*, 107: 1–8. DOI: 10.1016/j.ygeno.2015.11.003

Heeger, F., Bourne, E.C., Baschien, C., Yurkov, A., Bunk, B., Spröer, C., Overmann, J., Mazzoni, C.J. & Monaghan, M.T. (2018). Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. *Mol. Ecol. Resour.*, 18, 1500–1514. DOI: 10.1111/1755-0998.12937

HELCOM (2015). 3-1-Add.1. List of Target Species currently in use in the HELCOM area. Baltic Marine Environment Protection Commission, Maritime Group, MARITIME 15-2015, Klaipeda, Lithuania, 23–25 November 2015. Retrieved from <https://portal.helcom.fi/meetings/MARITIME%2015-2015-245/MeetingDocuments/3-1-Add1%20List%20of%20Target%20Species%20currently%20in%20use%20in%20the%20HELCOM%20area.pdf>

HELCOM (2021). *Guidelines for monitoring of phytoplankton species composition, abundance and biomass*. Updated November 2021. 22 p. Retrieved from <https://helcom.fi/wp-content/uploads/2020/01/HELCOM-Guidelines-for-monitoring-of-phytoplankton-species-composition-abundance-and-biomass.pdf>

Henriques, A., Cereija, T., Machado, A. & Cerca, N. (2012). *In silico* vs *in vitro* analysis of primer specificity for the detection of *Gardnerella vaginalis*, *Atopobium vaginae* and *Lactobacillus* spp. *BMC Research Notes*, 5, 637. DOI: 10.1186/1756-0500-5-637

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. DOI: 10.1038/ismej.2011.41

Hillis, D.M. & Dixon, M.T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.*, 66, 411–453. DOI: 10.1086/417338

Hleap, J.S., Littlefair, J.E., Steinke, D., Hebert, P.D.N. & Cristescu, M.E. (2021). Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes. *Mol. Ecol. Resour.*, 21, 2190–2203. DOI: 10.1111/1755-0998.13407

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. DOI: 10.3389/fmicb.2016.00679

Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D., Wilmes, P. & Andersson, A.F. (2014). Systematic design of 18S rRNA gene primers for

- determining eukaryotic diversity in microbial consortia. *PLoS ONE*, 9, e95567. DOI: 10.1371/journal.pone.0095567
- Huson, D.H., Beier, S., Flade, I., Górská, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J. & Tappu, R. (2016). MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.*, 12, e1004957. DOI: 10.1371/journal.pcbi.1004957
- Huws, S.A., Edwards, J.E., Kim, E.J. & Scollan, N.D. (2007). Specificity and sensitivity of eubacterial primers utilized for molecular profiling of bacteria within complex microbial ecosystems. *J. Microbiol. Methods.*, 70, 565–569. DOI: 10.1016/j.mimet.2007.06.013.
- Illumina (2020). *16S Sequencing Method Guide. Taxonomic profiling of bacterial communities and microbiomes through next-generation sequencing*. Retrieved from https://jornades.uab.cat/workshopmrama/sites/jornades.uab.cat/workshopmrama/files/16s_sequencing.pdf
- Illumina (n.d.) *Illumina sequencing platforms*. Retrieved (12 Dec 2022) from <https://www.illumina.com/systems/sequencing-platforms.html>
- Jaanus, A., Andersson, A., Olenina, I., Töming, K. & Kaljurand, K. (2011). Changes in phytoplankton communities along a north–south gradient in the Baltic Sea between 1990 and 2008. *Boreal Environ. Res.*, 16(Suppl. A), 191–208. Retrieved from <http://www.borenv.net/BER/archive/pdfs/ber16/ber16A-191.pdf>
- Jakobsen, H.H., Carstensen, J., Harrison, P.J. & Zingone, A. (2015). Estimating time series phytoplankton carbon biomass: Inter-lab comparison of species identification and comparison of volume-to-carbon scaling ratios. *Est. Coast. Shelf Sci.*, 162, 143–150. DOI: 10.1016/j.ecss.2015.05.006
- Jerney, J., Hällfors, H., Oja, J., Reunamo, A., Suikkanen, S. & Lehtinen, S. (2022). Guidelines for using environmental DNA in Finnish marine phytoplankton monitoring. Improved biodiversity assessment through method complementation. *Reports of the Finnish Environment Institute* 40/2022, 69 pp. Retrieved from <https://helda.helsinki.fi/handle/10138/351131>
- Jeunen, G.J., Knapp, M., Spencer, H.G., Taylor, H.R., Lamare, M.D., Stat, M., Bunce, M. & Gemmell, N.J. (2019). Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. *Ecol. Evol.*, 9, 1323–1335. DOI: 10.1002/ece3.4843
- John, U., Fensome, R.A. & Medlin, L.K. (2003). The application of a molecular clock based on molecular sequences and the fossil record to explain biogeographic distributions within the *Alexandrium tamarense* "species complex" (Dinophyceae). *Mol. Biol. Evol.*, 20, 1015–1027. DOI: 10.1093/molbev/msg105
- Joo, H.M., Kim, K.E., Park, J.S., Kim, H.-J., Yang, E.J., Lee, T.-K., Cho, K.-H., Jung, J., Lee, Y., Kang, S.-H., et al. (2022). Ecological responses of core phytoplankton by latitudinal differences in the Arctic Ocean in late summer revealed by 18S rDNA metabarcoding. *Front. Mar. Sci.*, 9, 879911. DOI: 10.3389/fmars.2022.879911
- Jurgensone, I., Carstensen, J., Ikaunieca, A. & Kalveka, B. (2011). Long-term changes and controlling factors of phytoplankton community in the Gulf of Riga (Baltic Sea). *Estuar. Coast.*, 34, 1205–1219. DOI: 10.1007/s12237-011-9402-x
- 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. Retrieved from <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>

