Acta Universitatis Agriculturae Sueciae

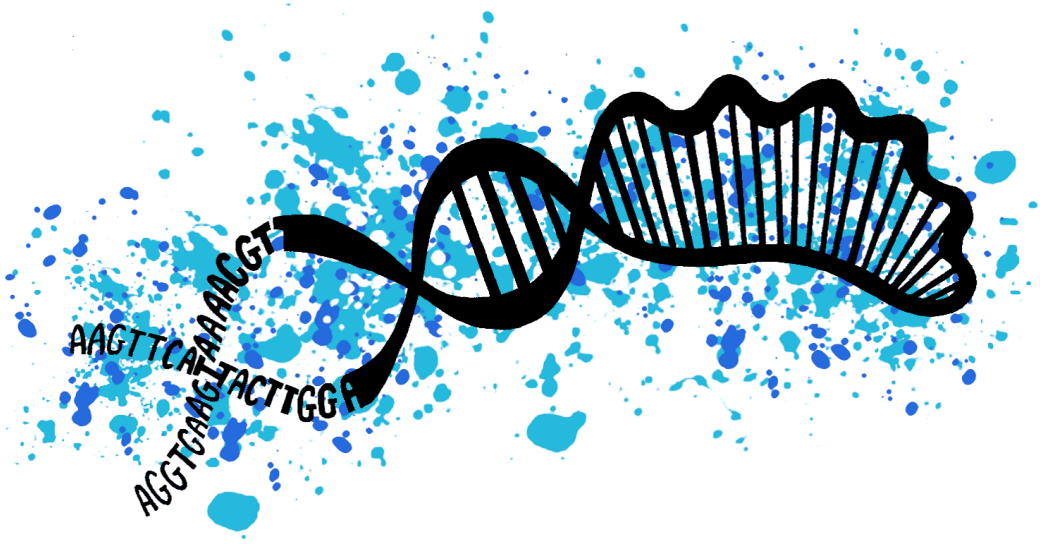
SLU

Doctoral Thesis No. 2021:55
Faculty of Natural Resources and Agricultural Sciences

New methods for improving water management

Exploring the role of diatoms in ecosystems

Bonnie Bailet



New methods for improving water management

Exploring the role of diatoms in ecosystems

Bonnie Bailet
Faculty of Natural Resources and Agricultural Sciences
Department of Aquatic Sciences and Assessment
Uppsala

SLU
SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Uppsala 2021

Acta Universitatis agriculturae Sueciae
2021:55

Cover: Diatom structure unfolding into DNA
(Illustration: B. Bailet)

ISSN 1652-6880
ISBN (print version) 978-91-7760-788-5
ISBN (electronic version) 978-91-7760-789-2
© 2021 Bonnie Bailet, Swedish University of Agricultural Sciences
Uppsala
Print: SLU Service/Repro, Uppsala 2021

New methods for improving water management - Exploring the role of diatoms in ecosystems

Abstract

Diatoms are photosynthetic microalgae which are well known to react quickly to environmental changes and are widely used as bioindicators within Water Framework Directive. However, diatom identification by light microscopy is time-consuming and error-prone, creating a need for alternative, faster and more reproducible methods for characterising diatom communities. This thesis describes the development and application of two such methods: a molecular approach using metabarcoding and a trait-based approach using combinations of different diatom morphological traits. Both approaches were shown to have potential for use in environmental assessment, as they revealed the response of diatoms to environmental change and enabled better representation of the diatom community. However, results from the two approaches were not directly comparable to results obtained by conventional microscopy and diatom DNA and trait data should not be used with tools calibrated for morphotaxa data. Discrepancies in molecular output data between laboratories revealed a strong need for standardisation and best-practice guidelines. Nutrient indices developed for each novel approach performed well and showed potential for use in assessment of total phosphorus levels in Fennoscandia freshwater. The molecular-based index showed differences in performance for stream and lake samples. The trait-based index performed equally well for both types of site, but current limitation of this index is that selection of meaningful traits must be done for each new gradient. Further work is needed on diatom traits and their response to environmental changes, to obtain reproducible and comparable molecular outputs and to continue the development of new indices calibrated to diatom DNA and trait data.

Keywords: ecological status, eutrophication, Bacillariophyta, morphological identification, 18S-V4, rbcL, ISU, ASV, trait classification, ecological guilds.

Author's address: Bonnie Bailet, Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment, PO Box 7050, SE750 07, Uppsala, Sweden.

Nya metoder för att förbättra vattenförvaltningen - utforska kiselalgernas roll i ekosystem

Abstract

Kiselalger är fotosyntetiska alger och används väl som bioindikatorer inom vattendirektivet inom ramen för övervakning och miljöbedömning av europeiska vatten. Men deras identifiering med ett ljussmikroskop mikroskopiskt är tidskrävande och har felkällor, vilket skapar ett behov av alternativa, snabbare och mer reproducerbara metoder för att karaktärisera kiselalgssamhället ökat. I denna avhandling testar vi utvecklingen och tillämpningen av två nya tillvägagångssätt: ett molekylärt tillvägagångssätt med metabarkodningsmetoden och ett egenskapsbaserat tillvägagångssätt med kombinationer av olika morfologiska drag. Båda tillvägagångssätten har potential för användning i miljöbedömning eftersom de visat starka responser från kiselalgerna på miljöförändringar, en bättre representation av kiselalgssamhället. Resultat från de två nya tillvägagångssätten är emellertid inte jämförbara med dem i den traditionella mikrokopimetoden. Alltså, kiselalgernas DNA-data och drag ska inte användas med verktyg kalibrerade för taxonomiska data. Skillnader i molekylära utdata mellan laboratorier skiljer sig åt och att det finns ett stort behov av standardisering och riktlinjer för bästa praxis. De nya indexen fungerar bra och visar potential för bedömning av totala fosforhalter i Fennoscandias sötvatten. Det molekylärbaserade indexet visade en differentiell prestanda i bäckar och sjöar. Det egenskaper indexet å andra sidan fungerade lika bra för båda typerna av vattenförekomster, men nuvarande begränsningar för detta index är att valet av meningsfulla egenskaper måste göras för varje ny gradient. Ytterligare arbete behövs med kiselalger egenskaper och deras reaktion på miljön för att utveckla mer reproducerbara och jämförbara molekylära utdata och för att hjälpa utvecklingen av nya index kalibrerade till kiselalgernas DNA och egenskaper.

Nyckelord: ekologisk status, eutrofiering, Bacillariophyta, morfologisk identifiering, 18S-V4, rbcL, ISU, ASV, egenskapsklassificering, ekologiska guilds.

De nouvelles méthodes améliorant la gestion de l'eau – explorer le rôle des diatomées dans les écosystèmes

Résumé

Les diatomées sont des algues photosynthétiques et utilisées comme bioindicateurs pour l'évaluation environnementale des eaux européennes. Cependant, leur identification au microscope optique est chronophage et sujette à des erreurs, générant un besoin de méthodes alternatives, plus rapides et plus reproductibles pour caractériser les communautés des diatomées. Dans cette thèse, nous testons le développement et l'application de deux nouvelles approches : une approche moléculaire utilisant la méthode du « metabarcoding » et une approche basée sur les traits utilisant des combinaisons de différents traits morphologiques des diatomées. Les deux nouvelles approches ont un bon potentiel d'utilisation dans l'évaluation environnementale car elles reflètent clairement la réponse des diatomées aux changements environnementaux et offrent une meilleure représentation de la communauté des diatomées. Cependant, les résultats des deux nouvelles approches ne sont pas comparables à ceux de l'approche de microcopie traditionnelle et les données ADN et de traits de diatomées ne devraient pas être utilisées avec des outils calibrés pour les données de microscopie. En outre, nous avons constaté un fort besoin de standardisation et de lignes directrices sur les meilleures pratiques à adopter. Un indice écologique pour chaque approche a aussi été développé et testé. Les nouveaux indices montrent un fort potentiel pour l'évaluation des niveaux de phosphore total dans les eaux douces de la Fennoscandie. L'indice moléculaire a montré une performance différentielle entre les sites de cours d'eau et de lacs. L'indice basé sur les traits, en revanche, a donné de bons résultats pour les deux types de sites, mais des traits représentatifs doivent être sélectionnés pour chaque nouveau gradient. De futures travaux devront approfondir notre compréhension des traits des diatomées et de leur réponse à l'environnement, permettre des résultats moléculaires plus reproductibles et comparables, et aider au développement de nouveaux indices calibrés sur les données d'ADN et de traits des diatomées.

Mots clés: état écologique, eutrophisation, Bacillariophyta, identification morphologique, 18S-V4, rbcL, ISU, ASV, classification des traits, guildes écologiques.

Dedication

À mon Papi Caillou, qui était là au début de cette aventure et qui aurait aimé être là pour la fin.

“The truth is everyone is winging it.

So, I say spread your wings and follow your dreams.”

Charlie Mackesy.

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Bailet, B**., Bouchez A., Franc A., Frigerio, J.M., Keck, F., Karjalainen S.M., Rimet F., Schneider, S. & Kahlert, M. (2019). Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. *Metabarcoding & Metagenomics* 3:21-35

II. **Bailet, B**., Apothéloz-Perret-Gentil, L., Baričević A., Chonova, T., Franc, A., Frigerio, J.M., Kelly, M., Mora, D., Pfannkuchen, M., Proft, S., Ramon, M., Vasselon, V., Zimmermann, J. & Kahlert, M. (2020). Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Science of the Total Environment* 745, 140948.

III. **Bailet, B**., McKie, B., Truchy, A., Muotka, T., Jyväsjärvi, J., Soininen, J., Huusko, A., Teittinen A. & Kahlert, M. The impact of multiple stressors on diatom traits, species composition and molecular diversity in stream ecosystems – experimental effects of light, flow, and nutrient manipulations. (*Manuscript*).

IV. Kahlert, M., **Bailet, B**., Chonova, T., Karjalainen, S.M., Schneider, S. & Tapolczai, K. Same same, but different: The response of diatoms to environmental gradients in Fennoscandian streams and lakes – barcodes, traits and microscopic data compared. (*Ecological indicators - Accepted manuscript*).

Papers I-II are reproduced with the permission of the publishers.

The contribution of Bonnie Bailet to the papers included in this thesis was as follows:

I. Data curation, Conceptualisation, Formal analysis, Methodology, Investigation, Validation, Visualisation, Writing of original draft, Writing review and editing.

II. Data curation, Conceptualisation, Formal analysis, Methodology, Investigation, Validation, Visualisation, Writing of original draft, Writing review and editing.

III. Data sampling and treatment, Data curation, Formal analysis, Methodology, Investigation, Validation, Visualisation, Writing of original draft, Writing review and editing.

IV. Data curation, Conceptualisation, Writing review and editing.

Abbreviations

ASV	Amplicon sequence variant
COST	European cooperation in science and technology
DNA	Deoxyribonucleic acid
HTS	High-Throughput Sequencing
IPS	Indice de polluosensibilité spécifique (Index of Pollution Sensitivity)
ISU	Individual sequence unit
NMDS	Non-metric multidimensional scaling
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
PGM	Personal genome machine
RDA	Redundancy analysis
TOC	Total organic carbon
TotN	Total nitrogen
TotP	Total phosphorus
WFD	Water Framework Directive

1. Explanation of frequently used terms and abbreviations

Environmental assessment

Environmental assessment is a structured approach of focused and systematic study to identify and evaluate the impact of environmental changes and to help regulate or compensate for the negative effects on ecosystems and the services they provide. It combines regular data collection (see "Monitoring") and analysis. It includes the development of methods for data collection (what is collected and how), the development of predictive models, data quality, reproducibility, storage and accessibility, but also further development of environmental monitoring and assessment.

(Bio)monitoring

Ideally, monitoring refers to collection of environmental data in a long-term, spatially comprehensive and standardised way. Biomonitoring is the process of tracking changes in living organisms.

Taxonomy

Taxonomy is the scientific study of defining, naming and classifying groups of biological organisms, living or extinct, based on shared characteristics. For identification and description, organisms are grouped as entities called 'taxa' (singular 'taxon'). The term 'taxonomic identification' is often used to refer to recognition of the identity of an organism. However, in this thesis, the process of recognising a diatom's identity is distinguished

from description or classification of diatom taxa. For the sake of clarity, the term “diatom identification” is used throughout the thesis.

Diatom identification

Diatom identification is the recognition and naming of a certain diatom unit by taxon-specific characters. In this thesis, taxon refers to any diatom unit identified by its taxonomic name, including diatom units identified to below species level and above genus level.

- traditionally by morphological characteristics

Diatoms are a group of photosynthetic aquatic organisms (algae) with the special feature of having a cell wall (called the frustule) made from silica. The form and structures of the frustule differ between diatom genera and species, and the frustule is therefore generally used for diatom identification. The cell is first oxygenised to clean the frustule, then observed under a light microscope and identified by comparing the frustule structures with descriptions and pictures in standard literature.

- by (meta)barcoding

Barcoding is a molecular method of entity identification using a short DNA fragment from a specific gene. Metabarcoding is the parallel identification of multiple taxa to evaluate the entire community composition. In diatom metabarcoding, diatom taxa are identified by assessing the similarity of one or more certain DNA sequences, called DNA markers, to another sequence from a known diatom taxon.

Traits

The most widely used definition of a trait is that by Violle *et al.* (2007): “any morphological, physiological or phenological measurable feature at the individual level”.

Bioinformatics pipeline

A well-formulated definition is given by Roy *et al.* (2018): "A bioinformatics pipeline is composed of a wide array of software algorithms to process raw sequencing data and generate a list of annotated sequence variants. Bioinformatics pipelines are either designed and developed by a vendor with or without customization by the laboratory or entirely developed by the laboratory."

Reference database

A reference database is a library of reference data, physically (e.g. to store reference specimens) or digitally. It combines and archives information from multiple sources, to make it readily available to users. Typical examples are reference databases of DNA barcodes, which gather barcode information from a wide number of taxa into a single repository to enable molecular identification of taxa.

2. Introduction

2.1 Environmental assessment and monitoring of water bodies

Water is one of the most precious resources on Earth. Water bodies such as oceans, lakes and rivers play host to complex ecosystems that provide humans with irreplaceable benefits, such as supporting services (nutrient cycling, primary production), regulatory services (climate regulation, buffer zones), provisioning services (freshwater, raw material) and cultural services (recreation, tourism, education). However, water resources world-wide are often impacted by human activities (pollution, overuse, landscape alteration, etc.) which can alter, sometimes irremediably, the ecosystems within. For example, eutrophication of water bodies is one of the most common stressors in Europe (Carvalho *et al.*, 2019, Poikane *et al.*, 2021). Remediation and rehabilitation of water bodies is costly, slow and sometimes impossible. Therefore environmental assessment of water resources is important to ensure prevention or early detection of problems in unaltered water bodies and to monitor progress in water bodies undergoing remediation. Regional, national and international agreements, e.g. the European Union (EU) Water Framework Directive (WFD: European Union, 2000) or the US Clean Water Act (CWA: United States, 1972), aim to protect global water resources and ensure that human impacts are kept under control. Several types of monitoring data are required under these agreements for environmental assessment of water bodies and all rely on regular collection of standardised data. Some are chemistry-based, such as target and non-target screening of pollutants, and others are biology-based, such as calculating biological indicators based on the fauna found in water bodies. The EU WFD requires

environmental assessment to be conducted with the focus on biology, i.e. monitoring and analysis of data on organisms. Methods usually focus on collection of data for one target group, such as fish species, macroinvertebrates, macrophytes or diatoms. Diatoms, which are microscopic algae with complex silica structures, are widely used as bioindicators of freshwater bodies, since they are photosynthetic microorganisms and are thus the first to reflect changes occurring in an ecosystem. Within the WFD, environmental assessment using diatom communities is recognised as a robust and reliable method (Kelly *et al.*, 2014, Kelly *et al.*, 2009). Diatoms are frequently used to assess many different environmental changes (such as surface water acidity or hydrological and climate changes (Smol and Stoermer, 2010).

2.2 Conventional identification of diatom taxa

There is a great diversity of diatoms, which are usually identified based on features of their silica structures. These structures allow for identification of recent or current communities, but also for identification of older samples, e.g. from sediment cores, since they remain intact for quite some time after the death of the cell. The number of diatom species in existence is often discussed and constantly changing, since diatom taxonomy is subject to constant alterations, but according to recent estimates there are 12,000-30,000 species (Guiry, 2012, Mann and Vanormelingen, 2013). Some diatom species are also very small or distinguishable from one another only in terms of specific minor features. Therefore, identification of diatoms under a microscope, while proven to be efficient and reliable for environmental assessment, requires extensive expertise and is time-consuming. It may also be difficult for agencies in need of monitoring to acquire access to a diatom expert or to the funds necessary to assess the diatom communities in numerous samples. Even when it is performed, morphological identification of diatom species is prone to uncertainties, often leading to discrepancies in results between laboratories (Kahlert *et al.*, 2016, Kahlert *et al.*, 2009, Kahlert *et al.*, 2012, Werner *et al.*, 2016). For future water biomonitoring, there is thus a need to develop new methods that are as reliable as morphological identification of organisms, but that are faster, cheaper and more reproducible. This thesis focuses on two novel alternative approaches for characterisation of freshwater benthic diatoms within the scope of

environmental assessment and monitoring of Fennoscandian streams and lakes: the molecular metabarcoding approach and the traits-based approach.

2.3 Barcoding and Metabarcoding

With the rise of molecular methods in the past 20 years or so, barcoding of species has been put forward as an alternative method for identification of organisms. Barcoding enables sequencing of specific species based on identification of barcode regions, i.e. regions of DNA that display sufficient diversity to distinguish species from one another and that are flanked by regions which are sufficiently conserved to support creation of a matching primer (Zimmermann *et al.*, 2011). Thousands of taxa have already been sequenced, and the barcode sequences have been deposited in reference libraries like the National Centre for Biotechnology Information (NCBI) GenBank (Sayers *et al.*, 2019) or the Barcode Of Life Database (BOLD: Ratnasingham and Hebert, 2007). In many cases, but not all, those taxa were deposited also with a species name and additional information, but such metainformation is often not complete or correct especially for small organisms such as diatoms (Rimet *et al.*, 2021). The subsequent development of High-Throughput Sequencing (HTS) has enabled massive parallel sequencing and the emergence of associated technologies capable of sequencing longer fragments of DNA has improved the quality of the results obtained (Hajibabaei *et al.*, 2007, Hebert *et al.*, 2003, Moritz and Cicero, 2004, Stoeckle, 2003). It has also enabled production of large amounts of data. Combining HTS with barcoding gave rise to the metabarcoding method, which allows for identification of the entire community in a sample, instead of a single organism (Abarca *et al.*, 2014, Taberlet *et al.*, 2012, Visco *et al.*, 2015). This in turn has enabled use of environmental DNA or eDNA, defined by Pawlowski *et al.* (2020) as "DNA isolated from environmental samples, in contrast to genomic DNA that is extracted directly from specimens", which is a less invasive method than sampling large organisms such fish and can also be used for microscopic organisms such as diatoms, with greater speed and precision (Pawlowski *et al.*, 2018).

2.4 Molecular-based approach for diatom identification

For diatoms, the sampling method in the molecular approach is similar to that used to collect live samples for microscopic identification (i.e. brushing stones to collect biofilm). The two most widely used barcode regions for diatom sequencing are the V4 region of the nuclear-encoded 18S gene (18S ribosomal RNA) and a fragment of the plastid gene rbcL (ribulose-1,5-bisphosphate carboxylase-oxygenase). These markers have been shown to capture most of the community diversity when used with the metabarcoding approach (Pawlowski *et al.*, 2018, Vasselon *et al.*, 2017, Zimmermann *et al.*, 2015). The sequenced fragment is then compared against reference sequences to be identified to a species name, a step commonly called 'taxonomic assignment'. The question of which reference database to use for taxonomic assignment often arises and, while there are many diatom barcodes available in the BOLD or GenBank databases, many laboratories have created their own reference libraries from their own molecular identifications of diatoms (e.g. Zimmermann *et al.*, 2014). In fact, while the public databases hold a great number of barcodes, many of these were sequenced with older technologies or the reference specimen identification was doubtful or not done at all. One open-access diatom database, diat.barcode, previously R::syst (Rimet *et al.*, 2016, Rimet *et al.*, 2019), is dedicated to identification of diatom communities from natural samples. It compiles diatom reference barcodes from cultured specimens from INRA (Institut National de la Recherche Agronomique, France) and from the NCBI database (on both the 18S and rbcL DNA markers) and is continually curated and updated. However, while choice of reference database is often studied and discussed, the difficulties created by the broad diversity of bioinformatics pipelines now available have not yet received much attention.

2.5 Trait-based approaches to characterise diatom communities

An alternative approach to morphological taxa-based indices is to study the usability of trait-based diatom indices. Diatom biomonitoring traditionally relies on taxonomic units (genus or species), because the ecological indices available, such as the Indice de Polluo-sensitivite specifique (IPS: Cemagref, 1982), Trophic Diatom Index (TDI: Kelly and Whitton, 1995), the Periphyton Index of Trophic status (PIP) and

Acidification Index Periphyton (PIT and AIP: Schneider and Lindstrøm, 2009, Schneider and Lindstrøm, 2011), are constructed around taxa-specific ecological values. However, defining the ecological profile of a taxon is complex, especially in the case of rare species or cosmopolitan species found in different ecoregions, because a taxon's response to its environment may vary with geography or habitat. Therefore, use of these taxa-based indices in different ecoregions needs to be carried out with caution concerning the robustness of the environmental assessment and it should be borne in mind that they still carry misconceptions from the morphology approach (Tapolczai *et al.*, 2016). Taxonomical relationships can also be a poor predictor of ecological similarity and, even within a taxon, life forms may vary substantially, and functional adaptations can depend on the local conditions (Kruk *et al.*, 2010). Physiological and morphological properties define the capacity of an organism to compete and adapt in a particular habitat, and there is often high redundancy at the species level in terms of adaptive strategies (Kelly, 2013). On the other hand, use of ecological guilds, i.e. groups composed of taxa with similar adaptive strategies that exploit the same kind of resources in comparable ways, has the potential to simplify the ecological assessment (Tapolczai *et al.*, 2016) and overcome the challenges of taxa identification. The use of trait-based approaches can provide valuable additional information on organisms' response to environmental changes and such approaches would benefit from wider geographical application, since similar traits are found in similar ecosystems (Stevenson *et al.*, 2010), simplifying the intercalibration process. Trait-based classification for biomonitoring has seen an increase in development and applications over the past decade, for example with fish (Logez *et al.*, 2013), macroinvertebrates (Dolédec and Statzner, 2008, Borja *et al.*, 2009), macrophytes (Orfanidis *et al.*, 2003, Wells *et al.*, 2007), and phytoplankton (Bichoff *et al.*, 2018, Kruk *et al.*, 2010, Moser *et al.*, 2017, Neif *et al.*, 2017, Padisák *et al.*, 2006). However, application of such approaches is still relatively new for diatoms and there is much less information available, often scattered across the literature, on benthic diatom traits in comparison with organism groups (Rimet and Bouchez, 2012).

