



The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding

Kálmán Tapolczai^{a,b,*}, Géza B. Selmeczy^b, Beáta Szabó^c, Viktória B-Béres^{d,e}, François Keck^f, Agnès Bouchez^f, Frédéric Rimet^f, Judit Padisák^b

^a Premium Postdoctoral Research Program, Hungarian Academy of Sciences, Budapest, Hungary

^b Research Group of Limnology, Center for Natural Science, University of Pannonia, Egyetem u. 10, H-8200 Veszprém, Hungary

^c Centre for Ecological Research, Balaton Limnological Institute, Klebelsberg Kuno u. 3, H-8237 Tihany, Hungary

^d Centre for Ecological Research, Danube Research Institute, Department of Tisza Research, 18/c. Bem square, H-4026 Debrecen, Hungary

^e MTA Centre for Ecological Research, GINOP Sustainable Ecosystems Group, 3. Klebelsberg Kuno str., H-8237 Tihany, Hungary

^f CARTELE USMB INRAE, 75 bis Avenue de Corzent 74200 Thonon-les-Bains France

ARTICLE INFO

Keywords:

Diatom
Environmental DNA
ESV
Land-use
Metabarcoding
Stream

ABSTRACT

Land-use imposes an important potential threat on the aquatic ecosystems of riverine habitats. In this study, DNA metabarcoding was used to assess the effect of land-use on diatom assemblages, with a special focus on cropland area as an integrative proxy for several direct-acting pressures. The so-called taxonomy-free approach was tested using exact sequence variants (ESVs) without pre-clustering and without assigning them to traditional taxonomy. Our hypothesis is that the taxonomy-free approach gives better interpretation of the effect of agricultural land-use and it provides a more efficient index to indicate agricultural stress than the classical method with taxonomical clustering. Classical ordination techniques (PCA, NMDS) were performed to study the effects of cropland area proportion on diatom assemblages and a modified version of the Zelinka-Marvan equation for the index development. Results showed that (i) although ESVs provided better results when studying land-use effects on diatoms, taxonomic assignment after analysis was necessary to give ecological interpretations and that (ii) a better performing index could be developed by using the taxonomy-free approach. By using ESVs without taxonomic assignment, information on the ecology of sequences belonging to the same species and of unassigned sequences could be kept. New types of clustering methods are welcome in the future of biomonitoring where the delimitation of taxonomic units should be refined based on a higher emphasis on their ecology rather than on morphological or genetical criteria.

1. Introduction

Land-use contributes substantially to the threats altering riverine ecosystems worldwide (Vörösmarty et al., 2010). As intensive agricultural activity on the catchments has been permanently expanding, it has become a major source of quality degradation of watercourses (Moss, 2008). It, among others, affects aquatic ecosystems through several direct-acting factors, e.g. nitrogen fertilizers (He et al., 2011), pesticides (DeLorenzo et al., 2001), river morphology alteration (Pedersen, 2009), salinization (Blinn and Bailey, 2001), reduced riparian vegetation (Broetto et al., 2017), and sedimentation (Naden et al., 2016). Thus, the assessment of such an integrative environmental pressure as agricultural

land-use is of great importance.

Diatoms are well studied indicator organisms and the effect of agricultural land-use on both lotic and lentic communities is reported (Blinn and Bailey, 2001; Bradshaw et al., 2006; Miettinen et al., 2005; Pan et al., 2004; Stenger-Kovács et al., 2020) or indicated by traditional diatom indices (Walsh and Wepener, 2009). Diatom indices are routinely used for biomonitoring purposes and are implemented in water management legislation frameworks (European Commission, 2000; U.S. Congress, 1972). These indices are traditionally computed based on the (in some cases poorly known) autecological properties of the taxa identified via microscopic examination. This method however encompasses some uncertainties that have been discussed in a number of

* Corresponding author.

E-mail address: tapolczai.kalman@almos.uni-pannon.hu (K. Tapolczai).

<https://doi.org/10.1016/j.ecolind.2020.107322>

Received 13 July 2020; Received in revised form 21 December 2020; Accepted 27 December 2020

Available online 5 January 2021

1470-160X/© 2020 The Author(s).

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

studies (e.g. Kelly et al., 2009; Schneider et al., 2017; Tapolczai et al., 2016). Microscopical identification of diatoms is time consuming, requires high-level expertise and up to date identification literature to cope with the fast-changing and relatively unstable taxonomy. The resulting taxa lists thus often suffer a subjective bias that makes the intercomparison of studies rather challenging (Kahlert et al., 2009).

Recent developments in DNA metabarcoding and high-throughput sequencing (HTS) offers a cost-effective way requiring less human effort for the simultaneous identification of a large quantity of DNA barcodes in hundreds of samples. Thus, taxa lists are obtained in a rather automatized and potentially standardised way. The advantage of the method for biomonitoring is evident and many papers, largely focusing on indicator organisms such as macroinvertebrates or diatoms, have addressed the implementation of metabarcoding in quality assessment (Bailet et al., 2019; Blackman et al., 2019; Cordier et al., 2020; Elbrecht et al., 2017; Keck et al., 2017; Serrana et al., 2019; Vasselton et al., 2017b). However, implementing this method as a standardized biomonitoring tool has its own challenges (Leese et al., 2016; Rivera et al., 2020) regarding the selection of the proper DNA extraction method (Vasselton et al., 2017a) and molecular barcode (Guo et al., 2015; Kermarrec et al., 2014), that needs to be supported by a curated and large reference database (Rimet et al., 2019). The different HTS techniques introduce further potential biases.

Most bioinformatic pipelines (Caporaso et al., 2010; Chen et al., 2013; Mahé et al., 2015; Schloss et al., 2009; Westcott and Schloss, 2015) cluster DNA sequences into operational taxonomic units (OTUs) based on sequence similarity in order to handle erroneous sequences or to reach a hypothetical species-level delimitation. Recently, it is becoming more adequate to use non-clustered taxonomic units, like amplicon sequence variants (ASVs) or exact sequence variants (ESVs) that went through a thorough denoising algorithm (Callahan et al., 2017; Nearing et al., 2018). Traditionally, these molecular taxonomic units are then assigned to taxonomy based on reference databases, which are however incomplete to varying degrees (Elbrecht et al., 2017; McGee et al., 2019; Rimet et al., 2018; Weigand et al., 2019) and are not completely independent from the traditional approach because the species in the database are also partly identified based on morphology. The so-called taxonomy-free method proposes an alternative approach so that molecular taxonomic units (OTU, ESV, ISU, etc.) are not assigned to taxa and thus, information carried by “unassigned” sequences are not lost. The approach has emerged only in the last years to study periphytic communities in streams but it already presented interesting insights into intraspecific variability and potential use in biomonitoring and bioassessment (Apothéloz-Perret-Gentil et al., 2017; Feio et al., 2020; Tapolczai et al., 2019). It permits the implementation of sequences into quality indices without taxonomic assignment and additionally makes it possible to study subspecies diversity or to enlighten hidden ecological diversity of the different genetic varieties within taxa. The further investigation of the method thus is very promising for both fundamental and applied scientific purposes in stream ecology.

Our study is the first investigating the influence of agricultural land-use (cropland area proportion) on the diatom assemblages through diatom ESVs, revealed by metabarcoding. We furthermore used the taxonomy-free approach for the first time to build a *de novo* model for assessing agricultural stress, based on the diatom assemblages. Hence, the study addresses the hypothesis that ESVs without taxonomic assignment are more efficient to study the effect of agricultural land-use on the diatom assemblages than ESVs clustered to taxonomy. We further hypothesise that the taxonomy-free approach performs better when developing an index predicting agricultural stress, by revealing hidden ecological differences of the ESVs belonging to the same taxa and by keeping the ecological information of unassigned sequences.

2. Materials and methods

2.1. Sampling network and study sites

In total, 130 samples from 67 sampling sites in watercourses were taken during two field sampling campaigns covering two catchment area subunits defined by the river basin management planning of Hungary (RBMP, 2015), in the Central and Western Transdanubian region in Hungary (Fig. 1). These are the catchments of the river Marcal with an area of 3,084 km² and river Rába covering an area of 5,600 km², including the catchment area of the Marcal as it is a part of the latter. The field sampling was carried out twice. The first period was in August and September 2018 (67 samples) and the second was in May and June 2019 (63 samples). The sampling periods were chosen so that they were preceded by a three- four week long period without strong disturbance events, e.g. heavy rains or floods (CEN, 2003). Special attention was taken so that the sampling network includes variability in terms of stream order and typology.

2.2. Environmental variables

Samples were assigned to typology and stream order based on the Hungarian RBMP classification (RBMP, 2015). Coordinates and altitude were noted on site together with the following parameters: width and composition (proportion of grass, shrubs and trees) of riparian vegetation, dominant substrate, coverage of the water surface by macrophytes, proportion of shaded surface area, morphology of the river. Composition and width of the riparian vegetation were used to derive a “riparian vegetation index” (Ripveg_index) so that higher values indicate wide vegetation dominated by trees and narrow or grass dominated riparian vegetation is indicated by low values. These parameters were estimated by naked eyes in the surrounding area of the sampling site. Additional site parameters were measured in situ: river width, water flow velocity, temperature, conductivity, dissolved oxygen concentration, oxygen saturation, and pH, using a Valeport Model 801 electromagnetic open channel flow meter and HQ40d Hach Lange multimeter and probes. At each sampling site, a water sample was taken for further laboratory analyses, during which nitrite (NO₂⁻), nitrate (NO₃⁻), ammonium (NH₄⁺), soluble reactive phosphorus (SRP), total phosphorus (TP), chloride (Cl⁻), soluble reactive silica (SRSi), carbonate (CO₃²⁻), bicarbonate (HCO₃⁻) and chemical oxygen demand (COD) were measured using titrimetric and spectrophotometric methods, following international standards (APHA, 2012; Wetzel and Likens, 2000). Nitrate concentration was measured using a Hach Lange ISENO3181 nitrate probe.

Additionally, land-use was analysed in a 1 km radius around the sampling points using the national, high-resolution ecosystem basemap of Hungary (Ministry of Agriculture, 2019; Tanács et al., 2019) and calculating the proportion of six land-use categories for each site: forest, cropland, grassland, urban, water, wetland. All parameters measured are presented in Table S.1, with their mean and standard deviation, or class proportion along the sampling network.