Karlson, B., Andersen, P., Arneborg, L., Cembella, A., Eikrem, W., John, U., West, J.J., Klemm, K., Kobos, J., Lehtinen, S., et al. (2021). Harmful algal blooms and their effects in coastal seas of Northern Europe. *Harmful Algae*, 102, 101989. DOI: 10.1016/j.hal.2021.101989

Kawato, M., Yoshida, T., Miya, M., Tsuchida, S., Nagano, Y., Nomura, M., Yabuki, A., Fujiwara, Y. & Fujikura, K. (2021). Optimization of environmental DNA extraction and amplification methods for metabarcoding of deep-sea fish. *MethodsX*, 8, 101238. DOI: 10.1016/j.mex.2021.101238

Kchouk, M., Gibrat, J.F. & Elloumi, M. (2017). Generations of sequencing technologies: From first to next generation. *Biol. Med. (Aligarh)*, 9, 395. DOI:10.4172/0974-8369.1000395

Kelly, R.P., Shelton, A.O. & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Sci. Rep.*, 9, 12133. DOI: 10.1038/s41598-019-48546-x

Ki, J.S. & Han, M.S. (2007). Informative characteristics of 12 divergent domains in complete large subunit rDNA sequences from the harmful dinoflagellate genus, *Alexandrium* (Dinophyceae). *J. Eukaryot. Microbiol.*, 54, 210–219. DOI: 10.1111/j.1550-7408.2007.00251.x

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. DOI: 10.3390/app9163272

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glöckner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.*, 41, e1, DOI: 10.1093/nar/gks808

Kohli, G.S., Neilan, B.A., Brown, M.V., Hoppenrath, M. & Murray, S.A. (2014). *Cob* gene pyrosequencing enables characterization of benthic dinoflagellate diversity and biogeography. *Environ. Microbiol.*, 16, 467–485. DOI: 10.1111/1462-2920.12275

Kolda, A., Ljubešić, Z., Gavrilović, A., Jug-Dujaković, J., Pikelj, K., Kapetanović, D. (2020). Metabarcoding Cyanobacteria in coastal waters and sediment in central and southern Adriatic Sea. *Acta Bot. Croat.*, 79, 157–169. DOI: 10.37427/botcro-2020-021. Retrieved from <https://hrcak.srce.hr/238381>

Kraft, K., Seppälä, J., Hällfors, H., Suikkanen, S., Ylöstalo, P., Anglès, S., Kielosto, S., Kuosa, H., Laakso, L., Honkanen, M., et al. (2021). First application of IFCB high-frequency imaging-in-flow cytometry to investigate bloom-forming filamentous cyanobacteria in the Baltic Sea. *Front. Mar. Sci.*, 8, 594144. DOI: 10.3389/fmars.2021.594144

Lahr, D.J.G. & Katz, L.A. (2009). Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-

- generation high-fidelity DNA polymerase. *BioTechniques*, 47, 857-866. DOI: 10.2144/000113219
- 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. DOI: 10.1111/1755-0998.13623
- Lenaers, G., Maroteaux, L., Michot, B. & Herzog, M. (1989). Dinoflagellates in evolution. A molecular phylogenetic analysis of large subunit ribosomal RNA. *J. Mol. Evol.*, 29, 40–51. DOI: 10.1007/BF02106180
- Leray, M., Knowlton, N., Ho, S.L., Nguyen, B.N. & Machida, R.J. (2019). GenBank is a reliable resource for 21st century biodiversity research. *P. Natl. Acad. Sci.*, 116, 22651–22656. DOI: 10.1073/pnas.1911714116
- Li, X.C., Huo, S., Zhang, J., Ma, C., Xiao, Z., Zhang, H., Xi, B., Xia, X. (2019). Metabarcoding reveals a more complex cyanobacterial community than morphological identification. *Ecol. Indic.*, 107, 105653. DOI: 10.1016/j.ecolind.2019.105653
- Liang, Z. & Keeley, A. (2013). Filtration recovery of extracellular DNA from environmental water samples. *Envir. Sci. Tech.*, 47, 9324–9331. DOI: 10.1021/es401342b
- Lindahl, O. (1986). *A dividable hose for phytoplankton sampling*. ICES Report of the Working Group on Exceptional Algal Blooms. International Council for the Exploration of the Sea, Hirtshals, Denmark, p. 31.
- Lindahl, B.D., Nilsson, R.H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjølner, R., Kõljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J. & Kausrud, H. (2013). Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New. Phytol.*, 199, 288–299. DOI: 10.1111/nph.12243
- MacNeil, L., Desai, D.K., Costa, M. & LaRoche, J. (2022). Combining multi-marker metabarcoding and digital holography to describe eukaryotic plankton across the Newfoundland Shelf. *Sci. Rep.*, 12, 13078. DOI: 10.1038/s41598-022-17313-w
- Magnussen, K. & Navrud, S. (2021). *Nytte og kostnader ved bruk av DNA-basert metodikk og miljø-DNA i miljøovervåking*. Rapport. Menon Economics. Menon-Publikasjon 90/2021. M-21152021. 56 pp. Retrieved from <https://www.menon.no/wp-content/uploads/2021-90-Nytte-og-kostnader-ved-bruk-av-DNA-basert-metodikk-og-miljo-DNA-i-miljoovervaking.pdf>
- Majaneva, M., Autio, R., Huttunen, M., Kuosa, H. & Kuparinen, J. (2009). Phytoplankton monitoring: the effect of sampling methods used during different stratification and bloom conditions in the Baltic Sea. *Boreal Environ. Res.*, 14, 313–322. Retrieved from <https://helda.helsinki.fi/bitstream/handle/10138/233444/ber14-2-313.pdf?sequence=1>
- Majaneva, M., Blomster, J., Müller, S., Autio, R., Majaneva, S., Hyytiäinen, K., Nagai, S. & Rintala, J.-M. (2017). Sea-ice eukaryotes of the Gulf of Finland, Baltic Sea, and evidence for herbivory on weakly shade-adapted ice algae. *Eur. J. Protistol.*, 57, 1–15. DOI: 10.1016/j.ejop.2016.10.005
- 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. DOI: 10.1038/s41598-018-23052-8