2.5.1 Diatom traits

Diatoms are fragile microscopic organisms, which makes identification of measurable traits difficult, particularly traits that are directly linked to

ecological functions. Nevertheless, diatoms exhibit a variety of functional traits (also called 'effect traits') and morphological traits (also called 'response traits'), such as capacity to fix environmental nitrogen, production of mucus or life forms such as colony formation, size and biovolume, among others. Identification of these key features can be used to estimate the diatom community and benefits from a direct link to ecological processes and ecosystem structure. Passy (2007) defined three ecological guilds for diatoms using such traits, later modified by Rimet and Bouchez (2012) linking them to environmental constraints such as nutrient availability and flow disturbances. Several studies have used this classification for biomonitoring and environmental assessment with freshwater diatoms (Berthon *et al.*, 2011, Rusanov and Khromov, 2016, B-Béres *et al.*, 2016, B-Béres *et al.*, 2017, Dong *et al.*, 2016, Lindholm *et al.*, 2018, Schneck *et al.*, 2017, Soininen *et al.*, 2016, Tang *et al.*, 2016). While this classification has shown promising results, the accuracy of the assessment is still limited by incomplete knowledge of the relationship between diatom traits and their functions. The challenge for diatomists, when seeking to build a robust trait-based evaluation system, is to define ecologically meaningful traits that can be used for the development of an adequate number of functional groups of diatoms, covering as many different habitat types as possible (Tapolczai *et al.*, 2016).

Taking all these aspects into account, there is a strong need for assessment, development and comparison of new tools based on these novel approaches for eutrophication assessments in Fennoscandian streams and lakes.

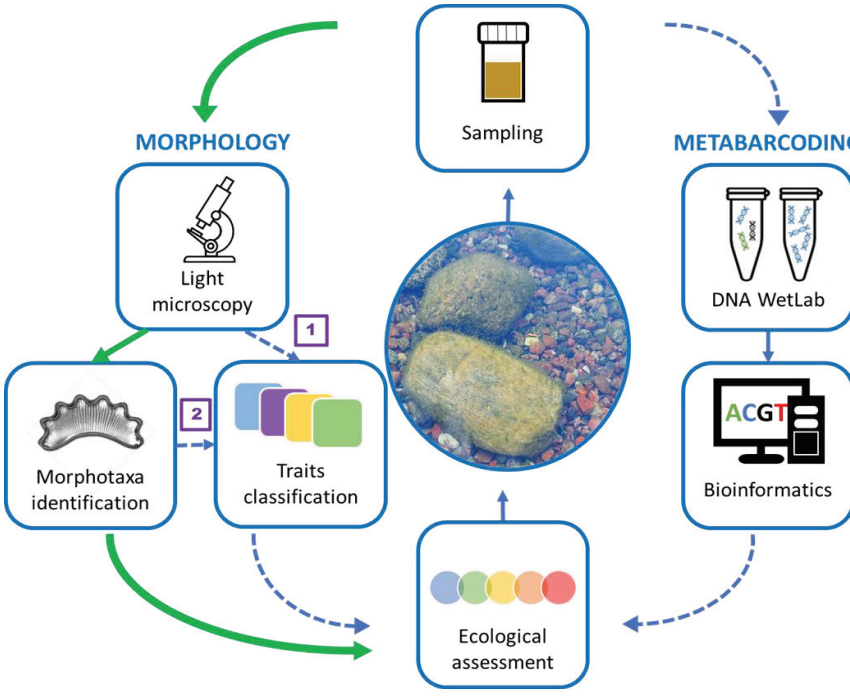


Figure 1. Schematic diagram of key steps in biological monitoring and environmental assessment using the conventional morphological identification approach, the trait classification approach and the molecular metabarcoding approach. Trait classification can be done directly under light microscope (1) or from the identified morphotaxa (2). Green arrows indicate current method of diatom environmental assessment, dotted blue arrows show potential novel approaches. 'DNA WetLab' refers to DNA extraction, PCR amplification and purification. Figure inspired by Pawlowski *et al.* (2018).

3. Aim and objectives

The aim of this thesis work was to investigate whether two novel promising approaches (metabarcoding and trait-based methods) for assessing the diversity of freshwater benthic diatoms can be used to improve freshwater conservation and management. Results obtained using the novel methods were compared against data from conventional diatom identification, in order to achieve a better understanding of the underlying methods themselves and better knowledge of diatom diversity and function. Such knowledge is necessary for further method development for assessment of diatoms and their use as indicators of eutrophication in streams and lakes.

Specific objectives of the work were to:

- Determine whether the molecular and trait-based novel approaches for assessing diatom diversity can be used in environmental assessment (**Papers I, III, IV**).
- Determine the implications of differences between the methods intended for routine use in research and environmental assessment (**Papers I, II, IV**).
- Explore the potential of trait-based assessment to link the diatom community response to changes in environmental conditions (**Papers III, IV**).
- Develop novel tools for environmental assessment based on the novel approaches (**Paper IV**).

4. Material & Methods

This chapter presents an overview of the key methods used in the work. For full details about the methods referred to here, see **Papers I-IV**.

4.1 Sample datasets and study design

In **Papers I, II** and **IV,** the diversity of freshwater benthic diatoms in natural samples collected from lakes and streams in Fennoscandia (Sweden, Finland, Norway) and Iceland was analysed. The samples used were existing samples collected within different regional and national monitoring projects between 2006 and 2017. For a subset of sites, temporal (several years) or spatial (several sites within one lake, or one stream reach) replicates were available. In **Paper I,** all sample replicates were included (181 samples), while in **Paper IV** replicate samples were removed (115 samples remaining). The sampling sites had been selected by the relevant monitoring projects to cover broad environmental gradients and different ecosystems, such as the alpine, boreal and nemoral ecoregions and gradients of forest, agriculture, and urban cover in river catchments (**Paper I**). Data on water chemistry parameters for the samples (alkalinity, conductivity, pH, total organic carbon (TOC), total nitrogen (TotN) and total phosphorus (TotP)) were obtained from the Swedish National database (https://miljodata.slu.se/MVM/), the Finnish Environmental Administration (Hertta system version 5.7), the Norwegian Institute for Water research (NIVA) database and, for Icelandic sites, from previous publications (Friberg *et al.*, 2009, Ólafsson *et al.*, 2010). A subset of 29 samples from this dataset was used in **Paper II**, comprising 14 samples from Sweden, 11 samples from Finland and four samples from Norway. The samples were selected to cover a broad range of environmental variables even with a reduced number of sites.

In **Paper III**, the potential of trait-based assessment to link the response of diatom assemblage community to changes in environmental conditions was explored using results from an experiment performed on a stream facility in Finland. A total of 192 biofilm samples (96 for microscopy identification and 96 for DNA extraction) were collected from tiles submerged in six experimental stream channels at the Kainuu Fisheries Research station (DESTRESS project, Paltamo, Finland) in September 2017.

4.1.1 Sampling

The natural samples assessed in **Papers I, II** and **IV** were collected in early autumn from submerged hard substrate in the lake or riverbed, following the European standard for diatom sampling (EN 13946:2014) (CEN, 2014). The fresh samples were preserved with 97% ethanol (final concentration approximately 70%) to protect the DNA from degradation in long-term storage (Stein *et al.*, 2013) and kept in darkness at room temperature. In **Paper III**, the samples for morphological identification were collected by brushing the tile with a toothbrush, rinsing with water from the stream and preservation of the sample in Lugol. The samples for DNA extraction were collected by scraping the tile with a scalpel blade and preserved in 95% ethanol (**Paper III**).

4.1.2 Study design

*European bioinformatics pipeline comparison (**Paper II**)*

In testing the performance of the novel molecular approach for diatom identification using the metabarcoding method in **Paper II,** the different algorithms (or 'bioinformatics pipelines') available to process raw DNA sequencing data were compared and a list of diatom taxon names was generated. The study formed part of a European collaboration programme (COST DNAqua-Net, Leese *et al.*, 2016) which made it possible to compare the bioinformatics tools used for diatom metabarcoding in different laboratories within Europe. Because all the pipelines included in the comparison were developed within the scope of diatom biomonitoring and environmental assessment, it was also possible to examine the potential of the molecular approach for possible implementation in environmental assessment legislation. The comparison, of a total of six pipelines, was designed to minimise any potential source of discrepancy external to the

bioinformatic process itself (see Material and Methods section in **Paper II**). Each pipeline was run as currently done by the laboratories, with as little adaptation as possible. This strict comparison allowed a focus on the bioinformatics processes only, bypassing discrepancies that can arise during sampling, DNA extraction, amplification, or sequencing, but also on the potential of the metabarcoding method with an array of different approaches (different DNA units, different steps, different algorithms). It was also decided to keep the reference database variable constant in the study, to eliminate its direct impact, which is often cited the main source of discrepancies between different pipelines (Kermarrec *et al.*, 2014, Kermarrec *et al.*, 2013, Rimet *et al.*, 2018, Zimmermann *et al.*, 2015, Kelly *et al.*, 2018a, Rivera *et al.*, 2018).

*DESTRESS (Disentangling the impacts of multiple Stressors on stream ecosystems) project (**Paper III**)*

The experiment described in **Paper III** was conducted in six artificial stream channels in the Kainuu Fisheries Research Station (Paltamo, northern Finland), as part of the DESTRESS project, a body of "cross-disciplinary work which tackles the multiple stressor problem by integrating novel experiments and intensive analysis of existing data within an explicit socio-economic context" (Fernkvist, 2020). The experiment took place in August-September 2017 and ran for 53 days in total (**Paper III**). Each of the channels is supplied with water draining from a nearby lake (Kivesjärvi), which is the main source of colonising organisms. For the duration of the experiment, each channel was divided into four sub-channels, resulting in 24 channel replicates to which different treatment combinations were applied (Figure 2). Three environmental variables were manipulated, namely nutrient input, flow variation and shading, which are related to three widespread pressures affecting boreal forest streams. Three levels of flow were manipulated in the channels (**Paper III**): (i) 'press' flow, where drought was introduced and maintained over a long period, (ii) 'pulse' flow, where water levels fluctuated between drying out and normal flow every other day, and (iii) 'control' flow (in respect to mean flow velocity typically observed in boreal streams). The flow disturbances were applied to 16 of the subchannels for nine days, followed by 20 days of normal 'control' flow for ecological recovery. Two levels of nutrient availability ('ambient' and 'enriched') were applied to the sub-channel level, nested within the six channels (Figure 2). Twelve of the

sub-channels were nutrient-enriched by delivering liquid nutrient solution into the system (phosphate (PO4) and nitrate (NO3)). The solution was continuously pumped out upstream of each sub-channel. The nutrient enrichment treatment was applied from the start of the experiment and lasted throughout the entire experiment (**Paper III**). Two levels of light availability ('open system' and 'shading') were applied to the sub-channel level, nested within the six channels (Figure 2). Twelve of the sub-channels were subjected to artificial shading, resembling shading levels of forested streams observed in Fennoscandia, by adding gardening 'shade fabric cloth' over the sub-channel to reduce light availability by approximately 70% (**Paper III**). The shading treatment was applied from the start of the experiment and lasted throughout the entire experiment (**Paper III**). The experiment consisted of a

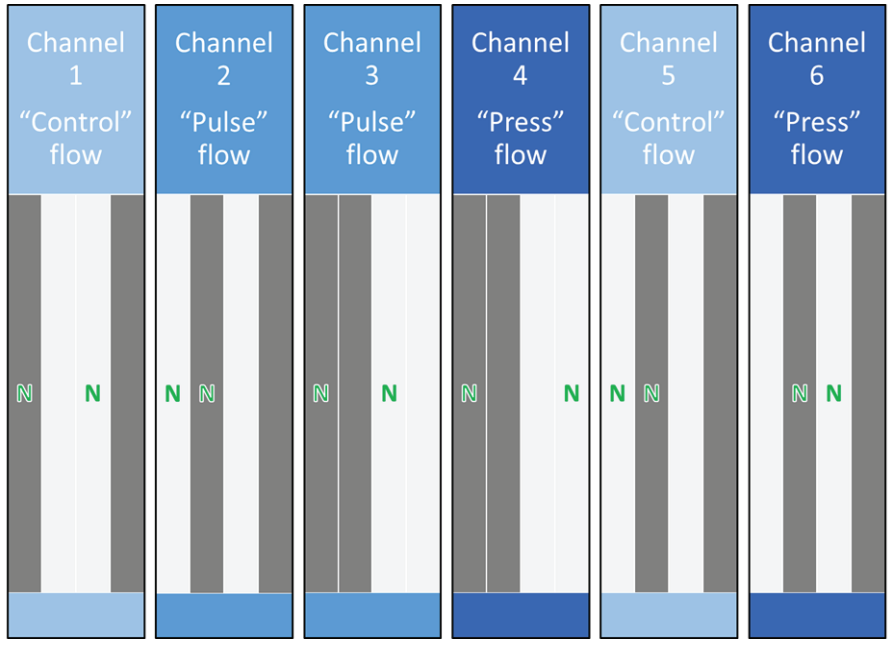


Figure 2. Channel set-up and treatment combinations used in the DESTRESS experiment (**Paper III**). Each channel was divided into four sub-channels. Flow treatment was applied at the channel level (indicated in blue), nutrient enrichment treatment was applied at the sub-channel level (indicated by a green "N") and shading treatment was applied at the sub-channel level (indicated in dark grey).

3x2x2 split-plot design with two replicates of each combination of the three treatments.

4.2 Diatom identification

Diatom identification by light microscopy and by metabarcoding was applied in all cases (**Papers I-IV**), in order to compare these two approaches.

4.2.1 Light microscopy

In **Papers I-IV**, the fresh samples were processed following the European standard (EN 14407:2014) (see the individual papers for details). Identification of diatom taxa to the lowest taxonomic level possible (mainly species) was performed in two different laboratories for the 180-samples dataset, at the Swedish University of Agricultural Sciences (SLU, Sweden), and the Finnish Environment Institute (SYKE, Finland). The results were harmonised before data analysis (**Papers I, II** and **IV**), to avoid significant differences due to different analysts or laboratories (Kahlert *et al.*, 2009). For the DESTRESS project samples (**Paper III**), morphological identification was performed at the University of Helsinki, Finland.

4.2.2 Metabarcoding

The method used for the metabarcoding approach was refined during the thesis work. Laboratory protocols for DNA extraction and amplification stayed constant overall, with only minor modifications, but sequencing technologies, bioinformatics pipelines and data analysis were subjected to several improvements between the different studies. This section provides a summary of these methods. Please refer to **Papers I-IV** for full details of the protocols used.

DNA extraction

DNA extraction and PCR amplification were performed at the laboratory of INRA CARRTEL in Thonon (France) for **Papers I** and **IV**, at the University of Geneva (Switzerland) for **Paper II** and at SLU for **Paper III**. There is a high level of organic carbon in Fennoscandia waters, which can inhibit the PCR reaction, so the NucleoSpin Soil Kit (Macherey-Nagel) was used for DNA extraction of all samples, following the manufacturer's recommendation (with one modification of the protocol according to

Vasselon *et al.* (2017), see Material and Methods section in **Paper I**). The starting sample for the DNA extraction process was a pellet collected after centrifuging a volume of 2 to 8 mL of biofilm sample at 13,000 rpm for 30 min. The DNA concentration in the DNA extracts was assessed in all cases, to calculate the dilution factor needed for the PCR.

PCR amplification with two DNA markers

Two different DNA barcodes were used in the analyses: the 18S-V4 region of the nuclear-encoded 18S gene and a 312 bp fragment of the plastid gene *rbc*L. These two DNA markers are well-known hyper-variable regions, which makes them interesting for barcoding. They are currently the most frequently used markers for molecular analysis of freshwater diatoms, because of their power to discriminate diatom communities, covering the three major diatom divisions, and for their balance between variability and conservation of the primer binding sites (Kermarrec *et al.*, 2014, Zimmermann *et al.*, 2011; **paper I**). For the 18S-V4 marker, the DIV4for and DIV4rev3 primer pair optimised for diatom metabarcoding by Visco *et al*. (2015) were used. For the *rbc*L marker, three pairs of degenerated primers modified after the Diat_rbcL_708F and R3 primer pair by Vasselon *et al*. (2017) were used.

In **Papers I, II** and **IV**, PCR amplification was performed in triplicate and later pooled to achieve sufficient DNA concentrations. For the samples used in these papers, the volume of DNA extract used for PCR amplification and the PCR conditions applied were adapted according to the DNA marker used: PCR amplification was performed according to the protocol in Vasselon *et al.* (2017) for the *rbc*L marker and according to Visco *et al.* (2015) for the 18S-V4 marker (**Paper I**). The DNA extractions and PCR amplification were re-done for **Paper II**, again following Vasselon *et al*. (2017), with two modifications (volume of DNA extract and number of PCR cycles) for the *rbc*L marker and following Visco *et al*. (2015) for the 18S-V4 marker. In **Paper III**, the protocol from Vasselon *et al*. (2017) was adapted and optimised for use with a cheaper and easier to use Taq polymerase (Phusion High fidelity PCR Mater Mix) (see Material and Methods section in **Paper III**).

Sequencing technologies

In **Papers I** and **IV**, the DNA libraries were sequenced using an Ion 318 Chip Kit V2 (Life Technologies) on an Ion Torrent Personal Genome Machine (PGM) at the Platform Genome Transcriptome (PGTB, Bordeaux, France). The Ion Torrent technology uses a semiconductor chip that captures chemical information during DNA sequencing (base incorporation induces a hydrogen ion release that changes the pH of the solution) and translates it into base information (ACGT). This technology, released in 2010, allows for faster and more scalable sequencing than the conventional light-based sequencers and was the best method available at the time of the study in **Paper I**. In **Paper II**, the DNA libraries were sequenced at the University of Geneva, on an Illumina MiSeq instrument using paired-end sequencing for 500 cycles with Standard kit v2. In Illumina sequencing, when a nucleotide base attaches to the DNA strand, a fluorescent dye is liberated and excited by diode lasers (at 530 & 660 nm) and recorded using digital cameras. Following their emergence, Illumina platforms quickly became the dominant technology in the sequencing industry (Quail *et al.*, 2012) and the MiSeq instrument, first presented in 2011, was the device used by most of the laboratories in the collaborative study described in **Paper II**. The technology is easier to use (kit available with the flow cell and reagents) and allows for sequencing of up to 96 samples in a single run and for longer read length (up to 2X 300 bp) with better accuracy (higher false positive rate in PGM technology). In **Paper III**, the DNA libraries were sequenced by the SNP&SEQ Platform (Uppsala Biomedicinskt centrum, Sweden), also with the Illumina MiSeq instrument using v2 sequencing chemistry kit.

Bioinformatics analyses

The bioinformatics analyses in **Paper I** were performed at the Mésocentre de Calcul Intensif Aquitain (MCIA) at the University of Bordeaux, France. Two Python programs were used in the bioinformatics pipeline in **Paper I**. These were: MPI-disseq, which computes a pair-wise distance matrix between the sample sequences and the reference barcodes, and Diagno-syst, which performs the taxonomic assignment (Frigerio *et al.*, 2016). The R-syst::diatoms (version 5) reference database (later renamed diat.barcode) (Rimet *et al.*, 2016, Rimet *et al.*, 2019) was used for taxonomic assignment because it includes sequences from the NCBI database and also unique

sequences from private INRA cultures. The pipeline was developed at UMR BioGeCo (INRA, France) and was an appropriate choice for the work in **Paper I** as it is optimised to work with sequence outputs from the PGBT platform. The pipeline is also designed to avoid as many heuristics as possible, *i.e*. to limit the number of user-chosen parameters that can impact the data output, which was an advantage for a first application of the metabarcoding method in Fennoscandia. The only user-chosen parameter in the pipeline is the range of the "sliding barcoding gap" which allows the user to obtain all possible taxonomic assignments at several levels of bp dissimilarity, for example between 0 bp and 10 bp difference (**Paper I**). Prior to matrix computation, a filtering step is also performed manually, to exclude too short or too long reads (**Paper I**). However, the absence of computation shortcuts means that the programs need high computation power and resources to run, which is not the best option for large-scale monitoring implementation.

Based on the above, in **Paper II** a decision was made to investigate all the different bioinformatics pipelines currently available for diatom metabarcoding. The Diagno-syst pipeline used in **Paper I** was included in a comparison with four other bioinformatics pipelines (Mothur, MetBan, Quiime, SLIM), which enabled assessment of the cost and benefit of low heuristics at the expense of computing power. In **Paper II**, all pipelines currently used by laboratories (from six different countries) involved in development of diatom metabarcoding for ecological assessment purposes (a collaboration that was possible thanks to the COST DNAqua net programme) were included in the comparison. This resulted in 13 different softwares/bioinformatics tools being implemented in the five pipelines tested (see Material and Methods section in **Paper II** for details). The bioinformatics analyses were performed by each laboratory in their respective institute (**Paper II**). The comparison was designed to allow assessment of strengths and weaknesses of different bioinformatics approaches (sequence filtering, DNA units, clustering and assignment algorithms and thresholds *etc.*) and to test the impact and importance of specific steps in the bioinformatics process. Since all pipelines were operated with the same set of raw sequences, the most important parameter to keep constant was the reference database used for the taxonomic assignment. The diat.barcode database was selected for use once again, since it had recently benefited from great curation and completion efforts (for the *rbc*L barcode).

One of the pipelines included in the comparison (MetBan) had a design that was not compatible with the diat.barcode reference database, but a decision was made to keep the original design rather than modify the pipeline (**Paper II**), as the main aim of the study was to compare the pipelines used by the different laboratories. For the same reason, one pipeline that was not designed for taxonomic assignment (SLIM) was used to perform a basic taxonomic assignment for comparison, but without any optimisation.

Different DNA units were used in the presented studies, as a consequence of the fast development within metabarcoding. These were: operational taxonomic units (OTUs), which are groups of phylogenetically close entities (sequences sharing 97-99% genetic similarity in the present case); amplicon sequence variants (ASVs) or exact sequence variants (ESVs), which are groups of exactly similar sequences (100% similarity); and individual sequence units (ISUs), which are composed of ASVs and background noise (a more limited filtering of sequences might leave some erroneous sequences from PCR and sequencing errors). After assessing the use of OTUs and ISUs in **Paper II**, the use of different DNA units (ASVs) was explored in **Papers III** and **IV** using another bioinformatics pipeline, DADA2 (version 1.14, Callahan *et al.*, 2016), which is gaining popularity in metabarcoding studies. This pipeline, first presented in 2016, was well-suited for the purposes of the studies, as it is optimised for Illumina sequences data (still dominant in the sequencing technology market) and proved to be an easy-to-use tool when working with ASV DNA units. The bioinformatics analyses were performed by the SLU Bioinformatics Infrastructure (SLUBI). In short, primers were removed using cutadapt version 2.3 (Martin, 2011) and only full amplicon reads with both forward and reverse primers attached were retained. Singletons were removed and taxonomic identification was done using the *assignTaxonomy* function from the DADA2 R package (available at https://github.com/fkeck/DADA2_diatoms_pipeline) and diat.barcode (version 7, Rimet *et al.*, 2019).

