

DOCTORAL (PHD) DISSERTATION

**Revealing diatom diversity, rare taxa and their environmental drivers through
complementary eDNA metabarcoding and microscopic approaches**

Bassam Hussein Ibrahim Abubaker

Doctoral School of Chemistry and Environmental Sciences

Supervisors:

Prof. Dr. Csilla Stenger-Kovács and Dr. Kálmán Tapolczai

DOI:10.18136/PE.2026.1010

University of Pannonia

Research Group of Limnology



Veszprém

2026

**Revealing diatom diversity, rare taxa and their environmental drivers through
complementary eDNA metabarcoding and microscopic approaches**

The thesis was prepared for the award of a doctoral degree (PhD) within the framework of the
Doctoral School of Chemistry and Environmental Sciences at University of Pannonia

in the discipline of **Environmental Sciences**

written by: **Bassam Hussein Ibrahim Abubaker**

Supervisors: **Prof. Dr. Csilla Stenger-Kovács and Dr. Kálmán Tapolczai**

I recommend the dissertation for acceptance: yes / no.

.....
supervisor

I recommend the dissertation for acceptance: yes / no.

.....
supervisor

I recommend the dissertation for peer review.

.....
chair of the DDHC

The PhD-candidate has achieved % at the public debate.

The composition of the Final Examination Committee:

chair:.....

reviewers:.....

members:.....

Veszprém,

.....
chair of the committee

Qualification of degree:

Veszprém,

.....
chair of the UDHC

Glossary of Key Terms

Term	Definition
Alpha diversity (α -diversity)	is the measure of species diversity (the number of different species and their relative abundances) within a single, specific area, habitat, or biological sample.
ASVs (Amplicon Sequence Variants)	Unique DNA sequences inferred from high-throughput sequencing reads that represent distinct biological entities, often corresponding to species or sub-species level taxa. They offer finer taxonomic resolution than traditional OTUs.
Beta Diversity	A measure of the differentiation of species composition among different sites or habitats within a region. It quantifies the turnover of species along environmental gradients or between communities.
Bioinformatic Pipeline	A series of computational steps and software tools used to process raw DNA sequencing data, including quality filtering, ASV inference, taxonomic assignment, and statistical analysis.
Biomonitoring	The use of biological organisms or their responses to assess environmental quality, particularly in aquatic ecosystems. Diatoms are one of the key organisms in freshwater biomonitoring.
Community Structure	The composition and relative abundance of different species within a biological community. It can be analyzed using various ecological metrics and ordination techniques like Non-metric Multidimensional Scaling (NMDS).
Cryptic Taxa	Species that are morphologically indistinguishable but genetically distinct. eDNA metabarcoding is particularly effective at detecting these hidden species.
Diatoms	A highly diverse group of microscopic, single-celled algae characterized by unique silica cell walls (frustules). They are primary producers in aquatic ecosystems and widely used as bioindicators.
Ecological Status	A classification of the quality of an aquatic ecosystem based on biological, hydromorphological, physical, and chemical elements. Diatoms are used to determine the ecological status of water bodies under directives like the WFD.

eDNA Metabarcoding	A molecular technique that involves extracting DNA from environmental samples (e.g., water, sediment) and then amplifying and sequencing specific genetic markers (barcodes) to identify multiple species present in a community.
Environmental Gradients	Gradual changes in environmental conditions (e.g., pH, nutrient concentration, conductivity) across a landscape or within a water body, which often influence species distribution and community composition.
High-throughput Sequencing	Also known as Next-Generation Sequencing (NGS), this technology allows for the rapid and cost-effective sequencing of millions of DNA fragments simultaneously, enabling comprehensive biodiversity assessments.
Morphological Analysis	The study of the form and structure of organisms. In diatoms, this involves examining the intricate patterns of their silica frustules under a microscope for identification.
Phytobenthos	The community of photosynthetic organisms (primarily algae, including diatoms) that live attached to or within the substrate of aquatic ecosystems. It is a key component of aquatic food webs.
Rare Species	Species that occur infrequently, in low abundance, or in unique habitats. Their ecological role and contribution to biodiversity are increasingly recognized as important for ecosystem functioning and resilience.
Rarity Types	A classification of rare species based on multiple dimensions, such as population size, habitat specificity, and geographic range. This framework helps to understand the ecological significance and conservation needs of rare taxa.
Reference Libraries	Databases containing DNA sequences of known, identified organisms. These libraries are crucial for assigning taxonomic names to ASVs generated through metabarcoding.
Relative Abundance	The proportional representation of a particular taxon within a biological sample or community, often expressed as a percentage or fraction of the total count or sequence reads
Water Framework Directive (WFD)	A European Union directive (2000/60/EC) that commits member states to achieve good qualitative and quantitative status of all water bodies. Diatoms are one of a key biological quality element for assessing ecological status under the WFD.

List of Figures

Figure 1. Diatom DNA metabarcoding workflow.....	33
Figure 2. Cumulative number of publications about diatom eDNA metabarcoding based on Web of Science data (3rd of March 2025).	34
Figure 3. Global distribution of diatom eDNA metabarcoding studies showing the number of publications.	35
Figure 4 Overview of surveyed publications: (A) habitats, (B) sample types, (C) target gene/region, and (D) DNA eextraction methods. Counts are shown as horizontal bars and scaled to the number of publications per category.	36
Figure 5. Sequencing platforms used in diatom DNA metabarcoding studies.....	37
Figure 6. Bioinformatic pipelines are frequently used in diatom DNA metabarcoding studies.	38
Figure 7. The investigated 100 sampling sites of the River Danube network system in Hungary and Croatia	58
Figure 8. (a) Trends in the number of rare species as the number of species increases in the samples (b) the rare species number in the genera with different species richness	61
Figure 9. The proportion of rare and common diatom species per sample	62
Figure 10. (a) Violin plots of rarity indices, (b) species number in the different rarity categories and in their combinations (PSI=population size index, HSI=habitat specificity index, GRI=geographical range index, RR=median of GRI, HSI and PSI; SP=small population species, HS=habitat specialists, GR=geographically restricted species)	62
Figure 11. Sørensen dissimilarities of the subcommunities of (a) small population species (SP); (b) habitat specialist species (HS); (c) geographically restricted species (GR) and (d) absolutely rare taxa (AR) compared to common species'(C). The contribution of (e) SP, (f) HS, (g) GR and (h) AR to the total community variability compared to common species'(C).....	63
Figure 12. Beta diversity and its components of the subcommunities of (a) small population species (SP); (b) habitat specialist species (HS), geographically restricted species (GR) and absolutely rare taxa (AR) compared to c-d) their complementary common species (C).....	64
Figure 13. Distribution of rare (a) small population species (SP); (b) habitat specialist species (HS), geographically restricted species (GR) and absolutely rare taxa (AR) and c-d) their complementary common species'(C) in the benthic diatom community networks. (Green: rare, red: common species; The code of the species can be found in the Supplement Table 1)	65
Figure 14. The location of the sampling sites (red dots) in the catchment areas of the rivers Raba and Marcal in western Hungary.	76
Figure 15. Proportion (%) of the successfully assigned ASVs (blue bars) and reads (green bars) to the different taxonomic levels.	82
Figure 16. Number of taxa detected by the three different identification methods (A) at the genus taxonomy level, and at the species taxonomy level (B).....	83
Figure 17. Taxa richness (A) and Shannon diversity (B) across the three methods: diatom microscopy, phytobenthos microscopy and diatom metabarcoding.	83
Figure 18. Distribution of correlation coefficients between morphological and metabarcoding abundance values for common diatom species.	84
Figure 19. Principal component analysis biplots showing the distribution of samples from the two sampling seasons (2018 - red dots, 2019 - blue dots) based on the environmental variables (blue arrows), presenting axes PC1 and PC2 (A), and PC1 and PC3 (B).....	85

Figure 20. Relationships between diversity metrics and PC1 for the phytobenthos based on microscopy (A, B), diatoms based on microscopy (C, D), and diatom ASVs based on metabarcoding (E, F). Figures show the quadratic fit (blue line) with 95% confidence intervals (shaded area)	87
Figure 21. Co-inertia analysis biplots illustrate the covariation between environmental variables and community composition for phytobenthos based on microscopy (A), diatoms based on microscopy (B) and diatom ASVs based on metabarcoding (C).	89

Table of Contents

Abstract	10
Absztrakt	12
الخلاصة	15
1. General introduction	17
1.1. General characteristics, taxonomy, physiology, and ecology of diatoms	17
1.2. Diatoms as indicators of ecological conditions	18
1.3. The Water Framework Directive and freshwater biomonitoring	19
1.4. Diatom-based ecological indices	20
1.5. Microscopy-based analysis of phytobenthos and diatoms	21
1.6. Environmental DNA metabarcoding of diatoms	23
1.7. Barcode markers used in diatom metabarcoding	24
1.8. Complementarity of microscopy and DNA metabarcoding	26
1.9. Rare diatoms and the neglected component of biodiversity	27
1.10. Knowledge gaps and dissertation rationale	29
2. Main objectives	30
3. Environmental DNA Metabarcoding of Diatoms: A Next-Generation Tool for Aquatic Biodiversity and Bioassessment	31
3.1. Introduction	32
3.2. Methods	33
3.3. Results	35
3.3.1. Global geographical distribution of diatom eDNA studies	35
3.3.2. Diatom eDNA metabarcoding workflow	36
3.4. Discussion	38
3.4.1. Sample collection and preservation	38
3.4.2. DNA extraction	41
3.4.3. Amplicon marker selection	42
3.4.4. High-throughput sequencing of amplicons	44
3.4.5. Bioinformatic pipelines and reference libraries	45
3.4.6. Taxa inventory correspondence between eDNA metabarcoding and morphological analyses	46
3.4.7. Diatom eDNA metabarcoding as a bio-assessment tool	49
3.4.8. Diatom eDNA metabarcoding as a forensic tool	50
3.6. Conclusion	52
4. Role of rare species in benthic diatom communities: patterns, processes and networks	54
4.1. Introduction	55

4.2. Materials and Methods	57
4.2.1 River network system, sampling and analyses of benthic diatoms	57
4.2.2. Classification of rarity and commonness	59
4.2.3. Statistical analyses	59
4.3. Results	60
4.3.1. Types of rarity in benthic diatom communities	60
4.3.2. Role of rare species in the variability of communities	62
4.3.3. Beta diversity of rare and common species	64
4.3.4. Network analyses of diatom communities	65
4.4. Discussion	66
4.4.1. Rarity types and patterns of benthic diatoms in river networks	66
4.4.2. Community variability	68
4.4.3. Beta-diversity in benthic diatom metacommunities	69
4.4.4. Community networks	70
4.5. Conclusions and future perspectives	71
5. Using complementary approaches to study diatom diversity and its environmental drivers	72
5.1. Introduction	73
5.2. Materials and Methods	75
5.2.1 Sampling sites	75
5.2.2 Environmental variables	76
5.2.3 Diatom sampling and laboratory procedure	76
5.2.3.1. Microscopy analysis of phytobenthos	77
5.2.3.2. Microscopy analysis of benthic diatoms	77
5.2.3.3. Molecular analysis of benthic diatoms	77
5.2.4. Bioinformatic pipeline	78
5.2.5. Statistical analysis	79
5.2.5.1. Data preparation	79
5.2.5.2. Comparison of biological inventories and diversity metrics	79
5.2.5.3. Abundance relationships between microscopy and metabarcoding	79
5.2.5.4. Environmental gradients	80
5.2.5.5. Co-inertia analysis of community–environment relationships	80
5.3. Results	80
5.3.1. Biological inventories	80
5.3.1.1. Morphological identification of the phytobenthos	80
5.3.1.2. Morphological identification of diatoms	81
5.3.1.3. eDNA metabarcoding of diatoms	81

5.3.1.4. Shared and unique taxa among the biological inventories	82
5.3.2. Diversity metrics compared among identification methods	83
5.3.3. Diatom species abundance analysis	84
5.3.4. Abiotic environment and its effect on algal communities	85
5.3.4.1. Environmental gradients	85
5.3.4.2. Response of community diversity metrics to environmental constraints	86
5.3.4.3 Response of community structure to environmental constraints	88
5.4. Discussion	89
5.4.1 Diversity metrics of the three biological inventories	89
5.4.2 Concordance and discrepancies in the estimation of species abundance	90
5.4.3 Environmental gradients shaping algal and diatom communities	91
5.4.4 Community–environment co-structure and implications for biomonitoring	92
5.5. Conclusion	93
Personal contribution to the conducted research	96
Dedication	99
Acknowledgment	100
References	101
Appendixes	122
Publication list	138
Bassam Abubaker CV	139

Abstract

Diatoms are one of the key organisms in freshwater biomonitoring. Light microscopy ensures continuity with established taxonomic frameworks and assessment indices, but it is constrained by counting effort and by the proportion of the community that remains observable after the slide's preparation. In contrast, DNA metabarcoding captures whole assemblages directly from environmental samples, offering greater sensitivity to low-abundance taxa and higher taxonomic resolution when amplicon sequence variants (ASVs) are retained.

The overall aim of this dissertation was to determine how complementary microscopic and molecular approaches can improve our understanding of diatom diversity, rarity patterns, and their relationships with environmental conditions. Specifically, the dissertation aimed to: (i) synthesize the current applications, methodological development, and limitations of diatom DNA metabarcoding; (ii) classify rare benthic diatom taxa and evaluate their contribution to community variability, beta diversity, and ecological network structure using microscopy-based data from a river network; and (iii) compare phytobenthos microscopy, diatom microscopy, and diatom metabarcoding in their ability to characterize biological inventories, diversity patterns, abundance relationships, and community responses to environmental gradients.

A systematic search of the Web of Science identified 128 papers on diatom DNA metabarcoding. The first chapter showed that each step of the DNA metabarcoding workflow, from sampling and DNA extraction to PCR, barcode selection, sequencing, and bioinformatic processing, can influence the final outcome. It also highlighted persistent limitations, including incomplete reference libraries, unresolved issues in abundance quantification, and variation among bioinformatic pipelines. Even so, diatom DNA metabarcoding has rapidly developed into an efficient approach that strengthens biomonitoring and broadens ecological interpretation.

The basic knowledge about rarity is missing in aquatic algal ecology, which could be the base of their conservation strategies. Rarity in benthic diatom communities was examined across a Danube River network in Hungary and Croatia using three dimensions: population size, habitat specificity, and geographic range. We found that rare and common species are of similar quantities on a regional scale. Genera with fewer taxa provided more rare taxa. The majority of the rare diatom species had small population size among which 45–47% were habitat specialists and geographically restricted, locating streams and river sections whose

protection would be outstandingly important. Rare diatom species provided 25–50% higher variability of the communities than common ones. Beta-diversity was also primarily determined by rare species and was mostly driven by species replacement especially in case of specialized and localized rare taxa.

A three-way comparison of phytobenthos microscopy, diatom microscopy, and diatom metabarcoding in the Marcal–Rába catchment of western Hungary used 130 samples from 67 sites to evaluate species inventories, richness, Shannon diversity, abundance rank agreement, and community–environment relationships. Metabarcoding produced the highest per-sample diatom richness and diversity, whereas both microscopy-based approaches yielded lower estimates. Phytobenthos microscopy captured the widest algal spectrum overall, but showed the lowest richness and diversity. Species overlap among methods was limited, and abundance correlations between microscopy and molecular data ranged from weak to strongly positive. Shannon diversity based on diatom ASVs showed a unimodal response along the main environmental gradient, driven largely by agricultural land use, conductivity, and nutrient enrichment. Co-inertia analysis indicated the strongest community–environment coupling for metabarcoding, followed by diatom microscopy and phytobenthos microscopy.

Overall, the findings indicate that microscopy remains essential for preserving taxonomic consistency and providing morphological evidence, while ASV-based metabarcoding enhances the detection of rare and cryptic taxa and reveals environmentally driven community patterns with greater sensitivity. Its broader implementation, however, will require further improvement of reference libraries and the adoption of more standardized, robust, and reproducible analytical workflows.

Absztrakt

A kovamoszatok (diatómák) az édesvízi biomonitorozás meghatározó szervezetei közé tartoznak. A fénymikroszkópos vizsgálatok biztosítják a kapcsolatot a hagyományos taxonómiai rendszerekkel és a már kialakított ökológiai állapotértékelési indexekkel, ugyanakkor az eredményeket korlátozza a számlálási ráfordítás, valamint az, hogy a tárgylemez előkészítése után a közösségnek csak egy része marad megfigyelhető. Ezzel szemben a DNS-metabarkódolás közvetlenül a környezeti mintákból képes feltárni a teljes közösséget, ezért érzékenyebben mutatja ki az alacsony abundanciájú taxonokat, és az amplikonszekvencia-variánsok (ASV-k) megtartásával nagyobb taxonómiai felbontást tesz lehetővé.

A disszertáció átfogó célja annak meghatározása volt, hogy a mikroszkópos és molekuláris módszerek egymást kiegészítő alkalmazása miként javíthatja a diatómaközösségek diverzitásáról, a ritkasági mintázatokról, valamint ezek környezeti tényezőkkel való kapcsolatáról alkotott ismereteinket. Ennek érdekében a disszertáció három fő célkitűzést fogalmazott meg: (i) a diatómák DNS-metabarkódolásának jelenlegi alkalmazásait, módszertani fejlődését és korlátait összegző szintézis elkészítését; (ii) a ritka bentikus diatómataxonok osztályozását, valamint a közösségi variabilitáshoz, a béta-diverzitáshoz és az ökológiai hálózatok szerkezetéhez való hozzájárulásuk értékelését egy folyóhálózat mikroszkópos adatai alapján; továbbá (iii) a fitobentosz-mikroszkópia, a diatóma-mikroszkópia és a diatóma DNS-metabarkódolás összehasonlítását a taxonlisták, a diverzitási mintázatok, az abundanciaviszonyok és a környezeti gradiensekre adott közösségi válaszok jellemzésében.

A Web of Science adatbázisban végzett szisztematikus kereséssel 128, diatómák DNS-metabarkódolásával foglalkozó publikációt azonosítottunk. Az első fejezetben bemutattuk, hogy a DNS-metabarkódolási munkafolyamat minden egyes lépése – a mintavételtől és a DNS-kivonástól kezdve a PCR-en, a genetikai marker kiválasztásán és a szekvenáláson át egészen a bioinformatikai feldolgozásig – befolyásolhatja a végső eredményeket. Az elemzés tartós módszertani korlátokat is feltárt, különösen a referencia-adatbázisok hiányosságait, az abundancia kvantitatív meghatározásával kapcsolatos bizonytalanságokat, valamint a különböző bioinformatikai feldolgozási eljárások közötti eltéréseket. Mindezek ellenére a diatómák DNS-metabarkódolása gyors fejlődésen ment keresztül, és olyan hatékony megközelítéssé vált, amely erősíti a biomonitorozást és szélesebb lehetőségeket kínál az ökológiai mintázatok értelmezésére.

A ritkaság alapvető ökológiai jellemzőiről továbbra is kevés ismerettel rendelkezünk a vízi algaközösségekben, holott ezek az ismeretek fontos alapot teremthetnek a természetvédelmi stratégiák kidolgozásához. A magyarországi és horvátországi Duna-vízgyűjtő folyóhálózat bentikus diatómaközösségeiben három dimenzió – a populációméret, az élőhely-specializáció és a földrajzi elterjedés – alapján vizsgáltuk a ritkaságot. Eredményeink szerint regionális léptékben a ritka és a gyakori fajok száma hasonló volt. Kimutattuk továbbá, hogy a kisebb fajszerű nemzetségek arányaiban több ritka fajt tartalmaztak. A ritka diatómafajok többségét kis populációméret jellemezte, és ezek 45–47%-a egyben élőhely-specialistának és földrajzilag korlátozott elterjedésűnek bizonyult. E fajok előfordulási mintázatai olyan patakokat és folyószakaszokat jelöltek ki, amelyek természetvédelmi szempontból kiemelt jelentőségűek lehetnek. A ritka fajok 25–50%-kal nagyobb mértékben járultak hozzá a közösségek variabilitásához, mint a gyakori fajok. A béta-diverzitás alakulását szintén elsősorban a ritka fajok határozták meg, és annak fő komponensét a fajkicserélődés adta, különösen a specialista és földrajzilag korlátozott ritka taxonok esetében.

A Nyugat-Magyarországon található Marcal–Rába vízgyűjtő 67 mintavételi helyéről származó 130 minta alapján három módszert – a fitobentosz-mikroszkópiát, a diatóma-mikroszkópiát és a diatóma DNS-metabarkódolást – hasonlítottunk össze. Az összehasonlítás során a taxonlistákat, a fajgazdagságot, a Shannon-diverzitást, az abundancia-rangsorok egyezését, valamint a közösségek és a környezeti változók közötti kapcsolatokat értékeltük. A DNS-metabarkódolás mintánként a legnagyobb diatóma-fajgazdagságot és diverzitást tárta fel, míg mindkét mikroszkópos módszer alacsonyabb értékeket eredményezett. A fitobentosz-mikroszkópia összességében a legszélesebb algataxon-spektrumot mutatta ki, ugyanakkor ez a módszer jelezte a legalacsonyabb fajgazdagságot és diverzitást. A három módszer által kimutatott fajok átfedése korlátozott volt, míg a mikroszkópos és molekuláris adatok közötti abundanciakorrelációk a gyengétől az erősen pozitívig terjedtek. A diatóma-ASV-k alapján számított Shannon-diverzitás unimodális választ mutatott a fő környezeti gradiens mentén, amelyet elsősorban a mezőgazdasági területhasználat, a vezetőképesség és a tápanyagterhelés alakított. A koinercia-elemzés a DNS-metabarkódolás esetében mutatta ki a legerősebb kapcsolatot a közösségösszetétel és a környezeti változók között; ezt a diatóma-mikroszkópia, majd a fitobentosz-mikroszkópia követte.

Összességében eredményeink azt mutatják, hogy a mikroszkópos vizsgálatok továbbra is nélkülözhetetlenek a taxonómiai következetesség fenntartásához és a morfológiai

bizonyítékok biztosításához, míg az ASV-alapú DNS-metabarkódolás érzékenyebbé teszi a ritka és kriptikus taxonok kimutatását, valamint részletesebben tárja fel a környezeti tényezők által meghatározott közösségi mintázatokat. A módszer szélesebb körű alkalmazásához ugyanakkor tovább kell fejleszteni a referencia-adatbázisokat, valamint egységesebb, robusztusabb és reprodukálhatóbb analitikai munkafolyamatokat kell kialakítani és alkalmazni.

الخلاصة

تُعدّ الدياتومات من المجموعات المحورية في الرصد الحيوي للأنظمة المائية العذبة، إلا أن الطريقة المستخدمة في تقييمها تؤثر بصورة كبيرة في الأنماط التي يتم الكشف عنها وفي الاستنتاجات التي يمكن التوصل إليها. يوفّر الفحص بالمجهر الضوئي طريقة تقليدية في الفحص مع الأطر التصنيفية المعتمدة ومؤشرات التقييم المختلفة، لكنه يبقى مقيداً بجهد العدّ وبالجزء من المجتمع الحي الذي يظل قابلاً للملاحظة بعد عمليات اعداد وتحضير الشرائح. في المقابل، يتيح الترميز الشريطي الأيضي المعتمد على الحمض النووي (DNA) توصيف المجتمعات كاملة مباشرة من العينات البيئية، مع حساسية أعلى للكشف عن الأصناف النادرة ودقة تصنيفية أكبر عند الاعتماد على متغيرات التسلسل (ASVs).

تمثل الهدف العام لهذه الأطروحة في تحديد الكيفية التي يمكن من خلالها للتكامل التكميلي بين الأساليب المجهرية والجزيئية أن يعزز فهم تنوع الدياتومات، وأنماط ندرتها، وعلاقاتها بالظروف البيئية. وعلى وجه التحديد، هدفت الأطروحة إلى: (1) تقديم تحليل تركيبى للتطبيقات الحالية والتطورات المنهجية والقيود المرتبطة بالترميز الشريطي للحمض النووي لدى الدياتومات؛ (2) تصنيف أصناف الدياتومات النادرة وتقييم مساهمتها في تباين المجتمعات، وبنية الشبكات البيئية، بالاعتماد على بيانات مجهرية من شبكة نهريّة؛ و(3) مقارنة مجهرية للدياتومات والترميز الشريطي الفوقي للدياتومات من حيث قدرتها على توصيف القوائم التصنيفية، وأنماط التنوع، والعلاقات بين الوفرة، واستجابات المجتمعات للتدرجات البيئية.

وقد حدّد بحثٌ منهجي في قاعدة بيانات Web of Science عدد 128 دراسة تناولت الترميز الشريطي الأيضي للدياتومات. وأظهرت هذه المراجعة أن كل خطوة في سير العمل، ابتداءً من أخذ العينات واستخلاص الحمض النووي، مروراً بتفاعل PCR واختيار القطعة الجينية والتسلسل، وانتهاءً بالمعالجة المعلوماتية الحيوية، يمكن أن تؤثر في النتيجة النهائية. كما أبرزت المراجعة عدداً من التحديات المستمرة، من بينها عدم اكتمال قواعد البيانات المرجعية، واستمرار عدم اليقين في تقدير الوفرة النسبية، والتباين بين مسارات التحليل المعلوماتي الحيوي. ومع ذلك، فقد تطور الترميز الشريطي الأيضي للدياتومات المعتمد على الحمض النووي البيئي بسرعة ليصبح أداة فعالة تعزز الرصد الحيوي وتوسع من إمكانيات التفسير البيئي.

وقد دُرست الندرة في مجتمعات الدياتومات القاعية عبر شبكة نهريّة ضمن حوض الدانوب في المجر وكرواتيا باستخدام ثلاثة أبعاد هي: حجم الجماعة، وتخصص الموطن، والنطاق الجغرافي. وعلى المستوى الإقليمي، ظهرت الأنواع النادرة والشائعة بأعداد متقاربة، في حين ساهمت الأجناس الأقل غنى بالأنواع بنسبة أعلى من الأنواع النادرة. واتّسمت معظم الأصناف النادرة بانخفاض حجم الجماعة، كما أن 45-47% منها كانت متخصصة في موائل محددة وذات انتشار جغرافي محدود، مما يشير إلى وجود مقاطع نهريّة ذات قيمة حفظية مرتفعة. وقد زادت الأصناف النادرة من تباين المجتمع بنسبة تراوحت بين 25 و50% مقارنةً بالأصناف الشائعة، وكان تنوع بيتا مدفوعاً أساساً بإحلال الأنواع النادرة، ولا سيما الأصناف المتخصصة وضيقة الانتشار. كما أشارت تحليلات الشبكات إلى أن بعض الأنواع النادرة قد تشغل أدواراً بنيوية مهمة، رغم أن مفهوم الندرة لا يزال حساساً لعدة مصادر من عدم اليقين.

كما أُجريت مقارنة ثلاثية بين مجهرية الفيتوبنتوس، ومجهرية الدياتومات، والترميز الشريطي الأيضي للدياتومات في حوض ماركال-رابا في غرب المجر، باستخدام 130 عينة من 67 موقعاً، بهدف مقارنة القوائم النوعية، وغنى الأنواع،

وتنوع شانون، ومدى توافق رتب الوفرة، والعلاقات بين المجتمع والعوامل البيئية. وأعطى الترميز الشريطي الأيضي أعلى قيم لغنى الدياتومات وتنوعها على مستوى العينة، في حين أعطت الطريقتان المجهريتان قيماً أقل. أما مجهرية الفيتوبنتوس فقد غطت أوسع طيف من الطحالب إجمالاً، لكنها أظهرت أدنى مستويات الغنى والتنوع. وكان التداخل النوعي بين الطرق محدوداً، في حين تراوحت الارتباطات بين الوفرة المجهرية والوفرة الجزيئية من ضعيفة إلى إيجابية قوية. وأظهر تنوع شانون المبني على ASVs في الدياتومات استجابةً أحادية القمة على طول التدرج البيئي الرئيسي، الذي تحدد بصورة كبيرة باستعمال الأراضي الزراعية، والتوصيلية الكهربائية، والإثراء بالمغذيات. كما أظهر تحليل التباين المشترك أن أقوى ارتباط بين بنية المجتمع والعوامل البيئية كان في بيانات الترميز الشريطي الأيضي، تليه مجهرية الدياتومات ثم مجهرية الفيتوبنتوس.

وبصورة عامة، تُظهر هذه النتائج أن الفحص المجهرى لا يزال أداة لا غنى عنها من أجل الاستمرارية التصنيفية والتفسير المورفولوجي، في حين أن الترميز الشريطي الأيضي المعتمد على ASVs يحسن القدرة على الكشف عن التنوع النادر والخفي، ويعزز التعرف على الأنماط التي تُشكلها العوامل البيئية. غير أن توسيع تطبيقه على نطاق أوسع سيعتمد على استمرار تطوير قواعد البيانات المرجعية، وعلى الوصول إلى معايير تحليلية أكثر متانة وقابليةً للتكرار.

1. General introduction

1.1. General characteristics, taxonomy, physiology, and ecology of diatoms

Diatoms are a highly diverse group of unicellular or colonial photosynthetic microorganisms belonging to the stramenopile lineage. Their defining feature is the frustule, a siliceous cell wall composed of two overlapping parts, known as valves. The shape, size, and ornamentation of these valves, including the arrangement of striae, pores, and raphe structures, provide the principal characteristics used in conventional taxonomic identification. Diatoms have traditionally been classified into broad morphological groups, particularly centric and pennate forms, although advances in molecular systematics have revealed evolutionary relationships that are not always apparent from morphology alone (Mann 2010).

Diatoms occur in almost all aquatic environments, including rivers, streams, lakes, wetlands, estuaries, seas, and oceans. They inhabit both littoral and pelagic zones and may occur as planktonic organisms in the water column or as benthic communities attached to stones, sediments, macrophytes, and other submerged surfaces (Malviya et al. 2016; Smol and Stoermer 2010). In running waters, benthic diatoms are major components of periphyton and biofilm communities, where they interact with bacteria, fungi, other algae, and small invertebrates.

Estimates of global diatom diversity vary considerably. Approximately 12,000 species have been formally described, whereas the actual number may range from 30,000 to 100,000 species (Guiry 2012; Mann and Vanormelingen 2013). This difference reflects the large number of poorly studied habitats, regional taxonomic gaps, cryptic diversity, and the continuing revision of species boundaries. Many morphologically similar populations previously treated as a single species may contain several genetically and ecologically distinct lineages. Conversely, some taxa described on the basis of subtle morphological differences may subsequently be reconsidered when additional molecular and ecological evidence becomes available. Diatom taxonomy is therefore dynamic, and species concepts continue to change as microscopy and molecular methods improve.

The physiology and life cycle of diatoms are closely connected to their ecological success. As photosynthetic primary producers, they convert inorganic carbon into organic matter and contribute to oxygen production, nutrient cycling, and aquatic food webs. Their

siliceous frustules also contribute to sediment formation, while benthic diatom biofilms can promote sediment stability and provide food and habitat for other organisms (B-Béres et al. 2023). Diatoms require dissolved silica for frustule formation, and their growth is also influenced by the availability of nitrogen, phosphorus, carbon, light, and other resources.

A characteristic feature of the diatom life cycle is the gradual reduction in average cell size that generally occurs during repeated vegetative division. Cell size is restored through sexual reproduction and auxospore formation, thereby maintaining viable populations over successive generations (Mann 2010). Variation in cell size and morphology can affect physiological performance, ecological preferences, and the amount of DNA contributed by individual cells. These characteristics are also relevant when comparing microscopy-based abundance estimates with sequence-read proportions generated by molecular methods.

Diatom species differ markedly in their environmental requirements. Their distributions are influenced by nutrient concentrations, pH, conductivity, salinity, temperature, oxygen availability, current velocity, light, substrate type, and hydromorphological conditions. Some species tolerate broad environmental ranges, whereas others occur within relatively narrow ecological limits. Consequently, changes in community composition can reflect changes in local environmental conditions over ecologically meaningful timescales (Desrosiers et al. 2013). These properties explain why diatoms have become central to freshwater ecological research and biomonitoring.

1.2. Diatoms as indicators of ecological conditions

Diatoms occupy a particularly important position among the biological groups used to assess the condition of freshwater ecosystems. Their short generation times allow communities to respond rapidly to changes in environmental conditions, while their species-specific ecological preferences permit relatively detailed interpretation of water quality. Their usefulness in rivers, streams, and lakes has been demonstrated across a wide range of geographical and ecological settings (Almeida et al. 2014; Lavoie et al. 2014; Rimet 2012; Rimet and Bouchez 2012; Schulte et al. 2024; Smol and Stoermer 2010; Stenger-Kovács et al. 2007; Stephen and Joshi 2010).

Individual diatom taxa differ in their sensitivity to nutrient enrichment, organic pollution, acidity, salinity, toxic substances, and hydrological alteration. Changes in their occurrence and relative abundance can therefore indicate both the nature and intensity of

environmental pressures. Species associated with nutrient-poor, well-oxygenated habitats may decline as eutrophication increases, whereas pollution-tolerant taxa may become more frequent or dominant. Community-level changes can consequently provide more information than measurements based on a single environmental variable, particularly when water chemistry varies over short periods.

Diatom assemblages are also useful because they integrate environmental conditions over the period during which a community develops. A chemical measurement describes water quality at the moment of sampling, whereas the biological community reflects the combined effect of environmental conditions experienced over a longer interval. This does not mean that biological and chemical assessments are interchangeable; rather, they provide complementary information about ecological condition.

The indicator value of diatoms extends beyond contemporary monitoring. Their siliceous valves may remain preserved in sediments, allowing past environmental conditions to be reconstructed. Changes in fossil diatom assemblages have therefore been used to investigate historical variation in nutrient conditions, acidity, salinity, water level, and climate. In both present-day monitoring and palaeoecological research, reliable interpretation depends on accurate identification and adequate ecological information for the recorded taxa.

1.3. The Water Framework Directive and freshwater biomonitoring

The European Water Framework Directive, Directive 2000/60/EC, established a common framework for the protection and management of European waters. One of its central objectives is to prevent further deterioration and to achieve at least good ecological and chemical status in surface waters. Ecological status is assessed by comparing observed conditions with those expected under reference or minimally disturbed conditions (EC Directive 2000).

The Water Framework Directive classifies ecological status into five categories: high, good, moderate, poor, and bad. The assessment is based primarily on biological quality elements, supported by hydromorphological and physicochemical information. Biological quality elements include phytoplankton, aquatic flora, benthic invertebrates, and fish, although their application differs among water-body categories. In rivers, diatoms are generally assessed as an important component of the phytobenthos biological quality element.

Diatoms are particularly suitable for this purpose because their communities respond predictably to several pressures considered under the Water Framework Directive, including eutrophication, organic enrichment, acidification, salinization, and changes in habitat conditions. Their widespread occurrence also allows comparable methods to be applied across many river types. In some situations, diatom assemblages alone provide ecological information comparable to that obtained from the wider phyto-benthos community, supporting their routine use in ecological status assessment (Kelly 1998; Kelly et al. 2009).

The Water Framework Directive does not prescribe a single diatom assessment method for all European countries. National systems have been developed according to regional species pools, river typologies, reference conditions, and management requirements. Intercalibration is therefore necessary to ensure that national classifications reflect comparable levels of ecological degradation. Differences in taxonomic practice, index composition, species nomenclature, and ecological classifications can nevertheless complicate comparisons among countries and monitoring programmes (Almeida et al. 2014; Kahlert et al. 2009).

Reliable ecological classification depends on the quality of the biological data used to calculate assessment metrics. Taxonomic inconsistencies, insufficient counting effort, or failure to detect ecologically informative species can influence index values and, in some cases, the final status class assigned to a water body. For this reason, methodological standardization and taxonomic harmonization are essential components of diatom-based monitoring.

1.4. Diatom-based ecological indices

Diatom indices convert information on species composition and relative abundance into numerical estimates of ecological condition. Most indices are based on the principle that species differ in their ecological optima, pollution sensitivity, and indicator value. Taxa with narrow environmental tolerances generally receive greater weight than broadly tolerant or poorly characterized species. The calculated index therefore represents the combined ecological information provided by the species within a sample.

Several diatom indices are routinely used in European monitoring. The Specific Pollution Sensitivity Index, or IPS, is one of the most widely applied and incorporates the relative abundance, pollution sensitivity, and indicator value of individual taxa (Descy and Coste 1991). Other examples include the Trophic Diatom Index, developed primarily to assess

nutrient enrichment (Kelly and Whitton 1995), the Biological Diatom Index used in France (Prygiel et al. 2002), and the Swiss Diatom Index, which was developed for the assessment of Swiss running waters (Hürlimann et al. 2007).

These indices have enabled diatom data to be incorporated into reproducible and operational assessment systems. However, their accuracy depends on several factors. Species must be correctly identified, their nomenclature must correspond to that used in the index, and reliable ecological values must be available. Recently described species, cryptic taxa, and revised species complexes may not be adequately represented in existing indices. In addition, ecological preferences may vary geographically, and a taxon classified as tolerant in one region may respond differently elsewhere.

Relative abundance is another important component of most conventional diatom indices. Microscopy-based assessments usually calculate abundance from the number of valves counted on a permanent slide. Molecular assessments, by contrast, frequently use the proportion of sequence reads assigned to each taxon. These two measures are not directly equivalent because read abundance may be affected by cell size, gene copy number, DNA extraction efficiency, primer affinity, and PCR amplification. Consequently, molecular and morphology-based index values may be significantly correlated while still producing different numerical scores or ecological classifications (Bailet et al. 2019; Pérez-Burillo et al. 2020; Tapolczai et al. 2024).

Molecular data also permit assessment approaches that do not require formal taxonomic assignment. In taxonomy-free methods, individual sequence variants are assigned ecological profiles directly from their relationships with environmental conditions. Such approaches retain information from unclassified sequences and from genetically distinct variants within a morphologically defined species. Taxonomy-free indices have produced promising results and may reduce some of the limitations caused by incomplete reference libraries (Apothélosz-Perret-Gentil et al. 2017, 2021; Tapolczai et al. 2021; Gregersen et al. 2023).

1.5. Microscopy-based analysis of phytobenthos and diatoms

Light microscopy remains the conventional method for identifying diatoms and assessing their relative abundance. It has generated extensive taxonomic and ecological knowledge and provides continuity with historical monitoring datasets and established diatom indices. The

development of microscopy during the nineteenth and twentieth centuries enabled increasingly detailed examination of frustule morphology and formed the basis of modern diatom taxonomy (Mann 2010).

Two microscopy-based approaches are particularly relevant to this dissertation. The first is the analysis of the wider phytobenthos community. In this approach, intact algae are examined, counted, and identified using methods such as those developed by Lund et al. (1958) and Utermöhl (1958). The method can record diatoms together with cyanobacteria, green algae, euglenoids, and other benthic algal groups. Depending on the objective, abundance may be expressed through cell counts, counting units, or biomass estimated from cell volume (Hillebrand et al. 1999).

The second approach specifically targets diatoms. Organic material is removed from the sample, commonly using an oxidation procedure, and the cleaned valves are mounted in a high-refractive-index medium on permanent slides (CEN 2002). Taxa are then identified from diagnostic valve characteristics, and relative abundance is estimated by counting a standardized number of valves. Counts of approximately 400 valves are generally considered sufficient for routine ecological assessment, although they cannot be expected to reveal the complete taxonomic inventory of a diverse sample (Prygiel et al. 2002).

Microscopy has several important advantages. It provides direct morphological evidence, permits the detection of taxa without reference DNA sequences, and allows analysts to evaluate valve condition, morphological variability, and unusual forms. It is also directly compatible with the taxonomic concepts and abundance data underlying most established monitoring indices. Permanent slides can be re-examined, compared among analysts, and stored as long-term records.

Nevertheless, microscopic identification is laborious and requires substantial taxonomic expertise, particularly in species-rich samples and in groups containing morphologically similar taxa (Pawlowski et al. 2016). The quality of identification may vary among analysts and laboratories, while continuing taxonomic revisions can reduce comparability among datasets collected at different times (Kahlert et al. 2009; Kulaš et al. 2022).

Routine counting effort also limits the probability of detecting low-abundance taxa. A species represented by only a few valves may remain unrecorded even when it is present in the

sample (Kermarrec et al. 2013; Tapolczai et al. 2024). Small, weakly silicified, damaged, or morphologically cryptic species may be particularly difficult to recognize. Chemical treatment can also destroy or distort delicate valves. These limitations can lead to an underestimation of diversity and reduce the detection of rare or ecologically specialized taxa (Pérez-Burillo et al. 2022).

Phytobenthos microscopy and cleaned-diatom microscopy also describe different components of the community. Phytobenthos analysis provides a broad view of algal composition, but taxonomic resolution and counting effort are divided among several algal groups. Cleaned-diatom analysis concentrates exclusively on diatom valves and generally permits more detailed species-level identification. The choice between these approaches should therefore depend on whether the objective is to characterize the full benthic algal community or to obtain a high-resolution inventory of diatoms.

1.6. Environmental DNA metabarcoding of diatoms

DNA metabarcoding has emerged as an important complement to morphology-based biodiversity assessment. Environmental DNA is broadly defined as DNA obtained directly from an environmental sample rather than from an isolated organism (Taberlet et al. 2012, 2018). In microbial communities, however, separating DNA originating from intact organisms from extracellular or degraded DNA is difficult. The interpretation of eDNA data therefore depends on the sampled material, the target organisms, and the ecological question being addressed (Pawlowski et al. 2020, 2021; Rodriguez-Ezpeleta et al. 2021).

Diatom metabarcoding involves extracting DNA from a sample, amplifying a selected barcode region, sequencing the resulting amplicons, and comparing the sequences with a reference library. The workflow generally includes sample collection and preservation, DNA extraction, PCR amplification, library preparation, high-throughput sequencing, bioinformatic processing, taxonomic assignment, and ecological analysis. Each stage can influence the final representation of the community.

Metabarcoding allows many samples and large numbers of sequences to be processed simultaneously. It can therefore reduce the dependence of routine identification on continuous microscopic examination and may increase the spatial and temporal coverage of biomonitoring programmes. The large sequencing depth generated by high-throughput platforms also

improves the probability of detecting taxa occurring at low abundance. Molecular methods may additionally distinguish genetic variants within morphologically similar or cryptic species complexes that cannot be reliably separated using light microscopy alone (Vasselon et al. 2019; Keck et al. 2022; Pérez-Burillo et al. 2021).

Bioinformatic processing may generate operational taxonomic units or amplicon sequence variants. Operational taxonomic units are produced by clustering sequences according to a selected similarity threshold, whereas ASVs represent exact sequences inferred after modelling and correcting sequencing errors. ASVs provide single-nucleotide resolution and can be compared directly among studies when the same barcode region is used (Callahan et al. 2016, 2017). They can also retain ecologically meaningful variation within nominal species and may reveal stronger relationships with environmental gradients than datasets aggregated exclusively at species level (Kahlert et al. 2021; Tapolczai et al. 2021).

The ecological interpretation of metabarcoding data depends strongly on the completeness and accuracy of reference libraries. A sequence can be assigned to a species only if a reliable and sufficiently similar reference sequence is available. Missing taxa may remain unclassified or may be assigned only to genus or a higher taxonomic rank. Incorrectly identified reference material can also propagate errors into subsequent studies. Curated databases such as diat.barcode have considerably improved the assignment of freshwater diatom sequences, but geographical and taxonomic gaps remain (Rimet et al. 2018, 2019).

Laboratory and analytical biases must also be considered. Differences in cell-wall properties may influence DNA extraction, while primer-template mismatches can cause unequal amplification among species. Sequence-read abundance is further affected by cell biovolume, gene copy number, PCR efficiency, sequencing depth, and bioinformatic filtering. Read proportions should therefore be interpreted as a relative molecular signal rather than as a direct count of cells or valves (Vasselon et al. 2018; Lamb et al. 2019).

