Original Articles

Same same, but different: The response of diatoms to environmental gradients in Fennoscandian streams and lakes – barcodes, traits and microscope data compared

Maria Kahlert [a,*], Bonnie Bailet [a], Teofana Chonova [b], Satu Maaria Karjalainen [c], Susanne C. Schneider [d], Kálmán Tapolczai [e]

[a] *Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment, PO Box 7050, SE – 750 07 Uppsala, Sweden*
[b] *INRAE, UMR CARRTEL, Thonon-les-Bains F-74200, France*
[c] *Finnish Environment Institute, Paavo Havaksentie 3, 90570 Oulu, Finland*
[d] *Norwegian Institute for Water Research, Gaustadalleen 21, 0349 Oslo, Norway*
[e] *Premium Postdoctoral Research Program, Hungarian Academy of Sciences, Budapest, Hungary, Research Group of Limnology, Center for Natural Science, University of Pannonia, Egyetem u. 10, H-8200 Veszprém, Hungary*

ARTICLE INFO

ABSTRACT

We developed and compared the performance of freshwater benthic diatom indices calculated from (i) traditional morphological species identification, (ii) Amplicon Sequence Variants (ASVs) obtained via DNA metabarcoding, and (iii) morphological traits to indicate eutrophication in rivers and lakes in Fennoscandia. Based on the results, we provided recommendations for the future routine use of diatom bioassessment tools in environmental monitoring and assessment. Our results show that ASVs are the most promising candidates to be used in environmental assessment. Indices based on ASVs correlated better with TotP concentrations than morphological taxa data, whereas the trait indices correlated least. We could see by studying the taxonomic assignments of the ASVs that traditional morphotaxa were divided up into several ASVs with different ecological profiles, which explained part of the better index performance and also encourages further studies on diatom diversity and ecological preferences. In general, ASV- and morphotaxon-specific optima differed slightly between streams and lakes, but were significantly correlated with each other. This means that it should be possible to develop a common index that is applicable in both streams and lakes, but boundary values with respect to TotP might need to be set separately for them. More knowledge on diatom traits is required to enable their use for environmental assessment.

1. Introduction

Freshwater benthic diatoms have been used in aquatic environmental assessment for >100 years (Kolkwitz and Marsson, 1908). Today, diatoms are part of standard bioassessment toolkits (Charles et al., 2021), for example for monitoring of water bodies in accordance with the United States Clean Water Act (1972) or the European Water Framework Directive (WFD; European Parliament and Council, 2000). Diatom indices have been developed for manifold pressures, including, for example, acidification or salinization (Smol and Stoermer, 2010). However, indices indicating eutrophication are most frequently used, likely because eutrophication still is considered the most common stressor in freshwaters (Carvalho et al., 2019; European Environment Agency, 2018; Poikane et al., 2021).

Diatom indices which currently are in use for freshwater monitoring are based on inventories of the taxa that occur at the sampling sites. Traditionally, light microscopy is used to identify diatoms by morphological characters, and to quantify the relative abundance of different taxa (Smol and Stoermer, 2010). This process is, however, time-consuming and correct species identification is difficult (Alers-García et al., 2021; Hicks et al., 2006; Mann, 1999). Even experts struggle with species identification because differences between species often are subtle, and the description of characters used for species differentiation may be vague or overlapping (Kahlert et al., 2019; Li et al., 2018; Mann, 1999). Intercalibration exercises have shown that differences in species identification among laboratories were due to misidentifications, but also because laboratories have established laboratory-internal

---

conventions on how to handle difficult and closely related taxa (Kahlert et al., 2009; Kahlert et al., 2012; Werner et al., 2016). Consequently, taxonomic consistency is a problem both within and between countries (Charles et al., 2021), and challenges related to diatom identification contribute more to the variability in reported diatom assemblages than differences in methods used for diatom sampling, sample preparation or the way of counting (Charles et al., 2021).

In contrast, metabarcoding is considered an objective method for diatom identification, because this DNA-based approach is automated, such that different laboratories should produce the same results if using the same methods, at potentially greater speed and precision than traditional diatom determination and counting (Pawlowski et al., 2018). The method has been benchmarked (Charles et al., 2021), and especially the *rbc*L marker has been shown to reflect diatom genetic diversity well, producing similar results as morphological species determination (Bailet et al., 2019; Kelly et al., 2020). Whereas earlier studies have used operational taxonomic units (OTU), that is, clusters of sequence variants, as a standard taxonomic unit of a barcoding analysis, recent studies have switched to the use of amplicon sequence variants (ASVs). In contrast to OTUs, ASVs are generated by applying a denoising algorithm that detects and eliminates sequences that are more likely to contain errors. ASVs produce reproducible and comprehensive results of very highly resolved diversity where even one nucleotide difference between barcodes can be detected (Callahan et al., 2017). Recent results have shown that ecological profiles can be assigned to genetic units in a similar way as to morphological taxa, and that this knowledge can be used to design molecular indices (Kelly et al., 2020; Pawlowski et al., 2018; Tapolczai et al., 2019).

Another alternative approach to the morphological species-based indices is to study the usability of trait-based diatom indices. Using easily measurable morphological traits such as size or growth form could potentially overcome the challenges of diatom identification as well. Furthermore, a trait-based index could potentially be used in a wider geographical region than a species-based index, thereby overcoming differences in taxa composition and environmental conditions that commonly complicate the intercalibration of species-specific indices (Soininen et al., 2016; Tapolczai et al., 2017). Additionally, traits may provide ecological explanations for how organisms respond to environmental constraints (Statzner and Bêche, 2010).

Whereas the use of ecophysiological traits, e.g. the ability of a species to use a resource or cope with a certain stressor (Violle et al., 2007), has a long history in diatom monitoring (e.g. Cemagref, 1982, Kelly, 1998; Kelly and Whitton, 1995; Kolkwitz and Marsson, 1908; Rott et al., 1999), the knowledge and use of morphological traits is relatively new for diatoms compared to other organism groups (Larras et al., 2021; Kruk et al., 2010; Statzner and Bêche, 2010). The few described diatom traits are compiled in Rimet and Bouchez (2012), and mostly refer to the size or growth form of diatom species. A few earlier studies indeed suggested that certain morphological traits of diatom species assemblages change with nutrient availability: the abundance of motile (fast moving) (Passy, 2007, Berthon et al., 2011, Soininen et al., 2016; Tapolczai et al., 2017) and high-profile (Passy, 2007) taxa increased with nutrient availability, whereas the abundance of low-profile (Berthon et al., 2011; Passy, 2007; Tapolczai et al., 2017) and small-sized taxa (B-Béres et al., 2016; Cattaneo et al., 1997; Lange et al., 2016) decreased. Potapova and Snoeijs (1997) have shown that the surface to volume ratio of diatom populations was high when resource uptake needed to be high. Furthermore, large taxa generally had a stronger response to environmental variables than small taxa (Snoeijs et al., 2002).

The present study was performed in Northern Europe which is not only a relatively understudied region regarding its diatom taxonomy, with large knowledge gaps about the presence, distribution and ecology of species (Bailet et al., 2019). Fennoscandia also presents an environmental gradient which potentially could have an impact on the diatom assemblage response to other stressors such as nutrients. Fennoscandia does not only cover a gradient of nutrients, conductivity and pH, but also a steep west-east gradient in the concentration of organic matter, from clear waters in Norway to colored waters in Sweden and Finland, caused by different humus outputs from the soils due to differences in climate, soil and vegetation type (Löfgren et al., 2003). Whereas some theoretical concepts have been suggested for the distribution of diatom traits due to nutrients, flow disturbance or toxins (Passy, 2007; Rimet and Bouchez, 2011; Rimet and Bouchez, 2012), the impact of color has so far not received much attention.

Here, we develop and compare the performance of indices calculated from (i) traditional morphological species identification, (ii) ASVs obtained via DNA metabarcoding, and (iii) morphological traits to indicate eutrophication in rivers and lakes in Fennoscandia. Based on the results, we provide recommendations for the future routine use of diatom bio-assessment tools in environmental monitoring and assessment.

2. Material & methods

2.1. Study sites

.

2.2. Collection of data

Data were collected from 76 stream and 39 lake sites in Fennoscandia (Sweden, Finland, Norway) and Iceland (Fig. 1). Those sites had been sampled in different projects, mainly for routine environmental monitoring programs (sampling years 2006–2016). Water chemistry parameters were extracted from the Swedish National database (https://miljodata.slu.se/mvm/) for the Swedish sites, from the Hertta system version 5.7 of Finnish Environmental Administration (http://www.syke.fi/en-US/Open_information) for the Finnish sites, and from databases hosted at the Norwegian Institute for Water Research for the Norwegian sites. For the sites in Iceland, water chemistry parameters were derived from literature (Friberg et al., 2009; Ólafsson et al., 2010). We compiled data for nutrients: total phosphorus (TotP) and total nitrogen (TotN), conductivity, total organic carbon (TOC) as measure for humic substances, alkalinity and pH. Diatoms were sampled in autumn following the European standard (EN 13946:2014 CEN, 2014a). The samples were preserved with 97% ethanol (final concentration approximately 70%) and kept in dark at room temperature until dividing for morphological analysis and DNA extraction.