4.2.3 Trait identification

Different terms for morphological features are used in the literature but, for the sake of simplicity, all diatom morphological features examined in this thesis are referred to as ‘traits. Ecological information on diatoms can be found in compiled open-access databases such as OMNIDIA (Lecointe *et al.*,

1993), the Diatoms of North America website (https://diatoms.org/) or the European Diatom Database (EDDI, Battarbee *et al.*, 2001). However, compared with other organisms, relatively little information about diatom traits is readily available. A study by Litchman and Klausmeier (2008) on the whole phytoplankton community provided the first categorised functional trait information for diatoms. A later study by Rimet and Bouchez (2012) provided a more comprehensive compilation of diatom traits, both functional and morphological. In **Papers III** and **IV**, morphological traits were assigned to the morphotaxa using the taxon-specific reference database developed by Rimet and Bouchez (2012, Appendix 1). At the time of the work described in **Papers III** and **IV**, that database included 1115 diatom taxa from Central Europe and, to our knowledge, was the most complete database available in open access. Among the 1332 diatom species used for routine environmental assessment of freshwaters in Sweden (Swedish Diatom List v.3.0; (Miljödata-MVM, 2021)), 637 are included in the database. However, trait information is missing for more than half of the species in genera such as *Pinnularia*, *Eunotia*, *Encyonemma* and *Stauroneis* and none of the species in the genus *Halamphora* is represented. On the other hand, genera such as *Nitzschia*, *Cymbella* and *Surirella* are well represented in the Rimet-Bouchez database. From the available traits in the database, four ecological guilds (high-profile, low-profile, motile, and planktonic) were used in **Paper III**. These traits were considered sufficiently descriptive for diatom morphology, because the guilds were defined based on a combination of several morphological features, particularly size range, and life habit (tube-forming, chain-forming, prostrate, motile), and ecological preferences of taxa regarding nutrient and flow disturbances. There were 175 diatom taxa in the dataset used in **Paper III**, of which 83 could be classified with the database. The remaining 92 were not included in the database and were labelled 'unknown guild'. In **Paper IV**, the ecological guild trait and biovolume (five categories) were used. An additional trait, surface to volume ratio (five categories), was calculated for each taxon, using the size values from the database and a simple equation (assuming a box shape for pennate diatoms and a cylinder shape for centric diatoms). This resulted in 14 trait categories and 100 potential unique combinations of traits. The taxa in the study covered 66 of these combinations and were referred to as 'fully combined traits' (**Paper IV**). As found in **Paper III**, not all the taxa in the dataset were included in the reference database, but in **Paper IV** biovolume,

size and ecological guild information for these, taken from literature, were added.

4.3 Ecological assessment

To test whether the metabarcoding method could be suitable for use in environmental assessment, in **Papers I** and **II** a comparison was made of diatom taxa list established using light microscopy and from metabarcoding data, and also of the index scores and ecological status classes calculated from those lists. In both papers, the IPS index (Cemagref, 1982) was used for the environmental assessment, since it is one of the most widely used diatom indices in Europe, it is intercalibrated in the EU WFD, it is the standard index for Sweden and it is also used in Finland, Norway and Iceland. The index calculations were performed using the official Swedish standard method for status classification, with the indicator values derived from the Swedish Freshwater Diatom List (Havs- och vattenmyndigheten, 2018). The classes used for ecological status of the water bodies ('Very Bad', 'Bad', 'Moderate', 'Good' and 'Very Good') were derived from the IPS scores following the intercalibrated class boundaries defined for Sweden in the WFD (Kelly *et al.*, 2014, Havs- och vattenmyndigheten, 2018).

4.4 Data analysis

To compare microscopy data and HTS data, the diatom valve and DNA read counts were transformed into relative abundances in all cases. For the statistical analysis, various multivariate methods commonly used in ecology were employed in **Papers I-IV.** The distribution of the data was tested in all cases and direct methods (RDA, CCA) were used to test the hypotheses and indirect methods (PCA, DCA, NMDS) to search for patterns, using unimodal-based method or linear-based methods when appropriate (see Material and Methods section in **Papers I-IV** for details). The analyses were usually performed using the R software (R Core Team, 2014), and occasionally using the PAST software (Hammer *et al.*, 2001) for the SIMPER analyses and the PC-ORD software (McCune and Mefford, 2011) for indicator species analyses. The JMP software (version 15.0.0, SAS Inc., Cary, NC, 1989-2019) was also used in **Paper III**, to create a linear mixed effect model. Different levels of data harmonisation and cleaning were used

when needed throughout the work (see Material and Methods section in **Papers I-IV** for details).

4.5 Assessment of error sources (**Paper I**)

In **Paper I**, a script using the R software that would return a specific code for each comparison of taxa in the two datasets (morphotaxa and taxonomic identification of DNA reads) was also designed. The aim was to estimate how often, on average, the molecular method would provide a taxonomic assignment that matched that obtained with light microscopy, both in terms of presence/absence and in terms of relative abundances of species, and to identify the source of any discrepancy between the methods (*e.g.* lack of reference barcode, non-detection by one or other method *etc.*). The codes provided valuable information on the origin of any identification discrepancies and for relative abundance comparisons between methods.

Table 1: Codes used in **Paper I** for the different types of deviation between the taxa lists generated by the microscopy approach and the molecular approach

MOL	Species found by molecular method only	
	H	Species not in the Fennoscandia taxa list
MOR	Species found with morphological method only	
	ND	Species represented in the DNA database, but no identification
	NR	Species not in the DNA reference database
AB1	Species found with both techniques, but with higher abundance in the microscopy approach	
AB2	Species found with both techniques, but with higher abundance in the molecular approach	
GM	Species found by both techniques with the same relative abundance	
G	Taxonomic identification stopped at the genus level	

4.6 Index design (**Paper IV**)

In designing a new total phosphorus (TotP) index in **Paper IV**, an existing method (Tapolczai *et al.*, 2019, Tapolczai *et al.*, 2021) was followed. The ecological values (optima, tolerances) of morphotaxa, ASVs and traits were calculated using a training dataset (75% of the samples). Using these ecological values, the index was calculated for the remaining 25% (test

dataset) of the samples. In order to overcome the limitation that the index values would only be calculated for 25% of the samples, the procedure described above was repeated 100 times, so that each sample was selected into the test dataset multiple times (25% chance at each iteration). This method also avoided inference of extreme values by chance. Optima value and tolerance value were defined for each morphotaxon, ASV and trait, using relative abundance data and presence/absence data (see Figure 5 in **Paper IV**). Because fewer samples and a narrower environmental gradient were available for the lake sites in the dataset, the stream sites were separated from the lake sites in both the training and test datasets. A total of 12 new indices were tested, with testing based on relative abundances values and on presence/absence values, separately for lake and stream sites, for each of the three biological units (2x2x3 = 12).

5. Results and Discussion

5.1 Can novel molecular and trait-based approaches for assessing diatom diversity be used in environmental assessment?

5.1.1 The molecular approach

In **Paper I**, an assessment was made of whether the molecular approach could be used for ecological assessment within the scope of freshwater biomonitoring, with the tools currently used within the EU WFD (index and ecological status classes). The results obtained were compared with the outputs of the conventional microscopy approach. Prior to the work in **Paper I**, the molecular approach had been tested in central Europe (Kermarrec *et al.*, 2014, Vasselon *et al.*, 2017, Visco *et al.*, 2015, Zimmermann *et al.*, 2015) with encouraging results, but only on small datasets (<50 samples), and had not been tested in Northern Europe. Moreover, while (Kermarrec *et al.*, 2013) had compared several DNA markers for eukaryote sequencing (SSU rDNA, *rbc*L and cox1), the two most widely used DNA markers for diatom metabarcoding (18S-V4 and *rbc*L) had not been compared within the scope of environmental assessment.

In **Paper I,** the diatom communities identified with the metabarcoding method and with the light microscopy method were used to calculate the IPS scores for the sampling sites, which were then compared. The IPS scores from the molecular approach (IPS scores: 8-20) covered a wider range than the microscopy values (IPS scores: 12-20), but both reflected the same trends, with lower scores for nutrient-enriched sites (**Paper I**). The IPS scores from the two approaches were correlated, based on R^2 coefficient (see

Figure 2 in **Paper I**), but the correlation was weak and the differences between the datasets were significant (Figure 3). Corresponding differences in index scores were observed by Vasselon *et al.* (2017) and Rivera *et al.* (2018) in comparisons of molecular and microscopy approaches.

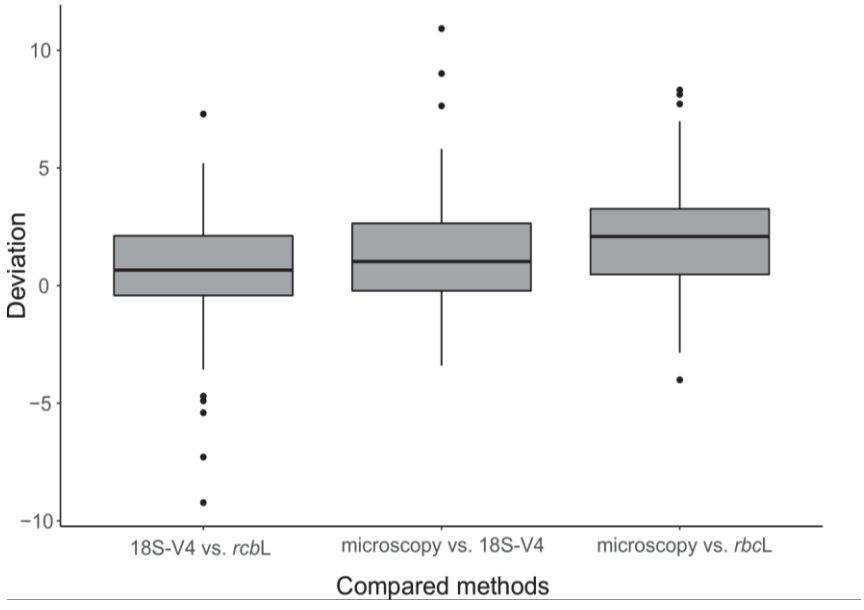


Figure 3. Deviation of the Index of Pollution Sensitivity (IPS) scores generated by the molecular approach from IPS scores generated by the microscopy approach. Deviation of IPS scores between DNA markers is also shown.

On the other hand, there were no significant differences in IPS scores between the results for the two DNA markers tested (18S-V4 and *rbc*L) (Figure 3). There were also no significant differences between countries or type of water body (river or lake), but statistical comparison was biased by an uneven number of samples from each country or from a certain water body type. One notable difference was that the IPS scores were closer to the values obtained with the microscopy approach when using the 18S-V4 marker for Norwegian samples, but when using the *rbc*L for Swedish and Finnish samples, possibly indicating differential performance of these DNA markers. Overall, the results indicated that the IPS index is robust enough to be used

with molecular data if accurate amplification and taxonomic identification of taxa is achieved.

In a second step, the samples were classified into the corresponding ecological status classes and the molecular and microscopy outputs were compared. Overall, the ecological status classes were significantly different between the molecular and the microscopy approaches (Figure 4). Similar ecological status was achieved for less than 50% of the samples (48% and 37% with the 18S-V4 and *rbc*L markers, respectively) (**Paper I**). The most concerning finding was that for 40-56% of the samples, use of the molecular method resulted in underestimation of the ecological status of the water body, mostly from 'Very Good' to 'Good', but also to lower status classes in some cases (Figure 4). The threshold between 'Good' and 'Moderate' status is especially critical, since the WFD prescribes remediation action for water bodies with status below 'Good' and underestimating the ecological status could therefore lead to unnecessary actions and costs. Overestimation of ecological status only occurred for 6-11% of the samples. The ecological status classes followed similar trends for river and lake sites, with just a few underestimations into 'Bad' and 'Very Bad' water quality classes for the river samples. This is probably due to the larger number of river samples, which covered a broader environmental gradient than the lake samples. In contrast to the IPS scores, the ecological status classes were also significantly different between the two DNA markers tested. The correlations found were lower than those reported by Visco *et al.* (2015) and Vasselon *et al.* (2017), but this can partly be explained by the much larger dataset used in **Paper I** and the different environmental gradient covered than at their sites in central Europe (different ecoregions). Ecological status classes are ultimately the most important information for stakeholders and decision makers, and these results were crucial in assessing the potential of the molecular method for environmental assessment purposes in Fennoscandia. Overall, in **Paper I**, the correlation between the results confirmed that the molecular approach, specifically the metabarcoding method, has potential for use in ecological assessment, but its outputs were significantly different from the microscopy values. The alternative method of using HTS data and metabarcoding for identification of benthic diatoms cannot be used at present to replace the conventional method of identification by light microscopy (currently accepted within the WFD) for ecological assessment of waters in Fennoscandia. Since HTS data may be too different from microscopy data to

be calibrated for the same tools (ecological index and ecological status classes), an assessment was made of whether both types of data give a similar response to environmental changes.

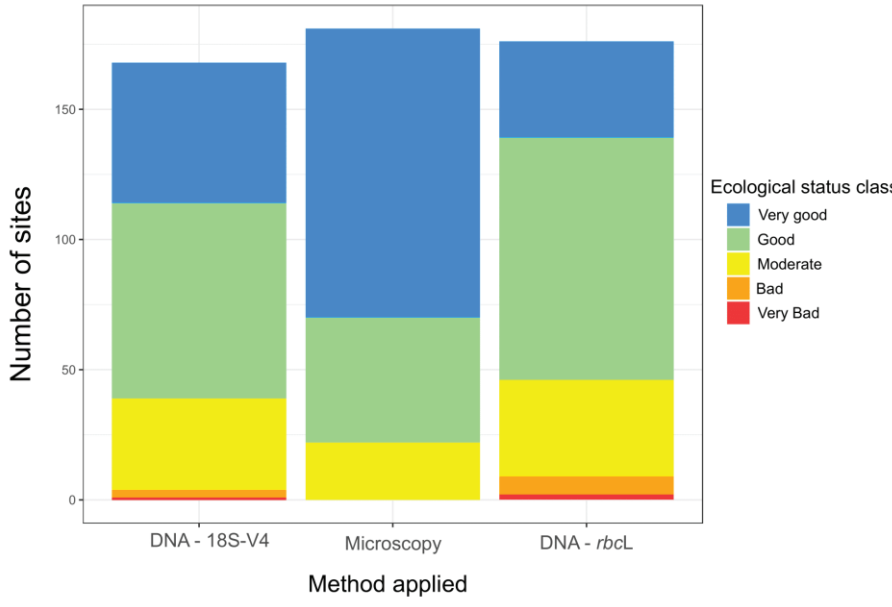


Figure 4. Proportion of each ecological status class assigned to the 180 Scandinavian sites when using the molecular approach (with the 18S-V4 and the rbcL DNA markers) and when using the microscopy approach. A few sites could not be classified for the molecular approach, because the few taxa detected did not have IPS values to calculate the site's index score.

This was done in **Paper III**, which examined the response of the diatom community to nutrient enrichment (a well-studied stressor in environmental assessment of freshwaters), but also the responses to relatively understudied factors (light/shading and flow), which could affect the response to nutrients. The results showed that the community variation reflected stressors applied to the ecosystem similarly between diatom communities quantified with the molecular approach (relative abundances of ASVs) and diatom communities quantified with the microscopy approach (relative abundances of morphotaxa). The response was similar between the microscopy and molecular approaches in the non-metric multidimensional scaling (NMDS) analysis (stress = 0.132 and 0.135, respectively) (**Paper III**). However, the total variation in assemblage structure of the morphotaxa was lower than for

the ASVs, and more of it could be explained by the environmental variables (see redundancy analysis (RDA) in **Paper III**), probably because the ASV dataset was composed of many more variables and was of greater complexity than the morphotaxa dataset. In terms of stressor-specific response, there were also differences between the two approaches. For the morphotaxa RDA analysis (microscopy approach), the first two axes together contributed 90% of the explanatory power. The first axis was negatively related to light exposure and positively related to the flow 'press' stressor, while the second axis was positively related to light exposure (**Paper III**). For the ASV data (molecular approach), the first two axes of the RDA analysis together accounted for 83% of the explanatory power, with the first axis also positively related to the flow stressors and the second axis also strongly negatively related to light exposure (**Paper III**). The axes were weakly correlated with the nutrient stressors in both approaches (short arrows for both datasets but in opposite directions; **Paper III**). Finally, the community composition characterised by ASVs (molecular approach) was significantly affected by all three experimental stressors, but always to a lesser extent than the morphotaxa community (microscopy approach). In both approaches, light was the strongest factor explaining community differences, followed by flow and then nutrients (see PERMANOVA analyses in **Paper III**). Thus, both datasets showed a similar clear response to the treatments applied, but with the morphotaxa structure variations somewhat better correlated to the environmental changes than the ASV structure variations. The ASVs data also nicely reflected environmental changes in **Paper IV**, but in that study the TotP gradient was a stronger predictor of community structure for the ASVs than for the morphotaxa (for correlation coefficients, see Figure 3 in **Paper IV**). Therefore, even if nutrients were not the main driving factor of community changes in the experiment described in **Paper III**, the dataset for Fennoscandia natural samples showed that the observed community changes on this large scale were well related to phosphorus levels, which is encouraging with respect to monitoring of eutrophication in freshwater bodies.

The IPS index score is calculated using taxa-specific values (Cemagref, 1982), as is the case for most frequently used diatom indices such as TDI (Kelly and Whitton, 1995), PIT and AIP (Schneider and Lindstrøm, 2009, Schneider and Lindstrøm, 2011). Therefore, another important aspect when considering the use of the different approaches in ecological assessment was

whether both approaches detected the same diatom taxa. In **Paper I**, the causes of differences in taxa detection between the molecular and microscopy approaches were investigated and it was found that the Shannon scores were significantly different. Interestingly, it was found that a significant number of taxa were only detected by the molecular approach (56 when using the 18S-V4 marker and 122 when using the *rbc*L marker), which represented 30-40% of the diatom community present (**Paper I**). Moreover, 6-8% of these identifications were taxa that are not included in the Swedish list of freshwater diatoms, *i.e.* had not been found previously in Sweden, and thus had no IPS values by which to include them in the index calculations (**Paper I**). This indicates that the molecular method can help detect and identify taxa that are currently not considered in environmental assessments in Sweden. In some cases, small or delicate species, were also only detected by the molecular method (**Paper I**). The higher taxonomic detection rate when using the molecular method can be explained by the methodology not relying on diatom silica cells (which can be too small or too fragile to identify after the treatment to make permanent slides) and by the number of sequences identified for each sample usually exceeding the 400 cells generally counted in light microscopy, which can enable better detection of rare species (Zimmermann *et al.*, 2015). Considering the known limitations of taxonomic assignment arising from choice of reference database and algorithm, relative abundance of DNA units (ASVs) without taxonomic assignment were used for the analyses in **Paper III**, which resulted in less data loss and better reproducibility of results. However, to get a better ecological understanding from the results, taxonomic assignment of the dominant ASVs in the dataset was also carried out, to compare their response to stressors with those of the dominant morphotaxa (**Paper III**). As in **Paper I**, there were significant differences in taxa detection between the morphological and metabarcoding approaches (**Paper III**). For example, several of the dominant diatom ASVs could not be identified with the reference database and different trends emerged in response to stressors, *e.g.* several distinct ASVs were identified as *Achnanthidium minutissimum,* but they showed opposing responses to light exposure (see Table 3 in **Paper III**). A similar trend was observed for the two ASVs identified as *Gomphonema saprophilum* (**Paper III**). Similar results have been found in other studies, with the molecular approach typically identifying more taxa than the light microscopy approach or with discrepancies in relative abundances (Rivera *et*

al., 2020, Rivera *et al.*, 2018, Vasselon *et al.*, 2017) indicating potential issues regarding the true identity of some reference barcodes or differences in genotype between reference barcodes and field populations (Kelly *et al.*, 2020). This confirms the better taxonomic coverage by the molecular method, but also highlights the possibility of detection of cryptic taxa with different ecological profiles. Some taxa usually merged into 'species complexes' because of difficulty in morphological identification (Kahlert *et al.*, 2019) were also clearly distinguished by the molecular method (*e.g.* *Fragilaria gracilis*) (**Paper III**).

Regarding the primary objective of the work in this thesis, the first application of the molecular approach (**Paper I**) showed that it is suitable for environmental assessment in Northern Europe (step-by-step methodology possible with HTS data to achieve IPS and ecological status classes), but it also showed that metabarcoding outputs cannot be directly compared to outputs from the microscopy approach. **Paper III** confirmed that the two approaches can be applied to assess environmental changes, but that HTS data and microscopy data show weak responses to nutrient enrichment. The response of HTS data (ASVs) was more complex and driven not only by the well-studied nutrient gradient, and probably also by other factors which were not assessed in **Paper III**. It can be concluded that, while HTS data and microscopy data for diatom communities both reflect environmental changes, HTS data should not be used with the current tools developed and calibrated for microscopy data (indices, ecological status classes). However, there is potential for development of new tools tuned to HTS data for environmental assessment, especially for evaluation of eutrophication status of natural water bodies. On the other hand, the molecular method enabled the detection of taxa often missed by light microscopy assessment (cryptic and rare taxa) and can be used to achieve a better representation of the diatom community complexity and new insights into diatom structural responses and ecological preferences. However, if some subtleties in diatom taxa can be identified with a molecular approach, others are missed without morphological identification, and it must be borne in mind that each approach carries its own uncertainties (Kelly *et al.*, 2018b, Kelly *et al.*, 2020).

5.1.2 The trait-based approach

Another novel approach tested in this thesis was the use of traits to quantify the diatom community (**Papers III** and **IV**). Morphological traits directly reflect the adaptation strategies of diatoms to their environment and similar traits should thus be detected in similar habitats, even in different geographical areas or different ecoregions. A trait-based approach could be especially useful in areas where knowledge of the taxa diversity of diatoms is limited, for example in tropical areas such as Mayotte Island in the Indian Ocean (Tapolczai *et al.*, 2017) or in Mexico (Mora *et al.*, 2017).

In **Paper III**, on quantifying the diatom community using morphological traits (specifically ecological guilds of diatoms), it was found that all three stressors applied significantly affected the community composition characterised by the traits (PERMANOVA analyses, light exposure: F = 74.7; 'press' flow disturbance: F = 33; 'pulse' flow disturbance: F = 27; nutrient enrichment: F = 12; $p<0.05$ in all cases, **Paper III**). This shows that the variation in diatom traits well reflected the environmental stressors in Scandinavian ecosystems, as has been observed previously for other ecoregions (B-Béres *et al.*, 2017, Dong *et al.*, 2016, Lindholm *et al.*, 2018, Schneck *et al.*, 2017, Soininen *et al.*, 2016, Tang *et al.*, 2016, Berthon *et al.*, 2011). The RDA results supported these findings, with low inertia (0.06) and 69% of the variance in the dataset explained by the four constrained axes. The first RDA axis was negatively correlated to light exposure and positively related to the "press" flow stressor, while the second axis was negatively related to light exposure and positively related to the nutrient stressor (see Figure 5). Overall, the trait-based approach revealed a stronger response to specific environmental stressors than when using morphotaxa quantification or DNA unit quantification methods (**Paper III**), which indicates potential for using trait quantification (by ecological guild classification of benthic diatoms) for environmental assessment. The study described in **Paper III** was the first application of diatom ecological guild classification in the Fennoscandia ecoregion and the observed guild responses were similar to those found for the US and central Europe (Rimet and Bouchez, 2012, Passy, 2007). The combination of light conditions and nutrient enrichment in particular was marked by a significant increase in motile diatoms (**Paper III**). This guild was positively related to the second RDA axis (Figure 5) and both the nutrient and light exposure stressors had a significant impact on its relative abundance in the samples (PERMNOVA analysis, nutrients: F =

96.7; light: F = 44.3; $p<0.05$ in both cases, **Paper III**). On other hand, the high-profile guild was strongly positively related, and the low-profile guild negatively related, to the first RDA axis (Figure 5). The PERMANOVA analysis confirmed that light conditions were significant for the relative abundances of the high-profile and low-profile guilds (F = 141.9 and F = 103.4, respectively; $p<0.05$ in both cases, **Paper III**). However, the low-profile diatoms were dominant in all samples and were largely composed of *Achnanthidium minutissimum* cells, so the response of that specific trait provided limited information. This means that, while the motile guild showed strong potential to detect changes in the nutrient gradient, the other ecological guilds only mildly reflected the nutrient enrichment factor and were mostly affected by light exposure, a factor rarely considered in environmental assessment.