Mäki, A., Salmi, P., Mikkonen, A., Kremp, A. & Tirola, M. (2017). Sample preservation, DNA or RNA extraction and data analysis for high-throughput phytoplankton community sequencing. *Front. Microbiol.*, 8, 1848. DOI: 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. DOI: 10.1038/s41598-021-95223-z

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R., Claverie, J.M., Decelle, J., et al. (2015). Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.*, 17, 4035–4049. DOI: 10.1111/1462-2920.12955

Menden-Deuer, S. & Lessard, E.J. (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnol. Oceanogr.*, 45, 569–579. DOI: 10.4319/lo.2000.45.3.0569

Monchy, S., Grattepanche, J.-D., Breton, E., Meloni, D., Sancier, G., Chabé, M., Delhaes, L., Viscogliosi, E., Sime-Ngando, T. & Christaki, U. (2012). Microplanktonic community structure in a coastal system relative to a *Phaeocystis* bloom inferred from morphological and tag pyrosequencing methods. *PLoS ONE*, 7, e39924. DOI: 10.1371/journal.pone.0039924

Morales, S.E. & Holben, W.E. (2009). Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by in silico analysis. *Appl. Environ. Microbiol.*, 75, 2677–2683. DOI: 10.1128/AEM.02166-08

Moreno-Pino, M., Krock, B., De la Iglesia, R., Echenique-Subiabre, I., Pizarro, G. & Vásquez, M., Trefault, N. (2018). Next Generation Sequencing and mass spectrometry reveal high taxonomic diversity and complex phytoplankton-phycotoxins patterns in Southeastern Pacific fjords. *Toxicon*, 151, 5–14. DOI: 10.1016/j.toxicon.2018.06.078

Muñoz-Colmenero, M., Sánchez, A., Correa, B., Figueiras, F.G., Garrido, J.L. & Sotelo, C.G. (2021). Evaluation of DNA extraction methods and bioinformatic pipelines for marine nano- and pico-eukaryotic plankton analysis. *Front. Mar. Sci.*, 7, 584253. DOI: 10.3389/fmars.2020.584253

Murray, S., Flø Jørgensen, M., Ho, S.Y.W., Patterson, D.J. & Jeremiin, L.S. (2005). Improving the analysis of dinoflagellate phylogeny based on rDNA. *Protist*, 156, 269–286. DOI: 10.1016/j.protis.2005.05.003

Nagai, S., Hida, K., Urusizaki, S., Takano, Y., Hongo, Y., Kameda, T. & Abe, K. (2016). Massively parallel sequencing-based survey of eukaryotic community structures in Hiroshima Bay and Ishigaki Island. *Gene*, 576(2 Pt 1), 681–689. DOI: 10.1016/j.gene.2015.10.026

Nagai, S., Sildever, S., Nishi, N., Tazawa, S., Basti, L., Kobayashi, T. & Ishino, Y. (2022). Comparing PCR-generated artifacts of different polymerases for improved accuracy of DNA metabarcoding. *Metabarcoding and Metagenomics*, 6, 27–39. DOI: 10.3897/mbmg.6.77704

Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E. & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding.

Mol. Ecol. Resour., 18, 927–939. DOI: 10.1111/1755-0998.12895

Nilsson, R.H., Andersson, A.F., Bissett, A., Finstad, A.G., Fossøy, F., Grosjean, M., Hope, M., Jeppesen, T.S., Køljalg, U., Lundin, D., et al. (2022). Introducing guidelines for publishing DNA-derived occurrence data through biodiversity data platforms. *Metabarcoding and Metagenomics*, 6, 239–244. DOI: 10.3897/mbmg.6.84960

Nishitani, G., Nagai, S., Hayakawa, S., Kosaka, Y., Sakurada, K., Kamiyama, T. & Gojobori, T. (2012). Multiple plastids collected by the dinoflagellate *Dinophysis mitra* through kleptoplastidy. *Appl. Environ. Microbiol.*, 78, 813–821. DOI: 10.1128/AEM.06544-11

Nübel, U., Garcia-Pichel, F. & Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl. Environ. Microbiol.*, 63, 3327–3332. DOI: 10.1128/aem.63.8.3327-3332.1997

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 Environmental Proceedings*, 166, 142 p. Retrieved from <https://epic.awi.de/id/eprint/30141/1/bsep106.pdf>

Oliver, A.K., Brown, S.P., Callahan, M.A. & Jumpponen, A. (2015). Polymerase matters: non-proofreading enzymes inflate fungal community richness estimates by up to 15%. *Fungal Ecol.*, 15, 86–89. DOI: 10.1016/j.funeco.2015.03.003

OSPAR Commission (2016). *Coordinated Environmental Monitoring Programme (CEMP) Guidelines: Phytoplankton monitoring*. Retrieved from https://mcc.jrc.ec.europa.eu/documents/OSPAR/CEMP_GuidelinesPhytoplanktonmonitoring.pdf

Ott, B.M., Litaker, R.W., Holland, W.C. & Delwiche, C.F. (2022). Using rDNA sequences to define dinoflagellate species. *PLoS ONE*, 17, e0264143. DOI: 10.1371/journal.pone.0264143