2.3. Diatom analysis & identification – Microscopy

Diatom morphological data were partly derived from national or regional databases (Sweden, Finland), and partly from analyses done for the study of Bailet et al. (2019) (Norway, Iceland). All analyses followed the European standard (EN 14407:2014 CEN, 2014b), and diatom identification was harmonized. In short, the samples were oxidised with hydrogen peroxide and the cleaned diatom valves then mounted with Naphrax (Brunel Microscope Ltd) on microscope slides. Identification to the lowest taxonomic level possible was done under a light microscope with interference contrast (1000 × magnification). At least 400 valves per sample were counted and identified using standard literature (Jarlman et al., 2016). Counts were then expressed as relative abundance. As diatom morphological identification by microscope is a large source of error, we made sure that taxa lists were comparable. All involved analysts had harmonized their way of diatom identification and counting following the recommendations of NorBAF (Kahlert et al., 2016), which ensured that the resulting taxa lists can be considered to contain no significant differences due to different laboratories (Kahlert et al., 2009). To simplify reading, we here use the term "morphotaxa" when referring to the diatom taxa identified by microscopy.
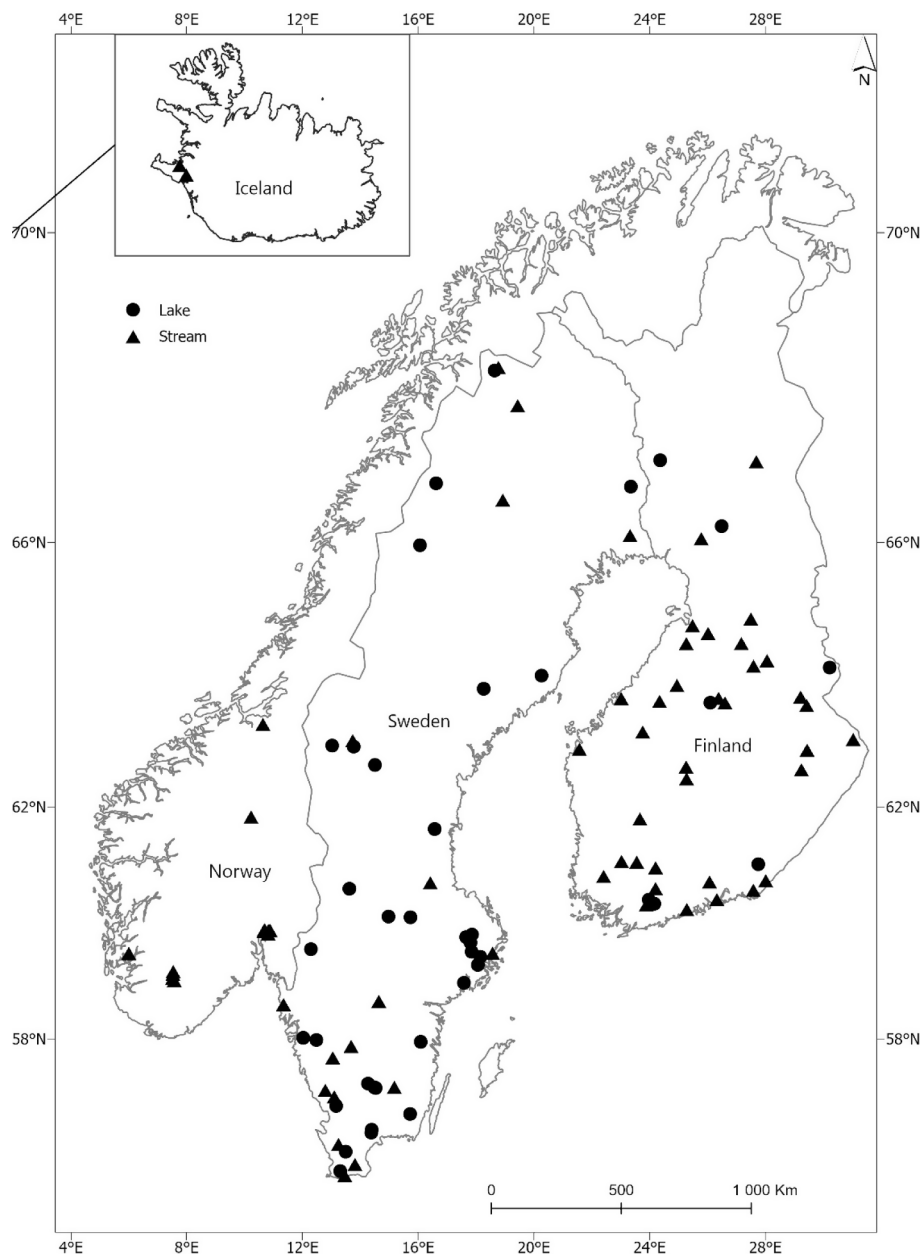
Fig. 1. Sampling sites (76 streams and 39 lakes) in Fennoscandia (Sweden, Finland, Norway) and Iceland.

2.4. Morphological trait assignment

Morphological traits were assigned to the morphotaxa using a taxon-specific value following the database of Rimet and Bouchez (2012). We used biovolume (five categories from very small to large; 1: <100 μm^3; 2: 100–300 μm^3; 3: 300–600 μm^3; 4: 600–1500 μm^3; 5: >1500 μm^3) and guilds (mainly based on growth-form: high-profile, low-profile, motile and planktonic). As not all Nordic taxa were found in the database, we added values for those taken from literature. Additionally, we calculated surface to volume ratios (S/V) for each taxon by using a simple equation assuming a box form of the cell for pennate diatoms, and a cylinder for central diatoms, taking size values from the database of Rimet and Bouchez (2012). S/V values were then also transformed to five categories (1: < 0.5, 2: 0.5–1; 3: 1–1.5; 4: 1.5–2; 5: > 2). Different terms are in use with respect to the definition of morphological characters, and the term guild comprises already a combination of several characters (Rimet and Bouchez, 2012). For the sake of simplicity, we refer in this article to all used characters as traits, i.e. a trait is here covering both guilds,

biovolume and S/V categories. Using the resulting 14 trait categories implied that a taxon could potentially be categorized into one of 100 unique combinations (5 biovolume categories * 5 S/V categories * 4 guilds), or into one of 20 combinations of a guild and a biovolume category, or guild and S/V category. The taxa found in our study comprised 66 of the 100 possible trait combinations. In the following text, we refer to the uncombined traits as "pure traits", to the 66 full combinations as "fully combined traits" and to the others as "partly combined traits". After testing all three variations, we mainly continued using the fully combined traits only (see further below).

2.5. Molecular analyses, amplicon sequence variants (ASVs)

2.5.1. DNA extraction and sequencing

The details for DNA extraction and sequencing can be found in Bailet et al. (2019). In short, DNA was extracted using the NucleoSpin Soil Kit (Macherey-Nagel), and PCR was performed on a 312 bp barcode on the *rbc*L plastid gene using the modified Diat_*rbc*L_708F and R3 primer pair

following (Vasselon et al., 2017a; Vasselon et al., 2017b). Sequencing was done by the Platform Genome Transcriptome (PGTB, Bordeaux, France) using Ion Torrent Personal Genome Machine (PGM).

2.5.2. Bioinformatics

Primers were removed from raw reads with cutadapt v2.3 (Martin, 2011). We retained only reads representing full amplicons with both forward and reverse primers. Then, ASVs were generated using the DADA2 pipeline (v1.14, Callahan et al., 2016), implemented in R, with default parameters. Singletons were removed during sample inference.

Taxonomy was assigned to ASVs with the *assignTaxonomy* function from the DADA2 package, which is an implementation of a naive Bayesian classifier method (Wang et al., 2007). Diat.barcode (version 7, Rimet et al., 2019) was used as a reference database for taxonomic assignment of ASVs.

2.6. Statistical analysis and data treatment

2.6.1. Environmental gradients

We first performed a Principal Components Analysis (PCA) of the sampling sites to show the environmental gradients of the water chemistry variables among the sites included in this study. Prior to analysis, the correlations (Pearson's r) between the environmental factors were checked (Table S2), and the alkalinity data were removed from the analyses due to a very high correlation with conductivity (r = 0.96). The remaining variables were then ln (TotP, conductivity) or square root (TotN, TOC) transformed to reach normal distribution, and were scaled to unit variance for the subsequent PCA analysis.

2.6.2. Diatom assemblage relationship to the environmental gradients

Then, we inspected the raw taxa data prior to the analyses of the assemblage structure. The molecular dataset contained on average 23,656 reads per sample, the maximum was 33,664 reads. ASVs with $\leq$ 10 reads in the total dataset were removed from further analyses to denoise the dataset. In the next step, we performed rarefaction in order to handle the bias potentially introduced by the different sequencing depth among samples. We used the *rrarefy* function of the vegan v2.5–7 package in R v4.0.4. Sampling size was set to the lowest read number of 8843 reads, and controlled by rarefaction curves (Figure S1). Then, read data were transformed to relative abundances (read numbers divided by the total number of 8843 reads). Microscope counts were not rarefied because the original total count number did not differ much between samples due to the standardized method. Rare ASVs, morphotaxa, traits and trait combinations which were present in one sample only were removed from further analyses to denoise the dataset and improve the detection of relationships between diatom assemblage composition and environmental factors.

