Applicability of diatom metabarcoding in the ecological status assessment of Hungarian lotic and soda pan habitats

Mónika Duleba [a], Angéla Földi [a], Adrienn Micsinai [b], Gábor Várbíró [c], Anita Mohr [b], Rita Sipos [b], Gyula Szabó [b], Krisztina Buczkó [a], Zsuzsa Trábert [a], Keve T. Kiss [a], Tibor Bíró [d], Edit Vadkerti [d], Éva Ács [a,d,*]

[a] Centre for Ecological Research, Institute of Aquatic Ecology, Karolina út 29., Budapest H-1113, Hungary
[b] Biomi Ltd, Gödöllő, Hungary
[c] Centre for Ecological Research, Institute of Aquatic Ecology, Department of Tisza Research, Debrecen, Hungary
[d] University of Public Service, Faculty of Water Sciences, Department of Water Supply and Sewerage, Baja, Hungary

A R T I C L E   I N F O

A B S T R A C T

Diatoms are widely applied in the ecological status assessment of aquatic ecosystems using indices calculated from pollution sensitivity and indicator values of species. Traditional, morphology-based identification of species requires in-depth taxonomic knowledge and expertise. Identifying taxa according to their barcode sequences obtained with high-throughput sequencing (metabarcoding) would be a promising alternative. In this pilot study, we tested the applicability of metabarcoding of benthic diatom assemblages for the ecological status assessment of Hungarian water bodies, comparing its performance to that of morphology-based identification of species. The barcode region of the *rbc*L gene was investigated in samples from running waters with various trophic states and unique lentic habitats, namely soda pans. For running waters, the Specific Pollution Sensitivity Index (IPS) and for soda pans the H index and the Indice Biologique Diatomées (IBD) were calculated from the morphology- and sequence-based abundances. The ecological quality ratio was determined for running waters. Overall, more infrageneric taxa were found with microscopy than with metabarcoding in the lotic and lentic samples. The correspondence between taxon lists detected by the two methods was relatively low but increased considerably in the case of morphologically dominant taxa. Community composition based on microscopy and DNA sequence analysis showed a significant correlation and was determined by the same main environmental drivers. Morphology-based indices strongly correlated with sequence-based indices. Both aspects indicated the same ecological status class for more than half of the lotic samples. In other cases, a status shift from good to moderate was frequent, a phenomenon that could prove problematic because the Water Framework Directive prescribes intervention for waters with a moderate or worse status. Considering discrepancies between the results obtained with microscopy and metabarcoding, using both methods in parallel could be proposed until the reference database has been suitably updated.

1. Introduction

Humans influence aquatic ecosystems in many ways through urbanisation (Søndergaard and Jeppesen, 2007), transport, agriculture and industrial activities. These involve physical (e.g. hydromorphological modifications; Elosegi and Sabater, 2013), chemical (various kinds of pollutions originating from point and diffuse sources; Nikanorov and Stradomskaya, 2007; Wen et al., 2017) and biological impacts (e.g. introduction of alien species; Richardson, 2011).

Considering the risk anthropogenic loading poses to aquatic ecosystems, it is important to detect its extent and effects reliably. Physicochemical measurements may be affected by pollution that was introduced just before sampling, but such intermittent pollution events have a less pronounced effect on indicator organism communities (Szczepocka and Szulc, 2009). It has been shown that communities of aquatic organisms such as macroinvertebrates (Johnson et al., 1993) or diatoms (Carayon et al., 2019; Hambrook Berkman and Porter, 2004; Makovinská et al., 2014) can integrate the effects of environmental pressures that have

lasted for longer periods and, therefore, are better indicators of water quality than physicochemical measurement values that reflect instantaneous conditions.

Diatoms are eukaryotic single-celled photosynthetic organisms that have a silica cell wall called frustule (Round et al., 1990). These algae constitute a frequently dominant group in aquatic ecosystems in pelagic and benthic habitats and play a significant role in biogeochemical cycling of elements, especially silicon (Burliga and Kociolek, 2016). They exhibit great species diversity (Mann and Vanormelingen, 2013), a short generation time (Round et al., 1990) and the ability to respond directly and sensitively to environmental changes (Lobo et al., 2016). These characteristics allow researchers to use diatoms as indicators in the ecological status assessment of aquatic ecosystems (Ector et al., 2004; Lobo et al., 2016), for which they have been applied since the late 1940s (Rimet, 2012). In 2000, in the European Union Water Framework Directive (European Commission, 2000) benthic diatoms were chosen as one of five biological elements based on which ecological status of surface waters should be assessed. This assessment uses diatom indices calculated from the pollution sensitivity and indicator values as well as relative abundances of diatom taxa in benthic communities (Berthon et al., 2011). The accurate identification of species is essential for the assessment. Traditionally, the identification of diatom species has been based on the morphological features of the frustule (Round et al., 1990) that can be investigated under a light microscope or a scanning electron microscope in the case of small or similar taxa. However, morphology-based species identification is labour-intensive and requires in-depth taxonomic knowledge and expertise (Rimet and Bouchez, 2012). Introduction of molecular methods to diatom community analysis offers a new approach allowing objective and specific species identification. Short DNA sequences called barcodes (Hebert et al., 2003) can be used to distinguish species. Metabarcoding combines barcodes and high-throughput sequencing (HTS) to identify species in community samples (Taberlet et al., 2012). A reliable and curated reference library in which the taxonomic assignment of sequences is unambiguous is required for this endeavour (Rimet et al., 2019; Rimet et al., 2016). This process builds more upon current state-of-the-art knowledge than another concept according to which sequences originated from metabarcoding without taxonomic assignment can be used in ecological status assessment (Tapolczai et al., 2021; Tapolczai et al., 2019). However, this requires a verification study to assign an ecological profile to these units that for many of the species are known.

An important question is the selection of the suitable DNA region that holds enough taxonomic information to provide species-level resolution. Until now, two possible candidates have been recommended for diatoms: the nuclear 18S rRNA gene (18S rDNA) (Visco et al., 2015; Zimmermann et al., 2015; Zimmermann et al., 2014) and the plastid *rbc*L gene (Kermarrec et al., 2014; Kermarrec et al., 2013b; Vasselon et al., 2017). Of the two, *rbc*L has the advantages of high taxonomic resolving power and the larger and more comprehensive reference database (Apothéloz-Perret-Gentil et al., 2020; Kermarrec et al., 2013a).

Several studies have been conducted to test metabarcoding in the ecological status assessment of waters of various areas (Bailet et al., 2019; Borrego-Ramos et al., 2021; Kelly, 2019; Mora et al., 2019; Mortágua et al., 2019). Metabarcoding of benthic diatom assemblages of Hungarian waters has been carried out in two studies. First, during the Fourth Joint Danube Survey (JDS4), phytobenthic samples from the entire section of the River Danube were compared based on both microscopy and environmental DNA (eDNA) analysis (Zimmermann et al., 2020). Tapolczai et al. (2021) applied a taxonomy-free approach without assigning exact sequence variants to traditional taxonomy to assess the effect of land-use on diatom assemblages.

In this paper we present a pilot study to assess the performance of morphological investigation and metabarcoding in the ecological status assessment of Hungarian waters from a diverse set of lotic environments and soda pans.

2. Materials and methods

2.1. Sampling

Samples were taken in the framework of a monitoring project with the aim of assessing the ecological status of Hungarian lotic and lentic environments based on benthic diatom assemblages. When selecting lotic environments, the principal selection criterion was to include samples from different trophic statuses. Trophic status was determined based on the total phosphorous concentration that significantly (p < 0.001) and negatively correlated with morphology-based ecological quality ratio (Pearson correlation coefficient: −0.42, Fig. 1). Most of the sampled running waters were calcareous, but four siliceous streams were among them. Besides, they have various catchment size (from small to very large), landscape (hilly-mountainous, hilly and lowland) and altitude (from high to low), as well as sediments grain size (coarse and medium-fine) (Supplementary Table 1). Lake samples were taken from soda pans, which represent a unique type of habitat in the Carpathian Basin (Tóth et al., 2014) regarding their physical and chemical characteristics, the inhabiting biota, distribution and vulnerability. Soda pans are the only natural saline waters in Central Europe, the Carpathian basin their westernmost occurrence in Eurasia (Horváth et al., 2013). These habitats provide specific environmental conditions such as high salinity, alkalinity and turbidity, hypertrophic conditions, in case of astatic ponds fluctuating water level and temperature (Boros et al., 2017) that maintain a unique biota consisting of taxa with specific tolerances and requirements. Many endemic and rare species can be found in these environments (Vad et al., 2017). The number of soda pans has seriously declined since the 18th century (Horváth et al., 2013) because of climate change and catchment activities e.g. water course diversion, dredging, excessive pumping of surface freshwater or ground water (Stenger-Kovács et al., 2014). Worse ecological status of soda pans in most cases means decreasing sodic character (Földi et al., 2018).

Ninety-three samples were selected for metabarcoding in parallel to morphological investigation. These involved 79 samples from rivers and streams selected along a trophic gradient as well as 14 samples from soda pans (Fig. 2, Supplementary Table 1). The samplings from rivers and streams were carried out from late March–early April until the end of April 2019; the samplings from soda pans were carried out in May–June 2019, except for Lake Kakasszéki, where the sample was taken at the beginning of October 2019 due to the drying of the lake in late spring–summer. Samplings were carried out according to MSZ EN

