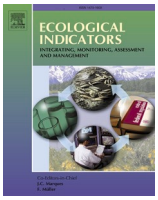
Ecological Indicators 160 (2024) 111883

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Molecular metrics to monitor ecological status of large rivers: Implementation of diatom DNA metabarcoding in the Joint Danube Survey 4

Kálmán Tapolczai [a,b,*], Teofana Chonova [c], Dana Fidlerová [d], Jarmila Makovinská [d], Demetrio Mora [e,f,g], Alexander Weigand [h], Jonas Zimmermann [f]

[a] HUN-REN Balaton Limnological Research Institute, Klebelsberg Kuno Street 3., H-8237 Tihany, Hungary
[b] National Laboratory for Water Science and Water Security, HUN-REN Balaton Limnological Research Institute, Tihany, Hungary
[c] Department Environmental Chemistry, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[d] Water Research Institute, Nábr. arm. gen. L. Svobodu 5, 81249 Bratislava, Slovakia
[e] Department of Microbial Ecology, Federal Institute of Hydrology, Am Mainzer Tor 1, 56068 Koblenz, Germany
[f] Botanischer Garten und Botanisches Museum Berlin, Freie Universität Berlin, Königin-Luise-Str. 6-8, 14195 Berlin, Germany
[g] Observatory for Climate, Environment and Biodiversity, Luxembourg Institute of Science and Technology, 41 rue du Brill, L-4422 Belvaux, Luxembourg
[h] National Museum of Natural History Luxembourg, 25 Rue Münster, L-2160 Luxembourg, Luxembourg

ARTICLE INFO

Keywords:
Bacillariophyta
Biomonitoring
Ecological quality assessment
Environmental DNA
Phytobenthos
*Rbc*L

ABSTRACT

The Joint Danube Survey (JDS) is a regular transnational survey to monitor the quality of the Danube and its main tributaries, in accordance with the EU Water Framework Directive. During the JDS4 in 2019, conventional methods to monitor selected biological quality elements were complemented with DNA metabarcoding.

All together 72 phytobenthos samples were collected along the Danube and its major tributaries within the JDS4, using light microscopy and DNA metabarcoding amplifying a fragment of the *rbc*L marker gene. The (i) applicability of DNA metabarcoding to identify diatom communities compared to microscopy; (ii) diversity metrics between DNA metabarcoding and microscopy analysis and (iii) the usability of DNA metabarcoding for routine monitoring and assessment of the Danube under future JDS surveys were investigated.

Diatom communities resulting from light microscopy and DNA metabarcoding assessments share 26.5% of all taxa which corresponds to 64.3% when considering relative abundances. Discrepancies originate from biases both from metabarcoding, e.g. missing taxa from the reference library, and from microscopy, e.g. overlooking of hardly visible taxa. Microscopy detected more taxa in total but metabarcoding revealed a higher alpha diversity, detecting also very rare taxa in a given sample. Molecular and microscopy based Specific Pollution Sensitivity Index (IPS) values correlated significantly but differences were detected at several sites due to the differences in community composition and the overestimation of large taxa by metabarcoding. Although both methods showed a decreasing trend of IPS along the Danube, the metabarcoding based IPS covered a higher range of quality classes indicating lower values for downstream sites and the tributaries.

We suggest that metabarcoding provides a standardisable and efficient tool in biomonitoring, being more distinctive among quality classes than microscopy. Due to the high sequencing depth, it is able to detect a higher diversity on genetic level in a time- and cost-efficient manner that should be implemented in future quality assessment tools. We recommend its use in future biomonitoring surveys, for now, as a complementary method to conventional ones.

1. Introduction

The monitoring and protection of our water bodies is an essential task for humanity in order to ensure good ecological quality and sustainable ecosystem functions and services, especially in an era when waters and their biota are exposed to several threats, e.g. climate change, environmental degradation and overexploitation (Heino et al., 2009; Tickner et al., 2020, Vörösmarty et al., 2010). Large rivers, like

* Corresponding author at: HUN-REN Balaton Limnological Research Institute, Klebelsberg Kuno Street 3., H-8237 Tihany, Hungary.
E-mail address: tapolczai.kalman@blki.hu (K. Tapolczai).

https://doi.org/10.1016/j.ecolind.2024.111883
Received 17 November 2023; Received in revised form 23 February 2024; Accepted 10 March 2024
Available online 13 March 2024

the Danube, whose river basin covers nineteen countries, are exposed to a diverse set of anthropogenic pressures and require international cooperation for efficient water management and ecological status assessment (Liška, 2015). According to the requirements of the European Water Framework Directive (WFD; European European Commission, 2000), the quality of the Danube in regards to several pollutants and biological quality elements are regularly monitored by transnational and national monitoring surveys. These are complemented by a comprehensive survey carried out every six years, known since 2001 as the Joint Danube Survey (JDS), and overseen by the International Commission for the Protection of the Danube River (ICPDR) through the Monitoring and Assessment Expert Group.

Phytobenthos is one of the biological quality elements defined by the WFD to assess the ecological quality status of freshwater bodies, including large rivers, with diatoms used in most EU countries as proxies for phytobenthos. Being suitable indicators of water quality, diatom indices are part of the water management monitoring toolkit worldwide to assess either specific pollutants or the general ecological status (Charles et al., 2021). Different indices were developed for different stressors and ecoregions but IPS (Specific Pollution sensitivity Index; Coste, 1982) is the most widely applied within Europe; and also used for the ecological quality assessment of the Danube (Fidlerová and Makovinská, 2021). IPS represents an early generation of diatom indices, not differentiating between organic and inorganic pollution but rather assessing a "general degradation" (Schneider et al., 2017). It is based on the weighted averaging equation of Zelinka and Marvan (1961): $\sum a_j v_j s_j / \sum a_j v_j$ where a_j is the relative abundance of the species j, v_j and s_j are its indicator (1–3) and sensitivity (1–5) values, respectively (Prygiel and Coste, 1993). It required the establishment of a database where values are predetermined for most species in the samples based on their autecology; their response in abundance and distribution related to the environment (Coste, 1982; Descy and Coste, 1991). Based on the scientific literature, IPS is an efficient metric to assess degradation, primarily organic pollution and eutrophication (Kelly, 2013; Prygiel and Coste, 1993). The proper identification of diatom species is thus necessary and is conventionally carried out by the detailed morphological examination of the valves with light microscopy. However, this method might entail several biases, e.g. being time consuming, and – in the case of some very similar species – rather challenging even for experts, leading to inconsistencies (Charles et al., 2021). These uncertainties require further work in the form of intercalibration practices to handle inconsistencies between labs (Kahlert et al., 2009).

The application of DNA metabarcoding to characterise diatom communities has been developed in the last decade and has become a promising tool for a potentially less biassed, cost- and time efficient diatom-based ecological quality monitoring and assessment, once a standardised methodology is set (Cordier et al., 2021; Pawlowski et al., 2020, 2018). The chloroplast marker gene *rbc*L is today considered the most efficient to characterise freshwater diatom communities on a molecular basis, outperforming other genes by providing better taxonomic resolution (Apothéloz-Perret-Gentil et al., 2021; Bailet et al., 2020; Kermarrec et al., 2013) and reference database coverage (Rimet et al., 2019). Instead of morphospecies, metabarcoding infers molecular identifiers, e.g. OTUs (Operational Taxonomic Units), MOTUs (Molecular OTUs), ASVs (Amplicon Sequence Variants), ESVs (Exact Sequence Variants), whose sequences can be assigned to taxonomy based on reference libraries (Weigand et al., 2019). However, there are alternative approaches that use the autecology of OTUs/ASVs to infer "taxonomy-free" indices, to bypass the limitations of Linnaean taxonomic assignment, which normally result in a large proportion of unclassified sequences, due to the incompleteness of taxonomic reference libraries for described diatoms (Apothéloz-Perret-Gentil et al., 2017; Feio et al., 2020; Tapolczai et al., 2021). It enables us to reveal a fine-scale hidden diversity of genetic variants.

Several studies have been published in which diatom communities and quality assessment indices obtained via DNA metabarcoding and microscopy are compared and inconsistencies are discussed (e.g. Bailet et al., 2019; Duleba et al., 2021; Kahlert et al., 2021; Kelly et al., 2020; Kulaš et al., 2022; Mortágua et al., 2019). Many of these studies revealed important differences in the identified species inventories, which were attributable to a higher accordance in the case of abundant and common species, but with molecular and morphology-based indices generally correlating well (Bailet et al., 2019; Duleba et al., 2021; Kelly et al., 2020; Mortágua et al., 2019; Pérez-Burillo et al., 2020; Vasselon et al., 2017).

Our objective in this study was to evaluate the applicability of DNA metabarcoding within the context of a large international monitoring program by analysing those aspects mentioned above. Diatom DNA metabarcoding was applied for the first time on phytobenthos samples from the Danube within the framework of the 4th Joint Danube Survey (JDS4), as detailed in Zimmermann et al. (2021) and this study. Here we analysed environmental and diatom data obtained with both microscopy and metabarcoding methods, in order to investigate (i) the applicability of DNA metabarcoding to identify diatom communities compared to microscopy; (ii) diversity metrics between DNA metabarcoding and microscopy analysis and (iii) the usability of DNA metabarcoding for routine monitoring and quality assessment of the Danube under future JDS campaigns, as well as for other large rivers of Europe.

2. Materials & methods

2.1. Study site and sampling for environmental variables

The Danube is Europe's 2nd largest river with its 2,850 km length, traversing nine countries (Germany, Austria, Slovakia, Hungary, Croatia, Republic of Serbia, Romania, Bulgaria, Ukraine) from the Black Forest until it flows into the Black Sea. Its river basin spans 801,463 km^2 being the 2nd largest in Europe, shared by 19 countries of which 14 are contracting parties of the ICPDR, namely Austria, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Germany, Hungary, Montenegro, Republic of Moldova, Republic of Serbia, Romania, Slovakia, Slovenia, Ukraine (Liška, 2015). Danubian sites are classified into the Danubian types (Moog et al., 2006) as follows: type 1 (2,581 river kilometers - rkm), type 2 (2,479.3–2,258 rkm), type 3 (2,204–2,008 rkm), type 4 (1,878–1,791 rkm), type 5 (1,707–1,532 rkm), type 6 (1,480–1,073 rkm), type 7 (was lacking of sampling sites, type 8 (852–488 rkm), type 9 (375–132 rkm), type 10 (17 rkm-0 rkm) and all tributaries were classified into one group. The Danubian types are also grouped into the major Danubian reaches, i.e. Upper Danube (types 1–4), Middle Danube (types 5 & 6) and Lower Danube (types 7–10).

Forty-three sites were sampled for environmental variables during July 2019, of which 16 were situated in tributaries and 27 located in the main channel of the Danube. Location and metadata of sampling sites are shown in Fig. S1-A and Table S1. The sampling for the physical and chemical parameters was carried out from surface water by the national teams of the participating laboratories based on standard operating procedures (Liška et al., 2021). The sampled variables used in this study are alkalinity, pH, biological oxygen demand (BOD_5), chlorophyll-*a* (chl-*a*), chemical oxygen demand (COD), total organic carbon (TOC), dissolved organic carbon (DOC), conductivity, dissolved oxygen (DO), oxygen saturation (OS), ammonium nitrogen (NH_4^+-N), nitrite nitrogen (NO_2^--N), nitrate nitrogen (NO_3^--N), total nitrogen (TN), phosphate phosphorus (PO_4^{3-}-P), total phosphorus (TP), suspended solids (SS), and temperature (Table S2).

