


Research Article

New DNA barcode reference data of freshwater diatoms (Bacillariophyceae) from Sweden: old acquaintances and new taxa

Maria Kahlert¹, Demetrio Mora^{2,3}, Wolf-Henning Kusber², Nélida Abarca², Jonas Zimmermann²

¹ Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden

² Botanischer Garten und Botanisches Museum, Freie Universität Berlin, Königin-Luise-Str. 6-8, 14195 Berlin, Germany

³ Luxembourg Institute of Science and Technology, 41 rue du Brill, 4422 Belvaux, Luxembourg

Corresponding author: Maria Kahlert (maria.kahlert@slu.se)



Academic editor:

Armando Espinosa Prieto

Received: 28 January 2026

Accepted: 28 April 2026

Published: 26 May 2026

Citation: Kahlert M, Mora D, Kusber W-H, Abarca N, Zimmermann J (2026) New DNA barcode reference data of freshwater diatoms (Bacillariophyceae) from Sweden: old acquaintances and new taxa. Metabarcoding and Metagenomics 10: e186778. <https://doi.org/10.3897/mbmg.10.186778>

Copyright: © Maria Kahlert et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Abstract

Introduction: DNA metabarcoding is currently under intense scrutiny for its future implementation in biodiversity monitoring and ecological assessment, particularly within the EU Water Framework Directive (WFD). However, before metabarcoding can be deployed for WFD reporting, challenges such as the incompleteness of reference databases need to be addressed. In this context, this study presents the results of a national barcoding project (FRESHBAR) focusing on benthic freshwater diatoms of Sweden, a key organism group regarding both ecology and environmental assessment. One major goal of the project was the publication and vouchering of all data, material, and results. **Methods:** A total of 312 diatom cultures were established, with a focus on oligotrophic and acid habitats. The cultures were sequenced for two barcodes (*rbcl* and 18SV4) and identified using light microscopy; selected strains were also studied by scanning electron microscopy. All data, including sampling metadata, barcode sequences, images, and voucher material, were published following the FAIR principles. Some cultures were archived in diatom culture collections. **Results:** Almost all strains were successfully sequenced and identified as 51 taxa across 17 genera, with the highest diversity in the genera *Eunotia* and *Fragilaria*. An attempt was also made to identify all taxa at the species level. Approximately half of the taxa were new to Diat.barcode, a curated diatom reference database. **Conclusion:** FRESHBAR represents the first large-scale effort to generate barcode reference sequences for Nordic benthic freshwater diatoms. As all data are publicly available, the added sequences and morphological information will help identify all taxa properly to the species level, unravel taxon relationships within diatoms, and improve reference databases for environmental monitoring and research.

Key words: 18SV4, DNA library, environmental monitoring, metabarcoding, molecular sequence data, *rbcl*, Water Framework Directive, water quality

Introduction

Freshwater ecosystems constitute only a relatively small amount of the Earth's surface, yet they harbor a disproportionate number of species. Globally, freshwaters are among the most threatened habitats, with biodiversity loss and ecosystem services impaired due to human-induced stressors associated with land use, chemical pollution, and warming (Strayer and Dudgeon 2010).

To gauge these anthropogenic impacts on biodiversity and ecological status, changes in the community composition of aquatic organisms are used in national assessments (Lindegarth et al. 2016). Benthic diatoms are key components of aquatic ecosystems and are essential parts of European assessment programs (Birk et al. 2012), as their taxonomy and autecology are well documented. Moreover, they can be found in almost every aquatic habitat; sampling is simple and nondestructive; they are species-rich in comparison to other organism groups; they respond rapidly and predictably to changes in their environment; and samples can be stored permanently for future use (Johnson et al. 1993; Smol and Stoermer 2010; Masouras et al. 2021).

However, despite the relatively good knowledge of the taxonomy of this organism group, traditional identification requires extensive experience and training and is time-consuming. This challenge of identification is not only costly; it is also a hindrance to the direct comparison of diatom results from different projects and laboratories. For species lists identified by morphological characters via light microscopy, extensive harmonization is needed (Kahlert et al. 2022; Jupke et al. 2023) because of varying identification sources, identification errors (Morales et al. 2001; Kahlert et al. 2009), and ongoing changes in diatom taxonomy (e.g., Mann and Vanormelingen 2013). The recent development of metabarcoding methods has opened the door for more cost-effective and automated identification, which would furthermore enable automated harmonization of datasets from different laboratories for large-scale studies (Pawlowski et al. 2018). Metabarcoding methods can also identify diatoms to a finer level than traditional morphological identification via light microscopy through the use of fine-level taxonomic units such as genetic variants (amplicon sequence variants/exact sequence variants (ASVs/ESVs)) for environmental assessment (Pérez-Burillo et al. 2021; Tapolczai et al. 2021). For diatoms, metabarcoding is already quite well developed and has been proposed as a future tool for biomonitoring as part of the EU Water Framework Directive (WFD) (European Parliament and Council of the European Union 2000; Pawlowski et al. 2018).

Diatom metabarcoding can rely on already established European standard methods for sampling biofilm from hard substrates (CEN 2014, 2025). These biofilm samples, if taken in suitable habitats as indicated by the standard, usually include sufficient living material such that the problems related to DNA degradation or contamination are generally limited (CEN 2025), which is an advantage over water samples often taken for environmental DNA (eDNA) analyses. The diatom metabarcoding community has agreed on two barcodes (*rbcL* and 18SV4) and has developed curated and publicly accessible databases, including the Diat.barcode reference library (Zimmermann et al. 2014; Rimet et al. 2019; Gelis et al. 2025). These databases focus on European species to match the needs of environmental assessment for the European Water Framework Directive (Weigand et al. 2019). The *rbcL* marker has been highlighted as the most promising barcode for the metabarcoding of diatoms because it better separates diatom species and poses fewer practical problems than the 18SV4 marker (Mann et al. 2010), whereas the latter could potentially open the way to integrated assessment of organism groups to provide a more holistic overview, as it can be used for parallel metabarcoding of organism groups other than only diatoms (Kelly et al. 2018). However,

the taxonomic resolution of 18SV4 for diatoms is generally lower, especially for closely related species.

However, a number of challenges remain before metabarcoding can become operational for environmental assessment, among them the standardization of laboratory protocols for DNA-based methods and the improvement of barcode libraries, especially for small organisms such as microalgae, including diatoms (Hering et al. 2018). Regarding diatoms, only approximately 15% of all European diatom species are represented in DNA reference databases (Weigand et al. 2019). Abundant species are covered more frequently, but rare and geographically restricted species are missing from the databases (Weigand et al. 2019). This incomplete coverage remains one of the major bottlenecks for the implementation of metabarcoding in routine assessment. As the curated databases have been developed in Central Europe (Zimmermann et al. 2014; Rimet et al. 2019), there is an urgent need to complete the curated diatom databases with taxa from other geographical regions. As an example, most species observed in Sweden are missing from these databases, especially those thriving in oligotrophic and acidic habitats. This can lead to a misclassification of the ecological status of these waters when metabarcoding data are used to calculate diatom indices for environmental assessment (Baillet et al. 2019).