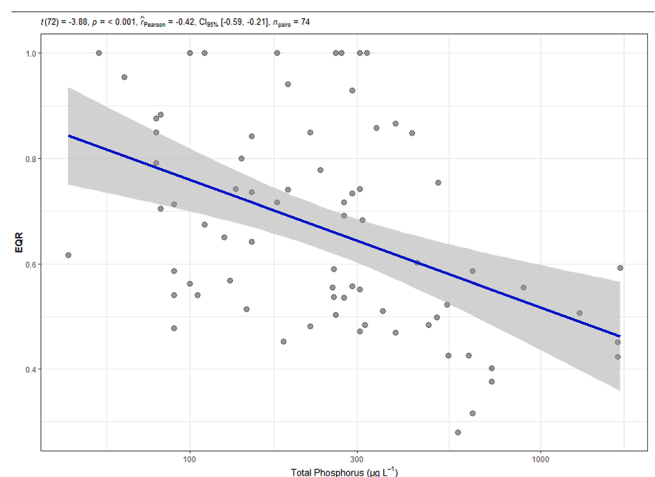**Fig. 1.** The relationship between the total phosphorous (TP) concentration and the ecological quality ratio (EQR) calculated based on the microscopic investigations. Results of statistical tests are indicated in the upper part of the figure: Student's *t* test with significance (p) value, Pearson correlation coefficient ($\hat{r}_{Pearson}$), 95% confidence interval ($CI_{95\%}$) and number of elements (n_{pairs}).
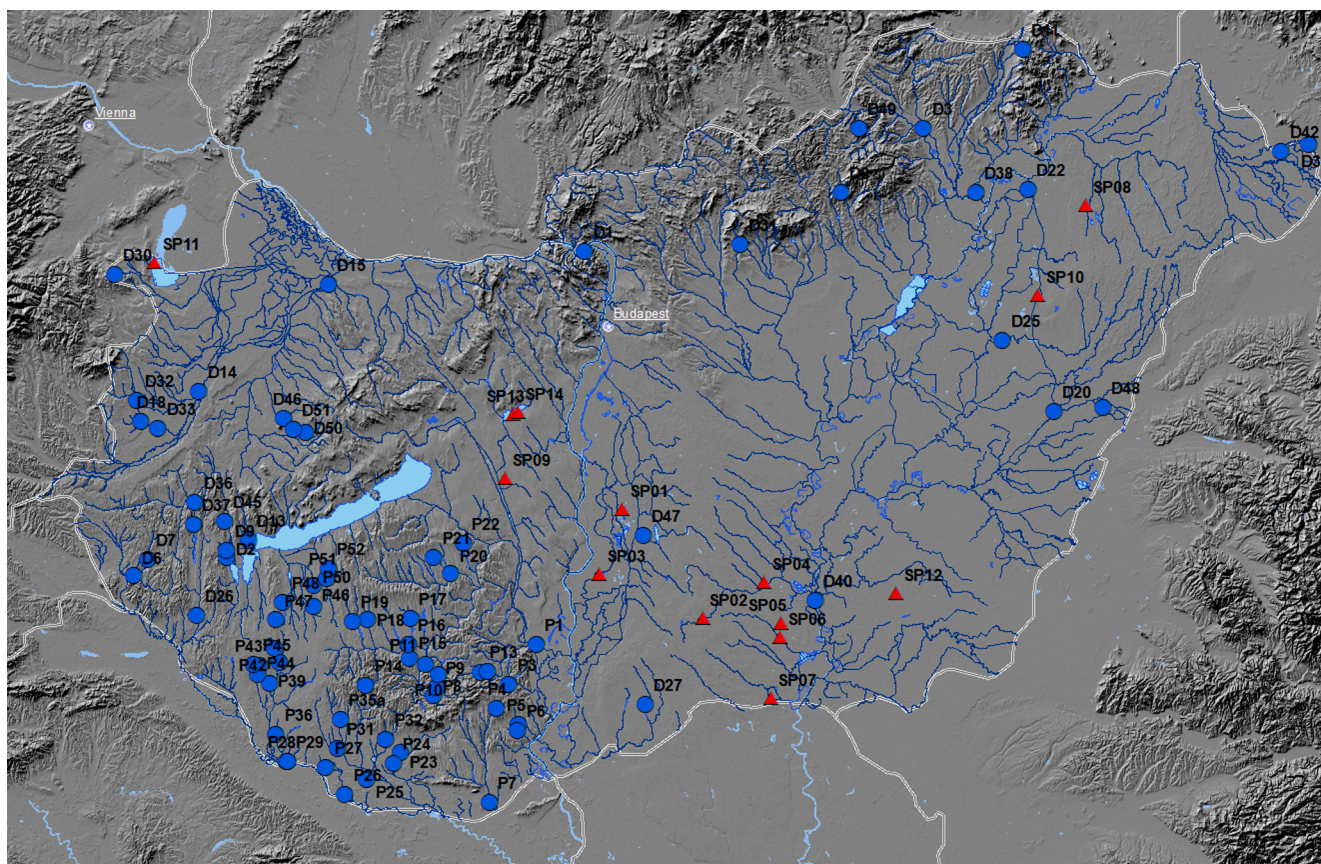
Fig. 2. Map of the sampling sites: blue dots represent rivers and stream and red triangles indicate soda pans. Description and codes of sampling sites can be found in Supplementary Table 1.

13946:2014 standard (Hungarian Standards Institution, 2014). Samples were taken from substrates most typical to the sampling site paying attention to that the sampled substrate should be in the euphotic zone and under water presumably for 4–6 weeks before the sampling. In lotic habitats, epilithic samples were taken from five randomly chosen cobbles; if these were absent, other available substrates, mainly emergent macrophytes, rarely dead-wood or artificial substrates (e.g. brick), were sampled. In the case of soda pans, epiphytic samples were taken from five randomly chosen stems of green common reed (*Phragmites australis* (Cav.) Trin. ex Steud.), a10-cm section of the stems starting at 10 cm below the water surface was cut. For all kinds of substrates, the five random repeats were integrated into one composite sample and biofilm from them was scraped with a toothbrush into tap water. The acquired slurry was homogenised and divided into two parts. For DNA analysis, 2–3 ml was pipetted into a 15-ml sterile plastic centrifuge tube that was filled with absolute ethanol (final ethanol concentration $\geq$ 70%), then stored at 4°C until processing. The rest of the slurry was preserved with formaldehyde (4% final concentration) for microscopic investigation (European Committee for Standardisation, 2002).

2.2. Environmental variables

Within the framework of the project, several environmental variables were measured, of which eight were used in our analyses. pH, dissolved oxygen content (Diss_O, mg l^{-1}) and conductivity (cond, μS cm^{-1}) were measured *in situ* with a portable multiparameter digital meter (Multi 350i-WTW, Germany). For further laboratory analysis, 1 L water sample was taken and stored cooled until processing. Samples were taken by tube sampler from the trophic layer (2.5 × Secchi depth) at the deepest parts of the lakes. In the case of water bodies where the maximum depth (Zmax) was < 2 m, the whole water column has been

sampled. In rivers samples were taken from the thalweg. From this sample, total nitrogen (TN, mg l^{-1}; MSZ EN 12260:2004; Hungarian Standards Institution, 2004), total phosphorous concentration (TP, mg l^{-1}, MSZ EN ISO 11885:2009; Hungarian Standards Institution, 2009), chlorophyll *a* concentration (CHA μg l^{-1}, MSZ ISO 10260:1993; Hungarian Standards Institution, 1993), chemical oxygen demand (COD mg l^{-1}, ISO 15705:2002; International Organization for Standardization, 2002) and biological oxygen demand (BOD, mg l^{-1}, MSZ ISO 6060:1991; Hungarian Standards Institution, 1992) were measured according to the national standard.

2.3. Morphological investigation

Diatom valves were cleaned using hydrogen peroxide and hydrochloric acid, washed in distilled water and then mounted with Naphrax mountant (European Committee for Standardisation, 2002). Diatom taxa were identified under a Zeiss Z2 Axio Imager microscope equipped with differential interference contrast (DIC) at a magnification of 1600×. The minimum number of values identified to the species or genus level was 400 for rivers and stream samples and 500 for soda pan samples.

In the case a species where identification under the light microscope was uncertain (because of their small size or similarity), we applied scanning electron microscopy. For this endeavour, cleaned valves were filtered onto an Isopore polycarbonate membrane filter with a 3-μm pore diameter (Merck Millipore), fixed on metal stubs with double-sided carbon tape, painted in spots with conductive silver paint (NO-VOC Silver Paint, SPI Supplies) and coated with gold using a rotary-pumped spatter coater Quorum Q150R S. The prepared samples were investigated with a Zeiss EVO MA 10 scanning electron microscope operated at 10 kV and 10 mm working distance using secondary electron detectors.

2.4. Metabarcoding

DNA was extracted from samples using the NucleoSpin Soil Kit (Macherey-Nagel) following the protocol by Vautier et al. (2020) with a modification that the mixture of the two kinds of lysis buffer provided in the kit was applied in a 1:1 ratio.

Polymerase chain reactions (PCR), library preparation and sequencing were performed by amplifying a 312 base pair (bp) partial target of the *rbc*L gene using gene-specific primers (Vasselon et al., 2017) with Illumina overhang P5/P7 adapters (Table 1).

Forward and reverse primers were mixed in equimolar quantities and used as the RbcL-Forward-Primer mix and the RbcL-Reverse-Primer Mix to amplify the 312 bp *rbc*L gene target in a single multiplex PCR. The final reaction volume was 25 µl, including 12.5 µl Q5® High-Fidelity 2X Master Mix (New England Biolabs), 5 µl each of the 2 µM Forward and the 2 µM Reverse Primer mixes, 1 µl of the DNA sample, and 1.5 µl of the molecular-biology-grade water (Lonza). The PCR was carried out in an Applied Biosystems Veriti 96-well thermal cycler with the following thermal profile: initial denaturation at 95°C for 3 min; 25 cycles of 98°C for 30 sec, 55°C for 40 sec and 72°C for 30 sec; and final extension at 72°C for 5 min. If this protocol returned poor results, the PCR reaction was repeated with an increased cycle number of up to 30.

The PCR products were purified using 1.0x AMPure XP magnetic beads (Beckman Coulter) according to the manufacturer's instructions and eluted in 25 µl of Buffer EB (Qiagen). Concentrations were quantified by using a Qubit™ 4 Fluorometer (Invitrogen) with the Qubit™ dsDNA HS Assay Kit. Cleaned PCR products were diluted to equimolar concentrations for indexing reactions.

Nextera DNA CD Indexes (96 Index, 96 Samples) were sourced from Merck with P5/P7 adapters and P7/P5 tags attached (Table 2). Index reactions were performed in a 50 µl final volume including 25 µl Q5® High-Fidelity 2X Master Mix (New England Biolabs), 5 µl of each 1 µM index primer, 5 µl of cleaned PCR product from the first PCR reaction and 10 µl of molecular-biology-grade water. The thermal profile was the following: initial denaturation at 95°C for 3 min; 8 cycle of 98°C for 30 sec, 55°C for 40 sec and 72°C for 30 sec; and final extension at 72°C for 5 min.

The PCR products were purified using 1.2x AMPure magnetic beads according to the manufacturer's instructions and eluted in 25 µl of Buffer EB. Quality and concentration control of the samples was performed with a Qubit™ 4 Fluorometer using the Qubit™ dsDNA HS Assay Kit reagents (concentration) and with an Agilent TapeStation System 4150 (Agilent) using the Agilent High Sensitivity D1000 ScreenTape Assay Reagents. Cleaned PCR products were diluted to a similar concentration and pooled properly. The final library pool was checked for size and concentration and diluted to a 4 nM starting concentration for the run on the Illumina MiSeq system. After denaturation and dilution, sequencing was performed using the Illumina MiSeq V2 (500) Reagent Kit and a 2 × 250 bp read length.

2.5. Bioinformatics analysis

For bioinformatics analysis, we used a workflow by Keck et al. (2019) that follows the official DADA2 pipeline, which is a custom script to process Illumina MiSeq HTS data directly (Callahan et al., 2016). The modified script includes specific needs and has proven to work well for diatom metabarcoding with *rbc*L (Keck et al., 2019).

Primer sequences were removed from R1 and R2 reads with cutadapt (Martin, 2011). The read quality profile was inspected and it indicated good quality. R1 and R2 reads were trimmed to 200 and 170 nucleotides (nt), respectively, and then filtered using the criteria of zero ambiguous nucleotides and a maximum of two expected errors. A parametric error model was learnt from the data by alternating estimation of the error rates and inference of sample composition until they converged on a jointly consistent solution. The estimated error rates were a good fit to the observed rates and, as expected, the error rates dropped as the quality increased. Then R1 and R2 reads were dereplicated and filtered with the core sample inference algorithm of DADA2. Paired forward and reverse reads were aligned and merged to one contig sequence. Chimaeras were removed. The read numbers of samples were tracked after each step of the pipeline. In one sample (D41), only raw reads were found; in another sample (D22), 47% of the merged sequences were lost after removing chimaeras. Therefore, these samples were removed from further analysis. The taxonomic assignment step was done using Diat. barcode version 9.2 (Rimet et al., 2019) as the reference database and a minimum bootstrap confidence of 75 for assigning a taxonomic level. This method searches short, 8-nt-long sequences in target and reference sequences and identifies target as the species of the database for which the sequence shows the greatest identity with the target. Read numbers were corrected according to cell biovolume using the correction factor developed by Vasselon et al. (2018).

2.6. Statistical analysis

We used relative abundances to compare the results of the morphological investigation and metabarcoding. In each sample, we investigated the number of taxa shared by the two approaches and the number of taxa that were dominant (their relative abundance was at least 5%) according to microscopy and recorded in the sample also by metabarcoding.

Number of taxa detected based on morphology as well as based on DNA sequences were illustrated on Venn diagram using 'draw.pairwise. venn' function in VennDiagram package (Chen, 2018) in R environment (R Core Team, 2020).

The Bray–Curtis dissimilarity index was calculated between samples identified based on morphology and DNA, then linear correlation between the two matrices was calculated using Mantel's permutation test with 999 permutations.