2.2. Phytobenthos sampling and microscopy analysis of samples

Benthic diatoms were sampled within the first two weeks of July 2019, in the same time period as the samples for the physical and chemical analyses of the water, following European standards (CEN, 2014a, 2018a). Samples were collected separately from both river banks

except sites where it was logistically not possible, resulting in a total of 72 samples. The length of selected sampling stretches was at least 10 m. Using a clean toothbrush at each site to prevent cross-contamination, samples were brushed off the light exposed surface of a minimum 10 cm^2 of the substrates, usually at least five stones occurring in the euphotic zone. Samples were divided into subsamples and preserved separately for microscopy analysis and molecular analysis, using 97 % ethanol to reach a final concentration of 70 %, complemented with deep freezing in the case of samples for molecular analysis. Sampled sites for environmental variables and phytobenthos slightly differed with some sites uniquely sampled either for environmental variables or phytobenthos (Fig. S1-B).

The subsequent microscopy analysis followed international standards (CEN, 2014b), including the hot hydrogen peroxide method to remove organic material from samples, the preparation of permanent slides and the counting and identification of a number of 300–500 diatom valves under a light microscope at 1000× magnification, to the lowest possible taxonomic level. Relative count abundance of taxa was calculated and used in subsequent analyses.

2.3. Molecular analysis of phytobenthos samples

2.3.1. Wet lab preparation of phytobenthos samples for High-Throughput sequencing

A volume of 2 to 4 mL of the defrosted phytobenthos sample suspensions were centrifuged at 13,000 x g for 30 min. The NucleoSpin® Soil kit of Macherey & Nagel (MN-Soil) was used to extract DNA from the resulting pellets. DNA concentrations were quantified using a Qubit 2.0 fluorometer (Invitrogen, Carlsbad, California, USA) and adjusted to a concentration of 20 ng μL^{-1} for PCR.

The 312 bp fragment of the plastid *rbc*L gene was used as a molecular marker and amplified by an equimolar mix of the modified versions of the Diat_rbcL_708F and R3 primers established by Vasselon et al. (2017); the Diat_rbcL_708F_1 (5′-AGGTGAAGTAAAAGGTTCWTACTTAAA-3′), Diat_rbcL_708F_2 (5′-AGGTGAAGTTAAAGGTTCWTAYTTAAA-3′) and Diat_rbcL_708F_3 (5′-AGGTGAAACTAAAGGTTCWTACTTAAA-3′) forward primers, and the R3_1 (5′-CCTTCTAATTTACCWACWACTG-3′) and R3_2 (5′-CCTTCTAATTTACCWACAACAG-3′) reverse primers. Each PCR mix was composed by 1 µL of extracted DNA, 0.75 U of Takara LA Taq® polymerase, 2.5 µL of 10X Buffer, 1.25 µL of 10 µM of primers Diat_rbcL_708F_1_2_3 and R3_1_2, 1.25 µL of 10 gL^{-1} BSA, 2 µL of 2.5 mM dNTP, and completed with molecular biology grade water. The PCR reaction conditions were initiated by a denaturation step at 95 °C for 15 min followed by a total of 30 cycles of 95 °C for 45 s (denaturation), 55 °C for 45 s (annealing), and 72 °C for 45 s (final extension). For each sample PCR was once repeated for technical replication. Purification of the samples was performed with 25 mL aliquots of the amplicons with HighPrep PCR Clean–up System (Magbio Genomics). Negative controls were conducted for the PCR as well as for the sequencing step.

Sequencing was conducted at the Berlin Botanic Garden in the BeGenDiv consortium on the Illumina MiSeq platform. A 2nd PCR (indexing PCR) was performed on the purified samples to ligate a unique combination of tags to the 5′ end of the primer. The indexing PCR reactions of 25 µL were comprised of 0.25 µL dNTP mix, 1 µL DMSO, 0.625 µL of each primer, 0.25 µL of Herculase, 5 µL Herculase II reaction buffer, 10 µL of template DNA, and 7.25 µL of HPLC grade water. We started the indexing PCR regime, with denaturation at 94 °C (2 min), 8 cycles of denaturation at 95 °C (20 s), annealing at 52 °C (30 s), elongation at 72 °C (30 s), and a final elongation at 72 °C (3 min). We purified the PCR products with HighPrep PCR paramagnetic beads and quantified them with Quant-iT PicoGreen dsDNA Assay Kits (Invitrogen). We prepared the library with MiSeq Reagent Kit V3 (Illumina, San Diego, California) following manufacturer instructions, such that samples were normalized to equal nM DNA concentrations. We then pooled the samples, denatured them to 4 nM, diluted them to 20 pM, mixed them with 5 % denatured and diluted PhiX (30 µL of PhiX and 570 µL of library), and loaded them onto the MiSeq cartridge.

2.3.2. Bioinformatic pipeline for sequencing data

Demultiplexed MiSeq reads were analysed with the DADA2 pipeline (Callahan et al., 2016) by adapting the settings to diatom metabarcoding sequence data (available on https://github.com/fkeck/DADA2_diatoms_pipeline). First, primers were removed from R1 and R2 reads using cutadapt 2.9 (Martin, 2011). Then, after checking the quality profiles of R1 and R2 reads, they were truncated to 200 and 170 nucleotides, respectively, in order to remove the last, poor quality nucleotides. Truncated sequences were filtered using criteria of 0 ambiguities ("N") and a maximum of expected errors (maxEE) of 2. The DADA2 error model was executed and it showed that estimated error rates fit well to the observed rates and the error rates decreased with increased quality. Reads were then dereplicated into individual sequence units, then ASVs were selected based on the error rate models and paired reads were merged into one sequence. Chimera and then singletons were removed. Sample size normalisation was performed to 13,552 reads, which choice was confirmed by the rarefaction curves. ASVs were assigned and clustered into taxonomy using the diat.barcode (v7) reference database, following European standards for reference barcoding library management (CEN, 2018b; Keck et al., 2019; Rimet et al., 2019), with the R package "diat.barcode" (Keck, 2020), using a minimum bootstrap confidence of 60 (Table S3). In total, 68 out of 72 samples were used for subsequent analyses. The remaining four samples were excluded due to the low number of sequencing reads obtained (Table S4). Relative read abundances of taxa were then calculated and used in subsequent analyses.

2.4. Statistical analyses

All statistical analyses and data management were carried out in R version 4.2.0 (R Core Team, 2022). Summary statistics for the environmental variables were calculated including minimum, 1st quartile, median, mean, 3rd quartile, and maximum values for the Danubian reaches and the tributaries (Table S2). In order to analyse spatial autocorrelation in the data, first, a spatial weight matrix was constructed based on the k-nearest neighbours (k = 3) method, using the GPS coordinates of the sampling sites and the spdep R package. We conducted lagged spatial autoregressive linear models using the lagsarlm function of the spdep package to study the values of each environmental variable in relation to the categorical predictors of Danubian reaches and tributaries, also accounting for spatial autocorrelation. All variables except pH, alkalinity, DO, OS, temperature showed strong skewness, thus they were logarithmic transformed to ensure a normal or near-normal distribution. We used the spatial autoregressive parameter (Rho) to assess for spatial effect. We conducted the analysis using the categorical variables in two ways, as (i) Danubian sites vs. tributary sites, and (ii) the three Danubian reaches.

Principal coordinates analysis (PCoA) was run for both diatom datasets (metabarcoding, microscopy) separately, to explore and display the diatom assemblage structure based on the Bray-Curtis distance measure, using the vegan R package. PERMDISP analysis was performed to study the homogeneity of dispersions of the Danubian reaches and the tributaries using the dissimilarity matrices of the two datasets based on Bray-Curtis dissimilarity with the. Similarity of percentage (SIMPER) analysis was run to compare taxa composition of the group of sites and thus, analyse the spatial pattern of the taxa.

Diversity metrics, including Shannon diversity, richness and evenness were additionally calculated with the R package diathor v0.1.0 (Nicolosi Gelis et al., 2022). Differences of diversity metrics between methods were tested with paired Wilcoxon signed-rank tests.

Specific Pollution sensitivity Index (IPS) values were calculated for the sampling sites with the diathor R package. Values obtained for the metabarcoding ($IPS_{metabarcoding}$) and the microscopy ($IPS_{microscopy}$) dataset were compared via Spearman's rank correlation because the

variables did not meet the assumption of normality for Pearson's correlation (Shapiro-Wilk's normality test, $p < 0.05$). Quality class boundaries of IPS scores are set as 15.2, 11.6, 8.1, 4.5 for the high/good, good/moderate, moderate/poor and poor/bad thresholds, respectively (Fidlerová and Makovinská, 2021). Based on a contingency table for the five quality classes, a chi-square test was performed to further analyse the relationship of the two methods. Taxon abundances between methods were also correlated and analysed in the context of cell sizes. Cell sizes were obtained with the diathor R package, using the database and classification of Rimet and Bouchez (2012), as follows: class 1 $<$ 100 μm^3, 100 $\mu m^3 \leq$ class 2 $<$ 300 μm^3, 300 $\leq$ class 3 $\mu m^3 <$ 600 μm^3, 600 $\mu m^3 \leq$ class 4 $<$ 1500 μm^3, class 5 $\geq$ 1500 μm^3. Differences in taxon abundances between methods in relation to their class sizes were compared with Kruskal-Wallis and Dunn's post-hoc tests.

To examine the extent to which differences in abundance of the same taxa between methods might play a role, Pearson's correlation of relative abundances of shared taxa was performed. To further study the effect of size the difference of the abundances of taxa belonging to the five size classes was further tested with Kruskal-Wallis rank and Dunn's post-hoc tests.

2.5. Phylogenetic analyses

To explore the accuracy in the taxonomic assignment of the most abundant species in the metabarcoding dataset, surprisingly assigned to the marine taxon *Navicula ramosissima*, the environmental sequences (i.e., six ASVs) initially assigned to this species were placed into a phylogenetic framework. Maximum Likelihood (ML) phylogenetic tree hypotheses were constructed based on an alignment containing 94 sequences, including those for the six ASVs assigned to this taxon, complemented by sequences obtained from NCBI. Of those, 61 were annotated as belonging to the genus *Navicula*, seven as *N. ramosissima*, as well as 22 sequences to the closely related genera *Haslea* and *Seminavis*. Five sequences of the genus *Eunotia* were used as outgroups to root the phylogenetic analysis. Sequences were aligned in MUSCLE (Edgar, 2004) as implemented in EMBL–EBI (Goujon et al., 2010). A first phylogenetic tree was calculated in RAxML v8 (Stamatakis, 2014) with the substitution model GTR + G + I in the CIPRES Science Gateway (Miller et al., 2010) based on a 1098-bp alignment, containing sequences of variable length, e.g., ASVs of only 263 bp. A second tree was calculated based on the previous alignment but with all sequences trimmed to 263 bp, to match the length of the ASVs, under the same substitution model for the first tree. To infer branch support, 1,000 rapid bootstrapping iterations were run for both trees, as implemented in RAxML v8 (Stamatakis, 2014). Phylogenetic trees were visualised in FigTree v.1.4.3. (Rambaut, 2016) and only bootstrap values $\geq$ 70 are shown. Genetic distances (uncorrected p-distances) among ASVs and reference sequences annotated to this species were calculated in MEGA X (Kumar et al., 2018) based on a 263 bp alignment.