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. DOI: 10.1111/1462-2920.13023

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.A., Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.*, 50(D1), D785–D794. DOI: 10.1093/nar/gkab776

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S.S, Cepicka, I., Decelle, J., Dunthorn, M., et al. (2012). CBOL Protist Working Group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.*, 10, e1001419. DOI: 10.1371/journal.pbio.1001419

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., et al. (2018). The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Sci. Total Environ.*, 637–638, 1295–1310. DOI: 10.1016/j.scitotenv.2018.05.002

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*. 2010, 71 pp. DOI: 10.5167/uzh-187800
- Piredda, R., Tomasino, M.P., D'Erchia, A.M., Manzari, C., Pesole, G., Montresor, M., Kooistra, W.H.C.F., Sarno, D. & Zingone, A. (2017). Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiol. Ecol.*, 93, fiw200. DOI: 10.1093/femsec/fiw200
- Pochon, X., Bott, N.J., Smith, K.F. & Wood, S.A. (2013). Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS ONE*, 8, e73935. DOI: 10.1371/journal.pone.0073935
- Quail, M.A., Otto, T.D., Gu, Y., Harris, S.R., Skelly, T.F., McQuillan, J.A., Swerdlow, H.P. & Oyola, S.O. (2012). Optimal enzymes for amplifying sequencing libraries. *Nat. Methods*. 9, 10–11. DOI: 10.1038/nmeth.1814.
- 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 Res.*, 41(D1), D590–D596. DOI: 10.1093/nar/gks1219
- Rimet, F., Gusev, E., Kahlert, M., Kelly, M.G., Kulikovskiy, M., Maltsev, Y., Mann, D.G., Pfannkuchen, M., Trobajo, R., Vasselon, V., et al. (2019). Diat.barcode, an open-access curated barcode library for diatoms. *Sci. Rep.*, 9, 15116. DOI: 10.1038/s41598-019-51500-6
- Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. DOI: 10.7717/peerj.2584
- 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. DOI: 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 (Version 1). *Zenodo*. DOI: 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 (Version 1). *Zenodo*. DOI: 10.5281/zenodo.5232772
- Salmaso, N., Vasselon, V., Rimet, F., Vautier, M., Elersek, T., Boscaini, A., Donati, C., Moretto, M., Pindo, M., Riccioni, G., et al. (2022). DNA sequence and taxonomic gap analyses to quantify the coverage of aquatic cyanobacteria and eukaryotic microalgae in reference databases: Results of a survey in the Alpine region. *Sci. Total Environ.*, 834, 155175. DOI: 10.1016/j.scitotenv.2022.155175
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *P. Natl. Acad. Sci. U.S.A.*, 74, 5463–5467. DOI: 10.1073/pnas.74.12.5463
- Santoferrara, L.F. (2019). Current practice in plankton metabarcoding: Optimization and error management. *J. Plankton Res.*, 41, 571–582. DOI: 10.1093/plankt/fbz041
- Sawaya, N.A., Djurhuus, A., Closek, C.J., Hepner, M., Olesin, E., Visser, L., Kelble, C., Hubbard, K. & Breitbart, M. (2019). Assessing eukaryotic biodiversity in the Florida

- Keys National Marine Sanctuary through environmental DNA metabarcoding. *Ecol. Evol.*, 9, 1029–1040. DOI: 10.1002/ece3.4742
- Schlaeppli, K., Bender, S.F., Mascher, F., Russo, G., Patrignani, A., Camenzind, T., Hempel, S., Rillig, M.C. & van der Heijden, M.G.A. (2016). High-resolution community profiling of arbuscular mycorrhizal fungi. *New Phytol.*, 212, 780–791. DOI: 10.1111/nph.14070
- 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. DOI: 10.1128/AEM.01541-09
- Scholin, C.A., Herzog, M., Sogin, M. & Anderson, D.M. (1994). Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). II. Sequence analysis of a fragment of the LSU rRNA gene. *J. Phycol.*, 30, 999–1011. DOI: 10.1111/j.0022-3646.1994.00999.x
- Sepulveda, A.J., Hutchins, P.R., Forstchen, M., McKeefry, M.N. & Swigris, A.M. (2020). The elephant in the lab (and field): Contamination in aquatic environmental DNA studies. *Front. Ecol. Evol.*, 8, 440. DOI: 10.3389/fevo.2020.609973
- Serret, P., Aligizaki, K., Davidson, K., Coricato, P., Gago, A., Jakobsen, H.H., Salas, R. & Silva, A. (2019). *Monitoring of toxin-producing phytoplankton in bivalve mollusc harvesting areas. Guide to good practice: Technical application*. European Union Reference Laboratory for Marine Biotoxins. Issue 1, 44 pp. Retrieved from https://www.aesan.gob.es/en/CRLMB/docs/docs/procedimientos/Phyto_Monitoring_Guide_DEC_2021.pdf
- Sherwood, A.R. & Presting, G.G. (2007). Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. *J. Phycol.*, 43, 605–608. DOI: 10.1111/j.1529-8817.2007.00341.x
- Silveira, S., Kawakami, Y., Kanno, N., Kasai, H., Shiimoto, 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. DOI: 10.1016/j.hal.2019.101631
- Silveira, S., Laas, P., Kolesova, N., Lips, I., Lips, U. & Nagai, S. (2021). Plankton biodiversity and species co-occurrence based on environmental DNA – a multiple marker study. *Metabarcoding and Metagenomics*, 5, 175–197. DOI: 10.3897/mbmg.5.72371
- Singer, E., Bushnell, B., Coleman-Derr, D., Bowman, B., Bowers, R.M., Levy, A., Gies, E.A., Cheng, J.F., Copeland, A., Klenk, H.P., et al. (2016). High-resolution phylogenetic microbial community profiling. *ISME J.*, 10, 2020–2032. DOI: 10.1038/ismej.2015.249
- Singer, G.A.C., Fahner, N.A., Barnes, J.G., McCarthy, A. & Hajibabaei, M. (2019). Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater. *Sci. Rep.*, 9, 5991. DOI: 10.1038/s41598-019-42455-9
- Slatko, B.E., Gardner, A.F. & Ausubel, F.M. (2018). Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.*, 122, 1–11. DOI: 10.1002/cpmb.59
- Smith, K.F., Kohli, G.S., Murray, S.A. & Rhodes, L.L. (2017). Assessment of the