1.7. Barcode markers used in diatom metabarcoding

The choice of barcode marker is a central decision in a metabarcoding study because it affects amplification success, taxonomic coverage, species-level resolution, and compatibility with reference databases. A suitable barcode must contain conserved regions that permit primer binding while also containing enough sequence variation to distinguish closely related taxa.

The selected fragment must also be compatible with the read length and error characteristics of the sequencing platform (Hebert et al. 2003; van der Loos and Nijland 2021).

Several genetic markers have been evaluated for diatoms, including the chloroplast ribulose-1,5-bisphosphate carboxylase large subunit gene, the nuclear 18S ribosomal RNA gene, the internal transcribed spacer region, and the mitochondrial cytochrome c oxidase subunit I gene (Evans et al. 2007; Hamsher et al. 2011; Moniz and Kaczmarzka 2010; Stoof-Leichsenring et al. 2020; Zimmermann et al. 2011).

The *rbcL* gene is currently one of the most widely used markers in freshwater diatom metabarcoding. It provides comparatively high taxonomic resolution and is supported by an extensive specialized reference library (Kermarrec et al. 2014). Short regions of approximately 263 or 312 base pairs have been developed for high-throughput sequencing. These fragments are sufficiently variable to discriminate many species while remaining short enough for reliable amplification and paired-end sequencing (Vasselon et al. 2017). The availability of curated *rbcL* reference sequences through diat.barcode has further contributed to its widespread use.

The 18S rRNA gene, particularly its V4 region, has also been used in diatom metabarcoding. Its conserved regions facilitate amplification across broad groups of eukaryotes, making it useful when the study aims to characterize wider microbial communities. However, its species-level resolution is often lower than that of *rbcL* for diatoms. Comparisons of the two markers have generally reported more successful genus- and species-level identification and stronger ecological relationships for *rbcL*, partly because of the greater coverage of diatom-specific reference sequences (Zimmermann et al. 2015; Bailet et al. 2019; Apothéloz-Perret-Gentil et al. 2021; Wang et al. 2024).

The internal transcribed spacer, or ITS, can contain substantial variation among closely related taxa and may provide high taxonomic resolution. Its use in broad community metabarcoding is nevertheless complicated by length variation, alignment difficulties, and uneven reference-library coverage. The mitochondrial COI marker is widely used for animal DNA barcoding and has also been investigated in diatoms. However, the development of broadly effective diatom primers and comprehensive COI reference databases has proved difficult, and the marker remains less widely used than *rbcL* or 18S in routine diatom metabarcoding.

No individual marker is optimal for every application. Marker selection must balance taxonomic resolution, primer universality, amplicon length, sample quality, sequencing technology, and reference-library completeness. Multi-marker approaches may increase taxonomic coverage, but they also increase laboratory costs and introduce additional difficulties when combining data generated from markers with different amplification efficiencies and taxonomic resolutions.

1.8. Complementarity of microscopy and DNA metabarcoding

Microscopy and metabarcoding characterize biological communities through fundamentally different signals. Microscopy records visible cells or valves and assigns taxonomic identities according to morphological characteristics. Metabarcoding records amplified DNA sequences and assigns them to taxonomic or molecular units through comparison with reference data. Differences between their results should therefore not automatically be treated as errors. Each method detects components of the community that may be missed or represented differently by the other.

Microscopy provides direct morphological evidence and maintains continuity with established taxonomic frameworks, reference collections, long-term datasets, and conventional ecological indices. It can identify taxa for which no reference barcode is available and permits the examination of morphological variability. However, its detection capacity is constrained by counting effort, sample preparation, taxonomic expertise, and the visibility of diagnostic features.

Metabarcoding offers higher throughput and generally greater sensitivity to low-abundance taxa. It can distinguish sequence variants within morphologically defined species and preserve ecological information that would otherwise be lost through taxonomic aggregation. Its main limitations are dependence on reference libraries and susceptibility to biases arising during DNA extraction, amplification, sequencing, and bioinformatic analysis. Metabarcoding may therefore detect a taxon without assigning it to species, whereas microscopy may identify the same organism morphologically even when no barcode is available.

The two methods also estimate abundance differently. Microscopy generally produces relative valve counts, while metabarcoding produces relative read abundance. A single large

diatom cell may contain more copies of the targeted gene and more DNA than a small cell, leading to disproportionate representation in sequence data. Cell-biovolume correction factors have consequently been proposed to improve the agreement between read proportions and microscopic counts (Vasselon et al. 2018; Mortágua et al. 2019; Kulaš et al. 2022). Nevertheless, extraction efficiency, primer bias, and copy-number variation cannot always be fully corrected.

Presence–absence data can reduce the influence of some quantitative biases, particularly when the primary objective is to compare species inventories or distribution patterns. However, this approach removes information on dominance and community evenness that is required for many established ecological indices. The selection of abundance-based or presence–absence data should therefore be determined by the ecological question rather than applied as a universal solution.

Comparative studies generally find the greatest agreement between microscopy and metabarcoding for abundant and frequently occurring taxa. Discrepancies are more pronounced among rare species, cryptic taxa, weakly silicified organisms, and species absent from reference libraries (Kermarrec et al. 2013; Bailet et al. 2020; Kulaš et al. 2022; Tapolczai et al. 2024). Metabarcoding often records higher richness within individual samples because thousands of sequences can be examined, whereas microscopy may produce a larger cumulative species inventory across an entire study when experienced taxonomists identify species that are missing from molecular databases.

Neither method can currently provide a complete and unbiased representation of diatom communities. Microscopy and metabarcoding are therefore most appropriately viewed as complementary approaches. Their combined interpretation can improve taxonomic coverage, clarify methodological discrepancies, and provide a more comprehensive understanding of diversity and environmental responses than either method used independently.

1.9. Rare diatoms and the neglected component of biodiversity

The need for improved biodiversity assessment has become increasingly urgent as freshwater ecosystems are exposed to pollution, land-use change, habitat fragmentation, hydromorphological alteration, eutrophication, salinization, and climate-related disturbances. Freshwater habitats support a disproportionately large share of global biodiversity but are

among the ecosystems most severely affected by human activities (Dudgeon et al. 2006; Stendera et al. 2012; Reid et al. 2019). The resulting biodiversity decline may alter primary production, nutrient cycling, resistance to biological invasions, and ecosystem stability.

Rare species are particularly vulnerable because small population sizes, restricted distributions, and specialized habitat requirements can reduce their capacity to persist under environmental change. Their local disappearance may remain unnoticed, especially in microbial and algal communities where detection is difficult. This does not imply that rare taxa are ecologically unimportant. They may represent unique evolutionary lineages, occupy specialized niches, contribute strongly to beta diversity, or provide functions that become important when environmental conditions change (Jousset et al. 2017; Violle et al. 2017).

Rarity is not a single ecological condition. Species may be rare because they occur at low abundance, occupy a limited geographical range, are restricted to specific habitats, or display combinations of these characteristics. Rabinowitz's framework distinguishes rarity according to population size, habitat specificity, and geographic range, thereby providing a more informative classification than abundance alone (Rabinowitz 1981). The criteria selected to distinguish rare and common species can substantially affect ecological interpretation, and studies based on a single threshold may combine species with very different ecological characteristics (Gaston 1994; Espeland and Emam 2011; Maciel 2021).

Rare diatoms remain poorly understood. Their detection is limited by routine microscopic counts, while their ecological preferences are difficult to determine because they occur in few samples or at low abundance. Although metabarcoding can improve the detection of low-abundance sequences, incomplete reference libraries may prevent these sequences from being assigned to species. Molecular detection alone also does not explain why a species is rare or what role it plays within a community.

The few studies addressing rarity in aquatic algal communities indicate that rare taxa can contribute substantially to community variability, species turnover, and ecological resilience (Padisák 1992; Padisák et al. 2010; Görgényi et al. 2023). Their importance should therefore be examined through ecological frameworks that consider their distributions and community relationships, rather than solely through improved detection. This forms the basis of Chapter II, in which microscopy-derived diatom data are used to classify rarity and examine the contribution of rare taxa to community variability, beta diversity, and network structure.

1.10. Knowledge gaps and dissertation rationale

Despite substantial progress in diatom ecology and biomonitoring, several important gaps remain. First, the rapid development of DNA metabarcoding has produced considerable variation in sampling, preservation, DNA extraction, marker selection, sequencing, bioinformatic processing, and taxonomic assignment. These methodological choices can influence the detected community and complicate comparisons among studies. Diatom metabarcoding research is also geographically uneven, with much greater representation in Europe, North America, and parts of East Asia than in many regions of Africa, South America, and the Middle East.

Second, reference libraries remain incomplete and geographically biased. Species without reliable barcode sequences cannot be assigned accurately, while incorrect or outdated taxonomic identifications may introduce errors into molecular inventories. Continued development and curation of reference libraries are therefore necessary before molecular methods can be applied consistently across regions and ecosystem types.

Third, basic ecological knowledge of rare diatom taxa remains limited. Rare species are frequently removed from ecological analyses or treated as sampling noise, despite their potential contribution to beta diversity, ecological networks, resilience, and conservation value. Their role cannot be understood solely by increasing detection sensitivity; it also requires multidimensional definitions of rarity and analyses conducted across sufficiently large spatial networks.

Finally, relatively few studies have compared phytobenthos microscopy, diatom microscopy, and ASV-based diatom metabarcoding using samples collected from the same sites. Such comparisons are needed to determine how methodological choice affects recorded richness, diversity, abundance patterns, and relationships between communities and environmental gradients.

This dissertation addresses these gaps through three complementary components. Synthesizes the geographical distribution, methodological development, applications, and limitations of diatom DNA metabarcoding (section 3). Uses microscopy-derived data to investigate rarity types and the contribution of rare benthic diatoms to community variability, beta diversity, and network structure (section 4). Compares phytobenthos microscopy, diatom

microscopy, and diatom metabarcoding in their capacity to characterize biological diversity and reveal community responses to environmental gradients (section 5).

2. Main objectives

The overall aim of this dissertation was to improve understanding of diatom diversity, rarity, and environmental responses through complementary microscopic and molecular approaches.

The dissertation pursued three main objectives:

1. To evaluate the development, methodological variability, applications, and remaining limitations of diatom DNA metabarcoding, particularly in relation to taxonomic detection, reference-library coverage, ecological assessment, and methodological standardization.
2. To investigate the ecological significance of rare benthic diatom species in river networks by classifying different forms of rarity and assessing their contribution to community variability, beta diversity, and ecological network structure.
3. To compare phytobenthos microscopy, diatom microscopy, and diatom DNA metabarcoding in their ability to characterize biodiversity and reveal community responses to environmental gradients.

These objectives were addressed in three corresponding chapters. Chapter I provides a systematic synthesis of diatom DNA metabarcoding research and evaluates the methodological factors affecting its application. Chapter II examines the distribution and ecological role of rare benthic diatoms using microscopy-derived data from a river network. Chapter III compares three identification approaches and assesses how differences among them influence estimates of diversity, abundance, and community–environment relationships.

3. Environmental DNA Metabarcoding of Diatoms: A Next-Generation Tool for Aquatic Biodiversity and Bioassessment

This chapter is a modified version of submitted manuscript to Metabarcoding and Metagenomic journal:

Abubaker, B., Stenger-Kovács, C., Rimet, F., & Tapolczai, K. Environmental DNA metabarcoding of diatoms: A next-generation tool for aquatic biodiversity and bioassessment. *Metabarcoding and Metagenomics*. 10, e168096. (Q1, Impact factor: 3.1)

3.1. Introduction

Environmental DNA (eDNA) is generally defined as the total pool of DNA extracted from environmental samples (Taberlet et al. 2012, 2018). Taberlet's broad concept emphasizes that eDNA is characterized by its source, with no limitations regarding its taxonomic composition or structural state (i.e. whether it is intracellular or extracellular). Pawlowski et al. (2020) recommended a two-level terminology to indicate the origin of the samples and the target taxa. Conversely, Rodriguez-Ezpeleta et al. (2021) suggested focusing on identifying the type of targeted DNA based on organismal and extra-organismal terminology, as this is inevitable for proper interpretation. However, separating organismal and extra-organismal DNA is rather challenging for microorganisms, so this approach is difficult to apply directly (Pawlowski et al. 2021). DNA metabarcoding has marked a new milestone in biological and ecological studies by enabling organisms to be identified based on their genetic material. The identification of DNA sequences using DNA barcoding is an innovative approach to organism identification, supported by continuous technological advances (Taberlet et al. 2012). DNA metabarcoding has become one of the most popular approaches among ecological scientists, with an increasing number of comprehensive best practice guides on the topic becoming available and the cost of DNA sequencing decreasing (Zinger et al. 2019). A comprehensive picture of an organism's entire genetic material can provide information on not only the community composition but also ecosystem functions and interactions.

The diatom DNA metabarcoding process as shown in (Fig. 1), consists of seven steps: (1) sampling and preservation, (2) DNA extraction, (3) PCR amplification, (4) library preparation (5) DNA sequencing, (6) the bioinformatics process and (7) data analysis. The effect of each step on the study's results has been assessed through more or less intensive study, resulting in the development of specific protocols.

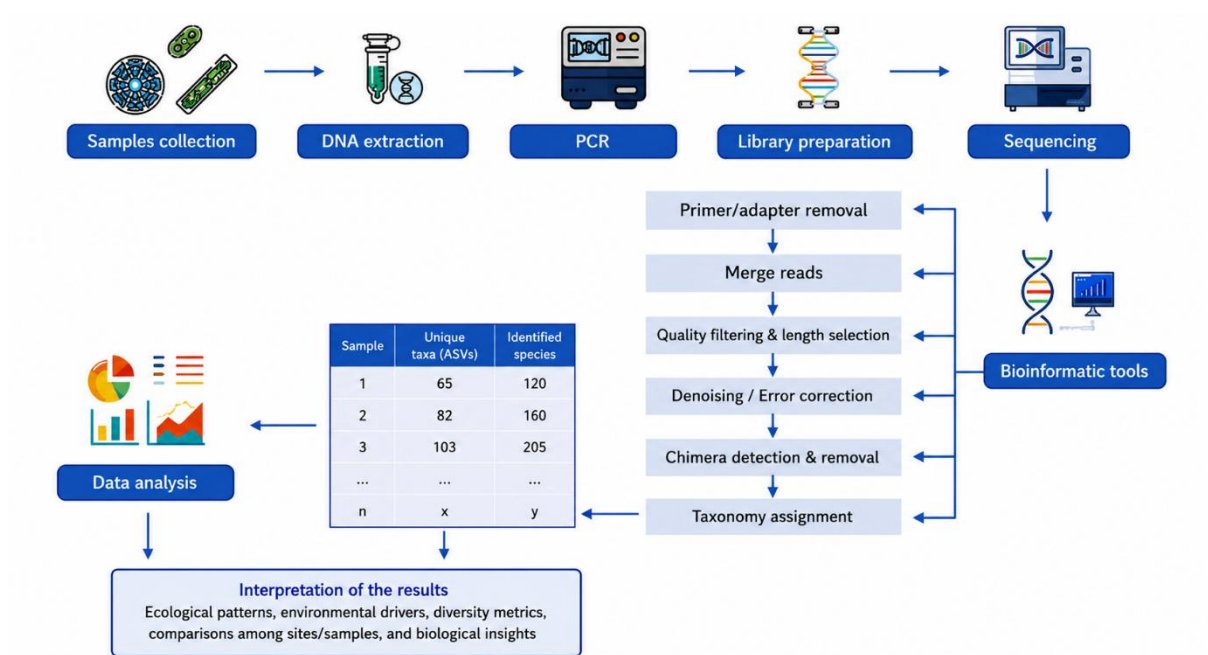


Figure 1. Diatom DNA metabarcoding workflow

3.2. Methods

To achieve an integrative coverage of the literature, more than 580 peer-reviewed papers were systematically collected to identify the relevant studies addressing diatom metabarcoding applications, insights and challenges. The Web of Science database was searched using the following prompt: (TS = (metabarcoding) OR TS = (meta-barcoding) OR TS = (High-throughput sequencing) OR TS = (High-throughput sequencing) OR TS = (next-generation sequencing) OR TS = (next generation sequencing) AND (TS = (diatom) OR TS = (diatoms)). A total of 582 papers were identified, and additional manual screening was applied to select the required studies, including the following criteria: (i) the study targets diatoms (more than 350 papers were excluded as they targeted microalgae, bacteria, or protists in general), (ii) the articles are primary research papers, excluding reviews, opinion papers, and dataset presentations). Finally, 128 publications were selected for subsequent analysis. Data from each study were extracted into a standardized table, including the following variables: publication year, study location (country), habitat, type of samples, method of DNA extraction, target gene region, sequencing technology, and bioinformatic pipelines. The data shows a significant increasing in DNA metabarcoding in the recent years (Fig.2).

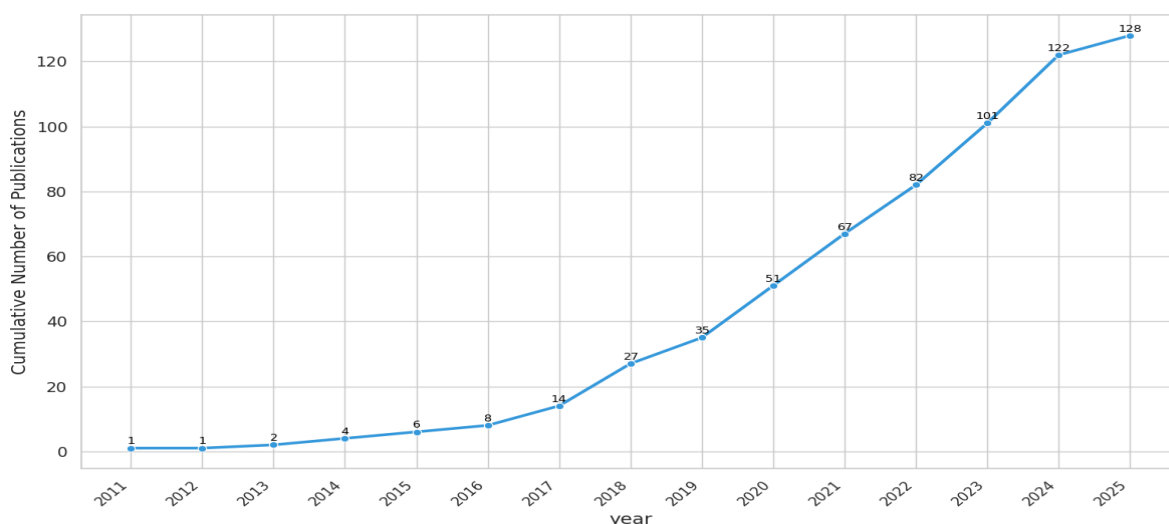


Figure 2. Cumulative number of publications about diatom eDNA metabarcoding based on Web of Science data (3rd of March 2025).

This chapter focused on the following themes:

i) *Adoption and coverage:*

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

ii) *Field and lab practices*

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

iii) *Genetic markers and reference sequences*

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

iv) *Bioinformatics pipelines*

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

v) *Detection performance*

Comparisons of diatom metabarcoding and microscopy as two identification methods for diatoms were analysed in terms of efficiency for successful taxa detection and diversity.

vi) *Ecological signal*

The efficiency of metabarcoding-based monitoring metrics (e.g. The Specific Pollution Sensitivity Index (IPS) and the Trophic Diatom Index (TDI) indices of ecological quality status) were studied in terms of their ability to capture environmental gradients or in comparison to morphology-based indices as references.

vii) *Applications*

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

(viii) *Limitations*

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

(ix) *Future prospects and priorities*

We outlined some perspectives addressing the development of measures to improve reference libraries, harmonise fieldwork to pipeline standards, and next-generation analytics.

3.3. Results

3.3.1. Global geographical distribution of diatom eDNA studies

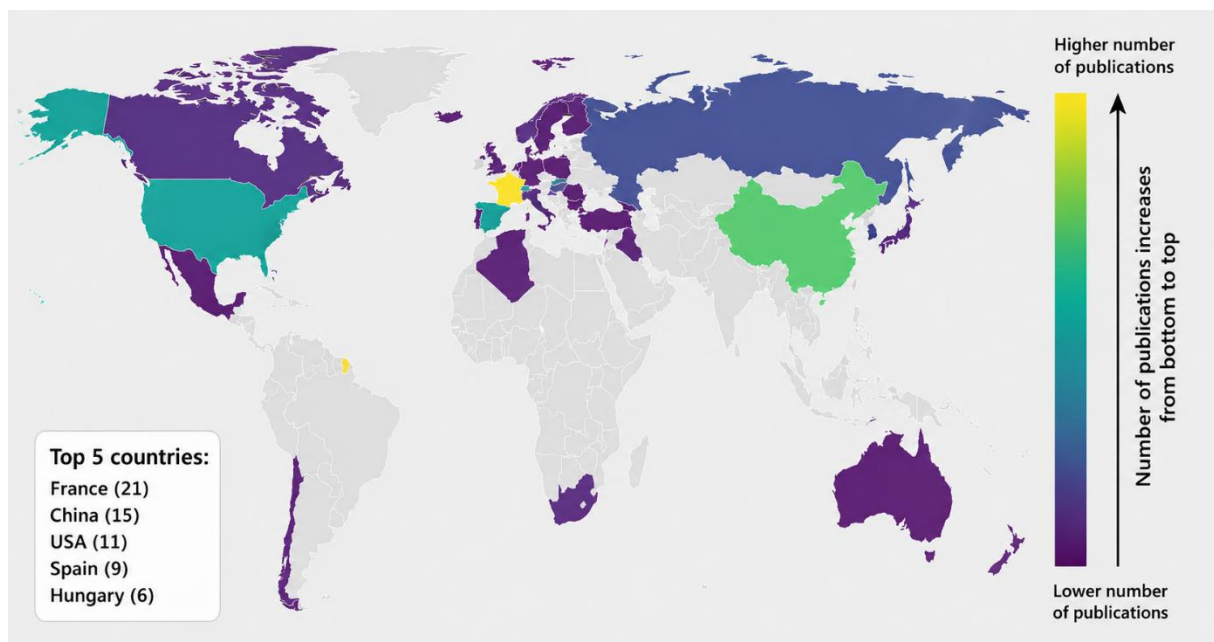


Figure 3. Global distribution of diatom eDNA metabarcoding studies showing the number of publications.

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

3.3.2. Diatom eDNA metabarcoding workflow

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

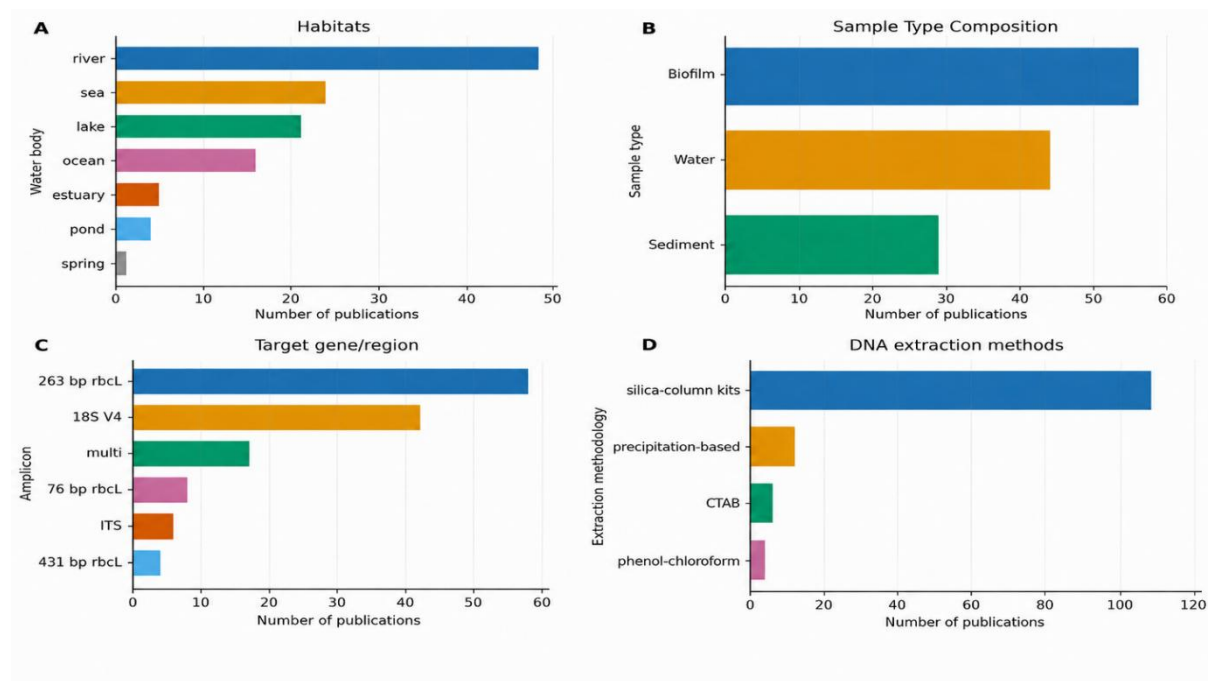


Figure 4 Overview of surveyed publications: (A) habitats, (B) sample types, (C) target gene/region, and (D) DNA extraction methods. Counts are shown as horizontal bars and scaled to the number of publications per category.

In addition to DNA extraction and marker selection, the sequencing technologies were studied. The bubble sizes in (Fig. 5) denote the number of studies per year using sequencing platforms, enabling a comparative view of temporal uptake. Sequencing platforms transitioned from early pyrosequencing and the occasional use of Ion Torrent and other Illumina platforms, to a sustained dominance of the Illumina MiSeq platform.

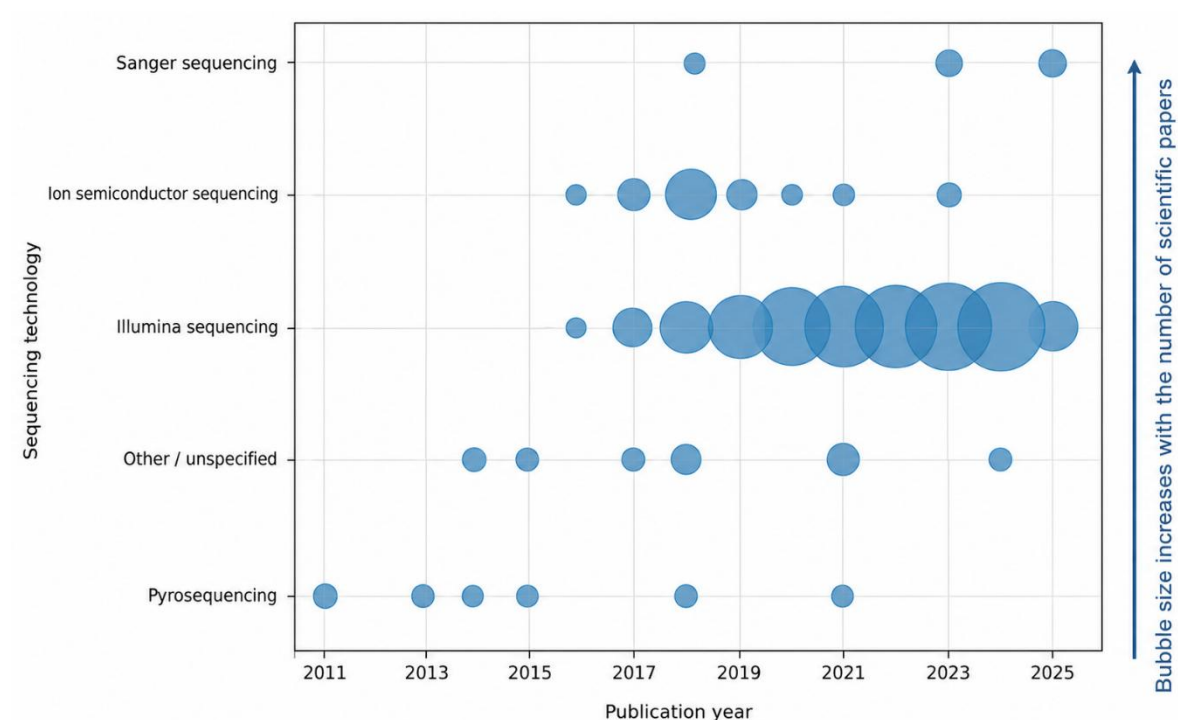


Figure 5. Sequencing platforms used in diatom DNA metabarcoding studies (Note: data for 2025 is partial until the 3rd of March 2025)

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

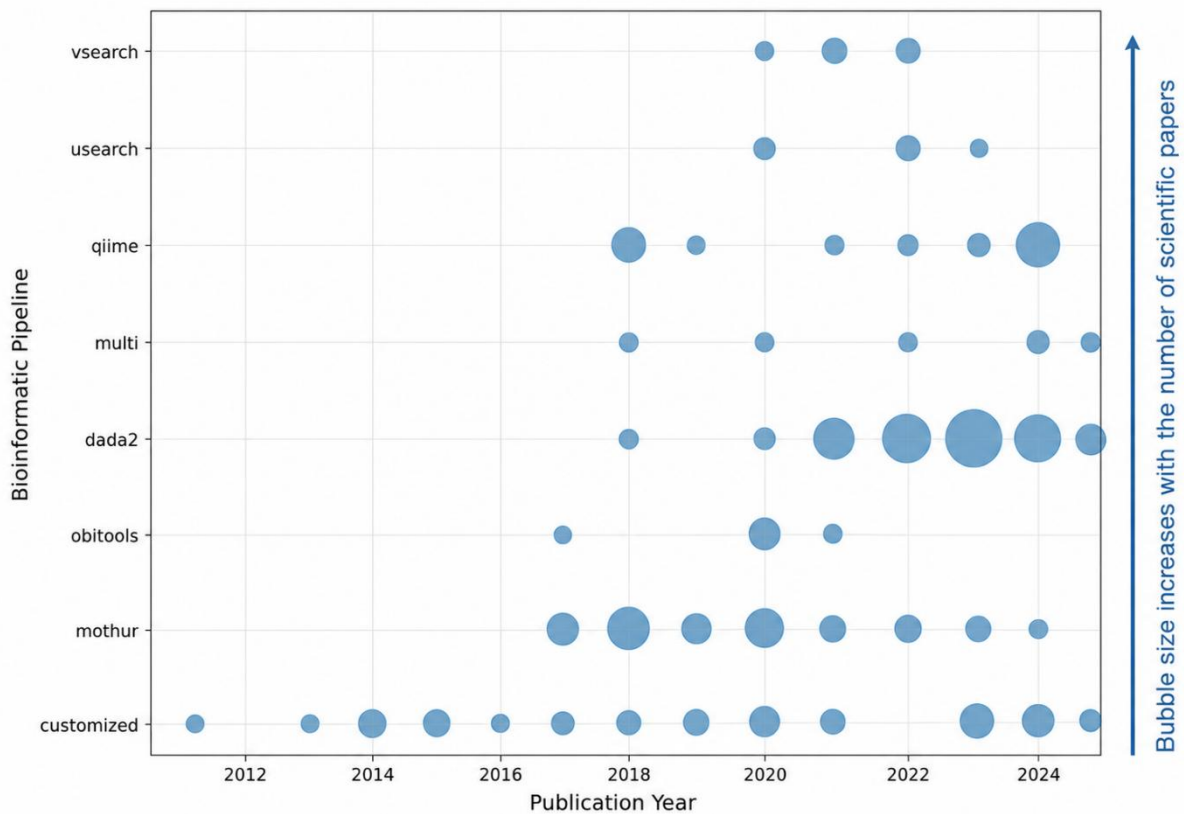


Figure 6. Bioinformatic pipelines are frequently used in diatom DNA metabarcoding studies (Note: data for 2025 is partial until the 3rd of March 2025)

3.4. Discussion

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

3.4.1. Sample collection and preservation

Sample collection and preservation represent critical initial steps in diatom DNA metabarcoding, where many elements of traditional sampling protocols can be retained, while specific modifications are required to ensure DNA integrity and prevent contamination. According to our literature analysis, the following aspects are generally considered: (i) maintaining the integrity of the DNA by using preservatives and keeping the samples in cool

or process the samples in the laboratory as soon as possible; (ii) doing it in a way that is as similar as possible to the sampling for the morphological analysis and (iii) potentially using the same samples for diatom analysis as are used for analysing other organisms. The sampling design in terms of location, frequency, and number of samples does not require any modifications to the classical methodology, rather, it depends on the aim of the study.

Benthic diatoms are typically collected from common natural substrates, such as macrophytes (Phiri et al. 2007), stones (Smucker et al. 2020), sediments (Liu et al. 2020) or artificial substrates (Nadir Solak et al. 2021). Biofilm sampling can be carried out in various ways, such as using a toothbrush or scalpel, or pipetting directly from loose sediment. This process is generally standardized in national and international protocols (CEN 2018). The same protocol can be used for DNA metabarcoding sampling, but more attention must be paid to avoid cross-contamination of samples (e.g. by wearing gloves or sterilizing sampling and storage equipment). The removed biofilm can be further treated with oxidizing agents and hydrochloric acid to prepare diatom frustules for microscopic analysis. However, it can also be used for direct DNA extraction. Although Lugol's iodine and formaldehyde are sometimes recommended for preserving samples for morphological analysis (Lund et al. 1958), they are not efficient for maintaining DNA integrity and ensuring the quality and quantity of extractable DNA (Mäki et al. 2017). The more commonly used ethanol preservation method can also be applied for this purpose. However, while microscopic analysis of diatoms requires only enough ethanol to fix the cells for morphological identification, e.g. 50% (Burfeid-Castellanos et al. 2022), a higher concentration (ca. 70%) is required for complete DNA preservation (Deiner and Altermatt 2014; Vasselon et al. 2017). Other storage methods include centrifuging the biofilm suspension, and then freezing the pellets (Visco et al. 2015) or using nucleic acid preservation to stabilize and protect RNA/DNA (Kelly et al. 2020).

Diatoms in the plankton are typically studied within phytoplankton communities but some metabarcoding studies are specifically targeting them (Turk Dermastia et al. 2023). In such cases, sampling can be conducted by filtering a given volume of water (Cui et al., 2023) or by using plankton nets. However, the latter is considered a bulk sample rather than an environmental one. These samples are preserved and stored using deep freezing (Hinlo et al. 2017), liquid nitrogen (Cui et al. 2023), or alcohol (Borrego-Ramos et al. 2025) until DNA extraction. Additionally, Maitland et al. (2020) investigated the feasibility of analysing diatoms using DNA extracted from kick-net bulk tissue samples which are typically used for the sampling of aquatic macroinvertebrates. Their study revealed no significant differences in

diatom community assemblages compared to those in periphyton samples collected from various substrates. This suggests that new sampling strategies could be adopted for ecological biomonitoring.

Sediment sampling for diatom DNA metabarcoding poses unique challenges compared to water-based eDNA collection, primarily due to factors such as water depth and sediment characteristics. Soft sediments, which are rich in DNA, are relatively straightforward to sample, whereas hard-bottom sediments present greater difficulties (Pawlowski et al. 2022). Once it is critical to preserve sediment eDNA samples maintain DNA integrity. Common preservation methods include deep-freezing (Visco et al. 2015), preservation in a nucleic acid preservation solution (e.g. RNAlater) (Rissanen et al. 2010), and preservation in ethanol (Sales et al. 2019). The aim of these methods is to prevent DNA degradation and ensure that samples remain suitable for downstream metabarcoding analysis over various storage durations.

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

From the reviewed studies, only one investigated the effect of sample storage methods on the identified diatom community composition, diversity and ecological quality index. Baricevic et al. (2022) tested three sample storage methods, including (1) preservation in 70%

ethanol and storage at 4°C, (2) preservation in a nucleic acid preservative and storage at -20°C and (3) centrifuging the biofilm and storage the remaining pellet at -20°C. Additionally, storing time was tested from one day up to one year, showing no significant effect of the preservation method or storage duration on the ecological index values and diversity metrics.

Regarding the origin of samples studied in the reviewed articles, we can conclude that among the seven habitat types shown in (Fig. 4A), rivers have been sampled most frequently. This likely reflects the fact that benthic diatoms are strongly dominant in riverine systems and thus in routine river biomonitoring programs and water-management frameworks. This has made diatoms a primary focus of scientific, regulatory, and stakeholder attention. Rivers are followed by sea and lake environments, which also show considerable representation. Oceans and estuaries have moderate detection levels, while ponds and springs show the lowest presence of diatom DNA metabarcoding studies. These habitats are understudied, so it would be interesting to conduct more studies targeting them. The data highlights the variability of diatom occurrence across water bodies and shows the importance of sampling diverse habitats in eDNA-based biomonitoring to capture a comprehensive picture of aquatic biodiversity.

3.4.2. DNA extraction

It is essential to conduct a comprehensive comparison of molecular protocols used in DNA workflows in order to identify and account for potential biases introduced by different methodologies. Such an understanding will help improve the reliability and consistency of future studies and ensure more accurate interpretations of DNA-based analyses (Darling and Mahon 2011). In this review, we identified widely used DNA extraction protocols in diatom DNA metabarcoding studies (Fig. 4D). Completely extracting DNA from environmental samples is crucial to obtain adequate quantities of high-quality genetic material for subsequent analyses. This is particularly important for single-celled organisms such as diatoms, as inefficient DNA extraction can result in the underrepresentation or even the complete absence of rare species with low cell numbers. Most of the reviewed studies used commercially available extraction kits. These kits typically involve the following steps: (i) Lysis or cell disruption to break open cells and release DNA using chemical, enzymatic and/or mechanical means. Most kits use silica columns or magnetic beads (ii) to bind DNA and separate it from other cellular components. Several washing steps with a buffer solution are then included to remove contaminants and preserve DNA binding. We found that the most commonly used kits are those designed for soil samples as these increase the yield and quality of DNA. Soils are

rich in PCR inhibitors, e.g. humic acids, so their removal is a key step. The two most popular kits are: Qiagen's DNeasy PowerSoil Kit and Macherey-Nagel's NucleoSpin Soil Kit.

A comparative study investigating the effect of five different DNA extraction kits/protocols on diatom community composition, diversity and quality index was carried out (Vasselon et al. 2017b). A precipitation-based protocol using aGenElute™-LPA kit from Sigma–Aldrich was compared with four commercial silica-column-based kits. All five methods were suitable for high-throughput sequencing (HTS), and the choice of extraction method did not significantly affect OTU richness. Similarly, a more recent study by Newbold et al. (2025) conducted comparative research on four popular DNA extraction kits, including Qiagen Blood and Tissue, Qiagen DNeasy PowerSoil Pro Kit, Thermo MagMAX Microbiome Ultra Nucleic Acid Isolation Kit and the Zymo Quick-DNA Fecal/Soil Microbe Miniprep Kit. Their performance was evaluated in terms of PCR and sequencing success. The study found that protocols involving mechanical lysis yielded higher PCR and sequencing success rates as well as increased cross-kingdom richness. However, the overall community composition was largely consistent across the different extraction methods, indicating that extraction methodology introduces minimal bias in multi-taxa DNA metabarcoding analyses.

3.4.3. Amplicon marker selection

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

The development of DNA barcoding for protists, including diatoms, was initially challenging due to their immense diversity and the absence of a universal marker. Pioneering efforts by groups such as the CBOL Protist Working Group, were crucial in establishing standardized barcoding approaches for eukaryotic diversity beyond the scope of traditional

taxonomic classification (Pawlowski et al. 2012). Their work, emphasizing both universal and group-specific barcodes, laid the groundwork for current advancements in eDNA metabarcoding for diverse protist groups, including diatoms (Pawlowski et al. 2012). Several early studies explored conserved genetic markers for DNA barcoding, including the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene (*rbcL*), the cytochrome c oxidase subunit I gene (COI), the internal transcribed spacer region (ITS), and the 18S ribosomal RNA gene (18S rRNA) (Evans et al. 2007; Hamsher et al. 2011; Moniz and Kaczmarek 2010; Stoof-Leichsenring et al. 2020; Zimmermann et al. 2011). Selecting an appropriate amplicon marker is a crucial step for improving taxonomic identification and phylogenetic resolution. For this reason, three molecular markers targeting the chloroplast genome, *rbcL*; the nuclear genome, small-subunit ribosomal DNA (SSU rDNA, including the 18S rRNA gene); and the mitochondrial genome, COI, have been investigated and mitochondrial (*COI*) genomes were investigated. Of these, the *rbcL* marker demonstrates strong discriminatory capacity and benefits from an extensive diatom reference database (Kermarrec et al. 2014). Significant progress has been made in designing diatom-specific primers that target the 263 bp region of *rbcL*, which is variable enough for species-level discrimination yet still amplifiable from degraded environmental DNA (Vasselon et al. 2017). The 263 bp region of *rbcL* is widely used in diatom metabarcoding studies, our literature review concluded. As shown in (Fig. 4C), over 50 publications use the 263 bp *rbcL* region, establishing it as a core marker in DNA metabarcoding, supported by its taxonomic resolution and growing database.

A study by Zimmermann et al. (2015), which applied NGS sequencing to the V4 region of the 18S rRNA gene for metabarcoding diatoms showed great potential for assessing water quality. Bailet et al. (2019) highlighted the use of two different DNA markers, namely the 18S-V4 and *rbcL* barcodes. Their results demonstrated that the molecular method could generate species inventories, diatom index scores, and ecological status classifications that were comparable to those derived from the morphology-based approach. Conversely, Apothéloz-Perret-Gentil et al. (2021) applied three analytical methods to estimate index values: (1) direct calculation from sequences assigned to taxa, (2) de novo calibration of ecovalues for all metabarcodes, and (3) training of predictive models using a supervised machine learning approach. The correlations obtained using the *rbcL* marker were higher than those obtained using the 18S rRNA marker. This observation shows the importance of a comprehensive reference database, which is more extensive for the *rbcL* marker. At the genus level, 80.4% of diatom genera were identified using the *rbcL* marker, compared with 66.7% using the 18S

rRNA gene. These results confirm the superiority of *rbcL* over 18S rRNA for DNA metabarcoding identification (Wang et al., 2024).

3.4.4. High-throughput sequencing of amplicons

HTS technologies have revolutionized biodiversity studies by significantly reducing costs and labour. Importantly, they have made it possible to sequence the DNA of entire communities from several samples in parallel (Taberlet et al. 2012). Compared to first-generation sequencing technologies, such as Sanger sequencing which were used for barcoding of single specimens or strains, this second generation of sequencing (formerly known as Next Generation Sequencing or NGS) has introduced novel opportunities for diatom biomonitoring (Dijk et al. 2014).

The landscape of eDNA metabarcoding is shaped by the choice of sequencing technologies: short-read or long-read. Traditional short-read platforms, such as those from Illumina and Ion Torrent, typically have a limit of 500 bp or less, meaning that specific variable regions must be targeted (Pawlowski et al. 2012). In contrast, long-read platforms from Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) can sequence full-length barcode regions (Chwalińska et al. 2025). Recent advances in platforms such as ONT GridION and PacBio platforms have improved the accuracy and accessibility of long-read metabarcoding. ONT in particular has gained attention for its ability to generate rapid full-length taxonomic profiles, providing a more comprehensive view of diatom community composition than short-read approaches (Jamy et al. 2020).