To reveal the relationship between environmental variables and the composition of diatom assemblages based on morphology and DNA, redundancy analysis (RDA) was performed. These analyses were performed in case of lotic samples. Number of soda pan samples was considered to be too low for this. RDA was carried out in the R environment (R Core Team, 2020) with the vegan package (Oksanen et al., 2018).

The map was created with ArcMap ESRI.

The following diatom indices were used: for rivers and streams, the Specific Pollution Sensitivity Index (IPS; Coste as cited in Cemagref, 1982) and the multimetric index IPSITI (Várbíró et al., 2012), for soda pans, the H index (Ziemann et al., 2001) and the Indice Biologique

Table 1
Primers used for the *rbc*L polymerase chain reaction.

Name	Illumina overhang P5 tag (5′ → 3′)	Locus-specific primer sequence (5′ → 3′)
Diat_rbcL_708F_1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	AGGTGAAGTAAAAGGTTCWTACTTAAA
Diat_rbcL_708F_2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	AGGTGAAGTAAAAGGTTCWTAYTTAAA
Diat_rbcL_708F_3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	AGGTGAAACTAAAAGGTTCWTACTTAAA
Name	Illumina overhang P7 tag (5′ → 3′)	Locus-specific primer sequence (5′ → 3′)
R3_1R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	CCTTCTAATTTACCWACWACTG
R3_2R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	CCTTCTAATTTACCWACAACAG

Table 2
Index primers used in polymerase chain reactions.

Name	Illumina P7 adapter (5′ → 3′)	i7 index (5′ → 3′)	Illumina P7 tag (5′ → 3′)
H701	CAAGCAGAAGACGGCATACGAGAT	TCGCCTTA	GTCTCGTGGGCTCGG
H702	CAAGCAGAAGACGGCATACGAGAT	CTAGTACG	GTCTCGTGGGCTCGG
H703	CAAGCAGAAGACGGCATACGAGAT	TTCTGCCT	GTCTCGTGGGCTCGG
H705	CAAGCAGAAGACGGCATACGAGAT	AGGAGTCC	GTCTCGTGGGCTCGG
H707	CAAGCAGAAGACGGCATACGAGAT	GTAGAGAG	GTCTCGTGGGCTCGG
H723	CAAGCAGAAGACGGCATACGAGAT	GAGCGCTA	GTCTCGTGGGCTCGG
H706	CAAGCAGAAGACGGCATACGAGAT	CATGCCTA	GTCTCGTGGGCTCGG
H712	CAAGCAGAAGACGGCATACGAGAT	TCCTCTAC	GTCTCGTGGGCTCGG
H720	CAAGCAGAAGACGGCATACGAGAT	AGGCTCCG	GTCTCGTGGGCTCGG
H710	CAAGCAGAAGACGGCATACGAGAT	CAGCCTCG	GTCTCGTGGGCTCGG
H711	CAAGCAGAAGACGGCATACGAGAT	TGCCTCTT	GTCTCGTGGGCTCGG
H714	CAAGCAGAAGACGGCATACGAGAT	TCATGAGC	GTCTCGTGGGCTCGG
Name	Illumina P5 adapter (5′ → 3′)	i5 index (5′ → 3′)	Illumina P5 tag (5′ → 3′)
H505	AATGATACGGCGACCACCGAGATCTACAC	GTAAGGAG	TCGTCGGCAGCGTC
H517	AATGATACGGCGACCACCGAGATCTACAC	GCGTAAGA	TCGTCGGCAGCGTC
H506	AATGATACGGCGACCACCGAGATCTACAC	ACTGCATA	TCGTCGGCAGCGTC
H503	AATGATACGGCGACCACCGAGATCTACAC	TATCCTCT	TCGTCGGCAGCGTC
H516	AATGATACGGCGACCACCGAGATCTACAC	CCTAGAGT	TCGTCGGCAGCGTC
H522	AATGATACGGCGACCACCGAGATCTACAC	TTATGCGA	TCGTCGGCAGCGTC
H513	AATGATACGGCGACCACCGAGATCTACAC	TCGACTAG	TCGTCGGCAGCGTC
H510	AATGATACGGCGACCACCGAGATCTACAC	CGTCTAAT	TCGTCGGCAGCGTC

Diatomées (IBD; Coste and Prygiel, 1998; Lenoir and Coste, 1996; Prygiel, 2000). The IPS, the Austrian Saprobic Index (SI; Rott et al., 1997) and the Austrian Trophic Index (TI; Rott et al., 1999) were calculated using OMNIDIA 6.0.8 (Lecointe et al., 1993). In addition, Van Dam halobity values of taxa (Van Dam et al., 1994) required for calculating the H index (Ács et al., 2015; Földi et al., 2018) were obtained from the database of the program. The IPSITI was calculated according to the following equation:

$$IPSITI = \frac{IPS + SI + TI}{3}.$$

For the running waters the ecological quality ratio (EQR) was calculated from the indices following the recommendations by Várbíró et al. (2012). Briefly, each sampled water body was rated to types according to landscape, altitude, geology, sediment and catchment size. The types of the sampled waters are provided in Supplementary Table 1. Adequate diatom index to be used was chosen according to type of water body (Table 3.). The value of the index indicated an ecological status class for which a transforming equation belonged to (Table 3). Index value (x) was converted into EQR (y) with this equation.

Number of soda pan samples was considered to be too low for drawing consequences on metabarcoding effect on ecological status assessment in case of these habitats.

The relationship between diatom indices as well as EQR values obtained based on the results of the two approaches – morphology and metabarcoding (with and without the correction factor) – was determined and Pearson's product-moment correlation were calculated using the package ggstatsplot (Patil, 2021) in the R environment (R Core Team, 2020). Scatterplots with statistical results were made using the function 'ggscatterstats', boxplots with the function 'ggbetweenstats'. The number of cases that occurred when the ecological class indicated by morphological investigation differed from the class indicated by metabarcoding was investigated and illustrated on a classification table.

To identify which taxa cause the main difference between morphology- and DNA-based indices, similarity percentage (SIMPER) analyses were carried with the PAST software (Hammer et al., 2001).

3. Results

3.1. Morphological investigations

For overall identification, there were 347 species belonging to 80 genera and 28 families in rivers and streams and 139 species and infraspecific taxa belonging to 34 genera and 17 families in soda pans. In lotic samples, there were 19–83 (mean 42) taxa per sample and in soda pan samples, there were 11–41 (mean 23) taxa per sample.

In rivers and streams, the five most abundant taxa were *Achnanthidium minutissimum* (Kützing) Czarnecki, *Navicula gregaria* Donkin, *Planothidium frequentissimum* (Lange-Bertalot) Lange-Bertalot, *Amphora pediculus* (Kützing) Grunow and *Planothidium lanceolatum* (Brébisson ex Kützing) Lange-Bertalot. In soda pans, the five most abundant taxa were *Nitzschia paleacea* (Grunow) Grunow in Van Heurck*, A. minutissimum*, *Nitzschia aurariae* Cholnoky, *Nitzschia perminuta* (Grunow) M. Peragallo and *Nitzschia reskoi* Ács, Duleba, C.E. Wetzel & Ector.

The morphology-based IPS values of rivers and streams ranged from 5.6 to 18.4, the H index of soda pans varied from 3.5 to 13.5 and the IBD ranged from 1.2 to 18.1. Using microscopic counting for calculation of indices, the trophic status of the running waters was: 21 high, 17 good, 34 moderate and 3 poor.

3.2. Metabarcoding

After the final step of the bioinformatics processing, there were a total of 1,049,315 reads (886,099 reads in rivers and streams and 163,216 reads in soda pans) with an average of 11,283 reads per sample (11,216 reads in rivers and streams and 11,659 reads in soda pans). These constituted a total of 1,412 amplicon sequence variants (ASVs) of which 277 (47,983 reads) did not belong to the phylum Bacillariophyta (seven non-Bacillariophyta ASVs were assigned at the species level, namely *Vaucheria bursata* (O.F. Müller) C.Agardh, a xanthophycean alga). Among the 1,135 Bacillariophyta ASVs, 612 (54%) were assigned at the species level, 198 (17%) at only the genus level and 38 (3%) at only the family level. The remaining (25%) were assigned at only the order or higher level.

In rivers and streams, 190 species were obtained (belonging to 69 genera and 8 families), with 5–65 species (mean 36) per sample. In soda pans, 118 species were obtained (belonging to 36 genera and 7 families), with 11–41 species (mean 23) per sample. In rivers and streams, *A. minutissimum, A. pediculus, Mayamaea permitis* (Hustedt) Bruder & Medlin, *N. gregaria* and *Nitzschia media* Hantzsch were the most abundant species after applying the correction factor. In soda pans, *A. minutissimum, Nitzschia supralitorea* Lange-Bertalot, *N. reskoi, Gomphonema saprophilum* (Lange-Bertalot & Reichardt) Abarca, R. Jahn, J. Zimmermann & Enke and *Nitzschia palea* (Kützing) W. Smith were the

Table 3

Indices, index boundaries and equations used to calculate the ecological quality ratio (EQR) according to the type of water. Sample information (type classifications) can be found in Supplementary Table 1. In equations x is the value of the diatom index to be transformed, y is the EQR value.

Type	Index[a]	Equation	Index boundaries	EQR[b] range
1S	Reference	IPS	$y = 0.1176 \times x - 1.0588$	≥ 16.4	
	High		$y = 0.1176 \times x - 1.0588$	≥ 15.8	$0.8 \leq EQR$
	Good		$y = 0.1176 \times x - 1.0588$	≥ 14.1	$0.6 \leq EQR < 0.8$
	Moderate		$y = 0.0426 \times x$	≥ 9.4	$0.4 \leq EQR < 0.6$
	Poor		$y = 0.0426 \times x$	≥ 4.7	$0.2 \leq EQR < 0.4$
	Bad		$y = 0.0426 \times x$	≥ 0	$0 \leq EQR < 0.2$
2S, 2M	Reference	IPSITI	$y = 0.0909 \times x - 0.4545$	≥ 14.3	
	High		$y = 0.0909 \times x - 0.4545$	≥ 13.8	$0.8 \leq EQR$
	Good		$y = 0.0909 \times x - 0.4545$	≥ 11.6	$0.6 \leq EQR < 0.8$
	Moderate		$y = 0.0519 \times x + 0.0017$	≥ 7.7	$0.4 \leq EQR < 0.6$
	Poor		$y = 0.0519 \times x + 0.0017$	≥ 3.9	$0.2 \leq EQR < 0.4$
	Bad		$y = 0.0519 \times x + 0.0017$	≥ 0	$0 \leq EQR < 0.2$
3S, 3M, 5S, 5M	Reference	IPSITI	$y = 0.25 \times x - 2.175$	≥ 12.2	
	High		$y = 0.25 \times x - 2.175$	≥ 11.9	$0.8 \leq EQR$
	Good		$y = 0.25 \times x - 2.175$	≥ 11.1	$0.6 \leq EQR < 0.8$
	Moderate		$y = 0.0541 \times x$	≥ 7.4	$0.4 \leq EQR < 0.6$
	Poor		$y = 0.0541 \times x$	≥ 3.7	$0.2 \leq EQR < 0.4$
	Bad		$y = 0.0541 \times x$	≥ 0	$0 \leq EQR < 0.2$
4L	Reference	IPSITI	$y = 0.0556 \times x - 0.1167$	≥ 16.6	
	High		$y = 0.0556 \times x - 0.1167$	≥ 16.5	$0.8 \leq EQR$
	Good		$y = 0.0556 \times x - 0.1167$	≥ 12.9	$0.6 \leq EQR < 0.8$
	Moderate		$y = 0.0465 \times x$	≥ 8.6	$0.4 \leq EQR < 0.6$
	Poor		$y = 0.0465 \times x$	≥ 4.3	$0.2 \leq EQR < 0.4$
	Bad		$y = 0.0465 \times x$	≥ 0	$0 \leq EQR < 0.2$
6S, 6M	Reference	IPSITI	$y = 0.125 \times x - 0.7$	≥ 12.5	
	High		$y = 0.125 \times x - 0.7$	≥ 12.0	$0.8 \leq EQR$
	Good		$y = 0.125 \times x - 0.7$	≥ 10.4	$0.6 \leq EQR < 0.8$
	Moderate		$y = 0.058 \times x - 0.0019$	≥ 6.9	$0.4 \leq EQR < 0.6$
	Poor		$y = 0.058 \times x - 0.0019$	≥ 3.5	$0.2 \leq EQR < 0.4$
	Bad		$y = 0.058 \times x - 0.0019$	≥ 0	$0 \leq EQR < 0.2$
7L, 8N	Reference	IPSITI	$y = 0.0465 \times x + 0.2279$	≥ 12.6	
	High		$y = 0.0465 \times x + 0.2279$	≥ 12.3	$0.8 \leq EQR$
	Good		$y = 0.0465 \times x + 0.2279$	≥ 8.0	$0.6 \leq EQR < 0.8$
	Moderate		$y = 0.0755 \times x - 0.0025$	≥ 5.3	$0.4 \leq EQR < 0.6$
	Poor		$y = 0.0755 \times x - 0.0025$	≥ 2.7	$0.2 \leq EQR < 0.4$

Table 3 (*continued*)

Type	Index[a]	Equation	Index boundaries	EQR[b] range
	Bad	$y = 0.0755 \times x - 0.0025$	≥ 0	$0 \leq EQR < 0.2$

[a] IPS, Specific Pollution Sensitivity Index; IPSITI, multimetric index calculated from the IPS, the Austrian Saprobic Index and the Austrian Trophic Index.
[b] EQR, ecological quality ratio.

most abundant species after the application of the correction factor.