Nonmetric Multidimensional Scaling (NMDS) analyses were performed followed by environmental variable fitting for each of the three diatom datasets (ASVs, morphotaxa, traits) to explore and display the diatom assemblage structure in relation to the environmental data. At this stage of the analyses, all three subdatasets for the traits (pure, partly and fully combined) were used to assess which of them would have the best relationship to the environmental variables, to decide for the further use in index development. For the NMDS, we used the *metaMDS* function of the vegan package in R 4.0.4. The distance measure was Bray-Curtis. We applied 20 random starts to reach a stable solution (default algorithm of *metaMDS*). The dimensionality of the data was assessed with a scree plot, using the lowest dimension where the stress was < 0.2, which was considered as good, resulting in three dimensions for each dataset. The stress value indicates the goodness of fit of the model to the data, with lower values implicating a better fit. The *metaMDS* function in R performs a PCA rotation on the final results so that axis 1 contains the greatest variance. Shepard stress plots were visualized in order to check how well the original dissimilarities were preserved in the reduced number of dimensions and we found satisfactory

R^2 values of the fit in each case ($R^2 = 0.97, 0.98, 0.99$ for ASVs, morphotaxa and fully combined traits, respectively). For the environmental variable fitting, we used the *envfit* function of the vegan package in R. The goodness of fit of the variables to the NMDS ordination was assessed using 999 permutations, and represented by the squared correlation coefficients. As the fully combined traits showed the best relationship to the environmental variables, only those were kept for further analyses. The *anosim* function of the vegan package in R was used to perform an analysis of similarities (ANOSIM) to test if the diatom assemblages differed significantly between streams and lakes, in each of the three datasets (ASVs, morphotaxa, fully combined traits), respectively.

To analyse if and how the assemblage structure was related to the measured environmental variables directly, we first used Detrended Correspondence Analysis (DCA) to analyse the type of response to be able to choose the suitable ordination method (unimodal or linear response). The DCA revealed quite long 1st axes for both ASVs and morphotaxa (Table S4), indicating an unimodal response of both (Šmilauer and Lepš, 2014). Based on the outcome of the DCA, we performed Canonical Correspondence Analysis (CCA) on all datasets (ASVs, morphotaxa, fully combined traits). Using forward selection, TotN was excluded from the analyses as it strongly correlated with TotP. For scaling options, we used centering with unit variance, and chose to optimize species. Monte Carlo tests were performed to test for a significant relationship between assemblage structure and environmental variables. DCAs and CCAs were performed using the functions *decorana* and *cca* of the vegan package in R. The biological datasets were Hellinger transformed for DCA and CCA analyses to tackle the problem of many variables having low counts and many zeros, giving them a lower weight. The outcomes of the NMDS plus *envfit*, and CCA analyses were used to compare the response of the three datasets to the measured environmental variables.

2.6.3. Development of nutrient indices

We then developed and compared diatom indices based on ASVs, morphotaxa, and fully combined traits, respectively, to test which of them was best related to nutrient enrichment, here measured as TotP concentrations. Indices were developed separately for streams and lakes as we expected the different water body types both to host different combinations of environmental factors, as well as somewhat differing diatom taxa and traits (Kahlert and Gottschalk, 2014). Even though it has been shown earlier that a diatom index can indicate TotP equally well in lakes and streams, there were subtle differences with respect to very low and very high TotP values (Kahlert and Gottschalk, 2014). We also tested index development based on two different ways of measuring diatom occurrence, i.e. we calculated ASV/morphotaxon/trait-specific optima on relative abundance in one trial, and on presence-absence data in another. In this way, we wanted to test which of the calculated ecological profiles would best reflect TotP concentrations.

The model building algorithm followed Tapolczai et al. (2019) and Tapolczai et al. (2021), a robust cross validation method during which a quality index is developed on a training dataset consisting of a randomly selected 75% of the samples. The index is then calculated on the remaining 25% test dataset. This cross validation procedure was performed 100x times for the three types of data, so each sample was sorted in the test dataset with a 25% chance at each iteration. The method resulted in multiple index values for each sample so a mean and standard deviation of the index scores for each sample could be obtained, and additionally, obtaining extreme values could be avoided.

During the index development on the training dataset, optimum and tolerance values along the TotP gradient were calculated for the biological units (ASVs, morphotaxa, fully combined traits) using the two approaches depending on the data type (abundance or presence-absence). Weighted means and weighted standard deviations were calculated, based on the relative abundances to infer optimum and tolerance values, respectively, for the abundance data. For presence-absence data, means and standard deviations were calculated using

the TotP values of samples where a given ASV/morphotaxon/trait was present. Only biological units that were present in at least three samples were kept in the training dataset, in order to ensure a relatively stable profile along the TotP gradient. The Zelinka-Marvan equation (Zelinka and Marvan, 1961) was adapted to calculate quality index scores on samples in the test dataset:

$$\text{Index}_{\text{ASV/morphotaxon/trait}} = \sum \left(a_i s_j v_j \right) \Big/ \sum \left(a_i v_j \right)$$

where *Index*_{ASV/morphotaxon/trait} is the index score of the test sample based on ASVs, morphotaxa and fully combined traits, respectively; a_j is the relative abundance of the jth ASV/morphotaxon/trait; s_j is the optimum value of the jth ASV/morphotaxon/trait; and v_j is the tolerance value of the jth ASV/morphotaxon/trait. Calculated index scores were then correlated with the TotP values to assess its efficiency. The index development resulted in four different indices for each biological unit: one index based on ASV/morphotaxon/trait-specific TotP optima and tolerance values modeled from presence-absence data, one index based on optima and tolerance values modeled from relative abundance data, both separately for streams and lakes. The index score for a site itself was in all cases calculated with the relative abundance data (a_j) of the ASVs/morphotaxa/traits present.

3. Results

3.1. Environmental gradient of the sampling sites

The sites in our study were well distributed along a nutrient gradient (Fig. 2), which was the main structuring environmental factor (correlations with the 1st PCA axis: TotP and TotN both r > 0.8, Table S3). Also conductivity was strongly correlated with the same axis (r = 0.75, Fig. 2, Table S3). The other strong gradient in our study was characterized by pH and TOC, both correlated to axis 2 but in opposing directions (Fig. 2, Table S3). The west-east gradient in the concentration of organic matter (from clear waters in Norway to brown waters in Finland) is clearly visible. Whereas the streams were distributed along the entire environmental gradient on both axes, lakes showed a more restricted distribution along the second PCA axis.
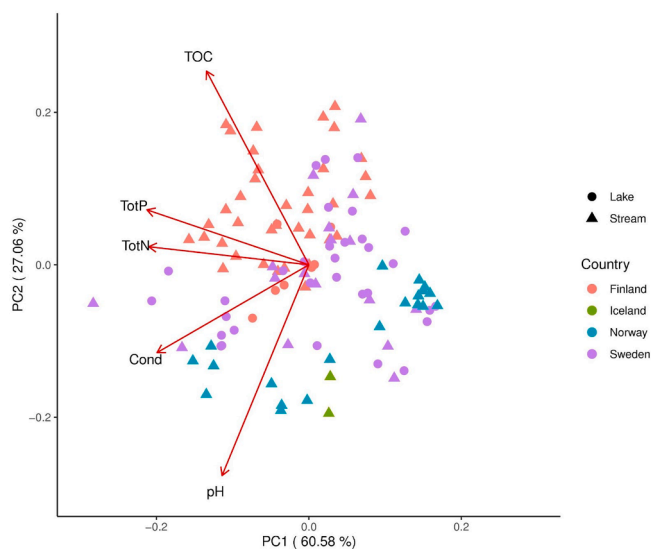

Fig. 2. PCA analysis of the 115 sampling sites clustered by their environmental variables: total organic carbon (TOC), total phosphorus (TotP), total nitrogen (TotN), conductivity (Cond) and pH.

3.2. Diatom assemblage structure and correlations to the environmental variables

We found 5467 ASVs, and 530 morphotaxa. The latter were classified into 66 full trait combinations. After the removal of the rare units to denoise the dataset and improve the modelling, 1941 ASVs, 368 morphotaxa and 64 full trait combinations remained, on which the analyses were performed. The seemingly high loss of units corresponded however to only 8.2, 1.1 and 0.2 % abundance of ASVs, morphotaxa and trait combinations, respectively. To ensure that no important information got lost due to the removal of the rare taxa, we performed all analyses also on the full dataset (data not shown). The results were basically the same, however with a higher uncertainty of the models.