The first NGS technology, 454 pyrosequencing, was one of the first high-throughput technologies to work with short read lengths, making it ideal for metabarcoding studies. It has already been used for diatom DNA metabarcoding studies (Kermarrec et al. 2013, 2014). Another early NGS platform, Ion Torrent, has also facilitated the development of DNA metabarcoding approaches targeting diatoms (e.g., Endo et al. 2018; Kahlert et al. 2021; Rivera et al. 2018). Illumina and its various instruments (e.g. MiSeq, HiSeq, NextSeq and NovaSeq) have become the dominant, widely used sequencing technology for metabarcoding (see Fig. 5), offering greater accuracy and throughput at a lower cost than previous technologies (Goodwin et al. 2016). Third-generation sequencing technologies, such as those from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) are working with increasingly reliable platforms for metabarcoding applications, and they may outperform Illumina sequencing in other respects. ONT provides the MinION sequencing tool, which can be used in the field, and it may enhance taxonomic resolution in community analysis (Stevens et al. 2023). However,

this technology has rarely been used for the community analysis of diatoms or other algae (Amamoo et al. 2025; Kaleli et al. 2025).

3.4.5. Bioinformatic pipelines and reference libraries

Bioinformatics plays a crucial role in analyzing DNA metabarcoding data by processing the sequencing datasets efficiently and providing meaningful results. This offers an efficient and scalable alternative for understanding ecological dynamics in response to environmental changes (Cordier et al. 2017). Various software tools have been applied for diatom metabarcoding, including packages such as Mothur (Schloss et al. 2009), QIIME (Caporaso et al. 2010), and DADA2 (Callahan et al. 2016). These tools encompass all stages from sequence filtering and clustering to taxonomic classification (Bailet et al. 2020; Rivera et al. 2020). Of the various tools employed in diatom DNA metabarcoding, DADA2 and Mothur were most commonly used pipelines (Fig. 6). In contrast, around 24 studies used customized packages, platforms, and datasets for the analysis.

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

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

Additionally, ASVs produced more favorable results when investigating the impacts of land use on diatoms. Furthermore, it was possible to develop an enhanced index using a taxonomy-free approach. This approach enabled valuable information about the ecological characteristics of sequences assigned to the same species, as well as unassigned sequences, to be retained. Looking ahead, integrating novel clustering techniques shows promise in the field

of biomonitoring. These methods could contribute to refining the delineation of taxonomic units, placing greater emphasis on ecological factors rather than exclusively relying on morphological or genetic criteria alone (Tapolczai et al. 2021). This suggests that incorporating the phylogenetic approach enhances the accuracy of ecological assessments based on DNA barcoding of freshwater diatoms. A study using a taxonomy-free approach achieved promising results in accurately assessing the ecological status of the examined sites (Apothéloz-Perret-Gentil et al. 2017). Specifically, the approach yielded a correct assessment for 77% of the analysed sites. A notable advantage of this method is that it enables the use of almost 95% of OTU for index calculation. By contrast, the taxonomic assignment approach only enabled the use of 35% of OTUs. This approach is both time-efficient and cost-effective for assessing the biological quality of surface waters. Values obtained from "taxonomy-free" approaches, involving molecular assignment and machine learning, showed a stronger correlation with the values of the morphological index values than values based on taxonomically assigned sequences. Correlations between the three analytical approaches were higher when *rbcL* was used compared to 18S rRNA (Apothéloz-Perret-Gentil et al. 2021).

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

3.4.6. Taxa inventory correspondence between eDNA metabarcoding and morphological analyses

Since the introduction of DNA metabarcoding to diatom research, several papers have studied the relevance of the methodology by validating resulting taxa inventory based on that of microscopy analysis. Firstly, the proportion of successfully assigned sequences is crucial for an accurate representation of the community. Based on the reviewed studies, the proportion of successful assignments is the lowest at the species rank, followed by the genus and the higher taxonomic ranks. If a sequence lacks a species match during taxonomic assignment, it is either misidentified or assigned to a higher rank. Early studies generally included a lower proportion of assigned sequences which is due to the continuous development of reference databases. One of the most used reference databases in diatom metabarcoding studies is diat.barcode, first published in 2012 with 3,549 barcodes of 715 species, while its latest version (Rimet et al.

2018b) contains 9,233 barcodes of 1,705 species. It contains sequences of barcoded strains from culture collections and several projects, many of which have not yet been deposited in GenBank, but also diatom sequences published on National Center for Biotechnology Information (NCBI). Species in understudied regions are less represented in the reference database, which results in lower taxonomic assignment success. Based on the diat.barcode entries, the geographical distribution of barcode sources is highly unbalanced. The origin of most of entries is unknown (28%), followed by France (16%), USA (9%), United Kingdom (7%) and the remaining 68% shared among 83 countries.

Important differences in the resulting communities have been observed in both composition and abundances with a varying number of shared taxa among taxonomic levels, year of study and the studied ecoregion. The number of shared species when comparing DNA metabarcoding and morphological approaches shows that there are still important discrepancies between methods when it comes to species composition. However, when the relative read number (as a proxy for abundance) was considered, the proportion of shared reads can be as high as 92% for species and 98% for genera (Tapolczai et al. 2024), clearly suggesting that inconsistencies are lower for abundant and common taxa. Besides the fact that abundant taxa are found with a higher probability using both methods, they are also more likely to be included in the reference barcode libraries. Additionally, the probability of misidentifying or neglecting abundant species with microscopy is less likely.

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

The challenges of biases related to reference libraries are well summarised by Keck et al. (2023). Some of these are relevant for diatom studies as well. Taxonomic mislabeling happens when the identification of the species added to the reference database is incorrect and the problem remains hidden if it affects species without further interest or if results are not validated by experts. Thus, supervision by taxonomists may still be necessary to disentangle such mistakes (e.g. Tapolczai et al. 2024). This shows that besides the completeness of the reference

library (Weigand et al. 2019), good-quality curation is almost as important (Keck et al. 2023). Missing taxa in the reference libraries are probably the most important bottleneck in the efficiency of diatom metabarcoding. This leads to no taxonomic assignment or, in the worst case, sequences being wrongly assigned to very closely related species (Couton et al. 2022; Tapolczai et al. 2025). It is thus of great importance to develop reference libraries by adding new entries with assigned taxonomy from cultures (Rimet et al. 2018b).

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

Some species, especially those that are weakly silicified are challenging to be seen under the microscope or they simply cannot sustain the chemical treatment for permanent slide preparation. A good example is *Fistulifera saprophila* which is more frequently detected by metabarcoding than by microscopy (Pérez-Burillo et al. 2020; Pissaridou et al. 2021; Tapolczai et al. 2024). This effect is similar for *Entomoneis*, *Amphipleura* or other species (Anslan et al. 2022; Kang et al. 2020).

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

3.4.7. Diatom eDNA metabarcoding as a bio-assessment tool

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

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

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

3.4.8. Diatom eDNA metabarcoding as a forensic tool

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

Major discrepancies were discovered between results generated by the routinely used Scanning Electronic Microscope (SEM) based diatom tests and 18S rRNA arrays. It was shown that the conventional SEM-based diatom testing method suffers from significantly low sensitivity, mainly due to the high loss of frustules during sample preparation. The adoption of diatom DNA arrays is expected to potentially revolutionize the forensic diagnosis of drowning

cases (Jiang et al. 2020). Liu et al. (2020) compared microscopic and DNA barcoding approaches in anatomical organs from 23 drowning victims and found that microscopic observations were weakly correlated with molecular analysis. These studies show that diatom DNA metabarcoding in forensic science is still in its early stages and requires further investigation to validate its reliability and standardize methodologies.

3.5. Limitations

The use of DNA in ecology and nature conservation has increased significantly over recent years, though this growth has not been matched by a full appreciation of its limitations. Some key challenges include (1) imperfect detection, (2) difficulties in quantifying absolute abundance, (3) issues with taxonomic assignment, (4) spatial and temporal dynamics of environmental DNA, (5) complexities in data analysis and interpretation, and (6) challenges in assessing ecological status (Beng and Corlett3 2020). A comprehensive understanding of these challenges is essential before DNA-based methods can confidently be adopted for routine biomonitoring. Despite promising results, several studies have highlighted potential biases in species detection and relative abundance data generated through DNA metabarcoding. Laboratory-based biases represent an important limitation in diatom metabarcoding. Differences in DNA extraction efficiency can influence the detected community composition because taxa vary in cell wall properties and susceptibility to lysis (Vasselon et al. 2017). Quantitative interpretation is also affected by biomass and biovolume-related biases, as differences among taxa can distort the relationship between read abundance and actual community structure (Vasselon et al. 2018). In addition, PCR amplification bias is a major source of error, since unequal primer affinity and amplification efficiency among taxa can skew relative abundances and lead to the underrepresentation of some species (Krehenwinkel et al. 2017). More generally, several studies have shown that sequence read proportions do not always reflect the original proportions of taxa in the sample because multiple sources of bias act simultaneously across the workflow (Lamb et al. 2019).

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

enhancing reference databases for habitat-specific species are essential steps forward (Kamberović et al. 2024).

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

3.6. Conclusion

The DNA metabarcoding in environmental sciences holds great promise, especially in biomonitoring and biodiversity assessment. As bioindicators, diatoms provide effective insights into ecological status, but traditional morphological identification methods need high-level expertise, cost and time-consuming. However, diatom DNA metabarcoding offers an alternative that captures comprehensive community profiles from environmental samples, allowing for rapid, cost-effective, and scalable biomonitoring. The following aspects could be a significant improvement for the diatom DNA metabarcoding communities:

i) Enhancing reference databases and taxonomic resolution, by focusing on coordinated international initiatives to complete the reference libraries. This process is essential for improving taxonomic assignments and comprehensive biodiversity assessments across different geographical regions and ecosystems. In the diat.barcode initiative, the scientific community organize themselves collectively, with a strong emphasis on community driven collaboration. Researchers from around the world share their diatom barcode data, and valuable feedback is provided whenever errors or inconsistencies are spotted. Importantly, everyone who

contributes to this effort is recognized as a co-author of diat.barcode, so the database remains open access and free, ensuring that it can serve as a shared, high-quality resource for researchers working with diatom DNA barcoding.

ii) Methodological standardization and quality assurance for routine monitoring, by establishing consensus protocols for each step in the metabarcoding workflow. These standardization efforts will be particularly important for regulatory applications, where consistency and reliability are important. Continuous effort from the scientific community like (Vasselon et al. 2025) and collaborative projects CEN, ISO, and IsteF ensure the consistency in methodological standardization.

iii) Advanced bioinformatic solutions represent a promising advancement for diatom eDNA metabarcoding. The shift from OTU-based methods to more precise ASV approaches has already demonstrated improved ecological resolution, also, utilizing taxonomy-free approaches were potentially enabling high resolution of diatom biodiversity and ecological status. Integration of machine learning algorithms and artificial intelligence could further enhance data analysis capabilities.

iv) Quantification and abundance estimation, addressing the challenge of quantitative assessment remains a significant hurdle for diatom metabarcoding. Current approaches primarily provide presence-absence or relative abundance data, which may not directly correlate with actual species abundance in the environment. Future research should focus on developing robust methods to correlate read abundance with actual species abundance.

4. Role of rare species in benthic diatom communities: patterns, processes and networks

This chapter is a modified version of the research article:

Stenger-Kovács, C., Korponai, J., **Abubaker, B.** *et al.* Role of rare species in benthic diatom communities: patterns, processes and networks. *Biodivers Conserv* **34**, 2095–2118 (2025). (Q1, impact factor: 3.1)
<https://doi.org/10.1007/s10531-025-03063-4>

4.1. Introduction

Rare species form an essential part of the biosphere (Jousset et al. 2017) and they are one of the key targets of conservation biology studies (Myers et al. 2000; Bonn and Gaston 2005). Although rare species can carry highly valuable ecological information, our current ecological knowledge, which can serve as the basis for further conservation strategies, originates mostly from the analysis of dominant species (Jousset et al. 2017). Specifically, rare species are often routinely removed from data-tables (e.g. Burgmer et al. 2007; Wang et al. 2018), as they are treated as noise or as distorting elements of multivariate analyses (Clarke and Green 1988), although this practice has been repeatedly criticized over the years (Goodall et al. 1969; Fore et al. 1996). The elimination of rare species, as a bad practice, might have a strong influence on ecological observations and theories and may compromise the sensitivity of community-based methods aimed at understanding ecological changes (Cao et al. 1998). To understand community structure (Cao et al. 1998) and ecological processes rarity cannot be neglected (Padisák et al. 2010), therefore an intensification of studies is urgently needed (e.g. Busseni et al. 2020).

Rare species (with limited population size and/or distribution) may have lower genetic diversity and narrower niche breadths than common species. Consequently, they are less able to adapt to changing environments (e.g. habitat destruction and loss, pollution, invasion of alien species; Fischlin et al. 2007; Vincent 2017) and are more vulnerable to extinction (Gaston 2008). Rare elements of metacommunities are exposed to neutral processes such as ecological drift caused by demographic stochasticity. Moreover, they have a limited probability of colonizing new habitats due to their small numbers and low reproduction rates (Iop et al. 2024). Based on their frequency-dependent fitness, rare species are more persistent, but they pay a price: they remain rare (Yenni et al. 2012). Indeed, remaining rare rather than going extinct may be a viable strategy for these “uncompetitive” species (Jousset et al. 2017). However, in a changing environment rare species that are considered irrelevant can reactivate and fulfill new and important compositional and functional key roles at the community level within ecosystems (e.g. carrying necessary traits and contributing to novel interspecific relationships; Fetzer et al. 2015; Jousset et al. 2017). Under favorable conditions these formerly rare species can therefore become dominant (Aanderud et al. 2015). In this way they can ensure the resilience and resistance (Tardy et al. 2014; Yan et al. 2015) of the community against environmental changes and the establishment of invasive species (Jousset et al. 2017). They occupy key niches and slow the establishment of neophytes, thus preserving ecosystem

stability, which was affirmed by experimental studies (Van Elsas et al. 2012; Vivant et al. 2013) because the elimination of rare species from the communities increased the likelihood of exotic species establishment.

The term 'rarity' serves as an 'umbrella' to describe different types of distribution patterns at various levels (Jeliaskov et al. 2022). There are a large number of approaches to identify rare species (Flather and Sieg 2007) but most are based on the species abundance (e.g. Tornés et al. 2021) and frequency of occurrence (e.g. Cao et al. 1998). Among these definitions, the best-known and most frequently used are Gaston's (1994) quartile criterion, which separates the lower quartiles in purely mathematical terms; and Rabinowitz's (1981) traditional scheme, which combines geographic range with habitat specificity and population size. The criteria used to define rare species can significantly influence the results of community analyses (Sgarbi and Melo 2018). It is also recognized that 60% of studies on rarity focus on only one dimension, and very few (18–24%; Maciel 2021) consider more, typically two or three dimensions (Espeland and Emam 2011).

Rare species may have a greater role in communities than expected simply from their relative abundance. The "rare species paradox" highlights a key point: these silent species require the greatest protection, yet we know the least about them (Lomba et al. 2010). Therefore, there is an urgent need to focus on rare species within ecosystems (e.g. Mouillot et al. 2013; Zizka et al. 2018) to discover their role in community structure and ecological processes (Allan, 2011) across extended spatial scales (Zizka et al. 2018) and dimensions (three axis) of rarity, applying the Rabinowitz' scheme (Maciel 2021).

For these reasons, we have set the following goals:

- (i) investigating rare benthic diatom species to discover the rarity types and patterns in a Central European river network system using standard sampling methods (CEN 2002) based on recently published rarity indices (Maciel et al. 2021);
- (ii) analyse the contribution of rare species to community variability and
- (iii) to beta-diversity;
- (iv) determining their role in community networks compared to common taxa and;
- (v) identifying the weaknesses of studies focusing on rare species.

4.2. Materials and Methods

4.2.1 River network system, sampling and analyses of benthic diatoms

For this study we selected the River Danube and its catchment basin as the second largest river network of Europe with historical, social and economic importance, where we have witnessed considerable water quality and ecological changes in the last decades (e.g. Abonyi et al. 2018, 2020; Mănoiu and Crăciun 2021; Buczkó et al. 2022). Altogether 100 sampling sites in two regions were selected in this river network system: 50 sites in Hungary and 50 in Croatia (Fig. 7, Supplement Table 1). The sampling sites differ in land use types, human inputs (nutrients, sediments and other contaminants), and consequently, in ecological status. Sampling was carried out in spring and autumn, 2022 (Supplementary Table 1). In a river network system, the river continuum ensures the possibility of easy spread of diatom species, with which we can exclude the effect of migration barriers in the developed community patterns and ecological processes.

The climate of Hungary is influenced by the oceanic, continental and Mediterranean climate effects therefore it is highly variable, and this is compounded by the diversity of topography. Most parts of the research area fall within the 9–11 °C mean temperature range. Due to the influence of the Mediterranean Sea, the annual precipitation varies between 500 and 750 mm on average. However, the year of sampling (2022) was the third warmest (higher with 1.45 °C than the climate normal value between 1901 and 2022) and 17th driest year (the country average was 497 mm with extreme distribution in space and time) in the last 122 years. Croatia's climate is determined by its location in the northern temperate latitudes. The most important factors influencing the climate in Croatia are the Adriatic and the Mediterranean, the Dinarides, the openness of the north-eastern regions to the Pannonian Plain and the diversity of terrestrial vegetation. There are therefore three main climate zones in Croatia: continental, mountain and coastal climate.

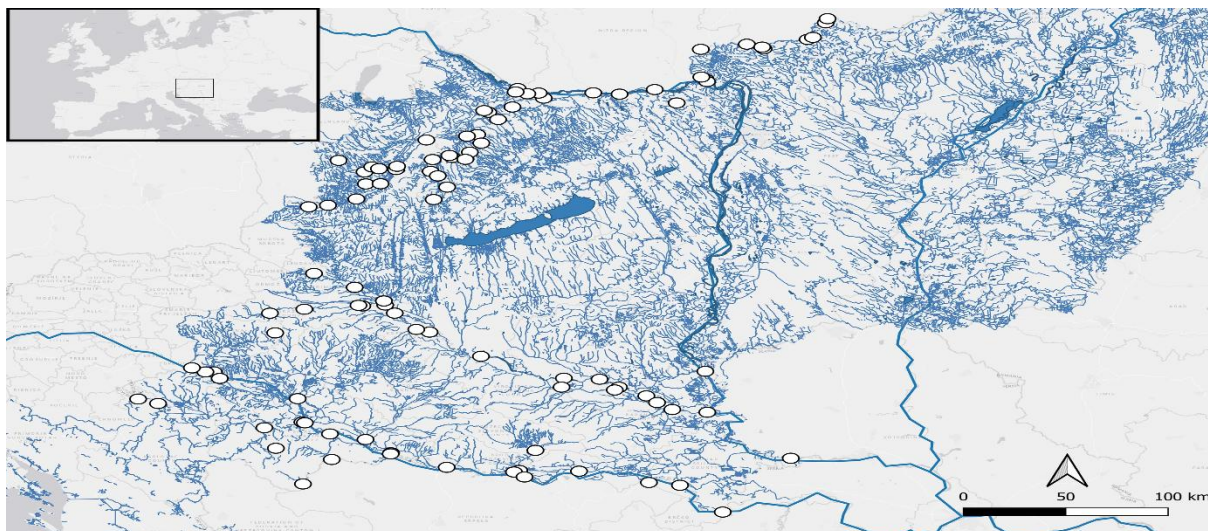


Figure 7. The investigated 100 sampling sites of the River Danube network system in Hungary and Croatia

The research stations in Croatia extend along the continental rivers with different average annual temperatures and precipitation. From colder and rainier areas (Kupa avg. an. temp. range 8.8–11.1 °C; avg. an. precipitation range 1112–2255 mm) to warmer and drier areas, depending on the average annual temperature and precipitation (Danube Ilok average annual temperature: 12.4 °C and average annual precipitation: 658 mm). In the area of the large rivers Sava and Drava, the average annual temperature range was between 11.2 and 12.3 °C and 10.6–11.4 °C, while the precipitation was 756–917 mm and 677–876 mm, respectively (data were provided by the Croatian Meteorological and Hydrological Service). The other investigated tributaries analyzed in this study had similar temperature and precipitation regimes within the specified range for the certain area. Deviations of the average air temperature in 2022 in Croatia, compared to the temperature between 1981 and 2010 were up to 2.1 °C higher. At all stations, the air temperature was higher than the average for this region and thermal conditions are described as extremely warm. Differences in the amount of precipitation in 2022 compared to the normal value of 1981–2010 were also observed but the analysis of the deviations in the amount of precipitation for the year 2022 shows that the precipitation amounts at most of the analyzed stations were in the range or below the annual average and in general the precipitation conditions in Croatia in 2022 are defined as normal.

Benthic diatom samples were taken from rock substrates with a toothbrush and preserved in ethanol. Altogether 183 samples were adequate for subsequent laboratory processing and analyses. Oxidation treatment of phytobenthos followed the hot hydrogen-peroxide method (CEN 2002), diatom valves freed from organic materials were embedded in

Pleurax© synthetic resin. Diatom taxa were identified to species level using up-to-date taxonomic guides (Lange-Bertalot et al. 2017; Bey and Ector 2013; Peeters and Ector 2017–2019; Ács et al. 2023). Taxonomic identification was made by two diatom experts of the same research group, who focused on different areas of the catchment basin. The continuous intercalibration between the two persons ensures the reliability and comparability of the diatom data (e.g. Kusel-Fetzmann et al. 1998). To determine the relative abundance of the species 400 valves per slide were counted in a light microscope (Zeiss Axio Imager A1, Plan-apochromat DIC lens).

4.2.2. Classification of rarity and commonness

Rabinowitz's approach (1981) has been applied for the classification of rare species in benthic diatom communities, stating that population size, habitat specificity and geographic range can be the criteria of separating rare species from common ones. Following this approach to identify rarities we applied Maciel's rarity indices (2021): GRI (geographical range index: inverse of the sum of the two extreme latitude and longitude range, where the species was found), HSI (habitat specificity index: inverse of the maximum habitat number, where the given species was found), PSI (population size index: inverse of the maximum population size of the given species found anywhere) as the three axis of rarity, and their combination RR (the median of GRI, HSI and PSI defines the absolute rare species) in the three-dimensional space. The index values range between 0 and 1 in case of all indices, where the closer the value to one, the rarer is the species. In case of PSI, the index value reaches one if the species was present with 4 individuals in the counted part of a sample, and we kept this maximum value if the number of individuals was less than 4, following the original scheme of Maciel (2021). After this identification and selection process, five subcommunities were separated and applied for further analyses, namely: subcommunity of (i) habitat specialists (HS), (ii) geographically restricted species (GR), (iii) small population species (SP), (iv) absolute rare species (AR) and (v) common species (C) as a complementary of the previous rare subcommunities. The complete species list with corresponding rarity categories was added to the paper as supplementary material (Supplement Table 2).

4.2.3. Statistical analyses

To avoid the effect of the abundance data of common species Sørensen dissimilarity based on binary data was calculated for the five subcommunities in pairs (rare species in different dimensions compared to common species) to discover the role of rare and common species in the community variability. The ratio of the Sørensen dissimilarities was applied to determine

the relative contribution of the subcommunities (rare and common) to the total community variability following the method of Shade et al. (2014). The Kruskal-Wallis test was used to compare whether the differences between the groups were significant. SDR (Similarity, richness Difference and Replacement) simplex method (Podani and Schmera 2011; Podani et al. 2018) was applied for the subcommunities to explore beta-diversity and its component highlighting the inherent structure of the metacommunity pattern.

Species co-occurrence network (SPIEC-EASI; SParse InversE Covariance Estimation for Ecological Association Inference; Kurtz et al. 2015) was established to characterize the structures of benthic diatom communities and to highlight the role of rare and common species in the association network. Prior to the network analysis, diatom abundance data were Hellinger transformed. During network computation, Meinshausen-Buhlmann's neighborhood selection was applied, and covariance selection was based on Stability Approach to Regularization Selection (StARS) with $\lambda = 40$. Closeness centrality was applied to making decisions between core and peripheral species. Closeness centrality measures how many steps are required to access every other node from a given node. Therefore, we assumed that the highest number of steps from the core species needed to reach the other species, while the periphery would take the lowest number of steps. Closeness values of 1.5 SD were applied as a threshold to establish core and peripheral species. Species with higher closeness than $\text{mean} + 1.5 \text{ SD}$ were considered as core, while species with $\text{mean} - 1.5 \text{ SD}$ were peripheral species. We characterized the community network by the mean degree of the nodes (representing the number of links of a node to other ones), closeness centrality (the number of steps required to access every other vertex from a given vertex) and betweenness (the number of the shortest paths going through a vertex or an edge) (Freeman 1979). All statistical analysis were done in R statistical environment (R Core Team 2024) using the “vegan” (Oksanen et al. 2022), “adespatial” (Dray et al. 2023), “spiec-easi” (Kurtz et al. 2024) and “igraph” (Csárdi et al. 2024) packages.

4.3. Results

4.3.1. Types of rarity in benthic diatom communities

A total of 409 diatom species from 84 genera were identified in the 183 samples. The number of the rare species significantly increased with the number of species found in samples (Fig. 8a). Besides, the number of the common species exceeded the number of rare species in all samples (Fig. 9). Most of the rare species originated primarily from genera represented by only

low species number (< 11) (Fig. 8b). Based on the calculated rarity indices, diatom species in the river network system of the Danube can be characterized by high population size index value, but low habitat specificity index value and low geographic range index (Fig. 10a). About 50% of benthic diatom species (204) were rare in certain terms (Supplementary Table 2). Most of them, 193 species, were rare based on their small population size, 91 were habitat specialists and 96 were geographically-restricted species (Fig. 8b, Supplementary Table 2). Most of the small population sized species (82) were rare based on even the other two axes, defined as absolutely rare taxa (representatives of e.g. *Amphipleura*, *Aneumastus*, *Anomoeoneis*, *Cymboppleura*, *Mastogloia*, *Placoneis* and *Stauroneis* genera). From the point of view of the population size 108 species were rare. Two were only geographically restricted (*Amphora minutissima*, *Fragilaria austriaca*) and there were no any species in the category of pure habitat specialist and in the combination of habitat specialist and small population size. Nine species (*Aulacoseira alpigena*, *Encyonema reichardtii*, *Eunotia* sp., *Fragilaria perminuta*, *Gomphonema angustatum*, *Gomphonema ristovskae*, *Navicula oligotrappenta*, *Navicula rhynchotella*, *Pantocsekiella costei*, *Sellaphora* sp.) were geographically restricted and habitat specialists, and three species (*Fragilaria famelica*, *Sellaphora seminulum* and *Placoneis clementoides*) were geographically restricted with small population sizes (Fig. 10b, Supplementary Table 2).

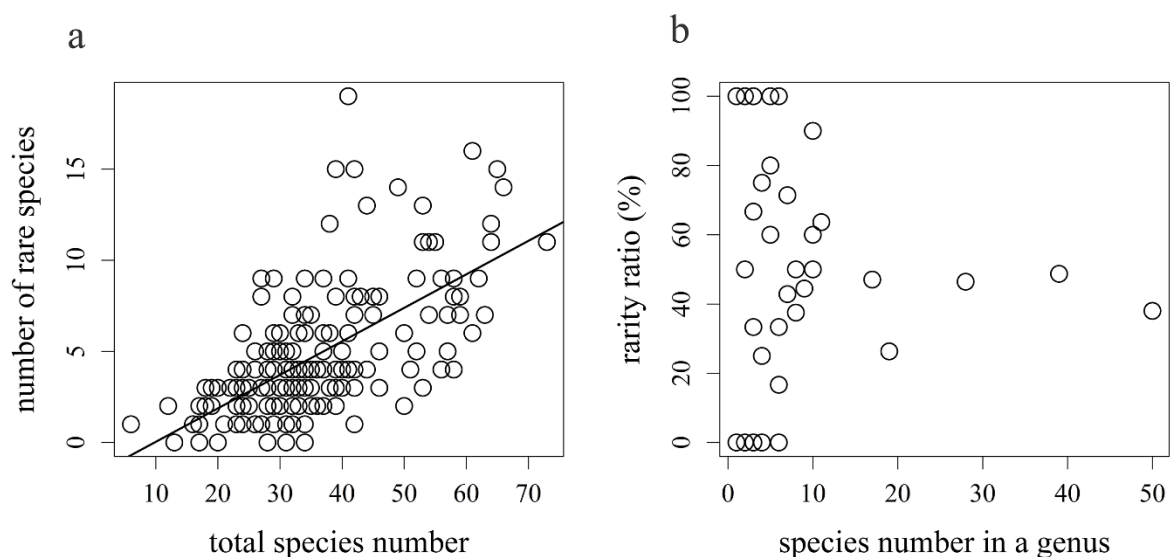


Figure 8. (a) Trends in the number of rare species as the number of species increases in the samples (b) the rare species number in the genera with different species richness

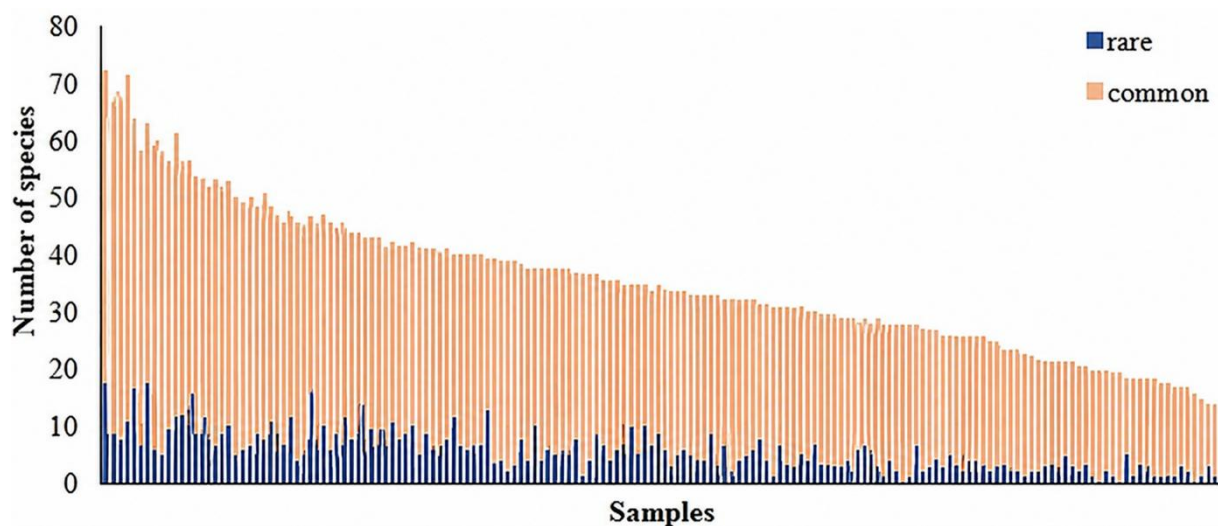


Figure 9. The proportion of rare and common diatom species per sample

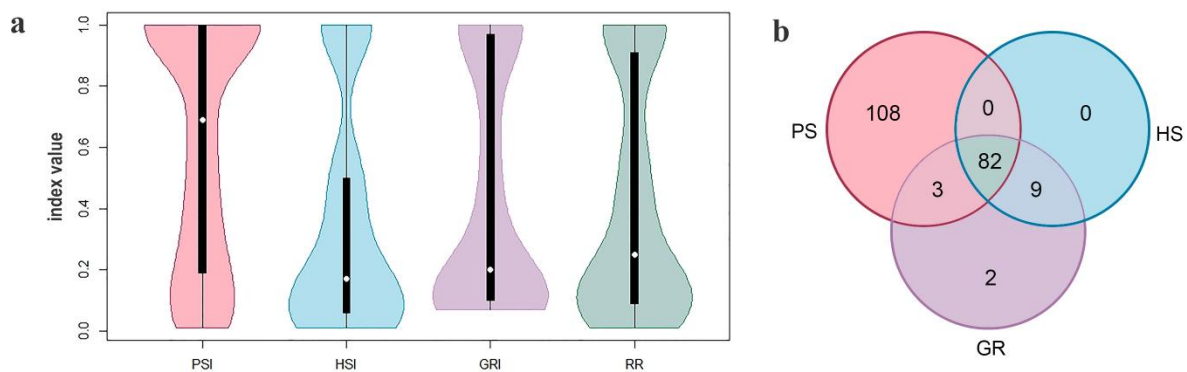


Figure 10. (a) Violin plots of rarity indices, (b) species number in the different rarity categories and in their combinations (PSI=population size index, HSI=habitat specificity index, GRI=geographical range index, RR=median of GRI, HSI and PSI; SP=small population species, HS=habitat specialists, GR=geographically restricted species).

4.3.2. Role of rare species in the variability of communities

Sørensen dissimilarities of the subcommunities of each rarity type are significantly higher than those of common species (Fig. 11. a, b, c, d). The variability of the common species' subcommunity is really close to the whole communities' (their ratio is ~ 1 ; Fig. 11. e, f, g, h), while the variability of the subcommunities of the rare species is higher than the whole community containing all of the species (their ratio is ~ 1.5 ; Fig. 11. e, f, g, h). Namely, the contribution of rare species to the total community variability is significantly higher than the common species' (Fig. 11. e, f, g, h), especially in the case of SP and GR with $\sim$ plus 50%, while this contribution is $\sim$ plus 40% in the case of AR and $\sim 25\%$ in case of the HS.

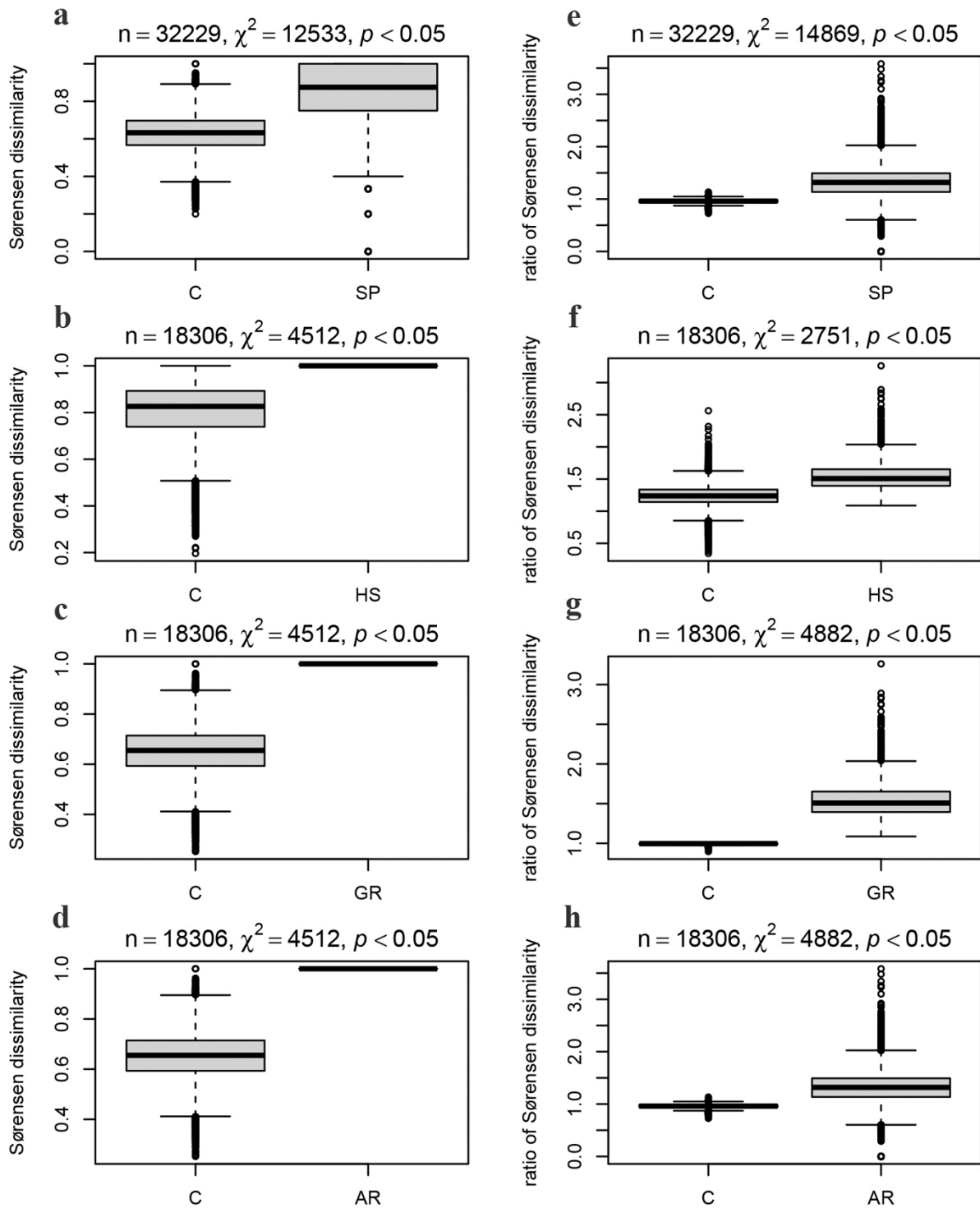


Figure 11. Sørensen dissimilarities of the subcommunities of (a) small population species (SP); (b) habitat specialist species (HS); (c) geographically restricted species (GR) and (d) absolutely rare taxa (AR) compared to common species' (C). The contribution of (e) SP, (f) HS, (g) GR and (h) AR to the total community variability compared to common species' (C).

4.3.3. Beta diversity of rare and common species

According to the pairwise comparisons, the variance (84.7%) of the SP is high due to the richness difference and replacement component of the beta-diversity (Fig. 12. a). According to the analyses, 48% of the species pool is continuously replacing and 36.6% of these species occurred only in one sampling site. The nestedness of small population size species is 51.9%, where the similarity component is low (15%) (Fig. 12. a). In the other three cases of rarity (HS, GR and AR) the variance of these rare species is extremely high (100%) due to the richness difference (20.3%) and replacement (79.7%) component of the beta-diversity (Fig. 12. b), while the similarity component of these submetacommunities is zero. In contrast, in the subcommunities of common species (as the complementary set of SP, HS, GR and RR [signed by SP C, HS C, GR C, RR C]) the part of the richness difference of the beta-diversity was lower (19%) than SP's ones (36.6%) but similar to HS, GR and AR's (20%). The 35–37% of the common species pool is responsible for the similarity of the subcommunities. However, the role of the replacement component is lower (44 and 45%) (Fig. 11.c, d) than that of rare species (Fig. 11a, b).

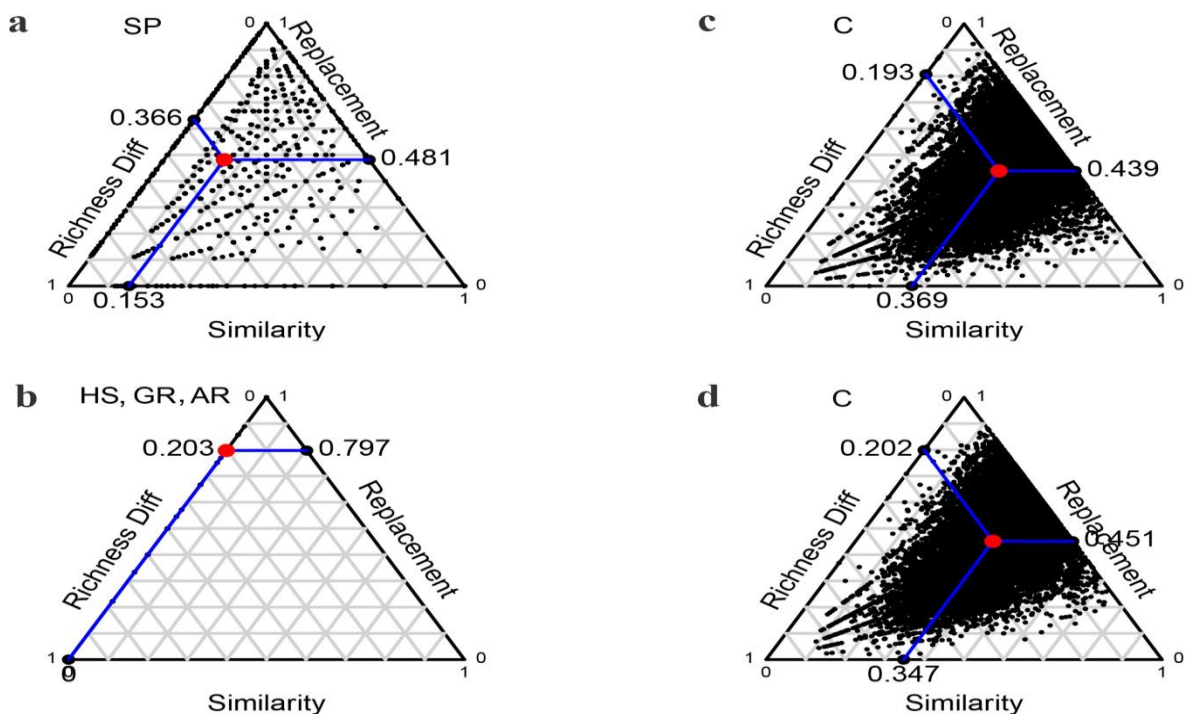


Figure 12. Beta diversity and its components of the subcommunities of (a) small population species (SP); (b) habitat specialist species (HS), geographically restricted species (GR) and absolutely rare taxa (AR) compared to c-d) their complementary common species (C)

4.3.4. Network analyses of diatom communities

A complex network with a number of connectivities was drawn for benthic diatom communities. The mean degree of nodes is 6.58. The mean closeness centrality is 0.0006, while the mean betweenness centrality is 584.9. Most of the species (356) in the communities have a neutral role. Most of the core species (total = 34) of the network are represented primarily by rare species (82% SP, 79% HR, 79% GR and 79% AR) e.g. *Amphipleura pellucida*, *Caloneis silicula*, *Mastogloia smithii* and *Placoneis gastrum*. Among the peripheral species (total number = 19) only HR, GR and AR species (e.g. *Gomphonema angustum*, *Gomphonema ristovskae*) can be found from the rare taxa, but they are less overrepresented (58%) (Fig. 13). Common species like e.g. *Amphora pediculus*, *Caloneis lancettula* and *Reimeria sinuata* are also in the peripheral part of the communities.

