MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 10: 689–718 (2026)
DOI: 10.3897/mbmg.10.168096

Review Article

DNA metabarcoding of diatoms: A next-generation tool for aquatic biodiversity and bioassessment

Bassam Abubaker[1,2], Csilla Stenger-Kovács[1,2,3], Frédéric Rimet[4], Kálmán Tapolczai[3,5]

1 *Center for Natural Sciences, Limnology Research Group, University of Pannonia, Veszprém, Hungary*
2 *HUN-REN-PE Limnoecology Research Group, Veszprém, Hungary*
3 *National Laboratory for Water Science and Water Security, Center for Natural Science, University of Pannonia, Veszprém, Hungary*
4 *UMR Carrtel, INRAE, Université Savoie Mont-Blanc, F-74200 Thonon les Bains, France*
5 *HUN-REN Balaton Limnological Research Institute, Tihany, Hungary*

Corresponding authors: Bassam Abubaker (bassam.abubaker.ba@gmail.com); Kálmán Tapolczai (tapolczai.kalman@blki.hu)

Abstract

DNA metabarcoding of diatoms is a next-generation biomonitoring tool, utilizing high-throughput sequencing technology for the identification and assessment of diatom biodiversity in aquatic ecosystems. Diatom DNA metabarcoding has been widely recognized as a powerful biomonitoring approach in comparison with the traditional morphological identification, which requires high expertise, is more costly and needs more time. This review provides a comprehensive overview of the methodological development, various applications, and challenges in this rapidly emerging field. A systematic search on the Web of Science database resulted in 128 papers directly related to diatom metabarcoding. Manual data extraction showed that elements of the metabarcoding workflow such as sampling, DNA extraction, PCR amplification, choice of barcode, high-throughput sequencing, and bioinformatic analyses have a direct influence on the results. This review highlights the influence of each workflow step on the reliability of the results.

Despite its considerable advantages, this novel approach faces persistent challenges, including incomplete reference libraries, abundance quantification, and the diversity of bioinformatics analyses that can influence the outcomes. Nevertheless, diatom DNA metabarcoding rapidly developed into an efficient tool, fundamentally transforming our understanding of aquatic ecosystems and enhancing global biomonitoring capabilities.

Key words: Diatom, environmental DNA, High-Throughput Sequencing (HTS), metabarcoding

Academic editor: Dirk Steinke
Received: 8 August 2025
Accepted: 19 May 2026
Published: 16 June 2026

Citation: Abubaker B, Stenger-Kovács C, Rimet F, Tapolczai K (2026) DNA metabarcoding of diatoms: A next-generation tool for aquatic biodiversity and bioassessment. Metabarcoding and Metagenomics 10: e168096. https://doi.org/10.3897/mbmg.10.168096

Introduction

DNA metabarcoding represents a new milestone in biological and ecological studies by identifying organisms based on their genetic material. Identifying organisms through DNA barcoding is an innovative approach supported by frequent technological advances (Taberlet et al. 2012, 2018). DNA metabarcoding can be applied to both environmental DNA (eDNA) and

bulk biological samples, such as cultures. It has become one of the most attractive approaches for ecological scientists, supported by an increasing number of comprehensive best practices guides and continuously decreasing costs especially for DNA sequencing (Zinger et al. 2019). The comprehensive picture of the entire genetic material can provide information not only on the community composition but on ecosystem functions and interactions as well (Handelsman et al. 1998).

Diatoms are a group of microscopic algae widespread across aquatic and even terrestrial ecosystems. They are considered the most diverse group of algae in benthic and pelagic habitats of both marine and freshwater habitats (Malviya et al. 2016). The diversity of diatoms is estimated at 30,000 up to 100,000 species (Mann and Vanormelingen 2013). They play a significant role in natural ecosystems as primary producers (Smol and Stoermer 2010), and major contributors to ecosystem services including nutrient cycling, habitat provision, oxygen production, climate regulation, and sediment stabilization (B-Béres et al. 2023). Several studies used benthic diatoms for the biomonitoring of streams, rivers and lakes, showing that they are excellent indicators of ecological status (Stenger-Kovács et al. 2007; Smol and Stoermer 2010; Stephen and Joshi 2010; Rimet 2012; Rimet and Bouchez 2012; Almeida et al. 2014; Lavoie et al. 2014; Schulte et al. 2024). The sensitivity and indicator value of diatom species can be assessed by their specific ecological optimum and tolerance to changing environmental conditions (e.g., nutrient concentrations, pH, and conductivity). Therefore, species diversity and community composition can provide valuable insight into both recent and past ecological conditions (Desrosiers et al. 2013). As a result, diatoms have become key bioindicators for ecological status assessment of lakes and rivers within legislative water management plans, e.g. the Water Framework Directive (WFD) (EC Directive 2000).

Morphological analyses of diatoms date back to the early 18th century although real advancements occurred mostly during the 19th and 20th century following the improvement of microscopes and the expansion of ecological and taxonomic studies (Mann 2010). Today, routine identification still relies mainly on morphological characteristics of their siliceous exoskeleton. This process consumes a lot of time, funds and requires substantial taxonomic expertise, which can limit their use for biomonitoring (Pawlowski et al. 2016). Additionally, diatom taxonomy is constantly evolving due to rapid advances in microscopy and molecular genetic methods, leading to the description of new taxa as well as the reclassification or merging of previously recognized ones. The resulting taxa lists often suffer from differing levels of expertise, which makes comparisons between studies rather challenging (Kahlert et al. 2009; Kulaš et al. 2022).

Diatom DNA metabarcoding, an active research field, has gained significant attention (Kermarrec et al. 2013; Zimmermann et al. 2015) as it is an efficient method for the parallel identification of multiple environmental samples. Consequently, several studies investigated its potential for diatom identification (Moniz and Kaczmarska 2010; Keck et al. 2022) and showed that it is a powerful tool to examine unknown diatom diversity, to expand our knowledge about their distribution patterns (Vasselon et al. 2019) and to

detect rare species (Zhan et al. 2013). In addition, it can be a reliable tool for WFD assessment (Pérez-Burillo et al. 2020). Current diatom DNA research in European rivers and lakes (Pawlowski et al. 2018) is focusing on the development of new water quality assessments and aims to benefit from the high taxonomic resolution to scale up the bioassessment process (Lobo et al. 2016; Tapolczai et al. 2024). However, several challenges remain to be addressed including the incompleteness of reference libraries, PCR amplification biases, abundance estimation biases, and variability in bioinformatic pipelines, all of which may substantially influence study outcomes (Keck et al. 2017; Pawlowski et al. 2018).

The diatom DNA metabarcoding workflow includes the following seven steps (Fig. 1): (1) sampling and preservation, (2) DNA extraction, (3) PCR amplification, (4) library preparation (5) DNA sequencing, (6) bioinformatic analysis and (7) data analysis. Each step has been studied more or less intensively to assess its effect on the study's outcomes, resulting in the development of certain protocols.

This review provides an in-depth analysis and comprehensive overview of the development and applicability of DNA metabarcoding analyses in diatom research, comparing this emerging and promising method with the traditional microscopic analyses and revealing synergies between the methods. Furthermore, we present a standard methodological workflow for diatom DNA metabarcoding, illustrating the efforts to standardize the procedural steps and to validate the results. We also explore the current state of implementing DNA-based diatom analysis in practical applications alongside an evaluation of the limitations that challenge its broader application. Finally, the review discusses future perspectives, emphasizing the potential of this innovative approach in advancing biomonitoring and biodiversity assessment.

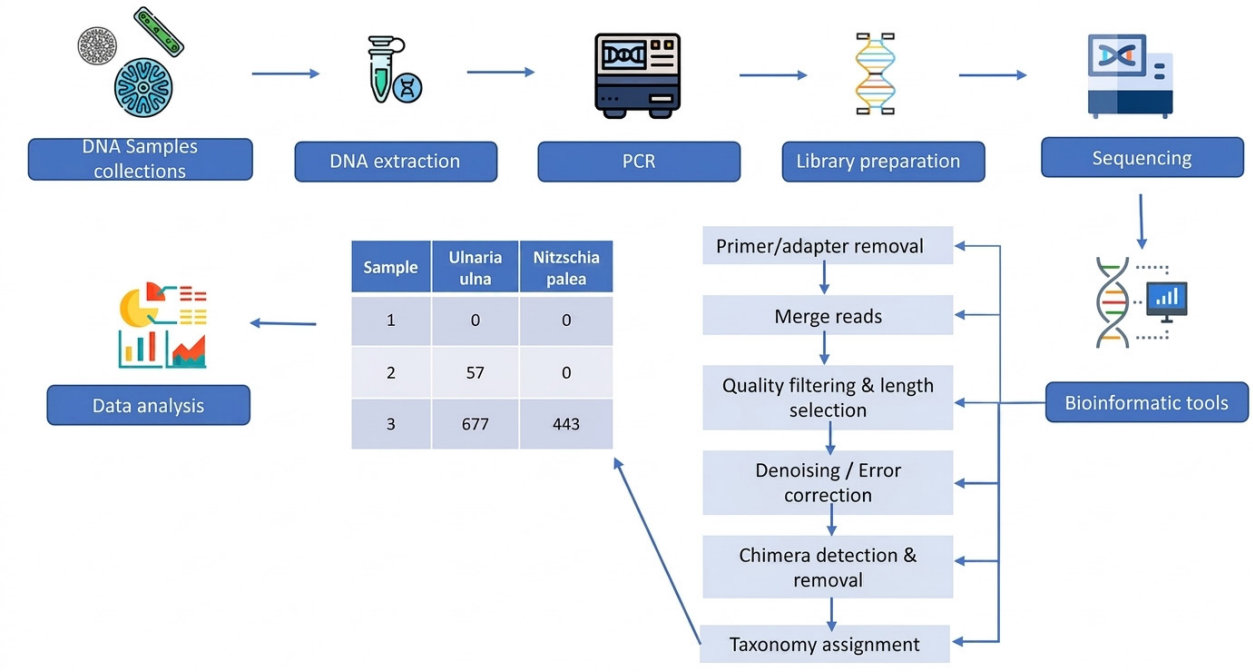

Figure 1. Diatom DNA Metabarcoding workflow

Methods

Systematic data collection was conducted involving more than 580 peer-reviewed scientific articles to identify relevant studies addressing diatom metabarcoding applications, insights and challenges. The search was performed using the Web of Science database accessed by using the following prompt: (TS=(metabarcoding) OR TS=(meta-barcoding) OR TS=(High throughput sequencing) OR TS=(High-throughput sequencing) OR TS=(next-generation sequencing) OR TS=(next generation sequencing)) AND (TS=(diatom) OR TS=(diatoms). A total of 582 papers were identified, and then additional criteria for the required studies were applied manually, including: (i) the study targeted diatoms (more than 350 papers were excluded due to targeting microalgae, bacteria, or protists in general), (ii) the articles were primary research papers, excluding reviews, opinion papers, and dataset presentations. As a result, 128 publications were selected for subsequent analysis. Data from each study were extracted into a standardized table (Suppl. material 1: table S1), including the following variables: publication year, study location (country), habitat, type of samples, method of DNA extraction, target gene region, sequencing technology, and bioinformatic pipelines (Fig. 2).