Another challenge coupled with the use of DNA as a basis for taxon identification is the continuous reshaping of diatom taxonomy—especially for freshwater benthic taxa—over the past decade. The integration of genetic data with traditional morphology (detailed light microscopy (LM) and scanning electron microscopy (SEM)) has revealed widespread cryptic and pseudocryptic diversity within morphospecies, prompting taxonomic revisions and the description of many new species (Abarca et al. 2014; Pinseel et al. 2019; Rimet et al. 2019; Zou et al. 2021). Such cryptic diversity has now been documented in many freshwater benthic genera, including, for example, *Achnantheidium*, *Cocconeis*, *Fistulifera*, *Fragilaria*, *Gomphonella*, *Gomphonema*, *Nitzschia*, and *Pinnularia* (Abarca et al. 2014; Jahn et al. 2019; Kahlert et al. 2019; Pinseel et al. 2019; Jahn et al. 2020; Kahlert et al. 2021b; Pérez-Burillo et al. 2021; Solak et al. 2021; Tapolczai et al. 2021). Diatoms are a taxon-rich organism group, and more research is needed to disentangle the hidden species diversity within diatom species complexes. Here, unialgal cultures are useful for analyzing the variability of morphological characteristics and assessing their stability, a task that is difficult to achieve in environmental samples where several species co-occur (Mohamad et al. 2022; Schimani et al. 2023). Cultures also provide voucher-linked barcode sequences that can be unambiguously assigned to specific taxa, which is essential for reference libraries.

To help populate diatom databases with new barcode sequences from Nordic freshwater species and to enable the inclusion of Nordic taxa in the ongoing taxonomic diatom revisions, the national barcoding project FRESHBAR was conducted during 2019–2023. The aim of the project was to establish as many diatom cultures as possible, generate their barcodes, identify them to the lowest taxonomic level possible, and make all results, data, and material publicly available for further research and use within environmental monitoring. FRESHBAR represents the first large-scale effort to generate barcode reference sequences for Nordic freshwater benthic diatoms. The project and its results are presented here.

Methods

Isolation and culture of diatoms

Most diatoms were isolated from fresh samples taken in coordination with the Swedish national monitoring program in 2019 and 2020. The study focused on oligotrophic and acidic streams but also used fresh samples from other aquatic habitats with differing ecology. A few isolates were generated from enrichment cultures taken in an earlier project and grown on agar plates (Gonçalves et al. 2018). In addition, a few clones had already been isolated in a pilot project in 2014–2016. The full list of sampling sites, including key environmental variables, is shown in Table 1. For each sampling site, more information can be retrieved from the national database of environmental monitoring data (Miljödata-MVM 2025). In total, 273 monoclonal cultures (Suppl. material 3: table S1) were established from single cells using the micropipette method (Kelly et al. 2018). Cultures were grown in WC (Guillard and Lorenzen 1972), PM (Stosch and Fecher 1979), or L16 (Lindström 1991) medium at room temperature. Photographs were taken from each living culture if possible. The cultures were subsequently harvested and sequenced and, in parallel, identified using morphological characters using LM in a first step. The clones were harvested when the biomass was sufficient for all further project-relevant analysis steps. The cultures were grown in triplicate, of which one culture was harvested for DNA analysis, one for morphological analysis, and one was allowed to grow further as a reserve. Some of the reserve cultures from 2019 were divided into several cultures if they grew well and were harvested more than once to ensure sufficient material for all analyses. The cultures from 2020 were harvested only once. In total, 312 harvested strains were sequenced, of which 39 were copies of some of the 273 unique clonal cultures. Isolation and cultivation were conducted according to the methods described in Mora et al. (2019).

DNA extraction, amplification, and sequencing

Cultured material was first centrifuged at 2000 rpm, and the culture medium was discarded by careful pipetting. DNA was isolated from the remaining pellet using the NucleoSpin® Plant II Mini Kit (Macherey–Nagel, Düren, Germany) following the manufacturer's instructions. DNA fragment size and concentrations were evaluated via gel electrophoresis (1.5% agarose gel) and NanoDrop® (PeqLab Biotechnology LLC; Erlangen, Germany), respectively. Amplification was conducted by polymerase chain reaction (PCR) following Zimmermann et al. (2011) for the marker 18SV4 and Abarca et al. (2014) for the *rbcL* marker. PCR products were then visualized in a 1.5% agarose gel and cleaned with MSB Spin PCRapace® (Invitex Molecular GmbH, Berlin, Germany) following the manufacturer's instructions. Concentrations of PCR products were measured using NanoDrop® (PeqLab Biotechnology) and normalized to $> 100 \text{ ng } \mu\text{L}^{-1}$ for sequencing. Sanger sequencing was conducted bidirectionally by Starseq® (GENterprise LLC; Mainz, Germany), with the same primers used for the amplifications. DNA extraction, amplification, and the pre-sequencing steps were carried out in the laboratory of the Botanic Garden and Botanical Museum Berlin (BGBM), Freie Universität Berlin, Germany. All DNA extractions, PCRs, and pre-sequencing steps were performed under sterile conditions and included negative controls to monitor potential contamination. DNA extraction, PCR amplification, and sequencing were performed

Table 1. Sample sites with information on the location, georeference, altitude, catchment, and water chemistry.