2.3. Diatom sampling and laboratory procedure

Periphyton samples were collected from the most representative substrate (stones, macrophytes or sediment, Table S.1) at each sampling site following European standards (CEN, 2018a). The upper surface of the substrates were scrubbed using a clean toothbrush and washed into a tray. In the case of sediment, the samples were collected by pipetting the superficial layer of the substrate. The samples were then homogenized by manual shaking and transferred into 20 mL tubes containing 96% ethanol for a final concentration of >70%. The samples were stored in 4 °C, in the dark until DNA extraction that was executed within one year following the first sampling period.

For the DNA extraction, samples were first manually homogenized, and 2 mL of each sample was used. The 2 mL samples were first

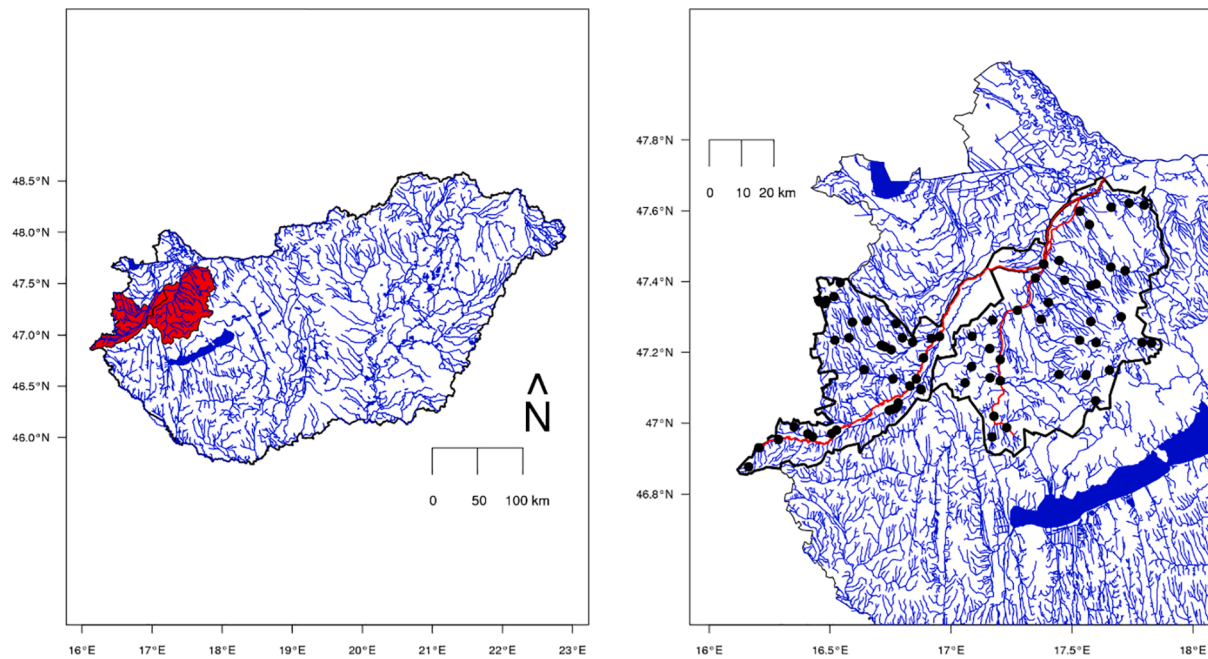


Fig. 1. The location of the sampling sites (black dots) in the catchment areas (borders indicated by black line) of the rivers Rába and Marcal (red lines) in Hungary. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

centrifuged at 13,000 rpm during 30 min at 4 °C in order to remove the preservative (ethanol) with the supernatant. The remaining pellet was used for total genomic DNA extraction using the Macherey-Nagel NucleoSpin® Soil kit extraction protocol with a final elution volume of 30 µL.

A short, 312 bp fragment of the Ribulose Biphosphate Carboxylase Large subunit (*rbcL*) chloroplast gene was used as the marker gene for PCR amplification. Primers used for amplification were the equimolar mix of three forward primers (Diat_rbcL_708F_1, Diat_rbcL_708F_2, Diat_rbcL_708F_3) and two reverse primers (R3_1, R3_2), after Vasselon et al. (2017b). Forward and reverse primers were tagged with the 5'-CTTCCCTACACGACGCTCTTCCGATCT-3' and the 5'-GGAGTTCAGAGTGTGCTCTTCCGATCT-3' Illumina adapters in order to prepare Illumina libraries in a dual-step PCR approach. During the first PCR step (PCR1), DNA samples were amplified in triplicates in a final volume of 25 µL using the tailed primers and the Takara LA Taq® polymerase and the reaction mix detailed in Table S.2. The PCR replicates of each DNA sample were then pooled together and transferred to the GenoToul Genomics and Transcriptomics platform (GeT-PlaGe, Auzeville, France). They performed a second PCR (PCR2) amplification using the purified PCR1 amplicons as template and the Illumina-tailed primers to add dual-index specific to the samples. They furthermore prepared the final pool that corresponds to an equimolar mix of the 130 PCR2 dual-indexed amplicons and carried out the sequencing of the pool using Illumina MiSeq sequencing platform with V3 paired-end sequencing kit (2×250bp).

2.4. Bioinformatics pipeline

The DADA2 pipeline (Callahan et al., 2016) was used for the bioinformatic treatment of the demultiplexed MiSeq reads. A pipeline adapted to diatom metabarcoding sequence data (available on https://github.com/fkeck/DADA2_diatoms_pipeline and in Data S.1) was applied with the following steps. The removal of primer sequences from R1 and R2 reads was done using cutadapt 2.9 (Martin, 2011). The quality profile of R1 and R2 reads were checked and the pattern showed good read quality and only the last 20–30 bps decreased in quality score (Fig. S.1). R1 and R2 reads were truncated to 200 and 170 nucleotides,

respectively, in order to remove the last, poor quality nucleotides. Truncated sequences were filtered using criteria of 0 ambiguities (“N”) and a maximum of expected errors (maxEE) of 2. An error model was executed showing that estimated error rates fit well to the observed rates and the error rates decrease with increased quality (Fig. S.2). R1 and R2 reads were then dereplicated into ISUs. ESVs were selected based on the error rate models determined by the DADA2 denoising algorithm and paired reads were merged into one sequence. Chimeras were then removed. Singletons in the dataset were also removed and sample size normalization was performed to read number 32,542, based on the rarefaction curves (Fig. S.3). Read numbers in each sample were tracked after each step of the bioinformatic pipeline and summarized in Table S.3. Treated data with the sequences and associated read numbers per samples are shown in Table S.4. Taxonomic assignment of ESVs was carried out with an adapted version (version 7) of the diat.barcode reference database following European standards for reference barcoding library management (CEN, 2018b; Keck et al., 2019; Rimet et al., 2019), with the R package “diat.barcode” (Keck, 2020), using a minimum bootstrap confidence of 75 for assigning a taxonomic level (Table S.5).

2.5. Data analysis and model development

2.5.1. Exploring the abiotic environment

In order to address our proposed hypotheses, preliminary analyses were done to study the relationships among abiotic variables. All data handling and analyses were carried out in R 3.6.2 (R Core Team, 2019) with the adequate packages. Principal component analysis (PCA) (Venables and Ripley, 2002) using the “prcomp” function was used to study the most important land-use categories responsible for the highest variation in the dataset. The relationship between land-use and continuous environmental variables was investigated via Pearson’s correlations on the log-transformed data. In addition to simple correlations, multiple linear regression (MLR) was conducted based on the combination of environmental variables. Parameters included in the MLR were chosen after model selection based on Akaike’s Information Criterion (AIC; Sakamoto et al., 1986) with stepwise model comparison run in both directions (Venables and Ripley, 2002). Non-parametric

Kruskal–Wallis test (Hollander and Wolfe, 1973) with post-hoc Dunn's test (Dunn, 1964) was used for the categorical variables.

2.5.2. Effect of land-use on diatom community

Based on the HTS data, two diatom datasets were created; an ESV and a species dataset. For the ESV data (Table S.6), the rarefied read numbers of the ESV dataset were transformed into relative abundances. For the species data (Table S.7), ESVs were clustered into species based on the taxonomic assignment detailed in Section 2.4. Unassigned ESVs were removed from this dataset. For both biological datasets, ESVs/species with a minimum of five occurrences were kept to maintain stable ecological profiles and Hellinger transformation (Legendre and Gallagher, 2001) was applied with the “decostand” function in the “vegan” package (Oksanen et al., 2019). Two non-metric multidimensional scaling (NMDS; Faith et al., 1987) analyses were run with the “metaMDS” function in “vegan” on both biological datasets, using Bray–Curtis dissimilarity index (Legendre and Legendre, 2012). Number of dimensions was set to 4 based on the inflexion point on a stress versus dimensionality plot (stress value < 0.2).

In order to address our first hypothesis on the effect of agricultural pressure, land-use variables were fitted onto the NMDS ordinations using the “envfit” function in the “vegan” package. Relevant ESVs/species were selected with the “ordiselect” function of the “goevg” package (Goral and Schellenberg, 2018), and were displayed on the ordination diagram. This was based on two criteria; (i) best fit to the gradient of significant environmental parameters and (ii) abundance. We set the proportion of taxonomic units with highest abundances to be displayed to 0.3 in case of ESVs and 0.5 in case of species. The proportion of species/ESVs with best fit to environmental factors to be displayed was set to 0.2 in both cases. Selection of these values was arbitrary but in accordance with other studies (Bohnenberger et al., 2018; Minerovic et al., 2020). A stricter abundance limit was set for ESVs as their abundance distribution is more skewed with rarer and fewer high-abundance ESVs. Important to note that in order to interpret results on ESV data, they were also assigned to taxa after conducting the analyses.