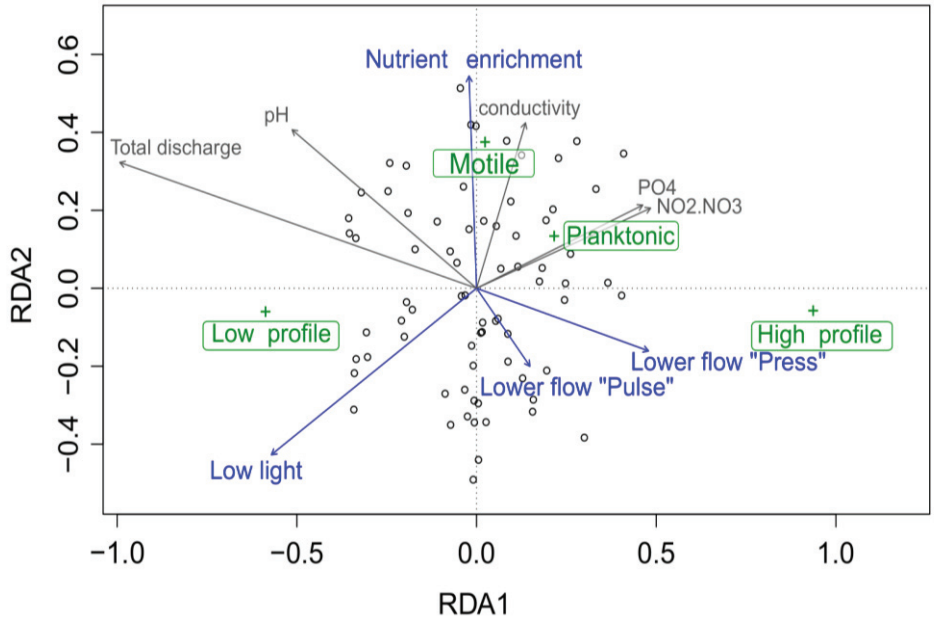


Figure 5. Results of redundancy analysis (RDA) on the trait (ecological guilds) dataset with stressors as explanatory variables and subsequent environmental fitting of physio-chemical parameters. Source: **Paper III**.

Similar observations of the strong impact of light exposure on the response of diatom ecological guilds have been made by Lange *et al.* (2011) and Stenger-Kovács *et al.* (2013). Therefore, even though the classification into

guilds is based on a combination of several morphological traits (Passy, 2007, Rimet and Bouchez, 2012), using only ecological guild trait to characterise the diatom community is not sufficient for environmental assessment, where the nutrient gradient is often the main focus. There are many other traits that could be measured, such as cell size or biovolume, to give a more precise description of the diatom community. In particular, surface to volume ratio has been associated with resource acquisition efficiency (Lange *et al.*, 2011, Snoeijs *et al.*, 2002) and could potentially reflect the diatom community response to changes in nutrient availability. Traits showing a strong response to the gradients applied will offer meaningful information, while traits that show a weak response and add little to the assessment should be removed to limit noise in the dataset. Some previous studies have tested the potential for using all available traits (B-Béres *et al.*, 2016), while others have highlighted specific traits that should be avoided, *e.g.* because of a weak response to the gradients tested or because of collinearity (Tapolczai *et al.*, 2017). Because no such investigation had been performed in Fennoscandia before, the work in this thesis, testing the response of a group of traits to diverse controlled stressors, was the first necessary step to best assess how to use trait data. The results obtained in **Paper III** provide evidence that the trait response is strongly shaped by light and flow, and potentially other factors, such as the acidity gradient. This, in turn, has an important impact on their response to nutrient gradients (Borchardt, 1996, Hill, 1996). Consequently, selection of relevant traits from sufficient data would allow for more precise investigation of environmental changes, for example with the development of new indices focusing on a specific environmental gradient. This was done in the study in **Paper IV**, where relevant traits to best reflect the nutrient gradient were carefully tested and selected and a preliminary TotP index was developed.

In **Paper IV**, the morphotaxa were categorised into 66 trait combinations (including ecological guilds, biovolume categories and surface to volume ratio categories, **Paper IV**). The traits did not significantly correlate to the environmental data when considered separately (Table S5 in **Paper IV**). However, when all traits were combined, the correlation to the environmental variables increased and was similar to that found for the

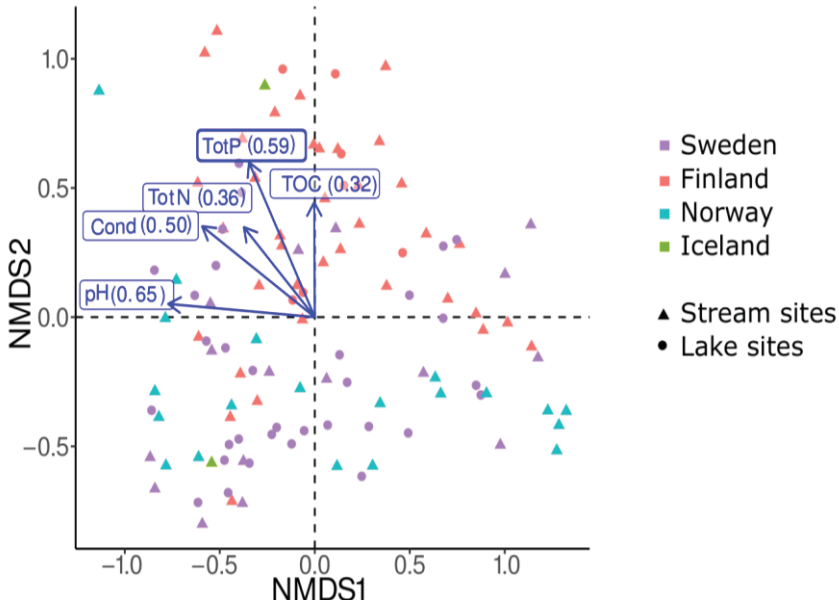


Figure 6. Results of non-metric multidimensional scaling (NMDS) analysis on the diatom trait dataset with environmental gradients as explanatory variables. Predictor values for each environmental variable are also shown. Source: **Paper IV**.

microscopy data (morphotaxa) and DNA data (ASVs) (Figure 6; Table S5 in **Paper IV**). There was a good correlation between the variation in the diatom community characterised by the 66 combinations of traits and the variation in the environmental gradients measured, especially the TotP gradient and the pH gradient (R^2= 0.59 and 0.65, respectively, Figure 6). The low-profile and high-profile diatom guilds did not seem to correlate significantly to any of the gradients measured, but the motile guild was strongly positively related to the first NMDS axis (Figure 11 in **Paper IV**), which correlated to most of the environmental variables and especially the nutrient gradient. In hindsight, after observing similar guild responses in **Paper III,** retaining only the motile guild and removing the other three guilds could have revealed a clearer response for the trait dataset in **Paper IV**. The two extreme categories of the surface to volume ratio (1 and 5) also strongly correlated to the first NMDS axis, in opposite directions, but no significant correlation was found for any of the other traits used. There were also no significant differences between the response of lake and stream communities when using the trait approach, in contrast to when using the microscopy and

molecular approaches (**Paper IV**). There was also some evidence that a few abundant key taxa dominated the response of specific trait groups, *e.g*. *Achnanthidium minutissimum* dominated the low-profile guild response, as in **Paper III**, but also the response of the smallest size class and largest surface to volume ratio traits.

In contrast to the study by Tapolczai *et al*. (2017), where the trait-based index covered a broader range of environmental variables compared with the morphotaxa-based index, the phosphorus gradient covered by the traits index in **Paper IV** was narrower than when using morphotaxa or DNA units (ASVs). Although the traits indicated a weaker relationship to the phosphorus gradient compared with morphotaxa index or molecular index, the results in **Paper IV** showed that it is possible to select meaningful traits and develop nutrient indices (in this case TotP) calibrated for diatom trait data. However, such traits need to be carefully selected, as some responses are stronger than others, and traits dominated by single species response should be handled carefully. Trait-based assessment provides a fine-tuned response to environmental changes, free of the functional redundancy present when characterising the diatom community using morphotaxa (Kelly, 2013). Moreover, although the traits-based approach does not circumvent the time-consuming process of analysis under light microscope, it does require significantly less expertise and could benefit from future implementation of automated image analysis techniques to improve the time efficiency (Burfeid-Castellanos *et al.*, 2020). Trait-based assessment of diatom communities, while not new, is still under development and is rarely used in environmental assessments. For example, no study has yet tried to acquire ecological status class from trait data. P**aper IV** confirmed that trait data can be used for environmental assessment, especially with the development of new calibrated indices. With further development, ecological status could be derived from these new indices and diatom trait data could potentially be used for environmental assessments required by national and international legislation. However, more research is still needed to further the current understanding of diatom traits and their response to environmental pressures.

To conclude, both novel approaches tested in this thesis have potential for use in environmental assessment of freshwaters in Fennoscandia and both involve methodology that requires less expertise than species-level morphological identification by light microscopy (Table 2). The molecular

approach also has the potential to avoid the long-existing issue of species concept (with the development of taxonomy-free approaches in *e.g.* **Papers III** and **IV**) and gives additional insights on diatom diversity and ecological preferences. The trait-based approach provides reliable information on direct and indirect impacts of environmental factors and meaningful traits can be carefully selected to best reflect the gradients under study. Both methods are still in need of further development and would benefit from further enrichment of reference databases (reference barcodes database and traits database) and the development of new indices calibrated for HTS and trait data.

Table 2. Potential of the novel metabarcoding and trait approaches for use in environmental assessment of freshwater bodies in Fennoscandia

Method potential	Metabarcoding	Traits
Easier application than morphological identification of species by microscopy	Yes	Yes
Reflects the response of diatom community to changes in the environment	Yes	Yes
Outputs comparable to those of microscopy	No	No
Outputs comparable to those of other novel method tested in this thesis	No	No
Can be used with current tools for environmental assessment	No	No
Can be used to develop new tools for environmental assessment	Yes	Yes
Provides information not provided by microscopy	Yes	Yes

5.2 Implications of differences between the methods for routine use in research and environmental assessment

5.2.1 Improvement of routine methodology

One of the clear advantages of the two novel approaches tested in this thesis is easier application, which is a very important aspect for potential implementation in routine monitoring. The trait approach still requires microscopic identification of diatom features, but with much less diversity, less variability and less expertise needed than in species-level identification. The molecular approach circumvents the time-consuming identification under microscope and benefits from a straightforward laboratory protocol that is easy to follow and replicate for the end-user, enabling better reproducibility of results. Besides, methods and technologies are evolving at a rapid pace and both approaches will surely benefit from future improvements as research continues. The molecular approach has also proven efficient for detection of additional species, providing more information on the complexity of the diatom community. **Papers I-IV** all revealed potential of the molecular method to identify species that are often missed in light microscopy-based assessments. Moreover, additional taxa were detected with all the different bioinformatics pipelines tested in this thesis (seven pipelines in total), which confirms the value of HTS data for better understanding the diatom community. Overall, dominant genera in the diatom community were detected rather similarly (presence/absence and abundance) by all seven pipelines tested and significant discrepancies were only observed in identification at species level (**Papers I** and **II**). In practice, only a few key taxa contribute most to the IPS score of individual sites in ecological assessment. Thus, detection and correct assessment of the abundance of these key taxa is sufficient to ensure reliable ecological assessment, thanks to the robustness of the IPS index. However, some of these key taxa are known to show differences in abundance quantification between the morphology and molecular approaches, apart from stochastic effects (Pawlowski *et al.*, 2018, Rivera *et al.*, 2020, Vasselon *et al.*, 2018) and are the major reason why it is difficult to use tools calibrated for the microscopy approach with molecular data. Work to fulfil the primary objective of this thesis revealed another interesting aspect of the molecular approach, *i.e.* that it can also detect cryptic species, as shown in **Papers III**

and **IV**. Several 'species complexes' have been defined to cope with cryptic species of diatoms when using taxonomy. While genetics properties do not necessarily capture the total variability of diatom communities, the molecular method helps to shed light on some of these complexes, *e.g*. the recent distinction of three new *Fragilaria* species previously known under the same species name (Kahlert *et al.*, 2019).

5.2.2 Overall low comparability of outputs at present

Variety of pipelines (molecular approach)

In **Paper I**, the bioinformatics pipeline selected for use (Diagno-syst) avoided the problem of user-chosen parameters, like clustering threshold for the molecular data, but it required much computer power to run on a large number of samples and might be difficult to implement in routine monitoring. The advantages and costs of using such bioinformatic pipelines needed to be assessed further, but the wide diversity of alternative pipelines available means that there is no go-to method with which to compare. In most cases, different laboratories seem to develop their own pipeline, suited to their own needs, and rarely compare their outputs with those of other laboratories.

The comparison of pipelines in **Paper II** revealed that every step of the bioinformatics process modifies the HTS data, so disparity between the outputs/results is visible as soon as the first filtering step (removing 'bad' sequences) is completed and keeps getting stronger as the analysis progresses. Moreover, parameters that are not necessarily obvious to the common user can have an important impact, *e.g*. handling of conflicts in taxonomic assignment or changes in pipeline function between software versions (**Paper II**). The diversity of bioinformatics pipelines included in the comparison in **Paper II** provided insights on different DNA units that can be used. For example, working with ISUs did not necessary lead to larger datasets than working with OTUs. Rather, it was the cleaning parameters that mattered the most (**Paper II**). The results also revealed greater importance of the choice of OTU algorithm compared with the choice of clustering threshold, as the differences in DNA unit assemblages were more marked between datasets of OTUs made with the same clustering threshold than between datasets created with different clustering thresholds (OTUs at 97% and OTUs at 99% threshold) or between datasets of different DNA units

(OTUs, ISUs and all sequences). In addition to pipeline design, the choice of reference database also had a major impact on the results (as in **Paper I**), but achieving taxonomic assignment is ultimately a constraint and carries misconceptions of the diatom communities from light microscopy identification. Thus the potential of the molecular approach may instead lie in alternative approaches, like taxonomy-free methods (Apothéloz-Perret-Gentil *et al.*, 2017, Apothéloz-Perret-Gentil *et al.*, 2020, Cordier *et al.*, 2018, Cordier *et al.*, 2019, Cordier *et al.*, 2021). In a nutshell, the results obtained in this thesis highlighted that the different pipelines used for benthic diatom identification in Europe have different designs, which impacts the outputs produced and hinders their comparability between laboratories and between countries, impeding development of the molecular approach as a new monitoring method under WFD regulations. There was also evidence that decisions made in pipeline design have implications for taxa assemblages, which in turn has implications for biotic indices calculations and, ultimately, for ecological assessment.

Variety of sequencing technologies (molecular approach)

DNA sequencing has been under extensive development for the past 20 years and several sequencing technologies are now available to end-users, including *e.g.* Illumina sequencing, Ion Torrent sequencing and PacBio sequencing (Quail *et al.*, 2012). Each technology developer regularly releases new sequencer models that offer improvements over the previous models, making DNA sequencing increasingly efficient and reliable. In the work described in this thesis, two of these technologies were used: Ion Torrent PGM (**Papers I** and **IV**) and Illumina Miseq (**Papers II** and **III**). However, any change in technology demands adaptation of the methodology, from sample preparation to bioinformatics processing, and output data from one technology are not always comparable to those from another in terms of read length, quality or format. Therefore for simplicity and comparability, end-users tend to use only one of the available technologies, considering its advantages and shortcomings for their research purposes. This wide diversity of technologies available, and the important difference between their outputs, limit the comparability of results across studies, laboratories and countries. The development of new tools requires massive training datasets and would benefit from combination of data from different countries to cover broad ecological gradients, which is not possible if such data are generated using different sequencing technologies. Besides, for implementation of a

molecular method in routine monitoring of *e.g.* freshwater eutrophication at national or international scale, it is also important to be aware that not all users may have the means to access the latest or most used technology.

Variety of DNA markers (molecular approach)

In **Papers I** and **II**, the molecular approach was tested using two different DNA markers (18S-V4 and *rbc*L). Both are commonly used for diatom metabarcoding but had never been directly compared prior to those studies. Overall, the two DNA markers used in **Paper I** showed similar trends and their outputs correlated to one another and to those obtained by microscopy, but there were still significant differences between the markers' datasets (in terms of taxonomic assignment, IPS scores and ecological status classes). On the other hand, differences in terms of relative abundances of taxa between the markers rarely occurred and were mostly found for the same taxa by the two markers (**Paper I**). Compared with use of the 18S-V4 marker, use of the *rbc*L marker resulted in a higher number of taxonomic assignments matching those made with the microscopy approach, but it also resulted in less exact estimates of ecological status, with a higher number of sites labelled as having 'Bad' or 'Poor' status (**Paper I**). It was also found that the ecological status of different freshwater sites was underestimated when using the two markers, most likely reflecting the taxa-specific amplification differences between the markers. Thus marker performance was directly linked to the taxa present in the samples. In **Paper II**, there was evidence that the discriminating power of the different DNA markers may vary between taxa, as some taxa seemed to be better detected and separated when using one or the other DNA marker, with different numbers of reads assigned to a specific taxon, and some taxa were only detected by one marker. Similar differences between these two DNA markers have been observed (Apothéloz-Perret-Gentil *et al.*, 2020, Zinger *et al.*, 2019) and they are not due solely to the quality and quantity of reads in the reference databases. While more reference barcodes for a taxon can allow for detection of wider genetic variability, the results in **Paper II** indicated that having many reference barcodes does not ensure detection of a taxon and that in fact one reference barcode may be sufficient for reliable detection (**Paper II**). Ultimately, preferential identification of diatom taxa by different DNA markers is still only partly understood and can be affected by external parameters during the PCR and sequencing stages (Alberdi *et al.*, 2018, Bálint *et al.*, 2016). In **Paper II**, better reproducibility of results was obtained when using the *rbc*L

marker (results better correlated to those obtained by microscopy and those of other molecular pipelines) in terms of taxa diversity and ecological assessment. Even pipelines optimised for the 18S-V4 marker performed better with the *rbc*L marker in this regard (**Paper II**). This could be attributable in part to the great curation and updating efforts performed to date for the *rbc*L marker in the diat.barcode reference database (which includes barcodes for both DNA markers, but the *rbc*L barcodes cover a larger number of taxa) and the lack of curation as yet for the 18S-V4 barcodes, and to the use of degenerated primer pairs compared with those used for the 18S-V4 marker (**Paper II**; see also Apothéloz-Perret-Gentil *et al.*, 2020, Vasselon *et al.*, 2017, Visco *et al.*, 2015). Moreover, as a hyper-variable region, the 18S-V4 marker may be more susceptible to insertion, deletion, and substitution mistakes (Bálint *et al.*, 2016, Boyer *et al.*, 2016, Brown *et al.*, 2015, Gaonkar *et al.*, 2018, Tedersoo *et al.*, 2018). Overall, both DNA markers perform well in molecular identification of diatoms and are currently used by many laboratories. Ultimately, for implementation in environmental assessment and routine monitoring, the choice of marker matters less than the selection of a well-curated reference database.

Lack of methodological independence (trait and molecular approaches)

The trait approach as applied in **Papers III** and **IV** is not entirely independent from the microscopy approach, because the trait information used is sorted by taxa-specific features in the database (Rimet and Bouchez, 2012). Thus, identified morphotaxa were used to infer the trait they exhibited. Another method could be to directly identify specific features (life form in terms of size, colony form, mucous secretion) using live material. This retains similarities to morphotaxa identification, however, since it needs be performed under a microscope and on fresh material, as many trait structures (colony form, stalk, mucous secretion) are not visible after the process of making permanent slides. Thus, while trait identification requires less time and less expertise than species-level identification, it is still demanding and can suffer from human bias and errors, which can be a source of discrepancies between laboratories and uncertainty in the results (Kahlert *et al.*, 2009, Kahlert *et al.*, 2012). Besides, most ecological information available on diatom traits is sorted according to taxa (as is the case *e.g.* for the database developed by Rimet and Bouchez (2012) used here) and is not always easy to access. To facilitate further development of the trait approach, diatom trait data and information must be made readily available, in an open-

access and user-friendly repository, *e.g.* combined with DNA barcode information in reference databases like diat.barcode or the EMBL Nucleotide Sequence Database (Baker *et al.*, 2000). Another current limitation of the trait-based approach is that it requires selection of traits that are meaningful for the gradient under study (**Paper IV**). To ensure stable and reliable responses of the trait data, this selection needs be done for each new gradient or if a broader range of gradients is covered. This is currently limiting the reproducibility and inter-study comparability of the method.

5.2.3 Best practice for the development of new tools

At the time of this thesis work, there was no existing tool for ecological assessment using diatom HTS or trait data, so application of the two novel approaches required adaptation of existing tools (*e.g,* IPS) or development of new tools. A few recent studies have explored the potential of index adaptation (Kelly *et al.*, 2020) and design of a new index for environmental assessment based on diatoms (Apothéloz-Perret-Gentil *et al.*, 2020). This thesis provides evidence that HTS data and morphotaxa data are fundamentally different and that they respond differently to environmental changes. Thus, developing new indices calibrated to this new type of data is a logical step forward in methodology improvement and better assessment reliability.

Indicator values

The work to develop a new index in **Paper IV** showed that this is not a simple process and is heavily influenced by the choice of training dataset and indicator values (optima and tolerance of each diatom unit). There is a trade-off between giving too much weight to a few very abundant taxa or to the presence of few cells even if not abundant. However, the results in **Paper IV** showed that index performance was better when optima and tolerance values were based on abundance data rather than presence/absence data, both for ASVs and for trait units.

Training datasets

P**aper IV** also revealed that, while the trait-based index performed equally well for stream and lake sampling sites, the molecular index performed better for stream sites. The diatom communities were significantly different between stream and lake samples, and thus it is

possible that separate indices for lakes and streams need to be developed. However, in a previous study the currently used national index IPS showed an equally strong response to nutrient enrichment at Scandinavian lake and stream sites with similar inter-community differences (Kahlert and Gottschalk, 2014), indicating that it is possible to have one index for both types of water body. It is likely that the difference observed in **Paper IV** was due to a higher number of stream sites (n=76) than lake sites (n=39) and the broader environmental gradient covered by these in the training dataset. This highlights the importance of extensive and balanced training datasets for index development, which may be difficult to obtain without access to large datasets from monitoring programmes. Besides, as mentioned previously, it is not always possible to combine datasets originating from different sources because of data incompatibility or because of the heavy harmonisation needed. For example, in order to use the full dataset of 180 Scandinavian samples in P**aper I**, data from different countries had to be combined, harmonisation of taxonomic names had to be performed to account for laboratory bias and only water chemistry parameters present in all the datasets could be retained. Such harmonisation is time-consuming and not always possible, and results in partial data loss. Similarly, only HTS data from the same sequencing technology can be combined to create large datasets and raw data should be used whenever possible, because differences in bioinformatics treatment can have consequences for data compatibility (*e.g*. the impossibility of combining ISU and OTU data).