In general, the comparison of the ten most abundant diatom taxa found in this study differed between the morphological and the molecular analysis, but also showed some similarities (Table S1). Looking down a microscope, the *Achnanthidium minutissimum* complex was most abundant in our study, followed by *Tabellaria flocculosa* (Roth) Kützing. With metabarcoding of the *rbc*L marker, *Melosira varians* Agardh was instead the most abundant taxon, also here followed by *T. flocculosa*. Some taxa were found in both top ten lists, such as *Fragilaria gracilis* Østrup, *Amphora pediculus* (Kützing) Grunow and *Aulacoseira ambigua* (Grunow) Simonsen, whereas others were not, such as *Eunotia incisa* Gregory (morphological list) or *Cymbella cymbiformis* Agardh (molecular list). The three most abundant morphological taxa represented>25% of all diatom counts, whereas the three most abundant ASVs only added up to 11%. On the other hand, several ASVs matched to the same taxon name, for example three ASVs represented *T. flocculosa* in the top ten list, adding up to about the same relative abundance as in the morphological taxa list. However, adding up all 67 ASVs of *A. minutissimum* resulted in a much lower total relative abundance for this taxon (2.1% vs. 18.6%). The relative abundance of the acidophilic genus *Eunotia* added up to about 10% relative abundance in both datasets. Overall, of the 1941 ASVs, 1069 ASVs could be assigned to 64 genera and to 190 species. 51 of the 64 genera, and 89 of the 190 species were found in the morphotaxa dataset as well.

When categorizing the taxa counted by microscope into the guild system following Rimet and Bouchez (2012), about one-third represented the low-profile guild and another one-third the high-profile guild (Table S6). The relative abundance of the motile guild was 15 %, while the planktonic guild was least abundant. Regarding biovolume, the smallest-sized diatoms (category 1) were most common (35%), followed by categories 2 (21%) and 4 (21%). Regarding surface to volume ratio, the category with the highest ratio (category 5) was most common (38%), followed by category 3 (32%). The most abundant morphotaxon *A. minutissimum* falls into the low-profile guild, with smallest cell size and largest S/V. The second most abundant taxon *T. flocculosa* falls into the high-profile guild, cell-size category 4 and S/V category 3.

The graphical representation of the assemblage relationships (NMDS) in combination with the overlay of the environmental variables showed that the main differences between assemblages reflected relatively well the environmental differences between sites for both ASVs and morphotaxa (Fig. 3). However, the graphs also show that some of the assemblage relationships especially correlated to axis 2 had little representation in environmental variables, especially so for the ASV data. Here, assemblages were significantly separated by habitat type, i.e. if they were sampled in streams or lakes (ANOSIM, Table S7). Stream and lake assemblages identified as morphotaxa were barely (but still significantly) separated (ANOSIM, Table S7). The stable solution of the NMDS was reached with a somewhat higher stress value for the ASV dataset than for the morphotaxa dataset, indicating more variation in the ASV assemblage structure (Fig. 3). The environmental variable fitting analysis indicated TotP and conductivity as potential strong predictors for the diatom assemblages both for ASVs and morphotaxa (Fig. 3, Table S5). However, pH was identified as the strongest predictor for the morphotaxa assemblage, whereas it had a rather weak
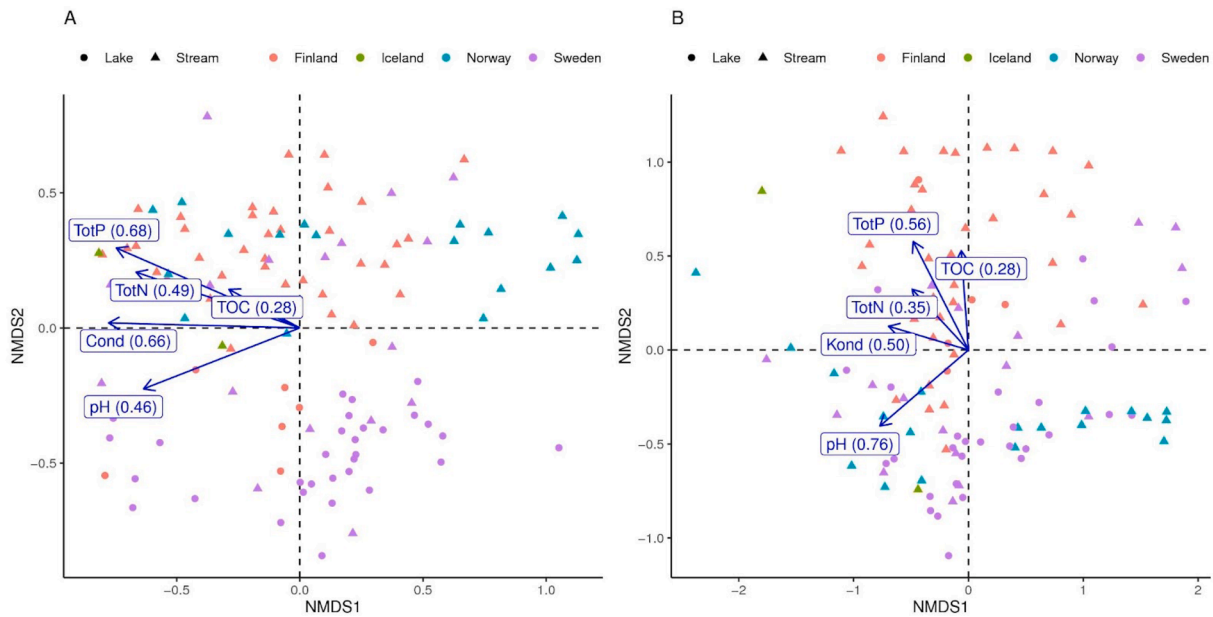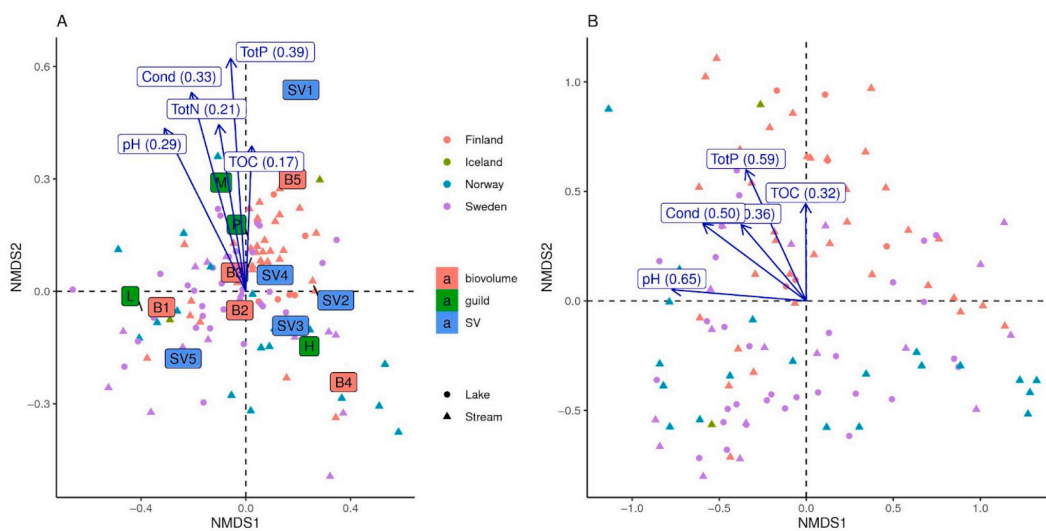
Fig. 3. Non-metric multidimensional scaling (NMDS) plot on diatom ASVs (A) and morphotaxa (B) assemblages data with fitted environmental variables. Stress values are 0.17 and 0.13 for the two NMDSs, respectively. Squared correlation coefficients are given for each environmental factor reflecting the strength of this factor as a predictor of the assemblage structure.

relationship to the structure of the ASVs (Fig. 3, Table S5). TOC showed the weakest relationship to the assemblage structure for both datasets. All correlations in both datasets obtained with environmental fitting were strongly significant (p < 0.001).

Analyzing the direct effect of the environmental variables, assemblage structure was preserved even in this constrained form, confirming that the measured variables captured main factors steering the taxon distribution (CCA, Monte Carlo permutation tests for axes 1, 999 randomizations: p < 0.01). For both ASVs and morphotaxa, pH, TotP and conductivity were strongly positively related to the 1st CCA axis (Table S8), whereas TOC was only weakly correlated to this axis, and instead correlated strongly to the 2nd axis. The explained variation was higher for the morphotaxa structure (13.45 %) than for the ASVs (8.41 %), and relatively more variation was explained by the first axis for the morphotaxa than for the ASVs (Table S8).

In contrast to ASVs and morphotaxa where the highest variation in structure correlated relatively well with the highest variation in water chemistry (Fig. 3), the highest variation in trait structure was basically uncoupled from the measured environmental variables when traits were analysed separately (pure trait dataset, Fig. 4A). Compared to ASVs and morphotaxa (Fig. 3), the environmental variables correlated clearly more with the second axis than with the first, and showed a weaker relationship with the pure trait composition than for ASVs or morphotaxa (Fig. 4A, Table S5). The most abundant guilds, low-profile and high-profile, were mainly correlated in opposing directions to the first axis which had little representation in environmental variables, whereas the motile guild was correlated to the second axis having the strongest relationship to the environmental variables. Also the smallest and the largest S/V categories were correlated to axis 2 in opposing directions. The smallest size category peaked in the same direction as the low-

Fig. 4. Non-metric multidimensional scaling (NMDS) plot on diatom morphological trait data with environmental variables included. A: Traits analysed separately (pure trait dataset). B: Traits analysed as the 64 full combinations of biovolume (B, 5 categories), surface to volume ratio (SV, 5 categories) and guilds (4 categories, H high-profile, L low-profile, M motile, P planktonic) present in the dataset (full combined trait dataset). Stress values of the NMDSs are 0.11 and 0.14, respectively.

profile guild on axis 1, whereas the second-largest size category peaked with the high-profile guild. No obvious correlation was found for the other trait categories.