2.5.3. A taxonomy-free predictive model and intraspecific variability

In order to address our second hypothesis, two indices for assessing cropland land-use as a proxy of agricultural pressure was constructed, based on the two biological datasets, so that higher values in a 0–5 scale indicates higher pressure. The index development followed a modified version of indices built in former studies by the authors (Tapolczai et al., 2019a, 2018). Here, we used a leave-one-out cross validation (LOOCV, Molinaro et al., 2005), during which the model developed on the training dataset (n–1 samples) is computed on the one left-out test sample. Thus, as many models are calculated as many samples (130) are present, always leaving another sample out. It results in one computed index value for each of the samples. At each model, optima and tolerance values for the ESVs/species in the training dataset were calculated. For this purpose, abundance values were identified along the cropland gradient for each ESV/species and their optima and tolerance were calculated by the abundance-weighted median and the 1-interquartile range (1-IQR), respectively (Cristóbal et al., 2014). These values were used so that higher values for optima indicate preference for higher cropland proportion and higher values for tolerance indicate narrower tolerance range along the cropland gradient. For each model, we got a database of ESVs/species with their optimum and tolerance values (Table S.8 and S.9). In order to compute the index value for agricultural pressure (I_A) on the test samples, a weighted average equation (Zelinka and Marvan, 1961) was used (Eq. (1)). It is computed using the optima (o) and tolerance (t) values from the database of the ESVs/species (i) together with the abundance (a) values of these ESVs/species found in the test sample. The test values are then correlated with the cropland area proportion.

$$I_A = \frac{\sum_{i=1}^n a_i \times o_i \times t_i}{\sum_{i=1}^n a_i \times t_i} \quad (1)$$

Eq. (1) - The index for agricultural pressure assessment is a sum of weighted averages of ESV/species optima weighted by ESV/species abundance and tolerance

In order to study how potential intraspecific variability influences the indices, species with more than one detected ESVs were identified and the ecological profile (abundance values) of each ESV within each species along the cropland land-use category was determined. For studying significance of difference in the ecological optima (calculated by weighted median) between ESVs of the same species, Kruskal–Wallis tests (Hollander and Wolfe, 1973) with post hoc Dunn's test (Dunn, 1964) were run and interpreted. The number of unassigned ESVs and their ecological profiles were also calculated for assessing the ecological importance they hold. Kruskal–Wallis test (Hollander and Wolfe, 1973) was performed to study if there are significant differences among the ecological profiles of unassigned ESVs.

3. Results

3.1. Agricultural land-use and related parameters

Principal component analysis performed on land-use data shows that the first two, most important principal axes explain the 65% of the total variation (Fig. 2). The results show the importance of cropland, presenting a strong and significant correlation with PC1 ($p < 0.05$, $r = -0.83$). Correlation matrix performed for the continuous environmental parameters shows that cropland land-use correlates significantly ($p < 0.05$) with several environmental factors, presented in Table 1, among which the most important ones are the forest land-use, altitude, ripveg_index, macrophyte coverage, HCO_3^- , SO_4^{2-} , conductivity and Cl^- . Based on the AIC criteria for model selection, the following continuous parameters were selected to include in the multiple linear regression: altitude, stream width, ripveg_index, DO, Cl^- , COD and shaded area. The

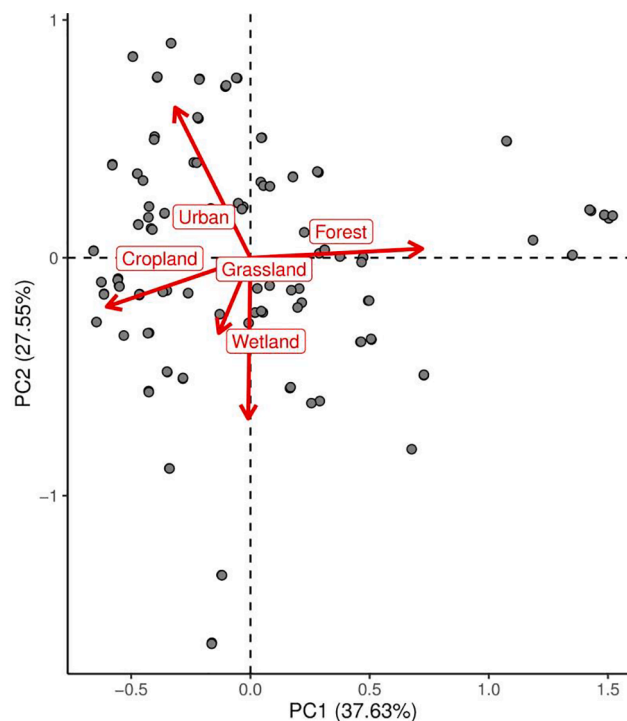


Fig. 2. Principal component analysis biplot based on land-use categories (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Pearson's correlation was computed between cropland land-use and continuous environmental variables. Only parameters with significant correlation are presented with the associated *r* and *p* values.

Cropland vs.	<i>r</i>	<i>p</i>	Cropland vs.	<i>r</i>	<i>p</i>
Forest	-0.77	<0.001	NO ₂ ⁻	0.27	<0.01
Altitude	-0.61	<0.001	NO ₃ ⁻	0.34	<0.001
Ripveg_index	-0.36	<0.001	Macrophyte coverage	0.39	<0.001
O ₂ saturation	-0.19	<0.05	HCO ₃ ⁻	0.42	<0.001
water flow	-0.18	<0.05	SO ₄ ²⁻	0.43	<0.001
NH ₄ ⁺	0.19	<0.05	Conductivity	0.50	<0.001
SRSi	0.25	<0.01	Cl ⁻	0.58	<0.001
SRP	0.25	<0.01			

multiple linear regression model (Table S.10) found altitude, Ripveg_index, Cl⁻ and COD as significant parameters ($R^2 = 0.46$, $p < 0.001$). Regarding the categorical variables, Kruskal-Wallis and post-hoc Dunn's test showed significant differences along the cropland gradient, except for stream order. Sediment and macrophytes as dominant substrates were found at a higher proportion of cropland (median = 0.55 and 0.42, respectively) than the stone substrate (median = 0.32). Straight stream morphology was also more common than meandering where the cropland proportion was higher (median = 0.40 and 0.30). Regarding river typology, streams with higher cropland proportions showed higher diversity, the only typological category was 2M that showed significantly lower values (median = 0.20) than the classes 3M, 3S, 6M and 6S (median = 0.36, 0.40, 0.53, 0.65, respectively).

3.2. Community structure and land-use categories

Average read number per sample in the raw database was 54,333.7 that was reduced to 45,141.92 during the bioinformatic process. Total number of ESVs were 2445 but this number was reduced to 667 by applying the limit of five occurrences, as presented in Section 2.5.2. This drop of a 72.72% in ESV number represents a drop of only 2.87% in relative abundance. The 667 different ESVs distributed with an average ESV number of 20.3 per sample. The minimum and maximum ESV number were 5 and 118, respectively. 366 ESVs (54.9%), representing 76.8% in relative abundance could be assigned to 140 species. An additional 100 ESVs were assigned to genus level with unknown species. The remaining 201 ESVs could be assigned to taxonomy levels above genus. The 466 ESVs assigned to at least genus level represent 51 different genera and 86.7% relative abundance.

The NMDS and environmental fitting analyses provided different results for the ESV and the species dataset. In the case of the ESV community, environmental fitting of the land-use categories showed significant correlations for cropland, forest, urban, wetland and grassland categories with R^2 values of 0.22, 0.24, 0.08, 0.11, 0.15, respectively (Fig. 3A). In case of species, only cropland, urban and wetland correlated significantly with the ordination with much lower, 0.08, 0.09 and 0.16 R^2 values, respectively. Most of the relevant ESVs correlated with the cropland and forest categories. In order to interpret the distribution of the most abundant and well correlating ESVs, their associated taxa names are also presented on Fig. 3B. Note that these are not the same as the taxa in the "species" dataset that was made by assigning and clustering ESVs into taxa before the analyses. Several unassigned ESVs were found correlating with both forest and cropland land-uses. Fewer ESVs correlated with the forest category than with the cropland category, including ESVs of species like *Amphora pediculus* Ehrenberg ex Kützing, *Cocconeis placentula* Ehrenberg, *Gomphonema pumilum* var. *pumilum* Ehrenberg, *Reimeria sinuata* (Gregory) Kocielek & Stoermer. More ESVs correlated with the cropland category including ESVs of the species *Navicula veneta* Kützing, *Planothidium caputum* Zimmermann & Jahn, *Gomphonema saprophilum* Ehrenberg, *Encyonema ventricosum* (Agardh) Grunow, *Ulnaria ulna* (Nitzsch) Compère, *Cymbella neocistula* Krammer, *Conticribra weissflogii* (Grunow) Stachura-Suchoples & Williams, *Sellaphora seminulum* (Grunow) Mann, *Cyclotella meneghiniana* Kützing and several species of *Nitzschia* as *N. palea* (Kützing) Smith, *N. paleacea* Grunow in Van Heurck, *N. fonticola* (Grunow) Grunow, and *N. linearis* Smith.

3.3. Assessment of agricultural stress and the importance of hidden intraspecific diversity

The model developed to assess agricultural stress performed differently based on the two types of input data; ESVs and species. Although the linear models were both significant ($p < 0.001$), the model developed on the ESVs had higher R^2 (0.42) and lower residual standard error (RSE = 0.60) (Fig. 4A) than the model based on species by grouping ESVs into taxa ($R^2 = 0.15$, RSE = 0.72) (Fig. 4B). We found five diatom species containing from 3 to 12 different ESVs within which significant differences (Kruskal-Wallis test and Dunn's post-hoc, with Bonferroni correction, $p < 0.05$) were found among the ESVs' optima along the cropland gradient. These species, together with the number of ESVs they possess, are *Cocconeis placentula* (9 ESVs), *Amphora pediculus* (12 ESVs), *Gomphonema pumilum* var. *pumilum* (3 ESVs), *Mayamaea permitis*

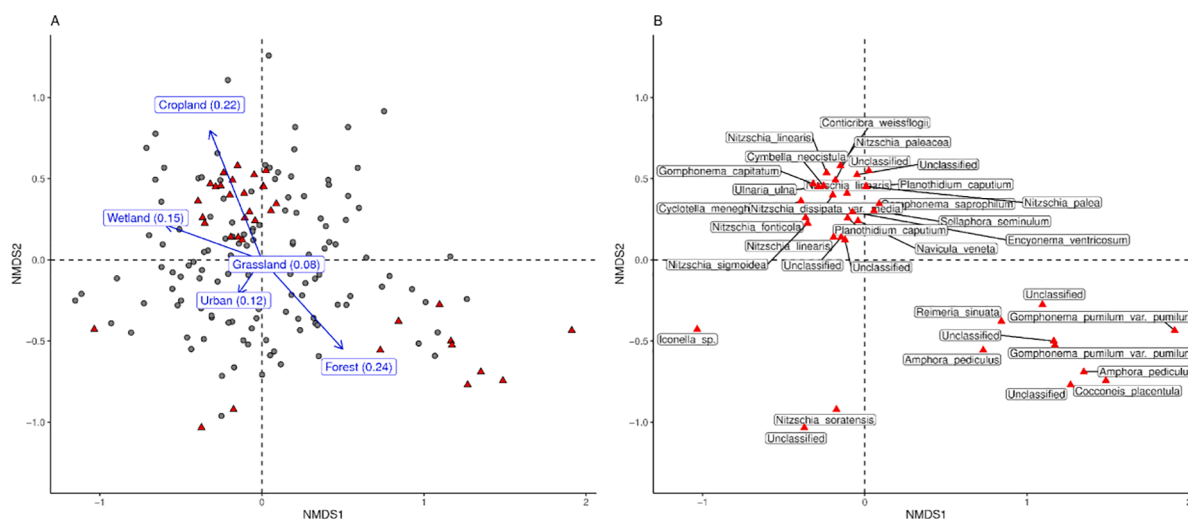


Fig. 3. (A) NMDS biplot with the samples (grey circles), important ESVs (red triangles) and the fitted land-use categories (blue arrows). (B) NMDS biplot presenting only the most important ESVs and their species assignment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