- metabarcoding approach for community analysis of benthic-epiphytic dinoflagellates using mock communities. *New Zeal. J. Mar. Fresh.*, 51, 555–576. DOI: 10.1080/00288330.2017.1298632
- Somervuo, P., Koskela, S., Pennanen, J., Nilsson, R.H. & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics*, 32, 2920–2927. DOI: 10.1093/bioinformatics/btw346
- 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. DOI: 10.1111/2041-210X.12721
- Sonnenberg, R., Nolte, A.W. & Tautz, D. (2007). An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Front. Zool.*, 4, 6. DOI: 10.1186/1742-9994-4-6
- Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey, E.S. & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Sci. Rep.*, 7, 12240. DOI: 10.1038/s41598-017-12501-5
- 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. DOI: 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. DOI: 10.1371/journal.pone.0189119
- Sze, M.A. & Schloss, P.D. (2019). The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. *mSphere*, 4, e00163-19. DOI: 10.1128/mSphere.00163-19
- Sze, Y., Miranda, L.N., Sin, T.M. & Huang, D. (2018). Characterising planktonic dinoflagellate diversity in Singapore using DNA metabarcoding. *Metabarcoding and Metagenomics*, 2, e25136. DOI: 10.3897/mbmg.2.25136
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.*, 21, 2045–2050. DOI: 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. DOI: 10.1093/oso/9780198767220.001.0001
- Tanabe, A.S., Nagai, S., Hida, K., Yasuike, M., Fujiwara, A., Nakamura, Y., Takano, Y. & Katakura, S. (2016). Comparative study of the validity of three regions of the 18S-rRNA gene for massively parallel sequencing-based monitoring of the planktonic eukaryote community. *Mol. Ecol. Resour.*, 16, 402–414. DOI: 10.1111/1755-0998.12459
- Tedersoo, L., Tooming-Klunderud, A. & Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol.*, 217, 1370–1385. DOI: 10.1111/nph.14776
- Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R.H., Kennedy, P.G., Yang, T., Anslan, S. &

- Mikryukov, V. (2022). Best practices in metabarcoding of fungi: From experimental design to results. *Mol. Ecol.*, 31, 2769–2795. DOI: 10.1111/mec.16460
- Tsiamis, K., Palialexis, A., Stefanova, K., Gladan, Ž.N., Skejic, S., Despalatovic, M., Cvitkovic, I., Drafacevic, B., Dulcic, J., Vidjak, O., et al. (2019). Non-indigenous species refined national baseline inventories: a synthesis in the context of the European Union's Marine Strategy Framework Directive. *Mar. Pollut. Bull.*, 145, 429–435. DOI: 10.1016/j.marpolbul.2019.06.012
- Utermöhl, H. (1958). Zur Vervollkommung der quantitativen phytoplankton-methodik. *Mitt. Int. Ver. Limnol.*, 9, 38.
- Vaulot, D., Geisen, S., Mahé, F. & Bass, D. (2022). pr2-primers: An 18S rRNA primer database for protists. *Mol. Ecol. Resour.*, 22, 168– 179. DOI: 10.1111/1755-0998.13465
- Vasselon, V., Domaizon, I., Rimet, F., Kahlert, M. & Bouchez, A. (2017). Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? *Freshw. Sci.*, 36, 162–177. DOI: 10.1086/690649.
- Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*, 73, 5261–5267. DOI: 10.1128/AEM.00062-07
- Wang, Z., Liu, L., Tang, Y., Li, A., Liu, C., Xie, C., Xiao, L. & Lu, S. (2022). Phytoplankton community and HAB species in the South China Sea detected by morphological and metabarcoding approaches. *Harmful Algae*, 118, 102297. DOI: 10.1016/j.hal.2022.102297
- Wasmund, N. & Siegel, H. (2008). Phytoplankton. In: Feistel, R., Nausch, G. & Wasmund, N. (eds.), *State and Evolution of the Baltic Sea, 1952–2005: A Detailed 50-Year Survey of Meteorology and Climate, Physics, Chemistry, Biology, and Marine Environment*. Wiley, pp. 441–481. DOI: 10.1002/9780470283134.ch15
- Wawrik, B., Paul, J.H. & Tabita, F.R. (2002). Real-time PCR quantification of *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. *Appl. Environ. Microbiol.*, 68, 3771–3779. DOI: 10.1128/AEM.68.8.3771-3779.2002
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. DOI: 10.1038/sdata.2016.18
- Wylezich, C., Nies, G., Mylnikov, A.P., Tautz, D. & Arndt, H. (2010). An evaluation of the use of the LSU rRNA D1-D5 domain for DNA-based taxonomy of eukaryotic protists. *Protist*, 161, 342–352. DOI: 10.1016/j.protis.2010.01.003
- Xie, Y. & Giesy, J.P. (2019). *Metabarcoding of Cyanobacteria Assembly*, Standard Operating Procedure UofS-ETL-EDNA-30, Version 1. Environmental Toxicology Laboratory, Toxicology Centre, University of Saskatchewan. Retrieved from <https://qwf.usask.ca/edna/documents/etl10030.pdf>
- Yan, D., Nishioka, J., Toyota, T. & Suzuki, K. (2022). Winter microalgal communities of the southern Sea of Okhotsk: A comparison of sea ice, coastal, and basinal seawater. *Progr. Oceanogr.*, 204, 102806. DOI: 10.1016/j.pocean.2022.102806
- 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. DOI: 10.3390/ijerph17207642
- Yoon, T., Kang, H., Kang, C., Lee, S.H., Ahn, D., Park, H. & Kim, H. (2016). Development of a cost-effective metabarcoding strategy for analysis of the marine phytoplankton community. *PeerJ*, 4, e2115. DOI: 10.7717/peerj.2115
- Yuan, J., Li, M. & Lin, S. (2015). An improved DNA extraction method for efficient and quantitative recovery of phytoplankton diversity in natural assemblages. *PLoS ONE*, 10, e0133060. DOI: 10.1371/journal.pone.0133060
- Zaiko, A., Greenfield, P., Abbott, C., von Ammon, U., Bilewitch, J., Bunce, M., Cristescu, M.E, Chariton, A., Dowle, E., Geller, J., et al. (2021). Towards reproducible metabarcoding data: Lessons from an international cross-laboratory experiment. *Mol. Ecol. Resour.*, 22, 519–538. DOI: 10.1111/1755-0998.13485
- Zhang, G.K., Chain, F.J.J., Abbott, C.L. & Cristescu, M.E. (2018). Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evol. Appl.*, 11, 1901–1914. DOI: 10.1111/eva.12694
- Zhang, Y., Pavlovskaya, M., Stoica, E., Prekrasna, I., Yang, J., Slobodnik, J., Zhang, X. & Dykyi, E. (2020). Holistic pelagic biodiversity monitoring of the Black Sea via eDNA metabarcoding approach: From bacteria to marine mammals. *Env. Internat.*, 135, 105307. DOI: 10.1016/j.envint.2019.105307
- Zimmermann, J., Jahn, R. & Gemeinholzer, B. (2011). Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Org. Divers. Evol.*, 11, 173. DOI: 10.1007/s13127-011-0050-6