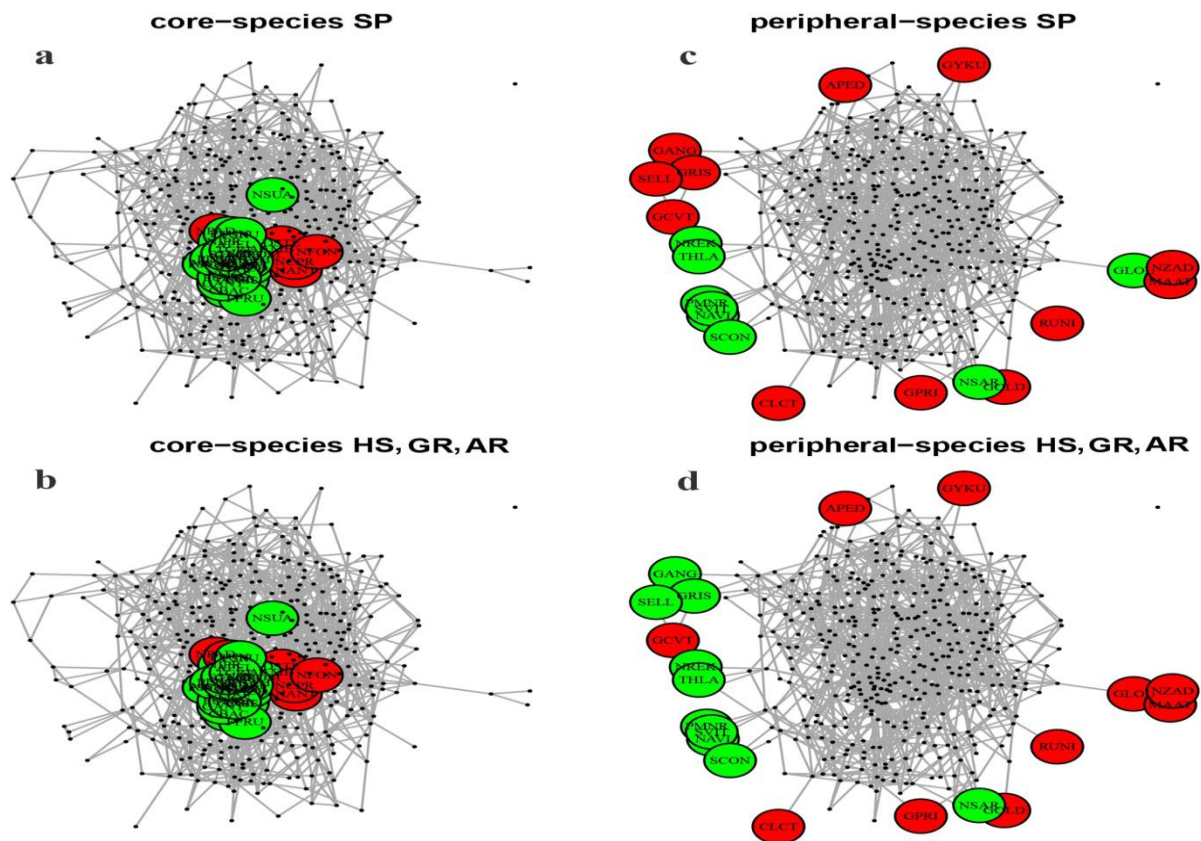


Figure 13. Distribution of rare (a) small population species (SP); (b) habitat specialist species (HS), geographically restricted species (GR) and absolutely rare taxa (AR) and c-d) their complementary common species' (C) in the benthic diatom community networks. (Green: rare, red: common species; The code of the species can be found in the Supplement Table 2)

4.4. Discussion

Benthic diatoms are a highly diverse group of biota (Arumugham et al. 2023) providing a number of ecosystem services (e.g. oxygen and primary production, sediment formation and stabilization, food for many other organisms, key role in nutrient cycling, habitat provisioning and climate control; B-Béres et al. 2023). Although so far even basic ecological information on rare diatom species is missing, they can be key elements of communities and ecological processes. This study has provided the first detailed results on the pattern of rare diatom species, their role in community structure, network and diversity. Investigation of rarity is particularly challenging in case of microorganisms that are not visible to the naked eye.

4.4.1. Rarity types and patterns of benthic diatoms in river networks

The “rare biosphere” represents an incredible huge part in the microbial communities (Busseni et al. 2020). Compilation of the Red List of diatoms it indeed seemed that number of rare taxa outnumber that of common ones (Cantonati et al. 2022). However, our study using standard protocols without “species-hunting” effort (discovering newer and newer habitats to find and describe new species or screening entire slides) did not support this experience, as common and rare diatom species were present in similar quantities (~ 50–50%) in communities on a regional scale. But in more species rich diatom samples the number of rare species was also higher, besides the fact that generally in a sample the species number of common species always considerably exceeded the number of rare species. This is consistent with what was found in case of amphipods and beetles: common species were presented in a higher proportion than rare species especially in species-rich aquatic communities (Bregović et al. 2019). These differences can originate from the sampling and analytical effort (Cermeño et al. 2014), which can be completely distinct if we (i) focus on the discovery of endangered species (Cantonati et al. 2022) or (ii) follow the standard protocols. These distinctive methods may shift rare species into the group of common species, but not vice versa (Bregović et al. 2019). In addition, the cutoff between rare and common species is always a bit arbitrary (Bregović et al. 2019).

Which genus provides the highest number of rare species? It can be assumed that the higher species number in a given taxonomic level can predict the higher probability of rare species, namely, in a species rich genus there can be more rare species. This assumption was equivocal for plants and insects (Ulrich 2005), but in case of diatoms we revealed an opposite phenomenon: primarily genera represented by < 11 taxa provided more rare taxa similarly to mammals and birds (Purvis et al. 2000).

Since habitat specificity, availability of suitable habitat(s), and dispersal capability will collectively drive evolution of potentially rare species (Kruckeberg and Rabinowitz 1985; Sgarbi and Melo 2018), we analyzed the structural rarity encompassing the three axes of rarity (population size, habitat specificity and geographic range; Murray et al. 2002). This approach provided ecologically meaningful analyses to detect the specialty of rare species which is generally masked under differentiation of “rare” and “common” species (Espeland and Emam 2011). Our results revealed that 95% of the rare species belonged to small population size taxa, of which 53% were generalist, widely distributed species in the region. Similar to running waters in Germany where small population species were found in different types of running waters (which may be ensured by their relatively rapid and easy dispersal; Mann and Vanormelingen 2013). This result is consistent with other observations reporting surprisingly wide distribution of a considerable part of the rare species supposing that they might be found virtually everywhere (Zizka et al. 2018). This is in line with diatom data in German running waters, where species with the lowest abundance occurred in several types of rivers (Hofmann et al. 2013). Only 31% of the species pool of this study was rare also in Germany (Hofmann et al. 2018) and this difference obviously originated from ecoregion dissimilarity. However, only 21% of the small population species can be found in the German species list (Hofmann et al. 2018), where the small population species (e.g. *Achnantheidium exile*, *A. gracillimum*, *Aneumastus stroesei*, *Aulacoseira pfaffiana* and *Diploneis parma*) were in different vulnerability categories (Hofmann et al. 2018). It should be noted in this context that the relative abundance of these species is not conditionally consistent with the status of the species in the German Red List.

However, many diatom species can affiliate to specific habitats (Cantonati et al. 2022). In our study, 45% of the rare diatoms were habitat specialists and 47% were restricted to a certain geographic area. Most of these specialized and localized species have small population sizes that defined them as absolutely rare taxa (members of e.g. *Amphipleura*, *Aneumastus*, *Anomoeoneis*, *Cymbopleura*, *Mastogloia*, *Placoneis* and *Stauroneis* genera) representing triple jeopardy from a conservation perspective. Based on the German Red List, species from *Cymbopleura* genus are definitely good examples for this threatened status (44 species are in endangered categories of the 55 listed species; Hofmann et al. 2018) but the members of the other genera have also low relative abundance and are restricted to only few types of aquatic habitats in Germany (Hofmann et al. 2013).

The geographically restricted and habitat specialist species assigned to streams (River Marcal, River Cinca, Szávai-channel in Hungary, River Kupa in Croatia) and river sections (Danube in Nagybjacs [Hungary], Danube in Basina and Ilok-most [Croatia]), whose protection can be outstandingly important. These sampling sites may offer high-quality special habitats providing refuges (DellaRocca et al., 2019) for rarities and maintaining important gene pools (Naughton-Treves et al. 2005). Therefore, habitat specialists should be preserved by (i) discovering why they are so unique and (ii) maintaining the properties of special habitat types, while for habitat generalists, we should ensure the configuration and connectivity of the patches (Pandit et al. 2009).

In terms of water quality, there are rather localized rare species (*Fragilaria austriaca*, *Placoneis clementoides*), which may indicate oligotrophic conditions (Krammer and Lange-Bertalot 1999; Lange-Bertalot, 2017) and high ecological status of the rivers. It must be noted that a significant part of the oligotrophic species appear in different categories of vulnerability in German water bodies (Hofmann et al. 2018). In case of these species, their rarity may have intrinsic causes due to their inherent biological and ecological characteristics (Flather and Sieg 2007). The maintenance of the good status of these rivers has a key importance in preserving these geographically restricted species, though we also found rare species, which are indicators of “bad” ecological status (like *Fistulifera saprophila*). It is common also in German rivers, where lots of eutrophic species are rare (Hofmann et al. 2018). If the water quality deteriorates, these species become common or even dominant in contrast with the general attitude that rare species have higher conservation value and are presumably threatened by extinction (Pimm et al. 2014). However, with adequate water management, i.e. we ensure the good ecological status according to Water Framework Directive [EC 2000]), we keep these species to be rare independently from their biology (Pärtel et al. 2005), simply because their distribution and abundance within this environment is restricted, since they are out of their ecological optimum ranges (Pärtel et al. 2005). Interestingly, there were some halophilic species (e.g. *Halamphora* species) assumed to be rare species in freshwaters, however, this time these species were already common through the river network system, which observation definitely indicates the phenomenon of the secondary salinization due to anthropogenic activities and climate change (Stenger-Kovács et al. 2023).

4.4.2. Community variability

Rare species, as less studied components of biodiversity, may considerably contribute to the diversity of microorganisms, as was clearly demonstrated in bacterial communities (Zizka et

al. 2018). Like rare bacteria, rare diatom species are the taxa that primarily maintain the variability of diatom communities as both small populations and geographically restricted diatom species are responsible for 50%, absolutely rare species for 40% and habitat specialists for 25% higher variability compared to the common species. These results also justify the impact of the rare biosphere (Sogin et al. 2006) based not on their biomass or abundance, but because they are the main components of the diversity (Busseni et al. 2020). They serve as a diversity reservoir of ecologically redundant species (Lynch and Neufeld 2015), which are like “a resting spore comes alive”, can respond fast to environmental changes and re-organise communities (Lynch and Neufeld 2015) thereby ensuring ecological resilience of the aquatic ecosystems. This means that with the protection of species diversity, we directly protect rare species (Jeliazkov et al. 2022), too. However, we must notice that both common and rare species are important components of the communities, but their relative contributions to the patterns of species diversity is questionable in many cases and has become a controversial issue among scientists (Gaston 2008; Šizling et al. 2009). A number of studies showed that common species determine the overall diversity patterns (e.g. Gaston 2010; Lennon et al. 2011). However, some other studies concluded the opposite results (Heegaard et al. 2013; Reddin et al. 2015). These different observations and conclusions can lead to different conservation strategies (protection of rare species or/and common species) to maintain biodiversity (Gaston 2010; Neeson et al. 2018). Therefore, the scientific community experiences these opposing views resulting in uncertainty, although these contradictory results might be derived from organism and habitat specificity of the studies.

4.4.3. Beta-diversity in benthic diatom metacommunities

Beta-diversity of rare diatom subcommunities is outstandingly higher than that of common species, which means that the diversity of diatom metacommunities is primarily determined by the neglected species. Although beta-diversity is driven mostly by species replacement both in case of rare and common species, it is higher for rare species, especially for habitat specialists, geographically restricted species and absolute rare taxa. This result is very similar to the case of terrestrial Amazonian tree species (Bispo et al. 2017) and aquatic invertebrates in ponds (Pandit et al. 2009), where common and rare species were exposed to different metacommunity processes. Notably, rare taxa were often reported to be habitat specialists (Bispo et al. 2017) and they responded first to environmental factors controlled by niche processes (Székely and Langenheder 2014; Liao et al. 2017; von Meijenfildt et al. 2023), followed the species sorting model (Pandit et al. 2009) as the species replacement occur along the environmental gradients.

However, given their inherently low abundances it can also be hypothesized that rare taxa might be regulated mainly by ecological drift which can also lead to the replacement of individuals or to stochastic loss (Jiao et al. 2017; Xue et al. 2018). However, in case of small population size the principal component of the beta-diversity was the richness difference, which highlighted the importance of e.g. species interactions or other ecological processes (Legendre et al. 2014). This suggests that even in the metacommunities of certain rarities divergent organizing processes can be characteristic.

4.4.4. Community networks

Microbial communities drive main ecosystem functions through a network built up by a wide range of species interactions (Alric et al. 2020). Network analyses of benthic diatom communities showed high and complex connectivities among species, which can ensure the network stability. Despite, in benthic diatom communities the number of common species equals that of rare species as was demonstrated above. Rare diatom species have a key role not only in maintaining variability and beta-diversity of communities, but also in the community structure. The seemingly unimportant diatom species along with the dominant ones were important attendees of the communities through their interspecific interactions similar to *Pinus massoniana* communities, where the community stability was determined by not only the dominant species but also the rare ones (Xiao et al. 2023). Furthermore, in the case of diatoms, they are not only important members, but 79–82% of the core species were “silent species” in the ecological network organization, which must be driven by the presence of rare species rather than their abundances. In contrast, well-known, common species (e.g. *Amphora pediculus*) have only a peripheral role in the community structure. Based on the similar number of rare diatom species in temporary and permanent streams, Tornés et al. (2021) supposed that these neglected species may provide the backbone and basic structural points of the communities, which has been justified now by this study. This phenomenon is also valid for bacterial communities (Jousset et al. 2017). These keystone species have a disproportionately large impact on the structure and functioning of an ecological community, consequently on ecosystems services (Dee et al. 2019), which was already demonstrated in other studies (Lyons et al. 2005; Jousset et al. 2017). According to Mouillot et al. (2013) 47% of fish species and 55% of tropical tree species with only one individual per sample on average likely support highly vulnerable functions. These results contradict the mass ratio hypothesis, that dominant species with high abundance are the base of the ecological processes (e.g. productivity) and drive the energy flow (Smith and Knaps 2003; Winfree et al. 2015). However, these less

abundant species even with key roles are more likely to disappear from the local- and regional species pools, than common ones (Flather and Sieg 2007). With their loss the overall species richness reduces which favours alien species to establish (Lyon and Schwartz 2001). Therefore, to focus on rarity and to maintain the structure of communities is a key issue. Furthermore, conservation of rare species even in highly diverse ecosystems has also a key importance (Mouillot et al. 2013), since these species as “ecological memory” will influence the present and future response of the community (Padisák 1992; Padisák et al. 2010).

4.5. Conclusions and future perspectives

In conclusion, we must step over the general use of species richness as metrics of biodiversity (Bregović et al. 2019). Using rarity (Crain and Tremblay 2014), beta-diversity (Yu et al. 2017) and the role of rare species in an ecological network can be an adequate strategy and key issue for understanding their ecology and conservation of aquatic communities and ecosystems. Rare species are the “hidden backbone” of the microorganisms (Jousset et al. 2017), which is especially true for benthic diatom communities. Therefore, we have to show self-consciousness to merge into research focusing on the rare biosphere (Jousset et al. 2017) to understand their distribution patterns (Jeliazkov et al. 2022) along the environmental gradients, to discover how they can buffer the changing environment as well as to determine their contribution to the ecosystem functions. There is an urgent need to answer these questions both by field observations, experimental or modeling studies by discovering new approaches and not insist on traditional statistical methods, which may be not applicable for rare species.

5. Using complementary approaches to study diatom diversity and its environmental drivers

This chapter is a modified version of the research article:

Abubaker, B., Stenger-Kovács, C. & Tapoleczai, K. Using complementary approaches to study diatom diversity and its environmental drivers. *Ecological Indicators*. Volume 187, 114970, ISSN 1470-160X. (2026). (**D1, impact factor: 7.4**). <https://doi.org/10.1016/j.ecolind.2026.114970>.

5.1. Introduction

Freshwater ecosystems are vital for global biodiversity, aesthetics, and economic value (Dudgeon et al., 2006). Freshwater biodiversity is facing severe threats from climate change, pollution, habitat degradation, and other anthropogenic pressures (Vörösmarty et al., 2010). Therefore, effective biomonitoring is essential for assessing ecological status and guiding conservation efforts (Smol and Stoermer, 2010; Stevenson and Smol, 2015). Benthic algae, especially diatoms, are widely used for bioassessment in aquatic ecosystems and are recognized as powerful bioindicators (Almeida et al., 2014; Rimet, 2012; Stenger-Kovács et al., 2007). In many contexts, diatoms alone can provide ecological status information comparable to that derived from the entire phytobenthos community, supporting their use as a reliable tool in water-quality assessments (Kelly, 1998; Kelly et al., 2009).

Diatoms are ubiquitous in aquatic environments and play crucial roles in primary production and nutrient cycling (Smol and Stoermer, 2010). Their sensitivity to environmental stressors including nutrient enrichment, organic pollution, acidification, and changes in hydrological conditions, makes them particularly valuable for evaluating water quality and ecosystem health (Hering et al., 2006; Smucker et al., 2022). Diatom-based indices are also integral components of regulatory monitoring programmes worldwide, including implementation under the European Water Framework Directive (WFD; Directive 2000/60/EC).

Traditionally, diatom-based biomonitoring has relied on morphological identification using light microscopy. This approach enables detailed species-level identification and has produced extensive historical datasets (Kwandrans et al., 1998; Potapova and Charles, 2002). However, morphological identification is time-consuming and requires high taxonomic expertise, which may vary among laboratories and lead to inconsistencies (Kahlert et al., 2009). In recent years, diatom metabarcoding has been increasingly applied for freshwater ecological quality assessment (Apothéloz-Perret-Gentil et al., 2021; Tapolczai et al., 2024; Wang et al., 2024; Abubaker et al., 2026) and has also been used in marine monitoring contexts (Bonačić et al., 2025). DNA metabarcoding, enabled by high-throughput sequencing, has emerged as a promising alternative or complement to morphology-based methods for assessing diatom diversity (Taberlet et al., 2018). This molecular approach offers several advantages, including faster sample processing, reduced dependence on specialist taxonomic expertise for routine identifications, and improved detection of cryptic or rare taxa that may be overlooked by

microscopy (Keck et al., 2022; Wang et al., 2024). Comparative evaluations generally indicate that integrating metabarcoding with traditional microscopy can improve the comprehensiveness of diatom identification and thereby strengthen freshwater ecological assessments (Zimmermann et al., 2015). Moreover, metabarcoding datasets can capture spatial and temporal patterns in diatom communities with fine resolution (Kutty et al., 2022).

At the same time, metabarcoding may miss some taxa that are detectable by conventional microscopy, particularly when reference libraries are incomplete; therefore, metabarcoding is best viewed as complementary rather than directly equivalent to traditional approaches (Keck et al., 2022). In addition, discrepancies between molecular and morphological inventories have been widely documented (Bailet et al., 2020; Duleba et al., 2021). Biases can also be introduced by molecular laboratory procedures, bioinformatic filtering steps, and limitations in available genetic reference data (Dulias et al., 2017). Although metabarcoding provides high-resolution sequence data, species-level assignments are often constrained by reference-library gaps, resulting in identifications at genus level or higher (Dulias et al., 2017; Kermarrec et al., 2014). By contrast, morphological analysis allows species-level identifications when diagnostic characters are visible and experts are available with high-level and up-to-date taxonomic knowledge. Importantly, diatom taxonomy is continually evolving with advances in microscopy and molecular systematics, leading to the description of new taxa and revisions of existing species concepts (Kahlert et al., 2009).

The accuracy and completeness of sequence reference databases are therefore critical, as incomplete databases limit taxonomic assignment and can reduce ecological interpretability (Dulias et al., 2017; Kermarrec et al., 2013). This highlights the need for standardized protocols and cross-laboratory validation to ensure the reliability and transferability of metabarcoding data (Vasselon et al., 2025).

Taxonomy-free approaches in diatom eDNA metabarcoding analyze sequence data without assigning reads to formal taxonomic names. Such approaches represent important progress in biomonitoring, particularly for relating community patterns to environmental variables and retaining ecological information that might otherwise be lost due to incomplete reference libraries (Apothéloz-Perret-Gentil et al., 2017). Similarly, Tapolczai et al. (2021) used exact sequence variants (ESVs) without taxonomic clustering to assess agricultural impacts on diatom communities, revealing cryptic ecological diversity within nominal species. Gregersen et al. (2023) compared taxonomy-free and taxonomy-assigned approaches for a diatom trophic-

level index and found that taxonomy-free indices predicted environmental variables more efficiently and produced lower average prediction error.

This study investigates community-level responses of benthic algae to environmental variation within a river network by comparing identification methods of (i) microscopic analysis of phytobenthos, (ii) microscopic analysis of diatoms and (iii) eDNA metabarcoding of diatoms. We expected the two diatom-based approaches to show clearer relationships with environmental variables than the broader phytobenthos microscopy dataset, because diatoms are the main group used in water quality assessment and they can be analysed with higher taxonomic resolution. Second, we expected diatom metabarcoding at the ASV level to detect more taxa within each sample and to show stronger links between community composition and environmental gradients than microscopy, because it can capture fine-scale variation and taxa. Third, we expected Shannon diversity to describe environmental patterns better than richness alone, because it includes both the number of taxa and their relative abundances, and therefore reflects community changes more fully across gradients such as nutrient enrichment, conductivity, and land use.

5.2. Materials and Methods

5.2.1 Sampling sites

A total of 130 samples were collected from 67 sites during two sampling campaigns: August–September 2018 (67 samples) and May–June 2019 (63 samples). The sites covered two catchments, the Marcal and Rába rivers, in the Central and Western Transdanubian region of Hungary (Fig. 14). Hungary's climate is influenced by oceanic, continental, and Mediterranean effects and is therefore highly variable. Most of the study area falls within a mean annual temperature range of 9–11 °C. Owing to Mediterranean influence, mean annual precipitation typically ranges between 500 and 750 mm.

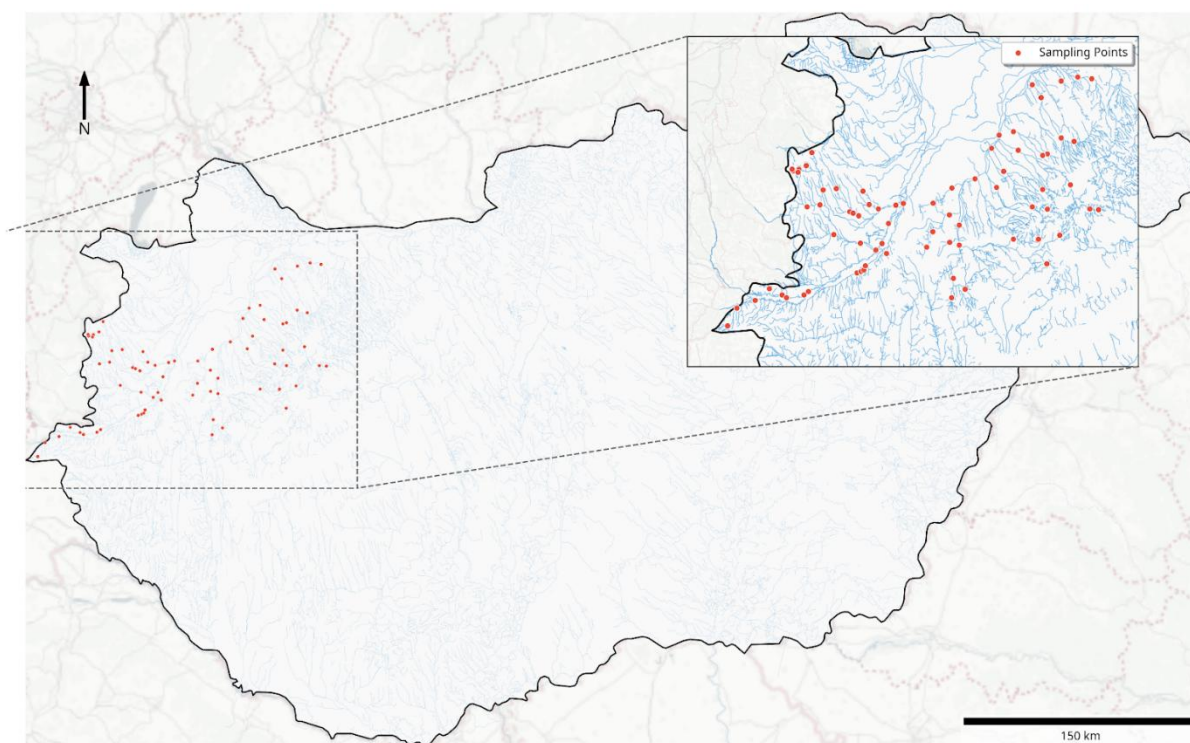


Figure 14. The location of the sampling sites (red dots) in the catchment areas of the rivers Raba and Marcal in western Hungary.

5.2.2 Environmental variables

Environmental parameters, including the extension and type of riparian vegetation, flow velocity, river width, conductivity, temperature, oxygen saturation, dissolved oxygen, and pH, were measured in situ using a Valeport Model 801 electromagnetic flow meter and an HQ40d Hach Lange multimeter. Water samples were collected at each sampling location for laboratory analyses. Parameters like nitrite (NO_2^-), nitrate (NO_3^-), ammonium (NH_4^+), soluble reactive phosphorus (SRP), total phosphorus (TP), chloride (Cl^-), soluble reactive silica (SRSi), carbonate (CO_3^{2-}), bicarbonate (HCO_3^-), and chemical oxygen demand (COD) were measured through titrimetric and spectrophotometric techniques, by standard protocols (APHA 2012; Wetzel and Likens, 2000).

5.2.3 Diatom sampling and laboratory procedure

Periphyton samples were collected at each sampling site according to European standards (CEN, 2018) from the most representative substrate (macrophytes, stones, or sediment). The upper surface of the substrates was scrubbed using a clean toothbrush and washed into a tray. For the sediment substrate, samples were collected by pipetting the uppermost layer. Three subsamples were collected at each site for the three different identification methods. After the

collected material was manually homogenized, one subsample was transferred into a 20 mL tube and preserved with Lugol's solution for the subsequent analysis of phytobenthos. Another subsample was transferred into a 20 mL tube and preserved with ethanol for the subsequent preparation of permanent diatom slides. A third subsample of ca. 5 mL was likewise transferred to a 20 mL tube containing 96% ethanol up to a final ethanol concentration of > 70%. All samples were stored at 4 °C in the dark until further analyses.

5.2.3.1. Microscopy analysis of phytobenthos

Benthic algal taxa were analysed according to the following: at least 400 settling units (cells, filaments, and colonies) were counted and identified in each sample using a Zeiss Axiovert 100 (Oberkochen, Germany) inverted microscope according to the classical methods developed by Lund et al. (1958) and Utermöhl (1958). Relative biomass was calculated based on cell volumes, considering their geometric forms, as described by Hillebrand et al. (1999). This calculation was performed using the Opticount cell counting software (Hepperle, 2008) and assuming a density of 1 g cm⁻³.

5.2.3.2. Microscopy analysis of benthic diatoms

The oxidation of phytobenthos was performed following the hot hydrogen peroxide method (CEN 2002), and the cleaned diatom valves were mounted in Pleurax© synthetic resin. Taxonomic identification to the species level was carried out using the latest identification keys (Lange-Bertalot et al., 2017; Bey and Ector, 2013; Peeters and Ector, 2017–2019). Relative abundances of diatom species were determined by counting at least 400 valves per slide under a light microscope (Zeiss Axio Imager A1, Plan-Apochromat DIC objective).

5.2.3.3. Molecular analysis of benthic diatoms

For DNA extraction, samples were again manually homogenized, and 2 mL of each sample was taken. The samples were then centrifuged at 13,000 g for 30 minutes at 4 °C. The resulting pellets were used for total genomic DNA extraction, following the Macherey-Nagel NucleoSpin® Soil Kit protocol, with a final elution volume of 30 µL. A fragment of 312 bp of *rbcL* gene, was selected as the gene marker for PCR amplification. Amplification was performed using an equally proportioned mixture of three forward primers (Diat_ *rbcL*_708F_1, Diat_ *rbcL*_708F_2, Diat_ *rbcL*_708F_3) and two reverse primers (R3_1, R3_2), as described by (Vasselon et al., 2017). Both forward and reverse primers were tagged with Illumina adapter sequences (5'-CTTTCCCTACACGACGCTCTTCCGATCT-3' and 5'-GGAGTTCAGACGTGTGCTCTTCCGATCT-3') for library preparation using a two-step

PCR approach. In the first PCR (PCR1), triplicate reactions of 25 µL were performed per DNA sample using the adapter-tailed primers and Takara LA Taq® polymerase. The triplicates were subsequently pooled and transferred to the GenoToul Genomics and Transcriptomics Platform (GeT-PlaGe, Auzeville, France), where a second PCR (PCR2) was conducted using the purified PCR1 products as templates. During PCR2, dual-indexing was added using Illumina-tailed primers. Finally, an equimolar pool of the 130 dual-indexed PCR2 amplicons was prepared and sequenced on an Illumina MiSeq platform using a V3 paired-end sequencing kit (2 × 250 bp).

5.2.4. Bioinformatic pipeline

Bioinformatic processing of the demultiplexed MiSeq reads was conducted using the DADA2 pipeline (Callahan et al., 2016), with a version specifically adapted for diatom metabarcoding data (available at https://github.com/fkeck/DADA2_diatoms_pipeline). Primer sequences were removed from both R1 and R2 reads using Cutadapt v2.9 (Martin, 2011). Quality profiles of the reads were examined, revealing high overall read quality. To eliminate low-quality regions, forward and reverse reads were truncated to 200 and 170 nucleotides, respectively. Sequences were filtered to exclude reads containing ambiguous bases (“N”) and those exceeding a maximum expected error rate (maxEE) of 2.

An error model was generated, indicating a strong fit between observed and estimated error rates, with error rates decreasing as sequence quality increased. Dereplication of R1 and R2 reads was performed to produce identical sequence units (ISUs). Amplicon sequence variants (ASVs) were inferred using the DADA2 denoising algorithm, and forward and reverse reads were subsequently merged. Chimeric sequences were identified and removed. Read counts were tracked throughout each step of the pipeline, with final read distributions and associated sequences. Taxonomic assignment of ASVs was conducted using *diat.barcode* R package reference database (Rimet, 2024), following European standards for barcoding library management (CEN, 2018b; Tapolczai et al., 2019; Rimet et al., 2019). Read numbers were transformed to relative reads by dividing the number of reads of ASVs in each sample by the total number of reads in that sample.

Metabarcoding data were used in two ways throughout subsequent analysis. To compare taxa inventories and diversity among the identification methods, ASVs were clustered into the finest taxonomic level to which they are assigned, summarizing their relative abundances. For example, all ASVs identified to the species *Achnanthes minutissimum* (Kützinger) Czarnecki were merged and their relative abundances were added. Similarly, all ASVs identified to e.g.

the genus *Navicula* without species level identification, were merged and their relative abundances were added. However, to study the relationship with the environment, ASVs were treated separately as taxonomic units to maximize ecological information (Kahlert et al., 2022; Tapolczai et al., 2021; Apothéloz-Perret-Gentil et al., 2021).

5.2.5. Statistical analysis

5.2.5.1. Data preparation

For all three biological datasets (phytobenthos microscopy, diatom microscopy and diatom metabarcoding), count data were converted to relative abundances per sample. For cross-method comparisons of diversity and inventories, metabarcoding ASVs were aggregated to the finest available taxonomic level (species or genus), and their relative abundances were summed. Mean and standard deviation of the values of environmental variables were calculated in each sampling period (Supplementary Table 3). The distribution of environmental variables was checked before the analyses. Strongly skewed variables (turbidity, flow velocity, NH_4^+ , NO_3^- , PO_4^{3-} , SRSi, SO_4^{2-} , TP, COD and forest land areas) were $\log(x + 1)$ transformed, whereas variables with approximately normal distributions (riparian vegetation, macrophyte coverage, temperature, conductivity, oxygen saturation, pH, Cl^- , cropland land use, urban land use, shaded area) were standardised without transformation.

5.2.5.2. Comparison of biological inventories and diversity metrics

For each dataset, the total number of taxa, the range and mean number of taxa per sample, and the identity of dominant taxa were calculated. Overlap and uniqueness among methods were quantified at the genus and species levels using Venn diagrams. For the metabarcoding dataset, we further summarised the proportion of ASVs and reads assigned to each taxonomic rank and the distribution of ASV numbers per species. Taxa (or ASV) richness and Shannon diversity (H') were computed for every sample and for each identification method. Differences in mean richness and Shannon diversity among methods were visualised with boxplots and tested using one-way analysis of variance (ANOVA). When the overall ANOVA was significant, Tukey's honestly significant difference (HSD) test was used for pairwise post-hoc comparisons.

5.2.5.3. Abundance relationships between microscopy and metabarcoding

To assess consistency in abundance estimates, we focused on diatom taxa detected by both microscopy and metabarcoding in at least three samples. For each shared taxon, a Spearman

rank correlation coefficient was calculated between its relative abundances obtained by the two methods across all sites. The resulting correlation coefficients were grouped into bins of width 0.1, and their frequency distribution was displayed as a bar chart to summarise the overall level of agreement.

5.2.5.4. Environmental gradients

To assess the response of diversity metrics to the environmental pressure, we modelled the responses of richness and Shannon diversity of the three biological datasets to the first axis of the PCA (PC1) conducted on the environmental variables. These axis values were used as the predictors in the models. We hypothesized a unimodal relationship between diversity metrics and PC1. To test this, quadratic regression models were fitted for each combination of the three datasets and the diversity metric. Model fit was assessed using Akaike's Information Criterion (AIC) to compare support across models. Adjusted R^2 and overall model significance were used to quantify the proportion of variance explained and the reliability of the relationship.

5.2.5.5. Co-inertia analysis of community–environment relationships

To assess congruence between the biological community compositions and the environment, three co-inertia analyses (CoIA) were conducted using the *coinertia()* function of the *ade4* R package (Dray and Dufour, 2007). The input for each of the CoIAs was two PCAs, one performed on one of the Hellinger-transformed community matrix (phytobenthos taxa, diatom taxa, diatom ASVs) and another performed on the standardised environmental data, using the *dudi.pca()* function of the *ade4* R package. Using the two PCAs, CoIA aims to find a linear combination and to maximize the covariance between the two matrices, revealing how their patterns are related. The test statistic of the CoIA is the co-inertia coefficient (RV), ranging from 0 (no relationship) to 1 (perfect covariation). The significance of the observed RV coefficient was tested with a Monte Carlo permutation test using 999 replicates and the *randtest()* function of the *ade4* R package.

5.3. Results

5.3.1. Biological inventories

5.3.1.1. Morphological identification of the phytobenthos

In total, the light microscopy identification of the phytobenthos using the Utermöhl method revealed 105 algae species and 109 genera. The number of taxa identified per sample ranged

from 1 to 47, with an average of 21. The ten most dominant phytobenthos species, in order, were: *Melosira varians* (C. Agardh), *Rhoicosphenia abbreviata* (Lange-Bertalot), *Achnantheidium minutissimum* (Czarnecki), *Nitzschia linearis* (W. Smith), *Ulnaria ulna* (Compère), *Trachelomonas oblonga* (Lemmermann), *Euglena sociabilis* (P. A. Dangeard), *Closterium moniliferum* (Ehrenberg ex Ralfs), *Nitzschia vermicularis* (Hantzsch), and *Cocconeis placentula* (Ehrenberg).

5.3.1.2. Morphological identification of diatoms

In total, 273 diatom species and 71 genera were identified by light microscopy. The number of species identified per sample ranged from 3 to 73, with an average of 33. The ten most dominant species, in order, were: *Cocconeis placentula*, *Amphora pediculus* (Grunow), *Achnantheidium minutissimum*, *Planothidium frequentissimum* (Lange-Bertalot), *Achnantheidium eutrophilum* (Lange-Bertalot), *Rhoicosphenia abbreviata*, *Navicula tripunctata* (Bory), *Nitzschia palea* (W. Smith), *Planothidium lanceolatum* (Lange-Bertalot), and *Navicula cryptotenella* (Lange-Bertalot).

5.3.1.3. eDNA metabarcoding of diatoms

A total of 5,867,971 reads of 2471 ASVs were obtained from sequencing after the bioinformatic pipeline had been performed. From the total 2471 ASVs, 894 (36.2%) and 1221 (49.4%) were identified to species and genus level, respectively. In terms of reads, 4,849,533 (82.6%) and 5,298,397 (90.3%) were identified to species and genus level, respectively. The proportion of assigned ASVs and reads across a range of taxonomic levels shows a clear trend of increasing assignment success as the taxonomic resolution broadens, culminating in 100% assignment for both ASVs and reads at the Empire level. Notably, the percentage of assigned reads consistently and substantially exceeds that of ASVs at every level, with the most significant difference occurring at the species level, with 82% for reads versus 36% for ASVs (Fig. 15). ASVs were taxonomically assigned to 222 diatom species and 75 genera. The number of corresponding ASVs per species followed a left-skewed distribution, with a majority of species being represented by one or two ASVs. Some species, however, were represented by several ASVs, e.g. *Achnantheidium minutissimum* (29 ASVs), *Mayamaea peritis* (K. Bruder & Medlin) (32 ASVs), *Sellaphora nigri* (Wetzel & Ector) (32 ASVs), and *Nitzschia palea* (41 ASVs). This corresponds to a range of the number of species identified per sample from 28 to 92, with an average of 60. The ten most detected species in all sites, in order, were: *Melosira varians*, *Cocconeis placentula*, *Achnantheidium minutissimum*, *Nitzschia linearis* (W. Smith), *Amphora*

pediculus (Grunow), *Navicula tripunctata*, *Nitzschia dissipata* var. *media* (Grunow), *Nitzschia palea*, *Planothidium lanceolatum*, *Ulnaria ulna*.

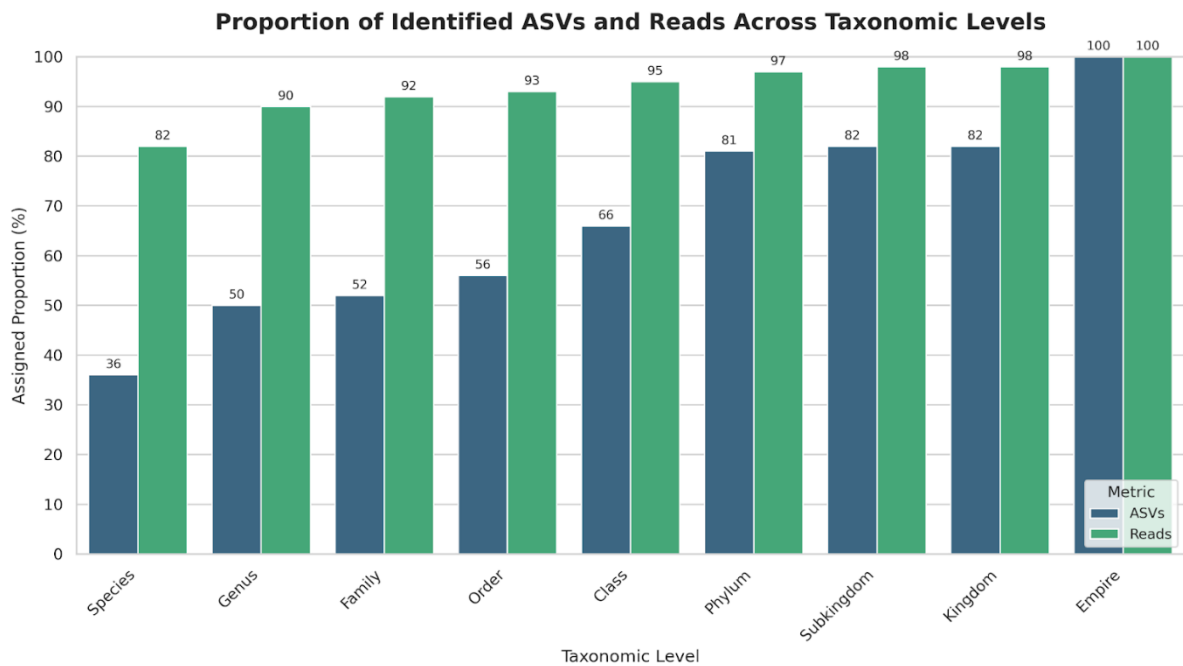


Figure 15. Proportion (%) of the successfully assigned ASVs (blue bars) and reads (green bars) to the different taxonomic levels.

5.3.1.4. Shared and unique taxa among the biological inventories

Metabarcoding of diatoms detected 22 genera (13.3%) that were not identified with microscopy. Microscopy of diatoms, in turn, recorded 16 genera (9.6%) that metabarcoding missed, while the broader algal dataset included 70 genera (42.2%) found only under the microscope. The overlap was limited: 32 genera (19.3%) were shared by all three approaches, and another 23 genera (13.9%) were common to metabarcoding and diatom microscopy (Fig. 16A). At the species level (Fig. 16B), metabarcoding identified 94 diatom species (21.2%) that were absent from microscopy, while diatom microscopy recorded 140 species (31.6%) not detected by metabarcoding. Both methods agreed on 100 species (22.6%), but only 28 species (6.3%) appeared in all three datasets. The microscopy analysis of the phytobenthos revealed 76 additional species.

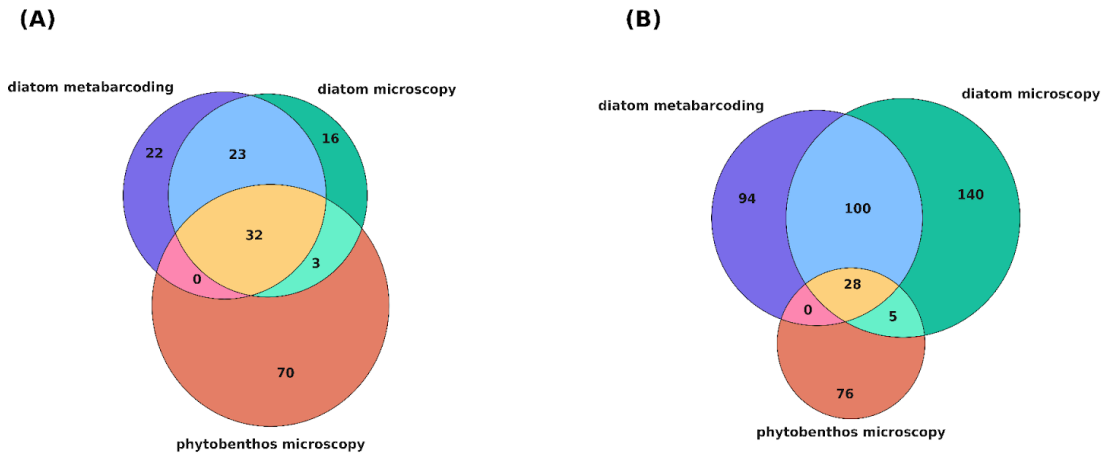


Figure 16. Number of taxa detected by the three different identification methods (A) at the genus taxonomy level, and at the species taxonomy level (B).

5.3.2. Diversity metrics compared among identification methods

Based on the data of the different methods, diversity metrics were estimated (Supplementary Table 4). For richness (Fig. 17A), diatom metabarcoding showed the highest median (60 taxa) per sample, diatom microscopy was intermediate (33 taxa), and phytobenthos microscopy had the lowest median (21 taxa).

For Shannon diversity (Fig. 17B), both diatom metabarcoding and diatom microscopy exhibited higher medians than phytobenthos microscopy, which has lower H' values. Mean Shannon values were 2.53 for diatom metabarcoding, 2.33 for diatom microscopy and 1.80 for phytobenthos microscopy.

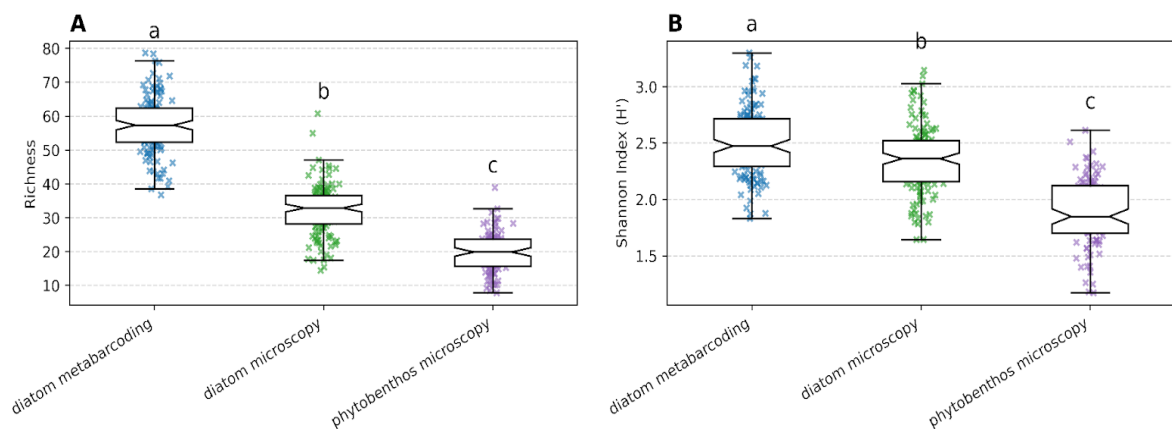


Figure 17. Taxa richness (A) and Shannon diversity (B) across the three methods: diatom microscopy, phytobenthos microscopy and diatom metabarcoding.