Stream	Site name	Site ID	Latitude decimal WGS84	Longitude decimal WGS84	Altitude [m]	water type	Upstream catchment area [km ²]	Landuse in upstream catchment [%]						Water chemistry (annual mean)				
								Forest	Sparse/no vegetation	Wetland	Water	Agriculture	Urban	pH	Absorbance [420/5]	Total-P [µg/l]	Total-N [µg/l]	Conductivity [mS/m]
Bastuån		1880	63.0938	13.8953	474	stream	42	28	70	2	0	0	0	6.4	0.079	4	147	1
Bjurbäcken	Bjurbäck	197	64.6773	20.3898	152	stream	41	76	18	4	1	0	0	5.7	0.365	19	614	4
Björnbackån		5056	60.7626	16.4215	183	stream	35	84	11	4	0	0	0	5.7	0.367	8	420	3
Brännkälsbäcken		48094	64.3228	21.2577	2	stream	24	84	6	0	1	8	0	4.5	0.508	24	640	10
Börumsbäcken		928	58.3475	16.6266	2	stream	29	80	1	0	2	18	0	7.0	0.373	93	1351	17
Enån	ned sågen	1900	59.8865	13.4886	96	stream	41	88	7	0	5	0	0	6.0	0.287	27	476	3
Fiskonbäcken	v.vid mynn	208	64.9916	15.2264	555	stream	99	17	68	11	3	0	0	6.6	0.049	4	122	2
Flarkån		48097	64.2498	21.0107	20	stream	147	79	3	1	0	16	0	4.4	0.329	27	890	11
Fyrisån		1E+05	59.8149	17.6704	1	stream	1988	56	6	1	1	31	4	7.5	0.177	62	2358	48
Hornsjöbäcken		1879	62.8450	17.3716	193	stream	40	84	12	0	4	0	0	6.9	0.089	4	196	4
Hågaån		1E+05	59.8420	17.5675	16	stream	107	63	3	1	0	30	3	7.5	0.160	65	3037	44
Härån (Storån)		1877	61.7360	16.4579	215	stream	21	92	8	0	0	0	0	6.4	0.264	7	328	3
Höjdabäcken		182	64.0344	16.9232	442	stream	5	78	0	8	14	0	0	5.9	0.157	5	207	1
Hökvattsån	Nr 1	213	63.8587	14.8808	367	stream	111	76	7	7	8	1	0	7.1	0.063	6	241	4
Kvarnån	Mynning (nederst)	202	63.4333	16.5292	151	stream	49	92	2	4	1	0	0	7.4	0.074	6	267	7
Källsjöbäcken	Finnbaracken	23833	61.6320	16.7018	282	stream	6	96	4	0	0	0	0	6.7	0.321	8	373	4
Källsjöån	Källsjöklack	1890	61.6318	16.7499	216	stream	18	93	6	0	1	0	0	6.1	0.336	11	451	2
Kärmsjöbäcken	Uppstr Lill-Kärmsjön	201	63.8676	16.8058	230	stream	25	60	34	3	3	0	0	6.6	0.138	6	268	3
Laxbäcken		1869	59.8536	15.4549	201	stream	9	94	4	2	0	0	0	4.8	0.445	10	597	3
Lekarån	uppströms bron	222	63.0750	13.7378	608	stream	90	10	86	1	3	0	0	6.7	0.018	1	82	1
Lillån, KIS54		2163	64.8485	20.9818	2	stream	40	65	4	0	0	31	0	4.2	0.267	32	1600	
Lillån, E4:an		2096	63.5972	19.7143	29	stream	14	89	0	11	0	0	0	4.8	0.441	14	522	3
Loån	Östanå andra bron 1	215	59.5551	18.5707	7	stream	89	81	1	0	9	7	1	7.3	0.163	26	901	13
Mattjokkbäcken	Uppstr väg	211	65.1739	18.2494	271	stream	125	46	17	29	7	0	0	6.5	0.131	10	438	2
Nygårdsbäcken		45898	60.2997	16.1848	172	stream	2	73	8	0	0	0	19	8.4	0.016	11	5350	143
Oradbäcken		5053	61.1679	13.4832	425	stream	19	48	3	49	0	0	0	6.0	0.323	12	284	3
Persbobäcken		45899	60.2693	16.2124	134	stream	49	66	12	0	8	5	10	7.2	0.045	8	1400	59
Rullshyttebäcken		21062	60.3025	16.1820	191	stream	5	62	9	0	6	0	23	7.4	0.103	17	1852	121
Rösjön 1 inl Väsman		1E+05	60.1998	15.1285	155	stream	2	86	0	0	12	0	2	7.7	0.091	88	1270	28
Silverbergasjön	1	4136	60.3541	15.5868	267	stream	3	87	1	0	12	0	0	5.3	0.238	7		
Stormrybäcken		194	62.2594	16.2680	414	stream	3	69	4	27	0	0	0	6.3	0.217	8	308	2
Storselsån	Storsele	1892	63.7363	18.1478	304	stream	74	75	11	6	8	0	0	6.8	0.152	6	272	3
Storsjön		3383	63.2983	14.4601	293	lake	12050	45	31	9	12	3	1	7.3	0.039	4	260	
Stortjärnsbäcken	Svartberget	193	64.2513	19.7779	242	stream	0.3	100	0	0	0	0	0	5.1	0.508	24	453	3
Stridbäcken	E4:an	2146	63.4836	19.2970	19	stream	13	93	4	3	0	0	0	6.6	0.217	7	247	3
Tomtsjön		27483	59.8036	18.8158	16	lake	2	65	0	0	23	12	0	7.6	0.073	21	743	12
Vapstälven	Skalmodal-kapellet	221	65.4433	14.5208	415	stream	600	30	54	7	9	0	0	7.3	0.011	1	61	4
Viskansbäcken		5049	62.4530	16.4385	79	stream	31	75	15	0	8	1	0	7.3	0.065	7	304	5
Ådalsån	Lyckemyran	1891	62.7446	17.1452	193	stream	50	80	5	15	1	0	0	6.7	0.379	9	354	3
Övre Klingens Utlopp		21111	60.4742	15.8466	138	stream	48	87	4	0	6	3	0	7.2	0.095	7		
Örvallsbäcken	Gravbacka	23830	61.6951	16.6633	114	stream	6	97	3	0	0	0	0	6.5	0.271	10	403	3
Övre Häggingan		23832	62.0116	13.3666	704	stream	21	9	81	11	0	0	0	6.0	0.089	3	117	2
Plogens utlopp		137	60.1561	15.3042	203	stream	1508	75			7		1	7.1	0.090			7
Prästvallsbäcken		4141	61.6329	16.7190	280	stream	13	85	5	0	8	2	0	6.9	0.336	8	402	4

once per strain. Sequencing was conducted in 2020 and was successful for almost all clonal cultures for both the *rbcL* and 18SV4 markers. Failed or low-quality sequences, if any, were excluded from downstream analyses.

Identification

The aim of the identification work was to give all established clonal cultures at least a genus name at the end of the project and to identify those missing from reference databases to the species level, if possible. The clones were photographed, measured, and studied using LM and SEM. LM analysis was performed for all cultures, whereas SEM analysis was restricted to a selection of taxa. The following literature was used for identification, including both standard literature and new scientific taxonomic publications specific to certain genera or taxonomic groups: *Achnantheidium* (Hofmann et al. 2011; Bey and Ector 2013), *Brachysira* (Van de Vijver et al. 2021a), *Eunotia* (Lange-Bertalot et al. 2011), *Fragilaria* (Lange-Bertalot and Ulrich 2014; Cantonati et al. 2019; Kahlert et al. 2019; Kahlert et al. 2021a; Van de Vijver et al. 2021b; Van de Vijver et al. 2022b; Van de Vijver et al. 2022c; Van de Vijver et al. 2023b), *Ulnaria* (Hofmann et al. 2011; Lange-Bertalot and Ulrich 2014), *Gomphonema* (Reichardt 1999; Hofmann et al. 2011; Abarca et al. 2014; Rose and Cox 2014; Abarca et al. 2020; Kahlert et al. 2021a), *Nitzschia* (Hofmann et al. 2011), *Pinnularia* (Lange-Bertalot and Metzeltin 1996; Krammer 2000; Blanco et al. 2012; Kollár et al. 2021; Kulikovskiy et al. 2023), *Tabellaria* (Heudre et al. 2021), *Frustulia* (Lange-Bertalot 2001; Urbankova and Vesela 2013; Kulichová and Fialová 2016; Urbánková et al. 2016), *Planothidium* (Hofmann et al. 2011; Van de Vijver et al. 2023a), other taxa (Krammer and Lange-Bertalot 2004; Hofmann et al. 2011).

To prepare the diatom samples for photography, preparations were made using a new rapid method (Trobajo and Mann 2019) recommended for material from cultures and for creating voucher material for deposition in natural history collections. The method was tested before use and proved to work very well. Taxa prioritized for SEM were those missing from reference databases or belonging to taxa with very few sequences. A total of 89 clones were prepared and photographed. The SEM work was carried out at BGBM with a high-resolution SU810 FE-SEM, Hitachi. The LM work was carried out at the Department of Aquatic Sciences and Assessment, SLU, using a Nikon Eclipse 80i light microscope with a Plan Apo objective (numerical aperture 1.40, 1000× magnification) with differential interference contrast (DIC).

In addition to the morphological analysis, the sequences were compared with reference datasets to assess their approximate phylogenetic placement and to evaluate the consistency between molecular clustering and morphological identifications. This analysis was intended as a supportive and exploratory step within the identification process, rather than as a formal phylogenetic framework. The reference database Diat.barcode for *rbcL* (Rimet et al. 2019) and the reference database for 18S from the Botanic Garden and Botanical Museum Berlin (BGBM), Freie Universität Berlin, Germany (Zimmermann et al. 2014), were used. Due to its overall better performance in delimiting species compared to the 18SV4 marker, the *rbcL* marker was used mainly as support for the identification work. In short, a multiple alignment was carried out with the new sequences (entire length for *rbcL* \geq 990 bp), using the Muscle algorithm (Edgar 2010) as implemented in MEGA6 (Tamura et al. 2013). This alignment was then added to the available multiple

alignment of the entire Diat.barcode library, which is maintained and updated continuously (Rimet et al. 2019; Gelis et al. 2025). The version with the short (263 bp) *rbcL* barcode (Rimet 2021) was used here to enable a rapid process, inserting the barcode region of the aligned sequences. A maximum likelihood phylogenetic analysis was then performed using RAxML v8.2.12 (Stamatakis 2014), applying the GTR Gamma model and 1,000 rapid bootstrap replicates.