Appendix

Table A1. Examples of guidelines and protocols for aquatic eDNA metabarcoding

Literature type	Environment	Target organism	Reference
Guidelines	Freshwater	Various	Pawłowski et al. (2020)
Guidelines	Freshwater	Various	Bruce et al. (2021)
Guidelines	Freshwater (pelagic and benthic)	Phytoplankton (eukaryotes and prokaryotes)	Elsersek et al. (2021)
Guidelines	Marine (pelagic)	Phytoplankton (eukaryotes and prokaryotes)	Karlson et al. (2018)
Guidelines	Marine (pelagic)	Phytoplankton (eukaryotes and prokaryotes)	Jerney et al. (2022)
Guidelines	Marine (pelagic)	HABs	Yarimizu et al. (2020)
Protocol (DNA-extraction)	Marine (pelagic)	Pro- and eukaryotes	Andersson et al. (2022)
Protocol (pipeline)*	Freshwater (pelagic)	Phytoplankton (eukaryotes)	Salmaso et al. (2021a)
Protocol (pipeline)*	Freshwater (pelagic)	Phytoplankton (prokaryotes)	Salmaso et al. (2021b)
Study	Marine (pelagic)	HABs	Silvever et al. (2019)

* part of the guidelines published by Elsersek et al. (2021).

Table A2. Overview of ribosomal rRNA genes, gene regions, and number of unique primer pairs used in marine metabarcoding studies targeting phytoplankton

Gene	Region targeted	Nr. of unique primer pairs	References
16S	V1-V2	2	Bolaños et al. (2020); Choi et al. (2020)
16S	V1-V3	1	Kolda et al. (2020)
16S	V3	2	Klindworth et al. (2013); Zhang et al. (2020)
16S	V3-V4	3	Herlemann et al. (2011); Li et al. (2019); Salmaso et al. (2022)
16S	V4	3	Caporaso et al. (2011); Apprill et al. (2015); Djurhuus et al. (2020)
16S	V4-V5	3	Parada et al. (2016); Djurhuus et al. (2017); MacNeil et al. (2022)
16S	cyanobacteria-specific V3-V4	1	Nübel et al. (1997)
16S	cyanobacteria-specific V6	1	Xie & Giesy (2019)
18S	Not specified	1	Mäki et al. (2017)
18S	diatom-specific V4	1	Zimmermann et al. (2011)
18S	dinoflagellate-specific	1	Kohli et al. (2014)
18S	V1-V3	2	Pochon et al. (2013); Tanabe et al. (2016)
18S	V4	15	Elwood et al. (1985); Bråte et al. (2010); Stoeck et al. (2010); Comeau et al. (2011); Hadziavdic et al. (2014); Hugerth et al. (2014); Balzano et al. (2015); Bradley et al. (2016); Piredda et al. (2017); Stat et al. (2017); Cordier et al. (2019); Latz et al. (2022); Salmaso et al. (2022); Wang et al. (2022); Yan et al. (2022)
18S	V4-V5	2	Sze et al. (2018); Joo et al. (2022)
18S	V6-V8	1	Latz et al. (2022)
18S	V7-V9	2	Nishitani et al. (2012); Tanabe et al. (2016)
18S	V9	3	Amaral-Zettler et al. (2009); Stoeck et al. (2010); Piredda et al. (2017); Latz et al. (2022)
18S-28S		4	Latz et al. (2022)
23S	not specified	2	Yoon et al. (2016); Sherwood & Presting (2007)
28S	D1-D2	2	Scholin et al. (1994), Sildever et al. (2019)
28S	D1-D3	1	Chai et al. (2018)
28S	D2-D3	1	Hamsher et al. (2011)
28S	dinoflagellate-specific D1-D2	1	Sildever et al. (2021)
COI	not specified	3	Jeunen et al. (2019); Sawaya et al. (2019); Bakker et al. (2019)
ITS	ITS1-ITS2	1	Fu et al. (2021b)
rbcL	not specified	1	Wawrick et al. (2002)