By comparing the three identification methods in both species' richness and Shannon diversity (Fig. 17A–B), we showed significant differences between the three approaches. Tukey's HSD tests revealed that all pairwise comparisons between methods were significant ($p < 0.05$).

5.3.3. Diatom species abundance analysis

For those taxa that were detected by both diatom microscopy and diatom metabarcoding methods in at least three samples, a Spearman rank correlation coefficient (r_s) was calculated between their relative abundances obtained by metabarcoding and by microscopy across all sites (Supplementary Table 5). The resulting r_s values were then grouped into bins of width 0.1 and plotted as a bar chart, with the x-axis showing the correlation bins and the y-axis showing the number of taxa in each bin (Fig. 18).

The top five diatom species showing strong rank-order agreement in abundance between metabarcoding and microscopy were *Cocconeis placentula*, *Amphora pediculus*, *Navicula tripunctata*, *Cocconeis pediculus*, *Eunotia bilunaris* (Ehrenberg). On the other hand, diatom species with weak or even slightly negative rank correlations were *Fallacia pygmaea* (Stickle & D.G. Mann), *Nitzschia sigmoidea* (W. Smith), *Nitzschia supralitorea* (Lange-Bertalot), *Fistulifera saprophila* (Lange-Bertalot & Bonik), and *Sellaphora seminulum* (D.G. Mann).

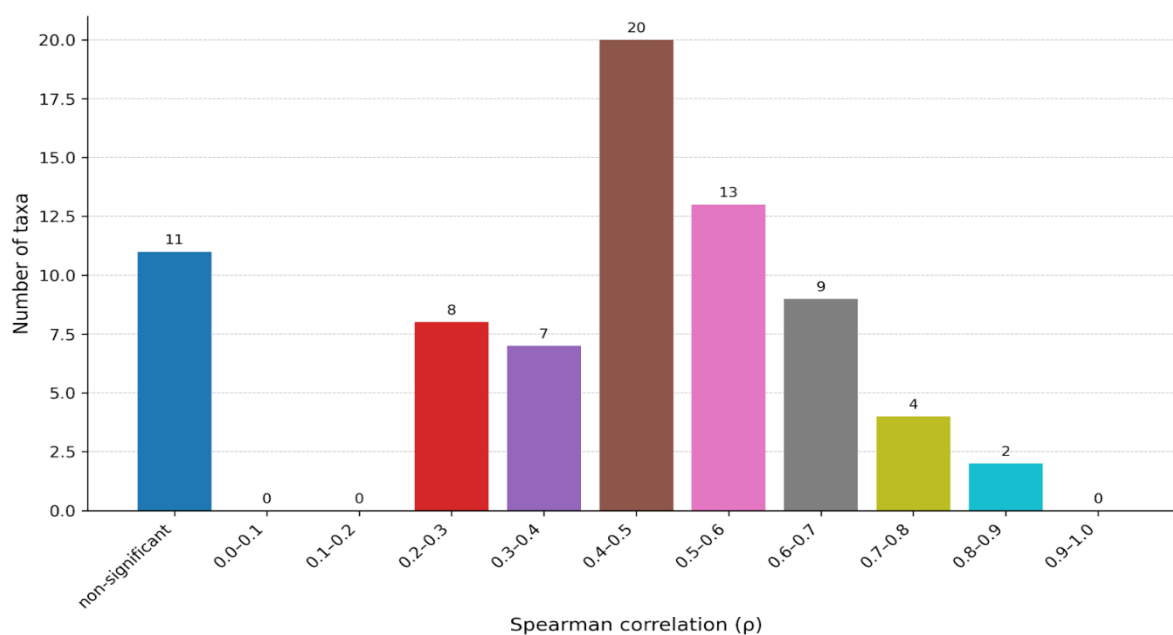


Figure 18. Distribution of correlation coefficients between morphological and metabarcoding abundance values for common diatom species.

5.3.4. Abiotic environment and its effect on algal communities

5.3.4.1. Environmental gradients

Main environmental gradients among the sampling sites and the two sampling seasons were studied through PCA (Fig. 19). The first three axes of the PCA explained 50.55% of the total variation among samples (PC1= 22.58%, PC2= 17.51%, PC3= 10.46%).

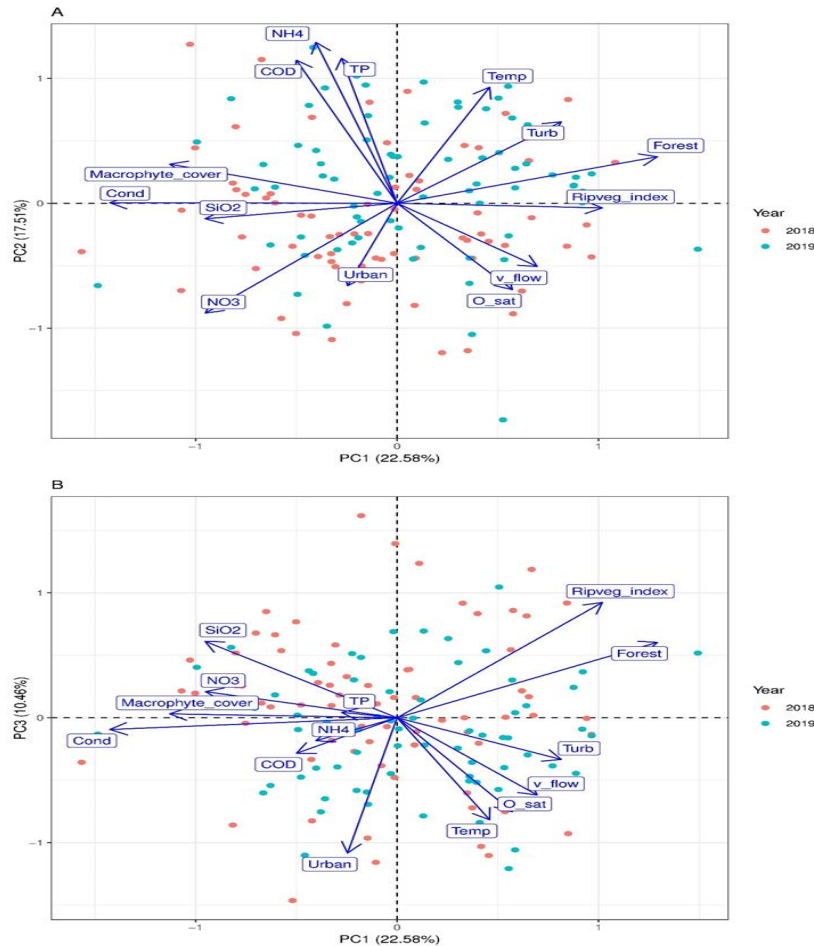


Figure 19. Principal component analysis biplots showing the distribution of samples from the two sampling seasons (2018 - red dots, 2019 - blue dots) based on the environmental variables (blue arrows), presenting axes PC1 and PC2 (A), and PC1 and PC3 (B).

The most important gradient (PC1) indicates a pressure of agricultural land use with associated variables related to salinity. Agricultural sites are characterized by elevated levels of conductivity, Cl^- , SO_4^{2-} and NO_3^- , reflecting higher inputs of e.g. fertilizers from surrounding croplands. In contrast, forested sites with extended riparian vegetation show markedly different conditions with low levels of the abovementioned variables. The second gradient (PC2) is associated with temperature, trophic state (TP), chemical oxygen demand (COD) and NH_4^+ , indicating trophic and organic pressure. The third axis shows a gradient of urbanization, associated with temperature as well. Although the PCA biplots do not show seasonal, we

detected significant differences in temperature, DO, turbidity, NH_4^+ , NO_2^- , NO_3^- , SRSi between the two sampling periods (Supplementary Table 3).

5.3.4.2. Response of community diversity metrics to environmental constraints

Both taxa/ASV richness and Shannon diversity, except for the Shannon diversity of the phytobenthos showed significant ($p < 0.05$) unimodal relationships with the main environmental gradient, which is represented by the first axis of the PCA (PC1) based on the environmental variables (Fig. 19). The strength of these relationships, however, varied among the three datasets. Model comparisons based on Akaike's Information Criterion (AIC) revealed that Shannon diversity consistently provided a better fit to the environmental gradient than taxa/ASV richness (Table 1). The lowest AIC value was obtained for Shannon diversity of diatom ASVs assessed by metabarcoding (AIC = 204.187, adjusted $R^2 = 0.10$, $p < 0.001$), indicating that this dataset most effectively captured the variation in Shannon diversity along the environmental pressure gradient. A weaker pattern was observed for phytobenthos Shannon diversity (AIC = 212.45, adjusted $R^2 = 0.06$, $p < 0.001$). In contrast, richness-based models showed much higher AIC values (977-1233) and weaker fits, even when their adjusted R^2 values were comparable or slightly higher in some cases.

Table 1. Comparison of quadratic models relating richness and Shannon diversity to the environmental gradient (PC1) across the three datasets. Lower AIC indicates better model support, adjusted R^2 and p-values indicate variance explained and model significance, respectively.

Dataset	Diversity metric	AIC	Adjusted R^2	p-value
Phytobenthos - microscopy	Richness	977.37	0.04	<0.05
	Shannon	212.45	0.06	n.s.
Diatom - microscopy	Richness	1000.75	0.02	<0.050
	Shannon	264.52	0.01	<0.050
Diatom - metabarcoding	Richness	1233.34	0.18	<0.001
	Shannon	204.3187	0.10	<0.001

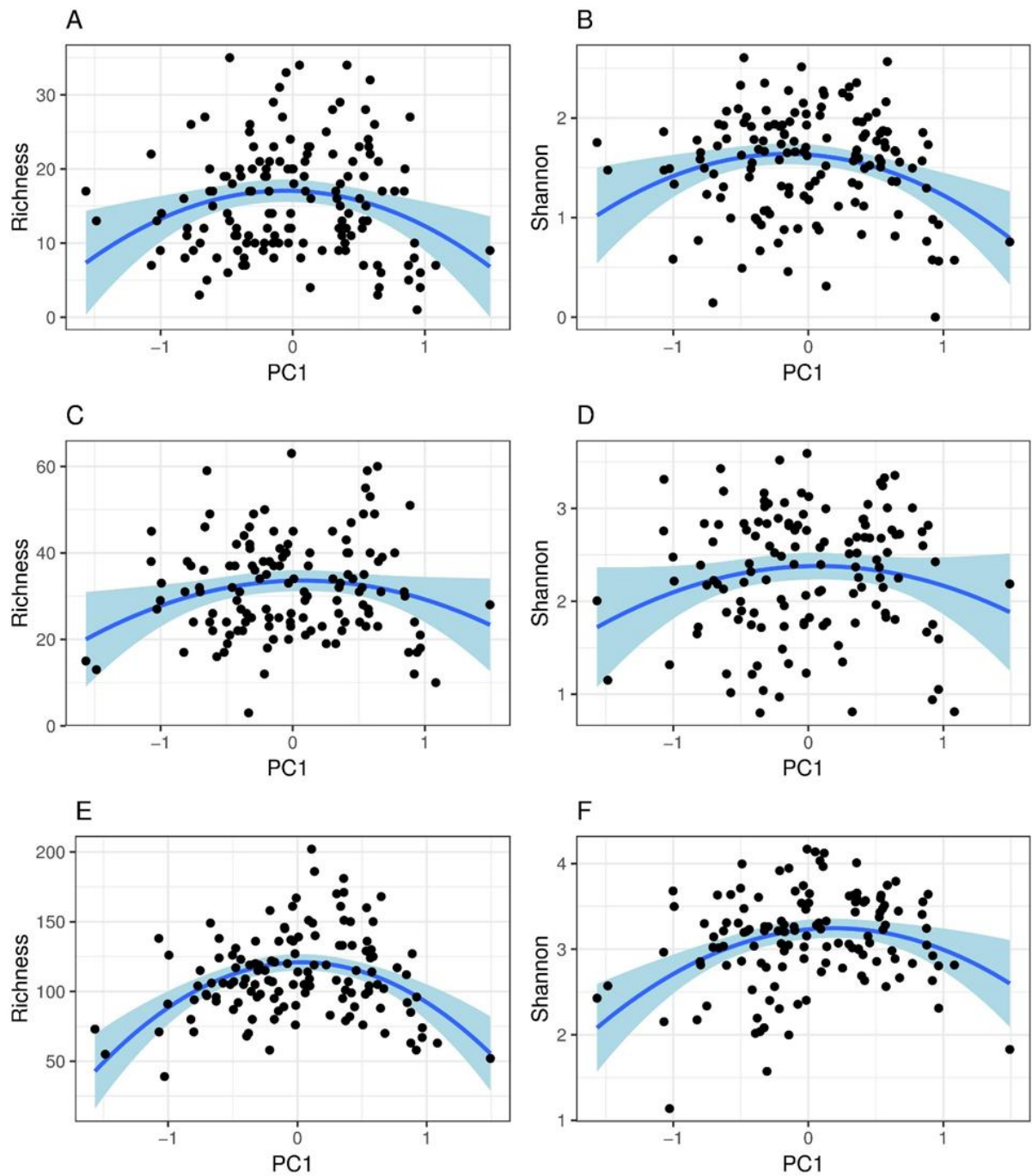


Figure 20. Relationships between diversity metrics and PC1 for the phytobenthos based on microscopy (A, B), diatoms based on microscopy (C, D), and diatom ASVs based on metabarcoding (E, F). Figures show the quadratic fit (blue line) with 95% confidence intervals (shaded area)

5.3.4.3 Response of community structure to environmental constraints

The co-structure between the community and the environmental data was significantly similar in the case of all three biological datasets ($p < 0.001$) according to the co-inertia analysis (Table 2, Fig. 21). The lowest RV value (0.19) was observed with the phytobenthos community dataset. At the same time, the co-structure was higher between the diatom community dataset identified with microscopy and the environmental variables (RV= 0.27). It was the highest between the diatom community identified with metabarcoding and the environmental variables (RV= 0.36).

Table 2. Co-inertia analysis results linking environmental variables with the three biological community datasets (phytobenthos, diatom-microscopy, and diatom-metabarcoding): RV coefficients and permutation test significance (p-values).

	Phytobenthos - microscopy	Diatom - microscopy	Diatom - metabarcoding
RV	0.19	0.27	0.36
p-value (Monte-Carlo test with 999 replicates)	<0.001	<0.001	<0.001

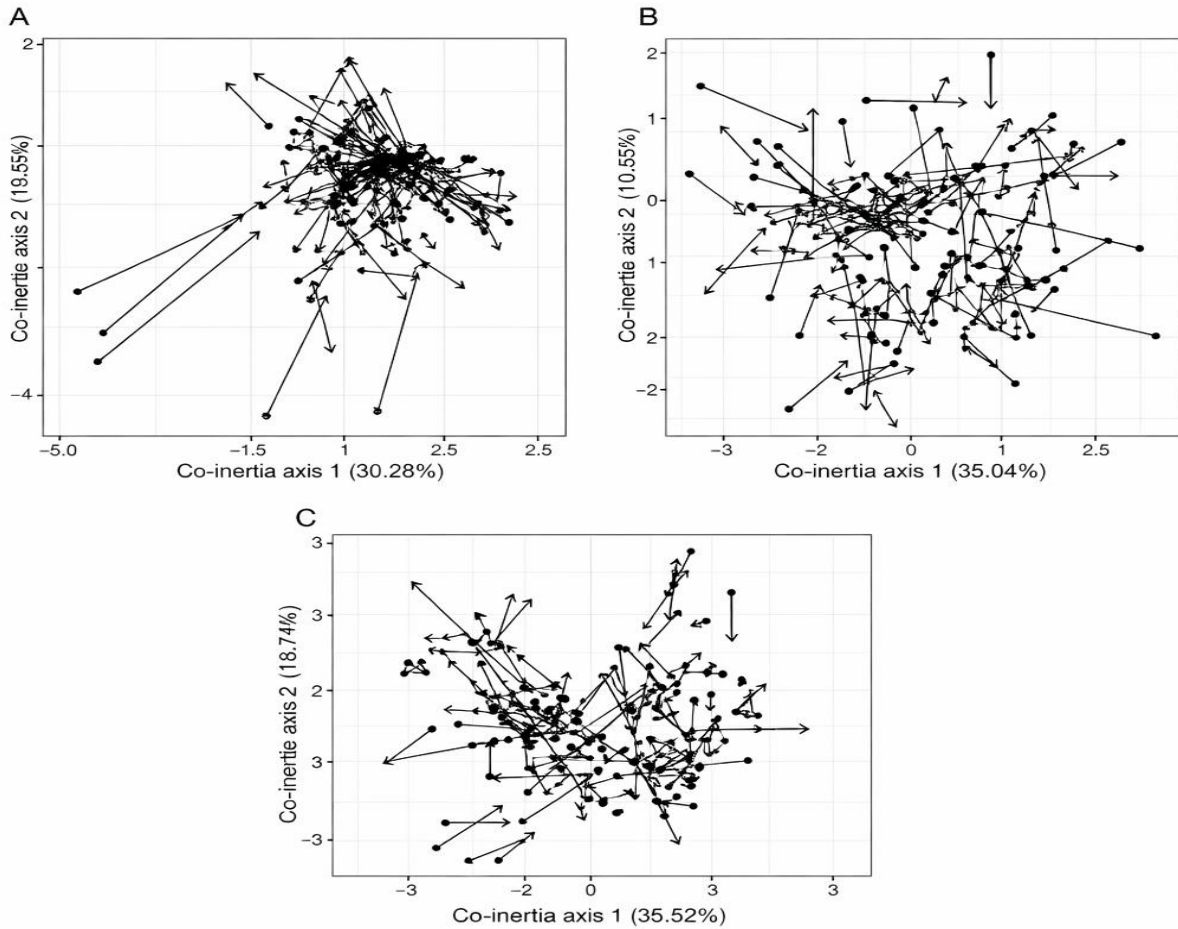


Figure 21. Co-inertia analysis biplots illustrate the covariation between environmental variables and community composition for phytobenthos based on microscopy (A), diatoms based on microscopy (B) and diatom ASVs based on metabarcoding (C).

Each point represents a sampling site, corresponding to positions based on the biological community and arrowheads correspond to positions in the environmental space (Fig. 21). The direction and length of the arrows indicate the strength and pattern of the shared structure between the two datasets; shorter arrows reflect stronger correspondence. Axis labels indicate the percentage of variation explained by axes 1 and 2.

5.4. Discussion

5.4.1 Diversity metrics of the three biological inventories

The three identification approaches provided a complementary diversity overview of benthic algae in the Marcal–Rába network. Phytobenthos microscopy captured a wide morphological range of algal genera, but provided the lowest species richness. Morphological diatom approach recorded more diatom species per sample than phytobenthos microscopy, while diatom

metabarcoding yielded the highest per-sample diatom richness. These patterns are broadly in line with previous studies showing that metabarcoding typically reveals more taxa than traditional microscopy, particularly among small, morphologically similar or rare diatoms that are easily overlooked under the microscope (Wang et al., 2024; Andersson et al., 2023; Zimmermann et al., 2015). However, when total richness across all samples was considered, phytobenthos microscopy detected the fewest taxa but with a wide taxonomic range. Morphological analysis of diatoms yielded a higher cumulative species richness across all samples (476 distinct taxa) compared to eDNA metabarcoding of diatoms (362 distinct taxa) - similar to other studies (Duleba et al., 2021; Kamberović et al., 2024) due to the limitations of the reference sequence libraries and highlights the effect of the experience level of the diatomist in identifying species in the microscope.

Shannon diversity showed a similar pattern, with both diatom-based approaches yielding higher values than the phytobenthos dataset. This suggests that, beyond detecting a greater number of taxa, diatom-focused methods also capture a more even distribution of abundances across species. These findings have important implications for biomonitoring and the interpretation of long-term trends. The distribution of ASV numbers per species, with a few species represented by many ASVs and most species by only one or two. Species such as *Achnanthes minutissimum* and *Nitzschia palea* were represented by numerous ASVs. This pattern supports the use of ASVs as ecologically meaningful units capable of capturing fine-scale community structure (Callahan et al., 2017), while at the same time underlining the need for continued curation and expansion of reference libraries.

Overall, the inventories originated from the different methods indicate that microscopy and metabarcoding should be viewed as complementary rather than competing tools (Keck et al., 2022; Bíró et al., 2022). Phytobenthos microscopy offers a broad, multi-taxa perspective, whereas diatom microscopy and diatom metabarcoding provide higher-resolution insights into this key bioindicator group.

5.4.2 Concordance and discrepancies in the estimation of species abundance

Correlations of species abundances between microscopy and metabarcoding for shared diatom taxa revealed a wide range of relationships, from weak to strong positive correlations. This suggests that, while many common diatom species show a reasonably consistent ranking of relative abundance between the two methods, others display clear discrepancies.

Our analysis was intentionally conservative, being restricted to taxa that occurred in at least three samples and were detected by both methods. For some taxa, strong positive correlations indicate that read counts scale relatively well correlated with microscopic abundance, especially in the case of dominant species like *Cocconeis placentula*, *Amphora pediculus* and *Navicula tripunctata*. On the other hand, numerous species with rare and low-abundance taxa showed negative correlations between the two methods, like *Fallacia pygmaea*, *Nitzschia sigmoidea*, and *Nitzschia supralitorea*. Vasselon et al. (2018) showed that applying a cell-biovolume-based correction factor to rbcL metabarcoding data improves the quantitative estimation of diatom taxa and brings molecular assessments closer to morphology-based water quality indices.

However, low or inconsistent correlations are likely driven by methodological and biological biases, including differences in DNA extraction efficiency, amplification success, gene copy number variation, and the challenges of consistently identifying morphologically similar taxa (Dulias et al., 2017; Keck et al., 2019).

5.4.3 Environmental gradients shaping algal and diatom communities

The PCA of environmental variables revealed a correlated gradient associated with agricultural land use, salinity and nutrient enrichment, separating agricultural sites from more forested, riparian-dominated areas. This confirms that land use and its influence on water chemistry represent the main environmental driver in the studied river network. A secondary gradient related to pH, temperature and trophic state reflected additional variation in habitat conditions, including altitudinal and hydrological differences among sites, such as oxygen concentrations and water flow velocities. In our dataset (Supplementary Table 6), *Reimeria sinuata* was positively associated with forest cover in the catchment, suggesting a preference for relatively undisturbed, shaded headwater sections with clear, well-oxygenated water. This pattern is consistent with previous studies that describe *R. sinuata* as a clean-water diatom occurring in oligotrophic to mesotrophic, often forested streams and headwaters, typically at low nitrogen concentrations and only moderate nutrient and conductivity levels (Veraart et al., 2008). On the other hand, diatom species like *Planothidium victorii* showed a positive relationship with both conductivity and cropland cover, indicating a preference for minerals and nutrients-enriched water of agricultural areas. This pattern agrees with previous studies that describe *P. victorii* as a nutrient-rich, semi-urban river catchment in New Zealand (Novis et al., 2012).

These gradients were also reflected in the relationship between the diversity of algal communities and the environment. The unimodal responses of both richness and Shannon diversity along the main environmental axis (PC1) are consistent with the classical view that algal diversity tends to peak at intermediate levels of disturbance and declines under heavily impacted conditions. However, the strength and clarity of these relationships varied among datasets and between diversity metrics (Huston, 1979; Groendahl and Fink, 2017). Quadratic models based on Shannon diversity were generally better supported (lower AIC) than those based on richness, particularly for diatom ASVs obtained from metabarcoding and for phytobenthos microscopy. This indicates that Shannon diversity provides a more informative descriptor of community responses to environmental stress than richness alone. In particular, the lowest AIC and the significant relationship observed for the Shannon diversity of diatom ASVs suggest that the metabarcoding dataset captured the relationship between the environmental variables and diversity most efficiently, because it can combine high taxonomic resolution with sensitivity to shifts in relative abundances across many taxa (Li et al., 2018; Pawlowski et al., 2018; Keck et al., 2022). In contrast, richness-based models consistently showed much higher AIC values, indicating weaker overall model support.

5.4.4 Community–environment co-structure and implications for biomonitoring

The co-inertia analyses provided an integrated view of how strongly community composition followed environmental gradients across the three datasets. The differences between all three CoIAs were highly significant as shown by the significant Monte-Carlo permutation tests and the aligned structures in the biplots (Fig. 21). However, the strength of this co-structure, as indicated by the RV coefficients, differed substantially among methods. The lowest RV was obtained for phytobenthos microscopy, indicating a relatively weaker coupling between the broad algal community and the measured environmental parameters. By contrast, diatoms are widely recognized as particularly responsive to key water-quality gradients (e.g., nutrients, ionic composition, organic enrichment), and diatom-based indicator frameworks explicitly rely on these tight environment–assemblage linkages, which aligns with the higher RV observed for both diatom datasets (van Dam et al., 1994). The stronger co-structure in the metabarcoding dataset further suggests that ASV-level resolution retains fine-scale ecological sorting that may be dampened in morphology-based inventories by taxonomic lumping and limited discrimination among cryptic or closely related taxa; this interpretation is consistent with the literature arguing that exact sequence variants provide higher-resolution (Callahan et al., 2017).

These findings are in line with recent work showing that taxonomy-free indices can outperform taxonomically aggregated metrics in predicting environmental variables and ecological status (Apothéloz-Perret-Gentil et al., 2020; Gregersen et al., 2023). From a biomonitoring perspective, the strong and significant co-structure between diatom ASVs and environmental variables suggests that eDNA metabarcoding can provide highly sensitive descriptors of ecological condition. When combined with robust statistical frameworks, such approaches could underpin new indices tailored to specific gradients (e.g. salinity, land-use intensity) and complement or, in some contexts, replace traditional morphology-based metrics.

5.5. Conclusion

Benthic algal diversity patterns in the Marcal–Rába river network were strongly method-dependent. Phytobenthos microscopy captured the broadest algal spectrum but showed the lowest richness and the weakest community environment coupling, whereas diatom-focused methods revealed higher taxonomic resolution within this key bioindicator group. Diatom metabarcoding at the ASV level yielded the highest within-site richness and showed relatively stronger relationships with the measured environmental variables than the microscopy-based datasets. At the same time, microscopy and metabarcoding recovered partly distinct taxonomic inventories, indicating that the two approaches provide complementary rather than interchangeable information. These results support the value of an integrated monitoring perspective in which metabarcoding enhances sensitivity to environmental gradients while microscopy remains important for maintaining continuity with established morphology-based assessments and for documenting taxa not recovered by molecular data. Future progress should focus on curated reference libraries, harmonised protocols, and the development of molecular indices that can be interpreted alongside existing monitoring frameworks.

6. General Discussion

Freshwater ecosystems are experiencing increasing anthropogenic pressures, making reliable biomonitoring essential for their conservation and management. Diatoms are among the most informative biological indicators of freshwater ecological status because of their high diversity, rapid response to environmental change, and central role in ecosystem functioning (Kelly and Whitton 1995; Smol and Stoermer 2010; B-Béres et al. 2023). This dissertation integrated a comprehensive review of diatom eDNA metabarcoding with two original research studies to evaluate current molecular approaches, investigate the ecological role of rare diatom species, and compare microscopy with eDNA metabarcoding for biodiversity assessment and ecological

interpretation. The review demonstrated that diatom metabarcoding has become an increasingly mature biomonitoring approach, supported by continuous improvements in sequencing technologies, bioinformatic pipelines, and reference databases (Kermarrec et al. 2014; Keck et al. 2017; Bailet et al. 2020; Keck et al. 2023). Nevertheless, several methodological challenges remain, including incomplete reference libraries, amplification bias, taxonomic inconsistencies, and limitations in abundance estimation, all of which influence ecological interpretation (Vasselon et al. 2018; Tapolczai et al. 2019; Keck et al. 2023). These challenges highlight that further standardization of laboratory procedures, analytical workflows, and reference databases is required before molecular approaches can be routinely implemented in freshwater monitoring programs (Bailet et al. 2020; Vasselon et al. 2025). The second chapter demonstrated that rare diatom species are fundamental components of freshwater biodiversity rather than incidental elements of communities. Although common species dominated local assemblages, rare taxa contributed disproportionately to community variability, beta diversity, species turnover, and co-occurrence network structure. Habitat specialists and geographically restricted taxa were particularly important in shaping metacommunity organization, emphasizing that conservation strategies should consider not only dominant species but also the diversity maintained by the rare biosphere (Sogin et al. 2006; Lynch and Neufeld 2015; Jousset et al. 2017). These findings support the view that protecting overall biodiversity inherently contributes to the conservation of ecologically valuable rare species (Jeliazkov et al. 2022). The comparison between microscopy and eDNA metabarcoding further demonstrated that the two approaches provide complementary rather than competing perspectives on freshwater biodiversity. Microscopy detected a larger regional species pool and remains indispensable for taxonomic validation and long-term compatibility with existing assessment systems. In contrast, eDNA metabarcoding consistently detected higher within-sample richness, identified cryptic and low-abundance taxa more efficiently, and showed stronger relationships with environmental gradients. The limited overlap between methods therefore reflects methodological differences in species detection rather than shortcomings of either approach, supporting previous studies demonstrating the complementarity of molecular and morphological approaches (Mora et al. 2019; Bailet et al. 2019; Keck et al. 2022). Furthermore, when sequence reads were considered as a proxy for relative abundance, agreement between methods improved substantially for abundant taxa, indicating that many discrepancies originate from the detection of rare species, incomplete reference databases, and differences in abundance estimation (Vasselon et al. 2018; Lamb et al. 2019). An important outcome of this dissertation is that the choice of assessment method influences ecological interpretation. Microscopy and eDNA metabarcoding identified similar

broad ecological patterns but differed in estimates of species richness, rarity, and species-environment relationships. These differences reflect the higher taxonomic resolution and detection sensitivity of metabarcoding together with the well-established taxonomic framework provided by microscopy. Consequently, integrating both approaches provides a more comprehensive understanding of freshwater diatom communities than either method alone and represents the most robust strategy for future biomonitoring applications (Apothéloz-Perret-Gentil et al. 2021; Kelly et al. 2020; Tapolczai et al. 2024). Overall, this dissertation demonstrates that advances in molecular ecology are expanding the capacity to characterize freshwater biodiversity while simultaneously emphasizing the continuing value of traditional taxonomy. Rather than replacing microscopy, eDNA metabarcoding should be viewed as a complementary tool that enhances biodiversity detection, improves ecological assessment, and provides new opportunities to investigate community assembly and the ecological significance of rare taxa. Continued development of curated reference libraries, standardized analytical workflows, and molecular ecological indices will facilitate the integration of these complementary approaches into routine freshwater biomonitoring and biodiversity conservation programs (Keck et al. 2023; Vasselon et al. 2025; Tapolczai et al. 2024).

Personal contribution to the conducted research

The research presented in this dissertation was conducted within the Research Group of Limnology at the University of Pannonia in collaboration with my supervisors, co-authors, laboratory staff, taxonomic experts, and collaborating institutions. My personal contribution differed among the three chapters, reflecting the nature of each study and the availability of previously collected data. I overall wrote all the dissertation with specific contributions are described below in the main sections.

Contribution to Section 3

Section 3 is based on the review article *Environmental DNA Metabarcoding of Diatoms: A Next-Generation Tool for Aquatic Biodiversity and Bioassessment*, for which I was the first author. My principal responsibility was the systematic collection, organization, analysis, and synthesis of the published literature on diatom DNA metabarcoding.

I conducted the systematic search of the Web of Science database and screened the 582 publications initially retrieved. Based on predefined inclusion criteria, I identified the 128 primary research articles included in the final review. I developed the structured database used to record information from each study, including publication year, geographical location, habitat type, sample type, preservation method, DNA extraction procedure, genetic marker, sequencing technology, bioinformatic pipeline, taxonomic assignment, and ecological application.

I extracted and checked the relevant information from the selected publications and carried out the descriptive synthesis of temporal, geographical, and methodological patterns. I prepared the figures presenting the development of the research field, geographical distribution of studies, habitat and sample types, DNA extraction methods, barcode markers, sequencing technologies, and bioinformatic pipelines.

I led the interpretation of the results and wrote the first complete draft of the manuscript. I was responsible for integrating the literature concerning sampling, preservation, DNA extraction, barcode selection, sequencing, bioinformatics, reference libraries, taxonomic correspondence with microscopy, ecological assessment, and forensic applications. I also coordinated the revision of the manuscript and incorporated the comments and suggestions provided by the co-authors and journal reviewers. My supervisors and co-authors contributed

to the conceptual refinement of the review, provided specialist knowledge, checked the scientific interpretation, and critically revised the manuscript.

Contribution to section 4

section 4 is based on the collaborative article *Role of Rare Species in Benthic Diatom Communities: Patterns, Processes and Networks*. My contribution to this study was primarily related to the investigation and the critical review and editing of the manuscript.

I contributed to the research activities associated with the river-network investigation and participated in the field and lab work including sampling and processing the samples in the lab for the chemical parameters.

I critically reviewed the manuscript, contributed to the interpretation of the ecological importance of rare diatom species, and assisted in revising the text before publication. In particular, I contributed to the discussion of the relationship between microscopy-based detection and the study of rare microbial taxa, and to the interpretation of how molecular approaches could complement microscopy in future studies of the rare biosphere.

The original conceptualization of the rarity framework, data curation, formal statistical analyses, and initial drafting of the manuscript were led by Prof. Csilla Stenger-Kovács and the principal co-authors. Therefore, my contribution to this chapter was collaborative rather than leading, and the work included in the dissertation is presented with full acknowledgement of the contributions of all co-authors.

Contribution to section 5

section 5 is based on the article *Using Complementary Approaches to Study Diatom Diversity and Its Environmental Drivers*, for which I was the first author. The original field sampling, environmental measurements, and production of the microscopy datasets were conducted within earlier research activities of the Research Group of Limnology. The samples used in this study were collected from 67 sites in the Marcal–Rába catchment during the 2018 and 2019 sampling campaigns.

My principal contribution was the integration and analysis of the phytobenthos microscopy, diatom microscopy, metabarcoding, and environmental datasets. I participated in

defining the comparative research questions and developed the analytical framework used to evaluate the performance and complementarity of the three identification approaches.

I carried out the bioinformatic processing of the metabarcoding data using the DADA2 pipeline adapted for diatom metabarcoding. This work included primer removal, inspection of read-quality profiles, sequence filtering, error-model construction, dereplication, ASV inference, merging of paired-end reads, chimera removal, tracking of retained reads, and taxonomic assignment using the diat.barcode reference database. I subsequently organized the ASV table and prepared both taxonomically aggregated and ASV-level datasets for the different analyses.

I conducted the statistical analyses used to compare the three approaches. These included analysis of variance and Tukey post-hoc comparisons of richness and Shannon diversity, Spearman rank correlations between microscopy-based relative abundance and metabarcoding read proportions, principal component analysis of the environmental variables, quadratic modelling of diversity responses along the main environmental gradient, and co-inertia analyses of community–environment relationships.

I prepared the figures and tables presenting taxonomic assignment, taxonomic overlap, diversity comparisons, abundance correlations, environmental gradients, quadratic diversity responses, and community–environment co-structure. I interpreted the results in relation to the methodological strengths and limitations of the three datasets and evaluated their implications for freshwater biomonitoring.

I wrote the initial manuscript and led its subsequent revision. My supervisors contributed to the study design, provided the existing biological and environmental datasets, supervised the analyses, supported the ecological interpretation, and critically revised the manuscript. Laboratory procedures, microscopic identifications, and sequencing activities were conducted with the support of research-group members, taxonomic specialists, and the sequencing platform described in the chapter.

Dedication

I dedicate this dissertation, first and foremost, to my mother and father, whose love, prayers, sacrifices, and unwavering belief in me have been the foundation of every step I have taken.

To my beloved wife Ala'a, I offer my deepest gratitude and love. This journey was never mine alone. You carried its weight with me every day, often more than anyone could see. Your patience, understanding, encouragement, and sacrifice gave me the strength to continue through the most difficult moments. You stood beside me when the path was long, uncertain, and exhausting, and your support made this achievement possible in ways words can never fully express.

To my dear children, Yamen and Yasmin, this work is especially dedicated to you. You gave up time, presence, and many quiet moments with me while I was pursuing this dream. Your love, innocence, and smiles were a source of light during the hardest days. I hope one day you will understand that every hour of effort was also for your future, and that this journey was shaped by my love for you.

I also dedicate this dissertation to all students who were never given a fair and safe chance to reach education because of war, conflict, fear, and displacement. There are countless bright minds whose dreams were interrupted, whose classrooms were taken away, and whose paths were made harder by circumstances beyond their control. This work is dedicated to them with sincere hope that one day education will be a safe right for every student, everywhere, and that no dream will be denied by violence or instability.

Acknowledgment

I would like to express my sincere gratitude to my supervisors, Prof. Csilla Stenger-Kovács and Dr. Kálmán Tapolczai, for their continuous support, guidance, and encouragement throughout my doctoral journey. I am deeply thankful for their trust, patience, and willingness to share their knowledge and experience with me at every stage of this work. Their academic guidance and constructive advice have greatly shaped this dissertation and my development as a researcher. I am especially grateful to them for giving me the opportunity to attend conferences, workshops, and scientific events that enriched my experience, broadened my perspective, and allowed me to grow both personally and professionally. These opportunities were invaluable for my academic development and for building connections within the scientific community.

I would also like to extend my sincere thanks to the research team, especially Edit Király, Rita Zsuga-Biró, Edina Lengyel, Judit Padisák, and Géza Selmczy, for their cooperation, support, and pleasant collaboration. Their help, kindness, and readiness to assist made many parts of this work easier and more enjoyable. I truly appreciate the positive working environment they helped create and their contribution throughout this period.

I would also like to express my heartfelt gratitude to Hungary for welcoming me during my doctoral journey and for giving me the opportunity to continue my studies in a supportive academic environment. I am equally thankful to the Hungarian people for their kindness, hospitality, and respect, which made my stay here a meaningful and memorable experience. The years I spent in Hungary will always remain an important part of my personal and academic life.

References

- Aanderud ZT, Jones SE, Fierer N, Lennon JT (2015) Resuscitation of the rare biosphere contributes to pulses of ecosystem activity. *Front Microbiol* 6:24. <https://doi.org/10.3389/fmicb.2015.00024>
- Abonyi A, Ács É, Hidas A, Grigorszky I, Várbíró G, Borics G, Kiss KT (2018) Functional diversity of phytoplankton highlights long-term gradual regime shift in the middle section of the Danube river due to global warming, human impacts and oligotrophication. *Freshw Biol* 63:456–472. <https://doi.org/10.1111/fwb.13084>
- Abonyi A, Kiss KT, Hidas A, Borics G, Várbíró G, Ács É (2020) Cell size decrease and altered size structure of phytoplankton constrain ecosystem functioning in the middle Danube river over multiple decades. *Ecosyst* 23:1254–1264. <https://doi.org/10.1007/s10021-019-00467-6>
- Abubaker, B., Stenger-Kovács, C., Rimet, F., & Tapolczai, K. Environmental DNA metabarcoding of diatoms: A next-generation tool for aquatic biodiversity and bioassessment. *Metabarcoding and Metagenomics*. 10, e168096. <https://doi.org/10.3897/mbmg.10.168096>
- Ács É, Bíró T, Duleba M, Földi A, Kiss KT, Orgoványi P, Trábert Z, Buczkó K (2023) Képes útmutató Magyarország Leggyakoribb Bevonatlakó Kovaalgáihoz (Bacillariophyceae). Ludovika Egyetemi Kiadó, Budapest
- Allan E, Weisser W, Weigelt A, Roscher C, Fischer M, Hillebrand H (2011) More diverse plant communities have higher functioning over time due to turnover in complementary dominant species. *Proc Natl Acad Sci USA* 108:17034–17039. <https://doi.org/10.1073/pnas.1104015108>
- Almeida, S. F., Elias, C., Ferreira, J., Tornés, E., Puccinelli, C., Delmas, F., ... Sabater, S. (2014). Water quality assessment of rivers using diatom metrics across Mediterranean Europe: A methods intercalibration exercise. *Science of the Total Environment*, 476, 768–776. <https://doi.org/10.1016/j.scitotenv.2013.11.144>
- Alric B, Ter Braak CJF, Desdevises Y, Lebretonchel H, Dray S (2020) Investigating microbial associations from sequencing survey data with co-correspondence analysis. *Mol Ecol Resour* 20:468–480. <https://doi.org/10.1111/1755-0998.13126>
- Amamoo, P. A., Ayiku, A. N. A., Oyawoye, F., Amoako, E. K., Wiafe, G., Botwe, B. O., & Quashie, P. K. (2025). Metabarcoding for the Molecular Diversity of Diatoms, Dinoflagellates and Cyanobacteria as a Relative Measure of Estuarine Health. SSRN Scholarly Paper No. 5136729. Social Science Research Network. <https://doi.org/10.2139/ssrn.5136729>
- American Public Health Association. (2012). Standard methods for the examination of water and wastewater (22nd ed.). American Public Health Association.
- Andersson, A., Zhao, L., Brugel, S., Figueroa, D., & Huseby, S. (2023). Metabarcoding vs microscopy: Comparison of methods to monitor phytoplankton communities. *ACS ES&T Water*, 3(8), 2671–2680. <https://doi.org/10.1021/acsestwater.3c00176>
- Anslan, S., Kang, W., Dulias, K., Wünnemann, B., Echeverría-Galindo, P., Börner, N., Schwarz, A., Liu, Y., Liu, K., Künzel, S., Kisand, V., Rioual, P., Peng, P., Wang, J., Zhu, L., Vences, M., & Schwalb, A. (2022). Compatibility of Diatom Valve Records With

- Sedimentary Ancient DNA Amplicon Data: A Case Study in a Brackish, Alkaline Tibetan Lake. *Frontiers in Earth Science* 10, 824656.
<https://doi.org/10.3389/feart.2022.824656>
- Apothéloz-Perret-Gentil, L., Bouchez, A., Cordier, T., Cordonier, A., Guéguen, J., Rimet, F., ... Pawlowski, J. (2021). Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. *Molecular Ecology*, 30(13), 2959–2968.
<https://doi.org/10.1111/mec.15646>
- Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., & Pawlowski, J. (2017). Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular Ecology Resources*, 17(6), 1231–1242.
<https://doi.org/10.1111/1755-0998.12668>
- Arumugham S, Joseph SJP, Gopinath PM, Nooruddin T, Subramani N (2023) Diversity and ecology of freshwater diatoms as pollution indicators from the freshwater ponds of Kanyakumari district, Tamilnadu. *Energy Nexus* 9:100164.
<https://doi.org/10.1016/j.nexus.2022.100164>
- B-Béres, V., Stenger-Kovács, C., Buczkó, K., Padisák, J., Selmečzy, G. B., Lengyel, E., & Tapolczai, K. (2023). Ecosystem services provided by freshwater and marine diatoms. *Hydrobiologia* 850(12–13), 2707–2733. <https://doi.org/10.1007/s10750-022-04984-9>
- Bailet, B., Apothéloz-Perret-Gentil, L., Baričević, A., Chonova, T., Franc, A., Frigerio, J.-M., ... Kahlert, M. (2020). Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Science of the Total Environment*, 745, 140948.
<https://doi.org/10.1016/j.scitotenv.2020.140948>
- Bailet, B., Bouchez, A., Franc, A., Frigerio, J. M., Keck, F., Karjalainen, S. M., Rimet, F., Schneider, S., & Kahlert, M. (2019). Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. *Metabarcoding and Metagenomics* 3, e34002.
<https://doi.org/10.3897/mbmg.3.34002>
- Baricevic, A., Chardon, C., Kahlert, M., Karjalainen, S. M., Pfannkuchen, D. M., Pfannkuchen, M., Rimet, F., Tankovic, M. S., Trobajo, R., Vasselon, V., Zimmermann, J., & Bouchez, A. (2022). Recommendations for the preservation of environmental samples in diatom metabarcoding studies. *Metabarcoding and Metagenomics* 6, 349–365.
<https://doi.org/10.3897/mbmg.6.85844>
- Beng, K. C., & Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: Opportunities, challenges and prospects. *Biodiversity and Conservation* 29(7), 2089–2121. <https://doi.org/10.1007/s10531-020-01980-0>
- Bey, M.-Y., & Ector, L. (2013). Atlas des diatomées des cours d'eau de la région Rhône-Alpes (Tome 5, pp. 751–971). DREAL Rhône-Alpes.
- Bíró, T., Duleba, M., Földi, A., Kiss, K. T., Orgoványi, P., Trábert, Z., ... Ács, É. (2022). Metabarcoding as an effective complement of microscopic studies in revealing the composition of the diatom community—a case study of an oxbow lake of Tisza River (Hungary) with the description of a new *Mayamaea* species. *Metabarcoding and Metagenomics*, 6, e87497. <https://doi.org/10.3897/mbmg.6.87497>

- Bispo PC, Balzter H, Malhi Y, Slik JF, Dos Santos JR, Rennó CD, Espírito-Santo FD, Aragão LE, Ximenes AC, Bispo PC (2017) Drivers of metacommunity structure diverge for common and rare Amazonian tree species. *PLoS ONE* 12:e0188300. <https://doi.org/10.1371/journal.pone.0188300>
- Bonačić, T., Arapov, J., Bušelić, I., Lepen Pleić, I., Milić Roje, B., Tomašević, T., ... Ninčević Gladan, Ž. (2025). Advancing the taxonomy of the diatom *Pseudo-nitzschia* through an integrative study conducted in the Central and Southeastern Adriatic Sea. *Plants*, 14(2), 245. <https://doi.org/10.3390/plants14020245>
- Bonn A, Gaston KJ (2005) Capturing biodiversity: selecting priority areas for conservation using different criteria. *Biodivers Conserv* 14:1083–1100. <https://doi.org/10.1007/s10531-004-8410-6>
- Borrego-Ramos, M., Bécarea, E., & Alba-Tercedor, J. (2025). DNA metabarcoding applications for freshwater biomonitoring: Current advances and future perspectives. *Science of the Total Environment*, 950, 175432.
- Bregović P, Fišer C, Zagmajster M (2019) Contribution of rare and common species to subterranean species richness patterns. *Ecol Evol* 9:11606–11618. <https://doi.org/10.1002/ece3.5604>
- Buczko K, Trábert Z, Stenger-Kovács C et al (2022) Rapid expansion of an aquatic invasive species (AIS) in Central-European surface waters; a case study of *Achnanthes delmontii*. *Ecol Ind* 135:108547. <https://doi.org/10.1016/j.ecolind.2022.108547>
- Burgmer T, Hillebrand H, Pfenniger M (2007) Effects of climate-driven temperature changes on the diversity of freshwater macroinvertebrates. *Oecologia* 151:93–103. <https://doi.org/10.1007/s00442-006-0542-9>
- Busseni G, Caputi L, Piredda R et al (2020) Large scale patterns of marine diatom richness: drivers and trends in a changing ocean. *Global Ecol Biogeogr* 29:1915–1928. <https://doi.org/10.1111/geb.13161>
- Cahoon, A. B., Huffman, A. G., Krager, M. M., & Crowell, R. M. (2018). A meta-barcoding census of freshwater planktonic protists in Appalachia – Natural Tunnel State Park, Virginia, USA. *Metabarcoding and Metagenomics* 1, e26939. <https://doi.org/10.3897/mbmg.2.26939>
- Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, 11(12), 2639–2643. <https://doi.org/10.1038/ismej.2017.119>
- Callahan, 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. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Cantonati M, Hofmann G, Spitale D, Werum M, Lange-Bertalot H (2022) Diatom red lists: important tools to assess and preserve biodiversity and habitats in the face of direct impacts and environmental change. *Biodivers Conserv* 31:453–477. <https://doi.org/10.1007/s10531-021-02339-9>
- Cantonati, M., Kelly, M. G., & Lange-Bertalot, H. (Eds.). (2017). *Freshwater benthic diatoms of Central Europe: Over 800 common species used in ecological assessment*. Koeltz Botanical Books.