The DNA-based IPS values of rivers and streams ranged from 3.7 to 18.6 using original read numbers and from 5.1 to 18.6 with cell-biovolume-corrected read numbers. Without read number correction, the H index of soda pans was from 5.5 to 14.9 and the IBD was from 2.2 to 14.2. With correction, the H index varied from 5.2 to 13.8 and the IBD varied from 2.3 to 18.7. Based on uncorrected sequence data, the ecological status of the running waters was: 5 high, 21 good, 45 moderate and 4 poor. When the data were corrected, 13 were high, 14 were good, 40 were moderate and 8 were poor.

3.3. Comparison of the morphological and metabarcoding analyses

Overall, there were 481 species and intraspecific taxa identified in lotic habitats and 153 in soda pans based on morphological and/or DNA sequence analysis. In case of both running waters and soda pans more taxa were found with microscope than with metabarcoding and about a quarter of taxa were detected with both methods (Fig. 3). In rivers and streams, 5%–47% (mean 25%) of taxa observed under there microscope were detected based on sequences; in the case of dominant species (relative abundance greater than 5%), this portion was 25%–100% (mean 70%). In soda pans, 17%–42% (mean 26%) of taxa identified morphologically were found based on metabarcoding (Fig. 4A); in the case of dominant species, this portion was 20%–100% (mean 56%) (Fig. 4B).

The Mantel test was applied to compare the structure of communities identified by the two methods. It revealed significant correlation for both the lotic samples (Pearson's product-moment correlation = 0.5143, p = 0.001) and soda pan samples (Pearson's product-moment correlation = 0.6704, p = 0.001). The correlation was stronger for lotic samples when corrected read numbers were compared to morphology-based abundances; for soda pan samples, however, the outcome was more or less the same (Pearson's product-moment correlation = 0.6261 and 0.6658, respectively, p = 0.001 for both).

RDA showed that the most important variables explaining the relationship between the diatom community and environmental variables in running waters were: TP, pH and conductivity in the case of morphology and TP (Fig. 5A, Table 4), pH and TN in the case of DNA sequences (Fig. 5B, Table 4). Both morphology-based and corrected DNA-based analyses showed a similar distribution of samples, namely large rivers formed a group overlapping with other running waters (Fig. 5).

For both lotic and lentic samples, the morphology-based indices showed a significant (p < 0.01) positive correlation with the DNA-based indices calculated either from uncorrected or cell-biovolume-corrected read numbers, correlation with corrected DNA-based indices was a bit stronger than with the uncorrected DNA-based indices (Figs. 6A and B, 7, Pearson correlation coefficients are indicated on the corresponding figures).

Morphology-based EQR values calculated for lotic habitats were usually higher than ecological quality ratio based on DNA sequences, EQR for uncorrected or corrected read numbers were more or less the same (Fig. 6C). Morphology-based EQR also showed a significant (p < 0.01) positive correlation with EQR for both uncorrected and corrected read numbers (Pearson correlation coefficients: 0.74 and 0.81, respectively).

Comparing the ecological status calculated using morphological and
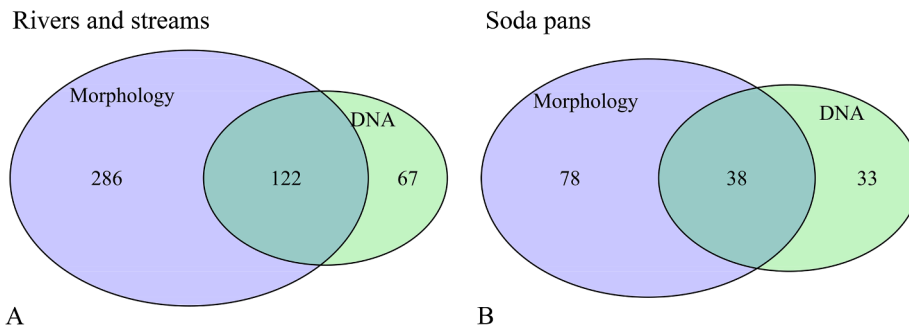
Fig. 3. Venn diagram of the number of taxa detected based on morphology under the microscope and with DNA sequence analysis in samples of rivers and streams (A) as well as soda pans (B).
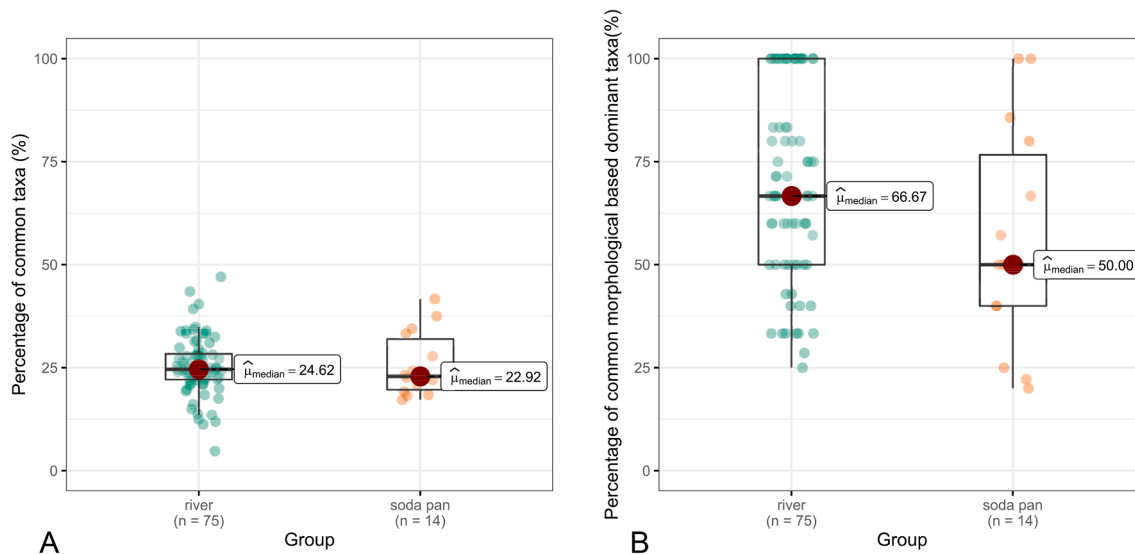
Fig. 4. (A) Boxplot of the number of taxa detected with both morphology and DNA sequence analysis. (B) Boxplot of the portion of taxa that were dominant in morphological analysis and were also detected based on DNA sequence analysis. Median values (μ) and number of elements (n) are indicated on the panels.
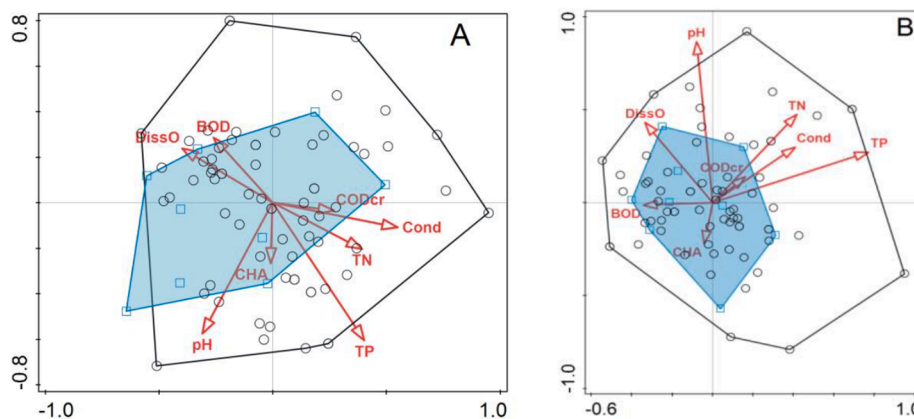
Fig. 5. Position of the samples in the factorial plane of the correlated environmental variables based on (A) morphology and (B) DNA sequence analysis. The arrow lengths represent the relative importance of each variable in the redundancy analysis and the direction of the arrows indicates variable correlations. The shaded areas indicate samples from large rivers.

DNA results of lotic habitats revealed that the status was the same in more than 60% of the samples (47); this number was slightly higher using the correction factor (49 samples). In the other samples, meta-barcoding generally showed one class worse (28% of the samples without correction, 29% with correction), namely a shift from the good to the moderate class occurred frequently. More than one class shift

(from high to moderate) was observed in a few samples. Cases when DNA indicated a better class than morphology were very rare (Table 5).

The ecological status indicated by morphology rather than DNA sequences seemed to correspond better to TP. Specifically, TP was 0.045–0.51 mg/l (mean 0.192 mg/l) in waters with a good or high status according to the morphological analysis, while it was 0.09–1.687 mg/l

Table 4
Variance explained by the environmental variables.