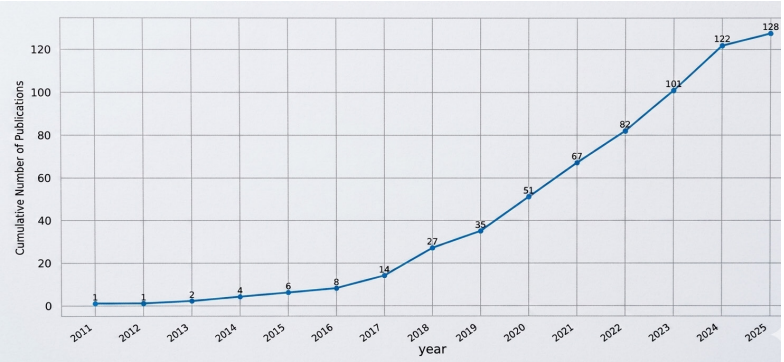

Figure 2. Cumulative number of publications on diatom DNA metabarcoding based on the Web of Science database (3 March 2025).

Our review focused on the following aspects of the analyzed studies:

i) Adoption and coverage:
The year, region (country) and habitat types the studies were employed in were registered to study possible geographical and temporal gaps in diatom metabarcoding studies and define potentially understudied environments in the literature.

ii) Field and lab practices
The substrates to sample, the preservation methods and additional wet-lab protocols were investigated regarding their effects on results in order to study the potential to develop standardized sampling and laboratory protocols.

iii) Genetic markers and reference sequences
The influence of genetic marker choice and the completeness of reference DNA sequence libraries on the taxonomic resolution and taxonomic assignment were studied.

iv) Bioinformatics pipelines
Various bioinformatics workflows are used for metabarcoding studies which can potentially affect the final diatom diversity metrics and community composition. We investigated the potential biases with a special focus on the choice of amplicon sequence variants (ASV) over operational taxonomic units (OTU) or the taxonomic assignment methods.

v) Detection performance
Comparisons of diatom metabarcoding and microscopy as two identification methods for diatoms were analyzed in terms of efficiency for successful taxa detection and diversity.

vi) Ecological signal
The efficiency of metabarcoding-based monitoring metrics (e.g. IPS, TDI indices of ecological quality status) was studied in terms of their ability to capture environmental gradients or in comparison to morphology-based indices as references.

vii) Applications
We overviewed the applied aspects of diatom DNA metabarcoding in different contexts such as ecological status assessments or forensic science.

viii) Limitations
Beside the potentials of diatom DNA metabarcoding, we also identified the limitations and challenges of the methodology. We assessed the factors that contribute to the reliability and challenges for adoption, including incomplete reference libraries, primer/marker biases, or the lack of standardization across other protocol steps.

ix) Future prospects and priorities
We outlined some perspectives addressing the development of measures to improve reference libraries, harmonize fieldwork to pipeline standards, and next-generation analytics.

Results

Global geographical distribution of diatom DNA studies

The 128 diatom DNA metabarcoding publications revealed that the investigated sampling sites are globally distributed, but heterogeneously with a focus on Europe, North America and East Asia (Fig. 3). France and China are among the most frequently sampled countries, followed by the USA, Spain, and several European countries including Hungary, Norway, Italy, Serbia and Switzerland. There are also contributions from countries such as Russia, South Korea and Japan. Some of the publications focused on less frequently represented regions including Africa (Algeria, South Africa), South America (Chile), and the Middle East (Iraq). This geographic distribution highlights both the global interest in DNA-based monitoring and the substantial geographic gaps that remain in large parts of the world, especially the huge difference in uptake between the global North and South.

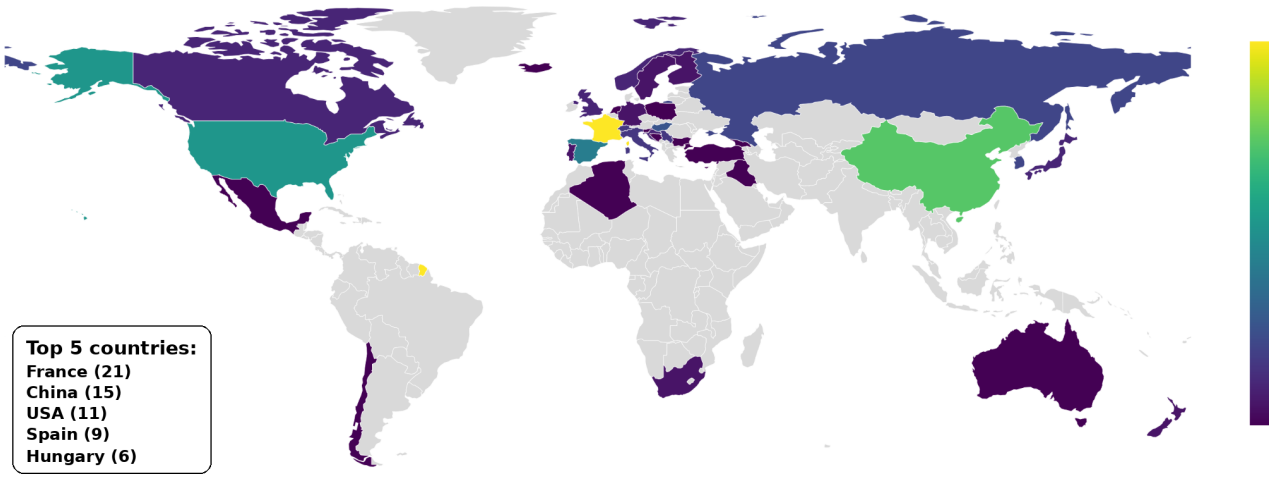

Figure 3. Global distribution of diatom DNA metabarcoding studies showing the number of publications.

Diatom DNA metabarcoding workflow

For this survey, we studied the methodological steps of the diatom DNA metabarcoding workflow. Fig. 4. shows the number of publications across water-body types (Fig. 4A), sample types (Fig. 4B), target gene regions/markers (Fig. 4C), and DNA extraction methods (Fig. 4D).

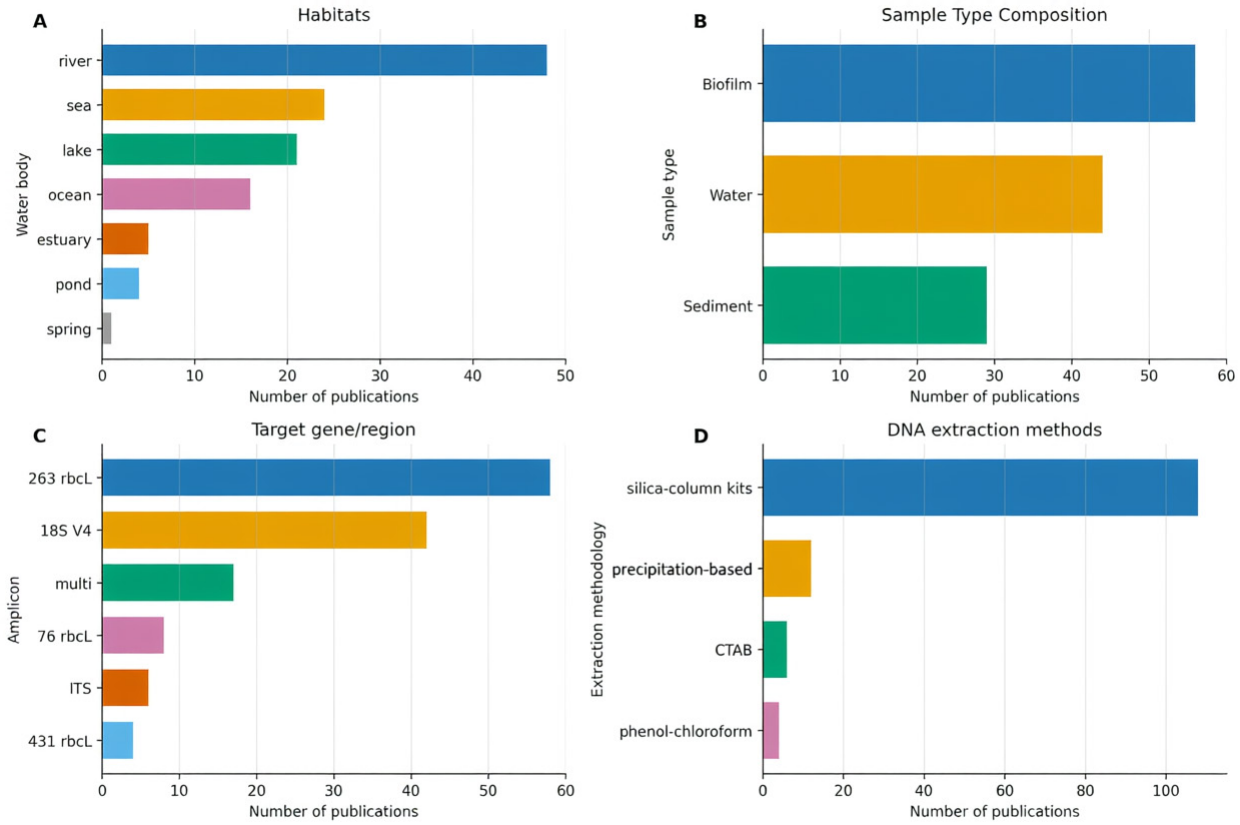

Figure 4. Overview of surveyed publications. **A** Habitats, **B** Sample types, **C** Target gene/region, and **D** DNA extraction methods. Counts are shown as horizontal bars and scaled to the number of publications per category.

In addition to DNA extraction and marker selection, we also looked at sequencing technologies. Bubble sizes in (Fig. 5) denote the number of studies per year using the sequencing platforms, enabling a comparative view of temporal uptake. Sequencing platforms transition from early pyrosequencing and occasional Ion Torrent (Ion semiconductor sequencing) and other Illumina platforms to a sustained dominance of Illumina MiSeq.

The distribution of the used pipelines in the reviewed studies (Fig. 6) shows that early studies frequently relied on customized workflows and legacy tools (e.g., mothur, OBITools), whereas from ~2018 onward, there is a progressive rise in DADA2 use, becoming the most used pipeline between 2020 and 2024, with smaller but consistent contributions from QIIME, v/usearch.