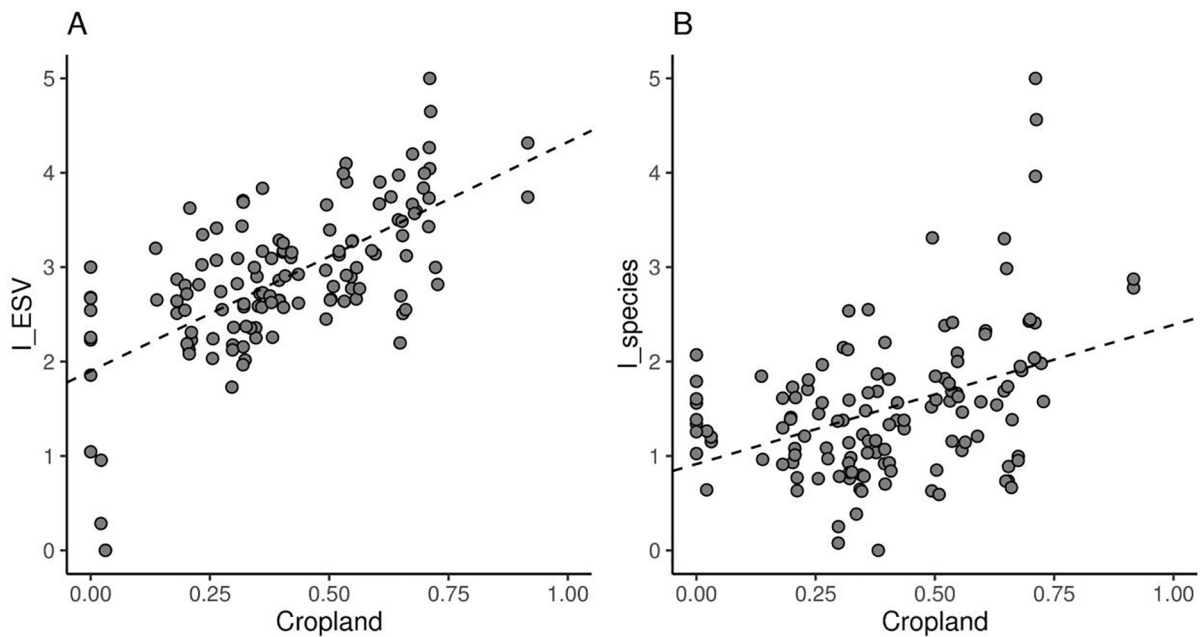


Fig. 4. Assessment index values correlated with the cropland category values based on ESVs (A) and species (B).

(Hustedt) K.Bruder & Medlin (9 ESVs), *Sellaphora minima* (15 ESVs) (Table S.11). Fig. 5 shows the ecological profiles of the ESVs for each of these species along the agricultural stress gradient together with the optimum and tolerance values of the species itself, containing all its ESVs. We found 301 ESVs that could not be assigned to any species, and thus are not involved in the species-based assessment index. These ESVs possessed significantly different ecological profiles with a high variety of optima and tolerance along the cropland land-use proportion gradient (Fig. 6).

4. Discussion

4.1. Characterization of the altered stream habitats due to land-use

Our results confirm previous studies that land-use represents an integrative proxy of the numerous environmental drivers directly or indirectly influencing aquatic communities (Allan, 2004). Out of the five land-use categories we tested, cropland area was the major anthropogenic alteration on stream catchments and thus, our study focused on it. As our findings also indicate, agricultural areas expand at the expense of forested areas which represent natural conditions in the study area (Borics et al., 2016). Agricultural land-use alters the natural habitats via

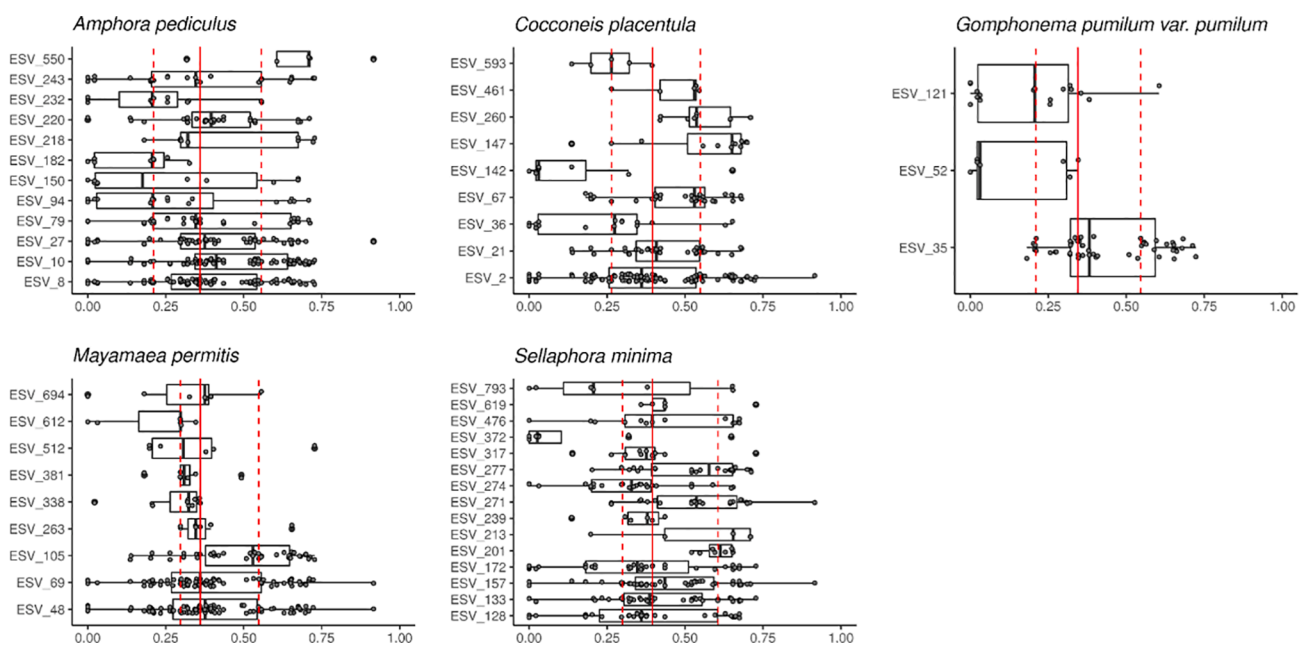


Fig. 5. Ecological profiles of species within which significant differences (Kruskal-Wallis and post-hoc Dunn's test, $p < 0.05$) were found. Medians show the ecological optimum, boxplot IQRs indicate ecological tolerance. Red dashed and solid lines represent the ecological tolerance and optimum for the entire species, involving all ESVs within them, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

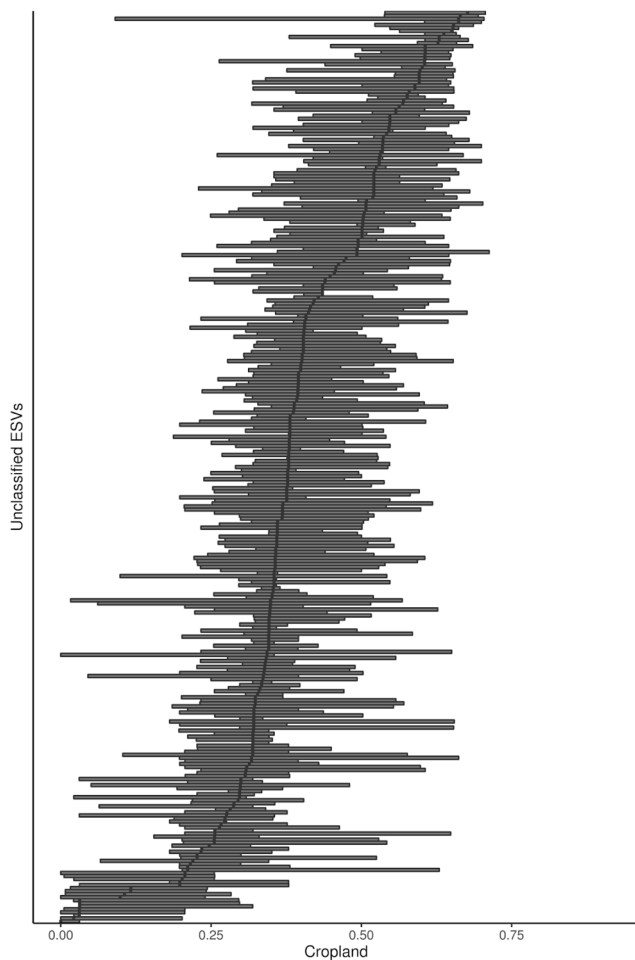


Fig. 6. Ecological profiles of unassigned ESVs along the cropland land-use category proportion ordered by median value. Medians represent ecological optima and the boxplot IQRs represent ecological tolerance of ESVs.

e.g. application of fertilizers that increase soluble salt concentrations in the water like Cl^- , SO_4^{2-} , or HCO_3^- and thus, the conductivity. Moreover, they increase availability of plant nutrients, especially that of nitrogen and phosphorus forms (Bullard, 1966). The physical alteration of these streams and their catchments involves a change in the morphology from meandering to straight channel-like streams and/or the reduction of the riparian vegetation (Cooper et al., 2013; Stenger-Kovács et al., 2006). The lack of shade along with the elevated nutrient concentrations lead to the formation of macrophytes in these streams (Mebane et al., 2014). The dense macrophyte vegetation can eventually create a light-limited environment for microalgae similarly to experiences in other forested areas (Celewicz-Goldyn et al., 2017). This can explain why we did not find correlation between the variable “shaded area” and the cropland area, however multiple linear regression showed the importance of this parameter combined with other ones. Agricultural activity influences the dominant substrates in streams that is also shown to have an important effect on benthic diatom assemblages. Elevated runoff and light exposure due to reduced forested area nearby the streams lead to sediment and macrophyte substrates instead of stones (Zeiger and Hubbart, 2019). According to the results on the typological classes, cropland area affects diverse typological classes regardless the type of bedrock, catchment size, and slope. The only class occurring at significantly lower cropland proportion values is 2M that is for mountainous small streams located at high altitude. 4L and 7L typologies showed transitional values in this regard. These are larger rivers, typically sites of the rivers Rába and Marcal where extended riparian vegetation is present.