A1 Extended discussion on primers

The 18S rRNA gene contains 9 variable regions (V1-V9) with the V2, V4, and V9 regions displaying the highest variability (Hadziavdic et al., 2014). At the same time, the V6 region is more conserved and thus usually not targeted (Hadziavdic et al., 2014). Based on the comparison of the V1-V3, V4-V5, and V7-V9 regions for eukaryotic plankton monitoring, Tanabe et al. (2016) suggest targeting the V1-V3 region due to the highest sequence variability and significantly higher identification power than the V4-V5 region. However, as the length of the V1-V3 region in eukaryotes ranges between 400 and 800 bp (Tanabe et al., 2016), it is currently not possible to sequence this region as the longest read length is limited to 2x300 bp by the Illumina MiSeq platform (Slatko et al., 2018).

The 28S rRNA gene shows more variability compared to the 18S rRNA gene (Hillis & Dixon, 1991) and may thus enhance the identification of some taxa (John et al., 2003; Murray et al., 2005; Sonnenberg et al., 2007; Wylezich et al., 2010), but see Forster et al. (2019). In eukaryotes, the gene contains 12 divergent regions (Hassouna et al., 1984; Lenaers et al., 1989) that have variable evolutionary rates (Hillis & Dixon, 1991), with the D1-D2 region being the most informative (based on dinoflagellates; Ki & Han, 2007). This region has been targeted for eDNA metabarcoding and HTS-based eukaryotic diversity studies using Roche 454 sequencing platforms (Elferink et al., 2017; Grzebyk et al., 2017) with read lengths up to 700 bp (single-end) (Kchouk et al., 2017). However, the 28S D1-D2 region with the length of 500–790 bp (in most phytoplankton phyla, together with the primers) is currently not possible to sequence with the MiSeq platform due to the read length limitation (Grzebyk et al., 2017).

A2 Glossary

Terms and explanations partly adapted from Pawlowski et al. (2020).

Term	Explanation
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 are used for phylogenetic studies of eukaryotes as their slow evolutionary rate makes them suitable to reconstruct ancient divergences.
Alignment	A way of arranging the DNA sequences to identify regions of similarity. Aligned sequences are typically represented as rows within a matrix so that identical or similar characters are aligned in successive columns. Global alignment tries to align every nucleotide in every sequence. Local alignment aligns parts of sequences.
Amplicon	A single-stranded DNA that is amplified in a PCR reaction. In the case of DNA barcoding, amplicon is the DNA barcode.
ASV	Amplicon Sequence Variant; a quality-controlled unique variant of the marker gene present in the sample.
Barcode	A barcode sequence is a short nucleotide sequence from a standard genetic locus for use in species identification.
BLAST	The Basic Local Alignment Search Tool; finds regions of local similarity between sequences.
Bulk sample	A sample consisting of whole organisms and their fragments originating from the environment and collected manually.

CEN	European Committee for Standardization.
Chimera	Genomic artefact created during PCR amplification by combining DNA fragments of different origins.
COI	Cytochrome c Oxidase subunit I.
Contamination	Presence of extraneous DNA, which does not originate from the sample.
Curation	In the case of reference libraries, curation means that each DNA barcode comes with comments on the reliability of its origin, i.e., a taxonomic expert has identified the specimen when DNA extraction was conducted, or an expert has checked for mislabelings (to wrong species).
DNA barcoding	Species identification method based on a short section of DNA (DNA barcode). Relies on a curated reference library that is populated with DNA barcodes of reliably identified specimens from different species.
DNA extraction	A laboratory process of chemical and physical steps to release and purify DNA from cells or other material.
DNA metabarcoding	Simultaneous identification of several species within a set of samples using DNA barcodes.
DNA template	Target DNA from a specimen, environment or sample. Specific part of this DNA will be amplified in a PCR reaction.
eDNA	Environmental DNA; free or particle-bound DNA found in an environment. Particle-bound DNA can be attached to surfaces of particles or it can be DNA inside detached cells (dead or alive). Thus, DNA inside single-celled organisms is included in eDNA by definition.
GBIF	Global Biodiversity Information Facility.
Genetic marker	A genomic DNA region (e.g. fragment of COI gene, or V9 region of 18S rRNA gene), which allows to identify species within a particular taxonomic group.
Genome	All genetic information of an organism (nucleotide sequences of DNA).
GES	Good Environmental Status.
HAB(s)	Harmful Algal Bloom; excessive algal growth that causes negative effects to other organisms.
HELCOM	Helsinki Commission; the 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).
INSDC	International Nucleotide Sequence Database Collaboration.
ISO	International Organization for Standardization.
ITS	Internal Transcribed Spacer.
Kit	A set of articles or equipment needed for a specific purpose. DNA extraction kit includes chemicals and plastics needed for DNA extraction.
LCA	Lowest Common Ancestor. In a phylogenetic tree, the lowest (deepest) node that has both species a and species b as descendants. In other words, the first node in a tree that connects the two species.
Marker gene	A gene that can be used to delineate between taxonomic lineages, i.e., identify different species.
MAG	A Metagenome-Assembled Genome is a single-taxon assembly based on one or more binned metagenomes that has been asserted to be a close representation to an actual individual genome (that could match an already existing isolate or represent a novel isolate).
MCE filter	Mixed Cellulose Ester 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.
MSFD	EU Marine Strategy Framework Directive.
Negative controls	Measures that allow tracking potential contamination during field sampling, DNA extraction, and PCR.