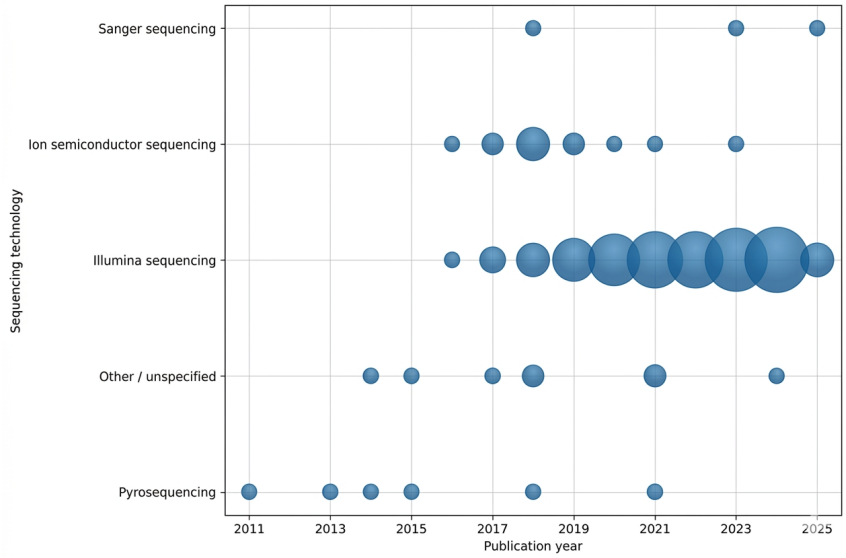

Figure 5. Sequencing platforms used in diatom DNA metabarcoding studies.

Figure 6. Bioinformatic pipelines used in diatom DNA metabarcoding studies.

Discussion

In this discussion, we synthesize the main findings of our review by examining the key methodological steps of diatom DNA metabarcoding and their influence on study outcomes. We focus on critical aspects including sample collection and preservation, DNA extraction, marker selection, sequencing technologies, and bioinformatic processing. Additionally, we evaluate the correspondence between metabarcoding and traditional morphological approaches, highlight current applications in biomonitoring and forensic science, and discuss the major limitations and future perspectives of this rapidly evolving field.

Sample collection and preservation

Sample collection and preservation represent critical initial steps in diatom DNA metabarcoding, where many elements of traditional sampling protocols can be retained, while specific modifications are required to ensure DNA integrity and prevent contamination.

We considered the following aspects: (i) maintaining DNA integrity by using preservatives, keeping the samples cool and/or process samples as soon as possible; (ii) minimizing sampling protocol differences compared to morphological analysis and (iii) potentially use the same samples for diatom analysis that are intended to be used for the analysis of other organisms. Sampling design elements such as locations, frequency, and number of samples do not require modifications of the classical methodology. The aim of the study is more important than the analysis methodology.

Benthic diatoms are generally collected from common natural sources, such as macrophytes (Phiri et al. 2007), stones (Smucker et al. 2020), sediments (Liu et al. 2020) or artificial substrates (Nadir Solak et al. 2021). The sampling of biofilm can be done in various ways, using a toothbrush, scalpels, or pipetting directly from loose sediment, and is generally standardized in national and international protocols (CEN 2018). Sampling for metabarcoding can be done using the same protocol but requires more rigor to avoid cross-contamination of samples (using gloves and disinfecting sampling and storage equipment). The collected biofilm can then be processed either by chemical cleaning with oxidizing reagents and hydrochloric acid to prepare diatom frustules for microscopic analysis, or it can be used directly as an environmental sample for DNA extraction. Lugol's iodine and formaldehyde are sometimes recommended for sample preservation in morphological analyses (Lund et al. 1958), however, formaldehyde degrades DNA (Mäki et al. 2017). The more commonly used ethanol preservation method is preferred especially when using DNA-based analysis. However, while the microscopic analysis of diatoms requires low concentrated ethanol to fix the cells for morphological identification, e.g. 50% (Burfeid-Castellanos et al. 2022), a higher concentration (min. 70%) is required for DNA preservation (Deiner and Altermatt 2014; Vasselon et al. 2017). Storage approaches can also include the centrifugation of biofilm suspension, followed by deep-freezing of the pellets (Visco et al. 2015) or the use of nucleic acid preservative solutions to stabilize and protect RNA/DNA (Kelly et al. 2020).

Planktic diatoms are usually studied as part of the phytoplankton community but some metabarcoding studies specifically targeted them (Turk Dermastia et

al. 2023). In such cases, sampling can be conducted by filtering a given volume of water (Cui et al., 2023) or by using plankton nets. Samples are preserved and stored using deep-freezing (Hinlo et al. 2017), liquid nitrogen (Cui et al. 2023), or alcohol (Borrego-Ramos et al. 2025) until DNA extraction. In addition, Maitland et al. (2020) tested the potential of analyzing diatoms from DNA extracted from kick-net bulk-tissue samples conventionally used for sampling aquatic macroinvertebrates. The study showed no significant differences in the diatom community assemblages compared to periphyton sample collected from different substrates, suggesting that this new sampling strategy can be used for ecological biomonitoring.

Sediment sampling for diatom eDNA metabarcoding presents unique challenges compared to water-based eDNA collection, primarily due to factors such as water depth and sediment characteristics. Soft sediments, rich in eDNA, are relatively straightforward to sample, whereas hard-bottom sediments are more difficult (Pawlowski et al. 2022). Once collected, the preservation of sediment eDNA samples is critical for maintaining DNA integrity. Common preservation methods include deep-freezing (Visco et al. 2015), preservation with a nucleic acid preservative solution (such as an RNAlater-style solution) (Rissanen et al. 2010), and preservation with ethanol (Sales et al. 2019). These methods aim to prevent DNA degradation and ensure the samples remain suitable for downstream metabarcoding analysis over various storage durations.

For palaeolimnological studies utilizing eDNA from sediments, sampling and preservation protocols are paramount to ensure the integrity and authenticity of ancient DNA (aDNA). Epp et al. (2019) outline key considerations for these procedures. To minimize contamination, subsampling of sediment cores, whether non-frozen or frozen, should be conducted in a clean, sheltered environment, ideally a climate-controlled chamber. For non-frozen cores, sterile scalpels and syringes are used to collect inner core samples, avoiding the outer edges where modern DNA contamination is more likely. For frozen cores, specialized tools like hole saws and drills are employed, with strict decontamination procedures for all equipment. A critical aspect of sampling involves working from the bottom to the top of the core, moving from older, presumably lower-concentration aDNA layers to younger, higher-concentration layers, to prevent contamination. Once collected, sediment samples for aDNA analysis should be stored at low temperatures, typically -20 °C or colder, to inhibit DNA degradation. Efficient aDNA extraction protocols must effectively remove inhibitors, such as humic acids, while maximizing DNA retention. A possible advantage of DNA-based methods is their ability to target different groups of organisms from the same environmental sample, saving time, cost and effort. In the case of the sediment substrates, the samples should be taken by pipetting the outer layer and manually shaking to homogenize them.

Only one of the reviewed studies investigated the effect of sample storage methods on the inferred diatom community composition, diversity and ecological quality index. Baricevic et al. (2022) tested three sample storage methods, including (1) preservation in 70% ethanol and storage at 4 °C, (2) preservation in nucleic acid preservative and storage at -20 °C and (3) centrifugation of biofilm and storage of the remaining pellet at -20 °C. Additionally, storage time ranging from one day to one year was tested, showing no significant effect on ecological index values or diversity metrics.

Regarding the origin of samples studied in the reviewed articles, we can conclude that among the seven habitat types shown in (Fig. 4A), rivers have been sampled most frequently. This likely reflects the fact that benthic diatoms are strongly dominant in riverine systems and thus in routine river biomonitoring programs and water-management frameworks. This has made diatoms a primary focus of scientific, regulatory, and stakeholder attention. Rivers are followed by sea and lake environments, which also show considerable representation. Oceans and estuaries show moderate use, while ponds and springs show the lowest number of diatom DNA studies. These habitats are generally understudied.

DNA extraction

A comprehensive comparison of molecular protocols used in DNA workflows is essential to identify and account for potential biases introduced by different methodologies. Such an understanding will help improve the reliability and consistency of future studies, ensuring more accurate interpretations of DNA-based analyses (Darling and Mahon 2011). In this review, we identified DNA extraction protocols that have been widely used in diatom DNA metabarcoding studies (Fig. 4D). The most complete extraction of DNA from environmental samples is of key importance in order to obtain genetic material of good quality and adequate quantity for the subsequent analyses. This is particularly important for single-celled organisms such as diatoms, where inefficient DNA extraction may lead to the underrepresentation or complete loss of rare species present in low cell numbers. Most of the reviewed studies used commercially available extraction kits. These include steps of (i) lysis or cell disruption to break open cells and release DNA by chemical, enzymatic or/and mechanical means. Most kits use silica columns or magnetic beads to (ii) bind DNA and separate it from other cellular components. To remove contaminants, (iii) several washing steps are included with a buffer solution to preserve DNA binding. We found that the most commonly used kits are those designed for soil samples in order to increase the yield and quality of DNA. Soils are rich in PCR inhibitors, e.g. humic acids, thus their removal is of key importance. The two most popular kits are Qiagen's DNeasy PowerSoil Kit and Macherey-Nagel's NucleoSpin Soil Kit.

A comparative study of DNA extraction steps investigated the effect of five different DNA extraction kits/protocols on diatom community composition, diversity and quality index (Vasselon et al. 2017b). A precipitation-based protocol using Sigma–Aldrich GenElute™-LPA was compared with four commercial silica-column-based kits. All five methods proved suitable for high-throughput sequencing (HTS), and the choice of extraction method did not significantly influence operational taxonomic unit (OTU) richness. Similarly, a more recent study conducted comparative research on four popular DNA extraction kits, including Qiagen Blood and Tissue, Qiagen DNeasy PowerSoil Pro Kit, Thermo MagMAX Microbiome Ultra Nucleic Acid Isolation Kit and the Zymo Quick-DNA Fecal/Soil Microbe Miniprep Kit (Newbold et al. 2025). Their performance was evaluated in terms of PCR and sequencing success. They found that mechanical lysis-based protocols yielded higher PCR and sequencing success rates and increased cross-kingdom richness. However, overall community compositions

were largely consistent across different extraction methods, suggesting that extraction methodology introduces minimal bias in multi-taxa DNA metabarcoding analyses.

Amplicon marker selection

In metabarcoding studies, it is challenging to find the right barcode region and primers that cover most of the diversity of organisms (Cahoon et al. 2018) and it can directly affect the metabarcoding results (Kezlya et al. 2023). A "good barcode" targets a universal gene present in all individuals of the studied community and must be taxonomically informative, capable of distinguishing between species by having a DNA region that mutates at an appropriate rate. This is crucial because modern biomonitoring programs and diatom-based biotic indices typically require species-level identification (Hebert et al. 2003). Additionally, the barcode should have conserved primer binding sites or utilize degenerate primers to ensure it can bind to the DNA of all organisms present in a sample (van der Loos and Nijland 2021). Finally, the barcode fragment size should match the sequencing capabilities of the technology used (Taberlet et al. 2018).