3. Results

3.1. Environmental gradients of the study sites

Raw abiotic data and summary statistics of the measured environmental parameters are presented with minimum, 1st quartile, median, mean, 3rd quartile, and maximum values, separately for the main reaches (Upper Danube, Middle Danube, Lower Danube) and the tributaries (Table S2). The spatial simultaneous autoregressive lag models showed positive and significant ($p < 0.05$) Rho values, indicating spatial clustering only for NO_3^--N and SS when the three Danube reaches were included as independent variables (Table 1). The regressions also indicate that some parameters (BOD5, TOC, TP, PO_4^{3-}-P and temperature) increased towards the Lower Danube sites, while oxygen concentration and saturation decreased. More significant effects were detected between Danube and tributary sites. Spatial autocorrelation (clustering) was detected in the case of BOD5, chl-*a*, DO, OS, DOC, NH_4^+-N, SS, temperature, TOC and TP. Additionally, a set of variables related to environmental pressures (COD, Cond, DOC, NH_4^+-N, NO_2^--N, PO_4^{3-}-P, TP, TOC) were present in higher concentrations in the tributary sites, while oxygen concentration and saturation were lower there.

3.2. Diatom data

A total of 1,599 distinct ASVs were identified in the 68 environmental samples. All ASVs could be assigned to at least Class level, 1,428 ASVs to Order and 1,314 ASVs to Family taxonomy levels. 1,248 ASVs (78 %) could be further assigned to 75 genera, of which 1,006 ASVs (63 %) could be assigned and clustered into 205 defined binomial species. The 75 identified genera correspond to 92 % of relative abundance in the total dataset and the identified 205 species correspond to 85 % of relative abundance.

Table 1
Partial results of the spatial simultaneous autoregressive lag models. The first set of analyses was conducted on each environmental variable as dependent and the Danube reaches as independent variables. The second set of variables was conducted on each environmental variable as dependent and being Danube or tributary site as the independent variable. Estimates for the Upper Danube and the Danube sites represent the reference (intercept) in the models to which the estimates on Middle and Lower Danube; and Tributaries were compared to. The models take into account the spatial factor represented by a spatial autoregressive parameter (Rho). Significant relationships ($p < 0.05$) are indicated by bold and asterisk.

Variables	Estimates			Rho	Estimates		Rho
	Upper Danube (Intercept)	Middle Danube	Lower Danube		Danube sites (Intercept)	Tributaries	
Alkalinity	1.86	−0.18	−0.29	0.38	2.80	+0.48	−0.01
lg (BOD5)	0.06	**+0.61***	**+0.63***	0.02	0.19	+0.17	**0.46***
lg (Chl-*a*)	1.45	+0.31	+0.43	0.17	1.09	+0.48	**0.43***
lg (COD)	1.71	+0.08	+0.24	0.15	1.63	**+0.46***	0.21
lg (Cond)	3.84	−0.06	−0.02	0.35	4.44	**+0.25***	0.23
DO	9.84	**−0.68***	**−1.42***	−0.12	4.47	**−0.87***	**0.48***
lg (DOC)	0.50	+0.10	**+0.48***	0.12	0.32	**+0.38***	**0.50***
lg (NH_4^+-N)	−3.14	−0.05	+0.23	0.16	−2.33	**+0.70***	**0.40***
lg (NO_2^--N)	−5.53	−0.11	+0.08	−0.14	−6.24	**+0.65***	−0.30
lg (NO_3^--N)	0.08	−0.14	−0.20	**0.67***	−0.05	−0.06	0.23
OS	116.28	−5.10	**−11.15***	−0.17	57.39	**−8.57***	**0.42***
pH	9.32	−0.23	−0.19	−0.15	7.64	+0.01	0.05
lg (PO_4^{3-}-P)	−4.13	**+0.85***	**+0.77***	0.02	−2.69	**+0.74***	0.30
lg (SS)	1.22	+0.47	+0.36	**0.53***	1.88	−0.12	**0.39***
Temp	15.67	**+2.74***	**+4.34***	0.22	10.27	+1.11	**0.54***
lg (TN)	0.24	−0.15	−0.16	0.40	0.28	+0.00	−0.22
lg (TOC)	0.58	+0.09	**+0.41***	0.22	0.37	**+0.43***	**0.51***
lg (TP)	−3.25	+0.57	**+0.54***	−0.07	−1.83	**+0.75***	**0.37***

From the morphological analyses, 360 species belonging to 76 genera were identified by light microscopy. The correspondence between microscopy and metabarcoding in terms of number of genera and species is shown in Fig. S2. The two methods found, in total, 89 genera of which 62 genera (69.7 %) are shared. 449 species were detected by the two methods of which 119 species (26.5 %) are present in both datasets. The 62 shared genera correspond to 98.4 % relative abundance of the total microscopy dataset and 91.8 % relative abundance of the total metabarcoding dataset. The 119 shared species correspond to 69.5 % relative count abundance of the total microscopy dataset and 64.3 % relative read abundance of the total metabarcoding data. Beside taxa fully identified/assigned to known species, both the microscopy and metabarcoding datasets used in subsequent analyses contained distinct species that were nonetheless not fully identified (e.g. *Navicula* sp.). Since it is not the same case as identifying them only on genus level, we kept these taxa, enlarging the microscopy and metabarcoding datasets to 385 and 244 species, respectively.

Taxonomic richness and evenness values between the two methods were significantly different ($p < 0.001$), with higher richness and lower evenness values for metabarcoding data than for microscopy data (Fig. S3A-B). Shannon index values were not different between the two methods ($p = 0.11$; Fig. S3C).

3.3. Spatial pattern of the diatom community compositions

In the case of both datasets, PCoAordinations showed similar results with forming a gradient from the Upper Danube towards the Lower Danube sites (Fig. 1). However, tributary sites do not form a distinct cluster different from communities of the Danube. PERMDISP analysis showed differences among the dispersions in the different groups (ANOVA, $p < 0.01$), in both datasets, tributary sites had a higher dispersion than the other groups.

SIMPER analysis on the Bray-Curtis dissimilarities revealed the contribution of individual taxa to the overall dissimilarity among the three reaches of the Danube (Table S5). Overall Bray-Curtis dissimilarity is 79.5 and 76.4 for the microscopy and the metabarcoding dataset, respectively. Fig. 2 shows the mean abundance in the different Danubian reaches of those taxa whose cumulative contribution to the overall dissimilarity reaches 80 %, i.e. 16 taxa for the microscopy and 19 taxa for the metabarcoding dataset, respectively. Most of these taxa were shared between both datasets, except *Navicula recens* (Lange-Bertalot) Lange-Bertalot and *Cocconeis euglypta* Ehrenberg for the microscopy data; and *Navicula ramosissima* (C.Agardh) Cleve and *Nitzschia dissipata* var. *media* (Hantzsch) Grunow for the metabarcoding data. The most contributing ones, *N. recens* and *N. ramosissima*, are unique to the microscopy and the metabarcoding data, respectively. These two taxa, however, showed a strong co-occurrence and correlation (Pearson's $r = 0.87$, $p < 0.001$). Shared taxa on this shortened list of the most important taxa with similar abundances are *Amphora pediculus* (Kützing) Grunow ex A.Schmidt (APED), *Achnanthidium delmontii* F.Pérès, R.le Cohu & A.Barthès (ADMO), *Achnanthidium minutissimum* (Kützing) Czarnecki (ADMI), *Cyclotella meneghiniana* Kützing (CMEN), *Eolimna minima* (Grunow) Lange-Bertalot & W.Schiller (EOMI), *Navicula cryptotenella* Lange-Bertalot (NCTE) and *Craticula subminuscula* (Manguin) Wetzel & Ector (CSNU). *Nitzschia palea* (Kützing) W.Smith (NPAL) and *Melosira varians* C.Agardh (MVAR) are both important in the microscopy and metabarcoding dataset, however much higher relative abundance was detected with metabarcoding. Beside some differences in relative abundances of some taxa, their abundance pattern along the three reaches are consistent.

3.4. Quality assessment of sampling sites

From 64 sites, we obtained both microscopy and metabarcoding data of diatom communities and inferred IPS quality values. A spatial pattern can be observed based on both methods with higher values at upstream sites and lower quality values at downstream sites but tributary sites have in general lower IPS value than Danubian sites (Fig. 3, Fig. S4). We found that there is a significant correlation between IPS values obtained with the two methods (Spearman's $r = 0.62$), and the chi-square test conducted on the contingency table of quality classes also showed significant relationship ($\chi 2 = 25.68$, $p < 0.01$). Important differences in quality classes were observed though (Fig. 4). The 64 samples of the microscopy dataset fall into only three quality classes, 26, 35 and 3 samples in the moderate, good and high quality classes, respectively. IPS values obtained from metabarcoding data represent all five classes with 3, 14, 19, 26 and 2 samples in the bad, poor, moderate, good and high classes, respectively. 32 samples (50 %) obtained the same quality class with the metabarcoding and the microscopy dataset. One quality class difference was observed for 25 samples (39 %), two class differences for 5 samples (8 %) and three quality class differences for 2 samples (3 %). In order to study the reasons for differences in the quality classes between methods, the community composition with species' relative abundances and their presence in the IPS database of each site, were analysed (Fig. 5, Fig. S5). Reasons for the incongruence at some sites were due to different diatom composition or/and differences in species' abundance or/and lack of IPS scores of some locally abundant species. Since diatom communities are usually composed of several rare taxa and a few strongly dominant ones, the presence or absence, or differences in the abundances of these few taxa strongly influenced the final index value in some cases.

When correlating the relative abundances of the shared taxa between the methods, a strong and significant relationship was detected (Pearson's $r = 0.57$, $p < 0.01$) (Fig. 6A). Investigating the differences in the

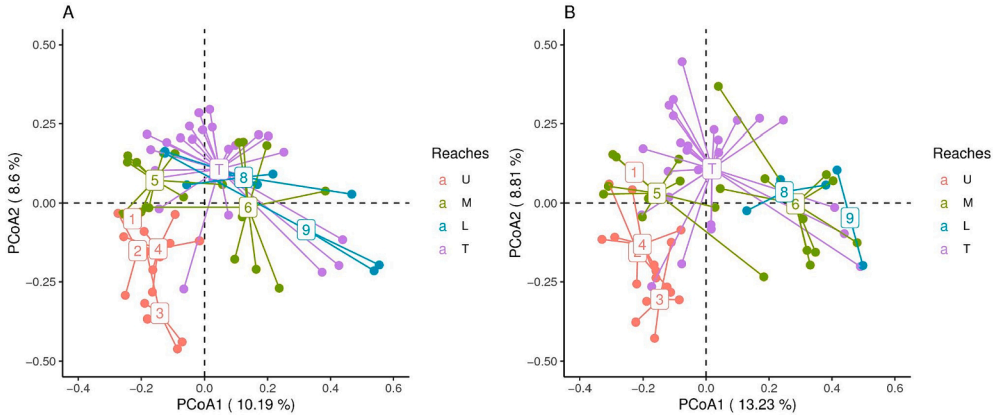


Fig. 1. PCoA ordination plots for communities detected with microscopy (A) and metabarcoding (B). Numbers indicate the Danubian types, while colours indicate Danubian reaches and tributary sites.