4.2. Relationship between diatom assemblages and land-use

The importance of land-use categories, mainly forest and cropland, in shaping the diatom assemblage were further confirmed by the NMDS analyses. Previous studies analysing the effect of land-use on diatoms (Blinn and Bailey, 2001; Stenger-Kovács et al., 2020; Walsh and Wepener, 2009) identified species with the traditional microscopical approach but the present study is the first where metabarcoding data with individual DNA sequences of diatoms was used in this aspect.

We furthermore found that environmental fitting provided better R^2 values when ESVs were used without clustering them into taxa due to mainly two reasons. Firstly, it enlightens the important differences regarding the ecology of sequences within the same species and how relevant they are in the community structure. Secondly, unassigned sequences are of important ecological value and their removal from the dataset leads to a major loss of information. We showed that several unassigned ESVs correlated significantly both with forest and cropland land-use categories. However, in order to be able to give ecological interpretation to our results, taxonomic assignment of ESVs after analysis was performed. It shows another advantage of ESVs over OTUs. Since OTUs are clusters of sequences that can change over studies, they are specific to a particular analysis and cannot be compared to the OTUs obtained in another study (Cordier et al., 2020; Tapolczai et al., 2018). Conversely, ESVs are single, denoised sequences, thus they can be the subject of intercomparison studies (Callahan et al., 2017).

We found characteristic taxa correlated with the cropland-forest land-use gradient. The siltation of the riverbed and the sediment substrate favours the occurrence of those motile taxa (Passy, 2007) that have adaptive advantage under such circumstances (*Nitzschia* spp., *Navicula veneta*, *Sellaphora seminulum*) to actively move towards the required conditions (Bahls, 1993; Dalu et al., 2020). Several other species at higher cropland level were high profile (Passy, 2007), epiphytic (Round et al., 1990) species (*Cymbella neocistula*, *Gomphonema capitatum* Ehrenberg, *G. saprophilum*, *Encyonema ventricosum*) that could outcompete other species by emerging from the dense biofilm to gain access to better nutrient concentrations and light availability (Leira et al., 2015; Rimet et al., 2015). *Cyclotella meneghiniana* and the euryhaline *Constrictibra weissflogii* (Grunow) Stachura-Suchoples & Williams are two planktic species but their presence in nutrient rich streams with elevated conductivity is already documented (B-Béres et al., 2017; Kiss et al., 2012; Sabater, 1990). Although flow velocity correlated weakly with the cropland area, we can suggest that this, or the presence of still sections in the streams could lead to the development of these species whose DNA is detected in the samples. At the opposite end of the gradient with low cropland but higher forest area, we found species characteristic to nutrient- and light-limited environments where the development of a thick autotrophic biofilm is limited (Smith et al., 2009). *Amphora pediculus*, *Cocconeis placentula* and *Reimeria sinuata* that were found in this environment are low profile (Passy, 2007) species positioned on the bottom layer of the biofilm attached strongly to the dominant stone substrate with the entire valve surface. Under such conditions environmental stress is the major assemblage-shaping force instead of competition that occurs in a dense biofilm (Tapolczai et al., 2016).

4.3. Assessment of agricultural pressure

Our hypothesis was further confirmed by the predictive models we developed. They showed that the ecological information obtained by ESVs provided an assessment index with a better prediction power than the index based on taxa after taxonomic assignment of the sequences. This recently developed strategy is often referred to as taxonomy-free approach (Apothéoz-Perret-Gentil et al., 2017; Feio et al., 2020; Tapolczai et al., 2018) because it neglects the classical delimitation of taxa, thus revealing ecological differences between molecular taxonomic units. In this way, the use of standardised denoising algorithms like DADA2 (Callahan et al., 2016) allows the creation of a large dataset

including the ecological information of unique sequences explored by different studies. Taxonomy-free approach proposes a solution for the incompleteness of reference libraries that normally leads to the loss of information held by unassigned sequences. Although the completion of such databases is rapidly improving (Rimet et al., 2019), it has remained an issue to be solved (Weigand et al., 2019). It is important to note that although this approach can be effective in regular monitoring and quality assessment, it can poorly support ecological interpretations. We showed that “post-analysis” taxonomic assignment and the knowledge of species’ traits helped to interpret their presence and relationship with the habitat they inhabit.

Further novelties of the developed indices compared to previous ones are the definition of the pressure gradient and the calculation of ESVs’ optima and tolerances. Former studies (Tapolczai et al., 2019, 2018) used ordination axes to define an integrative pressure gradient. However, axis values by themselves have little sense and it is hard to interpret what kind of conditions they indicate. In contrast, cropland area around the sampling site is a meaningful parameter that combines the effects of the many, directly acting variables.

In order to determine the ecological values of diatom species, it is common to calculate the mean value weighted by the abundance along the pressure gradient for the optimum and the weighted standard deviation for the tolerance (Ter Braak and Barendregt, 1986). This approach assumes a symmetric gaussian distribution of species’ abundance along the gradient. However species’ distribution often follows a right- or left skewed pattern where weighted mean under- or overestimates optima (Potapova et al., 2004). In this study, we handled this bias by applying the robust optimum (RO) method which assesses optimum and tolerance by the abundance-weighted median and the abundance-weighted interquartile range (IQR), respectively, giving a more realistic estimation (Cristóbal et al., 2014).

One reason for the better performance of the ESV-based index is the above mentioned cryptic diversity when genetic variations of the same morphospecies could possess different ecological adaptations (Rivera et al., 2018). Regarding the cropland area gradient, we found that in case of five species, significant differences in the abundance within the same species could be observed. While the optima of the species are all around 0.3–0.4 cropland proportion value, some ESVs within the species have significantly higher or lower values that is also true for the tolerances. We suggest that revealing this hidden ecological information helps in fine tuning the index. This is an issue that can be handled by the curation of the reference libraries. In this study, we used the version 7 of diat.barcode (Rimet et al., 2019) that reflects the taxonomy at that time. In version 8 of diat.barcode *Cocconeis placentula* and *Sellaphora minima* have been splitted to follow current taxonomy.

We furthermore suggest that the quantity of unassigned sequences is crucial even though they are normally removed from the biological dataset, taking the risk of eliminating a relevant part of the obtainable ecological information. The degree of this loss firstly depends on the studied geographical area. As the development of the diat.barcode reference database is mainly a French collaboration, its content also largely focuses on taxa common in that region. We could assign 54.9% of the ESVs to species, which is reasonable compared to other studies. Rivera et al. (2018) could assign only the 10% of their OTUs to species level using the R-Syst:diatom database v4 (Rimet et al., 2016), the predecessor of diat.barcode. They analysed samples collected from green turtles (*Chelonia mydas* Linnaeus, 1758) from Mayotte Island in the Indian-Ocean where we can assume a rather different diatom community than in Western-Europe, especially because they are from marine habitat. Better assignment proportion (32%) could be reached e.g. from Portuguese rivers (Mortágua et al., 2019), from rivers in Catalonia, Spain (48.6%) (Pérez-Burillo et al., 2020) or from a French river (57%) (Chonova et al., 2019). The 54.9% in our study however represented 76.8% regarding the abundance. The second factor influencing the importance of the unassigned ESV loss is the type of study. We chose to set a minimal occurrence of five (number of samples where a given ESV

is found) at the price of losing the 72.72% of the ESVs but it represented only the 2.87% in abundance. The same strategy in a study exploring diversity (Groendahl et al., 2017; Piredda et al., 2018; Zimmermann et al., 2015) or looking for rare taxa (Smith et al., 2017) would be inappropriate.

5. Conclusion

A rapid development on the implementation of metabarcoding methods in biomonitoring could be observed in recent years, stimulated partly by efficient international collaborations (e.g. Leese et al., 2016). These attempts largely focus on the issue to implement metabarcoding to replace already existing traditional methods (e.g. microscopic analysis) for a more standardised and efficient biomonitoring. In our study, we aimed to contribute in this development but also to give ecological interpretations.

After testing our hypothesis, we can conclude that (i) the effect of agricultural land-use could be better detected with individual sequences of diatoms (ESVs) than with diatom taxa and (ii) the index assessing agricultural stress we developed also worked with higher efficiency on diatom ESVs than diatom taxa.

Our findings explaining this higher efficiency with ESVs raise the often discussed question in the scientific literature of the appropriate taxonomic delimitation, as diatom morphology, on which their taxonomy still strongly relies, do not perfectly reflect the ecological requirements of the organisms (De Queiroz, 2007; Leliaert et al., 2014; Mann, 1999). On the one hand, traditional taxonomy is redundant since several morphospecies occupy largely overlapping niches (Rimet and Bouchez, 2012), on the other hand, the problem is raised by cryptic diversity (Amato et al., 2019; Rovira et al., 2015). Several attempts have been made to handle this incoherence with e.g. the use of traits or functional approaches (B-Béres et al., 2016; Lange et al., 2016; Stenger-Kovács et al., 2018). Novel genetic technologies, such as HTS provide solutions by clarifying diatom phylogeny (Blanco et al., 2017; Nakov et al., 2018) or revealing cryptic diversity (Kermarrec et al., 2013). Yet, it also produces a high-level of redundancy by providing hundreds of thousands of sequences that are either clustered into OTUs based on genetic similarity or used individually. Finally, these units are assigned to the traditional taxonomy or used without assignment as in this study. A potential direction could be the clustering of sequences based on their ecology but few attempts have been made for this generally (Preheim et al., 2013) and none for diatoms.

Funding

This work was supported by the Premium Postdoctoral Research Program of the Hungarian Academy of Sciences (PPD2018-026/2018); the COST Action DNAqua-Net (CA15219) supported by the COST (European Cooperation in Science and Technology); the National Research Development and Innovation Office (NKFIH K120595, NKFIH FK132142, TKP2020-IKA-07 under 2020-4.1.1-TKP2020); and the GINOP-2.3.2-15-2016-00019.

CRedit authorship contribution statement

Kálmán Tapolczai: Conceptualization, Methodology, Software, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Géza B. Selmeczy:** Methodology, Writing - review & editing. **Beáta Szabó:** Formal analysis, Writing - review & editing. **Viktória B-Béres:** Writing - review & editing. **François Keck:** Methodology, Software, Formal analysis, Writing - review & editing. **Agnès Bouchez:** Resources, Writing - review & editing, Funding acquisition. **Frédéric Rimet:** Resources, Writing - review & editing, Funding acquisition. **Judit Padisák:** Resources, Writing - review & editing, Funding acquisition.