Calibration for international use of the trait-approach

While traits should be similar under similar conditions and thus can in theory be used in similar habitats without calibration, the traits that are meaningful differ depending on the gradient considered. This means that new unit-specific indicator values (*e.g*. optima and tolerance range) have to be defined for new gradients and, in the case of a broader range, for a gradient for which an index has already been developed. Moreover, trait information on a specific type of habitat may not be available. For example, Fennoscandia sites are ecologically different from sites in central Europe or North America where the trait data used in **Papers III** and **IV** were collected and where the guild classification was developed (Rimet and Bouchez, 2012, Passy, 2007). This means that trait information for some key taxa can be missing or that the classification dynamics can be different. The overall dominance of the low-profile guild in most samples in both **Papers III** and **IV** was a good

example of such consequences. In a more distinctly different ecoregion, such as the tropics, careful calibration of the trait data might be needed prior to the development of any new trait-based index. Moreover, indices should always be refined, by introducing new sites to broaden the ecological gradient for which they are calibrated. They should also be checked regularly by testing their applicability at new sites, to ensure that new values are improving the index and not merely changing it (*e.g.* the recalibration done for TDI in Kelly *et al.* (2008)).

5.2.4 Standards and guidelines to ensure comparability and quality

For future implementation of novel methods into routine monitoring and environmental assessment, it is necessary to propose realistic standards or guidelines for end-users. These should provide an easily replicable and understandable framework, in order to minimise and control as much of the potential bias as possible and avoid erroneous conclusions (Zinger *et al.*, 2019). While standards for diatom sampling already exist (EN 13946:2014) (CEN, 2014), and can be applied to both the molecular and the trait approach, guidelines for the other aspects of the methodologies need to be defined. For the metabarcoding method, recommendations are needed on sample preservation, DNA extraction and amplification, sequencing and finally sequence analysis with bioinformatics tools. Previous and on-going studies have compared methods in order to make best-practice recommendations for sample processing (Stein *et al.*, 2013, Vasselon *et al.*, 2021), but few have investigated the bioinformatics aspect. The comparison of bioinformatics pipelines and their results in **Paper II** provided evidence that it is impossible to recommend one pipeline over another. It was evident that the use of different pipelines generated disparities in the output data, but also that the more similar the bioinformatic parameters, the more similar the results. For example, similar output datasets can be generated with strict cleaning of sequences or with strict taxonomic assignment. This confirms that the bioinformatics aspect of the metabarcoding method can be developed with best-practice recommendations (in the form of standards or guidelines) to ensure reproducibility of results and harmony in biomonitoring. Implementation of positive and negative controls in all steps, *e.g.* by including controlled mock data in the analysis (Elbrecht and Steinke, 2019, Siegwald *et al.*, 2017, Zinger *et al.*, 2019), would allow the integrity of bioinformatics outputs to be checked. It would also allow for broad

applicability of the molecular approach, not restricted to specific methods or pipelines. Further, it is important to advertise open-access resources to end-users, by including raw data and a repository for good reference databases for taxonomic assignment and bioinformatics scripts used. Similarly, guidelines for critical steps in the trait approach are needed, *e.g*. for meaningful trait selection, as well as open access to most complete trait data. Best-practice recommendations for designing new indices are also needed for stable performance and reliable results, such as adequate range of gradient covered, minimum number of samples in the training dataset for reliable calibration and minimum number of units (DNA units or trait units). Ways to establish new indicator values or ways in which to use units without indicator values should also be established. Last but not least, before implementation into official legislation, reference values for unaffected sites must be agreed.

Of course, both novel approaches are still under development and will see many new improvements and alternative methodologies in the future. Use of a molecular approach, while not yet calibrated for reliable ecological assessment of waterbodies in Fennoscandia, provides valuable information on the diatom communities. There is a crucial need to better understand diatom ecology and distribution, particularly on a larger scale (most studies are only local or regional). In **Paper I**, application of the newly tested molecular method for environmental assessment mostly suffered from the constraint of achieving taxonomic assignment, which was limited by reference database incompleteness. On the other hand, **Papers I** and **II** provided evidence that accurate detection and abundance assessment of dominant and key taxa is sufficient to calculate an accurate IPS score, but the powerful depth of identification enabled by the molecular method can potentially also be used for environmental research on diatom taxonomy and diversity. Besides, bypassing taxonomic names means the loss of ecological information such as trait features or key taxa identification. The importance of accurate taxa detection and quantification highlights the need for coordinated curation and completion efforts for open-access reference databases such as diat.barcode, to improve the depth of taxonomic assignment. Best-practice recommendations also need to be applied when adding or curating information into a reference database (Rimet *et al.*, 2021), because the work in this thesis provided evidence that having great quantity of reference barcodes in a database does not ensure better identification of

taxa (at least when working with Fennoscandia diatoms) and that the quality of the reference barcodes matters more. For the purposes of ecological assessment, the potential of other metrics should be tested for HTS data, and new indices should be designed and calibrated to this new type of data, such as the emerging taxonomy-free methods (Paper IV; Apothéloz-Perret-Gentil *et al.*, 2020, Cordier *et al.*, 2021). To achieve better representation of diatom communities (especially since microscopy ultimately still gives a biased representation), alternative approaches to diatom community assessment based on diatom life forms should be further investigated. Finally, further knowledge on diatom traits is needed, and the available databases should be regularly updated by adding traits for missing taxa, as done in **Paper IV**.

5.3 Exploring the potential of trait-based assessment to link diatom community response to changes in environmental conditions

The analyses in **Paper III** showed that more of the variation in the diatom community was explained by the stressors applied to the systems in the trait-based approach (65%), compared with the morphotaxa or molecular approaches (45% and 30%, respectively), highlighting the potential of this method of assessment to reflect the diatom community response in Fennoscandia. The analyses also showed that, of the three stressors applied, the light factor had the strongest impact in shaping the diatom communities present, followed by flow disturbances (pulse, press) and finally by nutrient enrichment. A strong impact of light availability on diatoms, or of parameters directly modifying ligh exposure, has been reported previously (Lange *et al.*, 2011, Stenger-Kovács *et al.*, 2013, Tapolczai *et al.*, 2017). Overall, in **Paper III,** nutrient enrichment was the stressor evoking the most unclear response from the diatom community. This implies that, even in assessments where the nutrient gradient is the main focus (*e.g.* eutrophication assessment), other factors, most likely not measured in routine monitoring, also strongly influence the diatom community and its response to nutrients. It has been shown previously that reduced algae cell energy in low light conditions impacts nutrient uptake efficiency (Falkner *et al.*, 1980, Wynne and Rhee, 1988) and that light intensities below saturation increase the minimum amount of nitrogen required by algae cells (Healy, 1985, Rhee and Gotham, 1981, Wynne and Rhee, 1986, Zevenboom *et al.*, 1980). Similarly, the

impact of nutrient enrichment on the diatom community will be stronger in low light conditions because of strong competition for resources (Lange *et al.*, 2011). Flow disturbances, especially drought events, also have a strong impact on benthic communities (Truchy *et al.*, 2020). Thus, these additional factors should be measured in monitoring whenever possible and their impact should be considered when assessing diatom responses to nutrients. On other hand, in **Paper III** the response of the communities to the directly measured water chemistry parameters (PO_4, NO_2, NO_3, conductivity, pH, total discharge) was much weaker than the response to the applied stressors and provided little information. Alternative parameters employed to assess nutrients, such as the commonly used TotP and TotN, could show a clearer response of the diatom traits. Total organic carbon could also be included, to give a better understanding of overall water clarity (and amount of light penetration).

In the trait-based approach, the concept of ecological guilds provided a useful framework for understanding the link between diatom traits and their functions. A decision was made to use only the ecological guild trait in **Paper III**, because the guilds are defined from a combination of life-form strategies relevant for the response of diatoms to nutrient gradients and water movements. The four guilds are based on morphological traits, which are also called 'response traits' since they directly reflect the response of a taxon to its environment. They are easier to observe than functional or 'effect' traits in small unicellular organisms. Passy (2007) and Rimet and Bouchez (2012) describe the ecological preferences of each guild regarding nutrient levels and flow disturbances. In **Paper III**, dominance of low-profile taxa (small, prostrate cells) was observed in all samples, with a major contribution of *Achnanthidium minutissimum*, a pioneer taxon well-suited for disturbed environments. It is likely that low-profile taxa, which are typical of nutrient-poor water, were already abundant in the water pumped into the experimental channels or in the channel substrate, and thus had an advantage in colonisation of the sampling tiles set in the experiment. The dominance of this guild in the samples made it difficult to evaluate its response to changing stressors and the results only provided a partial understanding of the situation, but agreed with expectations reported by Rimet and Bouchez (2012) for low-profile cells (adapted to nutrient-poor habitats and low flow velocities). Similarly, the expected higher colonisation of high-profile taxa with higher light availability, as opposed to shaded conditions, was observed

in **Paper III**. Finally, as mentioned previously, the abundance of taxa from the motile guild was clearly correlated with the nutrient gradient and significantly increased with nutrient enrichment if no flow disturbance was applied. Motile taxa can be expected to be well-suited to eutrophic conditions, since they can move through the thick biofilm formed by other algae in such nutrient-rich conditions but do not tolerate high flow velocities (Kelly *et al.*, 2009, Rimet and Bouchez, 2012). However, key taxa that are not yet included in the reference database need to be manually classified into trait categories from information available in the literature (and added to trait databases) to avoid important data loss and improve the trait-based approach. This complementation of database information using literature sources is time-consuming but was done for the trait-based assessment in **Paper IV** (see Material and Methods section in that paper).

While the study described in **Paper III** was conducted in an experimental setting, in **Paper IV** the response of diatom traits was assessed using a large dataset of natural samples covering a broad ecological gradient in Fennoscandia and Iceland (115 samples). In preliminary analyses, rather low correlations were observed between the trait responses and the environmental variables (see NMDS and CCA analysis in **Paper IV**). A few isolated traits followed specific environmental gradients, *e.g*. motile guild abundance was positively correlated with the TOC gradient, but no clear pattern was visible for all the categories of one specific type of trait. However, when several traits were combined to form groups, it became clear that the community structure was indeed driven by the environmental factors measured (TotP, TotN, pH, TOC, conductivity). The trait combination that correlated best with the environmental factors was the full combination of all traits and their categories, including different size, surface to volume ratio and ecological guild (**Paper IV**). This shows that diatom traits form a complex assemblage, just like taxa, and may not be completely understood separately, highlighting the importance of including several meaningful traits and of considering their combined response to get a better picture of the host ecosystem. Future studies considering diatom traits in more detail and monitoring additional environmental variables are needed for a more accurate understanding of the community response.

5.4 Initial development of novel tools for environmental assessment based on novel approaches

In **Paper IV**, the potential of the two novel approaches for development of a new nutrient index to be used in environmental assessment of Fennoscandian streams and lakes was tested. Three types of diatom community data (morphotaxa, ASVs and traits) were compared when developing the nutrient index. A study by Tapolczai *et al.* (2021) using ESVs and morphotaxa to develop an index for assessing the impact of land use on diatom communities in Hungary provided a framework on which to base index development in **Paper IV**.

First, the driving environmental factors in the dataset were identified. It was found that, while the diatom communities reflected some of the environmental variables quite well for all three approaches (see NMDS analyses in **Paper IV**), diatom structure was also affected by other factors not assessed in the study, *e.g.* flow rate, light availability or grazing (Lange *et al.*, 2011, Liess and Kahlert, 2007, Stenger-Kovács *et al.*, 2013, Tapolczai *et al.*, 2017). The morphotaxa were found to be better correlated to the pH gradient than to any other gradient. The percentage of variance explained by the environmental variables was higher for the morphotaxa data than for the other data tested (ASVs and traits), partly due to the high correlation to the strong pH gradient, which added to the explanatory power (**Paper IV**). However, the best correlation in common for all three approaches between the community structure and the environmental variables was with the TotP gradient. This justifies the development of an index directly linked to phosphorus levels in water in order to monitor eutrophication in streams and lakes, a widespread pressure in European waters (Carvalho *et al.*, 2019, Poikane *et al.*, 2021). When using molecular units (ASVs) more variation in the assemblage structure was observed (the NMDS solution was reached at a higher stress value) than with the morphotaxa, probably because there were more ASV variables than morphotaxa variables. This in turn contributed to the higher variance explained by the environmental variables for the morphological taxa (see CCA analysis in **Paper IV**), since there were fewer variables to constrain and explain for this ordination.

With the molecular approach, optima and tolerance values defined from the relative abundance of ASVs in the training dataset resulted in a nutrient index which performed better than when using values defined from the presence/absence data (**Paper IV**). There was a significant difference

between lake and stream samples, with the TotP optima values of stream ASVs and lake ASVs correlated, but in a far from 1:1 correlation (see Figure 7 in P**aper IV**). This difference can be explained by the smaller number of samples for lake sites and the narrower nutrient gradient covered by these compared with the samples for stream sites. This in turn had an impact on the index application, with slightly better performance of the index when tested on the stream sites compared with the lake sites (correlation coefficient between expected and observed TotP values: $R^2 = 0.8$ for lake sites, $R^2 = 0.9$ for stream sites, **Paper IV**). Overall, the ASV-based nutrient index showed very promising results. First, it performed better than the other two indices tested in **Paper IV** ($R^2 = 0.9$ for molecular units; $R^2 = 0.8$ for morphotaxa; $R^2 = 0.7$ for 'fully combined' traits), which confirms the potential of molecular data (in this case, ASVs) for environmental assessment, especially eutrophication assessment. Second, it provided valuable insights into diatom ecology, as it revealed differences and similarities between the dominant morphotaxa and the dominant AVSs in the datasets. For example, while *Achnanthidium minutissimum* was the dominant morphotaxon, the dominant ASV was identified as *Melosira varians*. In contrast, some taxa (*e.g*. *Tabellaria flocculosa*, *Fragilaria capucina* and *Amphora pediculus*) were dominant in both datasets. Discrepancies between the two approaches, in terms of detection and quantification, are already well known (Vasselon *et al.*, 2018; **Papers I-IV**). An important new finding made in the present work was that several distinct ASVs were identified as the same taxon, and in some cases with different geographical distribution, with different ASVs present in different countries. Another important finding was that different ASVs were assigned to the same taxon, but with different ecological profiles (different TotP optimum and tolerance values, Figure 5 in **Paper IV**). This confirms the potential of the molecular approach for providing a better understanding of diatom diversity and ecology. It also highlights the need to study morphology data and molecular data together in further development work, to understand how they relate to and differ from one another.

With the trait approach, the optima and tolerance values defined from the relative abundance of traits resulted in a nutrient index that performed better than when using values defined from the presence/absence of traits instead, similarly to the ASV and morphotaxa datasets (**Paper IV**). However, no significant differences between stream and lake sites were observed in terms of optimum and tolerance values, or in terms of index performance

(correlation coefficient between expected and observed TotP: R^2= 0.7 for the nutrient index on lake sites and R^2= 0.68 for the nutrient index on stream sites, **Paper IV**). On other hand, the trait-based nutrient index performed slightly worse than the other two indices tested in **Paper IV** (R^2 = 0.9 for ASVs; R^2 = 0.8 for morphotaxa; R^2 = 0.7 for 'combined traits'), but the approach still showed promising results. One possible reason for this difference in performance is that the trait dataset included significantly fewer variables (66 groups) than the morphotaxa dataset (530 taxa) or ASV dataset (5467 ASVs), possibly capturing less of the diatom community variation. A decision was made in this work to combine the traits into 66 groups, based on evidence that combinations of several traits showed a better relationship to the environmental variables than individual traits ("pure" and "fully combined traits" in NMDS analysis in **Paper IV**). The results also indicated that not all traits are useful to reflect a nutrient gradient and that selection of meaningful traits depends on the environmental gradient under study. Better knowledge is needed of the relationship between diatom traits and environmental factors. For nutrients, for example, it is known that cell size and surface to volume ratio are good indicators, as small cells are more abundant in nutrient-poor conditions thanks to their relatively higher nutrient uptake (Lange *et al.*, 2016), since high surface to volume ratio optimises the area exposed to the environment and allow for fast resource acquisition (Snoeijs *et al.*, 2002). However, it was observed in **Paper IV** that some taxa were clearly not only driven by the environmental variables measured. For example, *Achnanthidium minutissimum* was at the extreme range of size, surface to volume ratio and guild trait categories and often a dominant specimen in Fennoscandian diatom communities, and yet did not directly respond to the nutrient gradient. As a pioneer taxon, *A. minutissimum* is well-suited to resist many environmental disturbances and it is likely that it was responding to a gradient not measured in **Paper IV**. Similar observations have been made in other applications of trait-based approaches (Paper III; Cardinale *et al.*, 2006, Passy, 2007, Rimet and Bouchez, 2012, B-Béres *et al.*, 2016, Lange *et al.*, 2016). The impact of other factors, not routinely monitored, such as water movement, light and grazing, also needs to be better understood for further development of trait-based methods. Finally, expansion of available trait databases with additional relevant traits and combination of more meaningful traits should improve the trait approach for use in environmental assessment.

6. Conclusions and future perspectives

This thesis explored development and application of two novel approaches as alternatives to the conventional microscopy approach for environmental assessment using benthic diatoms. Investigations on whether novel molecular- and trait-based approaches could be used in environmental assessment revealed that diatom communities identified by both approaches responded well to environmental changes, but that diatom DNA and trait data are more complex than microscopy data and cannot be used with the same tools. Analyses on the implications of differences between the methods revealed that using DNA or trait methods could potentially represent an advantage for routine use both in research and in environmental assessment, circumventing the expertise and time needed for morphotaxa identification. Both novel approaches also provided valuable information not grasped by the microscopy approach, such as detection of cryptic taxa, unknown ecological profiles and fine-tuned responses of the diatom community to direct and indirect environmental impacts. However, both methods are currently limited by low inter-comparability of the outputs, a shortcoming originating from multiple sources, such as incomplete reference databases and the variety of tools used in laboratories. A study in an experimental setting exploring the links between diatom DNA and traits and diatom community response to environmental changes, represented by three environmental stressors, revealed that diatom ASVs and traits respond strongly to light and flow stressors, but more weakly to nutrient enrichment. However, both types of data showed a strong correlation to the total phosphorus gradient in natural samples, a finding used in development of novel tools for environmental assessment of eutrophication. New indices based on diatom ASVs, and on combinations of traits, performed well in

assessing the level of total phosphorus in natural freshwater samples, with indicator values derived from their relative abundances.

In conclusion, this thesis showed that novel molecular- and trait-based approaches have strong potential for index development for eutrophication assessment based on diatoms in Fennoscandia freshwaters. Both approaches also provide valuable additional information, benefit from a methodology requiring less expertise and can produce more reproducible results than the microscopy approach.

The findings in this thesis have implications for future development of molecular-based and trait-based approaches for characterising diatom communities and their application in monitoring and environmental assessment work. Environmental assessment of aquatic environments is important to preserve water resources, while biomonitoring is essential to detect environmental changes that might impact ecosystem biodiversity and ecosystem services supply, so together they provide a solid foundation for management programmes. Both processes would benefit greatly from use of new technologies and novel approaches to make them easier to apply by the end-user and provide more replicable and comparable results for decision-makers. For future development of molecular- and trait-based approaches, a better understanding of the response of DNA and trait data to environmental gradients is needed, as is progress in curation and completion of reference databases and best-practice guidelines to enable standardisation of the methodology. It has been suggested that application of the molecular method in environmental assessment could also benefit from a change from descriptive tools (like indices) toward predictive tools that include the underlying dynamics of ecosystems in their methodology (Makiola *et al.*, 2020). The light microscopy approach benefited from around 300 years of research to develop the methods and tools used today, while morphotaxa data have been used for more than 30 years to calibrate and recalibrate tools for environmental assessment using diatoms (Charles *et al.*, 1990, Kingston *et al.*, 1992). In comparison, diatom trait and diatom metabarcoding approaches are still in their infancy, but both methods already show as much potential for environmental assessment as conventional microscopy did 30 years ago. We must therefore keep exploring these two alternative approaches, to enable better, faster, and stronger monitoring of freshwater, our most precious resource.

7. References

ABARCA, N., JAHN, R., ZIMMERMANN, J. & ENKE, N. 2014. Does the Cosmopolitan Diatom Gomphonema parvulum (Kützing) Kützing Have a Biogeography? *PLOS ONE,* 9**,** e86885.

ALBERDI, A., AIZPURUA, O., GILBERT, M. T. P. & BOHMANN, K. 2018. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution,* 9**,** 134-147.

APOTHÉLOZ-PERRET-GENTIL, L., BOUCHEZ, A., CORDIER, T., CORDONIER, A., GUÉGUEN, J., RIMET, F., VASSELON, V. & PAWLOWSKI, J. 2020. 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. *Molecular Ecology,* n/a.

APOTHÉLOZ-PERRET-GENTIL, L., CORDONIER, A., STRAUB, F., ISELI, J., ESLING, P. & PAWLOWSKI, J. 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular Ecology Resources,* 17**,** 1231-1242.

B-BÉRES, V., LUKÁCS, Á., TÖRÖK, P., KÓKAI, Z., NOVÁK, Z., KRASZNAI, E., TÓTHMÉRÉSZ, B. & BÁCSI, I. 2016. Combined eco-morphological functional groups are reliable indicators of colonisation processes of benthic diatom assemblages in a lowland stream. *Ecological Indicators,* 64**,** 31-38.

B-BÉRES, V., TÖRÖK, P., KÓKAI, Z., LUKÁCS, Á., T-KRASZNAI, E., TÓTHMÉRÉSZ, B. & BÁCSI, I. 2017. Ecological background of diatom functional groups: Comparability of classification systems. *Ecological Indicators,* 82**,** 183-188.

BAKER, W., VAN DEN BROEK, A., CAMON, E., HINGAMP, P., STERK, P., STOESSER, G. & TULI, M. A. 2000. The EMBL nucleotide sequence database. *Nucleic acids research,* 28**,** 19-23.

BÁLINT, M., BAHRAM, M., EREN, A. M., FAUST, K., FUHRMAN, J. A., LINDAHL, B., O'HARA, R. B., ÖPIK, M., SOGIN, M. L., UNTERSEHER, M. & TEDERSOO, L. 2016. Millions of reads, thousands

of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews,* 40**,** 686-700.

BATTARBEE, R., JUGGINS, S., GASSE, F., ANDERSON, N., BENNION, H., CAMERON, N. G., RYVES, D., CHRISTINE, P., CHALIE, F. & TELFORD, R. 2001. European Diatom Database (EDDI). An information system for palaeoenvironmental reconstruction.