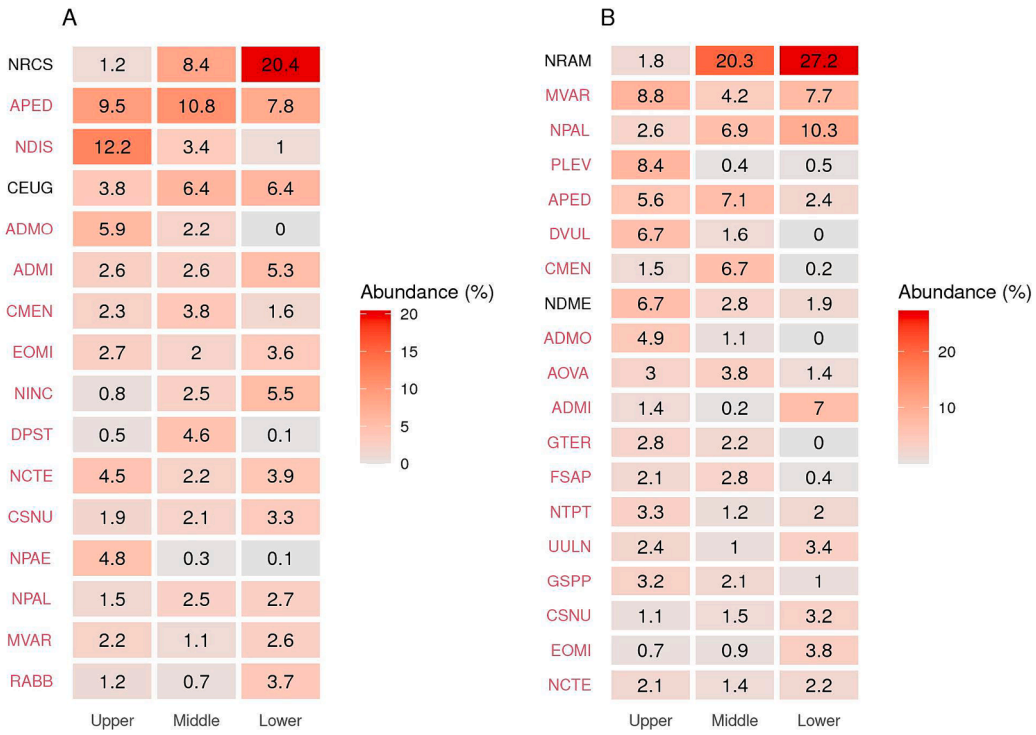


Fig. 2. Partial results from the SIMPER analysis performed separately for the microscopy (A) and the metabarcoding (B) dataset, showing the mean relative abundances of taxa in the three reaches of the Danube. Taxa are ordered in a decreasing order from top to the bottom according to their average contribution to overall dissimilarity. Only those taxa are presented whose cumulative contribution reaches 80%. Taxa in red letters are shared taxa between methods. For the taxa abbreviations, see Table S5.

abundances of taxa among the five size classes showed that the abundance of taxa belonging to classes 1–4 were underrepresented with metabarcoding compared to microscopy, while larger taxa belonging to class 5 were overrepresented with metabarcoding in terms of abundance (Fig. 6B).

3.5. *Phylogenetic placement of environmental sequences*

Phylogenetic analyses placed the six ASVs initially assigned to *N. ramosissima* by the bioinformatics pipeline in three different clades (A, B and C; bootstrap values ≥ 92 in Fig. 7; ≥78 in Fig. S6), along with reference sequences of this putative species contained in the diat.barcode database and also available in NCBI. Other available reference sequences for this species included in the phylogenetic reconstruction were placed in distinct clades in both phylogenetic trees (Figs 7, S6). Genetic distances (uncorrected p-distances) among the six ASVs varied from 0.38 to 4.56 %, and from 0 to 9.13 % among ASVs and reference sequences of *N. ramosissima* (Table S6). The most abundant ASV across the dataset (ASV_0001) was placed in a clade containing two other ASVs also assigned to this taxon (ASV_0383 and ASV_1268), together with an unidentified *Navicula* species (*Navicula* sp. HK558, MN977809) (Table S7), with genetic distances among these ASVs and reference sequence being 0.38 to 0.76 %, equivalent to only 1 to 2 bp difference.

4. Discussion

4.1. *Strong environmental gradient along the Danube river and its tributaries*

Although we analysed a relatively limited number of samples, the geographic and environmental gradients they covered were sufficient to fulfil our study objectives. We identified strong gradients in the environmental parameters measured from the water both along the length of the Danube and among the Danube and its tributaries. The most contrasting differences were found among the Danube and its tributaries where most of the parameters related to nutrients and organic matter enrichment are higher. Tributaries represent important strategic sampling points as they may be the source of potential pollution into the Danube as we could see at the Rusenski Lom tributary in Bulgaria, at the lower reach of the Danube catchment, similarly to former JDS campaigns (Liška et al., 2021). We also found that in the case of some parameters, positive spatial autocorrelation exists, i.e. sites that are closer to each other tend to have similar values. This effect should be always considered in monitoring programs.

4.2. *Diatom communities obtained by microscopy and metabarcoding analyses share abundant taxa but contain important discrepancies*

The 26.5 % and 69.7 % of shared species and genera between the identification methods are far from perfect match but the accordance is much better considering relative abundances, which match 91.8 % for the species and 98.4 % for the genera. These values correspond well with the literature where similar proportions were detected (Kulaš et al., 2022; Pérez-Burillo et al., 2022). The reason for this asymmetry between abundant and rare taxa in this aspect can have several components. First, the sampling effort in metabarcoding analysis, where it corresponds to the sequencing depth, is more efficient to find rare taxa than in the case of the microscopy counting. Although the counting of 400 valves has been set a long time ago in diatom studies to find a compromise between labour investment and representativity, rarefaction curves on the microscopy data often show undersampling (e.g. Anslan et al., 2022). Our results, similarly to some former studies (Mortágua et al., 2019; Rimet et al., 2018) detected ca. twice as many taxa per sample with metabarcoding than with microscopy. This higher richness per sample however does not necessarily mean higher total diversity, often, metabarcoding find less species in total compared to microscopy as it is shown by our results and former studies (Kulaš et al., 2022; Mora et al., 2019; Mortágua et al., 2019). It mainly depends on the

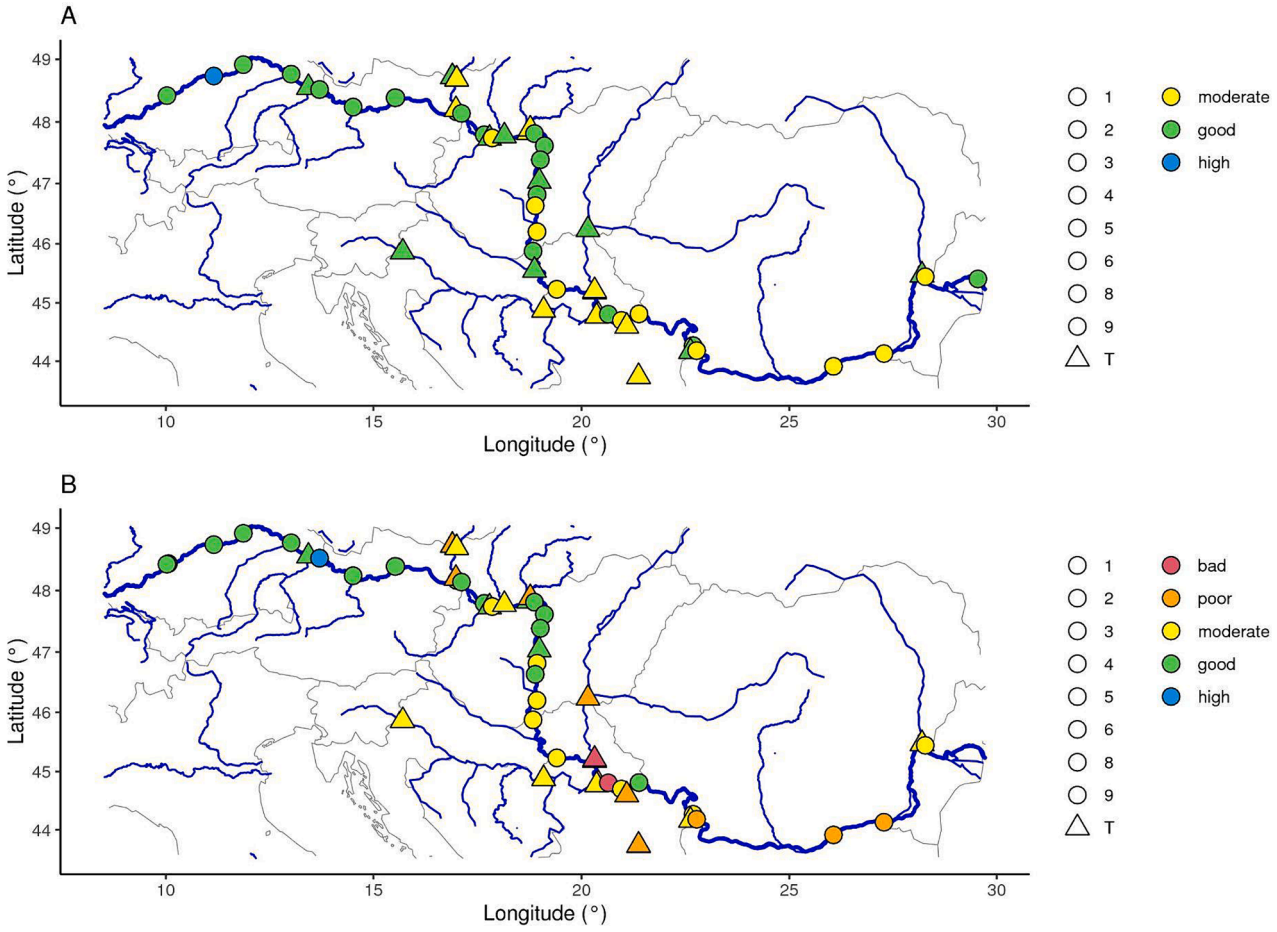


Fig. 3. Sampling sites on the Danube (1–9) and tributaries (T) coloured according to IPS quality notes, based on microscopy (A) and metabarcoding (B) data.

coverage of the reference sequence database which tends to involve sequences and species that are more abundant generally or in more studied regions (Weigand et al., 2019) and this is our second reason for the overrepresentation of abundant taxa. When a given region or habitat is understudied or possesses different diatom communities than the region which is better represented in the reference database, the proportion of successfully assigned OTUs/ASVs remains low (Rivera et al., 2018; Vasselon et al., 2017). Almost all the important taxa identified with the SIMPER analysis were shared taxa between methods, except for the most abundant ones which were identified as *Navicula recens* by microscopy and *N. ramosissima* by metabarcoding. Based on the well-correlating relative abundances of the two species from the two detection methods, we hypothesise they represent the same species, which is *N. recens* according to our microscopy identification. Our findings from the phylogenetic placement of ASVs initially assigned to this species, and genetic distance calculation support our assumption that they do not correspond to *N. ramosissima* but to three different species. This is supported by morphological observation of the two strains of *N. ramosissima* for which images are available (Table S7), which deviate from the morphologies illustrated for this species in identification monographs (Witkowski et al., 2000). Additionally, the distribution and ecology of *N. ramosissima* make its presence in the Danube unrealistic, since it is a marine species. Only the well-supported clade, labelled as A in both trees, containing the by far most abundant ASV (i.e., ASV_0001), and an unidentified *Navicula* species from a river in Florida, USA, most probably correspond to *N. recens*, a species also known to occur in North America (Potapova, 2009). This example points out that expert knowledge on the morphotaxonomy and ecological preferences is inevitable when such inaccuracies are detected. The reason for this can be that reference sequences are not correctly annotated taxonomically or sequences of *N. recens* are annotated as *N. ramosissima*. Unfortunately, as there are no sequences available for *N. recens* in any reference databases, we have no proof for our assumption. Thus, the permanent curation and the quality of the reference databases remain crucial for accurate metabarcoding analyses (Keck et al., 2023).