The development of DNA barcoding for protists, including diatoms, faced early challenges due to their immense diversity and lack of a universal marker. Pioneering efforts, such as those by the CBOL Protist Working Group, were crucial in establishing standardized barcoding approaches for eukaryotic richness beyond traditional kingdoms (Pawlowski et al. 2012). Their work, emphasizing both universal and group-specific barcodes, laid foundational groundwork that significantly informs current eDNA metabarcoding advancements for diverse protist groups, including diatoms (Pawlowski et al. 2012). Several early studies explored many conserved gene fragments like *rbcL*, *CO1*, ITS, and 18S DNA barcodes (Evans et al. 2007; Moniz and Kaczmarska 2010; Hamsher et al. 2011; Zimmermann et al. 2011; Stoof-Leichsenring et al. 2020). Choosing the amplicon marker is one of the crucial steps for better identification and phylogenetic analysis, for this reason, three molecular markers targeting the chloroplast (*rbcL*), nuclear (SSU rDNA), and mitochondrial (*CO1*) genomes have been investigated. Among these, *rbcL* demonstrated strong discriminatory capacity and benefited from an extensive reference database for diatoms (Kermarrec et al. 2014). For the *rbcL* amplicon marker, significant progress was made in designing diatom-specific primers targeting 263 bp region of *rbcL* that is variable enough for species-level discrimination while still amplifiable from degraded environmental DNA (Vasselon et al. 2017). Notably widely used is the 263 bp region of *rbcL* in diatom metabarcoding studies as concluded in our literature review search. As shown in (Fig. 4C), over 50 publications using 263 bp *rbcL* make it as a core maker in DNA metabarcoding, supported by its taxonomic resolution and growing database.

A study by Zimmermann et al. (2015) applying NGS sequencing of the V4 region of the 18S gene for metabarcoding diatoms holds significant promise for assessing water quality. Bailet et al. (2019) highlighted the use of two different DNA markers, namely the 18S-V4 and *rbcL* barcodes. The results showed the ability of the molecular method to generate species inventories, diatom indices scores, and ecological status classifications comparable to those derived from

the morphology-based approach. On the other hand, Apothéloz-Perret-Gentil et al. (2021) used three analytical methods to estimate index values: (1) direct calculation from sequences assigned to taxa, (2) *de novo* calibration of eco-values for all metabarcodes, and (3) training predictive models using a supervised machine learning approach. These correlations were higher for the *rbcL* marker than for 18S. This observation shows the significance of the reference database, which is more comprehensive for the *rbcL* marker. At the genus level, 80.4% of diatom genera have been identified by *rbcL*, while for 18S rDNA is 66.7%. These results ensure that *rbcL* is preferable for DNA metabarcoding identification over 18S rDNA (Wang et al., 2024).

High-throughput sequencing of amplicons

High-throughput sequencing (HTS) technologies have revolutionized biodiversity studies by significantly reducing costs and labor and more importantly, it made the DNA sequencing of entire communities from several samples in parallel possible (Taberlet et al. 2012). Thus, compared to the 1st generation of sequencing technologies, like Sanger sequencing which has been used for barcoding of single specimens or strains, this 2nd generation of sequencing (formerly Next Generation Sequencing - NGS) has introduced novel opportunities for diatom biomonitoring (Van Dijk et al. 2014).

The landscape of eDNA metabarcoding is shaped by the choice between short-read and long-read sequencing technologies. Traditional short-read platforms, such as Illumina and Ion Torrent, are typically limited to fragments shorter than 500 bp, requiring the targeting of specific variable regions (Pawlowski et al. 2012). In contrast, long-read platforms from Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) enable sequencing of full-length barcode regions (Chwalińska et al. 2025). Recent advances in platforms such as ONT GridION and PacBio have improved the accuracy and accessibility of long-read metabarcoding. In particular, ONT has gained attention for generating rapid full-length taxonomic profiles, offering a more comprehensive view of diatom community composition than short-read approaches (Jamy et al. 2020).

The first representative of NGS, 454 pyrosequencing, was one of the first technologies with high-throughput, working with short read length and ideal for metabarcoding studies. It was already used for diatom DNA metabarcoding studies (Kermarrec et al. 2013, 2014). Ion Torrent is another early NGS platform that has facilitated the development of DNA metabarcoding approaches targeting diatoms (e.g., Endo et al. 2018; Rivera et al. 2018; Kahlert et al. 2021). Illumina and its various instruments (e.g. MiSeq, HiSeq, NextSeq, NovaSeq) have been the dominant and widely used sequencing technology for metabarcoding (Fig. 5), providing higher accuracy, throughput and lower cost compared to former technologies (Goodwin et al. 2016). The 3rd generation of sequencing technologies, like Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) are working with an increasingly reliable platform for metabarcoding applications and other features may outperform Illumina sequencing. ONT provides a sequencing tool (MinION) that can be used for in-field workflow, and it may enhance taxonomic resolution in community analysis (Stevens et al. 2023). The technology has been seldom used for the community analysis of diatoms or other algae (Amamoo et al. 2025; Kaleli et al. 2025).

Bioinformatic pipelines and reference libraries

Bioinformatics plays a crucial role in analyzing DNA metabarcoding data by processing the sequencing datasets efficiently and providing a meaningful result that allows for an efficient and scalable alternative to understanding ecological dynamics in response to environmental changes (Cordier et al. 2017). Numerous software tools have been used for diatom metabarcoding, including packages such as Mothur (Schloss et al. 2009), QIIME (Caporaso et al. 2010), and DADA2 (Callahan et al. 2016), all of which include all steps from sequence filtering, clustering to taxonomy assignment (Bailet et al. 2020; Rivera et al. 2020). Among the different tools utilized in diatom DNA metabarcoding, DADA2 and Mothur were the most frequently used pipelines (Fig. 6). On the other hand, about 24 studies used customized packages, platforms and datasets for the analysis.

Metabarcoding bioinformatics can deliver entities such as Operational Taxonomic Units (OTUs), Molecular OTUs (MOTUs) or Amplicon Sequence Variants (ASVs), which can be taxonomically annotated using reference databases (Weigand et al. 2019). Bioinformatic pipelines generate single-nucleotide resolution ASVs that provide more precise ecological representations for biological diversity, taxonomic resolution and stronger correlations with environmental variables compared to the morphological approach (Callahan et al. 2017; Kahlert et al. 2021). Additionally, Unlike the Operational Taxonomic Units (OTUs), which are study-specific due to the clustering process, Amplicon Sequence Variants (ASVs) represent exact biological sequences that can be directly compared across different studies and datasets (Callahan et al., 2017).

Recent results confirmed that ecological profiles can be assigned to genetic units similarly to morphological taxa, which can be used to design molecular indices (Pawlowski et al. 2018; Tapolczai et al. 2019; Kelly et al. 2020). Given the frequent occurrence of unclassified sequences often due to gaps in reference data alternative methods have been proposed that use OTU/ASV-specific ecological responses to develop “taxonomy-free” indices (Apothéloz-Perret-Gentil et al. 2017; Tapolczai et al. 2021).

Additionally, ASVs yielded more favorable outcomes when investigating the impacts of land-use on diatoms. Furthermore, it was feasible to develop an enhanced index by employing a taxonomy-free approach. This approach allowed for the retention of valuable information regarding the ecological characteristics of sequences attributed to the same species, as well as unassigned sequences. Moving forward, the integration of novel clustering techniques holds promise in the field of biomonitoring. Such methods can contribute to the refinement of taxonomic unit delineation, placing greater emphasis on ecological factors rather than exclusively relying on morphological or genetic criteria (Tapolczai et al. 2021).

This suggests the effectiveness of incorporating the phylogenetic approach in enhancing the accuracy of ecological assessments based on DNA barcoding of freshwater diatoms. A study utilized a taxonomy-free approach with promising results in terms of accurately assessing the ecological status of the examined sites (Apothéloz-Perret-Gentil et al. 2017). Specifically, the approach yielded a correct assessment for 77% of the sites analyzed. One notable advantage of this method is that it allowed the utilization of nearly 95% of Operational

Taxonomic Units (OTUs) for index calculation. In contrast, the taxonomic assignment approach only enabled the utilization of 35% of the OTUs. This approach offers the advantages of being both time-efficient and cost-effective for assessing the biological quality of surface waters. The values obtained from "taxonomy-free" approaches, which involve molecular assignment and machine learning, showed a stronger correlation with the values of the morphological index compared to the values based on taxonomically assigned sequences. The correlations between the three analytical approaches were higher when using *rbcL* compared to 18S (Apothéloz-Perret-Gentil et al. 2021).

Finally, choosing the bioinformatics pipeline is important as shown in Bailet et al. (2020), where the results from six commonly used bioinformatics pipelines for diatom metabarcoding were compared across various European countries. The six pipelines showed differences in filtering, clustering, and taxonomic assignment all contributing to variations in outputs. Consequently, the chosen bioinformatics pipeline influences the results of ecological status assessments.

Taxa inventory correspondence between DNA metabarcoding and morphological analyses

Since the introduction of DNA metabarcoding to diatom research, several papers have studied the relevance of the methodology by validating resulting taxa inventory based on that of microscopy analysis. Firstly, the proportion of successfully assigned sequences is crucial for an accurate representation of the community. Based on the reviewed studies, the proportion of successful assignments is the lowest at the species rank, followed by the genus and the higher taxonomic ranks. If a sequence lacks a species match during taxonomic assignment, it is either misidentified or assigned to a higher rank. Early studies generally included a lower proportion of assigned sequences which is due to the continuous development of reference databases. One of the most used reference databases in diatom metabarcoding studies is diat.barcode, first published in 2012 with 3,549 barcodes of 715 species, while its latest version (Rimet et al. 2018b) contains 9,233 barcodes of 1,705 species. It contains sequences of barcoded strains from culture collections and several projects, many of which have not yet been deposited in GenBank, but also diatom sequences published on NCBI. Species in understudied regions are less represented in the reference database, which results in lower taxonomic assignment success. Based on the diat.barcode entries, the geographical distribution of barcode sources is highly unbalanced. The origin of most of entries is unknown (28%), followed by France (16%), USA (9%), United Kingdom (7%) and the remaining 68% shared among83 countries.

Important differences in the resulting communities have been observed in both composition and abundances with a varying number of shared taxa among taxonomic levels, year of study and the studied ecoregion. The number of shared species when comparing DNA metabarcoding and morphological approaches shows that there are still important discrepancies between methods when it comes to species composition. However, when the relative read number (as a proxy for abundance) was considered, the proportion of shared reads can be as high as 92% for species and 98% for genera (Tapolczai et al. 2024),

clearly suggesting that inconsistencies are lower for abundant and common taxa. Besides the fact that abundant taxa are found with a higher probability using both methods, they are also more likely to be included in the reference barcode libraries. Additionally, the probability of misidentifying or neglecting abundant species with microscopy is less likely.