BERTHON, V., BOUCHEZ, A. & RIMET, F. 2011. Using diatom life-forms and ecological guilds to assess organic pollution and trophic level in rivers: a case study of rivers in south-eastern France. *Hydrobiologia,* 673**,** 259-271.

BICHOFF, A., OSÓRIO, N. C., RUWER, D. T., DUNCK, B. & RODRIGUES, L. 2018. Trait structure and functional diversity of periphytic algae in a floodplain conservation area. *Brazilian Journal of Botany,* 41**,** 601-610.

BORCHARDT, M. A. 1996. 7 - Nutrients. *In:* STEVENSON, R. J., BOTHWELL, M. L. & LOWE, R. L. (eds.) *Algal Ecology.* San Diego: Academic Press.

BORJA, A., MILES, A., OCCHIPINTI, A. & BERG, T. 2009. Current status of macroinvertebrate methods used for assessing the quality of European marine waters: Implementing the Water Framework Directive. *Hydrobiologia,* 633**,** 181-196.

BOYER, F., MERCIER, C., BONIN, A., LE BRAS, Y., TABERLET, P. & COISSAC, E. 2016. obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources,* 16**,** 176-182.

BROWN, E. A., CHAIN, F. J. J., CREASE, T. J., MACISAAC, H. J. & CRISTESCU, M. E. 2015. Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecology and Evolution,* 5**,** 2234-2251.

BURFEID-CASTELLANOS, A. M., MARTÍN-MARTÍN, R. P., KLOSTER, M., ANGULO-PRECKLER, C., AVILA, C. & BESZTERI, B. 2020. Epiphytic diatom community structure and richness is determined by macroalgal host and location in the South Shetland Islands (Antarctica). PANGAEA.

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. *Nature Methods,* 13**,** 581-583.

CARDINALE, B. J., HILLEBRAND, H. & CHARLES, D. F. 2006. Geographic patterns of diversity in streams are predicted by a multivariate model of disturbance and productivity. *Journal of Ecology,* 94**,** 609-618.

CARVALHO, L., MACKAY, E. B., CARDOSO, A. C., BAATTRUP-PEDERSEN, A., BIRK, S., BLACKSTOCK, K. L., BORICS, G., BORJA, A., FELD, C. K., FERREIRA, M. T., GLOBEVNIK, L., GRIZZETTI, B., HENDRY, S., HERING, D., KELLY, M., LANGAAS, S., MEISSNER, K., PANAGOPOULOS, Y., PENNING, E., ROUILLARD, J., SABATER, S., SCHMEDTJE, U., SPEARS, B. M., VENOHR, M., VAN DE BUND, W. & SOLHEIM, A. L. 2019. Protecting and restoring Europe's waters: An analysis of the future development needs of the Water Framework Directive. *Science of The Total Environment,* 658**,** 1228-1238.

CEMAGREF 1982. Etudes des méthodes biologiques quantitative d'appréciation de la qualité des eaux. Bassin Rhône-Méditerranée-Corse. *In:* CENTRE NATIONAL DU MACHINISME AGRICOLE, D. G. R., DES EAUX ET DES FORÊTS. (ed.). Lyon, France.

CHARLES, D. F., BINFORD, M. W., FURLONG, E. T., HITES, R. A., MITCHELL, M. J., NORTON, S. A., OLDFIELD, F., PATERSON, M. J., SMOL, J. P., UUTALA, A. J., WHITE, J. R., WHITEHEAD, D. R. & WISE, R. J. 1990. Paleoecological investigation of recent lake acidification in the Adirondack Mountains, N.Y. *Journal of Paleolimnology,* 3**,** 195-241.

CORDIER, T., ALONSO-SÁEZ, L., APOTHÉLOZ-PERRET-GENTIL, L., AYLAGAS, E., BOHAN, D. A., BOUCHEZ, A., CHARITON, A., CREER, S., FRÜHE, L., KECK, F., KEELEY, N., LAROCHE, O., LEESE, F., POCHON, X., STOECK, T., PAWLOWSKI, J. & LANZÉN, A. 2021. Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. *Molecular Ecology,* 30**,** 2937-2958.

CORDIER, T., FORSTER, D., DUFRESNE, Y., MARTINS, C. I. M., STOECK, T. & PAWLOWSKI, J. 2018. Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources,* 18**,** 1381-1391.

CORDIER, T., LANZÉN, A., APOTHÉLOZ-PERRET-GENTIL, L., STOECK, T. & PAWLOWSKI, J. 2019. Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring. *Trends in Microbiology,* 27**,** 387-397.

DOLÉDEC, S. & STATZNER, B. 2008. Invertebrate traits for the biomonitoring of large European rivers: an assessment of specific types of human impact. *Freshwater Biology,* 53**,** 617-634.

DONG, X., LI, B., HE, F., GU, Y., SUN, M., ZHANG, H., TAN, L., XIAO, W., LIU, S. & CAI, Q. 2016. Flow directionality, mountain barriers and functional traits determine diatom metacommunity structuring of high mountain streams. *Scientific Reports,* 6**,** 24711.

ELBRECHT, V. & STEINKE, D. 2019. Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. *Freshwater Biology,* 64**,** 380-387.

EUROPEAN UNION 2000. Water Framework 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. *Official Journal (OJ L 327).*

FALKNER, G., HORNER, F. & SIMONIS, W. 1980. The regulation of the energy-dependent phosphate uptake by the blue-green alga Anacystis nidulans. *Planta,* 149**,** 138-43.

FERNKVIST, M. 2020. *DESTRESS (DisEntangling the impacts of multiple Stressors on stream ecosystems)* [Online]. Sveriges Lantbruksuniversitet.

Available: <https://www.slu.se/institutioner/ekonomi/forskning/projekt-sidlistning/projekt/destress/> [Accessed 2021-07-30].
FRIBERG, N., DYBKJÆR, J. B., OLAFSSON, J. S., GISLASON, G. M., LAREN, S. E. & LAURIDSEN, T. L. 2009. Relationships between structure and function in streams contrasting in temperature. *Freshwater Biology,* 54**,** 2051-2068.
FRIGERIO, J.-M., RIMET, F., BOUCHEZ, A., CHANCEREL, E., CHAUMEIL, P., SALIN, F., THÉROND, S., KAHLERT, M. & FRANC, A. 2016. diagno-syst: a tool for accurate inventories in metabarcoding. *arXiv:1611.09410 q-bio.QM*.
GAONKAR, C. C., PIREDDA, R., MINUCCI, C., MANN, D. G., MONTRESOR, M., SARNO, D. & KOOISTRA, W. H. C. F. 2018. Annotated 18S and 28S rDNA reference sequences of taxa in the planktonic diatom family Chaetocerotaceae. *PLOS ONE,* 13**,** e0208929.
GUIRY, M. 2012. How many species of algae are there? *Journal of Phycology,* 48.
HAJIBABAEI, M., SINGER, G. A. C., HEBERT, P. D. N. & HICKEY, D. A. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics,* 23**,** 167-172.
HAMMER, O., HARPER, D. & RYAN, P. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica,* 4**,** 1-9.
HAVS- OCH VATTENMYNDIGHETEN 2018. Kiselalger i sjöar och vattendrag. Vägledning för statusklassificering. *Havs- och vattenmyndighetens rapport.* Göteborg.
HEALY, P. M. 1985. The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics,* 7**,** 85-107.
HEBERT, P. D., CYWINSKA, A., BALL, S. L. & DEWAARD, J. R. 2003. Biological identifications through DNA barcodes. *Proc Biol Sci,* 270**,** 313-21.
HILL, W. 1996. 5 - Effects of Light. *In:* STEVENSON, R. J., BOTHWELL, M. L. & LOWE, R. L. (eds.) *Algal Ecology.* San Diego: Academic Press.
KAHLERT, M., ÁCS, É., ALMEIDA, S. F. P., BLANCO, S., DREßLER, M., ECTOR, L., KARJALAINEN, S. M., LIESS, A., MERTENS, A., VAN DER WAL, J., VILBASTE, S. & WERNER, P. 2016. Quality assurance of diatom counts in Europe: towards harmonized datasets. *Hydrobiologia,* 772**,** 1-14.
KAHLERT, M., ALBERT, R.-L., ANTTILA, E.-L., BENGTSSON, R., BIGLER, C., ESKOLA, T., GÄLMAN, V., GOTTSCHALK, S., HERLITZ, E., JARLMAN, A., KASPEROVIČIENĖ, J., KOKOCINSKI, M., LUUP, H., MIETTINEN, J., PAUNKSNYTE, I., PIIRSOO, K., QUINTANA, I., RAUNIO, J., SANDELL, B. & WECKSTRÖM, J. 2009. Harmonization is more important than experience—results of the first Nordic–Baltic diatom intercalibration exercise 2007 (stream monitoring). *Journal of Applied Phycology,* 21**,** 471-482.

KAHLERT, M. & GOTTSCHALK, S. 2014. Differences in benthic diatom assemblages between streams and lakes in Sweden and implications for ecological assessment. *Freshwater Science,* 33**,** 655-669.

KAHLERT, M., KELLY, M., ALBERT, R.-L., ALMEIDA, S., BEŠTA, T., BLANCO, S., COSTE, M., DENYS, L., ECTOR, L., FRÁNKOVÁ, M., HLUBIKOVA, D., IVANOV, P., KENNEDY, B., MARVAN, P., MERTENS, A., MIETTINEN, J., PICINSKA-FALTYNOWICZ, J., TISON-ROSEBERY, J., TORNÉS, E. & VOGEL, A. 2012. Identification versus counting protocols as sources of uncertainty in diatom-based ecological status assessments. *Hydrobiologia,* 695.

KAHLERT, M., KELLY, M. G., MANN, D. G., RIMET, F., SATO, S., BOUCHEZ, A. & KECK, F. 2019. Connecting the morphological and molecular species concepts to facilitate species identification within the genus Fragilaria (Bacillariophyta). *Journal of Phycology,* 55**,** 948-970.

KELLY, M. 2013. Data rich, information poor? Phytobenthos assessment and the Water Framework Directive. *European Journal of Phycology,* 48**,** 437-450.

KELLY, M., BOONHAM, N., JUGGINS, S., KILLE, P., MANN, D., PASS, D., SAPP, M., SATO, S. & GLOVER, R. 2018a. A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. *In:* ENVIRONMENT AGENCY (ed.). Horizon House, Deanery Road, Bristol, BS1 5AH, UK.

KELLY, M., JUGGINS, S., GUTHRIE, R., PRITCHARD, S., JAMIESON, J., RIPPEY, B., HIRST, H. & YALLOP, M. 2008. Assessment of ecological status in U.K. rivers using diatoms. *Freshwater Biology,* 53**,** 403-422.

KELLY, M., JUGGINS, S., PHILLIPS, G. & WILLBY, N. 2018b. Evaluation of benthic diatom classification in UK rivers using LM and NGS methods. Report number E18-56 (PO 4057194).

KELLY, M., KING, L., XED, CHATH, XE & IN, B. 2009. The conceptual basis of ecological status assessment using diatoms. *Biology and Environment: Proceedings of the Royal Irish Academy,* 109B**,** 175-189.

KELLY, M., URBANIC, G., ACS, E., BENNION, H., BERTRIN, V., BURGESS, A., DENYS, L., GOTTSCHALK, S., KAHLERT, M., KARJALAINEN, S. M., KENNEDY, B., KOSI, G., MARCHETTO, A., MORIN, S., PICINSKA-FAŁTYNOWICZ, J., POIKANE, S., ROSEBERY, J., SCHOENFELDER, I., SCHOENFELDER, J. & VARBIRO, G. 2014. Comparing aspirations: intercalibration of ecological status concepts across European lakes for littoral diatoms. *Hydrobiologia,* 734**,** 125-141.

KELLY, M. G., JUGGINS, S., MANN, D. G., SATO, S., GLOVER, R., BOONHAM, N., SAPP, M., LEWIS, E., HANY, U., KILLE, P., JONES, T. & WALSH, K. 2020. Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding. *Ecological Indicators,* 118**,** 106725.

KELLY, M. G. & WHITTON, B. A. 1995. The Trophic Diatom Index: a new index for monitoring eutrophication in rivers. *Journal of Applied Phycology,* 7**,** 433-444.

KERMARREC, L., FRANC, A., RIMET, F., CHAUMEIL, P., FRIGERIO, J.-M., HUMBERT, J.-F. & BOUCHEZ, A. 2014. A Next-Generation Sequencing Approach to River Biomonitoring Using Benthic Diatoms. *Freshwater Science,* 33**,** 349-363, 15.

KERMARREC, L., FRANC, A., RIMET, F., CHAUMEIL, P., HUMBERT, J. F. & BOUCHEZ, A. 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Molecular Ecology Resources,* 13**,** 607-619.

KINGSTON, J. C., BIRKS, H. J. B., UUTALA, A. J., CUMMING, B. F. & SMOL, J. P. 1992. Assessing Trends in Fishery Resources and Lake Water Aluminum from Paleolimnological Analyses of Siliceous Algae. *Canadian Journal of Fisheries and Aquatic Sciences,* 49**,** 116-127.

KRUK, C., HUSZAR, V. L. M., PEETERS, E. T. H. M., BONILLA, S., COSTA, L., LÜRLING, M., REYNOLDS, C. S. & SCHEFFER, M. 2010. A morphological classification capturing functional variation in phytoplankton. *Freshwater Biology,* 55**,** 614-627.

LANGE, K., LIESS, A., PIGGOTT, J. J., TOWNSEND, C. R. & MATTHAEI, C. D. 2011. Light, nutrients and grazing interact to determine stream diatom community composition and functional group structure. *Freshwater Biology,* 56**,** 264-278.

LANGE, K., TOWNSEND, C. R. & MATTHAEI, C. D. 2016. A trait-based framework for stream algal communities. *Ecology and Evolution,* 6**,** 23-36.

LECOINTE, C., COSTE, M. & PRYGIEL, J. 1993. "Omnidia": software for taxonomy, calculation of diatom indices and inventories management. *Hydrobiologia,* 269-270**,** 509-513.

LEESE, F., ALTERMATT, F., BOUCHEZ, A., EKREM, T., HERING, D., MEISSNER, K., MERGEN, P., PAWLOWSKI, J., PIGGOTT, J. J., RIMET, F., STEINKE, D., TABERLET, P., WEIGAND, A. M., ABARENKOV, K., BEJA, P., BERVOETS, L., BJÖRNSDÓTTIR, S., BOETS, P., BOGGERO, A., BONES, A. M., BORJA, Á., BRUCE, K., BURSIĆ, V., CARLSSON, J., ČIAMPOR, F., ČIAMPOROVÁ-ZATOVIČOVÁ, Z., COISSAC, E., COSTA, F., COSTACHE, M., CREER, S., CSABAI, Z., DEINER, K., DELVALLS, Á., DRAKARE, S., DUARTE, S., ELERŠEK, T., FAZI, S., FIŠER, C., FLOT, J.-F., FONSECA, V., FONTANETO, D., GRABOWSKI, M., GRAF, W., GUÐBRANDSSON, J., HELLSTRÖM, M., HERSHKOVITZ, Y., HOLLINGSWORTH, P., JAPOSHVILI, B., JONES, J. I., KAHLERT, M., KALAMUJIC STROIL, B., KASAPIDIS, P., KELLY, M. G., KELLY-QUINN, M., KESKIN, E., KÕLJALG, U., LJUBEŠIĆ, Z., MAČEK, I., MÄCHLER, E., MAHON, A., MAREČKOVÁ, M., MEJDANDZIC, M., MIRCHEVA, G., MONTAGNA, M., MORITZ, C., MULK, V.,

NAUMOSKI, A., NAVODARU, I., PADISÁK, J., PÁLSSON, S., PANKSEP, K., PENEV, L., PETRUSEK, A., PFANNKUCHEN, M. A., PRIMMER, C. R., RINKEVICH, B., ROTTER, A., SCHMIDT-KLOIBER, A., SEGURADO, P., SPEKSNIJDER, A., STOEV, P., STRAND, M., ŠULČIUS, S., SUNDBERG, P., TRAUGOTT, M., TSIGENOPOULOS, C., TURON, X., VALENTINI, A., VAN DER HOORN, B., VÁRBÍRÓ, G., VASQUEZ HADJILYRA, M. I., VIGURI, J., VITONYTĖ, I., VOGLER, A., VRÅLSTAD, T., WÄGELE, W., WENNE, R., WINDING, A., WOODWARD, G., ZEGURA, B., *et al.* 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes,* 2**,** e11321.

LIESS, A. & KAHLERT, M. 2007. Gastropod grazers and nutrients, but not light, interact in determining periphytic algal diversity. *Oecologia,* 152**,** 101-11.

LINDHOLM, M., GRÖNROOS, M., HJORT, J., KARJALAINEN, S. M., TOKOLA, L. & HEINO, J. 2018. Different species trait groups of stream diatoms show divergent responses to spatial and environmental factors in a subarctic drainage basin. *Hydrobiologia,* 816**,** 213-230.

LITCHMAN, E. & KLAUSMEIER, C. A. 2008. Trait-Based Community Ecology of Phytoplankton. *Annual Review of Ecology, Evolution, and Systematics,* 39**,** 615-639.

LOGEZ, M., BADY, P., MELCHER, A. & PONT, D. 2013. A continental-scale analysis of fish assemblage functional structure in European rivers. *Ecography,* 36**,** 80-91.

MAKIOLA, A., COMPSON, Z. G., BAIRD, D. J., BARNES, M. A., BOERLIJST, S. P., BOUCHEZ, A., BRENNAN, G., BUSH, A., CANARD, E., CORDIER, T., CREER, S., CURRY, R. A., DAVID, P., DUMBRELL, A. J., GRAVEL, D., HAJIBABAEI, M., HAYDEN, B., VAN DER HOORN, B., JARNE, P., JONES, J. I., KARIMI, B., KECK, F., KELLY, M., KNOT, I. E., KROL, L., MASSOL, F., MONK, W. A., MURPHY, J., PAWLOWSKI, J., POISOT, T., PORTER, T. M., RANDALL, K. C., RANSOME, E., RAVIGNÉ, V., RAYBOULD, A., ROBIN, S., SCHRAMA, M., SCHATZ, B., TAMADDONI-NEZHAD, A., TRIMBOS, K. B., VACHER, C., VASSELON, V., WOOD, S., WOODWARD, G. & BOHAN, D. A. 2020. Key Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science,* 7.

MANN, D. G. & VANORMELINGEN, P. 2013. An inordinate fondness? The number, distributions, and origins of diatom species. *J Eukaryot Microbiol,* 60**,** 414-20.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011,* 17**,** 3.

MCCUNE, B. & MEFFORD, M. J. 2011. *PC-ORD : multivariate analysis of ecological data, Version 6.0 for Windows.,* Oregon, U.S.A., Glenedan Beach (Or.) : MjM software design.

MILJÖDATA-MVM. 2021. *National data host lakes and watercourses, and national data host agricultural land* [Online]. Swedish University of Agricultural Sciences (SLU). Available: https://miljodata.slu.se/mvm/DataContents/Omnidia [Accessed 2017].

MORA, D., CARMONA, J., JAHN, R., ZIMMERMANN, J. & ABARCA, N. 2017. Epilithic diatom communities of selected streams from the Lerma-Chapala Basin, Central Mexico, with the description of two new species. *PhytoKeys,* 88**,** 39-69.

MORITZ, C. & CICERO, C. 2004. DNA Barcoding: Promise and Pitfalls. *PLOS Biology,* 2**,** e354.

MOSER, G. A. O., PIEDRAS, F. R., OAQUIM, A. B. J., SOUZA, D. S., LELES, S. G., DE LIMA, D. T., RAMOS, A. B. A., FARIAS, C. D. O. & FERNANDES, A. M. 2017. Tidal effects on phytoplankton assemblages in a near-pristine estuary: a trait-based approach for the case of a shallow tropical ecosystem in Brazil. *Marine Ecology,* 38**,** e12450.

NEIF, É. M., GRAEBER, D., RODRIGUES, L., ROSENHØJ-LETH, S., JENSEN, T. M., WIBERG-LARSEN, P., LANDKILDEHUS, F., RIIS, T. & BAATTRUP-PEDERSEN, A. 2017. Responses of benthic algal communities and their traits to experimental changes in fine sediments, nutrients and flow. *Freshwater Biology,* 62**,** 1539-1550.

ÓLAFSSON, J. S., INGIMUNDARDÓTTIR, G. V., HANSEN, I. & SIGURÐARDÓTTIR, S. G. 2010. *Smádýralíf í afrennslisvatni frá háhitasvæðunum við Kröflu, Ölkelduháls og í Miðdal í Henglinum.,* Reykjavík, Veiðimálastofnun.

ORFANIDIS, S., PANAYOTIDIS, P. & STAMATIS, N. 2003. An insight to the ecological evaluation index (EEI). *Ecological Indicators,* 3**,** 27-33.

PADISÁK, J., BORICS, G., GRIGORSZKY, I. & SORÓCZKI-PINTÉR, É. 2006. Use of Phytoplankton Assemblages for Monitoring Ecological Status of Lakes within the Water Framework Directive: The Assemblage Index. *Hydrobiologia,* 553**,** 1-14.

PASSY, S. I. 2007. Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany,* 86**,** 171-178.

PAWLOWSKI, J., APOTHÉLOZ-PERRET-GENTIL, L. & ALTERMATT, F. 2020. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology,* 29**,** 4258-4264.

PAWLOWSKI, J., KELLY-QUINN, M., ALTERMATT, F., APOTHÉLOZ-PERRET-GENTIL, L., BEJA, P., BOGGERO, A., BORJA, A., BOUCHEZ, A., CORDIER, T., DOMAIZON, I., FEIO, M. J., FILIPE, A. F., FORNAROLI, R., GRAF, W., HERDER, J., VAN DER HOORN, B., IWAN JONES, J., SAGOVA-MARECKOVA, M., MORITZ, C., BARQUÍN, J., PIGGOTT, J. J., PINNA, M., RIMET, F., RINKEVICH, B., SOUSA-SANTOS, C., SPECCHIA, V., TROBAJO, R., VASSELON, V.,

VITECEK, S., ZIMMERMAN, J., WEIGAND, A., LEESE, F. & KAHLERT, M. 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of The Total Environment,* 637-638**,** 1295-1310.

POIKANE, S., VÁRBÍRÓ, G., KELLY, M. G., BIRK, S. & PHILLIPS, G. 2021. Estimating river nutrient concentrations consistent with good ecological condition: More stringent nutrient thresholds needed. *Ecological Indicators,* 121**,** 107017.

QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 341.

R CORE TEAM 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

RATNASINGHAM, S. & HEBERT, P. D. N. 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular ecology notes,* 7**,** 355-364.

RHEE, G.-Y. & GOTHAM, I. J. 1981. The effect of environmental factors on phytoplankton growth: Temperature and the interactions of temperature with nutrient limitation1. *Limnology and Oceanography,* 26**,** 635-648.