Other reasons for discrepancies between methods lie in different selective biases related to them. For example, species with weakly silicified frustules which can hardly stand the standard H_2O_2 treatment to prepare permanent diatom slides, tend to occur with low abundances or might be completely missing from microscopy samples while they are detected with higher abundances with metabarcoding. One typical example for this is *Fistulifera saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot, a species that was highly underrepresented in the microscopy dataset compared to the metabarcoding in this and former studies as well (e.g. Pérez-Burillo et al., 2020). Another potential bias that may occur is related to the fact that we do not know if the frustule being counted under the light microscope originates from a living cell, or it is a remnant of an already dead cell. This effect has been seldom studied but they imply that the proportion of living-to-dead frustules may strongly vary with hydromorphology, however without an important effect on bioindication (Gillett et al., 2009, 2016; Wilson and Holmes, 1981). Similarly, we can suppose that metabarcoding detects the DNA of already dead cells or free DNA being transported from elsewhere. Although these signals can potentially produce false positive results, the relative biomass of living cells is so much higher that the final assessment can be influenced only by very small chance (Vasselon et al., 2019).

Further bias might be introduced by the different ways to estimate

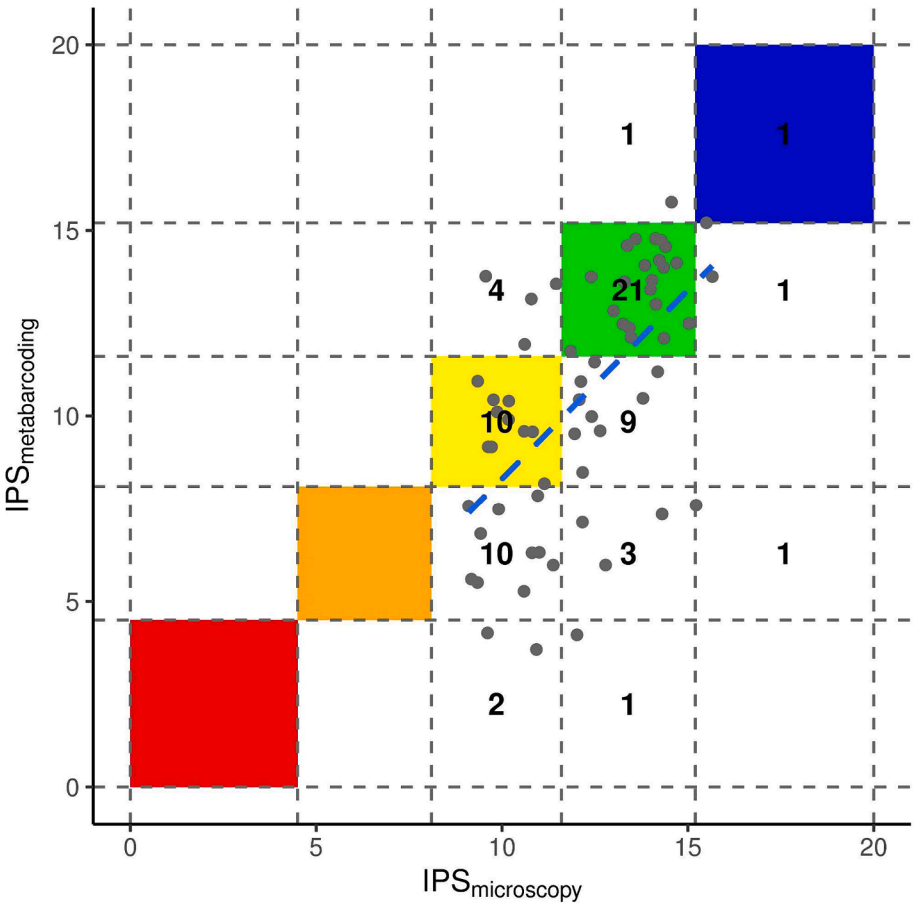


Fig. 4. Correlation of IPS values between data obtained with microscopy and metabarcoding (Spearman's $r = 0.62$, $p < 0.001$). Colours indicate quality classes, digits are the number of samples within quality classes.

relative abundances of taxa. While the conventional method uses the proportion of valves counted, in metabarcoding, we estimate abundance based on the proportion of read numbers of taxa, which is strongly related to cell biovolume (Vasselon et al., 2018). We showed that it results in the overestimation of large taxa in metabarcoding compared to microscopy. *Diatoma vulgaris* Bory de Saint-Vincent, *Melosira varians*, *Ulnaria ulna* (Nitzsch) P.Compère or *Pleurosira laevis* (Ehrenberg) Compère are characteristic examples for this phenomenon (e.g. Pérez-Burillo et al., 2022; Vasselon et al., 2018). However, given the large size plasticity of the diatom cell (Mohamad et al., 2022), its biovolume might not always correspond to genome size and gene copy numbers of the target marker (Martin et al., 2022). It remains, however, a question of debate which method reflects better the ecological relevance of a given taxon (Duarte et al., 1990; Laux and Torgan, 2015). Additionally, size can be strongly related to the identifiability of species and small celled-species, together with unstable taxonomy and hardly identifiable characteristics, it can easily hinder proper species determination (e.g. Potapova and Hamilton, 2007; Trobajo et al., 2013). Further identification biases are introduced in the case of morphologically similar distinct species where metabarcoding can give a solution (Kulaš et al., 2022). Identification biases during the microscopic analyses of samples can significantly contribute to uncertainties in bioassessment (Besse-Lototskaya et al., 2006; Kahlert et al., 2009).

4.3. IPS obtained by the different methods correlate significantly but affected by discrepancies in diatom communities

We found that in general both methods show a decrease in quality from the source towards the Danube delta and that tributary sites have in general lower IPS values than Danubian sites which was already shown in former JDS campaigns (Hlúbiková et al., 2014). The longitudinal decreasing quality of rivers is well known as the most common environmental pressures, i.e. elevated nutrient and organic matter concentrations downstream are well indicated by algal communities (Abonyi et al., 2012; Ács et al., 2003; Bellinger et al., 2013).

Although the correlation between the IPS values from the two methods correlated significantly, important differences were detected due to the differences in community composition and taxa abundances. This relationship was expected since former studies all showed significant but various correlations in different ecoregions. Although a correlation coefficient is a widely available metric and makes it easy to interpret the efficacy of our results in comparison to former ones, it is not an absolute measure to compare relationships. For instance, while Bailet et al. (2019) found an $r = 0.62$ between molecular (*rbc*L) and morphological IPS scores, only 37.5 % of the sites fell in the same ecological quality class, compared to the 50 % in this study. However the $r = 0.73$ with 63 % of sites classified into the same category (Duleba et al., 2021), the $r = 0.60$ with 56 % sites in the same classes (Mortágua et al., 2019), $r = 0.90$ with 69.8 % in the same class (Pérez-Burillo et al., 2020), $r = 0.83$ with 64 % in the same classes (Rivera et al., 2020) and $r = 0.72$ (Vasselon et al., 2017) are more or less higher values than the one we showed in our study. Reasons for these differences can be very diverse depending on the ecoregion, species inventory, the year of the study, etc. but we aim to give explanations below for the differences specific to our case.

IPS and several other autecological diatom indices are based on the idea that the average ecological value of taxa at a site, weighted by their relative abundances, represent the ecological quality of the site

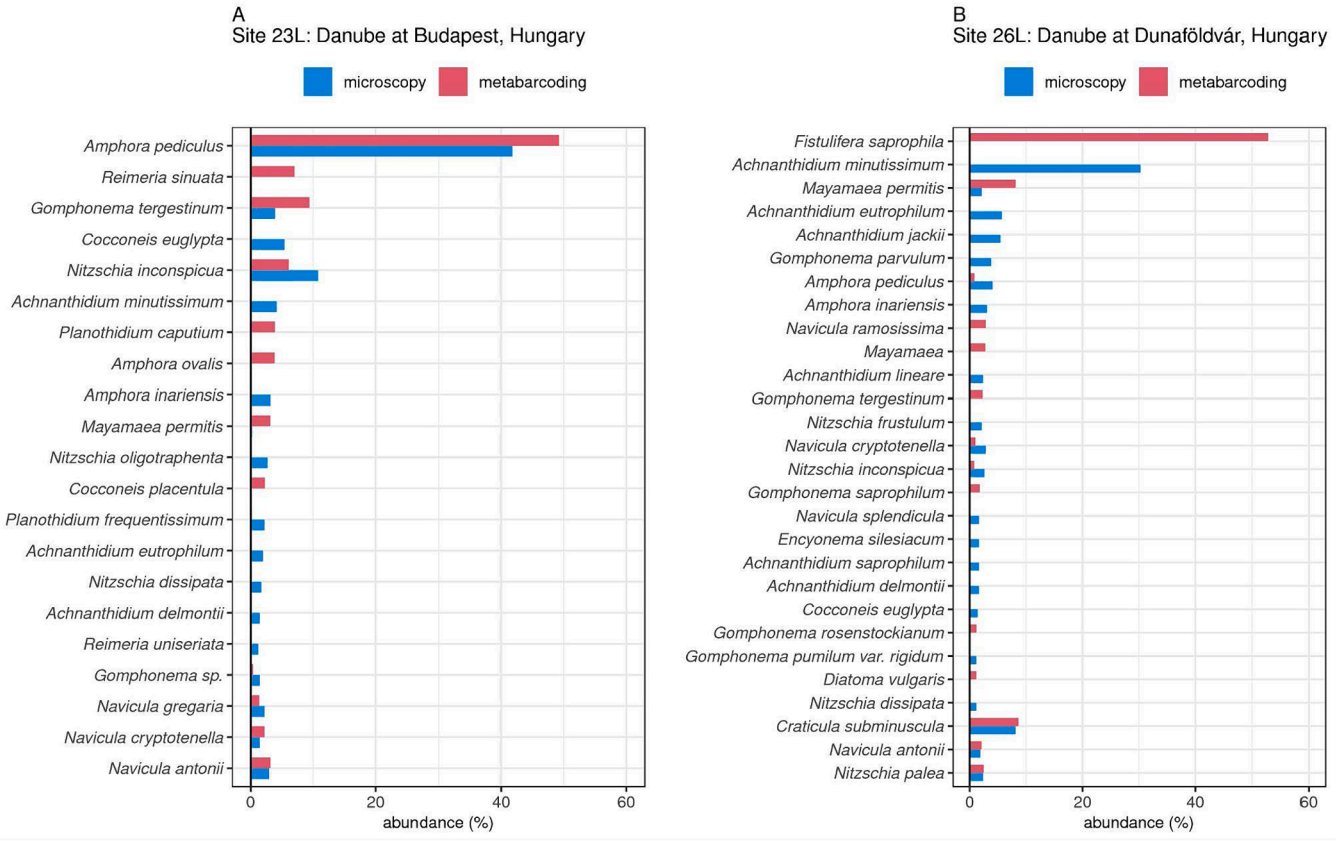


Fig. 5. Two examples of taxonomic (dis)similarity between methods. (A) Taxa composition (taxa with > 1 % abundance) of site 23L (Danube at Budapest, Hungary), which has been classified as "good quality" with both methods, and (B) site 26L (Danube at Dunaföldvár, Hungary), which has been classified as "high quality" with microscopy and "poor quality" with metabarcoding.