Molecular clustering was then checked to determine whether it supported the proposed identifications by searching for discrepancies between clustered sequences within the project and with reference sequences. In cases of a mismatch, the potential cause was analyzed by considering whether it might be a known taxonomic uncertainty or update based on relevant literature. Species names were not altered in the few mismatches found, as such work is better addressed in future taxonomic studies. The results are preliminary because diatom identification is a challenging expert task requiring dedicated phylogenetic work for individual genera.

During the identification work, it became clear that many sequences could not easily be matched with existing morphological species, which is why they have been named with the attribute “cf.” (*confer*). This means that the name is relatively certain, but some morphological characters did not match exactly with a species that has already been described. Another attribute used was “aff.” (*affinis*), which means that the name is highly uncertain, but the morphology is similar to that of a certain taxon. The use of qualifiers such as “cf.” and “aff.” reflects a conservative and expert-based taxonomic approach, acknowledging the current state of diatom taxonomy, where morphological and molecular boundaries are still being actively refined.

Data/vouchering

The aim was to store all results, including sequences, vouchers, metadata, living clones if possible, and the identification links to the respective pieces of information in public collections and databases to enable and facilitate future research. The guidance for the establishment of DNA reference libraries was followed according to Rimet et al. (2021), where previous diatom guidelines are included and updated (Zimmermann et al. 2014; Rimet et al. 2018).

Vouchers (raw material, microscopical slides, and SEM stubs) and data of physical objects are stored in the Herbarium Berolinense (B) and the collection management system for diatoms, “Micro Algae SpecimenTool” (BGBM 2016+ (continuously updated)), of the Herbarium Berolinense. All strains of the FRESHBAR project are stored under their strain number as a unique identifier, together with a subset of environmental data (e.g., georeference, collector, sampling date, locality name, and identifier, as given in Table 1) of the Swedish gathering. Each strain is linked to the accession number in B. This accession number is unique within the Herbarium Berolinense and is represented by a worldwide unique stable identifier, as agreed by CETAF (Groom et al. 2017). For example, for B B 40 0046697, see <https://herbarium.bgbm.org/object/B400046697>. This ensures long-term traceability and interoperability among biodiversity data infrastructures. New species that will be published based on FRESHBAR strains will be registered in the PhycoBank registration system (<https://phycobank.org>) for registration; see Turland et al. (2025) and linked by the stable identifier to the collection management system (see also

Suppl. material 3: table S1). Each natural history collection object derived from a living strain will be linked to images (LM and SEM), as far as these are available via a publicly available image server and have stable URLs. Each specimen record includes an identification history and links to publications citing that specimen, ensuring transparency and reproducibility of taxonomic decisions. The DNA is stored freeze-dried in the DNA biobank of the Botanic Garden and Botanical Museum, Freie Universität Berlin, and the associated metadata and identifiers are published in the portal of the Global Genome Biodiversity Network (GGBN (Eds.) 2011+ (continuously updated)) following their data standard.

The sequences were submitted to an International Nucleotide Sequence Database Collaboration (INSDC) database (NCBI GenBank (www.ncbi.nlm.nih.gov/Genbank)) along with strain identifiers and stable identifiers to vouchers in B. Additionally, the *rbcL* sequences will be provided for inclusion in the reference database Diat.barcode. For integration into global biodiversity infrastructures, the appropriate pathway for publishing the linked but distributed FRESHBAR data to the Global Biodiversity Information Facility (GBIF) will be selected based on the final data architecture.

Long-term storage of diatom cultures

Maintaining long-term diatom cultures is a labor-intensive effort that requires great expertise and resources. Therefore, culture collections were contacted and asked about the possibility of taking over the FRESHBAR cultures once all analyses were completed. The BCCM/DCG Culture Collection in Ghent, Belgium (the Belgian Coordinated Collections of Microorganisms (BCCM) consortium 2025), agreed to host some of the cultures, and the UTEX Culture Collection of Algae, USA, agreed to take care of some of the surviving cultures at the end of the project. The priority in these collaborations was to secure long-term preservation of as many strains as possible and to ensure their availability for future research. In the end, 36 clones for long-term culture/storage (cryopreservation) were sent to BCCM/DCG and 32 to UTEX. BCCM/DCG is sequencing deposited cultures for the markers ITS, LSU, *rbcL*, and SSU as part of its routine process of culture identification.

Results

Of the 312 harvested strains, 301 were successfully sequenced at BGBM for the *rbcL* marker and 299 for the 18SV4 marker. For 294 strains, both *rbcL* and 18SV4 sequences could be generated, but in a few cases, only one of them could be generated (see Suppl. material 3: table S1 for details). The *rbcL* sequences were preliminarily classified into 17 genera, of which *Achnantheidium* contained the most sequences (70), followed by *Fragilaria* (60) (Suppl. material 3: table S1; Table 2). The genera *Fragilaria* and *Eunotia* showed the highest intra-generic diversity based on *rbcL* clustering (Table 2; Suppl. materials 1, 2: figs S1, S2). The 301 *rbcL* sequences enlarged the number of reference sequences available in Diat.barcode v10 from Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) by 200%. In relation to the existing reference sequences per genus, the relative increase was highest for *Tabellaria*, followed by *Brachysira* and *Eunotia* (Table 2).

The morphological identification work revealed 51 different taxa within the 17 genera, most of them in the genera *Eunotia* and *Fragilaria* (Suppl. material 3: table S1; Table 2). Of the 51 taxa, 38 were identified to the species level; however, most of them had some grade of uncertainty (Suppl. material 3: table S1; Table 2). The molecular clustering, based on *rbcL* sequences, was generally consistent with the morphological identification at the genus level and, in most cases, also infrageneric (Suppl. materials 1, 2: figs S1, S2). Therefore, the genus-level identification was considered robust, as was the separation of the taxa within a genus. The decision to use identification labels indicating uncertainty in a taxonomic label was made to highlight the fact that this dataset needs more taxonomic work, which was not possible within the scope of the project. The challenges were multiple, with entire strain clusters not fitting perfectly with any described morphological species, other clusters where only some strains did not fit perfectly with respect to the morphological characters, or clusters where SEM data were lacking but would have been needed to resolve a species identification. The molecular clusters were, in some cases, quite well separated from other taxa but not always, and the relationship to the available reference sequences was not studied in detail using long sequences and the 18S results. Therefore, the analysis of how many new species the project added to Diat.barcode v10 ($n = 24$) is preliminary (Table 2). The number of new species names (18) was counted, plus the number of taxon clusters that did not receive a final or certain identification but clustered separately in relation to the available references in Diat.barcode v10, probably representing new species to the database as well (6).

Some highlights of the results are the relatively high proportion of *Eunotia* taxa and the *Tabellaria* and *Brachysira* sequences—those genera are poorly represented in diatom databases but occur frequently in Swedish freshwaters (Kahlert et al. 2018; Rimet et al. 2019). These genera are known to prefer oligotrophic and acidic conditions (Hofmann et al. 2011). For example, the conspicuous species *Eunotia myrmica* Lange-Bertalot has not previously been represented in any database (Fig. 1). In addition to the sequences, several morphological characteristics of living cultures were also documented in images, which is comparatively rare in diatom barcoding projects. In many cases, colony formation and the shape of these colonies were documented. In a few cases, the formation of auxospores and cells of different lengths in the life cycle was also documented. The images of the electron microscopic details also provide valuable information about the taxa in this study and how they differ from closely related taxa.