Morphology				DNA			
Simple	Term	Effects		Simple	Term	Effects	
name[a]	explains	%	p	name[a]	explains	%	p
TP	**2.4**	**1.8**	**0.004**	**TP**	**3.4**	**2.5**	**0.002**
pH	**2.4**	**1.8**	**0.006**	**pH**	**3.2**	**2.3**	**0.002**
Cond	**2.3**	**1.7**	**0.006**	**TN**	**2.8**	**2.1**	**0.008**
TN	2	1.5	0.032	Cond	2.7	2	0.012
CHA	1.8	1.4	0.092	CHA	2.2	1.6	0.048
BOD	1.8	1.3	0.088	DissO	1.9	1.4	0.118
CODcr	1.6	1.2	0.192	BOD	1.7	1.3	0.208
DissO	1.6	1.2	0.258	CODcr	1.4	1	0.45

[a] Diss_O, dissolved oxygen content (mg l^{-1}); Cond, conductivity (µS cm^{-1}); TN, total nitrogen (mg l^{-1}); TP, total phosphorous concentration (mg l^{-1}); CHA, chlorophyll a concentration (µg l^{-1}); COD, chemical oxygen demand (mg l^{-1}); BOD, biological oxygen demand (mg l^{-1}).
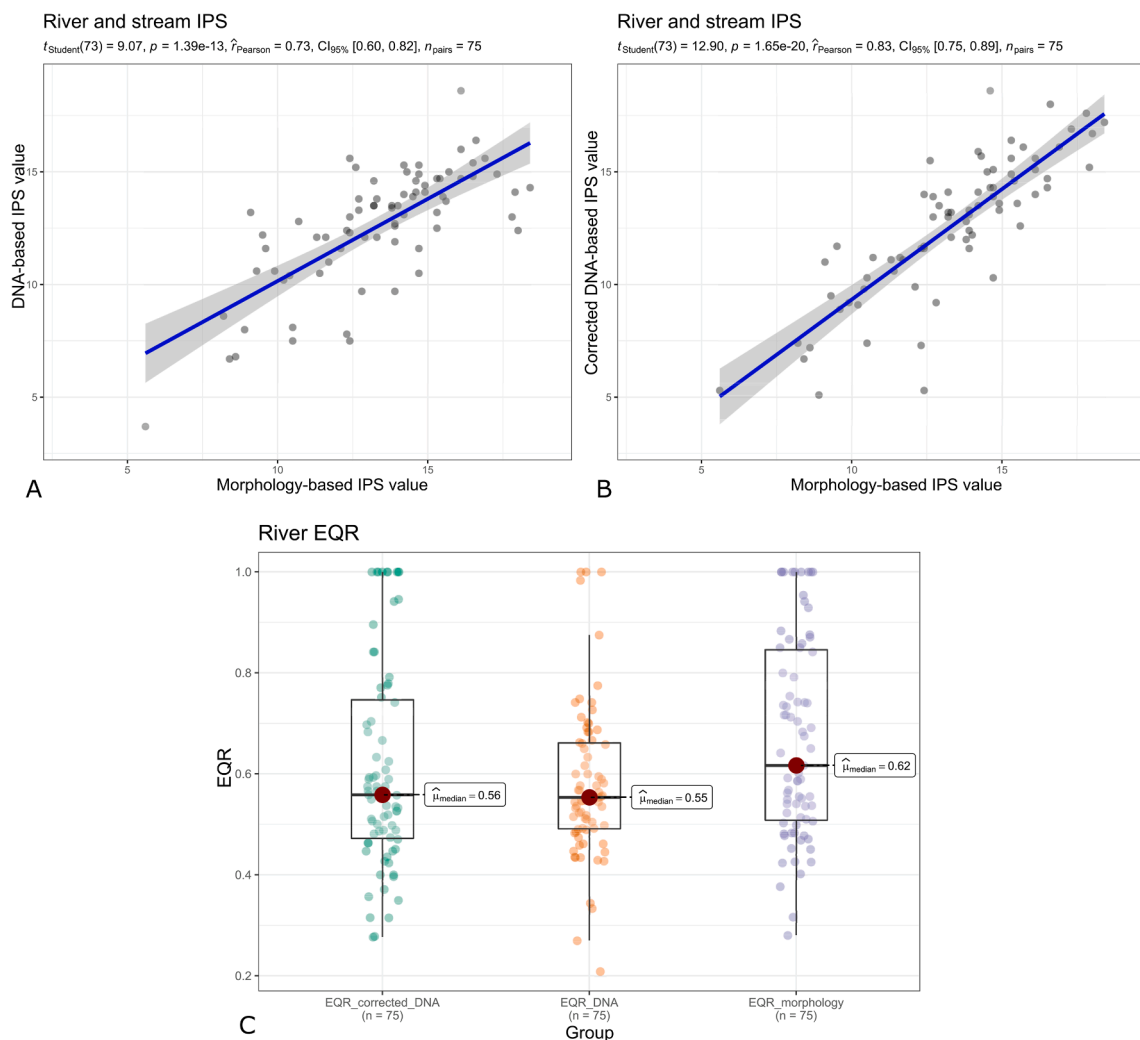

Fig. 6. (A, B) Scatterplots of the morphology-based and DNA-based Specific Pollution Sensitivity Index (IPS) values in rivers and streams. DNA-based indices were calculated using either uncorrected read numbers (A) or read numbers corrected with correction factor (B) developed by Vasselon et al. (2018). Results of statistical tests are indicated in the upper part of the panels A and B: Student's t test with significance (p) value, Pearson correlation coefficient ($\hat{r}_{Pearson}$), 95% confidence interval ($CI_{95\%}$) and number of elements (n_{pairs}). (C) Box plots illustrating the ecological quality ratio (EQR) values calculated from morphology as well as uncorrected and corrected DNA-based results. Median values (µ) and number of elements (n) are indicated on the panel.

(mean 0.508 mg/l) in habitats with a moderate or poor status. In samples that were assigned to a good or high status according to morphology but a moderate status based on corrected DNA data, TP was low (0.045–0.51 mg/l, mean 0.199 mg/l), similarly to those waters that had a good or high status based on both methods (0.055–0.385 mg/l, mean

0.189 mg/l). TP was higher in samples that indicated a moderate status with both methods: 0.09–1.687 mg/l (mean 0.508 mg/l).

SIMPER analysis showed that in running waters *A. pediculus, A. minutissimum, M. permitis, P. lanceolatum, P. frequentissimum* and *Planothidium victori* P.M. Novis, J. Braidwood & C. Kilroy contributed
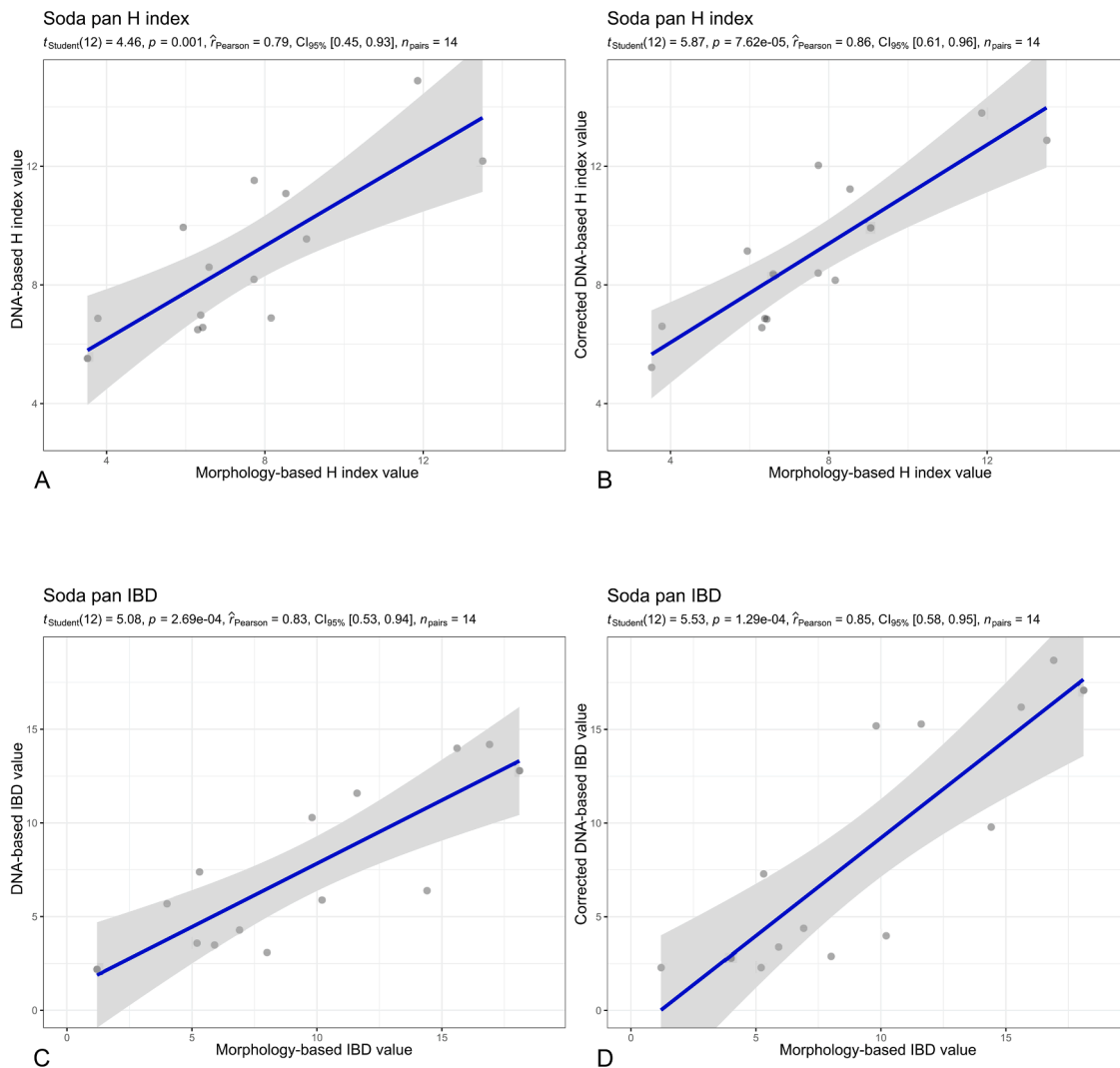
Fig. 7. Scatterplot of morphology-based and DNA-based H index values (A, B) and Indice Biologique Diatomées (IBD) (C, D) values in soda pans. For (A) and (C), DNA-based indices were calculated from uncorrected read numbers. For (B) and (D), DNA-based indices were calculated from read numbers corrected with the correction factor developed by Vasselon et al. (2018). Results of statistical tests are indicated in the upper part of the panels: Student's *t* test with significance (p) value, Pearson correlation coefficient ($\hat{r}_{Pearson}$), 95% confidence interval ($CI_{95\%}$) and number of elements (n_{pairs}).

Table 5
Classification table of ecological status classes of lotic habitats indicated by morphological and DNA sequence results. A green background indicates no shift, a yellow background shows a one-class shift and an orange background shows a two-class shift.

		Uncorrected DNA-based ecological status class					Corrected DNA-based ecological status class				
		High	Good	Moderate	Poor	Bad	High	Good	Moderate	Poor	Bad
Morphology-based ecological status class	High	5	11	5	0	0	12	7	2	0	0
	Good	0	9	8	0	0	1	6	10	0	0
	Moderate	0	1	31	2	0	0	1	28	5	0
	Poor	0	0	1	2	0	0	0	0	3	0
	Bad	0	0	0	0	0	0	0	0	0	0

more than 5% of the discrepancies between morphology- and DNA-based community composition in the samples that had a worse ecological status according to DNA sequences. Among these, *A. minutissimum* had the highest IPS sensitivity value and the lowest TI and SI sensitivity values, indicating good quality in contrast to *M. permitis*, which had the lowest IPS and the highest TI and SI sensitivity values, indicating worse status. In soda pans, *N. paleacea*, *N. perminuta*, *N. supralitorea*, *A. minutissimum*, *G. saprophilum* and *N. reskoi* contributed more than 5% of the discrepancies. All of these species are oligohalobic with the exception of *N. reskoi*, which is mesohalobic. The abundance of oligohalobic taxa decreased while the abundance of mesohalobic species increased the H index value.

4. Discussion

As shown in previous studies, metabarcoding offers a promising opportunity to complement microscopic evaluation of lentic and lotic algal communities (Mora et al., 2019; Mortágua et al., 2019). However, there are several steps that should be standardised and validated in the HTS procedure to use it in the ecological status assessment, including the choice of the barcode region, sample processing (DNA extraction, PCR primers, amplification protocol, sequencing technology) and bioinformatics analysis (see Rivera et al. (2018), and references therein). Although efforts have been made by several authors to develop standard methods, a comparison of these factors was not the scope of the present study. For our investigation, we chose the *rbc*L gene due to its high taxonomic resolution power (Apothéloz-Perret-Gentil et al., 2020) and the curated database, Diat.barcode (Rimet et al., 2019), contains more diatom *rbc*L sequences (4,713) that belong to more taxa than 18S rDNA (2,563 sequences of 792 taxa). We applied the barcode region, extraction and amplification suggested by Vasselon et al. (2017) and bioinformatics analysis suggested by Keck et al. (2019) on samples from rivers and streams of different biological types and soda pans. We compared these data to the results of morphological investigation that are currently accepted and used for diatom-based monitoring in Hungary (Szilágyi et al., 2008).