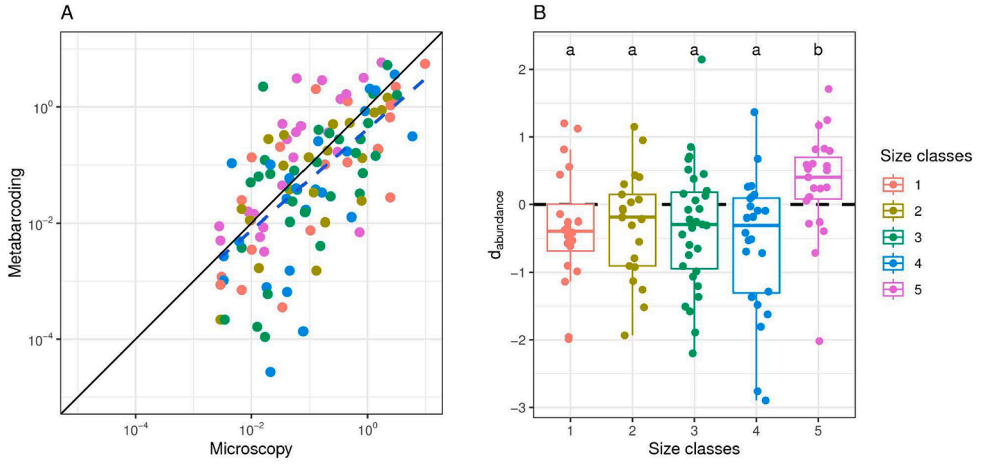


Fig. 6. (A) Correlation of shared taxa relative abundances measured by metabarcoding and microscopy ($R^2 = 0.33$, $p < 0.01$). Colours indicate the five size classes (class 1 < 100 μm^3, 100 μm^3 ≤ class 2 < 300 μm^3, 300 μm^3 ≤ class 3 < 600 μm^3, 600 μm^3 ≤ class 4 < 1500 μm^3, class 5 ≥ 1500 μm^3). (B) The difference of the $\log_{10}$ transformed abundance data (i.e. distance from slope = 1) between metabarcoding and microscopy tested with Kruskal-Wallis rank test ($p < 0.01$) and Dunn's post-hoc test. Letters above boxplots indicate significance of differences.

(Stevenson et al., 2010). These indices thus are strongly sensitive to abundant taxa while low-abundant taxa only minimally influence the final value. The overestimation of the relative abundance of large species thus leads to important differences in quality classes between methods. It was very prominent in the case of *P. laevis*, a large species with low value in IPS, indicating strong pressure. It was detected with high abundances (20–80 %) in the Moravian sites (sites 11R, 12L, 12R, 13L) and in the Váh river site (site 19R) with metabarcoding, while remained undetected or with only low abundances with microscopy. The strong underclassification of the right bank of the Dyje tributary (site 11R) by metabarcoding was due to the strong dominance of *P. laevis* (>80 %) which was however not detected by microscopy. However, knowing that this site had one of the highest values of nutrient concentrations, the result based on metabarcoding is more reliable. Our results on *P. laevis* are in accordance with the findings of Pérez-Burillo et al. (2020) who showed that this species contributed the most to

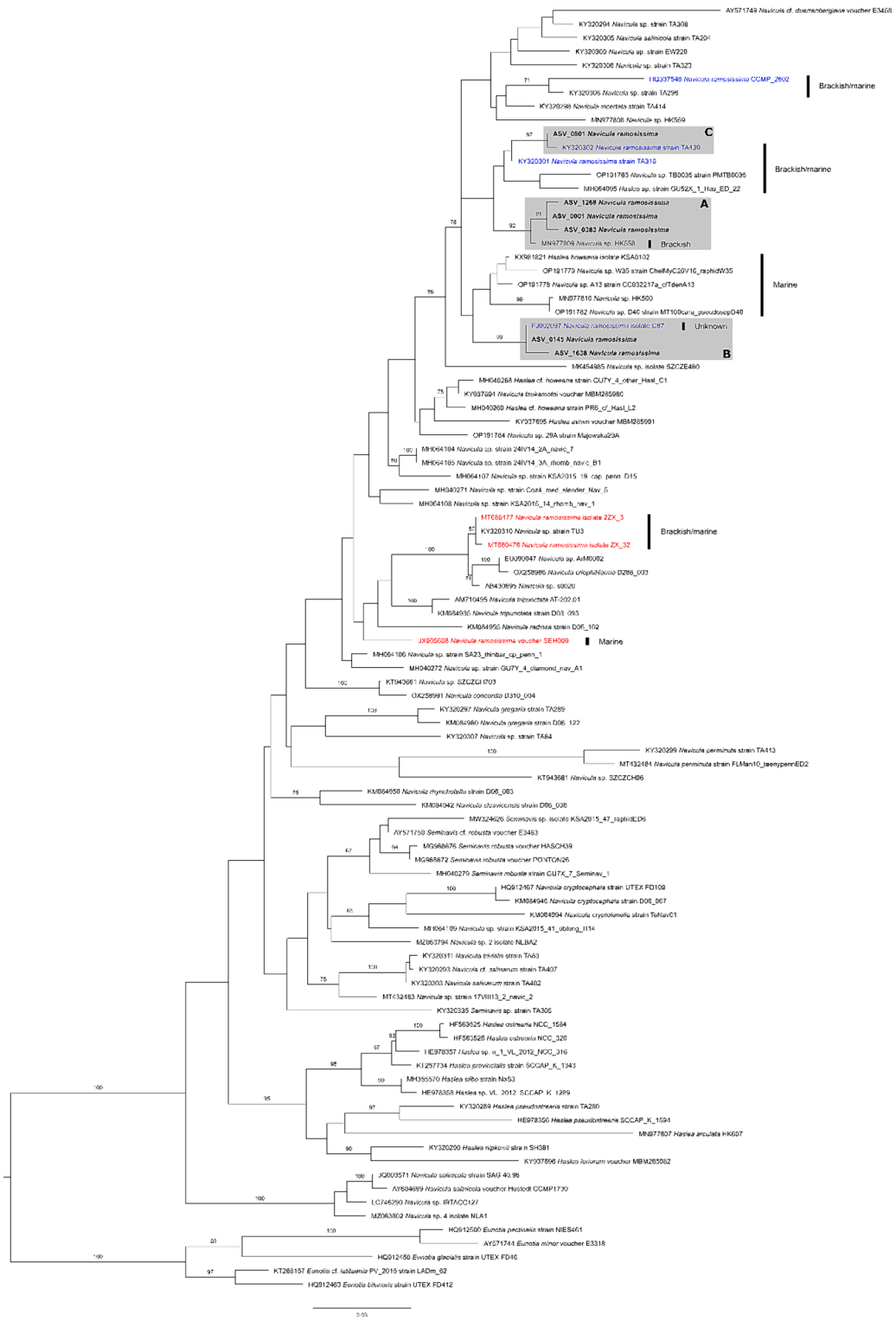


Fig. 7. Phylogenetic tree based on a 1098 position alignment of the plastid gene *rbc*L, with 89 sequences in the ingroup (*Navicula*, *Haslea* and *Seminavis*), including the six ASVs (in bold) initially assigned to *Navicula ramosissima*. The tree was rooted with five sequences of the genus *Eunotia*. The three supported clades containing the six ASVs are labelled as A, B and C. In blue are the four reference sequences of *N. ramosissima* contained in diat.barcode v7, which was used for the taxonomic assignment of environmental sequences. In red are other sequences annotated as belonging to *N. ramosissima*, available from NCBI. Indication of the origin (marine, brackish) of the strains annotated as *N. ramosissima* is given if known, as well as for other strains in neighbouring clades. Only bootstrap values ≥ 70 are shown. Scale bar = number of substitutions per site.

negatively affect IPS scores. Another contradictory site with three quality classes difference was the left bank of the danube at Dunaföldvár, Hungary (site 26L) where the microscopy dataset was dominated by *Achnanthidium minutissimum* (30 %) while *F. saprophila* highly dominated the metabarcoding dataset (>50 %). Because of the big difference in their IPS quality value (5 and 2, respectively), this resulted in a big difference in the site's IPS value as well. As both species only occurred in one of the two datasets with similarly high abundance compared to other taxa, it may suggest that the same taxon was labelled differently. However, experienced diatomist would rarely confuse the two species with each other. In the seven cases when quality score decreased two classes with the metabarcoding data, it was due to one or a few good indicators with strong abundance differences between the datasets. At the left bank of the Danube at Szob, Hungary (site 22L) microscopy analysis detected the dominance of *Cocconeis euglypta* (26.8 %) with *Nitzschia palea* (12.7 %), while metabarcoding detected 33.0 % of *N. palea* but no *C. euglypta*. The non-detection of *C. euglypta* in the metabarcoding dataset is related to the taxonomic annotation of this species' reference sequences in Diat.barcode v7, which are annotated as *Cocconeis placentula*. Later curation efforts changed the annotation to *C. euglypta* in Diat.barcode v10. This issue is caused by the widespread ambiguity in the identification of species within the *C. placentula* group, including *C. euglypta*, due to the difficulties in the observation of distinctive morphological characteristics under light microscopy. The misinterpretation of the concepts of these closely related species has led to several species being grouped under the umbrella name *C. placentula* (Jahn et al., 2020; Mora et al., 2022). Similarly, at the River Tisza in Serbia (site 33R), IPS class based on the metabarcoding data was "bad" while "moderate" for the microscopy data, because of the high abundance of *N. palea* (22.2 %), *F. saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot (5.15 %) and *Mayamaea* Lange-Bertalot spp. (14.8 %). Former studies also showed that *N. palea* is often represented by higher abundance in metabarcoding studies (Mora et al., 2019). The Tisza tributary at the Hungarian-Serbian border (site 32L) was another contradictory site with dominance of planktic species; *Aulacoseira pusilla* (F.Meister) Tuji & Houk, *Cyclostephanos invisitatus* (Hohn & Hellermann) Theriot, Stoermer & Håkasson, *Cyclotella meduanae* Germain and *Cyclotella meneghiniana* in the microscopy dataset which gives a "good" quality class. On the other hand, the two most dominant taxa in the metabarcoding data of this site, *Stephanodiscus* sp. and *Navicula ramosissima* (probably *N. recens*) have no IPS value associated and the "poor" site value mainly originates from the 3rd and 4th most abundant *Skeletonema potamos* (C.I.Weber) Hasle (IPS = 3) and *N. palea* (IPS = 1). Both *C. meneghiniana* and *N. palea* are known to have discrepancies in their abundances assessed by the two methods, leading to different quality scores (Bailet et al., 2019). The two-class difference between methods at the Danube at Banatska Palanka, Serbia (site 40L) is due to *C. meneghiniana* identified by metabarcoding (45.6 %), while only minorly with microscopy (0.7 %), and two *Amphora* species; *A. pediculus* (33.6 %) and *A. copulata* (Kützing) Schoeman & R.E.M.Archibald (19.1 %) that were detected only with microscopy.