RIMET, F., AYLAGAS, E., BORJA, A., BOUCHEZ, A., CANINO, A., CHAUVIN, C., CHONOVA, T., CIAMPOR JR, F., COSTA, F. O., FERRARI, B. J. D., GASTINEAU, R., GOULON, C., GUGGER, M., HOLZMANN, M., JAHN, R., KAHLERT, M., KUSBER, W.-H., LAPLACE-TREYTURE, C., LEESE, F., LELIAERT, F., MANN, D. G., MARCHAND, F., MÉLÉDER, V., PAWLOWSKI, J., RASCONI, S., RIVERA, S., ROUGERIE, R., SCHWEIZER, M., TROBAJO, R., VASSELON, V., VIVIEN, R., WEIGAND, A., WITKOWSKI, A., ZIMMERMANN, J. & EKREM, T. 2021. Metadata standards and practical guidelines for specimen and DNA curation when building barcode reference libraries for aquatic life. *Metabarcoding and Metagenomics,* 5**,** e58056.

RIMET, F. & BOUCHEZ, A. 2012. Life-forms, cell-sizes and ecological guilds of diatoms in European rivers. *Knowl. Managt. Aquatic Ecosyst.***,** 01.

RIMET, F., CHAUMEIL, P., KECK, F., KERMARREC, L., VASSELON, V., KAHLERT, M., FRANC, A. & BOUCHEZ, A. 2016. R-Syst::diatom: an open-access and curated barcode database for diatoms and freshwater monitoring. *Database (Oxford),* 2016.

RIMET, F., GUSEV, E., KAHLERT, M., KELLY, M. G., KULIKOVSKIY, M., MALTSEV, Y., MANN, D. G., PFANNKUCHEN, M., TROBAJO, R., VASSELON, V., ZIMMERMANN, J. & BOUCHEZ, A. 2019. Diat.barcode, an open-access curated barcode library for diatoms. *Scientific Reports,* 9**,** 15116.

RIMET, F., VASSELON, V., A.-KESZTE, B. & BOUCHEZ, A. 2018. Do we similarly assess diversity with microscopy and high-throughput sequencing? Case of microalgae in lakes. *Organisms Diversity & Evolution,* 18.

RIVERA, S. F., VASSELON, V., BOUCHEZ, A. & RIMET, F. 2020. Diatom metabarcoding applied to large scale monitoring networks: Optimization of bioinformatics strategies using Mothur software. *Ecological Indicators,* 109**,** 105775.

RIVERA, S. F., VASSELON, V., JACQUET, S., BOUCHEZ, A., ARIZTEGUI, D. & RIMET, F. 2018. Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment. *Hydrobiologia,* 807**,** 37-51.

RUSANOV, A. G. & KHROMOV, V. M. 2016. Longitudinal distribution of periphyton algae in the Moskva river under eutrophication. *Water Resources,* 43**,** 513-521.

SAYERS, E. W., CAVANAUGH, M., CLARK, K., OSTELL, J., PRUITT, K. D. & KARSCH-MIZRACHI, I. 2019. GenBank. *Nucleic Acids Research,* 48**,** D84-D86.

SCHNECK, F., LANGE, K., MELO, A. S., TOWNSEND, C. R. & MATTHAEI, C. D. 2017. Effects of a natural flood disturbance on species richness and beta diversity of stream benthic diatom communities. *Aquatic Ecology,* 51**,** 557-569.

SCHNEIDER, S. & LINDSTRØM, E.-A. 2009. *Bioindication in Norwegian rivers using non-diatomaceous benthic algae: The acidification index periphyton (AIP)*.

SCHNEIDER, S. & LINDSTRØM, E.-A. 2011. *The periphyton index of trophic status PIT: a new eutrophication metric based on non-diatomaceous benthic algae in Nordic rivers*.

SIEGWALD, L., TOUZET, H., LEMOINE, Y., HOT, D., AUDEBERT, C. & CABOCHE, S. 2017. Assessment of Common and Emerging Bioinformatics Pipelines for Targeted Metagenomics. *PLOS ONE,* 12**,** e0169563.

SMOL, J. P. & STOERMER, E. F. 2010. *The Diatoms: Applications for the Environmental and Earth Sciences, Second Edition*, Cambridge University Press.

SNOEIJS, P., BUSSE, S. & POTAPOVA, M. 2002. The importance of diatom cell size in community analysis. *Journal of Phycology,* 38**,** 265-281.

SOININEN, J., JAMONEAU, A., ROSEBERY, J. & PASSY, S. I. 2016. Global patterns and drivers of species and trait composition in diatoms. *Global Ecology and Biogeography,* 25**,** 940-950.

STEIN, E. D., WHITE, B. P., MAZOR, R. D., MILLER, P. E. & PILGRIM, E. M. 2013. Evaluating Ethanol-based Sample Preservation to Facilitate Use of DNA Barcoding in Routine Freshwater Biomonitoring Programs Using Benthic Macroinvertebrates. *PLOS ONE,* 8**,** e51273.

STENGER-KOVÁCS, C., LENGYEL, E., CROSSETTI, L., ÜVEGES, V. & PADISAK, J. 2013. Diatom ecological guilds as indicators of temporally changing stressors and disturbances in the small Torna-stream, Hungary. *Ecological Indicators,* 24**,** 138-147.

STEVENSON, R. J., PAN, Y. & VAN DAM, H. 2010. Assessing environmental conditions in rivers and streams with diatoms. *In:* STOERMER, E. F. & SMOL, J. P. (eds.) *The Diatoms: Applications for the Environmental and Earth Sciences.* 2 ed. Cambridge: Cambridge University Press.

STOECKLE, M. 2003. Taxonomy, DNA, and the Bar Code of Life. *BioScience,* 53**,** 796-797.

TABERLET, P., COISSAC, E., POMPANON, F., BROCHMANN, C. & WILLERSLEV, E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol Ecol,* 21**,** 2045-50.

TANG, T., JIA, X., JIANG, W. & CAI, Q. 2016. Multi-scale temporal dynamics of epilithic algal assemblages: evidence from a Chinese subtropical mountain river network. *Hydrobiologia,* 770.

TAPOLCZAI, K., BOUCHEZ, A., STENGER-KOVÁCS, C., PADISÁK, J. & RIMET, F. 2016. Trait-based ecological classifications for benthic algae: review and perspectives. *Hydrobiologia,* 776**,** 1-17.

TAPOLCZAI, K., BOUCHEZ, A., STENGER-KOVÁCS, C., PADISÁK, J. & RIMET, F. 2017. Taxonomy- or trait-based ecological assessment for tropical rivers? Case study on benthic diatoms in Mayotte island (France, Indian Ocean). *Science of The Total Environment,* 607-608**,** 1293-1303.

TAPOLCZAI, K., KECK, F., BOUCHEZ, A., RIMET, F., KAHLERT, M. & VASSELON, V. 2019. Diatom DNA Metabarcoding for Biomonitoring: Strategies to Avoid Major Taxonomical and Bioinformatical Biases Limiting Molecular Indices Capacities. *Frontiers in Ecology and Evolution,* 7.

TAPOLCZAI, K., SELMECZY, G. B., SZABÓ, B., B-BÉRES, V., KECK, F., BOUCHEZ, A., RIMET, F. & PADISÁK, J. 2021. The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. *Ecological Indicators,* 122**,** 107322.

TEDERSOO, L., TOOMING-KLUNDERUD, A. & ANSLAN, S. 2018. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytologist,* 217**,** 1370-1385.

TRUCHY, A., SARREMEJANE, R., MUOTKA, T., MYKRÄ, H., ANGELER, D. G., LEHOSMAA, K., HUUSKO, A., JOHNSON, R. K., SPONSELLER, R. A. & MCKIE, B. G. 2020. Habitat patchiness, ecological connectivity and the uneven recovery of boreal stream ecosystems from an experimental drought. *Global Change Biology,* 26**,** 3455-3472.

UNITED STATES 1972. Clean Water Act (PL 92-500). *In:* 92E500, P. L. (ed.).

VASSELON, V., ÁCS, É., ALMEIDA, S., ANDREE, K., APOTHÉLOZ-PERRET-GENTIL, L., BAILET, B., BARICEVIC, A., BEENTJES, K., BETTIG, J.,

BOUCHEZ, A., CAPELLI, C., CHARDON, C., DULEBA, M., ELERSEK, T., GENTHON, C., HURTZ, M., JACAS, L., KAHLERT, M., KELLY, M., LEWIS, M., MACHER, J., MAURI, F., MOLETTA-DENAT, M., MORTÁGUA, A., PAWLOWSKI, J., PÉREZ BURILLO, J., PFANNKUCHEN, M., PILGRIM, E., PISSARIDOU, P., PORTER, J., RIMET, F., STANIC, K., TAPOLCZAI, K., THEROUX, S., TROBAJO, R., VAN DER HOORN, B., VASQUEZ HADJILYRA, M., WALSH, K., WANLESS, D., WARREN, J., ZIMMERMANN, J. & ZUPANČIČ, M. 2021. The Fellowship of the Ring Test: DNAqua-Net WG2 initiative to compare diatom metabarcoding protocols used in routine freshwater biomonitoring for standardisation. *ARPHA Conference Abstracts,* 4: e65142.

VASSELON, V., BOUCHEZ, A., RIMET, F., JACQUET, S., TROBAJO, R., CORNIQUEL, M., TAPOLCZAI, K. & DOMAIZON, I. 2018. Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. *Methods in Ecology and Evolution,* 9**,** 1060-1069.

VASSELON, V., RIMET, F., TAPOLCZAI, K. & BOUCHEZ, A. 2017. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecological Indicators,* 82**,** 1-12.

VISCO, J. A., APOTHÉLOZ-PERRET-GENTIL, L., CORDONIER, A., ESLING, P., PILLET, L. & PAWLOWSKI, J. 2015. Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. *Environmental Science & Technology,* 49**,** 7597-7605.

WELLS, E., WILKINSON, M., WOOD, P. & SCANLAN, C. 2007. The use of macroalgal species richness and composition on intertidal rocky seashores in the assessment of ecological quality under the European Water Framework Directive. *Marine pollution bulletin,* 55**,** 151-61.

WERNER, P., ADLER, S. & DREßLER, M. 2016. Effects of counting variances on water quality assessments: implications from four benthic diatom samples, each counted by 40 diatomists. *Journal of Applied Phycology,* 28.

WYNNE, D. & RHEE, G.-Y. 1986. Effects of light intensity and quality on the relative N and P requirement (the optimum N:P ratio) of marine planktonic algae. *Journal of Plankton Research,* 8**,** 91-103.

WYNNE, D. & RHEE, G. Y. 1988. Changes in alkaline phosphatase activity and phosphate uptake in P-limited phytoplankton, induced by light intensity and spectral quality. *Hydrobiologia,* 160**,** 173-178.

ZEVENBOOM, W., DE GROOT, G. J. & MUR, L. R. 1980. Effects of light on nitrate-limited Oscillatoria agardhii in chemostat cultures. *Archives of Microbiology,* 125**,** 59-65.

ZIMMERMANN, J., ABARCA, N., ENK, N., SKIBBE, O., KUSBER, W.-H. & JAHN, R. 2014. Taxonomic Reference Libraries for Environmental

Barcoding: A Best Practice Example from Diatom Research. *PLOS ONE,* 9**,** e108793.

ZIMMERMANN, J., GLÖCKNER, G., JAHN, R., ENKE, N. & GEMEINHOLZER, B. 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources,* 15**,** 526-542.

ZIMMERMANN, J., JAHN, R. & GEMEINHOLZER, B. 2011. Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution,* 11**,** 173.

ZINGER, L., BONIN, A., ALSOS, I. G., BÁLINT, M., BIK, H., BOYER, F., CHARITON, A. A., CREER, S., COISSAC, E., DEAGLE, B. E., DE BARBA, M., DICKIE, I. A., DUMBRELL, A. J., FICETOLA, G. F., FIERER, N., FUMAGALLI, L., GILBERT, M. T. P., JARMAN, S., JUMPPONEN, A., KAUSERUD, H., ORLANDO, L., PANSU, J., PAWLOWSKI, J., TEDERSOO, L., THOMSEN, P. F., WILLERSLEV, E. & TABERLET, P. 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology,* 28**,** 1857-1862.

8. Popular science summary

Water, especially freshwater, is our most precious resource. All organisms need water to survive, and it also provides multiple provisioning, supporting, regulatory, and cultural services. It is important to protect freshwater bodies like streams and lakes from degradation or to restore them to a good state when needed. Water bodies are also complex ecosystems, with a thriving biodiversity of macro-organisms, such as macrophytes, fish and invertebrates, and of microorganisms, such as bacteria and microalgae. A specific group of microalgae, called diatoms, have a silica skeleton called a 'frustule'. The form and structure of the frustule differ between diatom genera and species, so it can be used for diatom identification. Diatoms, like other photosynthetic algae, are primary producers, harvesting the energy from light and producing organic matter. They are also ubiquitous, so they can be found in almost any type of water, anywhere around the world. For all these reasons, they are excellent candidates as bioindicators. Because they are at the bottom of the food chain, they are the first affected if something in the environment changes. In fact, the Water Framework Directive, the European Union legislation governing monitoring and environmental assessment of European freshwaters, prescribes collection of diatom data for evaluating the ecological status of freshwater sites. Monitoring and environmental assessment of freshwater is important because the data collected make it possible to identify and evaluate the impact of environmental changes and avoid possible negative effects on ecosystems. Data on diatom community are generally collected by identifying about 400 frustules under a light microscope for each site, which is time-consuming and requires expertise on diatom taxonomy, so researchers have begun investigating easier, cheaper and faster alternatives. This thesis describes the development and application of two novel approaches to identify diatom communities: i) use of diatom DNA in a molecular approach called 'metabarcoding' and ii) use of measurable morphological and physiological

features or ‘traits’ of diatoms. The results are very encouraging because they show that the variations in diatoms DNA and in diatom traits both clearly reflect environmental changes, so they can be used for environmental assessment. Both methods also provide information on diatom communities that conventional microscopy cannot reveal. However, diatom DNA and trait data are more complex than microscopy data and cannot be used with the ecological tools (*e.g*. ecological indices) developed for microscopy data. This thesis presents a new index, calibrated to be used with diatom DNA and diatom traits, for estimating the levels of nutrients in natural freshwater samples. This index, which could be used to detect eutrophication (when water is too enriched in minerals and nutrients) performed well for stream and lake water samples on which it was tested, reliably estimating the level of total phosphorus in the water.

The novel molecular and trait approaches tested here are still under development and need further improvement. For example, the results from the molecular approach are affected by the method used for analysis of the DNA sequences using computer programs (bioinformatics). If two laboratories use different bioinformatics methods, they will obtain different results and will not be able to compare outcomes. Therefore, unless guidelines are introduced to standardise work across laboratories, it will be impossible to implement the molecular approach at regional, national or international scale, as required by the Water Framework Directive. The main problem with the trait approach is that it requires finding meaningful traits which show a response to the environmental variable being measured. Also any new index needs to be tested with different types of sample, and calibrated or corrected if needed. Future work is needed to identify more clearly the response of diatom DNA and traits, to improve the methodology of the two approaches, to make results more reproducible and to develop new ecological tools. Work is just getting started but, with the strong potential shown by the two novel approaches studied in this thesis, the future of environmental assessment of freshwater using diatoms looks very promising!

9. Populärvetenskaplig sammanfattning

Vatten, särskilt sötvatten, är vår mest värdefulla resurs. Alla organismer behöver vatten för att överleva, men förutom det bistår sötvattnet också med många tjänster, till exempel tillhandahållande av resurser (mat, dricksvatten, energi), stödtjänster (primärproduktion, näringscykling), regleringstjänster (klimatreglering) och kulturtjänster (estetisk, rekreation). Det är viktigt att skydda vattenförekomster, som bäckar och sjöar, från degradering eller att återställa dem till ett gott skick när det behövs. Vattenförekomster är också komplexa ekosystem, med en blomstrande biologisk mångfald av makroorganismer, som makrofyter, fisk och ryggradslösa djur och mikroorganismer, som bakterier och små alger. En specifik grupp av dessa mikroskopiska alger kallas kiselalger, och det speciella med dem är att de har ett kiseldioxidskelett som kallas "frustula". Kiselalger, liksom andra fotosyntetiska alger, är primärproducenter som får energi från ljus och som producerar organiskt material. De är också allestädes närvarande, så de kan hittas i nästan alla typer av vatten, var som helst i världen. Allt det här gör dem till utmärkta kandidater som bioindikatorer: eftersom de är längst ner i näringskedjan kommer de att bli de första som påverkas om något i miljön förändras och på så vis kan vi spåra förändringar i kiselalgerna för att upptäcka förändringar i miljö. Faktum är att vattendirektivet, lagstiftningen som övervakar miljöövervakning och bedömning av europeiska sötvatten, innehåller insamling av kiselalgsdata för utvärdering av ekologisk status för en plats. Miljöövervakningen och bedömningen av sötvatten är mycket viktig eftersom den tillåter oss att med hjälp av insamlade data identifiera och utvärdera effekterna av miljöförändringar och undvika eventuella negativa effekter på ekosystemet och de tjänster som tillhandahålls. Insamling av data om kiselalgssamhällen sker dock traditionellt genom att identifiera cirka 400 individer under ett ljusmikroskop för varje plats. Detta är tidskrävande och kräver expertis om kiselalgstaxonomi. På grund av det här har studier börjat titta på möjliga enklare, billigare och snabbare

alternativ. I denna avhandling fokuserar vi på utvecklingen och tillämpningen av två nya tillvägagångssätt för att identifiera kiselalgssamhällena: Användningen av kiselalgs-DNA via ett molekylärt tillvägagångssätt som kallas 'metabarkodning' och användningen av kiselalgers mätbara morfologiska egenskaper, kallade 'traits'. Våra resultat är mycket uppmuntrande eftersom de visar att både kiselalgs-DNA och kiselalgernas traits visar en tydlig respons på miljöförändringar, vilket gör att vi kan använda dem för miljöbedömning. Dessutom ger de information om kiselalgerna som den traditionella mikroskopimetoden inte förmår. Kiselalgers DNA och traits är dock mer komplexa än mikroskopidata och kan inte användas med de ekologiska verktygen (till exempel ekologiska index) som utvecklats för mikroskopi. Så vi började utveckla nya index, kalibrerade för att användas med kiselalgs-DNA och kiselalgernas traits. Mer specifikt utvecklade vi ett index som gör att vi kan uppskatta näringsnivåerna i vattnet, så det kan användas för att upptäcka övergödning när vattnet är för berikat med mineraler och näringsämnen. De index vi utvecklade med hjälp av kiselalgernas DNA och traits fungerade bra i de strömmande vatten och sjöar där vi testade det och uppskattade tillförlitligt halterna av totalt fosfor i vattnet. Men för närvarande är båda de nya metoderna fortfarande under utveckling och behöver ytterligare metodförbättringar. Resultaten från det molekylära tillvägagångssättet, till exempel, förändras mycket av metoden som används för bioinformatikdelen, analysen av biologiska data som DNA-sekvenseras med hjälp av datorprogram. Så om två laboratorier använder olika bioinformatiska metoder kommer de att få olika resultat och resultaten kommer inte att kunna jämföras. Detta innebär att om vi inte tillhandahåller riktlinjer för att standardisera metoden i laboratorierna, kommer det att vara omöjligt att implementera den molekylära metoden på regional, nationell eller internationell nivå som de andra metoderna som ingår i vattendirektivet. Det finns liknande problem för tekniken med de traits eftersom du till exempel måste hitta vilka egenskaper som visar en respons på miljövariabeln du mäter. Ett nytt index måste också testas i olika typer av prover, för att kalibreras och korrigeras vid behov. I framtiden behövs ytterligare arbete för att bättre förstå responsen från kiselalgernas DNA och traits, förbättra metodiken för att ge mer reproducerbara resultat och utveckla nya ekologiska verktyg. Vi har precis börjat, men med den starka potential som de två nya tillvägagångssätten visar ser framtiden för miljöbedömning av sötvatten med hjälp av kiselalger ser lovande ut!

10. Résumé pour public non-expert

L'eau, en particulier l'eau douce, est notre ressource la plus précieuse. Non seulement tous les organismes vivants ont besoin d'eau pour survivre, mais elle nous fournit également de nombreux services, d'approvisionnement, de soutien, de régulation et culturels. Il est important de protéger les cours d'eau et les lacs de la dégradation ou de les restaurer à un bon état s'ils ont été détériorés. Les espaces aquatiques sont également des écosystèmes complexes, avec une biodiversité florissante de macro-organismes, comme les macrophytes, les poissons et les invertébrés, et de micro-organismes, comme les bactéries et les petites algues. Un groupe spécifique de ces algues microscopiques est appelé « diatomées », qui ont la particularité d'avoir un squelette de silice appelé « frustule ». Les diatomées, comme les autres algues photosynthétiques, sont des producteurs primaires, récoltant l'énergie de la lumière et produisant de la matière organique. Elles sont également omniprésentes, de sorte qu'elles peuvent être trouvées dans presque tous les types d'eau, partout dans le monde. Pour toutes ces raisons, ce sont d'excellentes candidates comme bioindicateurs : parce qu'elles sont au bas de la chaîne alimentaire, elles seront les premières touchées si quelque chose dans l'environnement change. En effet, la directive-cadre sur l'eau - la législation qui encadre la surveillance et l'évaluation environnementale des eaux douces européennes- inclut la collecte de données sur les diatomées pour évaluer l'état écologique d'un site aquatique. Le suivi et l'évaluation environnementale de l'eau douce sont très importants car ils permettent, à partir des données collectées, d'identifier et d'évaluer l'impact des changements environnementaux et d'éviter d'éventuels effets négatifs sur l'écosystème et les services qu'il fournit. Cependant, la collecte de données sur la communauté des diatomées se fait traditionnellement en identifiant environ 400 d'entre elles au microscope optique pour chaque site. Cela prend du temps et nécessite une expertise sur la taxonomie des diatomées. C'est pourquoi de récentes études se sont tournées vers des alternatives

potentielles, moins chères, plus faciles, et plus rapides. Dans cette thèse, je présente le développement et l'application de deux nouvelles approches pour identifier les communautés de diatomées : l'utilisation de l'ADN de diatomées via une approche moléculaire appelée « metabarcoding » et l'utilisation de caractéristiques morphologiques et physiologiques mesurables des diatomées, appelées « traits ». Nos résultats sont très encourageants car ils montrent que l'ADN et les traits des diatomées offrent une réponse claire aux changements environnementaux, et nous pouvons donc les utiliser pour l'évaluation environnementale. De plus, ils fournissent des informations sur les communautés de diatomées que l'approche microscopique traditionnelle ne capture pas. Cependant, les données d'ADN et de traits de diatomées sont plus complexes que les données de microscopie et ne peuvent pas être utilisées avec les outils écologiques (comme les indices écologiques par exemple) développés pour la microscopie. Nous avons donc commencé à développer de nouveaux indices, calibrés pour être utilisés avec l'ADN et les traits des diatomées. Plus précisément, nous avons développé un indice qui nous permet d'estimer les niveaux de nutriments dans l'eau, et pourra donc être utilisé détecter l'eutrophisation, lorsque l'eau est trop enrichie en minéraux et nutriments. Les indices que nous avons développés en utilisant l'ADN et les traits des diatomées ont donné de bons résultats dans les cours d'eau et les lacs où nous les avons testés, estimant de manière fiable les niveaux de phosphore total dans l'eau. Actuellement, les deux nouvelles approches sont toujours en cours de développement et doivent être améliorées dans leur méthodologie. Les résultats de l'approche moléculaire, par exemple, changent beaucoup selon la méthode utilisée pour la partie bioinformatique, c'est à dire l'analyse de données biologiques comme les séquences d'ADN à l'aide de programmes informatiques. Ainsi, si deux laboratoires utilisent des méthodes bioinformatiques différentes, ils obtiendront des résultats différents et ne pourront pas les comparer. Cela signifie qu'à moins que nous ne fournissions des recommandations pour standardiser la méthode entre les laboratoires, il sera impossible d'appliquer l'approche moléculaire à l'échelle régionale, nationale ou internationale, comme les autres méthodes incluses dans la directive-cadre sur l'eau. Il y a des problèmes similaires avec l'approche des traits, par exemple, parce qu'il faut trouver quels traits montrent une réponse à la variable environnementale qui est mesurée. Et un nouvel indice doit également être testé dans différents types de sites, pour le calibrer et le corriger si nécessaire. À l'avenir, des

travaux supplémentaires sont nécessaires pour mieux comprendre la réponse de l'ADN et des traits des diatomées, pour améliorer la méthodologie afin d'obtenir des résultats plus reproductibles et pour développer de nouveaux outils écologiques. Nous ne faisons que commencer, mais avec le fort potentiel montré par les deux nouvelles approches, l'avenir de l'évaluation environnementale de l'eau douce à l'aide de diatomées s'annonce prometteur!