NIS	Non-Indigenous Species.
OBIS	Ocean Biodiversity Information System.
OSPAR	The Convention for the Protection of the Marine Environment of the North-East Atlantic.
OTU	Operational Taxonomic Unit. A cluster of sequences grouped by similarity that is 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. Sometimes referred to as molecular OTU (MOTU).
PCR	Polymerase Chain Reaction. Amplification of specific DNA region. Requires DNA template, polymerase, forward and reverse primers, nucleotides (DNA bases), suitable chemical environment (buffer), and cations (e.g. Mg ²⁺ and K ⁺ ions). Carried out in a thermal cycler that heats and cools the reaction: heats to high temperature to denature DNA (breaks double-stranded DNA to single-stranded DNAs), cools to lower temperature for primers to anneal to single-stranded DNA, heats to middle temperature where the polymerase builds a copy of the single-stranded DNA by adding nucleotides based on the DNA template. This cycle is repeated several times.
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 builds a new DNA strand from the DNA template in PCR reaction.
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. Primer anneals to single-stranded DNA template, and the polymerase starts adding nucleotides complementary to the template from the end of the primer in the PCR reaction.
PS filter	PolyetherSulfone filter.
rbcL gene	Ribulose-Bisphosphate Carboxylase gene.
Reads	A common term used for DNA sequences generated during high-throughput sequencing.
Reference database/library	A collection of DNA barcodes from different species. Can be well-curated, meaning that an expert of each organism group has identified the species when DNA barcode was collected from the specimen, or not so well-curated, meaning that the reliability of each DNA barcode is not checked by experts.
Replicate	Repeated DNA sampling or PCR amplification to estimate the variability associated with the method and control the consistency of obtained results.
rRNA gene	Ribosomal RNA gene; a gene that codes ribosomal RNA, present both in prokaryotes and eukaryotes.
Sanger sequencing	Sanger sequencing is a method of DNA sequencing that involves electrophoresis and is based on the random incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication. The method was developed by the two-time Nobel Laureate Frederick Sanger and his colleagues in 1977, hence the name the Sanger Sequence.
Sequencing	The process of determining the order of nucleotides in a DNA strand.
Sterivex filter	A type of enclosed filter. Filter is enclosed in a capsule during filtration and DNA extraction. Sterivex filter is combined with a syringe to press the sample through the filter.
Taxonomic assignment	The taxonomic identification of DNA sequences based on reference databases.
WFD	EU Water Framework Directive.

About this publication

DNA metabarcoding

Guidelines to monitor phytoplankton diversity and distribution in marine and brackish waters

Jacqueline Jerney, Heidi Hällfors, Hans Jakobsen, Iveta Jurgensone, Bengt Karlson, Anke Kremp, Sirpa Lehtinen, Markus Majaneva, Kristian Meissner, Veera Norros, Sirje Sildever, Sanna Suikkanen, Karolin Teeveer

ISBN 978-92-893-7519-1 (PDF)

ISBN 978-92-893-7520-7 (ONLINE)

<http://dx.doi.org/10.6027/temanord2023-505>

TemaNord 2023:505

ISSN 0908-6692

© Nordic Council of Ministers 2023

Cover photo: USGS/Unsplash

Published: 12/4/2023

Disclaimer

This publication was funded by the Nordic Council of Ministers. However, the content does not necessarily reflect the Nordic Council of Ministers' views, opinions, attitudes or recommendations.

Rights and permissions

This work is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0>.

Translations: If you translate this work, please include the following disclaimer: This translation was not produced by the Nordic Council of Ministers and should not be construed as official. The Nordic Council of Ministers cannot be held responsible for the translation or any errors in it.

Adaptations: If you adapt this work, please include the following disclaimer along with the attribution: This is an adaptation of an original work by the Nordic Council of Ministers. Responsibility for the views and opinions expressed in the adaptation rests solely with its author(s). The views and opinions in this adaptation have not been approved by the Nordic Council of Ministers.

Third-party content: The Nordic Council of Ministers does not necessarily own every single part of this work. The Nordic Council of Ministers cannot, therefore, guarantee that the reuse of third-party content does not infringe the copyright of the third party. If you wish to reuse any third-party content, you bear the risks associated

with any such rights violations. You are responsible for determining whether there is a need to obtain permission for the use of third-party content, and if so, for obtaining the relevant permission from the copyright holder. Examples of third-party content may include, but are not limited to, tables, figures or images.

Photo rights (further permission required for reuse):

Any queries regarding rights and licences should be addressed to:

Nordic Council of Ministers/Publication Unit

Ved Stranden 18

DK-1061 Copenhagen

Denmark

pub@norden.org

Nordic co-operation

Nordic co-operation is one of the world's most extensive forms of regional collaboration, involving Denmark, Finland, Iceland, Norway, Sweden, and the Faroe Islands, Greenland and Åland.

Nordic co-operation has firm traditions in politics, economics and culture and plays an important role in European and international forums. The Nordic community strives for a strong Nordic Region in a strong Europe.

Nordic co-operation promotes regional interests and values in a global world. The values shared by the Nordic countries help make the region one of the most innovative and competitive in the world.

The Nordic Council of Ministers

Nordens Hus

Ved Stranden 18

DK-1061 Copenhagen

pub@norden.org

Read more Nordic publications on www.norden.org/publications