5. Conclusion

The potentials and practical advantages of DNA metabarcoding makes it a very promising replacement method for the classical methodology. However, former studies show comparisons with very diverse results for which the reasons vary from one study area to another. As the Danube and its catchment has a great international importance in Europe, one of our main objective was to test the applicability of metabarcoding within the framework of the Joint Danube Survey, as a potential substitute method for microscopy. Based on simple comparison metrics (correlation, number of sites classified into the same quality classes), the two IPS indices had important differences and metabarcoding worked more as a complementary method than a potential replacement.

An important finding of our study is that the molecular IPS had quality notes on a wider range including poor and bad quality sites as well, especially in the tributaries. This sensitivity of the index to these sites is very important as tributaries are potential pollution sources for the Danube.

Similar comparison studies on microscopy and metabarcoding regarding diatoms usually find good accordance in the composition and abundance of the common taxa. The unexpected misassignment of the most abundant species in our dataset however warns us again that the permanent curation of the reference database is still an important challenge and metabarcoding should be still used under taxonomic expert supervision. More comprehensive and taxonomically curated DNA barcode reference libraries will constantly increase the proportion of species-level detections in the future.

The diatom indices implemented in the WFD however are sometimes criticised for their simplicity and the lack of important ecological aspects (Kelly et al., 2009; Schneider et al., 2017). Today, diatom DNA metabarcoding can provide good quality, high resolution genetic data using a barcode of the *rbc*L marker gene with good taxonomic coverage with a well-curated reference sequence database with more than 8,000 entries up to date. Yet, issues concerning measures of abundance in connection to cell size need to be tackled. We suggest that molecular indices should profit from the benefits the method can provide, e.g. high sequencing depth and the detection of high diversity in a potentially standardised and comparable way among studies. Additionally, finding other barcodes for a wider taxonomic coverage of the phytobenthos, metabarcoding would not be limited to diatoms but could give a more holistic expanded view on the algal biofilm.

CRediT authorship contribution statement

Kálmán Tapolczai: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Teofana Chonova:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation. **Dana Fidlerová:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation. **Jarmila Makovinská:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation. **Demetrio Mora:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alexander Weigand:** Writing – review & editing, Validation, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Jonas Zimmermann:** Writing – review & editing, Validation, Project administration, Methodology, Investigation, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank the International Commission for the Protection of the Danube River (ICPDR), the several JDS4 expert and task groups, and the national fieldwork teams for the organisation and management of JDS4, furthermore the collection of the samples analysed in this comparative study. The authors are grateful to Jana Bansemer and Juliane Bettig for the work in the molecular lab at the BGBM Berlin. Apart from the fundings to carry out the JDS4 campaign by ICPDR, this study was partly

supported by the RRF-2.3.1-21-2022-00008 project, the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA) and the NKFIH project FK146760.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2024.111883.

References

Abonyi, A., Leitão, M., Lançon, A.M., Padisák, J., 2012. Phytoplankton functional groups as indicators of human impacts along the river loire (France). In: Salmaso, N., Naselli-Flores, L., Cerasino, L., Flaim, G., Tolotti, M., Padisák, J. (Eds.), Phytoplankton Responses to Human Impacts at Different Scales. Springer, Netherlands, Dordrecht, pp. 233–249. https://doi.org/10.1007/978-94-007-5790-5_18.

Ács, É., Szabó, K., Kiss, K.T., Hindák, F., 2003. Benthic algal investigations in the Danube river and some of its main tributaries from Germany to Hungary. Biol.-Bratisl. 58, 545–554.

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. Front. Earth Sci. 10.

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. Mol. Ecol. Resour. 17, 1231–1242. https://doi.org/10.1111/1755-0998.12668.

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

Bailet, B., Bouchez, A., Franc, A., Frigerio, J.-M., Keck, F., Karjalainen, S.-M., Rimet, F., Schneider, S., Kahlert, M., 2019. Molecular versus morphological data for benthic diatoms biomonitoring in northern Europe freshwater and consequences for ecological status. Metabarcoding Metagenomics 3, e34002.

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. Sci. Total Environ. 745, 140948 https://doi.org/10.1016/j.scitotenv.2020.140948.

Bellinger, B.J., Angradi, T.R., Bolgrien, D.W., Jicha, T.M., Hill, B.H., Reavie, E.D., 2013. Longitudinal variation and response to anthropogenic stress in diatom assemblages of the lower Mississippi River, USA. River Syst 21, 29–54.

Besse-Lototskaya, A., Verdonschot, P.F.M., Sinkeldam, J.A., 2006. Uncertainty in diatom assessment: sampling, identification and counting variation. In: Furse, M.T., Hering, D., Brabec, K., Buffagni, A., Sandin, L., Verdonschot, P.F.M. (Eds.), The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods, Developments in Hydrobiology. Springer, Netherlands, Dordrecht, pp. 247–260. https://doi.org/10.1007/978-1-4020-5493-8_17.

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

CEN, 2014a. Water quality - Guidance for the routine sampling and preparation of benthic diatoms from rivers and lakes (No. EN 13946).

CEN, 2014b. Water quality - Guidance for the identification and enumeration of benthic diatom samples from rivers and lakes (No. EN 14407).

CEN, 2018a. Water quality - CEN/TR 17245 - Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses.

CEN, 2018b. Water quality - CEN/TR 17244:2018 Technical report for the management of diatom barcodes.

Charles, D.F., Kelly, M.G., Stevenson, R.J., Poikane, S., Theroux, S., Zgrundo, A., Cantonati, M., 2021. Benthic algae assessments in the EU and the US: striving for consistency in the face of great ecological diversity. Ecol. Indic. 121, 107082 https://doi.org/10.1016/j.ecolind.2020.107082.

Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A., Chariton, A., Creer, S., Frühe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J., Lanzén, A., 2021. Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. Mol. Ecol. 30, 2937–2958. https://doi.org/10.1111/mec.15472.

Coste, M., 1982. Étude des méthodes biologiques d'appréciation quantitative de la qualité des eaux. Rapp, Cemagref QE Lyon-AF Bassin Rhône Méditerranée Corse.

Descy, J.-P., Coste, M., 1991. A test of methods for assessing water quality based on diatoms. SIL Proc. 1922–2010 (24), 2112–2116. https://doi.org/10.1080/03680770.1989.11899905.

Duarte, C.M., Marrasé, C., Vaqué, D., Estrada, M., 1990. Counting error and the quantitative analysis of phytoplankton communities. J. Plankton Res. 12, 295–304.

Duleba, M., Földi, A., Micsinai, A., Várbíró, G., Mohr, A., Sipos, R., Szabó, G., Buczkó, K., Trábert, Z., Kiss, K.T., Bíró, T., Vadkerti, E., Ács, É., 2021. Applicability of diatom metabarcoding in the ecological status assessment of Hungarian lotic and soda pan habitats. Ecol. Indic. 130, 108105 https://doi.org/10.1016/j.ecolind.2021.108105.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Commission, European, 2000. Directive 2000/60/EC of the European Parliament and of the council of 23rd october 2000 establishing a framework for community action in the field of water policy. Off. J. Eur. Communities 327, 1–72.

Feio, M.J., Serra, S.R.Q., Mortágua, A., Bouchez, A., Rimet, F., Vasselon, V., Almeida, S.F. P., 2020. A taxonomy-free approach based on machine learning to assess the quality of rivers with diatoms. Sci. Total Environ. 722, 137900 https://doi.org/10.1016/j.scitotenv.2020.137900.

Fidlerová, D., Makovinská, J., 2021. Phytobenthos, in: Joint Danube Survey 4 Scientific Report: A Shared Analysis of the Danube River. International Commission for the Protection of the Danube River, pp. 65–72.

Gillett, N.D., Oudsema, M.E., Steinman, A.D., 2016. Live diatoms as indicators of urban stormwater runoff. Environ. Monit. Assess. 189, 37. https://doi.org/10.1007/s10661-016-5747-8.

Gillett, N., Pan, Y., Parker, C., 2009. Should only live diatoms be used in the bioassessment of small mountain streams? Hydrobiologia 620, 135–147. https://doi.org/10.1007/s10750-008-9624-5.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., Lopez, R., 2010. A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Res. 38, W695–W699. https://doi.org/10.1093/nar/gkq313.

Heino, J., Virkkala, R., Toivonen, H., 2009. Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. Biol. Rev. 84, 39–54. https://doi.org/10.1111/j.1469-185X.2008.00060.x.

Hlúbiková, D., Makovinská, J., Fidlerová, D., 2014. Phytobenthos, in: Joint Danube Survey 3: A Comprehensive Analyses of Danube Water Quality. International Commission for the Protection of the Danube River, pp. 100–109.

Jahn, R., Abarca, N., Kusber, W.-H., Skibbe, O., Zimmermann, J., Mora, D., Jahn, R., Abarca, N., Kusber, W.-H., Skibbe, O., Zimmermann, J., Mora, D., 2020. Integrative taxonomic description of two new species of the <italic>Cocconeis placentula</italic> group (Bacillariophyceae) from Korea based on unialgal strains. Algae 35, 303–324. https://doi.org/10.4490/algae.2020.35.8.1.

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., Simola, H., Sundberg, I., Vilbaste, S., Weckström, J., 2009. Harmonization is more important than experience—results of the first nordic-Baltic diatom intercalibration exercise 2007 (stream monitoring). J. Appl. Phycol. 21, 471–482. https://doi.org/10.1007/s10811-008-9394-5.

Kahlert, M., Bailet, B., Chonova, T., Karjalainen, S.M., Schneider, S.C., 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. Ecol. Indic. 130, 108088 https://doi.org/10.1016/j.ecolind.2021.108088.

Keck, F., Couton, M., Altermatt, F., 2023. Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Mol. Ecol. Resour. 23, 742–755. https://doi.org/10.1111/1755-0998.13746.

Keck, F., Rimet, F., Vasselon, V., Bouchez, A., 2019. A ready-to-use database for DADA2: Diat.barcode_rbcL_312bp_DADA2. https://doi.org/10.15454/HNI1EK.

Keck, F., 2020. diatbarcode: Access the diat.barcode database. R package version 0.0.0.9000.

Kelly, M., 2013. Data rich, information poor? phytobenthos assessment and the water framework directive. Eur. J. Phycol. 48, 437–450. https://doi.org/10.1080/09670262.2013.852694.

Kelly, M.G., Juggins, S., Mann, D.G., Sato, S., Glover, R., Boonham, N., Sapp, M., Lewis, E., Hany, U., Kille, P., Jones, T., Walsh, K., 2020. Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding. Ecol. Indic. 118, 106725 https://doi.org/10.1016/j.ecolind.2020.106725.