Microscopy and metabarcoding showed significant correlation regarding both the community compositions (indicated by Mantel test) and diatom indices (indicated by Pearson correlation); in addition, the distribution of the samples was similar and the same main driver (TP) was revealed. Other studies have shown similar significant correlations between morphology- and DNA-based diatom indices in samples from various environments, for example, rivers in various conditions in France and Switzerland (the Swiss Diatom Index, DI-CH, Apothéloz-Perret-Gentil et al., 2020), the United Kingdom (the Trophic Diatom Index for Lakes, TDIL; Kelly et al., 2020), Catalan rivers (the IPS; Pérez-Burillo et al., 2020), streams and lakes of Fennoscandia (the IPS; Bailet et al., 2019) and shallow ponds within the Duero river basin (the TDI; Borrego-Ramos et al., 2021). Similarly to the mentioned authors, we found discrepancies between community compositions obtained with different analyses. Several taxa observed microscopically could not be detected with metabarcoding and vice versa, a factor that considerably affected the index calculation.

Overall, more taxa were detected with morphological analysis than with metabarcoding. This finding is consistent with Mortágua et al. (2019) and Bailet et al. (2019) and is in contrast to Kelly et al. (2020). There was about 25% overlap between the number of species detected microscopically and with metabarcoding; however, most of the deviating species had a low abundance. Multiple explanations could be given regarding the difference in detecting taxa by two different methods. Incompleteness of the reference database, incomplete extraction of DNA from cells, the presence of substances inhibiting PCR amplification, non-annealing of primers, observation of empty frustules, inhomogeneous distribution of taxa in the samples and morphologically variable taxa with low sequence variation can lead to biases, resulting in several taxa that could only be found with or in higher abundance based on

microscopy. On the other hand, species detected only or in higher abundance based on DNA sequencing could be explained by the destruction of weak cell walls during cleaning, cryptic or semi-cryptic species, detection of extracellular DNA or resting stages as well as inhomogeneous distribution of taxa in the samples. These possibilities are discussed in the following sections.

4.1. The reference database

Incompleteness of the reference database is likely the main reason for the differences between the taxonomic composition identified based on morphological features and DNA sequences. Several authors have encountered such a phenomena (Apothéloz-Perret-Gentil et al., 2020; Bailet et al., 2019; Borrego-Ramos et al., 2021; Rivera et al., 2018). A large part of the taxa detected by microscopic counting in the riverine and soda pan samples (60% and 32%, respectively) was not recorded in the database: *Aulacoseira pusilla* (Meister) Tuji et Houki, *Luticola acidoclinata* Lange-Bertalot, *Navicula metareichardtiana* Lange-Bertalot & Kusber nom. nov., *Nitzschia archibaldii* Lange-Bertalot, *Nitzschia liebethruthii* Rabenhorst, *Nitzschia palea* (Kützing) W. Smith var. *debilis* (Kützing) Grunow in Cleve & Grunow, *Nitzschia sociabilis* Hustedt, *Nitzschia subtilis* Grunow in Cleve et Grunow and *Sellaphora atomoides* Wetzel & Van de Vijver were dominant in more than one of the running water samples, and *Craticula elkab* (O. Müller ex O. Müller) Lange-Bertalot Kusber & Cocquyt, *Halamphora kevei* Levkov and *Ulnaria grunowii* (Lange-Bertalot et Ulrich) Cantonati et Lange-Bertalot in Kusber et al. were dominant in more than one of the soda pan samples. Other taxa not recorded in the database were rare species that occurred in only a few samples and were dominant in no more than one sample. We expected that a higher proportion of the species found in rivers and streams would be recorded in the database because the main species coverage is for temperate species (Bailet et al., 2019). Presumably, these taxa could be found in the sequences identified at the genus or higher level (Mortágua et al., 2019). In some cases, incompleteness did not mean the lack of species but the lack of a sequence variant. Although clustering sequences into operational taxonomic units (OTUs) before taxonomic assessment can cause loss of genetic information (Keck et al., 2017), searching for exact matches (as the DADA2 pipeline does) can also be a limitation. This was the case with *N. paleacea* in some samples in which this species was not found searching for exact matches with records of the database, but searching sequences identified at the genus level in the NCBI GenBank revealed 99% similarity. In the case of *Tabularia fasciculata* (Agardh)Williams et Round, only two sequence variants were available in the reference database and these proved to be too short for comparison. Primer specificity or preferential amplification of some taxa could also cause differences (Bailet et al., 2019; Kermarrec et al., 2013a).

Another question is the intraspecific variability of the studied region and the coincidence between sequence differences and morphological subspecies/varieties/forms (Mora et al., 2019). For many species, a sequence of only one intraspecific taxon was available in the database, but another intraspecific taxon of the species was observed under the microscope. This factor is significant especially when different variants indicate different ecological status. For example, *N. palea* var. *debilis* indicates better ecological status than other *N. palea* varieties and was dominant in two samples based on morphology, but its sequences were identified as *N. palea*. Moreover, some subspecies and varieties were later raised to the species level based on only morphological analysis – for example, *N. media* and *M. permitis* according to Reichardt (2018) – or were reassigned to other species – for example, *Nitzschia palea* var. *tenuirostris* became *Nitzschia capeitellata* var. *tenuirostris* (Bukhtiyarova, 1995).

Bailet et al. (2019) pointed out that misidentifications in the reference database also leads to discrepancies between morphology- and DNA-based species compositions. Therefore, they emphasised the importance of continuous verification and curation of the reference database. This is particularly important considering the changing

taxonomy of diatoms. In several of our samples, *N. reskoi* was a dominant species that was not found among sequences when using a previous version of Diat.barcode (v7). Instead *Nitzschia frustulum* (Kützing) Grunow was in high abundance based on sequences. This was because the TCC949 strain was identified as *N. frustulum* and it was updated as *N. reskoi* in the later version. Using version 9.2 allowed us to identify these sequences as *N. reskoi*.

4.2. Misidentification of similar or small species and cryptic species

Misidentification can occur in microscopic investigation (Bailet et al., 2019), especially in the case of cryptic species (morphologically similar but different based on molecular methods; Mann and Evans, 2007) or when the distinguishing feature between similar taxa is not unambiguous in its description (Borrego-Ramos et al., 2021) or it can be seen only with an electron microscope. *Nitzschia dissipata* (Kützing) Rabenhorst and its former variant, *N. media* – now raised to the species level – sometimes cannot be easily separated (it depends on the closeness of raphe keel to valve edge, valve linearity; (Reichardt, 2018). This situation presumably occurred when *N. dissipata* was found based on microscopic analysis and *N. media* was identified according to sequence analysis. *P. frequentissimum* was often identified in our samples morphologically while *P. victori* was detected with metabarcoding. *P. victori* resembles *P. frequentissimum*, and the main difference between them is their *rbc*L sequence (Jahn et al., 2017; Novis et al., 2012).

4.3. Environmental effect on morphology and phenotypic plasticity

Environmental variables can affect valve ontogeny resulting in morphological changes (Cox, 2014). This can make microscopic identification more difficult but does not affect the DNA sequence.

Extreme environmental conditions like low current velocity, drought, high temperature, ultraviolet radiation, mechanical damage by grazers, crowding, high salinity, a high level of nutrients, heavy metal contamination, toxic compounds or culturing can modify valve morphology in various ways, causing abnormal cells that can be used as indicators of environmental stress (Falasco et al., 2009, and references therein). In some of our samples, teratological forms of some species (*A. minutissimum, Gomphonema micropus* Kützing, *Navicula tripunctata* (O.F. Müller) Bory and *Nitzschia inconspicua* Grunow) were observed under the microscope in low proportions. In all of these samples, the normal form of the species was also found based on both aspects. The causes of teratological development of valves are still not fully understood. Previously, researchers had considered it was not induced by genetic changes because abnormal cells restored their normal morphology after sexual reproduction (Falasco et al., 2009). However, this theory has been challenged by Coquillé and Morin (2019), whose experiments on *Gomphonema gracile* Ehrenberg suggest the heritability of the teratological character. Therefore, it is still unknown which DNA regions are affected by teratology and, therefore, it is unknown whether it could be detected with metabarcoding.

4.4. Cell biovolume, life form and cell wall resistance

Researchers have shown that the copy number of the chloroplast-encoded *rbc*L gene is related to the cell biovolume; therefore, a correction factor has been developed to avoid this problem (Vasselon et al., 2018). When we applied this correction, we mitigated differences between abundances of species obtained with the two methods, particularly in the case of the small-celled *A. minutissimum*, whose abundance is often underestimated based on *rbc*L read numbers. These findings are consistent with the results reported by Bailet et al. (2019); Mortágua et al. (2019); and Borrego-Ramos et al. (2021) and confirm the benefit of applying the correction factor.

Several diatom species can form different kinds of colonies (Rimet and Bouchez, 2012) and this life form could affect the distribution of cells of a taxon in samples. For example, *Melosira varians* Agardh forms chains; thus, not only cells but chains can get into the subsample dedicated to microscopy or another set aside for DNA analysis, leading to larger differences in relative abundances.

Detectability of species also depends on the thickness and resistance of their cell wall. The differences between thin-walled species like *Fistulifera saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot and *M. permitis* are hard to observe under a light microscope; moreover, they can be destroyed during oxidising cleaning of frustules. By contrast, its wall enhances the extraction of its DNA, a fact that resulted in higher abundance based on sequences compared with morphology in our investigation as well as the studies by Kelly et al. (2020); Pérez-Burillo et al. (2020); and Borrego-Ramos et al. (2021).

4.5. Biases related to the properties of the samples

Organisms in a biofilm live in a matrix of extracellular polymeric substances (EPS) containing proteins, exopolysaccharides, metal ions, nucleic acids, lipids, and humic substances (Kumar et al., 2020). This matrix can retain extracellular DNA that can be detected only with metabarcoding and empty frustules that can be observed under a microscope (Borrego-Ramos et al., 2021; Mora et al., 2019; Rivera et al., 2018).

In some of our samples, species that were recorded in the reference database and were found dominant under the microscope were not detected with metabarcoding. This was the case with *A. minutissimum*. It was found in 76 lotic and 8 soda pan samples with microscope and/or metabarcoding and in 63 and 6 samples, respectively, with both methods. In most cases when only one of the methods detected it, *A. minutissimum* had a low abundance (<1%). However, in three samples it was dominant (10.2%–41.7%) based on morphological analysis but not detected based on DNA sequences. Moreover, other *Achnanthidium* sequences were not observed in those samples. Because *A. minutissimum* has a high sensitivity value (5), its absence could cause shifts in the DNA-based IPS, indicating a status worse than the morphology-based index. We hypothesise that the valves seen under the microscope belonged to dead or dying cells with DNA that was too degraded and PCR primers could not anneal to it. Mora et al. (2019) and Rivera et al. (2018) discussed dead cells as a possible source of deviation between morphological and metabarcoding results. They considered dead cells to be transported from a location other than the target assemblage (Mora et al., 2019) or settle from plankton (Rivera et al., 2018). Considering that *A. minutissimum* is a prostrate, low-profile diatom (Rimet and Bouchez, 2012) and a pioneer species during colonisation (Johnson et al., 1997), it cannot be ruled out that its cells gradually died as other species immigrated but its valves remained in the biofilm matrix. Another explanation could be that primers perhaps could not anneal because of mutations in binding sites.

Diatoms can form resting stages to overcome periods adverse to growth (Kooistra et al., 2007), from which DNA could be extracted (Piredda et al., 2017), but species could not be identified in this form under the light microscope. In the samples there can be substances such as polysaccharides, humic acids that can inhibit PCR amplification but were not removed during DNA purification (Semenov, 2021).