The physical material of all FRESHBAR clonal cultures has been deposited in the BGBM database and herbarium by the project partner Botanic Garden and Botanical Museum Berlin (BGBM), Freie Universität Berlin, Germany. BGBM is responsible for storing material and information for all established clones and for making them traceable and usable for further analysis (BGBM 2016+; search for FRESHBAR in the field “Collections”). All results are available with open access, including the sequences, which are publicly available from INSDC, and the *rbcL* sequences will be incorporated into the next version of Diat.barcode (Rimet et al. 2019). Images are deposited at SLU (Kahlert et al. 2025; Kahlert et al. 2026). The complete information about all clones, sequences, and other material, such as images, is also saved internally at the Department of Aquatic Sciences and Assessment, SLU, and can be ordered by sending a request to Maria Kahlert (maria.kahlert@slu.se).

Table 2. FRESHBAR results. Identifications for the 301 strains of benthic freshwater diatom species from Sweden with successfully generated *rbcL* sequences (one per strain) and contributions to the *rbcL* reference database Diat.barcode v10. The number of *rbcL* strains (= sequences) per taxon (genus and infrageneric) and the available number of sequences in Diat.barcode v10 for each genus are shown. Taxa new to Diat.barcode are highlighted. The taxon identifications represent preliminary identifications based on LM/SEM and *rbcL* sequences. Species designations marked with “cf.” or “aff.” indicate taxonomic uncertainty. *Four taxa included a few strains identified as “cf.” due to some morphological characters not fitting perfectly (nr. of “cf.” strains marked with *).

Genus	Number of strains per genus			Number of strains per taxon cluster (infrageneric species/taxon)	taxon new to Diat.barcode vs 10?	nr of available sequences of the respective genus in Diat. barcode v10
<i>Achnantheidium</i>	70	<i>A. minutissimum</i> complex		70	0	186
<i>Fragilaria</i>	60	<i>F. campyla</i>	(Hilse) Van de Vijver, Kusber & D.M.Williams	8	1	167
		<i>F. cf. acerosa</i>	Van de Vijver, C.E.Wetzel, Jarlman & Ector	1	1	
		<i>F. cf. capucina</i>	Desmazières	2	1	
		<i>F. cf. heatherae</i>	Kahlert & M. G. Kelly	1	0	
		<i>F. radians</i> /cf. <i>radians</i> *	(Kützing) D.M.Williams & Round	11/1*	0	
		<i>F. species unknown 1</i>		5	1	
		<i>F. species unknown 2</i>		11	1	
		<i>F. species unknown 3</i>		10	1	
		<i>F. tridentina</i> /cf. <i>tridentina</i> *	Cantonati & Lange–Bertalot	6/1*	1	
		<i>F. sp.</i>		4	NA	
<i>Ulnaria</i>	9	<i>U. cf. grunowii</i>	(Lange-Bertalot & S.Ulrich) Cantonati & Lange-Bertalot	2	1	47
		<i>U. cf. delicatissima</i>	(W.Smith) M.Aboal & P.C.Silva	4	1	
		<i>U. cf. danica</i>	(Kützing) Compère & Bukhtiyarova	2	1	
		<i>U. sp.</i>		1	NA	
<i>Gomphonema</i>	53	<i>G. aff. drutlingense</i>	E.Reichardt	4	1	198
		<i>G. cf. "nordicum 1"</i>		2	1	
		<i>G. cf. "nordicum 2"</i>		19	1	
		<i>G. cf. angustatum</i>	(Kützing) Rabenhorst	8	1	
		<i>G. cf. narodoense</i>	R. Jahn, N.Abarca, J. Zimmermann & Enke	3	0	
		<i>G. cf. lagenula</i>	Kützing	3	0	
		<i>G. montanum</i> /cf. <i>montanum</i> *	Schumann	10/1*	1	
		<i>G. sp.</i>		3	NA	
<i>Eunotia</i>	51	<i>E. bilunaris</i>	(Ehrenberg) Schaarschmidt	11	0	105
		<i>E. cf. seminulum</i>	Nörpel-Schempp & Lange-Bertalot	3	1	
		<i>E. flexuosa</i> / <i>pseudoflexuosa</i> / <i>latitaenia</i> – group		8	0	
		<i>E. glacialispinosa</i>	Lange-Bertalot & Cantonati	3	1	
		<i>E. myrmica</i>	Lange-Bertalot	2	1	
		<i>E. implicata</i>	Nörpel, Lange-Bertalot & Alles	3	0	
		<i>E. implicata</i> or <i>minor</i>		6	1	
		<i>E. incisa</i>	W.Smith ex W.Gregory	2	1	
		<i>E. pectinalis</i>	(Kützing) Rabenhorst	1	0	
		<i>E. sp.</i>		12	NA	
<i>Tabellaria</i>	17	<i>T. flocculosa</i> /cf. <i>flocculosa</i> *	(Roth) Kützing	9/3*	0	10
		<i>T. procera</i>	Heudre & al.	5	1	
<i>Pinnularia</i>	14	<i>P. aff. bertrandii</i>	Krammer	7	1	392
		<i>P. biceps</i> var. <i>gibberula</i>	(Hustedt) Krammer	2	1	
		<i>P. viridiformis</i> MT2	Krammer	5	0	
<i>Nitzschia</i>	12	<i>N. cf. acidoclinata</i>	Lange-Bertalot	2	0	447
		<i>N. cf. gracilis</i>	Hantzsch	7	0	
		<i>N. sp.</i>		3	NA	
<i>Encyonema</i>	4	<i>E. vulgare</i>	Krammer	3	1	40
		<i>E. cf. silesiacum</i>	(Bleisch) D.G.Mann	1	0	
<i>Brachysira</i>	2	<i>B. cf. microcephala</i>	(Grunow) Compère	2	0	4
<i>Craticula</i>	2	<i>C. buderi</i>	(Hustedt) Lange-Bertalot	2	0	16
<i>Asterionella</i>	1	<i>A. formosa</i>	Hassall	1	0	8
<i>Cyclotella</i>	1	<i>C. meneghiniana</i>	Kützing	1	0	63
<i>Frustulia</i>	1	<i>F. crassinervia-saxonica</i> complex		1	0	28

Genus	Number of strains per genus			Number of strains per taxon cluster (infrageneric species/ taxon)	taxon new to Diat. barcode vs 10?	nr of available sequences of the respective genus in Diat. barcode v10
<i>Navicula</i>	2	<i>N. sp.</i>		2	NA	232
<i>Planothidium</i>	1	<i>P. lanceolatum</i>	(Brébisson ex Kützing) Lange-Bertalot	1	0	44
<i>Staurosira</i>	1	<i>S. cf. venter</i>	(Ehrenberg) Cleve & J.D.Möller	1	0	33