- Cao Y, Williams DD, Williams NE (1998) How important are rare species in aquatic community ecology and bioassessment? *Limnol Oceanogr* 43:1403–1409.
<https://doi.org/10.4319/lo.1998.43.7.1403>
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., & ... Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5), 335–336.
<https://doi.org/10.1038/nmeth.f.303>
- CEN (2018). CEN/TR 17245: Water quality. Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. CEN/TC 230/WG23, Aquatic Macrophytes and Algae: 1–8..
- CEN (European Committee for Standardization). (2002). Water quality—Guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers (prEN 13946:2002).
- CEN (European Committee for Standardization). (2018a). Water quality—Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses (CEN/TR 17245:2018).
- CEN (European Committee for Standardization). (2018b). Water quality—Technical report for the management of diatom barcodes (CEN/TR 17244:2018).
- Cermeño P, Teixeira IG, Branco M, Figueiras FG, Marañón E (2014) Sampling the limits of species richness in marine phytoplankton communities. *J Plankton Res* 36:1135–1139.
<https://doi.org/10.1093/plankt/fbu033>
- Clark, D. E., Pilditch, C. A., Pearman, J. K., Ellis, J. I., & Zaiko, A. (2020). Environmental DNA metabarcoding reveals estuarine benthic community response to nutrient enrichment – Evidence from an in-situ experiment. *Environmental Pollution* 267, 115472. <https://doi.org/10.1016/j.envpol.2020.115472>
- Clarke KR, Green RH (1988) Statistical design and analysis for a 'biological effects' study. *Mar Ecol Prog Ser* 46:213–226. <https://doi.org/10.3354/meps046213>
- Cordier, T., Esling, P., Lejzerowicz, F., Visco, J., Ouadahi, A., Martins, C., Cedhagen, T., & Pawlowski, J. (2017). Predicting the Ecological Quality Status of Marine Environments from eDNA Metabarcoding Data Using Supervised Machine Learning. *Environmental Science & Technology* 51(16), 9118–9126. <https://doi.org/10.1021/acs.est.7b01518>
- Couton, M., Lévêque, L., Daguin-Thiébaud, C., Comtet, T., & Viard, F. (2022). Water eDNA metabarcoding is effective in detecting non-native species in marinas, but detection errors still hinder its use for passive monitoring. *Biofouling* 38(4), 367–383.
<https://doi.org/10.1080/08927014.2022.2075739>
- Crain BJ, Tremblay RL (2014) Do richness and rarity hotspots really matter for Orchid conservation in light of anticipated habitat loss? *Divers Distrib* 20:652–662.
<https://doi.org/10.1111/ddi.12179>
- Csárdi G, Nepusz T, Traag V, Horvát Sz, Zanini F, Noom D, Müller K (2024) Igraph: network analysis and visualization in R. R Package version 203.
<https://doi.org/10.5281/zenodo.7682609>

- Cui, Z., Liu, S., Xu, Q., Zhao, Y., & Chen, N. (2023). Differential ecological adaptation of diverse *Chaetoceros* species revealed by metabarcoding analysis. *Environmental DNA* 5(6), 1332–1350. <https://doi.org/10.1002/edn3.455>
- Dahiya, P., Makwana, M. D., Chaniyara, P., & Bhatia, A. (2024). A comprehensive review of forensic diatomology: Contemporary developments and future trajectories. *Egyptian Journal of Forensic Sciences* 14(1), 2. <https://doi.org/10.1186/s41935-023-00378-7>
- Darling, J. A., & Mahon, A. R. (2011). From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research* 111(7), 978–988. <https://doi.org/10.1016/j.envres.2011.02.001>
- Dee LE, Cowles J, Isbell F, Pau S, Gaines SD, Reich PB (2019) When do ecosystem services depend on rare species? *Trends Ecol Evol* 34:746–758. <https://doi.org/10.1016/j.tree.2019.03.010>
- Deiner, K., & Altermatt, F. (2014). Transport Distance of Invertebrate Environmental DNA in a Natural River. *PLoS ONE* 9(2), e88786. <https://doi.org/10.1371/journal.pone.0088786>
- Della Rocca F, Bogliani G, Breiner FT, Milanese P (2019) Identifying hotspots for rare species under climate change scenarios: improving saproxylic beetle conservation in Italy. *Biodivers Conserv* 28:433–449. <https://doi.org/10.1007/s10531-018-1670-3>
- Descy, J. P., & Coste, M. (1991). A test of methods for assessing water quality based on diatoms. *Internationale Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen*. <https://www.tandfonline.com/doi/abs/10.1080/03680770.1989.11899905>.
- Desrosiers, C., Leflaive, J., Eulin, A., & Ten-Hage, L. (2013). Bioindicators in marine waters: Benthic diatoms as a tool to assess water quality from eutrophic to oligotrophic coastal ecosystems. *Ecological Indicators* 32, 25–34. <https://doi.org/10.1016/j.ecolind.2013.02.021>
- Directive, W. F. (2000). EU Water framework directive. EC Directive 60..
- Dray S, Bauman D, Blanchet G et al (2023) adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3–23. <https://CRAN.R-project.org/package=adespatial>
- Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., ... Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81(2), 163–182. <https://doi.org/10.1017/S1464793105006950>
- Duleba, M., Földi, A., Micsinai, A., Várbiro, G., Mohr, A., Sipos, R., ... Ács, É. (2021). Applicability of diatom metabarcoding in the ecological status assessment of Hungarian lotic and soda pan habitats. *Ecological Indicators*, 130, 108105. <https://doi.org/10.1016/j.ecolind.2021.108105>
- Dulias, K., Stoof-Leichsenring, K. R., Pestryakova, L. A., & Herzschuh, U. (2017). Sedimentary DNA versus morphology in the analysis of diatom–environment relationships. *Journal of Paleolimnology*, 57(1), 51–66. <https://doi.org/10.1007/s10933-016-9926-y>

- Endo, H., Ogata, H., & Suzuki, K. (2018). Contrasting biogeography and diversity patterns between diatoms and haptophytes in the central Pacific Ocean. *Scientific Reports* 8(1), 10916. <https://doi.org/10.1038/s41598-018-29039-9>
- Epp, L. S., Zimmermann, H. H., & Stoof-Leichsenring, K. R. (2019). Sampling and Extraction of Ancient DNA from Sediments. In: Shapiro B, Barlow A, Heintzman PD, Hofreiter M, Pajmians JLA, Soares AER (Eds) *Ancient DNA: Methods and Protocols*. Springer, 31–44. https://doi.org/10.1007/978-1-4939-9176-1_5
- Espeland EK, Emam TM (2011) The value of structuring rarity: the seven types and links to reproductive ecology. *Biodivers Conserv* 20:963–985. <https://doi.org/10.1007/s10531-011-0007-2>
- European Parliament and of the Council of the European Union. (2000, October 23). Directive 2000/60/EC establishing a framework for Community action in the field of water policy (Water Framework Directive). *Official Journal of the European Communities*, L 327, 1–73.
- 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(3), 349–364. <https://doi.org/10.1016/j.protis.2007.04.001>
- Fetzer I, Johst K, Schawe R, Banitz T, Harms H, Chatzinotas A (2015) The extent of functional redundancy changes as species’ roles shift in different environments. *Proc Natl Acad Sci USA* 112:14888–14893. <https://doi.org/10.1073/pnas.1505587112>
- Fischlin A, Midgley GF, Price J, Leemans R, Gopal B, Turley C, Rounsevell MDA, Dube P, Tarazona J, Velichko AA (2007) Ecosystems, their properties, goods and services, in: *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, pp 211–272
- Flather CH, Sieg CH (2007) Species rarity: definition, causes and classification. In: Raphael MG, Molina R (eds) *Conservation of rare or little-known species: biological, social, and economic considerations*. Island, Washington, pp 40–66
- Fore LS, Karr JR, Wisseman RW (1996) Assessing invertebrate responses to human activities: evaluating alternative approaches. *J N Am Benthol Soc* 15:212–231. <https://doi.org/10.2307/1467949>
- Freeman LC (1979) Centrality in social networks conceptual clarification. *Social Networks* 1:215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Gaston KJ (1994) *Rarity*. Springer, Dordrecht. <https://doi.org/10.1007/978-94-011-0701-3>
- Gaston KJ (2008) Biodiversity and extinction: the importance of being common. *Prog Phys Geogr* 32:73–79. <https://doi.org/10.1177/0309133308089499>
- Gaston KJ (2010) Valuing common species. *Science* 327:154–155. <https://doi.org/10.1126/science.1182818>
- Goodall DW (1969) A procedure for recognition of uncommon species combinations in sets of vegetation samples. *Vegetatio Acta Geobot* 18:19–35. <https://doi.org/10.1007/BF00332827>

- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Gregersen, R., Pearman, J. K., Atalah, J., Waters, S., Vandergoes, M. J., Howarth, J. D., ... Wood, S. A. (2023). A taxonomy-free diatom eDNA-based technique for assessing lake trophic level using lake sediments. *Journal of Environmental Management*, 345, 118885. <https://doi.org/10.1016/j.jenvman.2023.118885>
- Guiry, M.D. (2012), HOW MANY SPECIES OF ALGAE ARE THERE?. *J. Phycol.*, 48: 1057-1063. <https://doi.org/10.1111/j.1529-8817.2012.01222.x>
- Guri, G., Ray, J. L., Shelton, A. O., Kelly, R. P., Præbel, K., Andruszkiewicz Allan, E., Yoccoz, N., Johansen, T., Wangenstein, O. S., Hanebrekke, T., & Westgaard, J. I. (2024). Quantifying the Detection Sensitivity and Precision of qPCR and ddPCR Mechanisms for eDNA Samples. *Ecology and Evolution* 14(12), e70678. <https://doi.org/10.1002/ece3.70678>
- Hamsher, S. E., Evans, K. M., Mann, D. G., Pouličková, A., & Saunders, G. W. (2011). Barcoding Diatoms: Exploring Alternatives to COI-5P. *Protist* 162(3), 405–422. <https://doi.org/10.1016/j.protis.2010.09.005>
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270(1512), 313–321. <https://doi.org/10.1098/rspb.2002.2218>
- Heegaard E, Gjerde I, Sætersdal M (2013) Contribution of rare and common species to richness patterns at local scales. *Ecography* 36:937–946. <https://doi.org/10.1111/j.1600-0587.2013.00060.x>
- Hepperle, D. (2008). OptiCount: Software for enumeration of plankton and particles (Version 8) [Computer software]. SequentiX – Digital DNA Processing. https://science.d-mix.de/software_opticount.php
- Hering, D., Johnson, R. K., Kramm, S., Schmutz, S., Szoszkiewicz, K., & Verdonschot, P. F. (2006). Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: A comparative metric-based analysis of organism response to stress. *Freshwater Biology*, 51(9), 1757–1785. <https://doi.org/10.1111/j.1365-2427.2006.01610.x>
- Hillebrand, H., Dürselen, C.-D., Kirschtel, D., Pollinger, U., & Zohary, T. (1999). Biovolume calculation for pelagic and benthic microalgae. *Journal of Phycology*, 35(2), 403–424. <https://doi.org/10.1046/j.1529-8817.1999.3520403.x>
- Hinlo, R., Gleeson, D., Lintermans, M., & Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PLoS ONE*, 12(6), e0179251. <https://doi.org/10.1371/journal.pone.0179251>
- Hofmann G, Lange-Bertalot H, Werum M (2013) Diatomeen Im Süßwasser-Benthos von Mitteleuropa: Bestimmungsflora Kieselalgen für die ökologische Praxis: Über 700 der Häufigsten Arten und Ihre Ökologie. Koeltz Sci Books. 908 pp <https://www.rote-liste-zentrum.de/de/Limnische-Kieselalgen-Bacillariophyta-1772.html>
- Hofmann G, Lange-Bertalot H, Werum M, Klee R (2018) Rote Liste und Gesamtartenliste der limnischen Kieselalgen (Bacillariophyta) Deutschlands. In Metzger D., Hofbauer N., Ludwig G. & Matzke Hajek G. (eds.): Rote Liste gefährdeter Tiere, Pflanzen und Pilze

- Deutschlands. Band 7: Pflanzen. Münster: Landwirtschaftsverlag. Naturschutz und Biologische Vielfalt 70: 601–708
- Hürlimann, J., Liechti, P., Niederhauser, P., & Suisse. Office fédéral de l'environnement (2007). Methoden zur Untersuchung und Beurteilung der Fließgewässer: Kieselalgen Stufe F (flächendeckend). Bundesamt für Umwelt, BAFU..
- Huston, M. (1979). A general hypothesis of species diversity. *The American Naturalist*, 113(1), 81–101. <https://doi.org/10.1086/283366>
- Iop S, Caldart VM, Vélez-Martin E, dos Santos TG, Prado PI, De Patta Pillar V, Cechin SZ (2024) Niche and neutral-based processes differ in importance for common and rare species in a metacommunity of anurans in subtropical grasslands. *Hydrobiologia* 851:2357–2371. <https://doi.org/10.1007/s10750-023-05461-7>
- Jeliazkov A, Gavish Y, Marsh CJ, Geschke J, Brummitt N, Rocchini D, Haase P, Kunin WE, Henle K (2022) Sampling and modelling rare species: conceptual guidelines for the neglected majority. *Glob Change Biol* 28:3754–3777. <https://doi.org/10.1111/gcb.16114>
- Jiang, L., Xiao, C., Zhao, J., Jiang, T., Lin, J., Xu, Q., Liu, C., & Cai, W. (2020). Development of 18S rRNA gene arrays for forensic detection of diatoms. *Forensic Science International* 317, 110482. <https://doi.org/10.1016/j.forsciint.2020.110482>
- Jiao S, Luo Y, Lu M, Xiao X, Lin Y, Chen W, Wei G (2017) Distinct succession patterns of abundant and rare bacteria in Temporal microcosms with pollutants. *Environ Pollut* 225:497–505. <https://doi.org/10.1016/j.envpol.2017.03.015>
- Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, Küsel K, Rillig MC, Rivett DW, Salles JF (2017) Where less May be more: how the rare biosphere pulls ecosystems strings. *ISME J* 11:853–862. <https://doi.org/10.1038/ismej.2016.174>
- Kahlert, M., Albert, R. L., Anttila, E. L., Bengtsson, R., Bigler, C., Eskola, T., ... Weckström, J. (2009). Harmonization is more important than experience—Results of the first Nordic–Baltic diatom intercalibration exercise 2007 (stream monitoring). *Journal of Applied Phycology*, 21(4), 471–482. <https://doi.org/10.1007/s10811-008-9394-5>
- Kahlert, M., Bailet, B., Chonova, T., Karjalainen, S. M., Schneider, S. C., & Tapolczai, K. (2021). Same same, but different: The response of diatoms to environmental gradients in Fennoscandian streams and lakes – barcodes, traits and microscope data compared. *Ecological Indicators* 130, 108088. <https://doi.org/10.1016/j.ecolind.2021.108088>
- Kahlert, M., Karjalainen, S.-M., Keck, F., Kelly, M., Ramon, M., Rimet, F., ... Zimmermann, J. (2022). Co-occurrence, ecological profiles and geographical distribution based on unique molecular identifiers of the common freshwater diatoms *Fragilaria* and *Ulnaria*. *Ecological Indicators*, 141, 109114. <https://doi.org/10.1016/j.ecolind.2022.109114>
- Kaleli, A., Ozbayram, E. G., Zengin, Z., & Akcaalan, R. (2025). Investigation of coastal benthic diatom communities using the rbcL marker with nanopore sequencing in Urban Marine environments. *Journal of Phycology*. <https://doi.org/10.1111/jpy.70035>
- Kamberović, J., Gligora Udovič, M., Kulaš, A., Tapolczai, K., Orlić, S., Jusufović, A., Gajić, A., Žutinić, P., Ahmić, A., & Kalamujić Stroić, B. (2024). The diatom diversity and ecological status of a tufa-depositing river through eDNA metabarcoding vs. a morphological approach—A case study of the Una River (Bosnia and Herzegovina). *Microorganisms*, 12(8), 1722. <https://doi.org/10.3390/microorganisms12081722>
- Kang, W., Anslan, S., Börner, N., Schwarz, A., Schmidt, R., Künzel, S., Rioual, P., Echeverría-Galindo, P., Vences, M., Wang, J., & Schwalb, A. (2020). Sample-size and

- bioinformatics independent, consistent patterns among diatom metabarcoding data from lake sediments. *Authorea*. <https://doi.org/10.22541/au.158750300.04703373>
- Kato-Unoki, Y., Mayama, S., Kurihara, A., & Minamoto, T. (2024). Development of specific detection assays from environmental DNA of invasive diatom *Cymbella janischii* in Japan. *Limnology* 25(3), 247–254. <https://doi.org/10.1007/s10201-024-00751-y>
- Keck, F., & Kahlert, M. (2019). Community phylogenetic structure reveals the imprint of dispersal-related dynamics and environmental filtering by nutrient availability in freshwater diatoms. *Scientific Reports*, 9(1), 11590. <https://doi.org/10.1038/s41598-019-48125-0>
- Keck, F., Blackman, R. C., Bossart, R., Brantschen, J., Couton, M., Hürlemann, S., ... Altermatt, F. (2022). Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. *Molecular Ecology*, 31(6), 1820–1835. <https://doi.org/10.1111/mec.16364>
- Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources* 23(4), 742–755. <https://doi.org/10.1111/1755-0998.13746>
- Keck, F., Vasselon, V., Tapolczai, K., Rimet, F., & Bouchez, A. (2017). Freshwater biomonitoring in the Information Age. *Frontiers in Ecology and the Environment* 15(5), 266–274. <https://doi.org/10.1002/fee.1490>
- Kelly, M. G. (1998). Use of the trophic diatom index to monitor eutrophication in rivers. *Water Research*, 32(1), 236–242. [https://doi.org/10.1016/S0043-1354\(97\)00157-7](https://doi.org/10.1016/S0043-1354(97)00157-7)
- Kelly, M. G., & Whitton, B. A. (1995). The Trophic Diatom Index: A new index for monitoring eutrophication in rivers. *Journal of Applied Phycology* 7(4), 433–444. <https://doi.org/10.1007/BF00003802>
- 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. *Ecological Indicators* 118, 106725. <https://doi.org/10.1016/j.ecolind.2020.106725>
- Kelly, M., Bennion, H., Burgess, A., Ellis, J., Juggins, S., Guthrie, R., ... Yallop, M. (2009). Uncertainty in ecological status assessments of lakes and rivers using diatoms. *Hydrobiologia*, 633(1), 5–15. <https://doi.org/10.1007/s10750-009-9872-z>
- Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J.-M., Humbert, J.-F., & Bouchez, A. (2014). A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, 33(1), 349–363. <https://doi.org/10.1086/675079>
- Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Humbert, J.-F., & Bouchez, A. (2013). Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. *Molecular Ecology Resources*, 13(4), 607–619. <https://doi.org/10.1111/1755-0998.12105>
- Kezlya, E., Tseplik, N., & Kulikovskiy, M. (2023). Genetic Markers for Metabarcoding of Freshwater Microalgae: Review. *Biology* 12(7), 1038. <https://doi.org/10.3390/biology12071038>
- Kim, Y. S., Yun, H. S., Lee, J. H., Lee, K. L., Choi, J. S., Won, D. H., Kim, Y. J., Kim, H. S., & Yoon, H. S. (2024). Comparison of Metabarcoding and Microscopy Methodologies to Analyze Diatom Communities in Five Estuaries Along the Southern Coast of the Korean Peninsula. *Microbial Ecology* 87(1), 95. <https://doi.org/10.1007/s00248-024-02396-x>

- Krammer K, Lange-Bertalot H (1999) Süßwasserflora von Mitteleuropa. Teil. 1: Naviculaceae. Spektrum Akademischer
- Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, 7, 17668. <https://doi.org/10.1038/s41598-017-17333-x>
- Kruckeberg AR, Rabinowitz D (1985) Biological aspects of endemism in higher plants. *Annu Rev Ecol Syst* 16:447–479. <https://doi.org/10.1146/annurev.es.16.110185.002311>
- Kulaš, A., Gligora Udovič, M., Tapolczai, K., Žutinić, P., Orlić, S., & Levkov, Z. (2022). Diatom eDNA metabarcoding and morphological methods for bioassessment of karstic river. *Science of the Total Environment* 829, 154536. <https://doi.org/10.1016/j.scitotenv.2022.154536>
- Kurtz Z, Müller C, Miraldi E, Bonneau R (2024) SpiecEasi: Sparse Inverse Covariance for Ecological Statistical Inference. R package version 1.1.3. <https://github.com/zdk123/SpiecEasi>
- Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA (2015) Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol* 11:e1004226. <https://doi.org/10.1371/journal.pcbi.1004226>
- Kusel-Fetzmann EL, Naideno W, Russev B (1998) Plankton und Benthos der Donau. *Ergebnisse Donau- Forschung*, vol 4. IAD, Wien, pp 1–376
- Kutty, S. N., Loh, R. K., Bannister, W., & Taylor, D. (2022). Evaluation of a diatom eDNA-based technique for assessing water quality variations in tropical lakes and reservoirs. *Ecological Indicators*, 141, 109108. <https://doi.org/10.1016/j.ecolind.2022.109108>
- Kwandrans, J., Eloranta, P., Kawecka, B., & Wojtan, K. (1998). Use of benthic diatom communities to evaluate water quality in rivers of southern Poland. *Journal of Applied Phycology*, 10(2), 193–201. <https://doi.org/10.1023/A:1008087114256>
- Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, 28(2), 420–430. <https://doi.org/10.1111/mec.14920>
- Lange-Bertalot, H., Hofmann, G., Werum, M., Cantonati, M., & Kelly, M. G. (2017). *Freshwater benthic diatoms of Central Europe: Over 800 common species used in ecological assessment (Vol. 942)*. Koeltz Botanical Books.
- Lavoie, I., Campeau, S., Zugic-Drakulic, N., Winter, J. G., & Fortin, C. (2014). Using diatoms to monitor stream biological integrity in Eastern Canada: An overview of 10 years of index development and ongoing challenges. *Science of The Total Environment* 475, 187–200. <https://doi.org/10.1016/j.scitotenv.2013.04.092>
- Legendre P (2014) Interpreting the replacement and richness difference components of beta diversity. *Glob Ecol Biogeogr* 23:1324–1334. <https://doi.org/10.1111/geb.12207>
- Lennon JJ, Beale CM, Reid CL, Kent M, Pakeman RJ (2011) Are richness patterns of common and rare species equally well explained by environmental variables? *Ecography* 34:529–539. <https://doi.org/10.1111/j.1600-0587.2010.06669.x>
- Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J., & Zhang, X. (2018). Application of environmental DNA metabarcoding for predicting anthropogenic pollution in rivers.

- Environmental Science & Technology, 52(20), 11708–11719.
<https://doi.org/10.1021/acs.est.8b03869>
- Liao J, Cao X, Wang J, Zhao L, Sun J, Jiang D, Huang Y (2017) Similar community assembly mechanisms underlie similar biogeography of rare and abundant bacteria in lakes on Yungui plateau, China. *Limnol Oceanogr* 62:723–735.
<https://doi.org/10.1002/lno.10455>
- Liu, M., Zhao, Y., Sun, Y., Wu, P., Zhou, S., & Ren, L. (2020). Diatom DNA barcodes for forensic discrimination of drowning incidents. *FEMS Microbiology Letters* 367(17), fnaa145. <https://doi.org/10.1093/femsle/fnaa145>
- Lobo, 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, 245–271. https://doi.org/10.1007/978-3-319-31984-1_11
- Lomba A, Pellissier L, Randin C, Vicente J, Moreira F, Honrado J, Guisan A (2010) Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biol Conserv* 143:2647–2657.
<https://doi.org/10.1016/j.biocon.2010.07.007>
- Lund, J. W. G., Kipling, C., & Le Cren, E. D. (1958). The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia*, 11(2), 143–170. <https://doi.org/10.1007/BF00007865>
- Lynch MD, Neufeld JD (2015) Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* 13:217–229. <https://doi.org/10.1038/nrmicro3400>
- Lyons KG, Brigham CA, Traut BH, Schwartz MW (2005) Rare species and ecosystem functioning. *Conserv Biol* 19:1019–1024. <https://doi.org/10.1111/j.1523-1739.2005.00106.x>
- Lyons KG, Schwartz MW (2001) Rare species loss alters ecosystem function–invasion resistance. *Ecol Lett* 4:358–365. <https://doi.org/10.1046/j.1461-0248.2001.00235.x>
- Mäki, A., Salmi, P., Mikkonen, A., Kremp, A., & Tirola, M. (2017). Sample Preservation, DNA or RNA Extraction and Data Analysis for High-Throughput Phytoplankton Community Sequencing. *Frontiers in Microbiology* 8, 1848.
<https://doi.org/10.3389/fmicb.2017.01848>
- Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., De Vargas, C., Bittner, L., Zingone, A., & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world’s ocean. *Proceedings of the National Academy of Sciences* 113(11), E1516–25. <https://doi.org/10.1073/pnas.1509523113>
- Mann, D. G. (2010). Discovering diatom species: Is a long history of disagreements about species-level taxonomy now at an end? *Plant Ecology and Evolution* 143(3): 251–264.
<https://doi.org/10.5091/plecevo.2010.405>
- Mann, D. G., & Vanormelingen, P. (2013). An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species. *Journal of Eukaryotic Microbiology* 60(4), 414–420. <https://doi.org/10.1111/jeu.12047>
- Mănoiu VM, Crăciun AI (2021) Danube river water quality trends: A qualitative review based on the open access web of science database. *Ecohydrol Hydrobiol* 21:613–628.
<https://doi.org/10.1016/j.ecohyd.2021.08.002>

- Marciel EA (2021) An index for assessing the rare species of a community. *Ecol Ind* 124:107424. <https://doi.org/10.1016/j.ecolind.2021.107424>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10–12.
- Moniz, M. B. J., & Kaczmarek, I. (2010). Barcoding of Diatoms: Nuclear Encoded ITS Revisited. *Protist* 161(1), 7–34. <https://doi.org/10.1016/j.protis.2009.07.001>
- Mora, D., Abarca, N., Proft, S., Grau, 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. *Freshwater Science* 38(3), 448–464. <https://doi.org/10.1086/704827>
- Mortágua A, Vasselon V, Oliveira R, Elias C, Chardon C, Bouchez A, Rimet F, João Feio M, F.P. Almeida S (2019). Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. *Ecological Indicators* 106, 105470. <https://doi.org/10.1016/j.ecolind.2019.105470>
- Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, Kulbicki M, Lavergne S, Lavorel S, Mouquet N (2013) Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol* 11:e1001569. <https://doi.org/10.1371/journal.pbio.1001569>
- Muhammad, B. L., Sin, Y., & Ki, J. S. (2021). DNA metabarcoding reveals multiple co-occurring species in *Stephanodiscus* spring diatom blooms in a temperate freshwater river. *Marine and Freshwater Research* 72(9), 1353–1364. <https://doi.org/10.1071/MF20254>
- Munn, M. D., Waite, I., & Konrad, C. P. (2018). Assessing the influence of multiple stressors on stream diatom metrics in the upper Midwest, USA. *Ecological Indicators* 85, 1239–1248. <https://doi.org/10.1016/j.ecolind.2017.09.005>
- Murray BR, Thrall PH, Gill AM, Nicotra AB (2002) How plant life-history and ecological traits relate to species rarity and commonness at varying Spatial scales. *Austral Ecol* 27:291–310. <https://doi.org/10.1046/j.1442-9993.2002.01181.x>
- Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- Nadir Solak, C., Ljubimir, S., Kaleli, A., Dupčić Radić, I., Hafner, D., & Car, A. (2021). Comparison of benthic diatom community structures on natural and artificial substrates in marine lake (Adriatic Sea). *Acta Adriatica* 62(1), 21–44. <https://doi.org/10.32582/aa.62.1.1>
- Naughton-Treves L, Holland MB, Brandon K (2005) The role of protected areas in conserving biodiversity and sustaining local livelihoods. *Annu Rev Environ Resour* 30:219–252. <https://doi.org/10.1146/annurev.energy.30.050504.164507>
- Neeson TM, Doran PJ, Ferris MC, Fitzpatrick KB, Herbert M, Khoury M, Moody AT, Ross J, Yacobson E, McIntyre PB (2018) Conserving rare species can have high opportunity costs for common species. *Glob Change Biol* 24:3862–3872. <https://doi.org/10.1111/gcb.14162>
- Newbold, L. K., Taylor, J. D., Thorpe, A. C., Warren, J., Walsh, K., & Read, D. S. (2025). DNA Extraction Methodology has a Limited Impact on Multitaxa Riverine Benthic Metabarcoding Community Profiles. *Environmental DNA* 7(3), e70102. <https://doi.org/10.1002/edn3.70102>

- Nistal-García, A., García-García, P., García-Girón, J., Borrego-Ramos, M., Blanco, S., & Bécares, E. (2021). DNA metabarcoding and morphological methods show complementary patterns in the metacommunity organization of lentic epiphytic diatoms. *Science of The Total Environment* 786, 147410. <https://doi.org/10.1016/j.scitotenv.2021.147410>
- 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, Simpson G, Blanchet F et al (2022) *vegan: Community Ecology Package*. R package version 2.6-4. <https://cran.r-project.org/web/packages/vegan/index.html>
- Padisak, J. (1992). Seasonal succession of phytoplankton in a large shallow lake (Balaton, Hungary)—A dynamic approach to ecological memory, its possible role and mechanisms. *Journal of ecology*, 217-230. <https://doi.org/10.2307/2261008>
- Padisák, J., Hajnal, E., Krienitz, L., Lakner, J., & Üveges, V. (2010). Rarity, ecological memory, rate of floral change in phytoplankton—and the mystery of the Red Cock. *Hydrobiologia*, 653(1), 45-64.
- Pandit SN, Kolasa J, Cottenie K (2009) Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology* 90:2253–2262. <https://doi.org/10.1890/08-0851.1>
- Pärtel M, Kalamees R, Reier Ü, Tuvi E-L, Roosaluuste E, Vellak A, Zobel M (2005) Grouping and prioritization of vascular plant species for conservation: combining natural rarity and management need. *Biol Conserv* 123:271–278. <https://doi.org/10.1016/j.biocon.2004.11.014>
- Pawlowski, J., Apothéloz-Perret-Gentil, L., & Altermatt, F. (2020). Environmental DNA: What’s behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology* 29(22), 4258–4264. <https://doi.org/10.1111/mec.15643>
- Pawlowski, J., Apothéloz-Perret-Gentil, L., & Altermatt, F. (2021). Environmental versus extra-organismal DNA. *Molecular Ecology* 30(19), 4606–4607. <https://doi.org/10.1111/mec.16144>
- Pawlowski, J., Bruce, K., Panksep, K., Aguirre, F. I., Amalfitano, S., Apothéloz-Perret-Gentil, L., Baussant, T., Bouchez, A., Carugati, L., Cermakova, K., Cordier, T., Corinaldesi, C., Costa, F. O., Danovaro, R., Dell’Anno, A., Duarte, S., Eisendle, U., Ferrari, B. J. D., Frontalini, F., & Fazi, S. (2022). Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. *Science of The Total Environment* 818, 151783. <https://doi.org/10.1016/j.scitotenv.2021.151783>
- Pawlowski, J., Christen, R., Lecroq, B., Bachar, D., Shahbazkia, H. R., Amaral-Zettler, L., & Guillou, L. (2012). Eukaryotic richness in the abyss: Insights from pyrotag sequencing. *PLoS ONE*, 6(4), e18169. <https://doi.org/10.1371/journal.pone.0018169>
- Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., ... Kahlert, M. (2018). The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment*, 637, 1295–1310. <https://doi.org/10.1016/j.scitotenv.2018.05.002>

- Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology* 55, 12–25. <https://doi.org/10.1016/j.ejop.2016.02.003>
- Peeters, V., & Ector, L. (2017–2019). *Atlas des diatomées des cours d’eau du territoire bourguignon* (Vols. 1–3). DREAL Bourgogne–Franche-Comté.
- Pérez-Burillo, J., Mann, D. G., & Trobajo, R. (2022). Evaluation of two short overlapping rbcL markers for diatom metabarcoding of environmental samples: Effects on biomonitoring assessment and species resolution. *Chemosphere* 307, 135933. <https://doi.org/10.1016/j.chemosphere.2022.135933>
- Pérez-Burillo, J., Trobajo, R., Leira, M., Keck, F., Rimet, F., Sigró, J., & Mann, D. G. (2021). DNA metabarcoding reveals differences in distribution patterns and ecological preferences among genetic variants within some key freshwater diatom species. *Science of The Total Environment* 798, 149029. <https://doi.org/10.1016/j.scitotenv.2021.149029>
- Pérez-Burillo, J., Trobajo, R., Vasselon, V., Rimet, F., Bouchez, A., & Mann, D. G. (2020). Evaluation and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of Mediterranean rivers. *Science of The Total Environment* 727, 138445. <https://doi.org/10.1016/j.scitotenv.2020.138445>
- Phiri, C., Day, J., Chimbari, M., & Dhlomo, E. (2007). Epiphytic diatoms associated with a submerged macrophyte, *Vallisneria aethiopica*, in the shallow marginal areas of Sanyati Basin (Lake Kariba): A preliminary assessment of their use as biomonitoring tools. *Aquatic Ecology* 41(2), 169–181. <https://doi.org/10.1007/s10452-006-9073-z>
- Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344:1246752. <https://doi.org/10.1126/science.1246752>
- Pissaridou, P., Vasselon, V., Christou, A., Chonova, T., Papatheodoulou, A., Drakou, K., Tziortzis, I., Dörfliinger, G., Rimet, F., Bouchez, A., & Vasquez, M. I. (2021). Cyprus’ diatom diversity and the association of environmental and anthropogenic influences for ecological assessment of rivers using DNA metabarcoding. *Chemosphere* 272, 129814. <https://doi.org/10.1016/j.chemosphere.2021.129814>
- Podani J, Ódor P, Fattorini S, Strona G, Heino J, Schmera D (2018) Exploring multiple presence-absence data structures in ecology. *Ecol Model* 383:41–51. <https://doi.org/10.1016/j.ecolmodel.2018.05.012>
- Podani J, Schmera D (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. *Oikos* 120:1625–1638. <https://doi.org/10.1111/j.1600-0706.2011.19451.x>
- Potapova, M. G., & Charles, D. F. (2002). Benthic diatoms in USA rivers: Distributions along spatial and environmental gradients. *Journal of Biogeography*, 29(2), 167–187. <https://doi.org/10.1046/j.1365-2699.2002.00668.x>
- Prygiel, J., Carpentier, P., Almeida, S., Coste, M., Druart, J. C., Ector, L., Guillard, D., Honoré, M. A., Iserentant, R., Ledeganck, P., Lalanne-Cassou, C., Lesniak, C., Mercier, I., Moncaut, P., Nazart, M., Nouchet, N., Peres, F., Peeters, V., Rimet, F., & Zydek, N. (2002). Determination of the biological diatom index (IBD NF T 90–354): Results of an intercomparison exercise. *Journal of Applied Phycology* 14(1), 27–39. <https://doi.org/10.1023/A:1015277207328>

- Purvis A, Gittleman JL, Cowlishaw G, Mace GM (2000) Predicting extinction risk in declining species. *Proc R Soc Lond B* 267:1947–1952.
<https://doi.org/10.1098/rspb.2000.1234>
- R Core Team (2024) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Rabinowitz D (1981) Seven forms of rarity. In: Synge H (ed) *Biological aspects of rare plant conservation*. Wiley, New York, pp 205–217
- Reddin CJ, Bothwell JH, Lennon JJ (2015) Between-taxon matching of common and rare species richness patterns. *Glob Ecol Biogeogr* 24:1476–1486.
<https://doi.org/10.1111/geb.12372>
- Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T., ... & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological reviews*, 94(3), 849-873. <https://doi.org/10.1111/brv.12480>
- Rimet, F. (2012). Recent views on river pollution and diatoms. *Hydrobiologia*, 683(1), 1–24.
<https://doi.org/10.1007/s10750-011-0949-0>
- Rimet, F., & Bouchez, A. (2012). Biomonitoring river diatoms: Implications of taxonomic resolution. *Ecological Indicators* 15(1), 92–99.
<https://doi.org/10.1016/j.ecolind.2011.09.014>
- Rimet, F., Abarca, N., Bouchez, A., Kusber, W. H., Jahn, R., Kahlert, M., Keck, F., Kelly, M. G., Mann, David., G., Piuze, A., Trobajo, R., Tapolczai, K., Vasselon, V., & Zimmermann, J. (2018a). The potential of High-Throughput Sequencing (HTS) of natural samples as a source of primary taxonomic information for reference libraries of diatom barcodes. *Fottea* 18(1), 37–54. <https://doi.org/10.5507/fot.2017.013>
- Rimet, F., Barry-Martinet, R., Briand, J.-F., Chonova, T., Gassiole, G., Gusev, E., Kahlert, M., Karthick, B., Keck, F., Kelly, M., Kochoska, H., Kulikovskiy, M., Levkov, Z., Maltsev, Y., Mann, D., Nayak, P., Nicolosi Gelis, M. M., Pfannkuchen, M., Trobajo, R., ... Bouchez, A. (2024). 2024-05-28-Diat.barcode_release-version 12.1 [Data set]. Recherche Data Gouv. <https://doi.org/10.57745/9G8KKT>
- Rimet, F., Chonova, T., Gassiole, G., Gusev, E., Kahlert, M., Keck, F., Kelly, M., Kochoska, H., Kulikovskiy, M., Levkov, Z., Maltsev, Y., Mann, D., Nicolosi Gelis, M. M., Pfannkuchen, M., Trobajo, R., Vasselon, V., Vidakovic, D., Wetzel, C., Zimmermann, J., & Bouchez, A. (2018b). Diat.barcode, an open-access barcode library for diatoms (Version 18) [Data set]. Recherche Data Gouv.
<https://doi.org/10.15454/TOMBYZ/RGCYSY>
- Rimet, F., Gusev, E., Kahlert, M., Kelly, M. G., Kulikovskiy, M., Maltsev, Y., ... Bouchez, A. (2019). Diat.barcode, an open-access curated barcode library for diatoms. *Scientific Reports*, 9(1), 15116. <https://doi.org/10.1038/s41598-019-51500-6>
- Rissanen, A. J., Kurhela, E., Aho, T., Oittinen, T., & Tirola, M. (2010). Storage of environmental samples for guaranteeing nucleic acid yields for molecular microbiological studies. *Applied Microbiology and Biotechnology* 88(4), 977–984.
<https://doi.org/10.1007/s00253-010-2838-2>
- Rivera, 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: Complementary to understand turtle behavior. *PLOS ONE* 13(4), e0195770.
<https://doi.org/10.1371/journal.pone.0195770>