5. Conclusions

In the present study, we evaluated the performance of metabarcoding in ecological status assessment of Hungarian waters and compared it to the validated morphology-based method. We found significant correlations between the two methods regarding community composition, diatom indices and the relationship with environmental variables. However, there were considerable discrepancies in species composition that can only be partly mitigated using correction according to cell biovolume. Discrepancies in species compositions led to contradictory ecological status assessment: a considerable part of the running water

samples DNA-based indices indicated moderate status while morphology-based indices indicated good status. This is problematic due to the fact that the Water Framework Directive prescribes intervention for waters in moderate or worse status (European Commission, 2000). Thus, in its present form, metabarcoding can not be recommended to be applied in ecological status assessment for identification of species in place of the morphology-based method.

There are several possible explanations to the differences and considering these factors the metabarcoding process can be improved. One of the major problems is the incompleteness of the reference database which could be improved with extensive sampling campaigns with special emphasis of peculiar habitats inhabited by diatoms with special requirements. Soda pans are among such environments where loss of the soda character is the main threat; thus, accurate identification of species and information on their halobity tolerance is particularly important for correct ecological status assessment. Therefore, complementing the reference database to allow the identification of such specific taxa is necessary.

There are techniques to acquire barcode sequence of a given taxon: clonal culturing, single cell sequencing, direct Sanger sequencing as well as high throughput sequencing of environmental samples with low species diversity (Rimet et al., 2018). Efforts should be made to improve the efficiency of these techniques or develop new ones.

Improvement of metabarcoding should involve taxonomic investigations. Taxonomic assignment of records in the database is also morphology-based. Misidentification of database records cannot be completely ruled out, but it can be minimized if the taxonomic assignment is carried out by taxonomists using both light and electron microscope. Decision which taxa should be valid is also requires taxonomical expertise. The situation could be eased if the description of new taxa would be based on both morphological and molecular investigations. Species complexes possibly containing cryptic species also need taxonomic revisions based on DNA sequences as it was initiated for e.g. *Nitzschia palea* (Trobajo et al., 2010), *Sellaphora pupula* (Evans et al., 2007), *Gomphonema parvulum* (Kermarrec et al., 2013a).

We propose using both microscopy and metabarcoding in parallel until the reference database has been suitably updated. Moreover, considering the effects of environment on valve morphogenesis a DNA extraction method when valves could be recovered and investigated under microscope (a method for communities similar to the single-cell method by Lang and Kaczmarska (2011) in parallel with DNA analysis would be beneficial.

In this study, we presented the use of metabarcoding for identification of species. However, there are taxonomy-free approaches when DNA sequences are used without taxonomic assignment in ecological status assessment (Tapolczai et al., 2021; Tapolczai et al., 2019). It does not depend on and thus is not biased by morphological analysis, moreover, all sequences can be used not only those that could be assigned at species level. However, it requires regular and reliable data on environmental variables to draw reliable consequences on their effects on molecular taxonomic units. Development of an all-new database and method may take a longer time than relying on already available knowledge and practice.

CRediT authorship contribution statement

Mónika Duleba: Investigation, Formal analysis, Writing – original draft. **Angéla Földi:** Investigation. **Adrienn Micsinai:** Methodology, Conceptualization. **Gábor Várbíró:** Formal analysis, Writing – original draft. **Anita Mohr:** Methodology, Investigation. **Rita Sipos:** Methodology, Investigation. **Gyula Szabó:** Formal analysis. **Krisztina Buczkó:** Investigation. **Zsuzsa Trábert:** Investigation. **Keve T. Kiss:** Investigation. **Tibor Bíró:** Writing - review & editing. **Edit Vadkerti:** Writing - review & editing. **Éva Ács:** Conceptualization, Funding acquisition, Supervision, 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.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.108105.

References

Ács, É., Borics, G., Boda, P., Csányi, B., Duleba, M., Engloner, A., Erős, T., Földi, A., Grigorszky, I., György, Á.I., Kiss, K.T., K. Szilágyi, E., Lukács, B.A., Nagy-László, Z., Pozderka, V., Sály, P., Szalóky, Z., Szekeres, J., Trábert, Z., Várbíró, G., 2015a. Magyarország felszíni vizeinek ökológiai állapotértékelő módszerei. Magy. Kémikusok Lapja 70, 374–380.

Apothéloz-Perret-Gentil, L., Bouchez, A., Cordier, T., Cordonier, A., Guéguen, J., Rimet, F., Vasselon, V., Pawlowski, J., 2020. Monitoring the ecological status of rivers with diatom eDNA metabarcoding: a comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. Mol. Ecol. 00, 1–10. 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. https://doi.org/10.3897/mbmg.3.34002.

Berthon, V., Bouchez, A., Rimet, F., 2011. Using diatom life-forms and ecological guilds to assess organic pollution and trophic level in rivers: a case study of rivers in south-eastern France. Hydrobiologia 673, 259–271. https://doi.org/10.1007/s10750-011-0786-1.

Boros, E., V.-Balogh, K., Vörös, L., Horváth, Z., 2017. Multiple extreme environmental conditions of intermittent soda pans in the Carpathian Basin (Central Europe) | Elsevier Enhanced Reader. Limnologica 62, 38–46.

Borrego-Ramos, M., Bécares, E., García, P., Nistal, A., Blanco, S., 2021. Epiphytic diatom-based biomonitoring in mediterranean ponds: traditional microscopy versus metabarcoding approaches. Water 13, 1351. https://doi.org/10.3390/w13101351.

Bukhtiyarova, L.N., 1995. New taxonomic combinations of Diatoms (Bacillariophyta). Algologia 5, 417–424.

Burliga, A.L., Kociolek, J.P., 2016. Diatoms (Bacillariophyta) in Rivers, in: Necchi JR, O. (Ed.), River Algae. Springer International Publishing, Cham, pp. 93–128. 10.1007/978-3-319-31984-1_5.

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.

Carayon, D., Tison-Rosebery, J., Delmas, F., 2019. Defining a new autoecological trait matrix for French stream benthic diatoms. Ecol. Indic. 103, 650–658. https://doi.org/10.1016/j.ecolind.2019.03.055.

Cemagref, 1982. Etude des méthodes biologiques quantitativesd'appréciation de la qualité des eaux, Rapport Division Qualité des Eaux Lyon. Agence financière de Bassin Rhône-Méditerranée-Corse, Pierre-Bénite, France.

Chen, H., 2018. VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20 [WWW Document]. URL https://rdrr.io/cran/VennDiagram/ (accessed 8.3.21).

Coquillé, N., Morin, S., 2019. Fitness of teratological morphotypes and heritability of deformities in the diatom Gomphonema gracile. Ecol. Indic. 106, 105442 https://doi.org/10.1016/j.ecolind.2019.105442.

Coste, M., Prygiel, J., 1998. Mise au point de l'Indice Biologique Diatomée, un indice diatomique pratique applicable au réseau hydrographique français. Eau Ind. Nuis. 40–45.

Cox, E.J., 2014. Diatom identification in the face of changing species concepts and evidence of phenotypic plasticity. J. Micropalaeontol. 33, 111–120. https://doi.org/10.1144/jmpaleo2014-014.

Ector, L., J.C., K., Charles, D., Denys, L., M.S.V., D., Manoylov, K., Michelutti, N., Rimet, F., Smol, J., Stevenson, R., J.G., W., 2004. Workshop report. Freshwater diatoms and their role as ecological indicators. pp. 469–480.

Elosegi, A., Sabater, S., 2013. Effects of hydromorphological impacts on river ecosystem functioning: a review and suggestions for assessing ecological impacts. Hydrobiologia 712, 129–143. https://doi.org/10.1007/s10750-012-1226-6.

European Commission, 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.

European Committee for Standardisation, 2002. EN 13946 Water Quality. Guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers.

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

Falasco, E., Bona, F., Badino, G., Hoffmann, L., Ector, L., 2009. Diatom teratological forms and environmental alterations: a review. Hydrobiologia 623, 1–35. https://doi.org/10.1007/s10750-008-9687-3.

Földi, A., Ács, É., Grigorszky, I., Ector, L., Wetzel, C.E., Várbíró, G., Kiss, K.T., Dobosy, P., Trábert, Z., Borsodi, A.K., Duleba, M., 2018. Unexpected consequences of bombing. Community level response of epiphytic diatoms to environmental stress in a saline bomb crater pond area. PLoS ONE 13, e0205343. https://doi.org/10.1371/journal.pone.0205343.

Hambrook Berkman, J.A., Porter, S.D., 2004. An overview of algal monitoring and research in the U.S. Geological Survey's National Water Quality Assessment (NAWQA) Program. Diatom 20, 13–22.

Hammer, O., Harper, D., Ryan, P., 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontol. Electron. 4, 1–9.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B Biol. Sci. 270, 313–321. https://doi.org/10.1098/rspb.2002.2218.

Horváth, Z., Vad, C.F., Vörös, L., Boros, E., 2013. Distribution and conservation status of fairy shrimps (Crustacea: Anostraca) in the astatic soda pans of the Carpathian basin: the role of local and spatial factors. J. Limnol. 72, 9. https://doi.org/10.4081/jlimnol.2013.e9.

Hungarian Standards Institution, 2014. MSZ EN 13946:2014. Water quality. Guidance for the routine sampling and preparation of benthic diatoms from rivers and lakes. Hungarian Standards Institution, Budapest.

Hungarian Standards Institution, 2009. MSZ EN ISO 11885:2009. Water quality. Determination of selected elements by inductively coupled plasma optical emission spectrometry (ICP-OES) (ISO 11885:2007). Hungarian Standards Institution, Budapest.

Hungarian Standards Institution, 2004. MSZ EN 12260:2004. Water quality. Determination of nitrogen. Determination of bound nitrogen (TNb), following oxidation to nitrogen oxides. Hungarian Standards Institution, Budapest.

Hungarian Standards Institution, 1993. MSZ ISO 10260:1993. Water quality. Measurement of biochemical parameters. Spectrometric determination of the chlorophyll-a concentration. Hungarian Standards Institution, Budapest.

Hungarian Standards Institution, 1992. MSZ ISO 6060:1991. Water quality. Determination of the chemical oxygen demand. Hungarian Standards Institution, Budapest.

International Organization for Standardization, 2002. ISO 15705:2002. Water quality. Determination of the chemical oxygen demand index (ST-COD) - Small-scale sealed-tube method. International Organization for Standardization, Geneva.

Jahn, R., Abarca, N., Gemeinholzer, B., Mora, D., Skibbe, O., Kulikovskiy, M., Gusev, E., Kusber, W.-H., Zimmermann, J., 2017. Planothidium lanceolatum and Planothidium frequentissimum reinvestigated with molecular methods and morphology: four new species and the taxonomic importance of the sinus and cavum. Diatom Res. 32, 75–107. https://doi.org/10.1080/0269249X.2017.1312548.

Johnson, R., Wiederholm, T., Rosenberg, D.M., 1993. Freshwater biomonitoring using individual organisms, populations and species assemblages of benthic macroinvertebrates. Freshw. Biomonitor. Benthic Macroinvertebrates 40–158.

Johnson, R.E., Tuchman, N.C., Peterson, C.G., 1997. Changes in the Vertical Microdistribution of Diatoms within a Developing Periphyton Mat. J. North Am. Benthol. Soc. 16, 503–519. https://doi.org/10.2307/1468140.

Keck, F., Rimet, F., Vasselon, V., Bouchez, A., 2019. A ready-to-use database for DADA2: Diat.barcode_rbcL_312bp_DADA2.

Keck, F., Vasselon, V., Tapolczai, K., Rimet, F., Bouchez, A., 2017. Freshwater biomonitoring in the Information Age. Front. Ecol. Environ. 15, 266–274. https://doi.org/10.1002/fee.1490.