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.

Acknowledgement

We would like to thank all the colleagues at the Department of Limnology, University of Pannonia for their assistance in the field sampling and in the laboratory analyses of the water samples. We are also grateful to the colleagues at the molecular laboratory of INRAE UMR CARRTEL for their essential guidance and help.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2020.107322>.

References

- Allan, J.D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annu. Rev. Ecol. Syst.* 35 (1), 257–284. <https://doi.org/10.1146/annurev.ecolsys.35.120202.110122>.
- Amato, A., Kooistra, W.H.C.F., Montresor, M., 2019. Cryptic diversity: a long-lasting issue for diatomologists. *Protist* 170 (1), 1–7.
- APHA, 2012. *Standard Methods for the Examination of Water and Wastewater, 22nd ed.* APHA American Public Health Association, Washington.
- Apothéoz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., Pawlowski, J., 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Mol. Ecol. Resour.* <https://doi.org/10.1111/1755-0998.12668>.
- Bahls, L.L., 1993. Periphyton bioassessment methods for Montana streams. Water Quality Bureau, Department of Health and Environmental Sciences.
- Baillet, 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>.
- 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. Indic.* 64, 31–38. <https://doi.org/10.1016/j.ecolind.2015.12.031>.
- B-Béres, V., Török, P., Kókai, Z., Lukács, Á., T-Krasznai, E., Tóthmérész, B., Bácsi, I., 2017. Ecological background of diatom functional groups: comparability of classification systems. *Ecol. Indic.* 82, 183–188. <https://doi.org/10.1016/j.ecolind.2017.07.007>.
- Blackman, R.C., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., Egeter, B., Elbrecht, V., Filipe, A.F., Jones, J.L., Macher, J., Majaneva, M., Martins, F.M.S., Murria, C., Meissner, K., Pawlowski, J., Schmidt Yanez, P.L., Zizka, V.M.A., Leese, F., Price, B.W., Deiner, K., 2019. Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. *Metabarcoding Metagenomics* 3, 49–57. <https://doi.org/10.3897/mbmg.3.34735>.
- Blanco, S., Borrego-Ramos, M., Olenici, A., 2017. Disentangling diatom species complexes: does morphometry suffice? *PeerJ* 5, e4159. <https://doi.org/10.7717/peerj.4159>.
- Blinn, D.W., Bailey, P.C.E., 2001. Land-use influence on stream water quality and diatom communities in Victoria, Australia: a response to secondary salinization. *Hydrobiologia* 466, 231–244. <https://doi.org/10.1023/A:1014541029984>.
- Bohnenberger, J.E., Schneck, F., Crossetti, L.O., Lima, M.S., Motta-Marques, D.D., 2018. Taxonomic and functional nestedness patterns of phytoplankton communities among coastal shallow lakes in southern Brazil. *J. Plankton Res.* 40, 555–567. <https://doi.org/10.1093/plankt/fby032>.
- Borics, G., Ács, É., Boda, P., Boros, E., ErHos, T., Grigorszky, I., Kiss, K.T., Lengyel, S., Somogyi, B., Vörös, L., 2016. Water bodies in Hungary—an overview of their management and present state. *Hung. J. Hydrol.* 86, 57–67.
- Bradshaw, Emily G., Nielsen, Anne Birgitte, Anderson, N. John, 2006. Using diatoms to assess the impacts of prehistoric, pre-industrial and modern land-use on Danish lakes. *Reg. Environ. Change* 6 (1–2), 17–24. <https://doi.org/10.1007/s10113-005-0007-4>.
- Broetto, T., Tornquist, C.G., de Campos, B.-H.C., Schneider, J.C., Broetto, T., Tornquist, C.G., de Campos, B.-H.C., Schneider, J.C., 2017. Relationships between agriculture, riparian vegetation, and surface water quality in watersheds. *Rev. Bras. Ciênc. Solo* 41. <https://doi.org/10.1590/18069657rbc20160286>.
- Bullard, W.E., 1966. Effects of land use on water resources. *J. Water Pollut. Control Fed.* 38, 645–659.
- Callahan, Benjamin J, McMurdie, Paul J, Holmes, Susan P, 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11 (12), 2639–2643. <https://doi.org/10.1038/ismej.2017.119>.
- Callahan, Benjamin J, McMurdie, Paul J, Rosen, Michael J, Han, Andrew W, Johnson, Amy Jo A, Holmes, Susan P, 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13 (7), 581–583. <https://doi.org/10.1038/nmeth.3869>.
- Caporaso, J Gregory, Kuczynski, Justin, Stombaugh, Jesse, Bittinger, Kyle, Bushman, Frederic D, Costello, Elizabeth K, Fierer, Noah, Peña, Antonio Gonzalez, Goodrich, Julia K, Gordon, Jeffrey I, Huttley, Gavin A, Kelley, Scott T, Knights, Dan, Koenig, Jeremy E, Ley, Ruth E, Lozupone, Catherine A, McDonald, Daniel, Muegge, Brian D, Pirrung, Meg, Reeder, Jens, Sevinsky, Joel R, Turnbaugh, Peter J, Walters, William A, Widmann, Jeremy, Yatsunenko, Tanya, Zaneveld, Jesse, Knight, Rob, 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7 (5), 335–336. <https://doi.org/10.1038/nmeth.f.303>.
- Celewicz-Goldyn, Sofia, Kuczyńska-Kippen, Natalia, Heneberg, Petr, 2017. Ecological value of macrophyte cover in creating habitat for microalgae (diatoms) and zooplankton (rotifers and crustaceans) in small field and forest water bodies. *PLOS ONE* 12 (5), e0177317. <https://doi.org/10.1371/journal.pone.0177317>.
- CEN, 2018a. Water quality - CEN/TR 17245 - Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses.
- CEN, 2018b. Water quality - CEN/TR 17244:2018 Technical report for the management of diatom barcodes.
- CEN, 2003. Water Quality – Guidance Standard for the Routine Sampling and Pretreatment of Benthic Diatoms from Rivers (No. EN 13946). Geneva: Comité Européen de Normalisation.
- Chen, Wei, Zhang, Clarence K., Cheng, Yongmei, Zhang, Shaowu, Zhao, Hongyu, Casiraghi, Maurizio, 2013. A comparison of methods for clustering 16S rRNA sequences into OTUs. *PLOS ONE* 8 (8), e70837. <https://doi.org/10.1371/journal.pone.0070837>.
- Chonova, T., Kurmayer, R., Rimet, F., Labanowski, J., Vasselon, V., Keck, F., Illmer, P., Bouchez, A., 2019. Benthic diatom communities in an alpine river impacted by waste water treatment effluents as revealed using DNA metabarcoding. *Front. Microbiol.* 10 <https://doi.org/10.3389/fmicb.2019.00653>.
- Cooper, Scott D., Lake, P. Sam, Sabater, Sergi, Melack, John M., Sabo, John L., 2013. The effects of land use changes on streams and rivers in mediterranean climates. *Hydrobiologia* 719 (1), 383–425. <https://doi.org/10.1007/s10750-012-1333-4>.
- Cordier, T., Alonso-Sáez, L., Apothéoz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A., Chariton, A., Creer, S., Frihe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J., Lanzén, A., 2020. Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. *Mol. Ecol.* <https://doi.org/10.1111/mec.15472>.
- Cristóbal, Elena, Ayuso, Sergio Velasco, Justel, Ana, Toro, Manuel, 2014. Robust optima and tolerance ranges of biological indicators: a new method to identify sentinel of global warming. *Ecol. Res.* 29 (1), 55–68. <https://doi.org/10.1007/s11284-013-1099-9>.
- Dalu, Tatenda, Cuthbert, Ross N., Chavalala, Tiyisani L., Froneman, P. William, Wasserman, Ryan J., 2020. Assessing sediment particle-size effects on benthic algal colonisation and total carbohydrate production. *Sci. Total Environ.* 710, 136348. <https://doi.org/10.1016/j.scitotenv.2019.136348>.
- De Queiroz, K., 2007. Species Concepts and Species Delimitation. *Syst. Biol.* 56, 879–886. <https://doi.org/10.1080/10635150701701083>.
- DeLorenzo, Marie E., Scott, Geoffrey I., Ross, Philippe E., 2001. Toxicity of pesticides to aquatic microorganisms: a review. *Environ. Toxicol. Chem.* 20 (1), 84–98. <https://doi.org/10.1002/etc.v20:110.1002/etc.5620200108>.
- Dunn, Olive Jean, 1964. Multiple comparisons using rank sums. *Technometrics* 6 (3), 241–252.
- Elbrecht, Vasco, Vamos, Ecaterina Edith, Meissner, Kristian, Aroviita, Jukka, Leese, Florian, Yu, Douglas, 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods Ecol. Evol.* 8 (10), 1265–1275. <https://doi.org/10.1111/mee3.2017.8.issue-10:1111/2041-210X.12789>.
- Commission, European, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23rd October 2000 establishing a framework for Community action in the field of water policy. *Off. J. Eur. Communities* 327, 1–72.
- Faith, D.P., Minchin, P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69, 57–68.
- Feio, Maria João, Serra, Sónia R.Q., Mortágua, Andreia, Bouchez, Agnès, Rimet, Frédéric, Vasselon, Valentin, Almeida, Salomé F.P., 2020. A taxonomy-free approach based on machine learning to assess the quality of rivers with diatoms. *Sci. Total Environ.* 722, 137900. <https://doi.org/10.1016/j.scitotenv.2020.137900>.
- Goral, F., Schellenberg, J., 2018. *goveeg: Functions for Community Data and Ordinations.* R package version (4), 2.
- Groendahl, Sophie, Kahlert, Maria, Fink, Patrick, Doi, Hideyuki, 2017. The best of both worlds: a combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. *PLOS ONE* 12 (2), e0172808. <https://doi.org/10.1371/journal.pone.0172808>.
- Guo, L., Sui, Z., Zhang, S., Ren, Y., Liu, Y., 2015. Comparison of potential diatom 'barcode' genes (the 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *Int. J. Syst. Evol. Microbiol.* 65, 1369–1380. <https://doi.org/10.1099/ijs.0.000076>.
- He, Bin, Kanae, Shinjiro, Oki, Taikan, Hirabayashi, Yukiko, Yamashiki, Yosuke, Takara, Kaoru, 2011. Assessment of global nitrogen pollution in rivers using an integrated biogeochemical modeling framework. *Water Res.* 45 (8), 2573–2586. <https://doi.org/10.1016/j.watres.2011.02.011>.
- Hollander, M., Wolfe, D.A., 1973. *Nonparametric statistical methods*, 2nd ed. Wiley, New York, NY, USA.
- Johnson, L.T., Tank, J.L., Dodds, W.K., 2009. The influence of land use on stream biofilm nutrient limitation across eight North American ecoregions. *Can. J. Fish. Aquat. Sci.* 66, 1081–1094. <https://doi.org/10.1139/F09-065>.