Combining traits increased the correlation to the local water chemistry variables to the point that ASVs, morphotaxa and the fully combined traits, correlated similarly well in general, with some differences in which variable gave a better correlation (Fig. 4B, Table S5). Stream and lake assemblages were not separated in the trait ordinations, in contrast to the ASVs. Morphotaxa assemblages were only slightly, however significantly different between streams and lakes (Table S7).

3.3. Development of diatom nutrient (TotP) indices

Some morphotaxa consisted of several ASVs, and TotP optima differed among these ASVs. For example, there were three dominant ASVs within *T. flocculosa*, and the most abundant ASV2 had a much lower TotP optimum than the ASVs 3 and 10 (Fig. 5). Consequently, the TotP optimum value of the morphotaxon *T. flocculosa* was approximately the average of the three ASVs, and its tolerance range was wider than of any of the ASVs.

In a next step, we analysed if the ASV or morphotaxon specific optima differed between streams and lakes. We found that optima were correlated, but that variability was large. There was, however, not a 1:1 relationship. Instead, optima were higher in streams than in lakes in nutrient poor habitats, while the opposite was true in nutrient rich habitats (Fig. 6). Those results were similar for both the relative abundance, and the presence-absence based modeled optima.

All the calculated indices were related to TotP concentrations, but some were closer related than others (Fig. 7). Overall, ASV indices performed better than the respective indices for morphotaxa , which in turn performed better than the indices based on traits (full combined dataset was used only). All indices performed better when based on optima and tolerance values derived from relative abundance data than from presence-absence data. Finally, stream and lake indices developed from relative abundance data performed approximately equally well. In contrast, indices developed from presence-absence data performed better in streams than in lakes for morphotaxa and ASVs, and equally well for traits. The indices based on ASVs had not only the best correlations to TotP, they also overall had the steepest response, i.e. the index differences from low TotP indication to high TotP were largest when based on ASVs, which would separate sites with different TotP concentration best.

4. Discussion

Our results show that major changes in diatom composition in Fennoscandian streams and lakes are strongly related to changes in nutrient concentrations, especially TotP. This was true for all of the three studied biological units, namely molecular taxa identified as ASVs, morphotaxa, and morphological traits. This result justified the development of nutrient indices to assess eutrophication, the main pressure in European freshwaters. Overall, the relationships between the environmental variables and the diatom assemblage composition were quite strong compared to the global study of (Soininen et al., 2016), probably because the geographically smaller scale of our study implied less additional factors affecting diatom assemblages on large scales. This encourages the development of diatom indices specific for the Nordic region. Our results show furthermore that it is fully possible to develop diatom TotP indices for both ASVs and traits. Especially ASV indices are promising as they performed as good as or better than indices based on traditional morphotaxa in both streams and lakes. Indices based on traits had the weakest relationship with TotP.

Our results suggest that molecular analysis using ASVs is a promising tool for assessing eutrophication in Nordic streams and lakes. Especially the pressure-response relationship of the ASV TotP indices based on abundance derived ASV-specific optima developed in this study were strong compared to existing indices based on morphotaxa (Kelly et al., 2007; Kelly et al., 2014). Regarding the most commonly applied traditional diatom index in Fennoscandia, the IPS (Indice de Polluo-sensibilité Spécifique Cemagref, 1982), a previous study covering a broad ecological gradient of streams and lakes in Sweden found correlation coefficients of the IPS with TotP of $R^2 = 0.55$ for lakes and 0.65 for streams (Kahlert and Gottschalk, 2014), while our results for ASV were 0.83 and 0.87. While comparing correlation coefficients should be done with caution, our results nevertheless indicate that a method based on ASVs outperforms the currently applied IPS to indicate nutrient concentrations. We suggest that the different structure of the ASV dataset, i. e. the differences in the abundance of the main diatom taxa (Kelly et al., 2020), in combination with the higher number of ASVs representing a more fine-tuned diversity (Tapolczai et al., 2019), are the reason for the better performance of the ASV indices. Even if we did no in-depth study of the detailed ASV structure, we could see that an important cause for the high ASV number was the fact that many morphotaxa were separated into several ASVs. We propose that different ASVs may reflect different adaptations to environmental conditions, as reflected in the
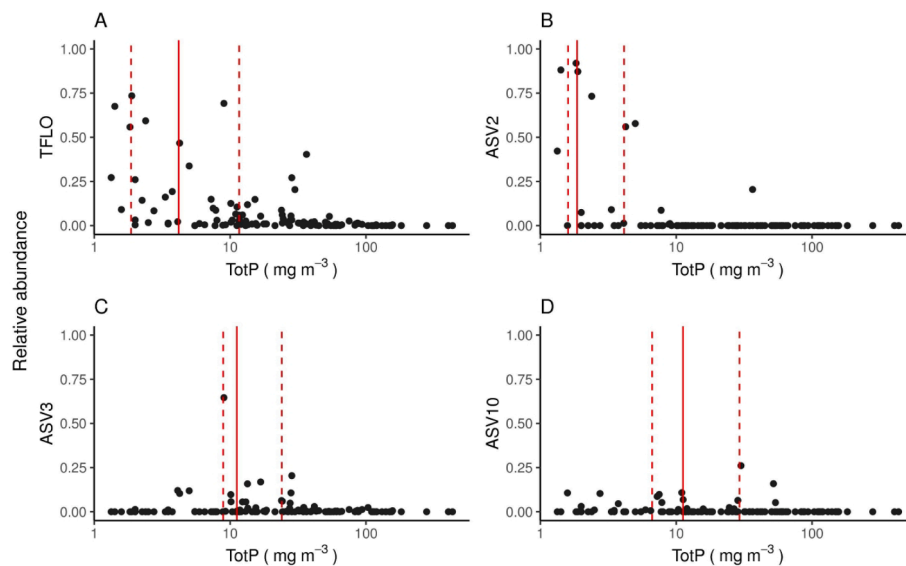

Fig. 5. Optimum and tolerance values for TotP of the microscopically observed taxon *Tabellaria flocculosa* (A) compared to the three most abundant ASVs within this taxon (B- D).
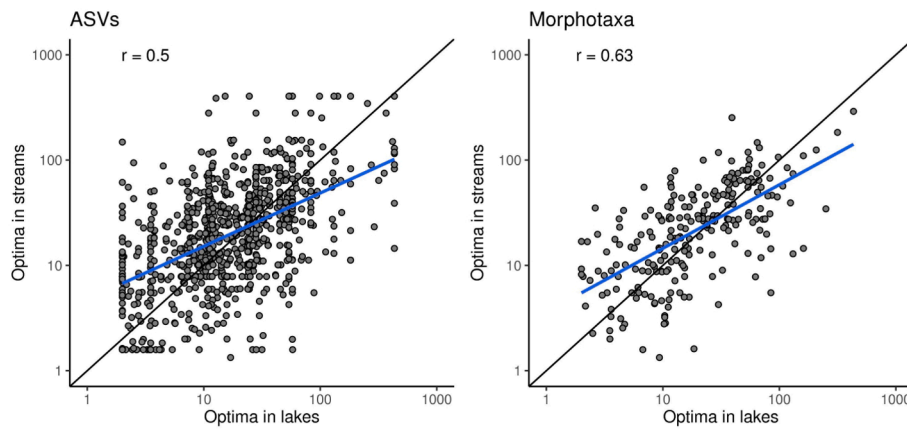
Fig. 6. Modeled taxon specific realized optima for TotP in streams versus modeled optima in lakes for the same taxon. Left: ASVs, right: Morphotaxa. (Optima were calculated based on relative abundance).
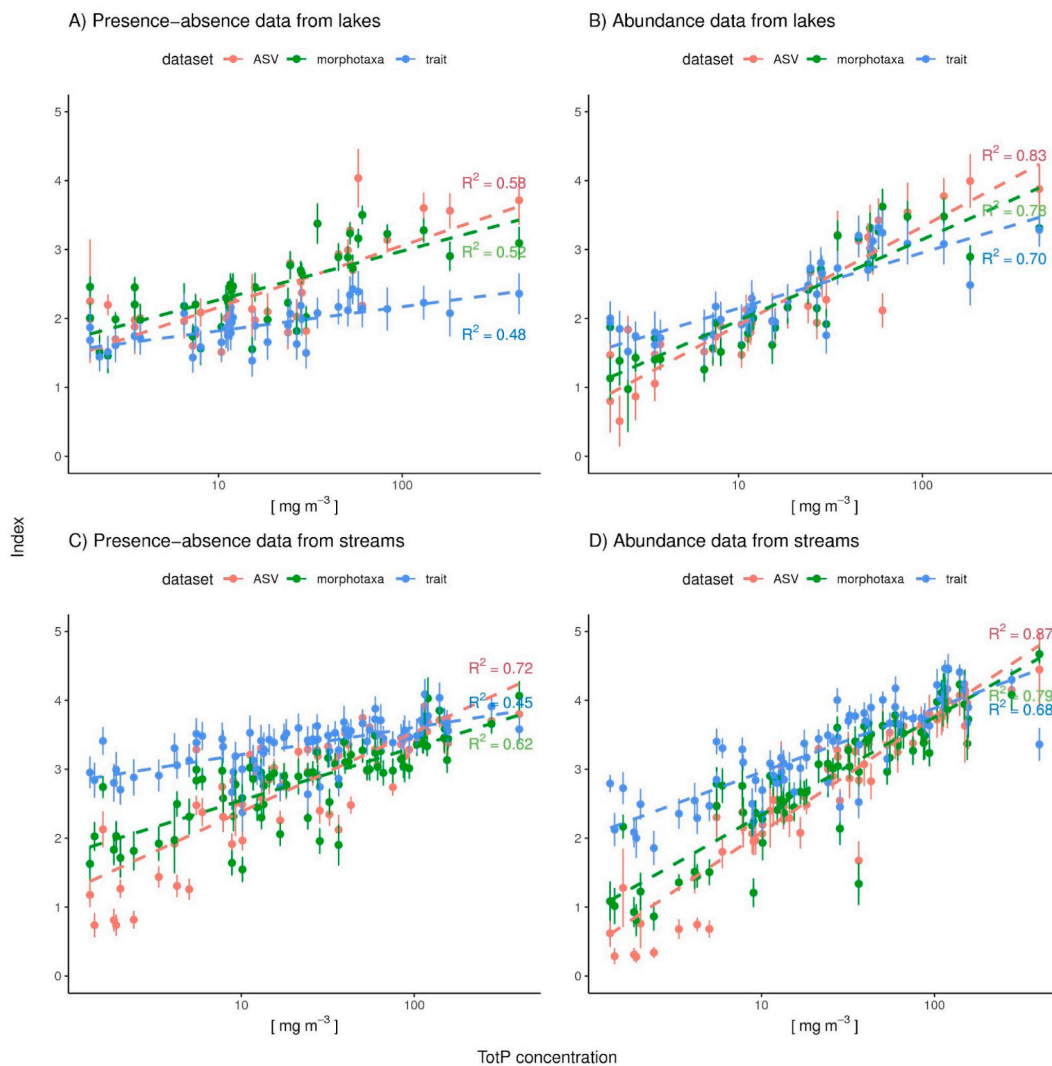