Kelly, M., 2019. Adapting the (fast-moving) world of molecular ecology to the (slow-moving) world of environmental regulation: lessons from the UK diatom metabarcoding exercise. Metabarcoding Metagenomics 3, e39041. https://doi.org/10.3897/mbmg.3.39041.

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.

Kermarrec, L., Bouchez, A., Rimet, F., Humbert, J.-F., 2013a. First Evidence of the Existence of Semi-Cryptic Species and of a Phylogeographic Structure in the Gomphonema parvulum (Kützing) Kützing Complex (Bacillariophyta). Protist 164, 686–705. https://doi.org/10.1016/j.protis.2013.07.005.

Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J.-M., Humbert, J.-F., Bouchez, A., 2014. A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshw. Sci. 33, 349–363. https://doi.org/10.1086/675079.

Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Humbert, J.F., Bouchez, A., 2013b. 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.

Kooistra, W.H.C.F., Gersonde, R., Medlin, L.K., Mann, D.G., 2007. CHAPTER 11 - The Origin and Evolution of the Diatoms: Their Adaptation to a Planktonic Existence. In: Falkowski, P.G., Knoll, A.H. (Eds.), Evolution of Primary Producers in the Sea. Academic Press, Burlington, pp. 207–249. https://doi.org/10.1016/B978-012370518-1/50012-6.

Kumar, J., Sharma, V.K., Parmar, S., Singh, P., Singh, R.K., 2020. Chapter 11 - Biofilm: A microbial assemblage on the surface—A boon or bane? In: Yadav, M.K., Singh, B.P. (Eds.), New and Future Developments in Microbial Biotechnology and

Bioengineering: Microbial Biofilms. Elsevier, Amsterdam, pp. 139–150. https://doi.org/10.1016/B978-0-444-64279-0.00011-6.

Lang, I., Kaczmarska, I., 2011. A protocol for a single-cell PCR of diatoms from fixed samples: method validation using Ditylum brightwellii (T. West) Grunow. Diatom Res. 26, 43–49. https://doi.org/10.1080/0269249X.2011.573703.

Lecointe, C., Coste, M., Prygiel, J., 1993. "Omnidia": software for taxonomy, calculation of diatom indices and inventories management. Hydrobiologia 269, 509–513. https://doi.org/10.1007/BF00028048.

Lenoir, A., Coste, M., 1996. Development of a practical diatom index of overall water quality applicable to the French National Water Board Network. Presented at the International symposium, Volksbildungsheim Grilhof Vill. Universität Innsbruck, pp. 29–43.

Lobo, E.A., Heinrich, C.G., Schuch, M., Wetzel, C.E., Ector, L., 2016. Diatoms as Bioindicators in Rivers, in: Necchi JR, O. (Ed.), River Algae. Springer International Publishing, Cham, pp. 245–271. 10.1007/978-3-319-31984-1_11.

Makovinská, J., Hlúbiková, D., Fidlerová, D., 2014. Joint Danube Survey 3—Phytobenthos. International Commission for the Protection of the Danube River, Vienna, Austria.

Mann, D., Evans, K., 2007. 13 Molecular genetics and the neglected art of diatomics, in: Unravelling the Algae: The Past, Present, and Future of Algal Systematics, Systematics Association Special Volume Series. CRC Press, Boca Raton, pp. 231–265.

Mann, D.G., Vanormelingen, P., 2013. An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species. J. Eukaryot. Microbiol. 60, 414–420. https://doi.org/10.1111/jeu.12047.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. 17, 10–12. https://doi.org/10.14806/ej.17.1.200.

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. 38, 448–464. https://doi.org/10.1086/704827.

Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., João Feio, M., F.P. Almeida, S., 2019. Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. Ecol. Indic. 106, 105470. 10.1016/j.ecolind.2019.105470.

Nikanorov, A.M., Stradomskaya, A.G., 2007. Chronic pollution of freshwater bodies: data on accumulation of pesticides, oil products, and other toxic substances in bottom deposits. Water Resour. 34, 314–320. https://doi.org/10.1134/S0097807807030098.

Novis, P.M., Braidwood, J., Kilroy, C., 2012. Small diatoms (Bacillariophyta) in cultures from the Styx River, New Zealand, including descriptions of three new species. Phytotaxa 64, 11–45. https://doi.org/10.11646/phytotaxa.64.1.3.

Oksanen, J.F., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Wagner, H., 2018. Community Ecology Package. R package version 2.5-3.

Patil, I., 2021. Visualizations with statistical details: The "ggstatsplot" approach, 10.31234/osf.io/p7mku.

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.

Piredda, R., Sarno, D., Lange, C.B., Tomasino, M.P., Zingone, A., Montresor, M., 2017. Diatom Resting Stages in Surface Sediments: A Pilot Study Comparing Next Generation Sequencing and Serial Dilution Cultures. Cryptogam. Algol. 38, 31–46. https://doi.org/10.7872/crya/v38.iss1.2017.31.

Prygiel, J., 2000. Guide méthodologique pour la mise en oeuvre de l'Indice Biologique Diatomées NF T 90–354. Agence de l'Eau Artois-Picardie, Amiens.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reichardt, E., 2018. Die Diatomeen im Gebiet der Stadt Treuchtlingen. Selbstverlag der Bayerischen Botanischen Gesellschaft, München, Germany.

Richardson, D.M., 2011. Fifty Years of Invasion Ecology: The Legacy of Charles Elton. John Wiley & Sons, Chichester.

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

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

Rimet, F., Bouchez, A., 2012. Life-forms, cell-sizes and ecological guilds of diatoms in European rivers. Knowl. Manag. Aquat. Ecosyst. 1283–1299 https://doi.org/10.1051/kmae/2012018.

Rimet, F., Chaumeil, P., Keck, F., Kermarrec, L., Vasselon, V., Kahlert, M., Franc, A., Bouchez, A., 2016. R-Syst::diatom: an open-access and curated barcode database for diatoms and freshwater monitoring. Database 2016. https://doi.org/10.1093/database/baw016.

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. 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.

Rott, E., Hofmann, G., Pall, K., Pfister, P., Pipp, E., 1997. Indikatorlisten für aufwuchsalgen in Österreichischen fliessgewässern. Teil. 1: Saprobielle Indikation.

Bundesministerium für Land- und Forstwirschaft, Wasserwirtschaftskataster, Wien, Austria.

Rott, E., Pipp, E., Pfister, P., Van Dam, H., Orther, K., Binder, N., Pall, K., 1999. Indikatorlisten für aufwuchsalgenin Österreichischen fliessgewässern. Teil 2: Trophie indikation. Bundesministerium für Land- und Forstwirschaft, Wasserwirtschaftskataster, Wien, Austria.

Round, F.E., Crawford, R.M., Mann, D.G., 1990. Diatoms: Biology and Morphology of the Genera, 1st ed. Cambridge University Press, Cambridge, UK.

Semenov, M.V., 2021. Metabarcoding and metagenomics in soil ecology research: achievements, challenges, and prospects. Biol. Bull. Rev. 11, 40–53. https://doi.org/10.1134/S2079086421010084.

Søndergaard, M., Jeppesen, E., 2007. Anthropogenic impacts on lake and stream ecosystems, and approaches to restoration. J. Appl. Ecol. 44, 1089–1094. https://doi.org/10.1111/j.1365-2664.2007.01426.x.

Stenger-Kovács, C., Lengyel, E., Buczkó, K., Tóth, F., Crossetti, L., Pellinger, A., Zámbóné Doma, Z., Padisák, J., 2014. Vanishing world: alkaline, saline lakes in Central Europe and their diatom assemblages. Inland Waters 4, 383–396. https://doi.org/10.5268/IW-4.4.722.

Szczepocka, E., Szulc, B., 2009. The use of benthic diatoms in estimating water quality of variously polluted rivers. Oceanol. Hydrobiol. Stud. 38, 17–26. https://doi.org/10.2478/v10009-009-0012-x.

Szilágyi, F., Ács, É., Borics, G., Halasi-Kovács, B., Juhász, P., Kiss, B., Kovács, T., Müller, Z., Lakatos, G., Padisák, J., Pomogyi, P., Stenger-Kovács, C., Szabó, K.É., Szalma, E., Tóthmérész, B., 2008. Application of water framework directive in Hungary: development of biological classification systems. Water Sci. Technol. 58, 2117–2125. https://doi.org/10.2166/wst.2008.565.

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

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

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. 10.1016/j.ecolind.2020.107322.

Tóth, A., Horváth, Z., Vad, C.F., Zsuga, K., Nagy, S.A., Boros, E., 2014. Zooplankton of the European soda pans: Fauna and conservation of a unique habitat type. Int. Rev. Hydrobiol. 99, 255–276. https://doi.org/10.1002/iroh.201301646.

Trobajo, R., Mann, D.G., Clavero, E., Evans, K.M., Vanormelingen, P., McGregor, R.C., 2010. The use of partial cox1, rbcL and LSU rDNA sequences for phylogenetics and species identification within the Nitzschia palea species complex (Bacillariophyceae). Eur. J. Phycol. 45, 413–425. https://doi.org/10.1080/09670262.2010.498586.

Vad, C.F., Péntek, A.L., Cozma, N.J., Földi, A., Tóth, A., Tóth, B., Böde, N.A., Móra, A., Ptacnik, R., Ács, É., Zsuga, K., Horváth, Z., 2017. Wartime scars or reservoirs of biodiversity? The value of bomb crater ponds in aquatic conservation | Elsevier Enhanced Reader. Biol. Conserv. 209, 253–262.

Van Dam, H., Mertens, A., Sinkeldam, J., 1994. A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. Netherland J. Aquat. Ecol. 28, 117–133. https://doi.org/10.1007/BF02334251.

Várbíró, G., Borics, G., Csányi, B., Fehér, G., Grigorszky, I., Kiss, K.T., Tóth, A., Ács, E., 2012. Improvement of the ecological water qualification system of rivers based on the first results of the Hungarian phytobenthos surveillance monitoring. Hydrobiologia 695, 125–135. https://doi.org/10.1007/s10750-012-1120-2.

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. 9, 1060–1069. https://doi.org/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.

Vautier, M., Vasselon, V., Chardon, C., Rimet, F., Bouchez, A., Domaizon, I., 2020. DNA extraction from environmental biofilm using the NucleoSpin® Soil kit (MACHEREY-NAGEL).

Visco, J.A., Apothéloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillet, L., Pawlowski, J., 2015. Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. Environ. Sci. Technol. 49, 7597–7605. https://doi.org/10.1021/es506158m.

Wen, Y., Schoups, G., van de Giesen, N., 2017. Organic pollution of rivers: Combined threats of urbanization, livestock farming and global climate change. Sci. Rep. 7, 43289. https://doi.org/10.1038/srep43289.

Ziemann, H., Kies, L., Schulz, C.-J., 2001. Desalinization of running waters: III. Changes in the structure of diatom assemblages caused by a decreasing salt load and changing ion spectra in the river Wipper (Thuringia, Germany). Limnologica 31, 257–280. https://doi.org/10.1016/S0075-9511(01)80029-3.

Zimmermann, J., Abarca, N., Enke, N., Skibbe, O., Kusber, W.-H., Jahn, R., 2014. Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. PLoS ONE 9, e108793. https://doi.org/10.1371/journal.pone.0108793.

Zimmermann, J., Glöckner, G., Jahn, R., Enke, N., Gemeinholzer, B., 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Mol. Ecol. Resour. 15, 526–542. https://doi.org/10.1111/1755-0998.12336.

Zimmermann, J., Mora, D., Tapolczai, K., Proft, S., Chonova, T., Rimet, F., Bouchez, A., Fidlerová, D., Makovinská, J., Weigand, A., 2020. Metabarcoding of phytobenthos samples. International Commission for the Protection of the Danube River, Vienna, Austria.