Kelly, M., King, L., Ní Chatháin, B., 2009. The conceptual basis of ecological-status assessments using diatoms. Biol. Environ. Proc. r. Ir. Acad. 109, 175–189. https://doi.org/10.3318/BIOE.2009.109.3.175.

Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Humbert, J.F., Bouchez, A., 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. Mol. Ecol. Resour. 13, 607–619. https://doi.org/10.1111/1755-0998.12105.

Kulaš, A., Udovič, M.G., Tapolczai, K., Žutinić, P., Orlić, S., Levkov, Z., 2022. Diatom eDNA metabarcoding and morphological methods for bioassessment of karstic river. Sci. Total Environ. 154536 https://doi.org/10.1016/j.scitotenv.2022.154536.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Laux, M., Torgan, L.C., 2015. Which metric to choose? differences between abundance and biomass responses to environmental conditions in a planktonic diatom community. Hydrobiologia 744, 63–76. https://doi.org/10.1007/s10750-014-2056-5.

Liška, I., Wagner, F., Sengl, M., Deutsch, K., Slobodník, J., Paunovic, M., 2021. Joint Danube Survey 4 Scientific Report: A Shared Analysis of the Danube River. Austria Vienna.

Liška, I., 2015. Managing an International River Basin Towards Water Quality Protection: The Danube Case, in: Liška, I. (Ed.), The Danube River Basin, The Handbook of Environmental Chemistry. Springer, Berlin, Heidelberg, pp. 1–19. https://doi.org/10.1007/698_2015_388.

Martin, J.L., Santi, I., Pitta, P., John, U., Gypens, N., 2022. Towards quantitative metabarcoding of eukaryotic plankton: an approach to improve 18S rRNA gene copy number bias. Metabarcoding Metagenomics 6, e85794.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. Doi: 10.14806/ej.17.1.200.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). Presented at the 2010 Gateway Computing Environments Workshop (GCE), pp. 1–8. Doi: 10.1109/GCE.2010.5676129.
Mohamad, H., Mora, D., Skibbe, O., Abarca, N., Deutschmeyer, V., Enke, N., Kusber, W.-H., Zimmermann, J., Jahn, R., 2022. Morphological variability and genetic marker stability of 16 monoclonal pennate diatom strains under medium-term culture. Diatom Res. 37, 307–328. https://doi.org/10.1080/0269249X.2022.2141346.
Moog, O., Sommerhäuser, M., Robert, S., Battisti, T., Birk, S., Hering, D., Ofenböck, T., Schmedtje, U., Schmidt-Kloiber, A., Vogel, B., 2006. Typology of the Danube River based on „top-down “and „bottom-up “approaches, in: Austrian Committee Danube Research. 36 Th Conference International Association for Danube Research. Vienna.
Mora, D., Abarca, N., Proft, S., Grau, J.H., 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. Freshw. Sci. 000–000 https://doi.org/10.1086/704827.
Mora, D., Stancheva, R., Jahn, R., 2022. Cocconeis czarneckii sp. nov. (Bacillariophyta): a new diatom species from Lake Okoboji (Iowa, USA), based on the strain UTEX FD23. Phycologia 61, 60–74. https://doi.org/10.1080/00318884.2021.2003684.
Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., João Feio, M., Almeida, F.P., S.,, 2019. Applicability of DNA metabarcoding approach in the bioassessment of portuguese rivers using diatoms. Ecol. Indic. 106, 105470 https://doi.org/10.1016/j.ecolind.2019.105470.
Nicolosi Gelis, M.M., Sathicq, M.B., Jupke, J., Cochero, J., 2022. DiaThor: R package for computing diatom metrics and biotic indices. Ecol. Model. 465, 109859 https://doi.org/10.1016/j.ecolmodel.2021.109859.
Pawlowski, J., Apothéloz-Perret-Gentil, L., Mächler, E., Altermatt, F., 2020. Environmental DNA applications for biomonitoring and bioassessment in aquatic ecosystems. Doi: 10.5167/UZH-187800.
Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M.J., Filipe, A.F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan Jones, J., Sagova-Mareckova, M., Moritz, C., Barquín, J., Piggott, J.J., Pinna, M., Rimet, F., Rinkevich, B., Sousa-Santos, C., Specchia, V., Trobajo, R., Vasselon, V., Vitecek, S., Zimmerman, J., Weigand, A., Leese, F., Kahlert, M., 2018. The future of biotic indices in the ecogenomic era: integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637–638, 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002.
Pérez-Burillo, J., Trobajo, R., Vasselon, V., Rimet, F., Bouchez, A., Mann, D.G., 2020. Evaluation and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of Mediterranean rivers. Sci. Total Environ. 727, 138445 https://doi.org/10.1016/j.scitotenv.2020.138445.
Pérez-Burillo, J., Valoti, G., Witkowski, A., Prado, P., Mann, D.G., Trobajo, R., 2022. Assessment of marine benthic diatom communities: insights from a combined morphological–metabarcoding approach in Mediterranean shallow coastal waters. Mar. Pollut. Bull. 174, 113183 https://doi.org/10.1016/j.marpolbul.2021.113183.
Potapova, M., Hamilton, P.B., 2007. Morphological and ecological Variation within the achnanthidium minutissimum (bacillariophyceae) species Complex1. J. Phycol. 43, 561–575. https://doi.org/10.1111/j.1529-8817.2007.00332.x.
Potapova, M., 2009. Navicula recens, in: Diatoms of North America. Retrieved November 15, 2023, from http://diatoms.org/species/navicula_recens.
Prygiel, J., Coste, M., 1993. The assessment of water quality in the Artois-Picardie water basin (France) by the use of diatom indices. Hydrobiologia 269, 343–349. https://doi.org/10.1007/BF00028033.
Rambaut, A., 2016. FigTree.
Rimet, F., Bouchez, A., 2012. Life-forms, cell-sizes and ecological guilds of diatoms in european rivers. Knowl. Manag. Aquat. Ecosyst. 01 https://doi.org/10.1051/kmae/2012018.
Rimet, F., Vasselon, V.A., Keszte, B., Bouchez, A., 2018. Do we similarly assess diversity with microscopy and high-throughput sequencing? case of microalgae in lakes. Org. Divers. Evol. 1–12 https://doi.org/10.1007/s13127-018-0359-5.
Rimet, F., Gusev, E., Kahlert, M., Kelly, M.G., Kulikovskiy, M., Maltsev, Y., Mann, D.G., Pfannkuchen, M., Trobajo, R., Vasselon, V., Zimmermann, J., Bouchez, A., 2019. Diat.barcode, an open-access curated barcode library for diatoms. Sci. Rep. 9, 1–12. https://doi.org/10.1038/s41598-019-51500-6.
Rivera, S.F., Vasselon, V., Jacquet, S., Bouchez, A., Ariztegui, D., Rimet, F., 2018. Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment. Hydrobiologia 807, 37–51. https://doi.org/10.1007/s10750-017-3381-2.
Rivera, S.F., Vasselon, V., Bouchez, A., Rimet, F., 2020. Diatom metabarcoding applied to large scale monitoring networks: optimization of bioinformatics strategies using mothur software. Ecol. Indic. 109, 105775 https://doi.org/10.1016/j.ecolind.2019.105775.
Schneider, S.C., Hilt, S., Vermaat, J.E., Kelly, M., 2017. The “forgotten” ecology behind ecological status evaluation: re-assessing the roles of aquatic plants and benthic algae in ecosystem functioning. In: Cánovas, F.M., Lüttge, U., Matyssek, R. (Eds.), Progress in Botany, Vol. 78. Progress in Botany. Springer International Publishing, Cham, pp. 285–304. https://doi.org/10.1007/124_2016_7.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stevenson, R.J., Pan, Y., Van Dam, H., 2010. Assessing environmental conditions in rivers and streams with diatoms. In: The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press, Cambridge, pp. 57–85.
Tapolczai, K., Selmeczy, G.B., Szabó, B., B-Béres, V., Keck, F., Bouchez, A., Rimet, F., Padisák, J., 2021. The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. Ecol. Indic. 122, 107322. Doi: 10.1016/j.ecolind.2020.107322.
Tickner, D., Opperman, J.J., Abell, R., Acreman, M., Arthington, A.H., Bunn, S.E., Cooke, S.J., Dalton, J., Darwall, W., Edwards, G., Harrison, I., Hughes, K., Jones, T., Leclère, D., Lynch, A.J., Leonard, P., McClain, M.E., Muruven, D., Olden, J.D., Ormerod, S.J., Robinson, J., Tharme, R.E., Thieme, M., Tockner, K., Wright, M., Young, L., 2020. Bending the curve of global freshwater biodiversity loss: an Emergency recovery plan. BioScience 70, 330–342. https://doi.org/10.1093/biosci/biaa002.
Trobajo, R., Rovira, L., Ector, L., Wetzel, C.E., Kelly, M., Mann, D.G., 2013. Morphology and identity of some ecologically important small nitzschia species. Diatom Res. 28, 37–59. https://doi.org/10.1080/0269249X.2012.734531.
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 Ecol. Evol. Doi: 10.1111/2041-210X.12960
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). Ecol. Indic. 82, 1–12. https://doi.org/10.1016/j.ecolind.2017.06.024.
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 Metagenomics 3, e39646.
Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 467, 555–561. https://doi.org/10.1038/nature09440.
Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A.M., Willassen, E., Wyler, S.A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., Dijkstra, K.-D.-B., Eisendle, U., Freyhof, J., Gadawski, P., Graf, W., Haegerbaeumer, A., van der Hoorn, B.B., Japoshvili, B., Keresztes, L., Keskin, E., Leese, F., Macher, J.N., Mamos, T., Paz, G., Pešić, V., Pfannkuchen, D.M., Pfannkuchen, M.A., Price, B.W., Rinkevich, B., Teixeira, M.A.L., Várbíró, G., Ekrem, T., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: gap-analysis and recommendations for future work. Sci. Total Environ. 678, 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247.
Wilson, C.J., Holmes, R.W., 1981. The ecological importance of distinguishing between living and dead diatoms in estuarine sediments. Br. Phycol. J. 16, 345–349. https://doi.org/10.1080/00071618100650371.
Witkowski, A., Lange-Bertalot, H., Metzeltin, D., 2000. Diatom flora of marine coasts I. Iconographia Diatomologica 7. Rugell. ARG Gantner Verlag KG.
Zelinka, M., Marvan, P., 1961. Zur präzisierung der biologischen klassifikation der reinheit fliessender gewässer. Arch Hydrobiol 57, 389–407.
Zimmermann, J., Mora, D., Tapolczai, K., Proft, S., Chonova, T., Rimet, F., Bouchez, A., Fidlerová, D., Makovinská, J., Weigand, A., 2021. Metabarcoding of phytobenthos samples. In: Joint Danube Survey 4 Scientific Report: A Shared Analysis of the Danube River. International Commission for the Protection of the Danube River, pp. 145–156.

Further reading

Chardon, C., Vasselon, V., Vautier, M., Lacroix, S., Bouchez, A., Rimet, F., Domaizon, I., 2020. Diatom DNA library preparation for Illumina Miseq Sequencing using the rbcL marker gene [WWW Document]. protocols.io. URL https://www.protocols.io/view/diatom-dna-library-preparation-for-illumina-miseq-bd94i98w (accessed 2.28.22).