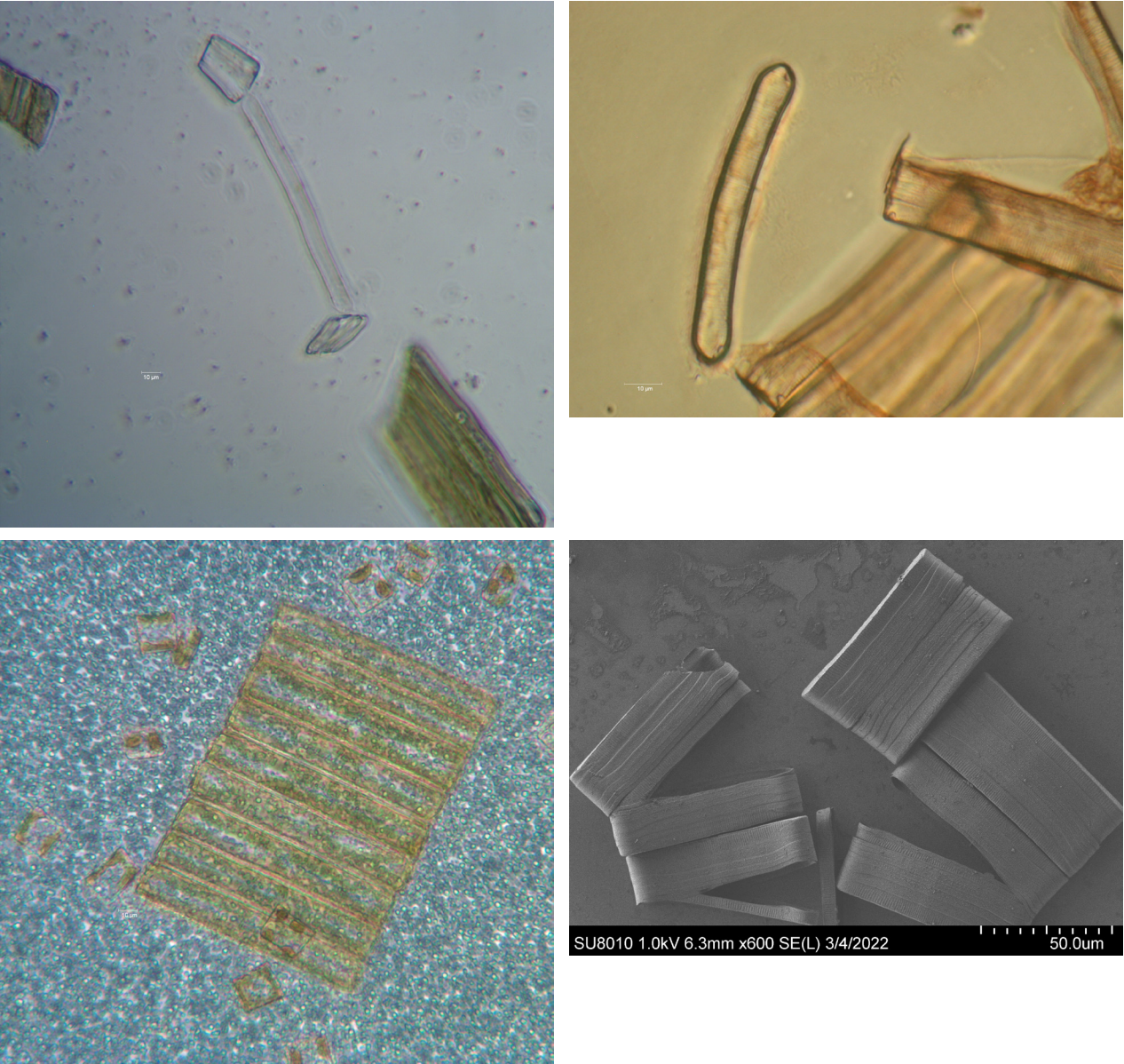


Figure 1. *Eunotia myrmica* Lange-Bertalot isolated in the FRESHBAR project.

The 36 clonal cultures sent to the BCCM/DCG collection in Belgium received the accession numbers DCG 1032, 1035–1045, 1093–1101, and 1211–1225 (Suppl. material 3: table S1; The Belgian Coordinated Collections of Microorganisms (BCCM) consortium 2025). Three cultures could not be recovered, while the remaining strains are now publicly available (Table 3). The agreement with BCCM/DCG is that the cultures are handed over to the collection with open access to both cultured material and any results they produce. BCCM/DCG has worked on purifying the cultures from contaminants by cultivating clone cultures from

healthy cells. BCCM/DCG has also analyzed four different barcodes (ITS, LSU, *rbcL*, SSU) for almost all cultures. Furthermore, BCCM/DCG has tested long-term storage of the cultures using cryopreservation. Most cultures are now cryopreserved, except for three that are now grown in liquid medium and four that died after some time. Fifteen different FRESHBAR taxa are now deposited at BCCM/DCG. The 32 clonal cultures sent to UTEX are not yet public. In the meantime, information can be obtained by contacting Maria Kahlert (maria.kahlert@slu.se).

Suppl. material 3: table S1 specifies the available results for the established clonal cultures. While it was not possible to obtain a complete set of results for every clonal culture, as much information as possible was collected and documented. For example, for the clonal culture SE037_1C11, identified as “*Fragilaria* unknown species nr. 3,” with voucher ID at BGBM B 40 0046536 (stable identifier: <https://herbarium.bgbm.org/object/B400046697>), two LM slides, a vial with cleaned material for further morphological analyses, and an SEM stub are available at BGBM. An LM image from the slide ID DSCN8146 is deposited at SLU and accessible with the title 1C11 DSCN8146 under the permanent link <https://hdl.handle.net/20.500.12703/6968> in the folder *Fragilaria* unknown species nr. 3. In the same deposit, there are also three images of living cells from the clonal culture (titles: isolate 191007 FGRA Vapstälven plate 1C11, isolate 1C11, and 5 cm dish 1c11 FGRA), as well as a series of SEM images, all found under the clonal culture ID 1C11. The culture has also been deposited in the BCCM/DCG Diatoms Collection under the accession number ID DCG 1211, and a private sample has been sent to Matt Ashworth at UTEX. The genetic material is vouchered at BGBM under the ID DB42560, and the *rbcL* and 18SV4 sequences were submitted to GenBank (BioProject ID PRJNA1454281). Information about the sampling locality can be found in Table 1. More information about the site, including biological data (i.e., diatom taxon composition), can be requested from the Miljödata-MVM database host (Miljödata-MVM 2025).

Table 3. FRESHBAR results for benthic freshwater diatom species from Sweden. Clones and new reference sequences (ITS, LSU, *rbcL*, SSU) deposited at the BCCM/DCG collection in Ghent, Belgium. Identified genera and taxon groups, as well as the number of included sequences and unique clones (some with replicates). * Unsuccessful attempt to either maintain a living culture or cryopreserve it.

Genus	Number of strains	Annotated name	Deposition
<i>Achnanthes</i>	10	<i>A. minutissimum</i> complex	cryopreservation
<i>Fragilaria</i>	1	<i>F. species</i> unknown 1	cryopreservation
	2	<i>F. species</i> unknown 3	cryopreservation
<i>Gomphonema</i>	2	<i>G. montanum</i>	culture (one dead*)
	2	<i>G. cf. "nordicum 2"</i>	cryopreservation
	1	<i>G. cf. narodoense</i>	cryopreservation
	2	<i>G. cf. angustatum</i>	cryopreservation
	1	<i>G. sp.</i>	cryopreservation
<i>Nitzschia</i>	1	<i>N. cf. gracilis</i>	cryopreservation
	1	<i>N. sp.</i>	cryopreservation
<i>Pinnularia</i>	1	<i>P. aff. bertrandii</i>	culture
	1	<i>P. viridiformis</i> MT2	culture
<i>Staurosira</i>	1	<i>S. cf. venter</i>	cryopreservation
<i>Tabellaria</i>	2	<i>T. flocculosa</i>	cryopreservation
<i>Ulnaria</i>	1	<i>U. cf. delicatissima</i>	culture
<i>Eunotia</i>	2	<i>E. flexuosa/pseudoflexuosa/latitaenia</i> – group	dead*
	1	<i>E. bilunaris</i>	dead*

Discussion

Many new diatom barcodes were generated during the project for the markers *rbcL* and 18SV4; they are now publicly available at INSDC and will be incorporated into the reference database Diat.barcode, which is available at INRAE in France (contact: Frédéric Rimet), and the database at the Botanic Garden and Botanical Museum Berlin (BGBM), Freie Universität Berlin (contact: Jonas Zimmermann). These are two of Europe's most important reference databases that have been developed to enable diatom metabarcoding for monitoring lakes and streams. Both databases are curated, which means that the sequences are quality assured, and the diatom species name has undergone a check by one or more taxonomic experts. Continuing the work to populate diatom databases with new barcodes is important to enable the ecological interpretation of data derived with molecular methods, e.g., for use in environmental monitoring, and also to contribute to the recent work of revisiting diatom taxonomy in light of genetic data (Kollár et al. 2025).