It is also a common phenomenon based on our observations that while the microscopy dataset finds, in total, more species considering all the samples in a study, local diversity (richness) is higher with metabarcoding. This can be explained by the sequencing depth, which is usually high enough (several tens of thousands of reads) to detect the entire diatom diversity in a sample. On the other hand, while counting 400 frustules under the microscope constitutes sufficiently representative data for quality assessment (Prygiel et al. 2002), it is not enough to detect rare or low-abundance taxa (Tapolczai et al. 2021; Anslan et al. 2022).

The challenges of biases related to reference libraries are well summarized by Keck et al. (2023). Some of these are relevant for diatom studies as well. Taxonomic mislabeling happens when the identification of the species added to the reference database is incorrect and the problem remains hidden if it affects species without further interest or if results are not validated by experts. Thus, supervision by taxonomists may still be necessary to disentangle such mistakes (e.g. Tapolczai et al. 2024). This shows that besides the completeness of the reference library (Weigand et al. 2019), good-quality curation is almost as important (Keck et al. 2023). Missing taxa in the reference libraries are probably the most important bottleneck in the efficiency of diatom metabarcoding. This leads to no taxonomic assignment or, in the worst case, sequences being wrongly assigned to very closely related species (Couton et al. 2022; Tapolczai et al. 2025). It is thus of great importance to develop reference libraries by adding new entries with assigned taxonomy from cultures (Rimet et al. 2018a).

Relative abundance in metabarcoding studies is estimated using relative read abundances (Mora et al. 2019), by rarefying sample read counts to a given (usually the lowest) read count (Kahlert et al. 2021). Alternatively, only presence/absence data are used (Zimmermann et al. 2014). It has been shown in laboratory experiments that gene copy numbers per cell correlate well with cell biovolume, suggesting that larger species are overrepresented in metabarcoding data regarding their relative abundances as well (Borrego-Ramos et al. 2021; Tapolczai et al. 2024). As the standard quantification method of microscopy data is relative cell count, neglecting cell sizes, it brings discrepancies when methods are compared. Several studies have applied a correction factor based on cell biovolume (Vasselon et al. 2018) to correct abundance values inferred from metabarcoding (Mortágua et al. 2019; Kulaš et al. 2022). This can be very important in the case of truly large taxa, like *Melosira varians* or *Pleurosira laevis*, which can often dominate metabarcoding species inventories or even change the quality grade for a given waterbody (Mortágua et al. 2019; Pérez-Burillo et al. 2020; Kim et al. 2024; Tapolczai et al. 2024).

Some species, especially those that are weakly silicified are challenging to be seen under the microscope or they simply cannot sustain the chemical treatment for permanent slide preparation. A good example is *Fistulifera saprophila* which is more frequently detected by metabarcoding than by microscopy (Pérez-Burillo et al. 2020; Pissaridou et al. 2021; Tapolczai et al. 2024).

This effect is similar for *Entomoneis*, *Amphipleura* or other species (Anslan et al. 2022; Kang et al. 2020).

Microscopy also has a lower ability to detect morphologically similar species which can contribute to further discrepancies. This is often the case with e.g. *Achnanthidium minutissimum*, *Planothidium frequentissimum* or the *Nitzschia palea* species complex (Nistal-García et al. 2021; Kulaš et al. 2022).

Diatom DNA metabarcoding as a bio-assessment tool

Diatoms are important biological indicators of the ecological status of water bodies (Smol and Stoermer 2010) and are widely used for ecological status assessment worldwide (e.g WFD). Many biotic indexes have been developed to assess environmental impacts (e.g. TDI in UK, the Swiss Diatom Indices [DI-CH] in Switzerland, IBD in France (Kelly and Whitton 1995; Prygiel et al. 2002; Hürlimann et al. 2007), however, the traditional way of using a light microscope to count and identify requires extensive taxonomic expertise. As a result, alternative methods using High-Throughput Sequencing (HTS) generated a higher taxonomic resolution and robust reference databases attracted the scientific community (Kermarrec et al. 2013; Visco et al. 2015; Zimmermann et al. 2015; Apothéloz-Perret-Gentil et al. 2017; Vasselon et al. 2017; Bailet et al. 2019; Kahlert et al. 2021; Rivera et al. 2020). Although differences are particularly concerning rare taxa, the results from both methodologies show strong correlations when assessing dominant species, and with ecological indices derived from molecular and morphological data (Bailet et al. 2019, 2020; Pérez-Burillo et al. 2020). The performance of DNA-based identification methods compared to traditional morphological identification of diatoms showed that integrating metabarcoding with traditional methods offers a comprehensive approach to diatom identification, enhancing the reliability and robustness of freshwater ecological status assessments (Zimmermann et al. 2015). The Specific Pollution Sensitivity Index (IPS - Descy and Coste 1991) is one of the most used indices in Europe. Molecular and microscopy-based IPS values correlated significantly making the DNA metabarcoding approach a strong candidate for integration into future ecological assessment frameworks (Tapolczai et al. 2024). Bailet et al. (2019) showed that IPS scores and ecological status assessments obtained by microscopy and metabarcoding were correlated, but still differed significantly because the two methods did not detect exactly the same taxa or estimate their relative abundances in the same way. The incompleteness of barcode reference databases and the stringency of bioinformatic pipelines are key factors contributing to these observed deviations.

Beyond methodological comparisons and taxonomic detection, several studies have used diatom DNA metabarcoding to examine ecological responses to environmental gradients and multiple stressors. These studies highlight the potential of the molecular approach not only for biodiversity assessment, but also for understanding stressor–community relationships in freshwater ecosystems. Diatom DNA metabarcoding studies quantitatively assessed the impacts of nutrients like total phosphorus (TP) and total nitrogen (TN) (Smucker et al. 2020), and monitored the changes in nutrient

enrichment, despite variations in environmental conditions (Clark et al. 2020). In the USA, an extensive study of 1788 streams and rivers found that diatom assemblage structure was strongly correlated with nitrogen concentration, total phosphorus, conductivity, and pH. These four variables were strong predictors of diatom assemblage structure in region-specific analyses, also indicating that diatom environment relationships, the significance of environmental variables, and variable ranges varied between ecoregions (Smucker et al. 2024). Furthermore, stressors like herbicides, nutrients, fungicides, streamflow and physical habitat have been investigated within an agriculture-dominated region. Such field studies can clarify the effects and interactions of stressors when the response variables are well-chosen, and stressors are measured with adequate resolution (Munn et al. 2018). DNA metabarcoding as an effective tool for resolving species identities in spring diatom blooms (Muhammad et al. 2021) revealed the concurrent presence of multiple diatom species during the blooms in the Yeongsan River. *Cyclotella* sp. was the most frequently detected species, closely followed by *Cyclostephanos dubius* and *Discostella* sp. Moreover, results from biomonitoring campaigns in Catalonia and France indicate that the wide ecological tolerances observed in certain diatom species stem from overlapping preferences among genetic variants (Pérez-Burillo et al. 2021). Interestingly, these genetic variants exhibit more restricted preferences and distributions when considered individually (Kahlert et al. 2021; Pérez-Burillo et al. 2021; Tapolczai et al. 2021). These studies highlighted the significance of investigating the ecological preferences of genetic variants within species complexes, a task made feasible through DNA metabarcoding. Overall, the molecular approach may provide a more comprehensive understanding of the ecological status, although it is limited by the incompleteness of the reference database, necessitating further updates to enhance routine diatom metabarcoding (Kulaš et al. 2022).

Diatom DNA metabarcoding as a forensic tool

Recent research has used diatom DNA metabarcoding in forensics, especially in drowning cases (Liu et al. 2020; Dahiya et al. 2024). Traditional approaches were enhanced by the integration of molecular techniques, advanced imaging, and the use of artificial intelligence which have introduced a wide range of promising new research avenues (Tambuzzi et al. 2024).

Major discrepancies were discovered between results generated by the routinely used SEM based diatom tests and 18S rRNA arrays. It was shown that the conventional SEM-based diatom testing method suffers from significantly low sensitivity, mainly due to the high loss of frustules during sample preparation. The adoption of diatom DNA arrays is expected to potentially revolutionize the forensic diagnosis of drowning cases (Jiang et al. 2020). Liu et al. (2020) compared microscopic and DNA barcoding approaches in anatomical organs from 23 drowning victims and found that microscopic observations were weakly correlated with the use of molecular analysis.

These studies show that diatom DNA metabarcoding in forensic science is still in its early stages and requires further investigation to validate its reliability and standardize methodologies.

Limitations

The use of DNA in ecology and nature conservation has increased significantly over recent years, though this growth has not been matched by a full appreciation of its limitations. Some key challenges include (1) imperfect detection, (2) difficulties in quantifying abundance, (3) issues with taxonomic assignment, (4) spatial and temporal dynamics of DNA, (5) complexities in data analysis and interpretation, and (6) challenges in assessing ecological status (Beng and Corlett 2020). A comprehensive understanding of these challenges is essential before DNA-based methods can confidently be adopted for routine biomonitoring. Despite promising results, several studies have highlighted potential biases in species detection and relative abundance data generated through DNA metabarcoding. Laboratory-based biases represent an important limitation in diatom metabarcoding. Differences in DNA extraction efficiency can influence the detected community composition because taxa vary in cell wall properties and susceptibility to lysis (Vasselon et al. 2017). Quantitative interpretation is also affected by biomass and biovolume-related biases, as differences among taxa can distort the relationship between read abundance and actual community structure (Vasselon et al. 2018). In addition, PCR amplification bias is a major source of error, since unequal primer affinity and amplification efficiency among taxa can skew relative abundances and lead to the underrepresentation of some species (Krehenwinkel et al. 2017). More generally, several studies have shown that sequence read proportions do not always reflect the original proportions of taxa in the sample because multiple sources of bias act simultaneously across the workflow (Lamb et al. 2019).

In bioinformatics, differences among approaches such as OTU clustering, chimera detection, denoising, and taxonomic assignment can affect ecological interpretation across studies. However, ASV-based methods reduce the biases associated with OTU clustering. This lack of consistency hinders the development of standardized, easily accessible bioassessment procedures suitable for routine freshwater monitoring (Tapolczai et al. 2019). The utility of diatom DNA metabarcoding for water quality monitoring and biodiversity assessment is clear; however, to maximize the approach's effectiveness, establishing standardized protocols and enhancing reference databases for habitat-specific species are essential steps forward (Kamberović et al. 2024).

Despite these limitations, some collective efforts and progress have been made in different aspects, Vasselon et al. (2025) highlighted the methodological foundation for developing standardized protocols like CEN/ISO standards. This demonstrates that diatom metabarcoding can be implemented reliably and comparably across labs, which is vital for routine biomonitoring in regulatory contexts. Also, recent updates of diat.barcode reflect ongoing efforts to improve reference library completeness. Side by side, huge efforts from European Committee for Standardization (CEN), the International Organization for Standardization (ISO), and collaborative initiatives such as the DNAquaIMG project reflect a coordinated push to standardize diatom-based biomonitoring using DNA metabarcoding. In 2018, CEN released two technical reports: one on managing diatom DNA barcodes (CEN/TR 17244) and another on metabarcoding-adapted sampling methods (CEN/TR 17245). A new 2025 CEN standard further integrates metabarcoding into routine sampling of benthic diatoms.