Fig. 7. Diatom nutrient indices versus TotP concentration [μgL^{-1}] based on presence-absence data from lakes (A), abundance data from lakes (B), presence-absence data from streams (C), and abundance data from streams (D).

optima presented for *T. flocculosa*. Consequently, the index based on ASVs was more closely related to TotP than the index based on morphotaxa. Our results fit well to recent studies which found that traditionally described diatom species based on morphological characters are

not necessarily represented by a given setup of barcodes with the same sequencing depth (Abarca et al., 2020; Kahlert et al., 2019; Kermarrec et al., 2013; Li et al., 2018; Vanormelingen et al., 2008). These discrepancies emerge from the different species concepts (Mann, 1999)

used by the conventional method based on the morphological characteristics of the organisms (morphospecies) and metabarcoding that delimitates taxa based on the genetic similarity between barcodes (genetic species). Thus, an in-depth comparison of the ASV and the morphotaxa structure was not possible in this study, since (a) the taxonomic assignment is still uncertain with our restricted knowledge on diatom taxonomy, and (b) the unidentified ASVs may either be new taxa, natural variability within a known taxon, or simply be due to an incomplete library. As a consequence, it currently is not possible to give a specific reason for the improved performance of the ASV dataset.

We suggest that the use of ASVs has many advantages. Apart from the higher number of ASVs, which may result in better index performance, it also automatically generates diatom units with reproducible identities. Our results indicate that it is possible to develop a useful index based on ASVs tailored to indicate eutrophication (measured as TotP concentrations). Such an index is important to assess if nutrients are the problem when a general degrading index is indicating worse than good ecological status. Furthermore, a TotP index is requested by water managers who wish to monitor the success of measures to reduce phosphorus.

One important result of our study is that the indices based on ASV/ morphotaxon/trait-specific optima and tolerance values derived from abundance data performed better than those where presence-absence data were used. There is a tradeoff between giving too much weight on random high occurrences and giving too much weight on random single occurrences. It clearly seems that the latter issue is most crucial for diatoms, because all our indices based on abundance optima performed better, irrespective of the diatom biological unit. Using presence-absence data gives more weight to the presence of a few cells, even if they never get abundant. Those taxa might indeed be a genuine part of the assemblage, but they also might just be washed into a site, alive or dead, and not really thrive there. Obviously, for the calculation of a TotP index at least, random high abundances were less of a problem.

Another insight from our study was that stream indices performed somewhat better than lake indices for the ASV and morphotaxa-data. Similar results have been found in a previous study in Sweden where the national diatom metric IPS was well correlated to TotP with strong relationships for both streams and lakes, however a small, but significant impact of the habitat was found (Kahlert and Gottschalk, 2014). More work is needed to understand if the differences between streams and lakes are of such importance that separate TotP indices should be developed. We currently cannot exclude that the observed differences just are an artifact of our dataset, as we had a lower number of sampled lakes than streams which probably generated less stable ecological profiles. Alternatively, the differences in index performance could also reflect real ecological differences, for example the different availability of phosphorus (P) in the TotP (Kahlert and Gottschalk, 2014). We observed that the realized diatom TotP optima were on average somewhat higher in streams than in lakes under oligotrophic conditions, and the opposite in nutrient rich waters. We assume that the available P actually was similar in both cases, but more P was bound to humic substances in streams, and to phytoplankton in lakes, increasing the realized diatom-specific TotP optima. The underlying causes are probably that streams in general tend to have a higher content of humic substances than lakes (Löfgren et al., 2003), and second that phytoplankton can reach high amounts of biomass especially in eutrophic lakes. It has earlier been shown that traditional diatom indices had better relationships to soluble phosphorus than to TotP (Poikane et al., 2021). However, as this form of phosphorus is not easily measurable, indices do not necessarily show this better relationship (Kelly et al., 2007). Algal P uptake can also be impacted by the prevailing N:P ratio, however pilot analyses had confirmed that the N:P ratio was not different between streams and lakes in our study. Last, but not least, it is possible that the differences in the diatom assemblages found in our study can partly explain the different index performance for streams and lakes. Diatom assemblages differed significantly between streams and lakes, especially in the ASV, but also in the morphotaxa dataset, which is

in agreement with earlier findings (Kahlert and Gottschalk, 2014). Probably the different environmental factors, for example the unidirectional flow in streams and the wave action in lakes, are selecting for different diatom taxa, which in turn might have a different response to TotP, but also might interact differently in the diatom assemblage impacting other taxa.

Indices based on traits currently performed slightly less than indices based on ASVs and morphotaxa. Nevertheless, our results show that traits generally are a promising tool for environmental assessment. However, even if we found that the full combination of all traits had the best relationship to the environmental variables and was thus used for index building, a subsequent study should develop an improved trait-based index by carefully selecting adequate traits with a good correlation to TotP, as also suggested by Tapolczai et al. (2017). Based on our findings, certain morphological traits are better candidates than others for the development of a nutrient index. The motile taxa were positively correlated with nutrients as expected, because they should be able to move towards limiting resources, and can live in sediments which are rich in resources (Passy, 2007; Rimet and Bouchez, 2012). Kelly et al. (2009) formulated a conceptual basis stating that motile diatoms are typical for eutrophic sites because they are better adapted to live in a three-dimensional biofilm built by filamentous green algae thriving under nutrient rich conditions. We also found that the taxa with the highest surface to volume ratio were negatively correlated with nutrients whereas the largest taxa and those with the smallest surface to volume ratio were positively correlated. Small-cell taxa have been hypothesized to be more abundant under nutrient-poor conditions due to a proposed higher nutrient acquisition, so the opposite should be true for large cells (Lange et al., 2016). More general, the surface to volume ratio (S/V) is a cell's relative area exposed to the environment (Snoeijs et al., 2002), and a high ratio has thus been related to a rapid resource acquisition (Lange et al., 2016; Potapova and Snoeijs, 1997; Snoeijs et al., 2002).