- Rivera, S. F., Vasselon, V., Bouchez, A., & Rimet, F. (2020). Diatom metabarcoding applied to large scale monitoring networks: Optimization of bioinformatics strategies using Mothur software. *Ecological Indicators* 109, 105775. <https://doi.org/10.1016/j.ecolind.2019.105775>
- Rodriguez-Ezpeleta, N., Morissette, O., Bean, C. W., Manu, S., Banerjee, P., Lacoursière-Roussel, A., Beng, K. C., Alter, S. E., Roger, F., Holman, L. E., Stewart, K. A., Monaghan, M. T., Mauvisseau, Q., Mirimin, L., Wangenstein, O. S., Antognazza, C. M., Helyar, S. J., de Boer, H., Monchamp, M. E., & Deiner, K. (2021). Trade-offs between reducing complex terminology and producing accurate interpretations from environmental DNA: Comment on “Environmental DNA: What’s behind the term?” by Pawlowski et al., (2020). *Molecular Ecology* 30(19), 4601–4605. <https://doi.org/10.1111/mec.15942>
- Sales, N. G., Wangenstein, O. S., Carvalho, D. C., & Mariani, S. (2019). Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. *Environmental DNA* 1(2), 198-207. <https://doi.org/10.1002/edn3.14>
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* 75(23), 7537–7541. <https://doi.org/10.1128/AEM.01541-09>
- Schulte, N. O., Craine, J. M., Leopold, D. R., Devitt, J. K., & Fierer, N. (2024). Diversity of freshwater algal assemblages across the United States as revealed by DNA metabarcoding. *Environmental DNA* 6(1), e477. <https://doi.org/10.1002/edn3.477>
- Sgarbi LF, Melo AS (2018) You don’t belong here: explaining the excess of rare species in terms of habitat, space and time. *Oikos* 127:497–506. <https://doi.org/10.1111/oik.04855>
- Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, Gilbert JA (2014) Conditionally rare taxa disproportionately contribute to Temporal changes in microbial diversity. *MBio* 5:10–1128. <https://doi.org/10.1128/mBio.01371-14>
- Šizling AL, Šizlingová E, Storch D, Reif J, Gaston KJ (2009) Rarity, commonness, and the contribution of individual species to species richness patterns. *Am Nat* 174:82–93. <https://doi.org/10.1086/599305>
- Smith MD, Knapp AK (2003) Dominant species maintain ecosystem function with non-random species loss. *Ecol Lett* 6:509–517. <https://doi.org/10.1046/j.1461-0248.2003.00454.x>
- Smol, J. P., & Stoermer, E. F. (Eds.). (2010). *The diatoms: Applications for the environmental and earth sciences*. Cambridge University Press.
- Smucker, N. J., Pilgrim, E. M., Nietch, C. T., Darling, J. A., & Johnson, B. R. (2020). DNA metabarcoding effectively quantifies diatom responses to nutrients in streams. *Ecological Applications* 30(8), e2205. <https://doi.org/10.1002/eap.2205>
- Smucker, N. J., Pilgrim, E. M., Nietch, C. T., Gains-Germain, L., Carpenter, C., Darling, J. A., Yuan, L. L., Mitchell, R. M., & Pollard, A. I. (2024). Using DNA metabarcoding to characterize national scale diatom-environment relationships and to develop indicators in streams and rivers of the United States. *Science of The Total Environment* 939, 173502. <https://doi.org/10.1016/j.scitotenv.2024.173502>

- Smucker, N. J., Pilgrim, E. M., Wu, H., Nietch, C. T., Darling, J. A., Molina, M., ... Yuan, L. L. (2022). Characterizing temporal variability in streams supports nutrient indicator development using diatom and bacterial DNA metabarcoding. *Science of the Total Environment*, 831, 154960. <https://doi.org/10.1016/j.scitotenv.2022.154960>
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored rare biosphere. *Proc Natl Acad Sci USA* 103:12115–12120. <https://doi.org/10.1073/pnas.0605127103>
- Stenger-Kovács C, B-Béres V, Buczkó K, Tapolczai K, Padisák J, Selmeczy GB, Lengyel E (2023) Diatom community response to inland water salinization: a review. *Hydrobiologia* 850:4627–4663. <https://doi.org/10.1007/s10750-023-05167-w>
- Stenger-Kovács, C., Buczko, K., Hajnal, E., & Padisák, J. (2007). Epiphytic, littoral diatoms as bioindicators of shallow lake trophic status: Trophic Diatom Index for Lakes (TDIL) developed in Hungary. *Hydrobiologia*, 589(1), 141–154. <https://doi.org/10.1007/s10750-007-0729-z>
- Stephen, C., & Joshi, L. (2010). Algal Biotechnology: An Emerging Resource with Diverse Application and Potential. In: Kole C, Michler CH, Abbott AG, Hall TC (Eds) *Transgenic Crop Plants*. Springer, 343–357. https://doi.org/10.1007/978-3-642-04812-8_9
- Stevens, B. M., Creed, T. B., Reardon, C. L., & Manter, D. K. (2023). Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. *Scientific Reports* 13(1), 9323. <https://doi.org/10.1038/s41598-023-36101-8>
- Stevenson, R. J., & Smol, J. P. (2015). Use of algae in ecological assessments. In J. D. Wehr, R. G. Sheath, & J. P. Kociolek (Eds.), *Freshwater algae of North America: Ecology and classification* (2nd ed., pp. 921–962). Academic Press. <https://doi.org/10.1016/B978-0-12-385876-4.00021-9>
- Stoof-Leichsenring, K. R., Dulias, K., Biskaborn, B. K., Pestryakova, L. A., & Herzsich, U. (2020). Lake-depth related pattern of genetic and morphological diatom diversity in boreal Lake Bolshoe Toko, Eastern Siberia. *PLOS ONE* 15(4), e0230284. <https://doi.org/10.1371/journal.pone.0230284>
- Székely AJ, Langenheder S (2014) The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *FEMS Microbiol Ecol* 87:102–112. <https://doi.org/10.1111/1574-6941.12195>
- Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21(8), 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Tambuzzi, S., Gentile, G., & Zoia, R. (2024). Forensic Diatom Analysis: Where Do We Stand and What Are the Latest Diagnostic Advances? *Diagnostics* 14(20): 2302. <https://doi.org/10.3390/diagnostics14202302>
- Tapolczai, K., Chonova, T., Fidlerová, D., Makovinská, J., Mora, D., Weigand, A., & Zimmermann, J. (2024). Molecular metrics to monitor ecological status of large rivers: Implementation of diatom DNA metabarcoding in the Joint Danube Survey 4. *Ecological Indicators*, 160, 111883. <https://doi.org/10.1016/j.ecolind.2024.111883>

- Tapolczai, K., Keck, F., Bouchez, A., Rimet, F., Kahlert, M., & Vasselon, V. (2019). Diatom DNA metabarcoding for biomonitoring: Strategies to avoid major taxonomical and bioinformatical biases limiting molecular indices capacities. *Frontiers in Ecology and Evolution*, 7, 409. <https://doi.org/10.3389/fevo.2019.00409>
- Tapolczai, K., Rimet, F., Ćirić, M., Ballot, A., Laplace-Tretyure, C., & Alric, B. (2025). A novel framework for phytoplankton biomonitoring: Trait assignment of 23S rRNA sequences. *Ecological Indicators* 173, 113361. <https://doi.org/10.1016/j.ecolind.2025.113361>
- Tapolczai, K., Selmeczy, G. B., Szabó, B., B-Béres, V., Keck, F., Bouchez, A., ... Padisák, J. (2021). The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. *Ecological Indicators*, 122, 107322. <https://doi.org/10.1016/j.ecolind.2020.107322>
- Tardy V, Mathieu O, Lévêque J, Terrat S, Chabbi A, Lemanceau P, Ranjard L, Maron P (2014) Stability of soil microbial structure and activity depends on microbial diversity. *Environ Microbiol Rep* 6:173–183. <https://doi.org/10.1111/1758-2229.12126>
- Tornés E, Colls M, Acuña V, Sabater S (2021) Duration of water flow interruption drives the structure and functional diversity of stream benthic diatoms. *Sci Total Env* 770:144675. <https://doi.org/10.1016/j.scitotenv.2020.144675>
- Turk Dermastia, T., Vascotto, I., Francé, J., Stanković, D., & Mozetič, P. (2023). Evaluation of the *rbcL* marker for metabarcoding of marine diatoms and inference of population structure of selected genera. *Frontiers in Microbiology* 14, 1071379. <https://doi.org/10.3389/fmicb.2023.1071379>
- Ulrich W (2005) Regional species richness of families and the distribution of abundance and rarity in a local community of forest hymenoptera. *Acta Oecol* 28:71–76. <https://doi.org/10.1016/j.actao.2004.10.003>
- Utermöhl, H. (1958). Zur vervollkommenung der quantitativen phytoplankton-methodik. *Mitteilungen der Internationalen Vereinigung für theoretische und angewandte Limnologie*, 9(1), 1–38. <https://doi.org/10.1080/05384680.1958.11904091>
- van Dam, H., Mertens, A., & Sinkeldam, J. (1994). A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherlands Journal of Aquatic Ecology*, 28(1), 117–133. <https://doi.org/10.1007/BF02334251>
- Van der Loos, L. M., & Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology* 30(13), 3270–3288. <https://doi.org/10.1111/mec.15592>
- Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics* 30(9), 418–426. <https://doi.org/10.1016/j.tig.2014.07.001>
- Van Elsas JD, Chiurazzi M, Mallon CA, Elhottová D, Křišťůfek V, Salles JF (2012) Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proc Natl Acad Sci USA* 109:1159–1164. <https://doi.org/10.1073/pnas.1109326109>
- Vasselon, V., Bouchez, A., Rimet, F., Jacquet, S., Trobajo, R., Corniquel, M., ... Domaizon, I. (2018). Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. *Methods in Ecology and Evolution*, 9(4), 1060–1069. <https://doi.org/10.1111/2041-210X.12960>

- Vasselon, V., Domaizon, I., Rimet, F., Kahlert, M., & Bouchez, A. (2017b). Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? *Freshwater Science* 36(1): 162–177. <https://doi.org/10.1086/690649>
- Vasselon, V., Rimet, F., Domaizon, I., Monnier, O., Reyjol, Y., & Bouchez, A. (2019). Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: Experience and developments from France Water Framework Directive networks. *Metabarcoding and Metagenomics* 3, e39646. <https://doi.org/10.3897/mbmg.3.39646>
- Vasselon, V., Rimet, F., Tapolczai, K., & Bouchez, A. (2017). Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecological Indicators*, 82, 1–12. <https://doi.org/10.1016/j.ecolind.2017.06.024>
- Vasselon, V., Rivera, S. F., Ács, É., Almeida, S. B., Andree, K. B., Apothéloz-Perret-Gentil, L., ... Paix, B. (2025). Proficiency testing and cross-laboratory method comparison to support standardisation of diatom DNA metabarcoding for freshwater biomonitoring. *Metabarcoding and Metagenomics*, 9, 1. <https://doi.org/10.3897/mbmg.9.133264>
- Veraart, A. J., Romani, A. M., Tornés, E., & Sabater, S. (2008). Algal response to nutrient enrichment in forested oligotrophic stream. *Journal of Phycology*, 44(3), 564–572. <https://doi.org/10.1111/j.1529-8817.2008.00503.x>
- Vincent H (2017) Experimental comparisons among very rare to widespread plant species of Switzerland. Doctoral dissertation, Philosophisch-naturwissenschaftliche Fakultät Universität Bern
- Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N. J., Cadotte, M. W., ... & Mouillot, D. (2017). Functional rarity: the ecology of outliers. *Trends in ecology & evolution*, 32(5), 356–367.
- 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. *Environmental Science & Technology* 49(13), 7597–7605. <https://doi.org/10.1021/es506158m>
- Vivante A-L, Garmyn D, Maron P-A, Nowak V, Piveteau P (2013) Microbial diversity and structure are drivers of the biological barrier effect against *Listeria monocytogenes* in soil. *PLoS ONE* 8:e76991. <https://doi.org/10.1371/journal.pone.0076991>
- von Meijenfeldt FB, Hogeweg P, Dutilh BE (2023) A social niche breadth score reveals niche range strategies of generalists and specialists. *Nat Ecol Evol* 7:768–781. <https://doi.org/10.1038/s41559-023-02027-7>
- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... Davies, P. (2010). Global threats to human water security and river biodiversity. *Nature*, 467(7315), 555–561. <https://doi.org/10.1038/nature09440>
- Wang N, Xiong J, Wang X et al (2018) Relationship between phytoplankton community and environmental factors in landscape water with high salinity in a coastal City of China. *Environ Sci Pollut Res* 25:28460–28470. <https://doi.org/10.1007/s11356-018-2886-1>
- Wang, B., Li, R., Lan, X., Kong, D., Liu, X., & Xie, S. (2024). Benthic diatom eDNA metabarcoding for ecological assessment of an urban river: A comparison with morphological method. *Ecological Indicators*, 166, 112302. <https://doi.org/10.1016/j.ecolind.2024.112302>

- Wang, H., Liu, K., He, Z., Chen, Y., Hu, Z., Chen, W., Leaw, C. P., & Chen, N. (2024). Extensive intragenomic variations of the 18S rDNA V4 region in the toxigenic diatom species *Pseudo-nitzschia multistriata* revealed through high-throughput sequencing. *Marine Pollution Bulletin* 201, 116198. <https://doi.org/10.1016/j.marpolbul.2024.116198>
- Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A. M., Willassen, E., Wyler, S. A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., & Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of The Total Environment* 678, 499–524. <https://doi.org/10.1016/j.scitotenv.2019.04.247>
- Wetzel, R. G., & Likens, G. E. (2000). *Limnological analyses* (3rd ed.). Springer. <https://doi.org/10.1007/978-1-4757-3250-4>
- Winfrey R, Fox JW, Williams NM, Reilly JR, Cariveau DP (2015) Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. *Ecol Lett* 18:626–635. <https://doi.org/10.1111/ele.12424>
- Xiao J, Zhong Z, Wang C, Li M, Wen Q, Lin X, Luo Z, Qian Z, Li M, Deng C (2023) Rare species are significant in harsh environments and unstable communities: based on the changes of species richness and community stability in different sub-assemblages. *Sustainability* 15:13994. <https://doi.org/10.3390/su151813994>
- Xue Y, Chen H, Yang JR, Liu M, Huang B, Yang J (2018) Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. *ISME J* 12:2263–2277. <https://doi.org/10.1038/s41396-018-0159-0>
- Yan Y, Kuramae EE, Klinkhamer PGL, Van Veen JA (2015) Revisiting the Dilution procedure used to manipulate microbial biodiversity in terrestrial dystems. *Appl Environ Microbiol* 81:4246–4252. <https://doi.org/10.1128/AEM.00958-15>
- Yenni G, Adler PB, Ernest SKM (2012) Strong self-limitation promotes the persistence of rare species. *Ecology* 93:456–461. <https://doi.org/10.1890/11-1087.1>
- Yu F, Skidmore AK, Wang T, Huang J, Ma K, Groen TA (2017) *Rhododendron* diversity patterns and priority conservation areas in China. *Divers Distrib* 23:1143–1156. <https://doi.org/10.1111/ddi.12607>
- Zhan, A., Hulák, M., Sylvester, F., Huang, X., Adebayo, A. A., Abbott, C. L., Adamowicz, S. J., Heath, D. D., Cristescu, M. E., & MacIsaac, H. J. (2013). High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution* 4(6), 558–565. <https://doi.org/10.1111/2041-210X.12037>
- Zimmermann, J., Abarca, N., Enk, N., Skibbe, O., Kusber, W. H., & Jahn, R. (2014). Taxonomic reference libraries for environmental barcoding: A best practice example from diatom research. *PLoS ONE* 9(9), e108793. <https://doi.org/10.1371/journal.pone.0108793>
- Zimmermann, J., Glöckner, G., Jahn, R., Enke, N., & Gemeinholzer, B. (2015). Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, 15(3), 526–542. <https://doi.org/10.1111/1755-0998.12336>
- Zimmermann, J., Jahn, R., & Gemeinholzer, B. (2011). Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution* 11(3), 173–192. <https://doi.org/10.1007/s13127-011-0050-6>

- Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., & Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* 28(8), 1857–1862.
<https://doi.org/10.1111/mec.15060>
- Zizka A, Steege HT, Pessoa MDCR, Antonelli A (2018) Finding needles in the haystack: where to look for rare species in the American tropics. *Ecography* 41:321–330.
<https://doi.org/10.1111/ecog.02192>

Appendixes

Supplementary table 1. Sampling sites in Croatia and Hungary, including site codes, locations, geographic coordinates, and sampling dates for the spring and autumn 2022 campaigns.

River	Code	Country	Site	GPS-N	GPS-E	Date	
						spring	autumn
Bednja	Bed 1	Croatia	Stažnjevec	46.240974	16.169081	07 Apr 2022	22 Sep 2022
Bednja	Bed 2	Croatia	Mali Bukovec	46.291004	16.73833	07 Apr 2022	22 Sep 2022
Drava	Drava 2	Croatia	Podravska Moslavina	45.790453	17.980645	05 May 2022	03 Oct 2022
Drava	Drava 3	Croatia	Gat, Petrovo selo	45.72775	18.319752	05 May 2022	03 Oct 2022
Drava	Drava 4	Croatia	Nard	45.667341	18.488948	05 May 2022	03 Oct 2022
Drava	Drava 5	Croatia	Petrijevci	45.617589	18.583732	05 May 2022	03 Oct 2022
Drava	Drava 6	Croatia	Osijek	45.572097	18.648607	05 May 2022	04 Oct 2022
Drava	Drava 7	Croatia	Sarvaš	45.553034	18.864247	05 May 2022	-
Drava	Drava 8	Croatia	Novo Virje	46.110226	17.152535	11 Apr 2022	23 Sep 2022
Drava	Drava 9	Croatia	Ledine Molvanske	46.128741	17.070945	11 Apr 2022	23 Sep 2022
Drava	Drava 10	Croatia	Donji Miholjac-Dravasabolc	45.783107	18.200611	05 May 2022	03 Oct 2022
Drava	Drava 11	Croatia	Terezino Polje-Barč	45.942418	17.46927	11 Apr 2022	23 Sep 2022
Drava	Drava 12	Croatia	Botovo-Ortilos	46.241306	16.938276	11 Apr 2022	23 Sep 2022
Drava	Drava 13	Croatia	Legrad	46.29715	16.879116	11 Apr 2022	22 Sep 2022
Duna	Dun-Bat	Croatia	Batina	45.874713	18.829242	05 May 2022	04 Oct 2022
Duna	Dun-Ilok	Croatia	Ilok	45.230603	19.37897	06 May 2022	04 Oct 2022
Karašica	Kar 1	Croatia	Črnkovci	45.708672	18.294293	05 May 2022	03 Oct 2022
Karašica	Kar 2	Croatia	Crnac - Krčenik	45.728336	17.966283	05 May 2022	-
Krapina	Kra 1	Croatia	Zaprešić	45.834771	15.822804	11 Apr 2022	16 Oct 2022
Krapina	Kra 2	Croatia	Krapina	46.104752	16.201174	07 Apr 2022	
Kupa	Kup 1	Croatia	Sisak	45.483641	16.36888	30 Mar 2022	-
Kupa	Kup 2	Croatia	Bubnjarci	45.645271	15.356443	11 Apr 2022	-
Kupa	Kup 3	Croatia	Ozalj	45.614169	15.479493	11 Apr 2022	11 Oct 2022
Mura	Mura 1	Croatia	Goričan	46.419121	16.691248	07 Apr 2022	22 Sep 2022
Mura	Mura 2	Croatia	Mursko Središće	46.51457	16.440522	07 Apr 2022	22 Sep 2022
Mura	Mura 3	Croatia	Veliki Pažut	46.325253	16.871193	11 Apr 2022	22 Sep 2022
Odra	Odr 1	Croatia	Sisak	45.500181	16.348948	30 Mar 2022	-
Orljava	Orl 1	Croatia	Pleternici	45.286226	17.805124	05 May 2022	03 Oct 2022
Orljava	Orl 2	Croatia	Lužani	45.145141	17.704758	06 May 2022	-
Plitvica	Plit 1	Croatia	Kučana Gornjeg	46.268257	16.380981	07 Apr 2022	22 Sep 2022
Plitvica	Plit 2	Croatia	Veliki Bukovec	46.294875	16.714555	07 Apr 2022	22 Sep 2022
Sava	Sava 1	Croatia	Županje	45.040493	18.695919	06 May 2022	24 Oct 2022
Sava	Sava 2	Croatia	Bosne	45.060156	18.5047	06 May 2022	24 Oct 2022

River	Code	Country	Site	GPS-N	GPS-E	Date	
						spring	autumn
Sava	Sava 3	Croatia	Broda	45.139754	18.073944	06 May 2022	24 Oct 2022
Sava	Sava 4	Croatia	Orljave, Sl. Kobaš	45.099068	17.737857	06 May 2022	24 Oct 2022
Sava	Sava 5	Croatia	Jasenovac	45.268811	16.915474	12 May 2022	
Sava	Sava 6	Croatia	Lukavec	45.401472	16.53887	12 May 2022	18 Oct 2022
Sava	Sava 7	Croatia	Galdovo	45.479194	16.384163	30 Mar 2022	18 Oct 2022
Sava	Sava 8	Croatia	Jankomir	45.788939	15.859353	12 May 2022	11 Oct 2022
Sava	Sava 9	Croatia	Drenje-Jesenice	45.862452	15.688058	12 May 2022	11 Oct 2022
Sava	Sava 10	Croatia	Rugvica	45.74637	16.229709	30 Mar 2022	-
Sava	Sava 10	Croatia	Rugvica	45.74637	16.229709	18 Oct 2022	-
Sava	Sava 11	Croatia	Vrbasa	45.13404	17.672908	06 May 2022	24 Oct 2022
Sava	Sava 12	Croatia	Topolje	45.647778	16.339606	30 Mar 2022	-
Sava	Sava 13	Croatia	Lonje i Trebeža	45.362241	16.757772	12 May 2022	25 Oct 2022
Sava	Sava 14	Croatia	Stare Gradiške	45.16701	17.25817	12 May 2022	25 Oct 2022
Sava	Sava 15	Croatia	Medsave	45.834317	15.771069	11 Apr 2022	11 Oct 2022
Sava	Sava 16	Croatia	Račinovci	44.851325	18.959543	06 May 2022	24 Oct 2022
Una	Una 1	Croatia	Uštica	45.261594	16.912873	12 May 2022	25 Oct 2022
Una	Una 2	Croatia	Hrvatska Kostajnica	45.222096	16.549762	30 Mar 2022	18 Oct 2022
Una	Una 3	Croatia	Bosanski Novi	45.049785	16.372592	30 Mar 2022	18 Oct 2022
Által-ér	ÁLT1	Hungary	Dunaalmás	47.731578	18.32382	06 Apr 2022	18 Oct 2022
Bitva	M26	Hungary	Nyárád	47.292183	17.373021	13 Apr 2022	26 Oct 2022
Cinca	M31	Hungary	Kemenesszentmárton	47.290994	17.171454	13 Apr 2022	26 Oct 2022
Cuhai-Bakony-ér	CUBA1	Hungary	Nagyszentjános	47.70461	17.856468	06 Apr 2022	18 Oct 2022
Csikvándi-Bakony-ér	M12	Hungary	Csikvánd	47.45951	17.44663	14 Apr 2022	-
Csömöc-Herpenyő	Csömöc1	Hungary	Sárvár	47.223252	16.946444	13 Apr 2022	25 Oct 2022
Csömöc-Herpenyő	Csömöc2	Hungary	Döröske	47.023333	16.701605	24 Oct 2022	12 Apr 2022
Damádsdi-patak	DAM01	Hungary	Ipolydamásdi	47.831161	18.853865	05 Apr 2022	18 Oct 2022
Dobroda-patak	DOB01	Hungary	Lítke	48.207931	19.594151	05 Apr 2022	17 Oct 2022
Duna	DUN1	Hungary	Szob	47.816204	18.862622	05 Apr 2022	18 Oct 2022
Duna	DUN2	Hungary	Nyergesújfalú	47.762361	18.540508	06 Apr 2022	18 Oct 2022
Duna	DUN3	Hungary	Szőny-Komárom	47.740468	18.161742	06 Apr 2022	18 Oct 2022
Duna	DUN4	Hungary	Gönyű	47.738873	17.825387	06 Apr 2022	18 Oct 2022
Duna	DUN5	Hungary	Nagybajcs	47.768855	17.695974	07 Apr 2022	19 Oct 2022
Fekete-víz	FEK01	Hungary	Patvarc	48.03547	19.21186	04 Apr 2022	17 Oct 2022
Gerence-patak	M17	Hungary	Takácsi	47.402118	17.471609	14 Apr 2022	25 Oct 2022
Hajagos	Hajagos1	Hungary	Vínár	47.305229	17.292561	13 Apr 2022	26 Oct 2022
Hosszú-víz	R25	Hungary	Csénye	47.22872	16.840857	13 Apr 2022	25 Oct 2022
Ipoly	IPO01	Hungary	Szécsény	48.097534	19.482762	04 Apr 2022	17 Oct 2022
Ipoly	IPO02	Hungary	Ipolytarnóc	48.23556	19.605588	05 Apr 2022	17 Oct 2022
Ipoly	IPO03	Hungary	Ipolyvece	48.065698	19.107886	17 Oct 2022	-

River	Code	Country	Site	GPS-N	GPS-E	Date	
						spring	autumn
Ipoly	IPO04	Hungary	Tésa	48.031432	18.82434	05 Apr 2022	18 Oct 2022
Ipoly	IPO05	Hungary	Ipolydamásdi	47.846207	18.827692	05 Apr 2022	18 Oct 2022
Kodó	M34	Hungary	Celldömölk	47.210633	17.159439	11 Apr 2022	-
Kozár-borzó	R22	Hungary	Bozsai-Bárdos	47.207085	16.752087	12 Apr 2022	25 Oct 2022
Lókos-patak	LÓK01	Hungary	Dejtár	48.04637	19.202505	04 Apr 2022	17 Oct 2022
Marcal	M16	Hungary	Malomsok	47.448847	17.383294	14 Apr 2022	25 Oct 2022
Marcal	M30	Hungary	Szergény	47.318717	17.274985	13 Apr 2022	26 Oct 2022
Marcal	M37	Hungary	Boba	47.180045	17.203338	11 Apr 2022	26 Oct 2022
Marcal	M46	Hungary	Gágánfa	47.019367	17.178018	11 Apr 2022	24 Oct 2022
Marcal	MAR1	Hungary	Rábapatoná	47.609778	17.514591	07 Apr 2022	19 Oct 2022
Ménés-patak	MÉN01	Hungary	Szécsény-Ludányhalászi	48.110939	19.515189	04 Apr 2022	-
Mosoni-Duna	MOS1	Hungary	Győrszentiván	47.733882	17.755304	07 Apr 2022	18 Oct 2022
Pándzsa	PÁN1	Hungary	Győr	47.645157	17.663524	07 Apr 2022	19 Oct 2022
Pápai-Bakony ér	M24	Hungary	Mezőlak	47.340737	17.401141	13 Apr 2022	25 Oct 2022
Perint	R07	Hungary	Gencsapáti	47.285246	16.591909	12 Apr 2022	24 Oct 2022
Rába	R38	Hungary	Csákánydoroszló	46.980498	16.526671	12 Apr 2022	24 Oct 2022
Rába	R45	Hungary	Rum	47.126851	16.85048	12 Apr 2022	24 Oct 2022
Rába	R50	Hungary	Sárvár	47.245063	16.95256	13 Apr 2022	25 Oct 2022
Rába	RAB1	Hungary	Rábapatoná	47.621064	17.488968	07 Apr 2022	19 Oct 2022
Rába	Rába 1	Hungary	Pápoc	47.423091	17.132876	13 Apr 2022	25 Oct 2022
Rátka-patak	R24	Hungary	Porpác	47.241012	16.798965	13 Apr 2022	-
Sokorói-Bakony-ér	SOB1	Hungary	Győrszemere	47.561259	17.574898	07 Apr 2022	19 Oct 2022
Sorok-patak	R17	Hungary	Gyanógeregye	47.125671	16.760266	12 Apr 2022	24 Oct 2022
Szávai-csatorna	SZAV1	Hungary	Kisbajcs	47.746114	17.683105	07 Apr 2022	19 Oct 2022
Torna-patak	első hely	Hungary	Kisberzsény	47.103812	17.261161	11 Apr 2022	24 Oct 2022
Únyi-patak	ÚNY1	Hungary	Sárisáp	47.67255	18.675955	06 Apr 2022	18 Oct 2022
Vörös-Patak	R34	Hungary	Vasszentmihály	46.970039	16.407761	12 Apr 2022	-

Supplementary table 2. List of benthic diatom species included in the rarity analysis, with taxonomic authorship, species codes, and classification according to the four rarity criteria: small population size (SP), habitat specificity (HS), geographic range restriction (GR), and overall rarity (AR). An “x” indicates that the species met the corresponding rarity criterion.

Species name	Authors	Code	SP	HS	GR	AR
<i>Achnantheidium affine</i>	(Grunow) Czarnecki	ACAF				
<i>Achnantheidium atomoides</i>	Monnier, Lange-Bertalot & Ector	ADAM				
<i>Achnantheidium catenatum</i>	(Bily & Marvan) Lange-Bertalot	ADCT	x	x	x	x
<i>Achnantheidium delmontii</i>	Peres, Le Cohu & Barthes	ADMO				
<i>Achnantheidium druartii</i>	Rimet & Couté in Rimet & al.	ADRU				
<i>Achnantheidium eutrophilum</i>	(Lange-Bertalot) Lange-Bertalot	ADEU				
<i>Achnantheidium exiguum</i>	(Grunow) Czarnecki	ADEG	x			
<i>Achnantheidium exile</i>	(Kützing) Heiberg	ADEX	x	x	x	x
<i>Achnantheidium gracillimum</i>	(Meister) Heiberg	ADGL	x	x	x	x
<i>Achnantheidium jackii</i>	Rabenhorst	ADJK				
<i>Achnantheidium latecephalum</i>	Kobayasi	ADLA				
<i>Achnantheidium lineare</i>	W. Smith	ACLI				
<i>Achnantheidium minutissimum</i>	(Kützing) Czarnecki	ADMI				
<i>Achnantheidium pyrenaicum</i>	(Hustedt) Kobayasi	ADPY				
<i>Achnantheidium rostrumpyrenaicum</i>	Jüttner & Cox	ARPY				
<i>Achnantheidium rivulare</i>	Potapova & Ponader	ADRI				
<i>Achnantheidium saprophilum</i>	(Kobayasi & Mayama) Round & Bukhtiyarova	ADSA	x	x	x	x
<i>Achnantheidium sp.</i>		ACHD				
<i>Achnantheidium straubianum</i>	(Lange-Bertalot) Lange-Bertalot	ADSB				
<i>Achnantheidium subatomus</i>	(Hustedt) Lange-Bertalot	ADSU				
<i>Adlafia minuscula</i>	(Grunow) Lange-Bertalot	ADMS	x			
<i>Amphipleura pellucida</i>	Kützing (Kützing)	APEL	x	x	x	x
<i>Amphora copulata</i>	(Kützing) Schoeman & Archibald	ACOP				
<i>Amphora inariensis</i>	Krammer	AINA				
<i>Amphora indistincta</i>	Levkov	AMID				
<i>Amphora lange-bertalotii</i>	Levkov & Metzeltin	AMLB	x			
<i>Amphora lange-bertalotii</i> var. <i>tenuis</i>	Levkov & Metzeltin	ALGT	x	x	x	x
<i>Amphora minutissima</i>	W. Smith	AMNU				
<i>Amphora ovalis</i>	(Kützing) Kützing	AOVA				
<i>Amphora pediculus</i>	(Kützing) Grunow	APED				
<i>Amphora sp.</i>		AMPH	x			
<i>Amphora subatomus</i>	Levkov	AMSM	x	x	x	x
<i>Aneumastus stroesei</i>	(Østrup) D.G. Mann & Stickle	ANSS	x	x	x	x
<i>Anomoeoneis sphaerophora</i>	(Ehrenberg) Pfitzer	ASPH	x	x	x	x
<i>Asterionella formosa</i>	Hassall	AFOR				
<i>Aulacoseira ambigua</i>	(Grunow) Simonsen	AAMB				
<i>Aulacoseira ambigua</i> f. <i>japonica</i>	Tuji & Williams	AUAJ	x	x	x	x
<i>Aulacoseira alpigena</i>	(Grunow) Krammer	AUAL		x	x	x
<i>Aulacoseira granulata</i>	(Ehrenberg) Simonsen	AUGR				
<i>Aulacoseira granulata</i> var. <i>angustissima</i>	(O. Müller) Simonsen	AUGA				
<i>Aulacoseira pfaffiana</i>	(Reinsch) Krammer	AUPF	x	x	x	x
<i>Aulacoseira pusilla</i>	(Meister) Tuji & Houki	AUPU				
<i>Bacillaria paxillifera</i>	(O.F. Müller) Hendey	BPAX				
<i>Brachysira neglectissima</i>	Lange-Bertalot	BNEG				
<i>Brachysira neoexilis</i>	Lange-Bertalot	BNEO	x	x	x	x
<i>Caloneis amphisbaena</i>	(Bory) Cleve	CAMP				
<i>Caloneis bacillum</i>	(Grunow) Cleve	CBAC	x	x	x	x

Species name	Authors	Code	SP	HS	GR	AR
<i>Caloneis fontinalis</i>	(Grunnow) Lange-Bertalot & Reichardt	CFON	x	x	x	x
<i>Caloneis lancettula</i>	(Schulz-Danzig) Lange-Bertalot & Witkowski	CLCT				
<i>Caloneis schumanniana</i>	(Grunow) Cleve	CSHU	x			
<i>Caloneis silicula</i>	(Ehrenberg) Cleve	CSIL	x	x	x	x
<i>Caloneis sp.</i>		CALO	x			
<i>Centrales</i>		CTRQ				
<i>Cocconeis disculus</i>	(Schumann) Cleve	CDIS	x			
<i>cocconeis neothumensis</i>	Krammer	CNTH	x			
<i>Cocconeis pediculus</i>	Ehrenberg	CPED				
<i>Cocconeis placentula sensu stricto</i>	Ehrenberg	COAG				
<i>Cocconeis pseudolineata</i>	(Geitler) Lange-Bertalot	COPL	x			
<i>Conticribra weissflogii</i>	(Grunow) Stachura-Suchoples & Williams	CTWE				
<i>Craticula accomoda</i>	(Hustedt) D.G. Mann	CRAC				
<i>Craticula buderi</i>	(Hustedt) Lange-Bertalot	CRBU				
<i>Craticula cuspidata</i>	(Kützing) D.G. Mann	CRCU	x			
<i>Craticula halophila</i>	(Grunow) D.G. Mann	CHAL	x			
<i>Craticula molestiformis</i>	(Hustedt) Lange-Bertalot	CMFO				
<i>Craticula subminuscule</i>	(Manguin) C.E. Wetzel & L. Ector	CSNU				
<i>Ctenophora pulchella</i>	(Ralfs ex Kützing) D.M. Williams & Round	CTPU				
<i>Cyclostephanos dubius</i>	(Fricke) Round	CDUB				
<i>Cyclostephanos invisitatus</i>	(Hohn & Hellerman) Theriot, Stoermer & Håkansson	CINV				
<i>Cyclotella atomus</i>	Hustedt	CATO				
<i>Cyclotella atomus var. gracilis</i>	Genkal & Kiss	CAGR	x	x	x	x
<i>Cyclotella distinguenda</i>	Hustedt	CDTG	x			
<i>Cyclotella meduanae</i>	Germain	CMDU	x			
<i>Cyclotella meneghiniana</i>	Kützing	CMEN				
<i>Cymatopleura apiculata</i>	W. Smith (Ralfs)	CYAP	x			
<i>Cymbella affinis</i>	Kützing	CAFF				
<i>Cymbella compacta</i>	Østrup	CCMP				
<i>Cymbella cymbiformis</i>	C. Agardh	CCYM	x			
<i>Cymbella excisa</i>	Kützing	CAEX				
<i>Cymbella excisiformis</i>	Krammer	CEXF	x			
<i>Cymbella neocistula</i>	Krammer	CNCI	x			
<i>Cymbella neolanceolata</i>	W. Silva	CNLC	x			
<i>Cymbella tumida</i>	(Brébisson) Van Heurck	CTUM				
<i>Cymbopleura frequens</i>	Krammer	CBFQ	x	x	x	x
<i>Cymbopleura naviculiformis</i>	(Auerswald) Krammer	CBNA	x	x	x	x
<i>Cymbopleura sp.</i>		CBPL	x	x	x	x
<i>Denticula tenuis</i>	Kützing	DTEN				
<i>Diadesmis conservacea</i>	Kützing	DCOF				
<i>Diatoma ehrenbergii</i>	Kützing	DEHR				
<i>Diatoma mesodon</i>	(Ehrenberg) Kützing	DMEN	x	x	x	x
<i>Diatoma moniliformis</i>	Kützing	DMON				
<i>Diatoma polonica</i>	M.Bak, H.Lange-Bertalot, J. Nosek, Z. Jakubowska & M.Kielbasa	DPOL				
<i>Diatoma tenuis</i>	C. Agardh	DITE				
<i>Diatoma vulgaris</i>	Bory	DVUL				
<i>Didymosphenia geminata</i>	(Lyngbye) M. Schmidt	DGEM	x			
<i>Diploneis calcicolafrequens</i>	Lange-Bertalot & Fuhrmann	DCFQ	x			
<i>Diploneis calcilacustris</i>	Lange-Bertalot & Fuhrmann	DCAL	x			
<i>Diploneis elliptica</i>	(Kützing) Cleve	DELL	x	x	x	x
<i>Diploneis fontanella</i>	Lange-Bertalot	DFON	x	x	x	x
<i>Diploneis marginestriata</i>	Hustedt	DMAR	x			
<i>Diploneis oblongella</i>	(Naegeli) Cleve-Euler	DOBL				
<i>Diploneis oculata</i>	(Brébisson) Cleve	DOCU	x			
<i>Diploneis parva</i>	Cleve	DPAR	x			
<i>Diploneis separanda</i>	Lange-Bertalot	DSEP	x	x	x	x
<i>Diploneis sp.</i>		DIPL	x	x	x	x
<i>Discostella asterocostata</i>	(Lin Xie & Cai) Houk & Klee	DATC				
<i>Discostella pseudostelligera</i>	(Hustedt) Houk & Klee emend. Genkal	DPSG				

Species name	Authors	Code	SP	HS	GR	AR
<i>Discostella stelligera</i>	(Cleve & Grun.) Houk & Klee	DSTE				
<i>Discostella woltereckii</i>	(Hustedt) Houk & Klee	DWOL				
<i>Dorofeyukea kotschyi</i>	(Grunow) Kulikovskiy, Kociolek, Tusset & T.Ludwig	DKOT	x			
<i>Ellerbeckia arenaria</i>	(Moore ex Ralfs) Crawford	EARE	x	x	x	x
<i>Encyonema caespitosum</i>	Kützing	ECAE				
<i>Encyonema langebertalotii</i>	Krammer	ELBV				
<i>Encyonema leibleinii</i>	(C. Agardh) W.J. Silva, R. Jahn, T.A. Veiga Ludwig & M. Menezes	ELEI	x			
<i>Encyonema minutum</i>	(Hilse) D.G. Mann	ENMI				
<i>Encyonema reichardtii</i>	(Krammer) D.G. Mann	ENRE		x	x	x
<i>Encyonema silesiacum</i>	(Bleisch) D.G. Mann	ESLE				
<i>Encyonema ventricosum</i>	(C. Agardh) Grunow	ENVE				
<i>Encyonema vulgare</i>	Krammer	EVUL	x			
<i>Encyonopsis microcephala</i>	(Grunow) Krammer	ENCM				
<i>Encyonopsis minuta</i>	Krammer & Reichardt	ECPM				
<i>Encyonopsis subminuta</i>	Krammer & Reichardt	ESUM				
<i>Encyonopsis sp.</i>		ENCP	x			
<i>Entomoneis paludosa</i>	(W. Smith) Reimer	EPAL	x			
<i>Epithemia gibba</i>	Kützing	EGBA	x			
<i>Epithemia sores</i>	Kützing	ESOR	x			
<i>Eunotia bilunaris</i>	(Ehrenberg) Schaarschmidt	EBLU				
<i>Eunotia formicina</i>	Lange-Bertalot	EFOM	x	x	x	x
<i>Eunotia minor</i>	(Kützing) Grunow	EMIN	x			
<i>Eunotia sp.</i>		EUNO		x	x	x
<i>Fallacia lenzii</i>	(Hustedt) Lange-Bertalot	FLEN				
<i>Fallacia mitis</i>	(Hustedt) D.G. Mann	FMIT	x			
<i>Fallacia pygmaea</i>	(Kützing) A.J. Stickle & D.G. Mann	FPYG				
<i>Fallacia subhamulata</i>	(Grunow) D.G. Mann	FSBH	x			
<i>Fallacia subcladula</i>	(Hustedt) D.G. Mann	FSLU	x			
<i>Fistulifera saprophila</i>	(Lange-Bertalot & Bonik) Lange-Bertalot	FSAP	x			
<i>Fragilaria aquaplus</i>	Lange-Bertalot & Ulrich	FAQU	x	x	x	x
<i>Fragilaria austriaca</i>	(Grunow) Lange-Bertalot	FAUT				
<i>Fragilaria capucina</i>	Desmarziers	FCAP				
<i>Fragilaria crotonensis</i>	Kitton	FCRO				
<i>Fragilaria famelica</i>	(Kützing) Lange-Bertalot	FFAM	x			
<i>Fragilaria gracilis</i>	Østrup	FGRA				
<i>Fragilaria mesolepta</i>	Rabenhorst	FMES				
<i>Fragilaria pararumpens</i>	Lange-Bertalot, G.Hofmann & Werum	FPRU	x	x	x	x
<i>Fragilaria pectinalis</i>	(O.F.Müller) Lyngbye	FPEC				
<i>Fragilaria perdelicissima</i>	Lange-Bertalot & Van de Vijver	FPDE	x			
<i>Fragilaria perminuta</i>	(Grunow) Lange-Bertalot	FPEM		x	x	x
<i>Fragilaria recapitulata</i>	Lange-Bertalot & Metzeltin	FRCP	x			
<i>Fragilaria rumpens</i>	(Kützing) Carlson	FRUM				
<i>Fragilaria saxoplantonica</i>	Lange-Bertalot & Ulrich	FSXP	x			
<i>Fragilaria sp.</i>		FRAG				
<i>Fragilaria tenera</i>	(W. Smith) Lange-Bertalot	FTEN				
<i>Fragilaria vaucheriae</i>	(Kützing) Petersen	FVAU				
<i>Frustulia vulgaris</i>	(Thwaites) De Toni	FVUL				
<i>Gomphonema acidoclinatum</i>	Lange-Bertalot & Reichardt	GADC	x			
<i>Gomphonema acuminatum</i>	Ehrenberg	GACU				
<i>Gomphonema angustatum</i>	(Kützing) Rabenhorst	GANG		x	x	x
<i>Gomphonema angustum</i>	C. Agardh	GANT	x	x	x	x
<i>Gomphonema augur</i>	Ehrenberg	GAUG	x			
<i>Gomphonema calcareum</i>	Cleve	GNLC	x	x	x	x
<i>Gomphonema capitatum</i>	Ehrenberg	GCAP	x	x	x	x
<i>Gomphonema clavatulum</i>	Reichardt	GCVT				
<i>Gomphonema clavatum</i>	Ehrenberg	GCLA				
<i>Gomphonema exilissimum</i>	(Grunow) Lange-Bertalot & Reichardt	DEXL	x			
<i>Gomphonema italicum</i>	Kützing	GITA	x			
<i>Gomphonema lateripunctatum</i>	Reichardt & Lange-Bertalot	GLAT	x			
<i>Gomphonema micropus</i>	Kützing	GMIC				