- Kahlert, Maria, Albert, Raino-Lars, Anttila, Eeva-Leena, Bengtsson, Roland, Bigler, Christian, Eskola, Tiina, Gälman, Veronika, Gottschalk, Steffi, Herlitz, Eva, Jarlman, Amelie, Kasperovicene, Jurate, Kokocinski, Mikolaj, Luup, Helen, Miettinen, Juha, Paunksnyte, Ieva, Piirsoo, Kai, Quintana, Isabel, Raunio, Janne, Sandell, Bernt, Simola, Heikki, Sundberg, Irene, Vilbaste, Sirje, Weckström, Jan, 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. <https://doi.org/10.1007/s10811-008-9394-5>.
- Keck, F., 2020. diatbarcode: Access the diat.barcode database. R package version 0.0.0.9000.
- Keck, F., Rimet, F., Vasselon, V., Bouchez, A., 2019. A ready-to-use database for DADA2: Diat.barcode_rbc1_312bp_DADA2. <https://doi.org/10.15454/HN11EK>.
- Keck, F., Vasselon, V., Tapolczai, K., Rimet, F., Bouchez, A., 2017. Freshwater biomonitoring in the Information Age. *Front. Ecol. Environ.* <https://doi.org/10.1002/fee.1490>.
- Kelly, Martyn, King, Lydia, Ní Chatháin, Bernadette, 2009. The conceptual basis of ecological-status assessments using diatoms. *Biol. Environ. Proc. R. Ir. Acad.* 109 (3), 175–189. <https://doi.org/10.3318/BIOE.2009.109.3.175>.
- Kerमारrec, Lenaig, Bouchez, Agnès, Rimet, Frédéric, Humbert, Jean-François, 2013. First evidence of the existence of semi-cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kützinger) Kützinger complex (Bacillariophyta). *Protist* 164 (5), 686–705. <https://doi.org/10.1016/j.protis.2013.07.005>.
- Kerमारrec, Lenaig, Franc, Alain, Rimet, Frédéric, Chaumeil, Philippe, Frigerio, Jean-Marc, Humbert, Jean-François, Bouchez, Agnès, 2014. A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshw. Sci.* 33 (1), 349–363. <https://doi.org/10.1086/675079>.
- Kiss, K., Klee, R., Ector, L., Acs, E., 2012. Centric diatoms of large rivers and tributaries in Hungary: morphology and biogeographic distribution. *Acta Bot. Croat.* 71, 311–363.
- Lange, Katharina, Townsend, Colin Richard, Matthaei, Christoph David, 2016. A trait-based framework for stream algal communities. *Ecol. Evol.* 6 (1), 23–36. <https://doi.org/10.1002/ece3.2016.6.issue-110.1002/ece3.1822>.
- Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., Mergen, P., Pawlowski, J., Piggott, J., Rimet, F., Steinke, D., Taberlet, P., Weigand, A., Abanador, K., Beja, P., Bervoets, L., Björnsdóttir, S., Boets, P., Boggero, A., Bones, A., Borja, A., Bruce, K., Bursić, V., Carlsson, J., Ciampor, F., Ciamporová-Zatovičová, Z., Coissac, E., Costa, F., Costache, M., Creer, S., Csabai, Z., Deiner, K., DelValls, A., Drakare, S., Duarte, S., Elerséki, T., Fazi, S., Fiser, C., Flot, J.-F., Fonseca, V., Fontaneto, D., Grabowski, M., Graf, W., Guðbrandsson, J., Hellström, M., Hershkovitz, Y., Hollingsworth, P., Japoshvili, B., Jones, J., Kahlert, M., Kalamujic Strojil, B., Kasapidis, P., Kelly, M., Kelly-Quinn, M., Keskin, E., Kóljalj, U., Ljubešić, Z., Maček, I., Mächler, E., Mahon, A., Marečková, M., Mejdandžic, M., Mircheva, G., Montagna, M., Moritz, C., Mulk, V., Naumoski, A., Navodaru, I., Padišák, J., Pálsson, S., Panksep, K., Penev, L., Petrušek, A., Pfannkuchen, M., Primmer, C., Rinkevich, B., Rotter, A., Schmidt-Kloiber, A., Segurado, P., Speksnijder, A., Stoev, P., Strand, M., Šulcius, S., Sundberg, P., Traugott, M., Tsigonopoulos, C., Turon, X., Valentini, A., van der Hoorn, B., Várbró, G., Vasquez Hadjilyra, M., Víguri, J., Vitonyté, I., Vogler, A., Vrålstad, T., Wägele, W., Wenne, R., Winding, A., Woodward, G., Zegura, B., Zimmermann, J., 2016. DNAqua-Net: developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Res. Ideas Outcomes* 2, e11321. <https://doi.org/10.3897/rio.2.e11321>.
- Legendre, Pierre, Gallagher, Eugene D., 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129 (2), 271–280. <https://doi.org/10.1007/s004420100716>.
- Legendre, P., Legendre, L., 2012. *Numerical ecology*, 3rd English edition.
- Leira, M., Filippi, M.L., Cantonati, M., 2015. Diatom community response to extreme water-level fluctuations in two Alpine lakes: a core case study. *J. Paleolimnol.* 53, 289–307. <https://doi.org/10.1007/s10933-015-9825-7>.
- Leliaert, Frederik, Verbruggen, Heroen, Vanormelingen, Pieter, Steen, Frederique, López-Bautista, Juan M., Zuccarello, Giuseppe C., De Clerck, Olivier, 2014. DNA-based species delimitation in algae. *Eur. J. Phycol.* 49 (2), 179–196. <https://doi.org/10.1080/09670262.2014.904524>.
- Mahé, F., Rognes, T., Quince, C., Vargas, C. de, Dunthorn, M., 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3, e1420. <https://doi.org/10.7717/peerj.1420>.
- Mann, David G., 1999. The species concept in diatoms. *Phycologia* 38 (6), 437–495. <https://doi.org/10.2216/i0031-8884-38-6-437.1>.
- Martin, Marcel, 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17 (1), 10. <https://doi.org/10.14806/ej.17.1.10.14806/ej.17.1.200>.
- McGee, K.M., Robinson, C.V., Hajibabaei, M., 2019. Gaps in DNA-Based Biomonitoring Across the Globe. *Front. Ecol. Evol.* 7 <https://doi.org/10.3389/fevo.2019.00337>.
- Mebane, Christopher A., Simon, Nancy S., Maret, Terry R., 2014. Linking nutrient enrichment and streamflow to macrophytes in agricultural streams. *Hydrobiologia* 722 (1), 143–158. <https://doi.org/10.1007/s10750-013-1693-4>.
- Miettinen, Juha O., Simola, Heikki, Grönlund, Elisabeth, Lahtinen, Jaana, Niinjoja, Riitta, 2005. Limnological effects of growth and cessation of agricultural land use in Ladoga Karelia: sedimentary pollen and diatom analyses. *J. Paleolimnol.* 34 (2), 229–243. <https://doi.org/10.1007/s10933-005-2542-x>.
- Minorovic, Alison D., Potapova, Marina G., Sales, Christopher M., Price, Jacob R., Enache, Mihaela D., 2020. 18S–V9 DNA metabarcoding detects the effect of water-quality impairment on stream biofilm eukaryotic assemblages. *Ecol. Indic.* 113, 106225. <https://doi.org/10.1016/j.ecolind.2020.106225>.
- Ministry of Agriculture, 2019. Development of an ecosystem basemap and data model. Hungarian: Ökoszisztéma alaptérkép és adatmodell kialakítása. <https://doi.org/DOI:10.34811/osz.alapterkep>.
- Molinario, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 21, 3301–3307. <https://doi.org/10.1093/bioinformatics/bti499>.
- Mortágu, 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. <https://doi.org/10.1016/j.ecolind.2019.105470>.
- Moss, Brian, 2008. Water pollution by agriculture. *Philos. Trans. R. Soc. B Biol. Sci.* 363 (1491), 659–666. <https://doi.org/10.1098/rstb.2007.2176>.
- Naden, P.S., Murphy, J.F., Old, G.H., Newman, J., Scarlett, P., Harman, M., Duerdoth, C. P., Hawczak, A., Pretty, J.L., Arnold, A., Laizé, C., Hornby, D.D., Collins, A.L., Sear, D.A., Jones, J.L., 2016. Understanding the controls on deposited fine sediment in the streams of agricultural catchments. *Sci. Total Environ.* 547, 366–381. <https://doi.org/10.1016/j.scitotenv.2015.12.079>.
- Nakov, Teofil, Beauieu, Jeremy M., Alverson, Andrew J., 2018. Insights into global planktonic diatom diversity: The importance of comparisons between phylogenetically equivalent units that account for time. *ISME J.* 12 (11), 2807–2810. <https://doi.org/10.1038/s41396-018-0221-y>.
- Nearing, J.T., Douglas, G.M., Comeau, A.M., Langille, M.G.I., 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ* 6, e5364. <https://doi.org/10.7717/peerj.5364>.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., H. Stevens, M.H., Szoecs, E., Wagner, H., 2019. *vegan: Community Ecology Package*. R package version 2.5–6. 2019.
- Pan, Yangdong, Herlihy, Alan, Kaufmann, Philip, Wigginton, Jim, van Sickle, John, Moser, Tom, 2004. Linkages among land-use, water quality, physical habitat conditions and lotic diatom assemblages: a multi-spatial scale assessment. *Hydrobiologia* 515 (1–3), 59–73. <https://doi.org/10.1023/B:HYDR.0000027318.11417.e7>.
- Passy, Sophia I., 2007. Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquat. Bot.* 86 (2), 171–178. <https://doi.org/10.1016/j.aquabot.2006.09.018>.
- Pedersen, Morten Lauge, 2009. Effects of channelisation, riparian structure and catchment area on physical habitats in small lowland streams. *Fundam. Appl. Limnol. Arch. Für Hydrobiol.* 174 (1), 89–99. <https://doi.org/10.1127/1863-9135/2009/0174-0089>.
- Pérez-Burillo, Javier, Trobajo, Rosa, Vasselon, Valentin, Rimet, Frédéric, Bouchez, Agnès, Mann, David 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., Claverie, J.-M., Decelle, J., de Vargas, C., Dunthorn, M., Edvardson, B., Ekrem, W., Forster, D., Kooistra, W.H.C.F., Logares, R., Massana, R., Mentrator, M., Not, F., Ogata, H., Pawlowski, J., Romac, S., Sarno, D., Stoeck, T., Zingone, A., 2018. Diatom diversity through HTS-metabarcoding in the coastal European seas. *Sci. Rep.* 8, 18059. <https://doi.org/10.1038/s41598-018-36345-9>.
- Potapova, Marina G., Charles, Donald F., Ponader, Karin C., Winter, Diane M., 2004. Quantifying species indicator values for trophic diatom indices: a comparison of approaches. *Hydrobiologia* 517 (1–3), 25–41.
- Preheim, Sarah P., Perrotta, Allison R., Martin-Platero, Antonio M., Gupta, Anika, Alm, Eric J., 2013. Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl. Environ. Microbiol.* 79 (21), 6593–6603. <https://doi.org/10.1128/AEM.00342-13>.
- R Core Team, 2019. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RBMP, 2015. Hungarian part of the Danube river basin - River Basin Management Plan. Hungarian: A Duna-vízgyűjtő magyarországi része; Vízgyűjtőgazdálkodási Terv; General Directorate of Water Management: Budapest, Hungary 1–666.
- Rimet, Frederic, Abarca, Nelida, Bouchez, Agnes, Kusber, Wolf-Henning, Jahn, Regine, Kahlert, Maria, Keck, Francois, Kelly, Martyn G., Mann, David, G., Piuz, Andre, Trobajo, Rosa, Tapolczai, Kalman, Vasselon, Valentin, Zimmermann, Jonas, 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 (1), 37–54. <https://doi.org/10.5507/fof.2017.013>.
- Rimet, Frédéric, Bouchez, Agnès, 2012. Biomonitoring river diatoms: Implications of taxonomic resolution. *Ecol. Indic.* 15 (1), 92–99. <https://doi.org/10.1016/j.ecolind.2011.09.014>.
- Rimet, F., Bouchez, A., Montuelle, B., 2015. Benthic diatoms and phytoplankton to assess nutrients in a large lake: Complementarity of their use in Lake Geneva (France–Switzerland). *Ecol. Indic.* 53, 231–239. <https://doi.org/10.1016/j.ecolind.2015.02.008>.
- Rimet, F., Chaumeil, P., Keck, F., Kerमारrec, 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 J. Biol. Databases Curation* 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, 1–12. <https://doi.org/10.1038/s41598-019-51500-6>.
- Rivera, S.F., Vasselon, V., Ballorain, K., Carpentier, A., Wetzel, C.E., Ector, L., Bouchez, A., Rimet, F., 2018. DNA metabarcoding and microscopic analyses of sea turtles biofilms: Complementarity to understand turtle behavior. *PLOS ONE* 13, e0195770. <https://doi.org/10.1371/journal.pone.0195770>.
- Rivera, Sinziana F., Vasselon, Valentin, Bouchez, Agnès, Rimet, Frédéric, 2020. Diatom metabarcoding applied to large scale monitoring networks: Optimization of bioinformatics strategies using Mothur software. *Ecol. Indic.* 109, 105775. <https://doi.org/10.1016/j.ecolind.2019.105775>.