Additionally, ISO and the DNAquaIMG project, are developing standardized protocols for DNA extraction. These initiatives aim to ensure consistency, comparability, and the broader adoption of molecular tools in aquatic biomonitoring.

Conclusion

DNA metabarcoding in environmental sciences holds great promise, especially in biomonitoring and biodiversity assessment. As bioindicators, diatoms provide effective insights into ecological status, but traditional morphological identification methods need high-level expertise and are costly and time-consuming. Diatom DNA metabarcoding offers an alternative that captures comprehensive community profiles from environmental samples, allowing for rapid, cost-effective, and scalable biomonitoring. After an extensive review of the related literature, we conclude that the following aspects could help to further improve DNA metabarcoding of diatom communities:

i) Further expanding and completing reference databases and taxonomic resolution by focusing on coordinated international initiatives. This process is essential for improving taxonomic assignments and comprehensive biodiversity assessments across different geographical regions and ecosystems. For the diat.barcode initiative, the scientific community organized itself collectively, with a strong emphasis on community driven collaboration. Researchers from around the world share their diatom barcode data, and valuable feedback is provided whenever errors or inconsistencies are spotted. Importantly, everyone who contributes to this effort is recognized as a co-author of diat.barcode, so the database remains open access and free, ensuring that it can serve as a shared, high-quality resource for researchers working with diatom DNA barcoding.
ii) Ensuring methodological standardization and quality assurance for routine monitoring, by establishing consensus protocols for each step in the metabarcoding workflow. These standardization efforts will be particularly important for regulatory applications, where consistency and reliability are paramount. Continuous effort of the scientific community like (Vasselon et al. 2025) and collaborative projects by CEN and ISO ensure consistency in methodological standardization.
iii) Advanced bioinformatic solutions represent a promising advancement for diatom DNA metabarcoding. The shift from OTU-based methods to more precise ASV approaches has already demonstrated improved ecological resolution, also, utilizing taxonomy-free approaches can potentially enable high resolution of diatom biodiversity and ecological status. The integration of machine learning algorithms and artificial intelligence could further enhance data analysis capabilities.
iv) Quantification and abundance estimation, addressing the challenge of quantitative assessment remains a significant hurdle for diatom metabarcoding. Current approaches primarily provide presence-absence or relative abundance data, which may not directly correlate with actual species abundance in the environment. Future research should focus on developing robust methods to correlate read abundance with actual species abundance.

Acknowledgements

We thank the support of the RRF2.3.1 21 2022 00008 project (CSK, BA, KT) within the framework of the Széchenyi Plan Plus program.

References

Almeida SFP, Elias CL, Ferreira J, Tornés E, Puccinelli C, Delmas F, Dörflinger G, Urbanič G, Marcheggiani S, Rosebery J, Mancini L, Sabater S (2014) Water quality assessment of rivers using diatom metrics across Mediterranean Europe: A methods intercalibration exercise. Science of The Total Environment 476: 768–776. https://doi.org/10.1016/J.SCITOTENV.2013.11.144

Amamoo PA, Ayiku ANA, Oyawoye F, Amoako EK, Wiafe G, Botwe BO, Quashie PK (2025) Metabarcoding for the Molecular Diversity of Diatoms, Dinoflagellates and Cyanobacteria as a Relative Measure of Estuarine Health. SSRN Scholarly Paper No. 5136729. Social Science Research Network. https://doi.org/10.2139/ssrn.5136729

Anslan S, Kang W, Dulias K, Wünnemann B, Echeverría-Galindo P, Börner N, Schwarz A, Liu Y, Liu K, Künzel S, Kisand V, Rioual P, Peng P, Wang J, Zhu L, Vences M, Schwalb A (2022) Compatibility of Diatom Valve Records With Sedimentary Ancient DNA Amplicon Data: A Case Study in a Brackish, Alkaline Tibetan Lake. Frontiers in Earth Science 10: 824656. https://doi.org/10.3389/feart.2022.824656

Apothéloz-Perret-Gentil L, Bouchez A, Cordier T, Cordonier A, Guéguen J, Rimet F, Vasselon V, Pawlowski J (2021) Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. Molecular Ecology 30(13): 2959–2968. https://doi.org/10.1111/mec.15646

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(6): 1231–1242. https://doi.org/10.1111/1755-0998.12668

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. https://doi.org/10.1016/j.scitotenv.2020.140948

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 and Metagenomics 3: e34002. https://doi.org/10.3897/mbmg.3.34002

Baricevic A, Chardon C, Kahlert M, Karjalainen SM, Pfannkuchen DM, Pfannkuchen M, Rimet F, Tankovic MS, Trobajo R, Vasselon V, Zimmermann J, Bouchez A (2022) Recommendations for the preservation of environmental samples in diatom metabarcoding studies. Metabarcoding and Metagenomics 6: 349–365. https://doi.org/10.3897/mbmg.6.85844

B-Béres V, Stenger-Kovács C, Buczkó K, Padisák J, Selmeczy GB, Lengyel E, Tapolczai K (2023) Ecosystem services provided by freshwater and marine diatoms. Hydrobiologia 850(12–13): 2707–2733. https://doi.org/10.1007/s10750-022-04984-9

Beng KC, Corlett RT (2020) Applications of environmental DNA (eDNA) in ecology and conservation: Opportunities, challenges and prospects. Biodiversity and Conservation 29(7): 2089–2121. https://doi.org/10.1007/s10531-020-01980-0

Cahoon AB, Huffman AG, Krager MM, Crowell RM (2018) A meta-barcoding census of freshwater planktonic protists in Appalachia – Natural Tunnel State Park, Virginia, USA. Metabarcoding and Metagenomics 1: e26939. https://doi.org/10.3897/mbmg.2.26939

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11(12): 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5): 335–336. https://doi.org/10.1038/nmeth.f.303

CEN (2018) CEN/TR 17245: Water quality. Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. CEN/TC 230/WG23, Aquatic Macrophytes and Algae, 1–8.

Chwalińska M, Karlicki M, Romac S, Not F, Karnkowska A (2025) From short to long reads: enhanced protist diversity profiling via Nanopore metabarcoding. bioRxiv, 2025-06. https://doi.org/10.1101/2025.06.17.660212

Clark DE, Pilditch CA, Pearman JK, Ellis JI, Zaiko A (2020) Environmental DNA metabarcoding reveals estuarine benthic community response to nutrient enrichment – Evidence from an in-situ experiment. Environmental Pollution 267: 115472. https://doi.org/10.1016/j.envpol.2020.115472

Cordier T, Esling P, Lejzerowicz F, Visco J, Ouadahi A, Martins C, Cedhagen T, Pawlowski J (2017) Predicting the Ecological Quality Status of Marine Environments from eDNA Metabarcoding Data Using Supervised Machine Learning. Environmental Science & Technology 51(16): 9118–9126. https://doi.org/10.1021/acs.est.7b01518

Couton M, Lévêque L, Daguin-Thiébaut C, Comtet T, Viard F (2022) Water eDNA metabarcoding is effective in detecting non-native species in marinas, but detection errors still hinder its use for passive monitoring. Biofouling 38(4): 367–383. https://doi.org/10.1080/08927014.2022.2075739

Cui Z, Liu S, Xu Q, Zhao Y, Chen N (2023) Differential ecological adaptation of diverse *Chaetoceros* species revealed by metabarcoding analysis. Environmental DNA 5(6): 1332–1350. https://doi.org/10.1002/edn3.455

Dahiya P, Makwana MD, Chaniyara P, Bhatia A (2024) A comprehensive review of forensic diatomology: Contemporary developments and future trajectories. Egyptian Journal of Forensic Sciences 14(1): 2. https://doi.org/10.1186/s41935-023-00378-7

Darling JA, Mahon AR (2011) From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. Environmental Research 111(7): 978–988. https://doi.org/10.1016/j.envres.2011.02.001

Deiner K, Altermatt F (2014) Transport Distance of Invertebrate Environmental DNA in a Natural River. PLoS ONE 9(2): e88786. https://doi.org/10.1371/journal.pone.0088786

Descy J-P, Coste M (1991) A test of methods for assessing water quality based on diatoms. Internationale Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen. https://www.tandfonline.com/doi/abs/10.1080/03680770.1989.11899905

Desrosiers C, Leflaive J, Eulin A, Ten-Hage L (2013) Bioindicators in marine waters: Benthic diatoms as a tool to assess water quality from eutrophic to oligotrophic coastal ecosystems. Ecological Indicators 32: 25–34. https://doi.org/10.1016/j.ecolind.2013.02.021

Directive WF (2000) EU Water framework directive. EC Directive 60.

Endo H, Ogata H, Suzuki K (2018) Contrasting biogeography and diversity patterns between diatoms and haptophytes in the central Pacific Ocean. Scientific Reports 8(1): 10916. https://doi.org/10.1038/s41598-018-29039-9

Epp LS, Zimmermann HH, Stoof-Leichsenring KR (2019) Sampling and Extraction of Ancient DNA from Sediments. In: Shapiro B, Barlow A, Heintzman PD, Hofreiter M, Paijmans JLA, Soares AER (Eds) Ancient DNA: Methods and Protocols. Springer, 31–44. https://doi.org/10.1007/978-1-4939-9176-1_5

Evans KM, Wortley AH, Mann DG (2007) An Assessment of Potential Diatom "Barcode" Genes (cox1, rbcL, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta). Protist 158(3): 349–364. https://doi.org/10.1016/j.protis.2007.04.001

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics 17(6): 333–351. https://doi.org/10.1038/nrg.2016.49

Hamsher SE, Evans KM, Mann DG, Poulíčková A, Saunders GW (2011) Barcoding Diatoms: Exploring Alternatives to COI-5P. Protist 162(3): 405–422. https://doi.org/10.1016/j.protis.2010.09.005

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & biology, 5(10): R245-R249. https://doi.org/10.1016/S1074-5521(98)90108-9

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270(1512): 313–321. https://doi.org/10.1098/rspb.2002.2218

Hinlo R, Gleeson D, Lintermans M, Furlan E (2017) Methods to maximise recovery of environmental DNA from water samples. PLoS ONE 12(6): e0179251. https://doi.org/10.1371/journal.pone.0179251

Hürlimann J, Liechti P, Niederhauser P, Suisse. Office fédéral de l'environnement (2007) Methoden zur Untersuchung und Beurteilung der Fliessgewässer: Kieselalgen Stufe F (flächendeckend). Bundesamt für Umwelt, BAFU.

Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F (2020) Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. Molecular Ecology Resources 20(2): 429–443. https://doi.org/10.1111/1755-0998.13117

Jiang L, Xiao C, Zhao J, Jiang T, Lin J, Xu Q, Liu C, Cai W (2020) Development of 18S rRNA gene arrays for forensic detection of diatoms. Forensic Science International 317: 110482. https://doi.org/10.1016/j.forsciint.2020.110482

Kahlert M, Albert R-L, Anttila E-L, Bengtsson R, Bigler C, Eskola T, Gälman V, Gottschalk S, Herlitz E, Jarlman A, Kasperoviciene J, Kokociński 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 inter-

calibration exercise 2007 (stream monitoring). Journal of Applied Phycology 21(4): 471–482. https://doi.org/10.1007/s10811-008-9394-5

Kahlert M, Bailet B, Chonova T, Karjalainen SM, Schneider SC, Tapolczai K (2021) Same same, but different: The response of diatoms to environmental gradients in Fennoscandian streams and lakes – barcodes, traits and microscope data compared. Ecological Indicators 130: 108088. https://doi.org/10.1016/j.ecolind.2021.108088

Kaleli A, Ozbayram EG, Zengin Z, Akcaalan R (2025) Investigation of coastal benthic diatom communities using the rbcL marker with nanopore sequencing in Urban Marine environments. Journal of Phycology 61(4): 870–883. https://doi.org/10.1111/jpy.70035

Kamberović J, Gligora Udovič M, Kulaš A, Tapolczai K, Orlić S, Jusufović A, Gajić A, Žutinić P, Ahmić A, Kalamujić Stroil B (2024) The Diatom Diversity and Ecological Status of a Tufa-Depositing River through eDNA Metabarcoding vs. A Morphological Approach—A Case Study of the Una River (Bosnia and Herzegovina). Microorganisms 12(8): 1722. https://doi.org/10.3390/microorganisms12081722

Kang W, Anslan S, Börner N, Schwarz A, Schmidt R, Künzel S, Rioual P, Echeverría-Galindo P, Vences M, Wang J, Schwalb A (2020) Sample-size and bioinformatics independent, consistent patterns among diatom metabarcoding data from lake sediments. Authorea. https://doi.org/10.22541/au.158750300.04703373

Keck F, Blackman RC, Bossart R, Brantschen J, Couton M, Hürlemann S, Kirschner D, Locher N, Zhang H, Altermatt F (2022) Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. Molecular Ecology 31(6): 1820–1835. https://doi.org/10.1111/mec.16364

Keck F, Couton M, Altermatt F (2023) Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources 23(4): 742–755. https://doi.org/10.1111/1755-0998.13746

Keck F, Vasselon V, Tapolczai K, Rimet F, Bouchez A (2017) Freshwater biomonitoring in the Information Age. Frontiers in Ecology and the Environment 15(5): 266–274. https://doi.org/10.1002/fee.1490

Kelly MG, Juggins S, Mann DG, 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. https://doi.org/10.1016/j.ecolind.2020.106725

Kelly MG, Whitton BA (1995) The Trophic Diatom Index: A new index for monitoring eutrophication in rivers. Journal of Applied Phycology 7(4): 433–444. https://doi.org/10.1007/BF00003802

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(1): 349–363. https://doi.org/10.1086/675079

Kermarrec L, Franc A, Rimet F, Chaumeil P, Humbert JF, Bouchez A (2013) Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. Molecular Ecology Resources 13(4): 607–619. https://doi.org/10.1111/1755-0998.12105

Kezlya E, Tseplik N, Kulikovskiy M (2023) Genetic Markers for Metabarcoding of Freshwater Microalgae: Review. Biology 12(7): 1038. https://doi.org/10.3390/biology12071038

Kim Y-S, Yun H-S, Lee J-H, Lee K-L, Choi J-S, Won DH, Kim YJ, Kim H-S, Yoon H-S (2024) Comparison of Metabarcoding and Microscopy Methodologies to Analyze Diatom

Communities in Five Estuaries Along the Southern Coast of the Korean Peninsula. Microbial Ecology 87(1): 95. https://doi.org/10.1007/s00248-024-02396-x

Kulaš A, Gligora Udovič M, Tapolczai K, Žutinić P, Orlić S, Levkov Z (2022) Diatom eDNA metabarcoding and morphological methods for bioassessment of karstic river. Science of the Total Environment 829: 154536. https://doi.org/10.1016/j.scitotenv.2022.154536

Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG (2017) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports 7(1): 17668. https://doi.org/10.1038/s41598-017-17333-x

Lamb PD, Hunter E, Pinnegar JK, Creer S, Davies RG, Taylor MI (2019) How quantitative is metabarcoding: A meta-analytical approach. Molecular Ecology 28: 420–430. https://doi.org/10.1111/mec.14920

Lavoie I, Campeau S, Zugic-Drakulic N, Winter JG, Fortin C (2014) Using diatoms to monitor stream biological integrity in Eastern Canada: An overview of 10 years of index development and ongoing challenges. Science of The Total Environment 475: 187–200. https://doi.org/10.1016/j.scitotenv.2013.04.092

Liu M, Zhao Y, Sun Y, Wu P, Zhou S, Ren L (2020) Diatom DNA barcodes for forensic discrimination of drowning incidents. FEMS Microbiology Letters 367(17): fnaa145. https://doi.org/10.1093/femsle/fnaa145

Lobo EA, Heinrich CG, Schuch M, Wetzel CE, Ector L (2016) Diatoms as Bioindicators in Rivers. In: Necchi Jr O (Ed) River Algae. Springer International Publishing, 245–271. https://doi.org/10.1007/978-3-319-31984-1_11

Lund JWG, Kipling C, Le Cren ED (1958) The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia 11(2): 143–170. https://doi.org/10.1007/BF00007865

Maitland VC, Robinson CV, Porter TM, Hajibabaei M (2020) Freshwater diatom biomonitoring through benthic kick-net metabarcoding. PLoS ONE 15(11): e0242143. https://doi.org/10.1371/journal.pone.0242143

Mäki A, Salmi P, Mikkonen A, Kremp A, Tiirola M (2017) Sample Preservation, DNA or RNA Extraction and Data Analysis for High-Throughput Phytoplankton Community Sequencing. Frontiers in Microbiology 8: 1848. https://doi.org/10.3389/fmicb.2017.01848

Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, Wincker P, Iudicone D, De Vargas C, Bittner L, Zingone A, Bowler C (2016) Insights into global diatom distribution and diversity in the world's ocean. Proceedings of the National Academy of Sciences 113(11): E1516-25. https://doi.org/10.1073/pnas.1509523113

Mann DG (2010) Discovering diatom species: Is a long history of disagreements about species-level taxonomy now at an end? Plant Ecology and Evolution 143(3): 251–264. https://doi.org/10.5091/plecevo.2010.405

Mann DG, Vanormelingen P (2013) An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species. Journal of Eukaryotic Microbiology 60(4): 414–420. https://doi.org/10.1111/jeu.12047

Moniz MBJ, Kaczmarska I (2010) Barcoding of Diatoms: Nuclear Encoded ITS Revisited. Protist 161(1): 7–34. https://doi.org/10.1016/j.protis.2009.07.001

Mora D, Abarca N, Proft S, Grau JH, Enke N, Carmona J, Skibbe O, Jahn R, Zimmermann J (2019) Morphology and metabarcoding: A test with stream diatoms from Mexico highlights the complementarity of identification methods. Freshwater Science 38(3): 448–464. https://doi.org/10.1086/704827

Mortágua A, Vasselon V, Oliveira R, Elias C, Chardon C, Bouchez A, Rimet F, João Feio M, FP Almeida S (2019) Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. Ecological Indicators 106: 105470. https://doi.org/10.1016/j.ecolind.2019.105470

Muhammad BL, Sin Y, Ki J-S (2021) DNA metabarcoding reveals multiple co-occurring species in Stephanodiscus spring diatom blooms in a temperate freshwater river. Marine and Freshwater Research 72(9): 1353–1364. https://doi.org/10.1071/MF20254

Munn MD, Waite I, Konrad CP (2018) Assessing the influence of multiple stressors on stream diatom metrics in the upper Midwest, USA. Ecological Indicators 85: 1239–1248. https://doi.org/10.1016/j.ecolind.2017.09.005

Nadir Solak C, Ljubimir S, Kaleli A, Dupčić Radić I, Hafner D, Car A (2021) Comparison of benthic diatom community structures on natural and artificial substrates in marine lake (Adriatic Sea). Acta Adriatica 62(1): 21–44. https://doi.org/10.32582/aa.62.1.1

Newbold LK, Taylor JD, Thorpe AC, Warren J, Walsh K, Read DS (2025) DNA Extraction Methodology has a Limited Impact on Multitaxa Riverine Benthic Metabarcoding Community Profiles. Environmental DNA 7(3): e70102. https://doi.org/10.1002/edn3.70102

Nistal-García A, García-García P, García-Girón J, Borrego-Ramos M, Blanco S, Bécares E (2021) DNA metabarcoding and morphological methods show complementary patterns in the metacommunity organization of lentic epiphytic diatoms. Science of The Total Environment 786: 147410. https://doi.org/10.1016/j.scitotenv.2021.147410

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirků M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EA, Nitsche F, Romeralo M, Saunders GW, Simpson AG, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLoS Biology 10(11): e1001419. https://doi.org/10.1371/journal.pbio.1001419

Pawlowski J, Bruce K, Panksep K, Aguirre FI, Amalfitano S, Apothéloz-Perret-Gentil L, Baussant T, Bouchez A, Carugati L, Cermakova K, Cordier T, Corinaldesi C, Costa FO, Danovaro R, Dell'Anno A, Duarte S, Eisendle U, Ferrari BJD, Frontalini F, Fazi S (2022) Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. Science of The Total Environment 818: 151783. https://doi.org/10.1016/j.scitotenv.2021.151783

Pawlowski J, Kelly-Quinn M, Altermatt F, Apothéloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Feio MJ, Filipe AF, Fornaroli R, Graf W, Herder J, van der Hoorn B, Iwan Jones J, Sagova-Mareckova M, Moritz C, 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. https://doi.org/10.1016/j.scitotenv.2018.05.002

Pawlowski J, Lejzerowicz F, Apotheloz-Perret-Gentil L, Visco J, Esling P (2016) Protist metabarcoding and environmental biomonitoring: Time for change. European Journal of Protistology 55: 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Pérez-Burillo J, Trobajo R, Leira M, Keck F, Rimet F, Sigró J, Mann DG (2021) DNA metabarcoding reveals differences in distribution patterns and ecological preferences among genetic variants within some key freshwater diatom species. Science of The Total Environment 798: 149029. https://doi.org/10.1016/j.scitotenv.2021.149029

Pérez-Burillo J, Trobajo R, Vasselon V, Rimet F, Bouchez A, Mann DG (2020) Evaluation and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of

Mediterranean rivers. Science of The Total Environment 727: 138445. https://doi.org/10.1016/j.scitotenv.2020.138445