11. Acknowledgements

This PhD was a four and a half year adventure and would not have been the same without the support of many wonderful people.

I am deeply grateful to **Maria Kahlert**, first because you gave me the opportunity to start this amazing journey with you. You were an amazing supervisor all along, and I would not have managed without your kindness, your guidance and your patience. I am so grateful for everything you taught me, about diatoms, about research and about life. I also thank my worldwide co-supervisor team for their support and excellent inputs: To **Brendan McKie**, for stepping in at every key moment of this PhD and for his help with SLU student life. If I ever make it to "Doctor", I will need to steal a blue police box, don't you think?; to **Agnes Bouchez** and **Alain Franc** for hosting me in their labs, taking care of me and teaching me how to work the wonders of diatom DNA – and for the French touch!; and to **Susanne Schneider**, for her encouragement and inputs throughout this journey, and especially in the writing process. I hope to finally meet you in real life one day!

I would like to thank all my co-authors: **Ana Baricevic**, **Kalman Tapolczai**, **Teofana Chonova**, **Demetrio Mora**, **Sebastian Proft**, **Mathieu Ramon**, **Frederic Rimet**, **Jean-Marc Frigerio**, **Satu-Maria Karjalainen**, **Timo Muotka**, **Jussi Jyväsjärvi**, **Janne Soininen**, **Ari Huusko** and **Anette Teittinen** for their help in bringing every paper to life. To my co-authors **Martin Pfannkuchen** and **Jonas Zimmermann**, thank you for hosting me in your institute for a short-term scientific mission. Your passion for your work greatly motivated me! A special thank you to **Francois Keck** for making my R scripts work, and to **Valentin Vasselon** for letting me follow him around at conferences and giving me confidence in my own work.

I also want to express my gratitude to the Department's researchers and employees, especially **Stina Drakare,** my section head, **Willem Goedkoop**

and **Lars Sonesten**, two great heads of department, **Richard Johnson**, **Ulf Gradin**, **Simon Hallstan**, **Moritz Buck, Prune Leroy**, **Marcus Korvela**, **Erika Kjellén** and **Birgitta Olandersson, Herman Paz** and **Ronald Rodríguez**, as well as **Domenico Simone** from SLUBI. And, of course, my gratitude to the algae laboratory research engineers too: **Isabel Quintana**, **Eva Herlitz**, **Anders Stehn, Magda-Lena Wiklund-Mckie** and **Putte Olsson,** for their help and kindness.

This adventure would not have been complete without amazing friends: all my love and thanks to **Elodie**, because she is simply the best; to **Nadine** and **Sara**, because they are incredible; to **Jasmina**, **Jenny** and **Frank**, for all the fun and good memories; to **Ande** and her family, for opening their house and their hearts to me. To semi-goddess **Amélie** and her family, for helping me survive in Sweden, in many ways.

Thanks to all current PhD students and post-docs: **Alberto**, **Björn**, **Božo**, **Claudia**, **Carlotta**, **Didier**, **Emma**, **Hannah**, **James**, **Kong**, **Malin**, **Oksana**, **Pilar**, **Rikard**, **Sanne**, **Sara**, **Sophia**, **Valentina** (most recently!), **Vera**, **Wiebke** and **Winnie**, because you are awesome and the best bunch one could ask to work with. And also, many thanks to past young researchers **Alex**, **Jacob**, **Joachim**, **José**, **Kristina**, **Lisa**, **Maidul**, **Maria B.**, **Matthias**, **Pablo** and **Pianpian**, because I missed you from the moment you left. So many "Nice to meet you" and so many "I'll miss you".

Finally, my deepest thanks to **my wonderful family**, Papa, Maman, Bradly, Laurent, Sophie, Jules and Mireille, for their support and love during the past five years.

No thanks to Covid-19 for making everything more complicated and ruining all the fun!

Appendix

Supplementary material 1. List of samples used in **Papers I, II and IV**. Samples were collected across Fennoscandia and Iceland. Sample ID, sampling location, country of origin, type of waterbody -River or Lake-, sampling month and sampling year are provided. Samples selected for the dataset used in **Paper II** are indicated in bold.

Sample ID	Sample location	Country	Waterbody	Sampling month	Sampling year
DAN 4.1	Atnasjøen	Norway	River	6	2015
DAN 4.2	Atnasjøen	Norway	River	9	2016
VIK 11.1	Utløp Fjellgardsvatn	Norway	River	6	2016
VIK 11.2	Utløp Fjellgardsvatn	Norway	River	9	2016
VIK 12.1	**Bekk fra Røyravatnet**	**Norway**	**River**	**6**	**2016**
VIK 12.2	Bekk fra Røyravatnet	Norway	River	9	2016
MR51	**Sagelva**	**Norway**	**River**	**9**	**2015**
Ostra st1	upstream WTP Valle	Norway	River	8	2016
Ostra st2	downstream WTP Valle	Norway	River	8	2016
Ostra st3	uppstream WTP Rysstad	Norway	River	8	2016
Ostra st4	downstream WTP Rysstad	Norway	River	8	2016
HOF 1	**Oslo**	**Norway**	**River**	**9**	**2016**

HOF 2	Oslo	Norway	River	9	2016
HOF3	Oslo	Norway	River	9	2016
FRO 1	Oslo	Norway	River	9	2016
FRO 3	**Oslo**	**Norway**	**River**	**9**	**2016**
OST1	Oslo	Norway	River	9	2016
ALN3	Oslo	Norway	River	9	2016
BRE	Oslo	Norway	River	9	2016
HOV1	Oslo	Norway	River	9	2016
ML1	**Njalakjaure**	**Sweden**	**Lake**	**x**	**2009**
ML10	Hällsjön	Sweden	Lake	8	2009
ML11	Sidensjön	Sweden	Lake	9	2009
ML12	Remmarsjön	Sweden	Lake	9	2007
ML13	Saxen	Sweden	Lake	9	2010
ML14	Täftesträsket	Sweden	Lake	9	2009
ML15	Allgjuttern	Sweden	Lake	9	2007
ML16	Brunnsjön	Sweden	Lake	9	2007
ML17	**Kyrksjön Hölö**	**Sweden**	**Lake**	**9**	**2009**
ML18	Ullnasjön	Sweden	Lake	9	2009

ML19	Valloxen	Sweden	Lake	10	2008
ML2	Tronnttjärnarna	Sweden	Lake	8	2009
ML20	**Oxundasjön**	**Sweden**	**Lake**	**9**	**2009**
ML21	Funbosjön	Sweden	Lake	10	2008
ML22	Ulvsjön	Sweden	Lake	8	2009
ML23	Ekoln	Sweden	Lake	x	2008
ML24	Immeln	Sweden	Lake	7	2007
ML25	Bäen	Sweden	Lake	9	2009
ML26	Östra Ringsjön	Sweden	Lake	9	2009
ML27	**Börringesjön**	**Sweden**	**Lake**	**9**	**2009**
ML28	Gärsjön	Sweden	Lake	9	2007
ML29	Härsvatten	Sweden	Lake	9	2007
ML3	**Vuolejaure**	**Sweden**	**Lake**	**8**	**2009**
ML30	Gyltigesjön	Sweden	Lake	x	2006
ML31	Fiolen	Sweden	Lake	x	2006
ML32	Älgarydssjön	Sweden	Lake	x	2006
ML34	Gyslättasjön	Sweden	Lake	x	2007
ML35	**Stensjön**	**Sweden**	**Lake**	**9**	**2007**

ML36	MVM-lake	Sweden	Pond	11	2017
ML4	Abiskojaure	Sweden	Lake	9	2007
ML5	Örvattnet	Sweden	Lake	10	2007
ML6	Pahajärvi	Sweden	Lake	8	2009
ML7	Stor-Backsjön	Sweden	Lake	8	2009
ML8	Gipsjön	Sweden	Lake	8	2009
ML9	Stor-Tjulträsket	Sweden	Lake	9	2009
MR1	Pipbäcken Nedre	Sweden	River	9	2013
MR10	Lekarån	Sweden	River	9	2013
MR11	Färgeån	Sweden	River	9	2013
MR12	Lindåsabäcken	Sweden	River	9	2013
MR13	Vemmenhög (M42)	Sweden	River	9	2011
MR14	Björnbackån	Sweden	River	10	2013
MR15	Norrhultsbäcken	Sweden	River	9	2013
MR16	Kolarebäcken	Sweden	River	9	2013
MR17	Skärån, Skäralid	Sweden	River	9	2013
MR18	Pipbäcken Nedre	Sweden	River	9	2014
MR19	Ö. Anråsälven	Sweden	River	9	2013

MR2	Ö. Anråsälven	Sweden	River	9	2014
MR20	Alep Uttjajåkkå	Sweden	River	9	2013
MR21	Kukkasjärvi	Sweden	River	9	2013
MR22	Akkarjåkkå	Sweden	River	9	2013
MR23	**Abiskojokk Röda Bron**	**Sweden**	**River**	**11**	**2013**
MR24	Lommabäcken Nedre	Sweden	River	9	2013
MR25	Tolångaån Tolånga	Sweden	River	9	2013
MR26	Loån	Sweden	River	10	2013
MR27	Lekarån	Sweden	River	9	2014
MR28	Färgeån	Sweden	River	9	2014
MR29	Lindåsabäcken	Sweden	River	9	2014
MR3	Alep Uttjajåkkå	Sweden	River	9	2014
MR30	Vemmenhög (M42)	Sweden	River	9	2014
MR31	Björnbackån	Sweden	River	9	2014
MR32	Norrhultsbäcken	Sweden	River	9	2014
MR33	Kolarebäcken	Sweden	River	9	2014
MR34	Skärån, Skäralid	Sweden	River	9	2014
MR35	**Pipbäcken Nedre**	**Sweden**	**River**	**9**	**2015**

MR36	Ö. Anråsälven	Sweden	River	9	2015
MR37	Kukkasjärvi	Sweden	River	9	2014
MR38	Akkarjåkkå	Sweden	River	9	2014
MR39	Abiskojokk Röda Bron	Sweden	River	9	2014
MR4	Kukkasjärvi	Sweden	River	9	2015
MR40	**Lommabäcken Nedre**	**Sweden**	**River**	**9**	**2014**
MR41	Tolångaån Tolånga	Sweden	River	9	2014
MR42	**Loån**	**Sweden**	**River**	**9**	**2014**
MR43	**Lekarån**	**Sweden**	**River**	**9**	**2015**
MR44	Färgeån	Sweden	River	9	2015
MR45	Lindåsabäcken	Sweden	River	9	2015
MR46	Vemmenhög (M42)	Sweden	River	9	2015
MR47	Björnbackån	Sweden	River	9	2015
MR48	Norrhultsbäcken	Sweden	River	9	2015
MR49	Kolarebäcken	Sweden	River	9	2015
MR5	**Akkarjåkkå**	**Sweden**	**River**	**9**	**2015**
MR50	Skärån, Skäralid	Sweden	River	9	2015
MR52	**Vemmenhög (M42)**	**Sweden**	**River**	**9**	**2007**

MR53	Vemmenhög (M42)	Sweden	River	9	2008
MR6	Abiskojokk Röda Bron	Sweden	River	9	2015
MR7	Lommabäcken Nedre	Sweden	River	9	2015
MR8	**Tolångaån Tolånga**	**Sweden**	**River**	**9**	**2015**
MR9	Loån	Sweden	River	9	2015
F101	**Hiidenvesi**	**Finland**	**Lake**	**10**	**2011**
F102	Hiidenvesi	Finland	Lake	10	2011
F103	Hiidenvesi	Finland	Lake	10	2011
F104	Iso Riihijärvi	Finland	Lake	9	2011
F105	**Iso Riihijärvi**	**Finland**	**Lake**	**11**	**2011**
F106	Iso Riihijärvi	Finland	Lake	9	2011
F107	Kivijärvi pohjoisosa	Finland	Lake	10	2011
F108	Kivijärvi pohjoisosa	Finland	Lake	10	2011
F109	Kivijärvi pohjoisosa	Finland	Lake	10	2011
F110	**Komujärvi**	**Finland**	**Lake**	**9**	**2011**
F111	Komujärvi	Finland	Lake	9	2011
F112	Komujärvi	Finland	Lake	9	2011
F113	Pasmajärvi	Finland	Lake	9	2011

F114	Pasmajärvi	Finland	Lake	9	2011
F115	**Pasmajärvi**	**Finland**	**Lake**	**9**	**2011**
F116	Pusulanjärvi eli Jäämäjärvi	Finland	Lake	9	2011
F117	**Pusulanjärvi eli Jäämäjärvi**	**Finland**	**Lake**	**9**	**2011**
F118	Pusulanjärvi eli Jäämäjärvi	Finland	Lake	9	2011
F119	Siika-Kämä	Finland	Lake	10	2011
F120	Siika-Kämä	Finland	Lake	10	2011
F121	Siika-Kämä	Finland	Lake	10	2011
F1	Aittojoki	Finland	River	11	2011
F10	Pohjajoki	Finland	River	9	2011
F11	Pohjajoki	Finland	River	10	2012
F12	Pohjajoki	Finland	River	9	2013
F13	Punkalaitumenjoki, Teikarla	Finland	River	11	2011
F14	**Punkalaitumenjoki, Teikarla**	**Finland**	**River**	**10**	**2012**
F15	Punkalaitumenjoki, Teikarla	Finland	River	9	2013
F16	Sikkilänjoki, Vähä-Jakama	Finland	River	11	2011
F17	Sikkilänjoki, Vähä-Jakama	Finland	River	10	2012

F18	Sikkilänjoki, Vähä-Jakama	Finland	River	9	2013
F19	Tarpianjoki 7	Finland	River	10	2011
F2	Aittojoki	Finland	River	11	2012
F20	Tarpianjoki 8	Finland	River	10	2012
F21	Tarpianjoki 9	Finland	River	9	2013
F22	Korpijoki 2	Finland	River	9	2013
F23	Kuohattijoki	Finland	River	9	2013
F24	Lanskinjoki, Ylä-Myllykoski	Finland	River	9	2013
F25	Luohuanjoki, Mikkola	Finland	River	9	2013
F26	Nuottipuro	Finland	River	10	2013
F27	Pusulanjoki, Ankelistonkoski	Finland	River	9	2013
F28	Koskenjoki	Finland	River	9	2013
F29	**Sipoonjoki, Brobölenkoski**	**Finland**	**River**	**9**	**2013**
F3	Aittojoki	Finland	River	10	2013
F30	**Taasianjoki, Holmankoski**	**Finland**	**River**	**8**	**2013**
F31	Teutjoki, Junttilankoski	Finland	River	9	2011
F32	Torasjoki, Raunukoski	Finland	River	9	2011

F33	**Vieresjoki, Repuli**	**Finland**	**River**	**9**	**2011**
F34	Onkamaanjoki RI	Finland	River	9	2013
F35	**Vuotosjoki**	**Finland**	**River**	**9**	**2011**
F36	Yläneenjoki P2, Vanhakartano	Finland	River	11	2011
F37	Maalahdenjoki, Kyrkbacken	Finland	River	9	2013
F38	Kruununpyynjoki, mylly	Finland	River	9	2013
F39	Lestijoki, Kallisenkoski	Finland	River	9	2013
F4	Luostanjoki 1	Finland	River	9	2011
F40	Murronjoki	Finland	River	11	2013
F41	Iso-Tainijoki 1	Finland	River	9	2011
F42	Taipaleenjoki, Siikakoski	Finland	River	9	2011
F43	**Haapajoki 32**	**Finland**	**River**	**9**	**2011**
F44	Malisjoki, Nivalankoski	Finland	River	9	2013
F45	Tyrnävänjoki	Finland	River	11	2011
F46	Kuorejoki 3	Finland	River	10	2013
F47	Vilajoki, Lohonkoski	Finland	River	9	2013
F48	Saunajoki	Finland	River	11	2013
F49	Mäntyjoki, Saunakoski	Finland	River	11	2011

F5	Luostanjoki 2	Finland	River	9	2012
F50	Rauanjoki 2	Finland	River	10	2011
F51	Nummenjoki, Lukkarinkoski	Finland	River	9	2013
F6	Luostanjoki 3	Finland	River	10	2013
F7	Muhosjoki, Mustakoski	Finland	River	8	2011
F8	Muhosjoki, Mustakoski	Finland	River	9	2012
F9	Muhosjoki, Mustakoski	Finland	River	9	2013
IS1	Ulfarsa	Iceland	River	x	2015
IS2	Hengill	Iceland	River	x	2015

Supplementary material 2. List of samples used in Paper III. The sample were collected at the Kainuu Fisheries Research Station (Paltamo, northern Finland) as part of the Destress project (see Material and Methods). Sample ID, Sampling channel, sampling sub-channel and treatment applied (Nutrient: Enriched or Ambient, Light exposure: Shade or Light, Flow disturbance: Press, Pulse or Control) are provided.

Sample ID	Channel	Sub-channel	Nutrients	Light Exposure	Flow disturbance
111	1	1	Enriched	Shade	Control
112	1	1	Enriched	Shade	Control
113	1	1	Enriched	Shade	Control
114	1	1	Enriched	Shade	Control
121	1	2	Ambient	Light	Control
122	1	2	Ambient	Light	Control
123	1	2	Ambient	Light	Control
124	1	2	Ambient	Light	Control
131	1	3	Enriched	Light	Control
132	1	3	Enriched	Light	Control
133	1	3	Enriched	Light	Control
134	1	3	Enriched	Light	Control
141	1	4	Ambient	Shade	Control
142	1	4	Ambient	Shade	Control
143	1	4	Ambient	Shade	Control
144	1	4	Ambient	Shade	Control
211	2	1	Enriched	Light	Pulse
212	2	1	Enriched	Light	Pulse
213	2	1	Enriched	Light	Pulse
214	2	1	Enriched	Light	Pulse
221	2	2	Enriched	Shade	Pulse
222	2	2	Enriched	Shade	Pulse
223	2	2	Enriched	Shade	Pulse
224	2	2	Enriched	Shade	Pulse
231	2	3	Ambient	Light	Pulse
232	2	3	Ambient	Light	Pulse

233	2	3	Ambient	Light	Pulse
234	2	3	Ambient	Light	Pulse
241	2	4	Ambient	Shade	Pulse
242	2	4	Ambient	Shade	Pulse
243	2	4	Ambient	Shade	Pulse
244	2	4	Ambient	Shade	Pulse
311	3	1	Enriched	Shade	Pulse
312	3	1	Enriched	Shade	Pulse
313	3	1	Enriched	Shade	Pulse
314	3	1	Enriched	Shade	Pulse
321	3	2	Ambient	Shade	Pulse
322	3	2	Ambient	Shade	Pulse
323	3	2	Ambient	Shade	Pulse
324	3	2	Ambient	Shade	Pulse
331	3	3	Enriched	Light	Pulse
332	3	3	Enriched	Light	Pulse
333	3	3	Enriched	Light	Pulse
334	3	3	Enriched	Light	Pulse
341	3	4	Ambient	Light	Pulse
342	3	4	Ambient	Light	Pulse
343	3	4	Ambient	Light	Pulse
344	3	4	Ambient	Light	Pulse
411	4	1	Enriched	Shade	Press
412	4	1	Enriched	Shade	Press
413	4	1	Enriched	Shade	Press
414	4	1	Enriched	Shade	Press
421	4	2	Ambient	Shade	Press
422	4	2	Ambient	Shade	Press
423	4	2	Ambient	Shade	Press
424	4	2	Ambient	Shade	Press
431	4	3	Ambient	Light	Press

432	4	3	Ambient	Light	Press
433	4	3	Ambient	Light	Press
434	4	3	Ambient	Light	Press
441	4	4	Enriched	Light	Press
442	4	4	Enriched	Light	Press
443	4	4	Enriched	Light	Press
444	4	4	Enriched	Light	Press
511	5	1	Enriched	Light	Control
512	5	1	Enriched	Light	Control
513	5	1	Enriched	Light	Control
514	5	1	Enriched	Light	Control
521	5	2	Enriched	Shade	Control
522	5	2	Enriched	Shade	Control
523	5	2	Enriched	Shade	Control
524	5	2	Enriched	Shade	Control
531	5	3	Ambient	Light	Control
532	5	3	Ambient	Light	Control
533	5	3	Ambient	Light	Control
534	5	3	Ambient	Light	Control
541	5	4	Ambient	Shade	Control
542	5	4	Ambient	Shade	Control
543	5	4	Ambient	Shade	Control
544	5	4	Ambient	Shade	Control
611	6	1	Ambient	Light	Press
612	6	1	Ambient	Light	Press
613	6	1	Ambient	Light	Press
614	6	1	Ambient	Light	Press
621	6	2	Enriched	Shade	Press
622	6	2	Enriched	Shade	Press
623	6	2	Enriched	Shade	Press
624	6	2	Enriched	Shade	Press

631	6	3	Enriched	Light	Press
632	6	3	Enriched	Light	Press
633	6	3	Enriched	Light	Press
634	6	3	Enriched	Light	Press
641	6	4	Ambient	Shade	Press
642	6	4	Ambient	Shade	Press
643	6	4	Ambient	Shade	Press
644	6	4	Ambient	Shade	Press

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:55

Diatoms are useful bioindicators but their identification using light microscopy is time consuming and error prone. Two alternative approaches to characterise the diatom community were developed and applied in this thesis: a DNA metabarcoding approach and a trait-based approach. Both approaches showed potential for environmental assessment but cannot be used with tools developed for microscopy data. New nutrient indices were developed for each novel approach and performed well for a use in assessment of total phosphorus levels in Fennoscandia freshwater.

Bonnie Bailet received her graduate education at the Department of Aquatic Sciences and Assessment at the Swedish University of Agricultural Sciences in Uppsala. Her M.Sc. degree in Oceanography specialised in Marine Biology was obtained at Aix-Marseille University, Marseille, France.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: http://pub.epsilon.slu.se/

ISSN 1652-6880
ISBN (print version) 978-91-7760-788-5
ISBN (electronic version) 978-91-7760-789-2