- Round, F.E., Crawford, R.M., Mann, D.G., 1990. *The diatoms: biology & morphology of the genera*. Cambridge University Press.
- Rovira, Laia, Trobajo, Rosa, Sato, Shinya, Ibáñez, Carles, Mann, David G., 2015. Genetic and Physiological Diversity in the Diatom *Nitzschia inconspicua*. *J. Eukaryot. Microbiol.* 62 (6), 815–832. <https://doi.org/10.1111/jeu.12240>.
- Sabater, Sergi, 1990. Composition and dynamics of a highly diverse diatom assemblage in a limestone stream. *Hydrobiologia* 190 (1), 43–53. <https://doi.org/10.1007/BF00020686>.
- Sakamoto, Y., Ishiguro, M., Kitagawa, G., 1986. *Akaike Information Criterion Statistics. Mathematics and its Applications*, Springer, Netherlands.
- Schloss, Patrick D., Westcott, Sarah L., Ryabin, Thomas, Hall, Justine R., Hartmann, Martin, Hollister, Emily B., Lesniewski, Ryan A., Oakley, Brian B., Parks, Donovan H., Robinson, Courtney J., Sahl, Jason W., Stres, Blaz, Thallinger, Gerhard G., Van Horn, David J., Weber, Carolyn F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75 (23), 7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
- Schneider, S.C., Hilt, S., Vermaat, J.E., Kelly, M., 2017. The “forgotten” ecology behind ecological status evaluation: re-assessing the roles of aquatic plants and benthic algae in ecosystem functioning. In: Cánovas, F.M., Lüttge, U., Matissek, R. (Eds.), *Progress in Botany*, Vol. 78. Progress in Botany. Springer International Publishing, Cham, pp. 285–304. https://doi.org/10.1007/124_2016_7.
- Serrana, Joeselle M., Miyake, Yo, Gamboa, Maribet, Watanabe, Kozo, 2019. Comparison of DNA metabarcoding and morphological identification for stream macroinvertebrate biodiversity assessment and monitoring. *Ecol. Indic.* 101, 963–972.
- Smith, Kirsty F., Kohli, Gurjeet S., Murray, Shauna A., Rhodes, Lesley L., 2017. Assessment of the metabarcoding approach for community analysis of benthic-epiphytic dinoflagellates using mock communities. *N. Z. J. Mar. Freshw. Res.* 51 (4), 555–576. <https://doi.org/10.1080/00288330.2017.1298632>.
- Stenger-Kovács, C., Körmendi, K., Lengyel, E., Abonyi, A., Hajnal, É., Szabó, B., Buczkó, K., Padišák, J., 2018. Expanding the trait-based concept of benthic diatoms: Development of trait- and species-based indices for conductivity as the master variable of ecological status in continental saline lakes. *Ecol. Indic.* 95, 63–74. <https://doi.org/10.1016/j.ecolind.2018.07.026>.
- Stenger-Kovács, Csilla, Lengyel, Edina, Sebestyén, Viktor, Szabó, Beáta, 2020. Effects of land use on streams: traditional and functional analyses of benthic diatoms. *Hydrobiologia* 847 (13), 2933–2946. <https://doi.org/10.1007/s10750-020-04294-y>.
- Stenger-Kovács, C., Padišák, J., Soróczki-Pintér, É., Ács, É., Borics, G., Buczkó, K., van Dam, H., 2006. The effect of hydro-morphological modifications of streamflow compositional features of attached diatom assemblages in Hungarian streams. *6th Int. Symp. Use Algae Monit. Rivers Hung. Balatonfüred* 12–16.
- Tanács, E., Belényesi, M., Lehoczki, R., Pataki, R., Petrik, O., Standovár, T., Pásztor, L., Laborczí, A., Szatmári, G., Zsolt, M., Bede-Fazekas, Á., Fodor, L.K., Varga, I., Zsembery, Z., Maucha, G., 2019. A national, high-resolution ecosystem basemap: methodology, validation, and possible uses. *Termvéd. Közlemények* 25, 34–58. <https://doi.org/10.17779/tvk-jnatconserv.2019.25.34>.
- Tapolczai, Kálmán, Bouchez, Agnès, Stenger-Kovács, Csilla, Padišák, Judit, Rimet, Frédéric, 2016. Trait-based ecological classifications for benthic algae: review and perspectives. *Hydrobiologia* 776 (1), 1–17. <https://doi.org/10.1007/s10750-016-2736-4>.
- 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., Vasselon, V., Bouchez, A., Stenger-Kovács, C., Padišák, J., Rimet, F., 2018. The impact of OTU sequence similarity threshold on diatom-based bioassessment: A case study of the rivers of Mayotte (France, Indian Ocean). *Ecol. Evol.* <https://doi.org/10.1002/ece3.4701>.
- Ter Braak, Cajo J.F., Barendregt, Leo G., 1986. Weighted averaging of species indicator values: Its efficiency in environmental calibration. *Math. Biosci.* 78 (1), 57–72. [https://doi.org/10.1016/0025-5564\(86\)90031-3](https://doi.org/10.1016/0025-5564(86)90031-3).
- U.S. Congress, 1972. Federal Water Pollution Control Act Amendments.
- Vasselon, Valentin, Domaizon, Isabelle, Rimet, Frédéric, Kahlert, Maria, Bouchez, Agnès, 2017a. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? *Freshw. Sci.* 36 (1), 162–177. <https://doi.org/10.1086/690649>.
- 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. Indic.* 82, 1–12. <https://doi.org/10.1016/j.ecolind.2017.06.024>.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S, Fourth Edition*. Springer, New York.
- Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. *Nature* 467, 555–561. <https://doi.org/10.1038/nature09440>.
- Walsh, G.W., Wepener, V., 2009. The influence of land use on water quality and diatom community structures in urban and agriculturally stressed rivers. *Water SA* 35. <https://doi.org/10.4314/wsa.v35i5.49184>.
- Weigand, H., Beermann, A.J., Ciampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A.M., Willassen, E., Wyler, S.A., Bouchez, A., Borja, A., Čiamporová-Zatovičová, Z., Ferreira, S., Dijkstra, K.-D.B., Eisendle, U., Freyhof, J., Gadawski, P., Graf, W., Haegerbaeumer, A., van der Hoorn, B.B., Japoshvili, B., Keresztes, L., Keskin, E., Leese, F., Macher, J.N., Mamos, T., Paz, G., Pešić, V., Pfannkuchen, D.M., Pfannkuchen, M.A., Price, B.W., Rinkevich, B., Teixeira, M.A.L., Várbíró, G., Ekrem, T., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Sci. Total Environ.* 678, 499–524. <https://doi.org/10.1016/j.scitotenv.2019.04.247>.
- Westcott, S.L., Schloss, P.D., 2015. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ* 3, e1487. <https://doi.org/10.7717/peerj.1487>.
- Wetzel, R.G., Likens, G.E., 2000. *Limnological analyses*. Verlag, New York. Inc NY 429.
- Zeiger, S.J., Hubbart, J.A., 2019. Characterizing Land Use Impacts on Channel Geomorphology and Streambed Sedimentological Characteristics. *Water* 11, 1088. <https://doi.org/10.3390/w11051088>.
- Zelinka, M., Marvan, P., 1961. Zur präzisierung der biologischen Klassifikation der Reinheit fließender Gewässer. *Arch Hydrobiol* 57, 389–407.
- 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>.