Phiri C, Day J, Chimbari M, Dhlomo E (2007) Epiphytic diatoms associated with a submerged macrophyte, *Vallisneria aethiopica*, in the shallow marginal areas of Sanyati Basin (Lake Kariba): A preliminary assessment of their use as biomonitoring tools. Aquatic Ecology 41(2): 169–181. https://doi.org/10.1007/s10452-006-9073-z

Pissaridou P, Vasselon V, Christou A, Chonova T, Papatheodoulou A, Drakou K, Tziortzis I, Dörflinger G, Rimet F, Bouchez A, Vasquez MI (2021) Cyprus' diatom diversity and the association of environmental and anthropogenic influences for ecological assessment of rivers using DNA metabarcoding. Chemosphere 272: 129814. https://doi.org/10.1016/j.chemosphere.2021.129814

Prygiel J, Carpentier P, Almeida S, Coste M, Druart J-C, Ector L, Guillard D, Honoré M-A, Iserentant R, Ledeganck P, Lalanne-Cassou C, Lesniak C, Mercier I, Moncaut P, Nazart M, Nouchet N, Peres F, Peeters V, Rimet F, Zydek N (2002) Determination of the biological diatom index (IBD NF T 90–354): Results of an intercomparison exercise. Journal of Applied Phycology 14(1): 27–39. https://doi.org/10.1023/A:1015277207328

Rimet F (2012) Recent views on river pollution and diatoms. Hydrobiologia 683(1): 1–24. https://doi.org/10.1007/s10750-011-0949-0

Rimet F, Abarca N, Bouchez A, Kusber W-H, Jahn R, Kahlert M, Keck F, Kelly MG, Mann David. G, Piuz A, Trobajo R, Tapolczai K, Vasselon V, Zimmermann J (2018a) The potential of High-Throughput Sequencing (HTS) of natural samples as a source of primary taxonomic information for reference libraries of diatom barcodes. Fottea 18(1): 37–54. https://doi.org/10.5507/fot.2017.013

Rimet F, Bouchez A (2012) Biomonitoring river diatoms: Implications of taxonomic resolution. Ecological Indicators 15(1): 92–99. https://doi.org/10.1016/j.ecolind.2011.09.014

Rimet F, Chonova T, Gassiole G, Gusev E, Kahlert M, Keck F, Kelly M, Kochoska H, Kulikovskiy M, Levkov Z, Maltsev Y, Mann D, Nicolosi Gelis MM, Pfannkuchen M, Trobajo R, Vasselon V, Vidakovic D, Wetzel C, Zimmermann J, Bouchez A (2018b) Diat. barcode, an open-access barcode library for diatoms. Recherche Data Gouv, V18. https://doi.org/10.1038/s41598-019-51500-6

Rissanen AJ, Kurhela E, Aho T, Oittinen T, Tiirola M (2010) Storage of environmental samples for guaranteeing nucleic acid yields for molecular microbiological studies. Applied Microbiology and Biotechnology 88(4): 977–984. https://doi.org/10.1007/s00253-010-2838-2

Rivera SF, Vasselon V, Ballorain K, Carpentier A, Wetzel CE, Ector L, Bouchez A, Rimet F (2018) DNA metabarcoding and microscopic analyses of sea turtles biofilms: Complementary to understand turtle behavior. PLoS ONE 13(4): e0195770. https://doi.org/10.1371/journal.pone.0195770

Rivera SF, 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. https://doi.org/10.1016/j.ecolind.2019.105775

Sales NG, Wangensteen OS, Carvalho DC, Mariani S (2019) Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. Environmental DNA 1(2): 198–207. https://doi.org/10.1002/edn3.14

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology 75(23): 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schulte NO, Craine JM, Leopold DR, Devitt JK, Fierer N (2024) Diversity of freshwater algal assemblages across the United States as revealed by DNA metabarcoding. Environmental DNA 6(1): e477. https://doi.org/10.1002/edn3.477

Smol JP, Stoermer EF (2010) The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press. https://doi.org/10.1017/cbo9780511763175.002

Smucker NJ, Pilgrim EM, Nietch CT, Darling JA, Johnson BR (2020) DNA metabarcoding effectively quantifies diatom responses to nutrients in streams. Ecological Applications 30(8): e2205. https://doi.org/10.1002/eap.2205

Smucker NJ, Pilgrim EM, Nietch CT, Gains-Germain L, Carpenter C, Darling JA, Yuan LL, Mitchell RM, Pollard AI (2024) Using DNA metabarcoding to characterize national scale diatom-environment relationships and to develop indicators in streams and rivers of the United States. Science of The Total Environment 939: 173502. https://doi.org/10.1016/j.scitotenv.2024.173502

Stenger-Kovács C, Buczkó K, Hajnal É, Padisák J (2007) Epiphytic, littoral diatoms as bioindicators of shallow lake trophic status: Trophic Diatom Index for Lakes (TDIL) developed in Hungary. Hydrobiologia 589(1): 141–154. https://doi.org/10.1007/s10750-007-0729-z

Stephen C, Joshi L (2010) Algal Biotechnology: An Emerging Resource with Diverse Application and Potential. In: Kole C, Michler CH, Abbott AG, Hall TC (Eds) Transgenic Crop Plants. Springer, 343–357. https://doi.org/10.1007/978-3-642-04812-8_9

Stevens BM, Creed TB, Reardon CL, Manter DK (2023) Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. Scientific Reports 13(1): 9323. https://doi.org/10.1038/s41598-023-36101-8

Stoof-Leichsenring KR, Dulias K, Biskaborn BK, Pestryakova LA, Herzschuh U (2020) Lake-depth related pattern of genetic and morphological diatom diversity in boreal Lake Bolshoe Toko, Eastern Siberia. PLoS ONE 15(4): e0230284. https://doi.org/10.1371/journal.pone.0230284

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

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21(8): 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tambuzzi S, Gentile G, Zoia R (2024) Forensic Diatom Analysis: Where Do We Stand and What Are the Latest Diagnostic Advances? Diagnostics 14(20): 2302. https://doi.org/10.3390/diagnostics14202302

Tapolczai K, Chonova T, Fidlerová D, Makovinská J, Mora D, Weigand A, Zimmermann J (2024) Molecular metrics to monitor ecological status of large rivers: Implementation of diatom DNA metabarcoding in the Joint Danube Survey 4. Ecological Indicators 160: 111883. https://doi.org/10.1016/j.ecolind.2024.111883

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: 409. https://doi.org/10.3389/fevo.2019.00409

Tapolczai K, Rimet F, Ćirić M, Ballot A, Laplace-Treyture C, Alric B (2025) A novel framework for phytoplankton biomonitoring: Trait assignment of 23S rRNA sequences. Ecological Indicators 173: 113361. https://doi.org/10.1016/j.ecolind.2025.113361

Tapolczai K, Selmeczy GB, 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. https://doi.org/10.1016/j.ecolind.2020.107322

Turk Dermastia T, Vascotto I, Francé J, Stanković D, Mozetič P (2023) Evaluation of the rbcL marker for metabarcoding of marine diatoms and inference of population structure of selected genera. Frontiers in Microbiology 14: 1071379. https://doi.org/10.3389/fmicb.2023.1071379

Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends in Genetics 30(9): 418–426. https://doi.org/10.1016/j.tig.2014.07.001

Van der Loos LM, Nijland R (2021) Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology 30(13): 3270–3288. https://doi.org/10.1111/mec.15592

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(4): 1060–1069. https://doi.org/10.1111/2041-210X.12960

Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A (2017b) Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshwater Science 36(1): 162–177. https://doi.org/10.1086/690649

Vasselon V, Rimet F, Domaizon I, Monnier O, Reyjol Y, Bouchez A (2019) Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: Experience and developments from France Water Framework Directive networks. Metabarcoding and Metagenomics 3: e39646. https://doi.org/10.3897/mbmg.3.39646

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. https://doi.org/10.1016/j.ecolind.2017.06.024

Vasselon V, Rivera SF, Ács É, Almeida SFP, Andree KB, Apothéloz-Perret-Gentil L, Bailet B, Baričević A, Beentjes KK, Bettig J, Bouchez A, Capelli C, Chardon C, Duleba M, Elersek T, Genthon C, Jablonska M, Jacas L, Kahlert M, Paix B (2025) Proficiency testing and cross-laboratory method comparison to support standardisation of diatom DNA metabarcoding for freshwater biomonitoring. Metabarcoding and Metagenomics 9: e133264. https://doi.org/10.3897/mbmg.9.133264

Visco JA, 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(13): 7597–7605. https://doi.org/10.1021/es506158m

Wang H, Liu K, He Z, Chen Y, Hu Z, Chen W, Leaw CP, Chen N (2024) Extensive intragenomic variations of the 18S rDNA V4 region in the toxigenic diatom species *Pseudo-nitzschia multistriata* revealed through high-throughput sequencing. Marine Pollution Bulletin 201: 116198. https://doi.org/10.1016/j.marpolbul.2024.116198

Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wyler SA, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Ekrem T (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Science of The Total Environment 678: 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247

Zhan A, Hulák M, Sylvester F, Huang X, Adebayo AA, Abbott CL, Adamowicz SJ, Heath DD, Cristescu ME, MacIsaac HJ (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods in Ecology and Evolution 4(6): 558–565. https://doi.org/10.1111/2041-210X.12037

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(9): e108793. https://doi.org/10.1371/journal.pone.0108793

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(3): 526–542. https://doi.org/10.1111/1755-0998.12336

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(3): 173–192. https://doi.org/10.1007/s13127-011-0050-6

Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, De Barba M, Dickie IA, Dumbrell AJ, Ficetola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Jumpponen A, Taberlet P (2019) DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology 28(8): 1857–1862. https://doi.org/10.1111/mec.15060

Supplementary material 1

Additional information

Authors: Bassam Abubaker, Csilla Stenger-Kovács, Frédéric Rimet, Kálmán Tapolczai

Data type: xlsx

Explanation note: The Data from 128 publications were extracted into a standardized table including the following variables: publication year, study location (country), habitat, type of samples, method of DNA extraction, target gene region, and sequencing technology.

Link: https://doi.org/10.3897/mbmg.10.168096.suppl1

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Artificial Intelligence (AI) use

The authors accept full responsibility for the content of the manuscript, including the disclosure of any use of AI.
No AI tools were used in the preparation of this manuscript.

Funding

The study was financed by the Hungarian National Research, Development and Innovation Office (NRDIO): K 137950 and FK 146760; and the János Bolyai research grant from the Hungarian Academy of Sciences.

Author contributions

All authors have contributed equally.

Author ORCIDs

B. Abubaker https://orcid.org/0000-0003-4659-5022
C. Stenger-Kovács https://orcid.org/0000-0001-6175-4904
F. Rimet https://orcid.org/0000-0002-5514-869X
K. Tapolczai https://orcid.org/0000-0003-1453-767X

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.