However, species identification of the clonal cultures was challenging. This reflects the current state of diatom taxonomy, where species concepts are still being actively revised and where both morphological and molecular boundaries are not yet fully resolved. As a result, reliable species identification often requires integrative approaches, including multiple molecular markers and detailed morphological analyses (e.g., SEM), ideally carried out within dedicated taxonomic studies focusing on individual genera or species complexes. With the help of all project partners and the most recent literature, 38 taxa were identified as close to species level as possible. Nevertheless, some clonal cultures could only be identified to genus level or to a certain taxon complex, as their morphology deviated considerably from that of any described species. Other clonal cultures matched a described species complex reasonably closely, with only minor characters not fitting, and were in these cases assigned to a species complex with some epithet of uncertainty. More time and knowledge are needed to understand whether a clonal culture represents a species that is new to science or whether it correlates with a species that has been described with too little morphological variation. More detailed analyses of the genetic variation within species and genera are needed, including how molecular lineages relate to ultra-structural morphology and how, in turn, morphological characters are connected to a taxon grouped together by molecular characters. There is an urgent need to combine research in molecular diatomology and traditional morphological taxonomy to unravel diatom biodiversity (Kollár et al. 2025). Diatom taxonomy is evolving at a rapid pace, but unfortunately, it is still common for new species to be described solely based on morphological characteristics (e.g., Heudre et al. 2021; Van de Vijver and Williams 2022). However, as such validly published species often are well documented and show significant morphological differences from other known species, there is likely high underlying biodiversity that requires molecular support for accurate delimitation, e.g., for *Brachysira* from Scandinavia (see Van de Vijver et al. 2022a). Linking molecular sequences to the morphology of organisms will be an ongoing scientific task. The data storage strategy bridges the gap between this publication and future taxonomic revisions. The sequences, stored in INSDC, link back to the voucher in Herbarium Berolinense by a stable resolvable identifier. Each voucher is linked to the clonal culture and has an identification history. If, for example, the *Achnantheidium*

minutissimum complex is resolved in the future or an unknown *Fragilaria* sp. is formally described as a species new to science, the data curation at BGBM will keep track and update the respective records. Further, novelties registered at PhycoBank are also linked to the relevant vouchers at natural history museums, ensuring long-term interoperability (Müller et al. 2022). The findings contribute to understanding whether species, existing or newly described, are truly distinct genetic entities or whether the definition of certain species should be reassessed. There are many ongoing research projects on this (Abarca et al. 2014; Kollár et al. 2019; Pinseel et al. 2020). The FRESHBAR data can now be included in such projects as results, as well as physical vouchers, and in a few cases, even algal strains are or will be available in publicly accessible databases and collections.

Many of the FRESHBAR clones belong to taxa that already exist in sequence reference databases. These references are nevertheless valuable because they originate from Sweden, and some of them are common here as well. Indeed, the dataset tripled the amount of reference *rbcL* sequences in Diat.barcode v10 with an origin in the Nordic countries. These sequences can now be used for Swedish metabarcoding with greater certainty than if there had only been sequences from other biogeographical areas (Baillet et al. 2019). Furthermore, the more sequences that are available for a taxon in a reference database, the more secure the annotation of a new sequence to a taxon name is (Baillet et al. 2019). In addition, the results contribute to increasing knowledge about their biogeographical distribution, taxonomy, and ecology (Vaulot et al. 2023). Finally, several of the identified taxa belong to diatom genera where taxonomic research is in full swing and where all sequences within these genera are valuable for research. This applies in particular to *Eunotia* (Vanormelingen et al. 2008; Lange-Bertalot et al. 2011) and *Fragilaria* (Kahlert et al. 2019), but also to *Gomphonema* (Abarca et al. 2020), *Pinnularia* (Kollár et al. 2021), and others. Many new species are described in these genera, and results are now available that can put Swedish taxa into context. With more sequences of the same species in the databases, knowledge about intraspecific variation can also be increased, species identification can be made more reliable, and environmental tolerances can be reassessed either at the species level or from genetic variants.

The *rbcL* barcodes generated in the FRESHBAR project with Sanger sequencing are relatively long (ca. 1000 bp) compared to the *rbcL* barcode used for metabarcoding purposes (263–331 bp) (Pérez-Burillo et al. 2022) and therefore enable more detailed studies of taxonomic resolution, lineage divergence, and intra- and interspecific genetic variation (Pérez-Burillo et al. 2022). Longer barcoding sequences have been shown to include more variations in their composition, and the possibility of being able to distinguish between species is then greater (Gueidan and Li 2022). Preferably, molecular research and assessment should use long sequences and several markers for species identification and delimitation (Gueidan and Li 2022). Changing from the current short-read sequencing platforms to long-read ones, such as Pacific Bioscience or Oxford Nanopore Technologies, for metabarcoding will further improve the taxonomic resolution of species (Pérez-Burillo et al. 2022; Chwalińska et al. 2025). However, it is discussed that high-quality DNA material is needed for the metabarcoding of longer markers, and such quality is not always available from environmental samples.

Each added sequence, even sequences of more common species, potentially facilitates identification when using a reference database in metabarcoding projects. However, it is important that barcoding sequences are taxonomically

accurately and consistently annotated to enable the further development of automated taxon detection for metabarcoding. Automated bioinformatic pipelines will otherwise detect a conflict when, for example, identical sequences have different names (Bailet et al. 2020). The results highlight that assigning a correct name to a new barcoding sequence remains challenging and that the gaps in taxonomic knowledge require further work, ideally involving close collaboration between molecular and morphological experts and including developers of bioinformatic pipelines to ensure that practical solutions are clearly shown. For example, using a short *rbcL* barcode for metabarcoding means that certain species cannot be separated based on the short sequence (Pérez-Burillo et al. 2022; Ács et al. 2025), and a pipeline should highlight which species are pooled behind an annotated name, as well as the probability of correctness.

Finally, to be able to use metabarcoding as decision support, knowledge must be available about reference communities and the difference between good and moderate status. The essential tools for this ecological assessment are metrics based on genetic data. The approaches to developing such metrics differ, as does their use of sequence reference databases. One approach is to directly translate molecular units to traditional species names via a reference database and then directly use knowledge of the ecological preferences of the morphotaxa and traditional diatom metrics for environmental assessment (Hering et al. 2018). Alternatively, or additionally, the biological unit used can be the molecular unit directly, either as unique or as clustered sequences (Keck et al. 2017; Hering et al. 2018; Kelly et al. 2020; Kelly et al. 2024). Direct use of the molecular units has been suggested to acknowledge and take advantage of the fact that diatom assemblage composition data generated using light microscopy and metabarcoding are different, both regarding taxonomic resolution and abundance estimation (Kelly et al. 2020). Indeed, it has been shown that genetic variants (e.g., ASVs and ESVs) can carry ecological information hidden by cryptic diatom diversity and that this information is well kept when building models to indicate environmental pressure-response relationships based on unassigned diatom sequences (Kelly et al. 2024). However, developing metrics based solely on DNA data to overcome the uncertainties of species identification will still require a good understanding of the ecological preferences of the molecular units used for the metric (Apothéoz-Perret-Gentil et al. 2020). A large effort will be necessary to ensure coverage of a range of stressor values at least as broad as that used for the development of the traditional methods (Pawlowski et al. 2018), and databases will be needed to host both the molecular units and their ecological profiles (Keck et al. 2017). Finally, it has been suggested to move away from environmental assessment based on single organism groups toward a more holistic analysis of the stress response of multiple organism groups (Pawlowski et al. 2018; Kelly et al. 2024). Indeed, it has been shown that such metrics might capture the environmental signal better than using diatoms alone (Simons et al. 2023). However, while such attempts require the use of other barcodes than organism-specific ones, specific and curated databases such as Diat.barcode will still be needed in the future to transform sequences into understandable pieces of biological information, connecting them to known information.