However, we found no clear order in the response to nutrients in the S/V categories, and even less so in the biovolume categories. Furthermore, the abundant low-profile and high-profile guilds were not related well to the nutrient gradient. We suggest that other factors than those we have measured in our study were steering the occurrence of those traits in the study sites. Similar results have been found in other studies trying to establish a diatom trait-based concept for environmental assessment, and especially disturbance has been mentioned as an important factor affecting diatom trait compositions (B-Béres et al., 2016; Cardinale et al., 2006; Lange et al., 2016; Passy, 2007; Rimet and Bouchez, 2012). In our study, the trait response to the main ordination axis, which had no good relationship to the measured environmental factors, was dominated by *A. minutissimum* with negative correlation to this axis, and *T. flocculosa* with positive correlation. *A. minutissimum*, falling in the category low-profile with smallest cell size and largest S/V, was dominating the response of these three trait groups, and was obviously not mainly responding to nutrients. Indeed, *A. minutissimum* is known as a pioneer taxon (Rimet and Bouchez, 2012), and is therefore abundant in frequently or recently disturbed environments, irrespective of nutrient concentrations (Kelly et al., 2009). *A. minutissimum* is also categorized as withstanding shorter periods of drought (Denys, 1991). There certainly are factors disturbing particular sites in our study, such as flow abrasion or grazing, which we do not have any data for. The second most abundant taxon was *T. flocculosa*, falling in the high-profile guild with a relative large biovolume (category 4) and an average S/V (category 3), responding negatively to nutrients. In contrast to *A. minutissimum*, *T. flocculosa* is no pioneer taxon, is more common in matured biofilms (Kelly et al., 2009), and is sensitive to drought (Denys, 1991). However, *T. flocculosa* also is reported to withstand disturbance by water movement, potentially even more so than *A. minutissimum* (Hasselquist et al., 2018). Apart from disturbance by water movement, light or grazing might have affected diatom trait composition, because the low-profile guild potentially should be better adapted to low light conditions and

grazing than the high-profile guild (Passy, 2007; Rimet and Bouchez, 2012). Unfortunately, such data are usually not part of routine monitoring programs from which our data originate. Therefore, more research is needed to complete our understanding of diatom traits, and especially more knowledge is needed to understand pressure-relationships, and the importance of other factors than nutrients (Kelly et al., 2009).

Still, the idea of using a morphological trait-based approach is appealing, because traits should be the overall manifestation of the adaption of diatoms to their environment. Consequently, traits should be similar in similar habitats of different geographical regions, and a trait-based index could be used also in regions where the diatom species diversity is not fully known (Soininen et al., 2016; Tapolczai et al., 2017). Indeed, we could show that habitat type (streams versus lakes) was no important factor steering the trait composition in our dataset, whereas it was important and significant for the ASV dataset, and also significant for the morphotaxa dataset even if less prominent. These results are in agreement with an earlier study (Kahlert and Gottschalk, 2014). On the other hand, an argument against the use of traits in our study is our result that the trait index represented a shorter gradient than the ASV and morphotaxa indices, which is in contrast to the study of (Tapolczai et al., 2017) where the trait index covered a wider distribution of the environmental variables compared to the morphotaxa-based one. One could also argue that a trait-based approach still does not overcome the large efforts of analysis by microscope, however, today there are alternatives using automated imaging methods which can be trained to identify and quantify traits (Burfeid-Castellanos et al., 2020).

In summary, the good response to TotP of the molecular taxa assessed as ASVs with the *rbc*L barcode opens up for the development of a molecular nutrient indicator. The ASVs showed in general a similar response pattern to nutrients as the well-known morphotaxa, even if the assemblage structure was different. The developed nutrient index on ASV basis correlated even better to TotP concentrations than the morphotaxa index, and therefore should be tested for its use in environmental assessment. The impact of TOC on the studied molecular data seemed to be minor compared to the impact of nutrients, which would be an advantage if nutrients are the focus of the analysis. Further studies are needed, however, to understand if TOC could have an impact on TotP availability, especially in oligotrophic systems. In addition, more work is needed to further develop the indices up to application maturity, which was not possible in the present study. An index should be "fine-tuned" to test if it would improve when only including ASVs/morphotaxa/traits with a narrow ecological tolerance range for the target nutrient TotP, and the index must be used in new sites to tests its performance under new combinations of environmental variables and diatom assemblages. If the index will be based on certain indicator ASVs/morphotaxa/traits instead of the whole assemblage, criteria must be established on a minimum number of units for a stable and reliable performance. Rules need to be established on how to deal with new units with unknown indicator value. The indicator values should also be published with open access, and the database Diat.barcode would be an optimal platform for ASVs. Last but not least, for the use within the Water Framework Directive to classify an ecological status, reference values for different water types must be developed.

CRediT authorship contribution statement

Maria Kahlert: Conceptualization, Methodology, Project administration, Writing - original draft. **Bonnie Bailet:** Data curation, Investigation, Writing - review & editing. **Teofana Chonova:** Investigation, Writing - review & editing. **Satu Maaria Karjalainen:** Investigation, Writing - review & editing. **Susanne C. Schneider:** Investigation, Methodology, Writing - review & editing. **Kálmán Tapolczai:** Formal analysis, Methodology, Writing - review & editing.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.108088.

References

Abarca, N., Zimmermann, J., Kusber, W.-H., Mora, D., Van, A.T., Skibbe, O., Jahn, R., 2020. Defining the core group of the genus Gomphonema Ehrenberg with molecular and morphological methods. Bot. Lett. 167 (1), 114–159.

United States Clean Water Act (PL 92-500), 1972. Public Law, p. 92e500.

Alers-García, J., Lee, S.S., Spaulding, S.A., 2021. Resources and Practices to Improve Diatom Data Quality. Limnol. Oceanogr. Bull. 30, 48–53.

B-Béres, V., Lukács, Á., Török, P., Kókai, Z., Novák, Z., T-Krasznai, E., Tóthmérész, B., Bácsi, I., 2016. Combined eco-morphological functional groups are reliable indicators of colonisation processes of benthic diatom assemblages in a lowland stream. Ecol. Ind. 64, 31–38.

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.

Berthon, V., Bouchez, A., Rimet, F., 2011. Using diatom life-forms and ecological guilds to assess organic pollution and trophic level in rivers: a case study of rivers in south-eastern France. Hydrobiologia 673 (1), 259–271.

Burfeid-Castellanos, A.M., Kloster, M., Cambra, J., Beszteri, B., 2020. Both hydrology and physicochemistry influence diatom morphometry. Diatom Res. 35 (4), 315–326.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11 (12), 2639–2643.

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 (7), 581–583.

Cardinale, B.J., Hillebrand, H., Charles, D.F., 2006. Geographic patterns of diversity in streams are predicted by a multivariate model of disturbance and productivity. J. Ecol. 94, 609–618.

Carvalho, L., Mackay, E.B., Cardoso, A.C., Baattrup-Pedersen, A., Birk, S., Blackstock, K. L., Borics, G., Borja, A., Feld, C.K., Ferreira, M.T., Globevnik, L., Grizzetti, B., Hendry, S., Hering, D., Kelly, M., Langaas, S., Meissner, K., Panagopoulos, Y., Penning, E., Rouillard, J., Sabater, S., Schmedtje, U., Spears, B.M., Venohr, M., van de Bund, W., Solheim, A.L., 2019. Protecting and restoring Europe's waters: an analysis of the future development needs of the Water Framework Directive. Sci. Total Environ. 658, 1228–1238.

Cattaneo, A., Kerimian, T., Roberge, M., Marty, J., 1997. Periphyton distribution and abundance on substrata of different size along a gradient of stream trophy de Montréal. Hydrobiologia 354, 101–110.

Cemagref, 1982. Etude des méthodes biologiques d'appréciation quantitative de la qualité des eaux. Q.E. Lyon-A.F.Bassion Rhône-Méditeranée-Corse 218.

CEN, 2014a. EN 13946:2014 Water quality – Guidance for the routine sampling and preparation of benthic diatoms from rivers and lakes CEN.

CEN, 2014b. EN 14407:2014 Water quality – Guidance for the identification and enumeration of benthic diatom samples from rivers and lakes. CEN.

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. Ind. 121, 107082.

Denys, L., 1991. A check-list of the diatoms in the Holocene deposits of the western Belgian coastal plain with a survey of their apparent ecological requirements. I. Introduction, ecological code and complete list. Geological Survey of Belgium. 1991/ 02 - 246.

European Environment Agency, 2018. European waters. Assessment of status and pressures 2018, EEA Report No 7/2018. Publications Office of the European Union, Luxembourg p. 85.

European Parliament, Council, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Union L327, pp. 1-73.

Friberg, N., Dybkjær, J.B., Olafsson, J.S., Gislason, G.M., Larsen, S.E., Lauridsen, T.l., 2009. Relationships between structure and function in streams contrasting in temperature. Freshwater Biol. 54, 2051–2068.

Hasselquist, E.M., Polvi, L.E., Kahlert, M., Nilsson, C., Sandberg, L., McKie, B.G., 2018. Contrasting responses among aquatic organism groups to changes in geomorphic complexity along a gradient of stream habitat restoration: implications for restoration planning and assessment. Water 10 (10), 1465.

Hicks, Y.A., Marshall, D., Rosin, P.L., Martin, R.R., Mann, D.G., Droop, S.J.M., 2006. A model of diatom shape and texture for analysis, synthesis and identification. Mach. Vis. Appl. 17 (5), 297–307.

Jarlman, A., Kahlert, M., Sundberg, I., Herlitz, E., 2016. Påväxt i sjöar och vattendrag – kiselalgsanalys. Version 3:2: 2016-01-20, Handledning för miljöövervakning. Undersökningstyp. Havs- och Vattenmyndigheten, Göteborg, p. 24.

Kahlert, M., Ács, E., Almeida, S.F.P., Blanco, S., Dreßler, M., Ector, L., Karjalainen, S.M., Liess, A., Mertens, A., van der Wal, J., Vilbaste, S., Werner, P., 2016. Quality assurance of diatom counts in Europe: towards harmonized datasets. Hydrobiologia 772 (1), 1–14.