Species name	Authors	Code	SP	HS	GR	AR
<i>Gomphonema minutum</i>	(C. Agardh) C. Agardh	GMIN				
<i>Gomphonema olivaceoides</i>	Hustedt	GOLD				
<i>Gomphonema olivaceum</i>	(Hornemann) Brébisson	GOLI				
<i>Gomphonella olivaceolacuum</i>	(Lange-Bertalot & Reichardt) Lange-Bertalot & Reichardt	GLOL	x			
<i>Gomphonema pala</i>	Reichardt	GOPA	x			
<i>Gomphonema parvulum</i>	(Kützing) Kützing	GPAP				
<i>Gomphonema pumilum</i> var. <i>pumilum</i>	(Grunow) Reichardt & Lange- Bertalot	GPUM	x			
<i>Gomphonema pumilum</i> var. <i>rigidum</i>	(Grunow) Reichardt & Lange- Bertalot	GPRI				
<i>Gomphonema saprophilum</i>	(Lange-Bertalot & Reichardt) Abarca, R. Jahn, J. Zimmermann & Enke	GPAS				
<i>Gomphonema cf.</i> <i>ristovskae</i>	Levkov & Tofilovska	GRIS		x	x	x
<i>Gomphonema sp.</i>		GOMP				
<i>Gomphonema subclavatum</i>	Grunow (Grunow)	GSCL				
<i>Gomphonema supertergestinum</i>	E. Reichardt	GSUT				
<i>Gomphonema tergestinum</i>	(Grunow) M. Schmidt	GTER				
<i>Gomphonema truncatum</i>	Ehrenberg	GTRU				
<i>Gomphosphenia holmquistii</i>	(Foged) Lange-Bertalot	GPHQ	x			
<i>Gomphosphenia lingulatifomis</i>	(Lange-Bertalot & Reichardt) Lange-Bertalot	GPLI	x	x	x	x
<i>Gyrosigma acuminatum</i>	(Kützing) Rabenhorst	GYAC	x			
<i>Gyrosigma attenuatum</i>	(Kützing) Rabenhorst	GYAT	x			
<i>Gyrosigma kuetzingii</i>	(Grunow) Cleve	GYKU				
<i>Gyrosigma obtusatum</i>	(Sullivan & Wormley) C. S. Boyer	GYOB	x			
<i>Gyrosigma sciotoense</i>	(W. S. Sullivan) Cleve	GSCI				
<i>Halamphora montana</i>	(Krasske) Levkov	HLMO				
<i>Halamphora normanii</i>	(Rabenhorst) Levkov	HNOR				
<i>Halamphora veneta</i>	(Kützing) Levkov	HVEN				
<i>Hannaea arcus</i>	(Ehrenberg) R.M. Patrick	HARC	x			
<i>Hantzschia abundans</i>	Lange-Bertalot	HAMB	x			
<i>Hantzschia amphioxys</i>	(Ehrenberg) Grunow	HAMD	x			
<i>Hippodonta capitata</i>	(Ehrenberg) Lange-Bertalot, Metzeltin & Witkowski	HCAP				
<i>Hippodonta neglecta</i>	Lange-Bertalot, Metzeltin & Witkowski	HNEG	x			
<i>Humidophila contenta</i>	(Grunow) Lowe, Kociolek, Johansen, Van de Vijver, Lange- Bertalot & Kopalová	HUCO				
<i>Iconella helvetica</i>	(Brun) Ruck & Nakov	IHEL	x			
<i>Iconella linearis</i>	(W. Smith) Ruck & Nakov	ILIN	x	x	x	x
<i>Iconella tenera</i>	(W. Gregory) Ruck & Nakov	ITEN	x			
<i>Karayevia clevei</i>	(Grunow) Bukhtiyarova	KCLE				
<i>Karayevia laterostrata</i>	(Hustedt) Bukhtiyarova	KALA	x	x	x	x
<i>Kolbesia ploenensis</i>	(Hustedt) Round & L. Bukhtiyarova ex Round	KPLN				
<i>Lemnicola hungarica</i>	(Grunow) Round & Basson	LHUN				
<i>Lindavia balatonis</i>	(Pantocsek) Nakov, Guillory, Julius, Theriot & Alverson	LBAL	x			
<i>Lindavia radiosa</i>	(Grunow) De Toni & Forti	LRAD	x			
<i>Lindavia tenuistriata</i>	(Hustedt) Nakov, Guillory, Julius, Theriot & Alverson	LDTE	x			
<i>Luticola acidoclinata</i>	Lange-Bertalot	LACD	x	x	x	x
<i>Luticola goeppertiana</i>	(Bleisch) D.G. Mann	LGOP				
<i>Luticola hlubikovae</i>	Levkov, Metzeltin & Pavlov	LHLU	x			
<i>Luticola mutica</i>	(Kützing) D.G. Mann	LMUT				
<i>Luticola saprophila</i>	Levkov, Metzeltin & Pavlov	LSAP				
<i>Luticola sp.</i>		LUTI				
<i>Luticola ventricifusa</i>	Lange-Bertalot	LVCF				
<i>Luticola ventricosa</i>	(Kützing) D.G. Mann	LVEN	x			
<i>Mastogloia smithii</i>	Thwaites	MSMI	x	x	x	x
<i>Mayamaea agrestis</i>	(Hustedt) Lange-Bertalot	MAGR	x	x	x	x
<i>Mayamaea atomus</i>	(Kützing) Lange-Bertalot	MAAT				
<i>Mayamaea perititis</i>	(Hustedt) K. Bruder & Medlin	MPMI				
<i>Melosira varians</i>	C. Agardh	MVAR				
<i>Meridion circulare</i>	(Gréville) C. Agardh	MCIR				

Species name	Authors	Code	SP	HS	GR	AR
<i>Meridion constrictum</i>	Ralfs	MCCO	x	x	x	x
<i>Navicula amphiceropsis</i>	Lange-Bertalot & Rumrich	NAAM				
<i>Navicula antonii</i>	Lange-Bertalot	NANT				
<i>Navicula associata</i>	Lange-Bertalot	NXAS	x			
<i>Navicula capitatoradiata</i>	Germain	NCPR				
<i>Navicula catalanogermanica</i>	Lange-Bertalot & Hofmann	NCAT				
<i>Navicula cataracta-rheni</i>	Lange-Bertalot	NCTT	x			
<i>Navicula caterva</i>	Hohn & Hellerman	NCTV	x	x	x	x
<i>Navicula cryptocephala</i>	Kützing	NCRY				
<i>Navicula cryptotenella</i>	Lange-Bertalot	NCTE				
<i>Navicula cryptotenelloides</i>	Lange-Bertalot	NCTO				
<i>Navicula erifuga</i>	Lange-Bertalot	NERI	x			
<i>Navicula escambia</i>	(Patrick) Metzeltin & Lange-Bertalot	NESC	x	x	x	x
<i>Navicula flandriae</i>	Van de Vijver & Mertens	NFDR				
<i>Navicula germainii</i>	Wallace	NGER	x			
<i>Navicula gottlandica</i>	Grunow	NGOT	x	x	x	x
<i>Navicula gregaria</i>	Donkin	NGRE				
<i>Navicula lanceolata</i>	(C. Agardh) Ehrenberg	NLAN				
<i>Navicula menisculus</i>	Schumann	NMEN				
<i>Navicula moskalii</i>	Metzeltin, Witkowski & Lange-Bertalot	NMOS	x	x	x	x
<i>Navicula oligotraphenta</i>	Lange-Bertalot & G. Hofmann	NOLI		x	x	x
<i>Navicula radiosa</i>	Kützing	NRAD				
<i>Navicula recens</i>	(Lange-Bertalot) Lange-Bertalot	NRCS				
<i>Navicula reichardtiana</i>	Lange-Bertalot	NRCH				
<i>Navicula reinhardtii</i>	(Grunow) Grunow	NREI				
<i>Navicula rhynchotella</i>	Lange-Bertalot	NRHT		x	x	x
<i>Navicula rostellata</i>	Kützing	NROS				
<i>Navicula salinarum</i> var. <i>rostrata</i>	(Hustedt) Lange-Bertalot	NSAR	x	x	x	x
<i>Navicula salinarum</i> var. <i>salinarum</i>	Grunow	NSAL	x			
<i>Navicula simulata</i>	Manguin	NSIA	x			
<i>Navicula slesvicensis</i>	Grunow	NSLE				
<i>Navicula</i> sp.		NAVI	x	x	x	x
<i>Navicula tenelloides</i>	Hustedt	NTEN	x			
<i>Navicula tripunctata</i>	(O.F.Müller) Bory	NTPT				
<i>Navicula trivialis</i>	Lange-Bertalot	NTRV				
<i>Navicula upsaliensis</i>	(Grunow) Peragallo	NUSA	x			
<i>Navicula veneta</i>	Kützing	NVEN				
<i>Navicula viridula</i>	(Kützing) Ehrenberg	NVIR	x			
<i>Navicula viridulacalcis</i>	Lange-Bertalot	NVCC	x			
<i>Navigeia decussis</i>	(Østrup) Bukhtiyarova	NGDU				
<i>Neidiomorpha binodis</i>	(Ehrenberg) Cantonati, Lange-Bertalot & Angeli	NDMB	x	x	x	x
<i>Neidiomorpha binodeformis</i>	(Krammer) Cantonati, Lange-Bertalot & Angeli	NDBF	x	x	x	x
<i>Neidium ampliutum</i>	(Ehrenberg) Krammer	NEAM	x	x	x	x
<i>Neidium cuneatiforme</i>	Levkov	NECF	x	x	x	x
<i>Neidium dubium</i>	(Ehrenberg) Cleve	NEDU	x			
<i>Nitzschia acicularis</i>	(Kützing) W.M. Smith	NACI				
<i>Nitzschia acidoclinata</i>	Lange-Bertalot	NACD	x	x	x	x
<i>Nitzschia acula</i>	Hantzsch ex Cleve & Grunow	NACU				
<i>Nitzschia adamata</i>	Hustedt	NZAD				
<i>Nitzschia amphibia</i>	Grunow	NAMP				
<i>Nitzschia archibaldii</i>	Lange-Bertalot	NIAR				
<i>Nitzschia bulnheimiana</i>	(Rabenhorst) H.L. Smith	NIBU	x			
<i>Nitzschia capitellata</i>	Hustedt	NCPL				
<i>Nitzschia clausii</i>	Hantzsch	NCLA	x			
<i>Nitzschia communis</i>	Rabenhorst	NCOM	x			
<i>Nitzschia costei</i>	Tudesque, Rimet & Ector	NYCO				
<i>Nitzschia dissipata</i>	(Kützing) Grunow	NDIS				
<i>Nitzschia diversa</i>	Hustedt	NDIV	x	x	x	x
<i>Nitzschia dubia</i>	W. Smith	NDUB				
<i>Nitzschia filiformis</i> var. <i>conferta</i>	(Richter) Lange-Bertalot	NFIC				
<i>Nitzschia fonticola</i>	Grunow	NFON				
<i>Nitzschia frustulum</i>	(Kützing) Grunow	NIFR				
<i>Nitzschia gracilis</i>	Hantzsch	NIGR				

Species name	Authors	Code	SP	HS	GR	AR
<i>Nitzschia graciliformis</i>	Lange-Bertalot & Simonsen	NIGF	x			
<i>Nitzschia heufleriana</i>	Grunow	NHEU				
<i>Nitzschia bacilliformis</i>	Hustedt	NIBF	x			
<i>Nitzschia inconspicua</i>	Grunow	NINC				
<i>Nitzschia intermedia</i>	Hantzsch	NINT				
<i>Nitzschia lacuum</i>	Lange-Bertalot	NILA	x			
<i>Nitzschia linearis</i>	(C. Agardh) W. Smith	NLIN				
<i>Nitzschia media</i>	Hantzsch	NIME				
<i>Nitzschia microcephala</i>	Grunow	NMIC				
<i>Nitzschia palea</i> var. <i>debilis</i>	(Kützing) Grunow	NPAD				
<i>Nitzschia palea</i> var. <i>palea</i>	(Kützing) W. Smith	NPAL				
<i>Nitzschia palea</i> var. <i>tenuirostris</i>	(Grunow) Lange-Bertalot	NPAT				
<i>Nitzschia paleacea</i>	(Grunow) Grunow	NPAE				
<i>Nitzschia pura</i>	Hustedt	NIPR	x	x	x	x
<i>Nitzschia pusilla</i>	(Kützing) Grunow	NIPU				
<i>Nitzschia radicula</i>	Hustedt	NZRA	x			
<i>Nitzschia recta</i>	Hantzsch	NREC				
<i>Nitzschia reskoi</i>	Acs, Duleba, C.E. Wetzel & Ector	NREK	x	x	x	x
<i>Nitzschia sigmoidea</i>	(Nitzsch) W. Smith	NSIO				
<i>Nitzschia sociabilis</i>	Hustedt	NSOC				
<i>Nitzschia soratensis</i>	E. Morales & Vis	NSTS				
<i>Nitzschia</i> sp.		NITZ	x			
<i>Nitzschia haloserbica</i>	Vidakovic, Ector, C.E. Wetzel & Krismanic	NIHAS	x	x	x	x
<i>Nitzschia subacicularis</i>	Hustedt	NSUA	x	x	x	x
<i>Nitzschia subtilis</i>	Grunow	NISU				
<i>Nitzschia supralitorea</i>	Lange-Bertalot	NZSU				
<i>Nitzschia solgensis</i>	Cleve-Euler	NSOL	x			
<i>Nitzschia tabellaria</i>	Grunow	NTAB				
<i>Nitzschia tenuis</i>	W. Smith	NITE				
<i>Nitzschia umbonata</i>	(Ehrenberg) Lange-Bertalot	NUMB	x			
<i>Nitzschia vermicularis</i>	(Kützing) Hantzsch	NVER				
<i>Nitzschia vermicularoides</i>	Lange-Bertalot	NVMO	x			
<i>Pantocsekiella costei</i>	(Druart & F. Straub) K.T. Kiss & Acs	PCOS	x	x	x	x
<i>Pantocsekiella ocellata</i>	(Pantocsek) K.T. Kiss & Acs	POCL				
<i>Paraplaconeis minor</i>	(Grunow) Lange-Bertalot	PMNR	x	x	x	x
<i>Pinnularia lundii</i>	Hustedt	PLUN	x			
<i>Pinnularia lundii</i> var. <i>linearis</i>	Krammer	PLLN	x	x	x	x
<i>Pinnularia microstauron</i>	(Ehrenberg) Cleve	PMIC	x	x	x	x
<i>Pinnularia perspicua</i>	Krammer	PPPC	x			
<i>Pinnularia subrupestris</i>	Krammer	PSRU	x	x	x	x
<i>Pinnularia viridis</i>	(Nitzsch) Ehrenberg	PVIR	x			
<i>Placoneis clementioides</i>	(Hustedt) Cox	PCLD	x			
<i>Placoneis gastrum</i>	(Ehrenberg) Mereschowsky	PGAS	x	x	x	x
<i>Placoneis paraelginensis</i>	Lange-Bertalot	PPAE	x	x	x	x
<i>Placoneis symmetrica</i>	(Hustedt) Lange-Bertalot	PSYM	x	x	x	x
<i>Placoneis undulata</i>	(Østrup) Lange-Bertalot	PUND	x	x	x	x
<i>Planothidium cavilanceolatum</i>	C.E. Wetzel, M.G. Kelly & Van de Vijver	PTCV	x			
<i>Planothidium frequentissimum</i>	(Lange-Bertalot) Lange-Bertalot	PLFR	x			
<i>Planothidium incuriatum</i>	C.E. Wetzel, Van de Vijver & Ector	PICU	x	x	x	x
<i>Planothidium lanceolatum</i>	(Brébisson ex Kützing) Lange-Bertalot	PTLA				
<i>Planothidium</i> sp.		PLTD				
<i>Planothidium robustum</i>	(Hustedt) Lange-Bertalot	PLRO	x	x	x	x
<i>Planothidium rostratum</i>	(Østrup) Lange-Bertalot	PRST				
<i>Planothidium rostratoholarcticum</i>	Lange-Bertalot & Båk	PROH				
<i>Planothidium straubianum</i>	C.E. Wetzel, Van de Vijver & Ector	PHST				
<i>Platessa conspicua</i>	(A. Mayer) Lange-Bertalot	PTCO				
<i>Praestephanos</i> cf. <i>triporus</i>	(Genkal & G.V. Kuzmin) Tuji & J.-S. Ki	PTPU	x	x	x	x
<i>Prestauroneis protracta</i>	(Grunow) Kulikovskiy & Glushchenko	PPTA				
<i>Prestauroneis protractoides</i>	(Hustedt) Q. Liu & Kociolek	PPRD				

Species name	Authors	Code	SP	HS	GR	AR
<i>Pseudofallacia tenera</i>	(Hustedt) Liu, Kociolek & Wang	PFTN	x	x	x	x
<i>Pseudostaurosira brevistriata</i>	(Grunow) Williams & Round	PSBR				
<i>Pseudostaurosira parasitica</i>	(W.Smith) Morales	PPRS				
<i>Pseudostaurosira subconstricta</i>	(Grunow) Kulikovskiy & Genkal	PUSB	x			
<i>Punctastriata lancettula</i>	(Schumann) Hamilton & Siver	PULA	x	x	x	x
<i>Reimeria sinuata</i>	Gregory (Kociolek & Stroemer)	RSIN				
<i>Reimeria uniseriata</i>	Sala Guerrero & Ferrario	RUNI				
<i>Rhoicosphenia abbreviata</i>	(C.Agardh) Lange-Bertalot	RABB				
<i>Rhoicosphenia lacustris</i>	Levkov	RLAC				
<i>Sellaphora atomoides</i>	(Grunow) Wetzel & Van de Vijver	SEAT				
<i>Sellaphora bacillum</i>	(Ehrenberg) D.G. Mann	SEBA	x			
<i>Sellaphora cosmopolitana</i>	(Lange-Bertalot) C.E. Wetzel & Ector	SCPO	x	x	x	x
<i>Sellaphora nigri</i>	(De Notaris) C.E. Wetzel & L. Ector	SNIG				
<i>Sellaphora pupula</i>	(Kützing) Mereschkowksy	SPUP				
<i>Sellaphora saugerresii</i>	(Desm.) C.E. Wetzel & D.G. Mann	SSGE				
<i>Sellaphora seminulum</i>	(Grunow) D.G. Mann	SSEM	x			
<i>Sellaphora sp.</i>		SELL		x	x	x
<i>Sellaphora stroemii</i>	(Hustedt) D.G. Mann	SSTM	x	x	x	x
<i>Sellaphora utermoehlii</i>	(Hustedt) C.E. Wetzel & D.G. Mann	SEUT				
<i>Sellaphora vitabunda</i>	(Hustedt) D.G. Mann	SVIT	x	x	x	x
<i>Simonsenia delognei</i>	(Grunow) Lange-Bertalot	SIDE	x			
<i>Stauroneis gracilior</i>	Reichardt	SGRL	x	x	x	x
<i>Stauroneis kriegei</i>	Patrick	STKR	x	x	x	x
<i>Stauroneis phoenicenteron</i>	(Nitzsch) Ehrenberg	SPHO	x	x	x	x
<i>Stauroneis separanda</i>	Lange-Bertalot & Werum	STSE	x	x	x	x
<i>Stauroneis smithii</i>	Grunow	SSMI	x			
<i>Staurosira construens</i>	Ehrenberg	SCON	x	x	x	x
<i>Staurosira lapponica</i>	(Grunow) Lange-Bertalot	SLPP	x	x	x	x
<i>Staurosira venter</i>	(Grunow) Lange-Bertalot	SSVE				
<i>Staurosirella ovata</i>	Morales	STOV	x			
<i>Staurosirella pinnata</i>	(Ehrenberg) D.M. Williams & Round	SPIN	x			
<i>Stephanodiscus hantzschii</i>	Grunow in Cleve & Grunow	SHAN				
<i>Stephanodiscus hantzschii f. tenuis</i>	(Hustedt) Håkansson & Stoermer	SHTe				
<i>Stephanodiscus medius</i>	Håkansson	SMED				
<i>Stephanodiscus minutulus</i>	(Kützing) Cleve & Moller	STMI				
<i>Stephanodiscus neoastrea</i>	Håkansson & B. Hicke	SNAS				
<i>Stephanodiscus parvus</i>	Stoermer & Håkansson	SPAV				
<i>Surirella angusta</i>	Kützing	SANG				
<i>Surirella brebissonii</i>	Krammer & Lange-Bertalot	SBRE				
<i>Surirella brebissonii var. kuetzingii</i>	Krammer & Lange-Bertalot	SBKU				
<i>Surirella lacrimula</i>	English	SLAC				
<i>Surirella librile</i>	(Ehrenberg) Ehrenberg	SULI	x			
<i>Surirella minuta</i>	Brébisson ex Kützing	SUMI	x			
<i>Surirella ovalis</i>	Brébisson	SOVI	x			
<i>Surirella suecica</i>	Grunow in Van Heurck	SSUE	x			
<i>Surirella terricola</i>	Lange-Bertalot & Alles	STER	x	x	x	x
<i>Surirella undulata</i>	(Ehrenberg) Ehrenberg	SUUN	x			
<i>Tabularia fasciculata</i>	(C. Agardh) D.M. Williams & Round	TFAS				
<i>Tabellaria flocculosa</i>	(Roth) Kützing	TFLO	x			
<i>Thalassiosira lacustris</i>	(Grunow) Hasle in Hasle & Fryxell	THLA	x	x	x	x
<i>Tryblionella angustata</i>	W. Smith	TANG	x			
<i>Tryblionella angustatula</i>	(Lange-Bertalot) Cantonati & Lange-Bertalot	TATU	x			
<i>Tryblionella apiculata</i>	Gregory	TAPI				
<i>Tryblionella brunoii</i>	(Lange-Bertalot) Cantonati & Lange-Bertalot	TBNO	x			
<i>Tryblionella calida</i>	(Grunow) D.G. Mann	TCAL	x			
<i>Tryblionella debilis</i>	Arnott ex O'Meara	TDEB	x			
<i>Tryblionella gracilis</i>	W.Smith	TGRL	x	x	x	x
<i>Tryblionella hungarica</i>	(Grunow) Frenguelli	THUN				
<i>Tryblionella levidensis</i>	W.Smith	TLEV				

Species name	Authors	Code	SP	HS	GR	AR
<i>Tryblionella salinarum</i>	(Grunow) Pantocsek	TSAL				
<i>Ulnaria acus</i>	Kützing) Aboal	UACU				
<i>Ulnaria biceps</i>	(Kützing) Compère	UBIC				
<i>Ulnaria capitata</i>	(Ehrenberg) Compère	UCAP	x	x	x	x
<i>Ulnaria danica</i>	(Kützing) Compère & Bukhtiyarova	UDAN	x	x	x	x
<i>Ulnaria grunowii</i>	(Lange-Bertalot & Ulrich) Cantonati & Lange-Bertalot	UGRU				
<i>Ulnaria ulna</i>	(Nitzsch) Compère	UULN				

Supplementary table 3. Environmental variables by sampling period

Environmental variable	September-October, 2018		May-June, 2019	
	Mean	SD	Mean	SD
Ripveg index (%)	39.95	26.53	39.91	25.95
Macrophyte cover (%)	30.70	34.72	27.37	33.03
Temperature (°C)*	16.24	3.29	18.04	4.19
Conductivity ($\mu\text{S cm}^{-1}$)	610.08	222.38	616.02	220.59
DO (mg L^{-1})*	8.73	1.90	8.06	1.78
O ₂ saturation (%)	89.64	19.07	86.03	18.15
pH	8.28	0.34	8.20	0.29
Turbidity (FNU)*	17.84	32.12	22.00	22.38
Flow velocity (m s^{-1})	0.17	0.17	0.20	0.18
NH ₄ ⁺ -N ($\mu\text{g L}^{-1}$)*	63.30	196.16	212.29	629.31
NO ₂ ⁻ -N ($\mu\text{g L}^{-1}$)*	17.16	21.33	58.24	105.56
NO ₃ ⁻ -N ($\mu\text{g L}^{-1}$)*	8679.83	8679.97	3393.52	2755.70
SRP ($\mu\text{g L}^{-1}$)	61.84	148.13	62.27	153.87
SRSi (mg L^{-1})*	11.60	9.37	7.91	2.59
SO ₄ ²⁻ (mg L^{-1})	59.25	49.43	66.91	63.24
TP ($\mu\text{g L}^{-1}$)	105.65	126.53	146.11	194.34
COD ($\text{mg O}_2 \text{ L}^{-1}$)	3.42	2.72	3.14	1.69
Cl ⁻ (mg L^{-1})	32.86	21.56	31.84	20.45
HCO ₃ ⁻ (mg L^{-1})	254.94	103.31	220.52	87.85

Supplementary table 4. Taxon richness and Shannon diversity by site

Site	Diatom metabarcoding		Diatom microscopy		Phytobenthos microscopy	
	Taxon richness	Shannon diversity	Taxon richness	Shannon diversity	Taxon richness	Shannon diversity
M801	59	2.106	23	1.605	34	2.325
M802	62	2.595	22	2.013	24	2.456
M804	52	1.743	19	2.102	20	1.762
M808	72	2.397	43	2.871		
M809	64	1.196	40	2.901	24	1.102
M810	68	3.542			47	1.483
M811	61	1.836	39	2.642	13	0.759
M812	56	3.144	32	2.543	12	0.980
M813	85	3.222	58	3.635	29	2.133
M815	63	2.536	22	2.225	16	2.102
M816	65	1.935	39	2.818	35	1.180
M817	67	3.268	32	2.590	24	2.380
M818	52	2.419	34	2.805	21	1.795
M820	59	2.394	29	2.395	26	2.094
M821	68	2.707	43	3.213	39	2.797
M823	68	1.847	47	3.174	31	1.156
M824	54	2.379				
M825	76	3.039	30	2.104		
M826	60	2.394	35	2.688		
M827	57	2.517	24	1.537		
M828	50	2.819	32	2.369		

M830	65	2.991	38	2.846		
M831	33	1.032	27	1.317		
M832	55	2.075	53	3.230		
M834			60	3.445		
M835	72	3.365	45	2.814		
M836	46	1.237	47	3.351		
M837	69	2.321	23	1.231		
M838	58	2.652	31	2.225		
M839	49	3.033	32	2.620		
M841	47	1.960	28	1.397		
M842	51	2.493				
M844	47	2.484	24	1.815		
M845	59	2.377	38	2.389		
M846			32	1.736		
M847	36	2.126				
M848	72	3.052	36	2.184		
M901	84	3.020	43	2.973	39	2.735
M902	64	2.868	33	2.777	21	2.358
M904	39	1.940	15	1.359	15	1.499
M908	53	2.484	24	2.132		
M909	79	3.355	40	2.690		
M910	58	2.954	25	2.076		
M911	68	2.832	42	2.946		
M912	63	2.969	33	2.217		
M913	64	3.454				
M915	63	2.566	44	2.855	10	1.945
M916	78	3.116	42	2.605	25	1.778
M917	63	2.854	34	1.824	23	1.800
M918	52	2.272	24	1.751		
M920	66	2.364	24	2.151		
M921			33	1.997		
M923	85	3.248	42	2.936	26	2.376
M924	54	2.406	23	1.729	21	0.917
M925	76	3.051	45	2.908	24	2.325
M926	89	3.006	45	2.971	40	1.170
M927	50	2.752	33	2.384	16	2.351
M928	62	2.712	28	2.253	16	2.338
M930	74	2.846	49	3.051	29	2.221
M931	42	1.870	17	1.649	19	1.837
M932	55	2.552	26	1.881	16	2.098
M934	61	2.638	40	2.438	16	1.897
M935	59	2.619	25	2.102	13	0.963
M936	45	2.272	19	1.897	20	1.328
M937	65	2.693	36	2.495	23	1.443
M938	40	1.941	27	1.729	14	0.999
M939	58	2.439	45	2.637	31	2.301
M941	82	3.187	33	2.691	19	1.949
M942	59	2.778	23	2.283	14	1.604
M944	35	1.655	25	2.464	19	1.880
M945	52	2.330	31	1.875	27	2.045
M946	58	2.562	47	2.834	32	2.401
M947	33	1.781	12	0.970	10	1.579
M948	59	2.208	22	1.214	13	1.756
R801	70	2.870	35	2.743	25	2.177
R803			17	2.425	1	0.000
R804	37	1.538	21	1.051	4	0.930
R805	29	1.893	10	0.811	8	0.577
R806	61	2.827	33	2.290	8	1.721
R807	74	3.098	27	1.789	15	0.765
R808	60	2.149	25	2.214	12	1.413
R810	50	2.792	28	2.797	17	2.313
R811	50	2.767	25	2.398	14	2.344

R816			43	2.338		
R817	69	2.693	43	3.090		
R818	67	2.875	40	3.099		
R820	63	1.629	37	2.590		
R821	57	3.019	24	1.998		
R822	49	2.996	16	1.016	10	1.060
R823			31	1.803	12	1.934
R824			3	1.040	14	2.020
R825	49	2.713	25	1.952	20	2.324
R826	73	2.529	41	2.834	39	2.564
R827					4	1.362
R828	51	2.772	47	2.974	15	2.211
R830	88	2.929	51	3.302	10	1.858
R833	59	2.727	36	2.715	26	1.955
R834					11	1.962
R835	58	2.137	23	2.262	23	2.317
R836	75	2.770	46	3.101	34	2.687
R838	57	2.488	52	2.809	33	2.347
R841	86	2.987	72	3.716	24	1.970
R842					9	1.960
R843	42	1.879	27	2.702	23	1.032
R844	91	2.922	33	1.785	26	2.450
R845	56	2.424	61	3.349	18	1.850
R846	78	2.461	29	1.915	28	2.467
R848	79	2.272	19	0.810	20	1.443
R850	47	2.396	40	2.732	22	2.028
R901	64	2.956	23	2.150		
R903	42	2.075	19	1.633		
R904	32	1.792	12	0.940		
R905	31	1.515	29	2.199		
R906	28	2.225	19	1.741		
R907	78	3.195	35	2.540		
R908	44	2.400	33	2.666		
R910	56	2.626	20	1.771		
R911	49	1.207	22	1.305		
R916	55	1.787	24	0.799	7	0.666
R917	61	2.434	27	2.020	25	2.057
R918	56	1.833	39	2.865	30	1.323
R920	46	2.285	25	2.231	19	1.563
R921	68	2.241	34	1.784	13	2.130
R922	50	2.116	20	1.536	13	1.947
R924					11	0.459
R925	45	2.291	19	1.346	28	2.397
R926	81	2.995			36	2.520
R927			62	3.374		
R928	69	2.961	41	2.874	19	2.657
R930	88	3.012	41	2.537	24	1.894
R933	50	2.860	58	3.304	32	1.951
R934					14	1.650
R935	63	2.377	24	1.821	18	1.906
R936	68	2.545	40	2.996	27	1.780
R938	64	2.691	53	3.005	40	2.726
R941	92	2.882			4	0.311
R942	60	2.976	51	3.143	13	2.210
R943	62	2.985	49	3.166	25	1.254
R944	84	2.914	37	2.599	27	2.387
R945	59	2.872	55	2.877	30	1.746
R946	60	2.746			7	1.295
R948	71	2.880	27	1.872	26	1.712
R950	70	2.548	42	3.025	23	1.775

Supplementary table 5. Taxon-level correlations between methods

<i>Taxon</i>	Samples positive in both methods (n)	Spearman's ρ	p-value	Significance ($\alpha = 0.05$)
<i>Cocconeis placentula</i>	102	0.885	1.04E-41	Significant
<i>Amphora pediculus</i>	104	0.854	7.59E-36	Significant
<i>Navicula tripunctata</i>	84	0.767	6.63E-25	Significant
<i>Cocconeis pediculus</i>	30	0.727	2.47E-21	Significant
<i>Eunotia bilunaris</i>	6	0.717	1.69E-20	Significant
<i>Rhoicosphenia abbreviata</i>	57	0.702	2.26E-19	Significant
<i>Planothidium lanceolatum</i>	62	0.687	2.52E-18	Significant
<i>Achnanthyidium minutissimum</i>	96	0.662	9.76E-17	Significant
<i>Planothidium frequentissimum</i>	75	0.660	1.35E-16	Significant
<i>Navicula gregaria</i>	74	0.642	1.59E-15	Significant
<i>Reimeria sinuata</i>	27	0.642	1.61E-15	Significant
<i>Navicula rostellata</i>	17	0.641	1.73E-15	Significant
<i>Melosira varians</i>	49	0.621	2.46E-14	Significant
<i>Nitzschia inconspicua</i>	47	0.613	6.44E-14	Significant
<i>Gomphonema micropus</i>	46	0.604	1.82E-13	Significant
<i>Cyclotella meneghiniana</i>	34	0.585	1.51E-12	Significant
<i>Karayevia ploenensis</i> var. <i>gessneri</i>	15	0.577	3.49E-12	Significant
<i>Nitzschia palea</i>	83	0.563	1.49E-11	Significant
<i>Diatoma moniliformis</i>	5	0.559	2.19E-11	Significant
<i>Cyclostephanos invisitatus</i>	5	0.545	8.64E-11	Significant
<i>Nitzschia amphibia</i>	31	0.534	2.28E-10	Significant
<i>Navicula capitatoradiata</i>	18	0.534	2.46E-10	Significant
<i>Navicula lanceolata</i>	40	0.524	5.64E-10	Significant
<i>Encyonema ventricosum</i>	12	0.522	6.76E-10	Significant
<i>Ulnaria ulna</i>	23	0.512	1.63E-09	Significant
<i>Navicula cryptotenella</i>	73	0.504	3.19E-09	Significant
<i>Lemnicola hungarica</i>	5	0.502	3.87E-09	Significant
<i>Navicula cryptocephala</i>	20	0.502	3.97E-09	Significant
<i>Gomphonema tergestinum</i>	4	0.496	6.49E-09	Significant
<i>Navicula veneta</i>	70	0.494	7.16E-09	Significant
<i>Hippodonta capitata</i>	32	0.493	8.14E-09	Significant
<i>Achnanthyidium pyrenaicum</i>	13	0.492	8.40E-09	Significant
<i>Aulacoseira granulata</i>	5	0.491	9.43E-09	Significant
<i>Navicula trivialis</i>	32	0.480	2.28E-08	Significant
<i>Discostella pseudostelligera</i>	3	0.475	3.21E-08	Significant
<i>Amphora ovalis</i>	17	0.475	3.30E-08	Significant
<i>Caloneis amphisbaena</i>	5	0.458	1.09E-07	Significant
<i>Cymbella neocistula</i>	4	0.455	1.45E-07	Significant
<i>Nitzschia linearis</i>	77	0.451	1.91E-07	Significant
<i>Cymbella tumida</i>	9	0.436	5.02E-07	Significant
<i>Sellaphora pupula</i>	8	0.436	5.17E-07	Significant
<i>Meridion circulare</i>	17	0.435	5.54E-07	Significant
<i>Nitzschia capitellata</i>	28	0.430	7.57E-07	Significant
<i>Amphora copulata</i>	28	0.428	8.53E-07	Significant
<i>Tryblionella apiculata</i>	38	0.415	2.06E-06	Significant

<i>Fragilaria gracilis</i>	8	0.407	3.33E-06	Significant
<i>Diatoma vulgare</i>	6	0.406	3.54E-06	Significant
<i>Navicula antonii</i>	55	0.402	4.57E-06	Significant
<i>Gyrosigma acuminatum</i>	25	0.376	1.98E-05	Significant
<i>Conticribra weissflogii</i>	5	0.370	2.76E-05	Significant
<i>Craticula subminuscula</i>	17	0.352	6.96E-05	Significant
<i>Navicula slesvicensis</i>	7	0.324	2.77E-04	Significant
<i>Halamphora veneta</i>	4	0.309	5.38E-04	Significant
<i>Gomphonema acuminatum</i>	10	0.309	5.39E-04	Significant
<i>Parlibellus protracta</i>	4	0.304	6.71E-04	Significant
<i>Gomphonema parvulum</i>	24	0.283	1.60E-03	Significant
<i>Nitzschia fonticola</i>	12	0.245	6.59E-03	Significant
<i>Frustulia vulgaris</i>	10	0.242	7.35E-03	Significant
<i>Nitzschia gracilis</i>	4	0.237	8.52E-03	Significant
<i>Tryblionella hungarica</i>	7	0.230	1.09E-02	Significant
<i>Caloneis fontinalis</i>	3	0.214	1.78E-02	Significant
<i>Gomphonema minutum</i>	15	0.208	2.14E-02	Significant
<i>Halamphora montana</i>	9	0.202	2.60E-02	Significant
<i>Tryblionella debilis</i>	3	0.171	6.00E-02	Not significant
<i>Nitzschia paleacea</i>	8	0.153	9.17E-02	Not significant
<i>Nitzschia draveillensis</i>	4	0.151	9.78E-02	Not significant
<i>Mayamaea permitis</i>	13	0.144	1.13E-01	Not significant
<i>Achnanthyrium eutrophilum</i>	13	0.122	1.82E-01	Not significant
<i>Nitzschia dissipata</i>	7	0.106	2.45E-01	Not significant
<i>Fallacia pygmaea</i>	3	0.087	3.39E-01	Not significant
<i>Nitzschia sigmoidea</i>	7	0.036	6.91E-01	Not significant
<i>Nitzschia supralitoria</i>	5	0.013	8.88E-01	Not significant
<i>Fistulifera saprophila</i>	4	-0.088	3.35E-01	Not significant
<i>Sellaphora seminulum</i>	7	-0.177	5.14E-02	Not significant

Supplementary table 6. Co-inertia analysis loadings

Environmental variable	Phytobenthos microscopy		Diatoms microscopy		Diatoms metabarcoding	
	RS1	RS2	RS1	RS2	RS1	RS2
Ripveg index	0.174	-0.017	0.170	0.119	0.196	0.067
Macrophyte cover	0.015	0.325	-0.337	0.257	-0.269	0.273
Temp	0.149	-0.255	0.172	0.068	0.150	0.154
Cond	-0.346	0.144	-0.421	-0.127	-0.459	-0.088
O sat	-0.307	-0.064	0.169	-0.314	0.123	-0.444
pH	-0.367	-0.061	-0.045	-0.348	-0.069	-0.399
Cl	-0.308	-0.096	-0.227	-0.197	-0.269	-0.080
Cropland	-0.164	0.102	-0.271	0.028	-0.321	0.018
Urban	-0.164	0.024	0.030	-0.288	0.003	-0.280
Shaded area	0.268	0.392	-0.107	0.478	-0.018	0.332
Turb	0.196	-0.162	0.185	0.198	0.162	0.069
v flow	-0.143	-0.100	0.163	-0.410	0.113	-0.393
NH4	-0.065	-0.363	-0.160	-0.011	-0.146	0.133
NO3	-0.204	0.372	-0.253	-0.077	-0.217	-0.172
PO4	-0.086	-0.109	-0.191	-0.062	-0.202	0.090
SiO2	0.097	0.421	-0.238	0.195	-0.189	0.157
SO4	-0.369	0.100	-0.365	-0.174	-0.380	-0.145

TP	0.025	-0.097	-0.122	0.105	-0.075	0.175
COD	-0.073	-0.318	-0.125	-0.065	-0.156	0.129
Forest	0.339	-0.102	0.277	0.168	0.320	0.177

Publication list

Abubaker, B., Stenger-Kovács, C. & Tapolczai, K. Using complementary approaches to study diatom diversity and its environmental drivers. *Ecological Indicators*. Volume 187, 114970, ISSN 1470-160X. (2026). <https://doi.org/10.1016/j.ecolind.2026.114970>.

Stenger-Kovács, C., Korponai, J., **Abubaker, B.** *et al.* Role of rare species in benthic diatom communities: patterns, processes and networks. *Biodivers Conserv* **34**, 2095–2118 (2025). <https://doi.org/10.1007/s10531-025-03063-4>

Lengyel, E., **Abubaker, B.**, Viktória, B., Udovič, M. G., Grmasha, R. A., Hubai, K., ... & Stenger-Kovács, C. (2025). Ecological drivers of rarity and commonness in freshwater benthic diatoms. *Global Ecology and Conservation*, e03991. <https://doi.org/10.1016/j.gecco.2025.e03991>

Abubaker, B., Stenger-Kovács, C., Rimet, F., & Tapolczai, K. (under second review): Environmental DNA metabarcoding of diatoms: A next-generation tool for aquatic biodiversity and bioassessment. *Metabarcoding and Metagenomics*. (Under review)

Baker, B.A., Tawabini, B., Nazal, M. *et al.* Efficiency of Thermophilic Bacteria in Wastewater Treatment. *Arab J Sci Eng* **46**, 123–128 (2021). <https://doi.org/10.1007/s13369-020-04830-x>

Conferences

Using complementary approaches to study the environmental drivers of diatom diversity
B Abubaker, C Stenger-Kovács, K Tapolczai
(03/08/2025 - 08/08/2025 27th International Diatom Symposium - South Africa)

Environmental drivers of diatom diversity: insights from DNA metabarcoding
B Abubaker, C Stenger-Kovács, K Tapolczaihe description
(17/03/2025 - 21/03/2025 AquaEcOmics Conference - France)

Responses of benthic algae to environmental constraints using different identification approaches
B Abubaker, C Stenger-Kovács, K Tapolczai
(24/09-26/09/2024 3rd International Conference: Lakes & reservoirs: hot spots and topics in limnology - Poland)

Bassam Abubaker CV

Environmental Scientist | Freshwater Ecology | eDNA Metabarcoding | Diatom Ecology

Bassam Abubaker is an environmental scientist specializing in freshwater ecology, environmental DNA (eDNA), DNA metabarcoding, diatom ecology, and aquatic biodiversity assessment. He completed his PhD in Environmental Sciences at the University of Pannonia, Hungary, where he conducted research within the Limnology Research Group under the supervision of Dr. Csilla Stenger-Kovács and Dr. Kálmán Tapolczai.

His doctoral research focused on integrating DNA metabarcoding and conventional microscopic approaches to investigate benthic diatom diversity, rare taxa, invasive species, and their relationships with environmental conditions in freshwater ecosystems. His work involved extensive field sampling across river networks in Hungary and Croatia and combined ecological, molecular, and statistical approaches to evaluate biodiversity patterns at multiple spatial and temporal scales.

During his PhD, Bassam developed strong expertise in environmental DNA sampling, DNA extraction, PCR-based metabarcoding, Illumina sequencing data analysis, bioinformatics, ecological statistics, and multivariate community analysis. His bioinformatics experience includes processing metabarcoding datasets using DADA2, taxonomic assignment of diatom sequences, management of amplicon sequence variants (ASVs), and integration of molecular biodiversity data with environmental datasets. He is also experienced in R-based statistical analysis and scientific data visualization, including diversity analysis, PCA, community–environment relationships, co-inertia analysis, and other multivariate ecological approaches.

His doctoral work has contributed to research on the application of eDNA metabarcoding for freshwater biodiversity assessment and biomonitoring, particularly its ability to complement traditional microscopy and improve the detection of rare and invasive taxa. His research has resulted in scientific publications covering diatom metabarcoding, rare species ecology, biodiversity assessment, and community responses to environmental gradients.

Bassam also has experience in field ecology, freshwater sampling, laboratory molecular methods, scientific writing, manuscript preparation, peer-reviewed publication, and student supervision. He has contributed to the supervision of Master's-level research involving invasive diatoms and combined microscopic and eDNA metabarcoding approaches.

Education

PhD in Environmental Sciences – University of Pannonia, Hungary
Master's Degree in Life Sciences – King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia
Bachelor's Degree in Biotechnology – An-Najah National University, Palestine

Technical Expertise

eDNA sampling and metabarcoding | DNA extraction and PCR | Illumina sequencing workflows | DADA2 | ASV-based analyses | R statistical programming | Multivariate statistics | Biodiversity indices | Community ecology | Freshwater field sampling | Diatom identification | Scientific writing and publication