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 (4), 471–482.

Kahlert, M., Gottschalk, S., 2014. Differences in benthic diatom assemblages between streams and lakes in Sweden and implications for ecological assessment. Freshwater Sci. 33 (2), 655–669.

Kahlert, M., Kelly, M., Albert, R.-L., Almeida, S.F.P., Bešta, T., Blanco, S., Coste, M., Denys, L., Ector, L., Fránková, M., Hlúbiková, D., Ivanov, P., Kennedy, B., Marvan, P., Mertens, A., Miettinen, J., Picinska-Fałtynowicz, J., Rosebery, J., Tornés, E., Vilbaste, S., Vogel, A., 2012. Identification versus counting protocols as sources of uncertainty in diatom-based ecological status assessments. Hydrobiologia 695 (1), 109–124.

Kahlert, M., Kelly, M.G., Mann, D.G., Rimet, F., Sato, S., Bouchez, A., Keck, F., Kroth, P., 2019. Connecting the morphological and molecular species concepts to facilitate species identification within the genus Fragilaria (Bacillariophyta). J. Phycol. 55 (4), 948–970.

Kelly, M., Bennett, C., Coste, M., Delmas, F., Denys, L., Ector, L., Fauville, C., Ferreol, M., Golub, M., Jarlman, A., Kahlert, M., Lucey, J., Ni Chathain, B., Pardo, I., Pfister, P., Picinska-Faltynowicz, J., Schranz, C., Schaumburg, J., Tison, J., van Dam, H., Vilbaste, S., 2007. WFD intercalibration technical report. Part 1 – Rivers. Section 3 – Phytobenthos, p. 76.

Kelly, M., King, L., xed, Chath, xe, in, B., 2009. The conceptual basis of ecological status assessment using diatoms. Biology and Environment: Proceedings of the Royal Irish Academy 109B, 175-189.

Kelly, M., Urbanic, G., Acs, E., Bennion, H., Bertrin, V., Burgess, A., Denys, L., Gottschalk, S., Kahlert, M., Karjalainen, S.M., Kennedy, B., Kosi, G., Marchetto, A., Morin, S., Picinska-Fałtynowicz, J., Poikane, S., Rosebery, J., Schoenfelder, I., Schoenfelder, J., Varbiro, G., 2014. Comparing aspirations: intercalibration of ecological status concepts across European lakes for littoral diatoms. Hydrobiologia 734 (1), 125–141.

Kelly, M.G., 1998. Use of the trophic diatom index to monitor eutrophication in rivers. Water Res. 32 (1), 236–242.

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. Ind. 118, 106725.

Kelly, M.G., Whitton, B.A., 1995. Trophic diatom index - a new index for monitoring eutrophication in rivers. J. Appl. Phycol. 7, 433–444.

Kermarrec, L., Bouchez, A., Rimet, F., Humbert, J.-F., 2013. First evidence of the existence of semi-cryptic species and of a phylogeographic structure in the Gomphonema parvulum (Kutzing) Kutzing Complex (Bacillariophyta). Protist 164, 686–705.

Kolkwitz, R., Marsson, M., 1908. Ökologie der pflanzlichen Saprobien. Ber. Dt. Botan. Ges. 26a, 505–519.

Kruk, C., Huszar, V.L.M., Peeters, E.T.H.M., Bonilla, S., Costa, L., Lurling, M., Reynolds, C.S., Scheffer, M., 2010. A morphological classification capturing functional variation in phytoplankton. Freshwater Biol. 55, 614-627.

Lange, K., Townsend, C.R., Matthaei, C.D., 2016. A trait-based framework for stream algal communities. Ecol. Evol. 6, 23–36.

Larras, F., Boyé, A., Faure, E., Aberle, N., Archambault, P., Bacouillard, L., Beisner, B.E., Bittner, L., Castella, E., Danger, M., Gauthier, O., Karp-Boss, L., Lombard, F., Maps, F., Stemmann, L., Thiébaut, E., Usseglio-Polatera, P., Vogt, M., Laviale, M., Ayata, S.-D., 2021. Functional trait-based approaches as a common framework for aquatic ecologists. Limnol Oceanogr 66, 965–994.

Li, C., Andrzej, W., Ashworth, M., Dąbek, P., Sato, S., Zgłobicka, I., Witak, M., 2018. The morphology and molecular phylogenetics of some marine diatom taxa within the Fragilariaceae, including twenty undescribed species and their relationship to Nanofrustulum. Opephora and Pseudostaurosira, Phytotaxa, p. 355.

Löfgren, S., Forsius, M., Andersen, T., 2003. The color of water – climate induced water color increase in Nordic lakes and streams due to humus. Nordic council of Ministers brochure 12 pp.

Mann, David G., 1999. The species concept in diatoms. Phycologia 38 (6), 437–495.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 17, 3 %J EMBnet journal.

Ólafsson, J.S., Ingimundardóttir, G.V., Hansen, I., Sigurðardóttir, S.G., 2010. Smádýralíf í afrennslisvatni frá háhitasvæðunum við Kröflu, Ölkelduháls og í Miðdal í Henglinum. Veiðimálastofnun VMST/10019.

Passy, S.I., 2007. Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. Aquat. Bot. 86 (2), 171–178.

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.

Poikane, S., Várbíró, G., Kelly, M.G., Birk, S., Phillips, G., 2021. Estimating river nutrient concentrations consistent with good ecological condition: More stringent nutrient thresholds needed. Ecol. Indicat. 121.

Potapova, M., Snoeijs, P., 1997. The natural life cycle in wild populations of Diatoma moniliformis (Bacillariophyceae) and its disruption in an aberrant environment. J. Phycol. 33, 924–937.

Rimet, F., Bouchez, A., 2011. Use of diatom life-forms and ecological guilds to assess pesticide contamination in rivers: lotic mesocosm approaches. Ecol. Ind. 11 (2), 489–499.

Rimet, F., Bouchez, A., 2012. Life-forms, cell-sizes and ecological guilds of diatoms in European rivers. Knowl. Manage. Aquat. Ecosyst. (406), 01.

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, 15116.

Rott, E., Dam, H.V., Pfister, P., Pipp, E., Pall, K., Binder, N., Ortler, K., 1999. Indikationslisten für Aufwuchsalgen. Teil 2: Trophieindikation, geochemische Reaktion, toxikologische und taxonomischen Anmerkungen. Publ. Wasserwirtschaftskataster, BMfLF 1–248.

Šmilauer, P., Lepš, J., 2014. Multivariate Analysis of Ecological Data using CANOCO 5, 2 ed. Cambridge University Press, Cambridge.

Smol, J.P., Stoermer, E.F.e., 2010. The Diatoms: Applications for the Environmental and Earth sciences. Cambridge University Press.

Snoeijs, P., Busse, S., Potapova, M., 2002. The importance of diatom cell size in community analysis. J. Phycol. 38, 265–281.

Soininen, J., Jamoneau, A., Rosebery, J., Passy, S.I., 2016. Global patterns and drivers of species and trait composition in diatoms. Glob. Ecol. Biogeogr. 25 (8), 940–950.

Statzner, B., Bêche, L.A., 2010. Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? Freshwater Biol. 55, 80–119.

Tapolczai, K., Bouchez, A., Stenger-Kovács, C., Padisák, J., Rimet, F., 2017. Taxonomy- or trait-based ecological assessment for tropical rivers? Case study on benthic diatoms in Mayotte island (France, Indian Ocean). Sci. Total Environ. 607-608, 1293–1303.

Tapolczai, K., Keck, F., Bouchez, A., Rimet, F., Kahlert, M., Vasselon, V., 2019. Diatom DNA metabarcoding for biomonitoring: strategies to avoid major taxonomical and bioinformatical biases limiting molecular indices capacities. Front. Ecol. Evol. 7.

Tapolczai, K., Selmeczy, G.B., Szabó, B., B-Béres, V., Keck, F., Bouchez, A., Rimet, F., Padisák, J., 2021. The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. Ecol. Ind. 122, 107322.

Vanormelingen, P., Chepurnov, V.A., Mann, D.G., Sabbe, K., Vyverman, W., 2008. Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of Eunotia bilunaris sensu lato (Bacillariophyta). Protist 159 (1), 73–90.

Vasselon, V., Domaizon, I., Rimet, F., Kahlert, M., Bouchez, A., 2017a. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: do DNA extraction methods matter? Freshwater Sci. 36 (1), 162–177.

Vasselon, V., Rimet, F., Tapolczai, K., Bouchez, A., 2017b. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). Ecol. Ind. 82, 1–12.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E., 2007. Let the concept of trait be functional! Oikos 116, 882–892.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. 73, 5261-5267.

Werner, P., Adler, S., Dreßler, M., 2016. Effects of counting variances on water quality assessments: implications from four benthic diatom samples, each counted by 40 diatomists. J. Appl. Phycol. 28 (4), 2287–2297.

Zelinka, M., Marvan, P., 1961. Zur Präzisierung der biologischen Klassifikation der Reinheit fliessender Gewässer. Arch. Hydrobiol. 57, 389–407.