The FRESHBAR results can now, together with the many other new molecular datasets currently being generated, be used to unravel not only diatom taxonomy but also information on the ecological profiles and geographical distribution of diatom units. Some work has already started using FRESHBAR results

(Dani et al. 2025; Kollár et al. 2026). Connecting studies on diatom ecological preferences and geographical distribution with molecular lineages is urgently recommended to improve species delimitation and enable future environmental monitoring and research to better interpret the ecological meaning of diatom assemblages and taxonomic relationships. Even if full completion of reference libraries with all available different sequences might not be realistic, populating them with new sequences will aid in understanding diatom assemblages and will make environmental assessment of understudied regions using molecular units less biased. There are many projects globally acquiring new sequences, and databases are continuously increasing their content (Epoch AI 2024). New diatom sequences are, furthermore, not only generated by the slow method of isolating and culturing single cells but also via metabarcoding projects directly (Rimet et al. 2018) and via direct sequencing of single cells, avoiding the need to culture (Hamilton et al. 2015). Finally, a project such as FRESHBAR is not only generating barcodes but also enables further research using the DNA extracts, other physical material, and the publicly available metadata, opening up continued use of the results for new purposes beyond those originally anticipated.

Acknowledgements

The authors thank Gunilla Ejdung and Kari Stange from the Swedish Environmental Protection Agency, as well as Michael Haldin from the Swedish Agency for Marine and Water Management, for their interest and support in developing DNA methods. The authors also thank everybody who contributed to the project, not least by sending fresh diatom samples. Here, special thanks go to Erika Melander, Daniel Larsson, Paul Andersson, Willem Goedkoop, and Guillaume Grosbois, who made fresh samples available. The authors thank Sara Goncalves and Bonnie Baillet for all the work they have completed on diatom cultures together with Demetrio Mora. The authors thank the staff at BGBM's Research Group Diatoms, especially Jana Bansemer and Juliane Bettig, for carrying out the DNA analyses and helping with the electron microscope and Gabriele Droege and Belen Escobari for making stored DNA and DNA-Bank metadata available. Maria Kahlert would especially like to thank Juliane Bettig, Agnes Kirchhoff, Oliver Skibbe, Maryam Malekmohammadi, and the entire staff in Berlin for making the stay in Berlin very enjoyable even during the Corona restrictions. Many thanks go to Olga Chepurnova and Peter Chaerle from BCCM/DCG and Matt Ashworth from UTEX for their help with the long-term storage of living cultures. Special thanks go to Åke Olson and Katarina Ihrmark from SLU for lending space in an -80°C freezer. The authors thank the reviewers, Martyn Kelly and an anonymous reviewer, and the subject editor, Armando Espinosa Prieto, for their careful reading of the manuscript and their constructive comments.

References

- Abarca N, Jahn R, Zimmermann J, Enke N (2014) Does the cosmopolitan diatom *Gomphonema parvulum* (Kützinger) Kützinger have a biogeography? PLOS ONE 9: e86885. <https://doi.org/10.1371/journal.pone.0086885>
- Abarca N, Zimmermann J, Kusber W-H, Mora D, Van AT, Skibbe O, Jahn R (2020) Defining the core group of the genus *Gomphonema* Ehrenberg with molecular and morphological methods. Botany Letters 167: 114–159. <https://doi.org/10.1080/23818107.2019.1694980>

- Ács É, Korponai JL, Kiss KT, Bíró T, Duleba M, Grigorszky I, Szekeres J, Vadkerti E, Somlyai I (2025) Integration of morphology-based and molecular techniques to study the most significant diatom order (Thalassiosirales) of phytoplankton in the River Danube. *Freshwater Biology* 70: e14360. <https://doi.org/10.1111/fwb.14360>
- Apothéloz-Perret-Gentil L, Bouchez A, Cordier T, Cordonier A, Guéguen J, Rimet F, Vasselon V, Pawlowski J (2020) Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. *Molecular Ecology* 30: 2959–2968. <https://doi.org/10.1111/mec.15646>
- Baillet B, Bouchez A, Franc A, Frigerio J-M, Keck F, Karjalainen S-M, Rimet F, Schneider S, Kahlert M (2019) Molecular versus morphological data for benthic diatoms bio-monitoring in Northern Europe freshwater and consequences for ecological status. *Metabarcoding and Metagenomics* 3: e34002. <https://doi.org/10.3897/mbmg.3.34002>
- Baillet B, Apothéloz-Perret-Gentil L, Baricevic A, Chonova T, Franc A, Frigerio J-M, Kelly M, Mora D, Pfannkuchen M, Proft S, Ramon M, Vasselon V, Zimmermann J, Kahlert M (2020) Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Science of The Total Environment* 745: 140948. <https://doi.org/10.1016/j.scitotenv.2020.140948>
- Bey MY, Ector L (2013) Atlas des diatomées des cours d'eau de la région Rhône-Alpes. Direction regionale de l'Environnement, de l'Amenagement et du Logement Rhone-Alpes, Lyon, 1182 pp.
- BGBM (2016+) [continuously updated] Micro Algae Collection Database at B. https://diatoms.bo.berlin/algae_bgbm/sptool/sptool.php [accessed 3rd of December 2025]
- Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund W, Zampoukas N, Hering D (2012) Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators* 18: 31–41. <https://doi.org/10.1016/j.ecolind.2011.10.009>
- Blanco S, Álvarez-blanco I, Cejudo-Figueiras C, Bécares E (2012) Contribution to the diatom flora of Cambodia: Five new recent freshwater taxa. *Journal of Systematics and Evolution* 50: 258–266. <https://doi.org/10.1111/j.1759-6831.2012.00189.x>
- Cantonati M, Angeli N, Lange-Bertalot H (2019) Three new *Fragilaria* species (Bacillariophyta) from low-conductivity mountain freshwaters (Alps and Apennines). *Phytotaxa* 404: 261–274. <https://doi.org/10.11646/phytotaxa.404.6.5>
- CEN (2014) EN 14407:2014 Water quality – Guidance for the identification and enumeration of benthic diatom samples from rivers and lakes. Comité Européen de Normalisation, Geneva, Switzerland, 13 pp.
- CEN (2025) EN 13946:2025 Water quality – Guidance standard for the routine sampling and preparation of benthic diatoms from rivers and lakes. Comité Européen de Normalisation, Geneva, Switzerland, 13 pp.
- Chwalińska M, Karlicki M, Romac S, Not F, Karnkowska A (2025) From short to long reads: enhanced protist diversity profiling via Nanopore metabarcoding. *Metabarcoding and Metagenomics* 9. <https://doi.org/10.3897/mbmg.9.163750>
- Dani M, Beszteri S, Castellanos AB, Schimani K, Skibbe O, Zimmermann J, Soares AR, Griesdorn L, Probst AJ, Kahlert M, Beszteri B (2025) Species delimitation within the *Achnantheidium minutissimum* complex (Bacillariophyta), based on morphological, molecular and ecophysiological approaches. *Journal of Phycology* 62: 25–43. <https://doi.org/10.1111/jpy.70124>
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>

- Epoch AI (2024) with major processing by Our World in Data. "Number of entries in biological sequence databases" [dataset]. Epoch AI, "Trends in Biological Sequence Data" [original data]. [Retrieved March 19, 2026 from] <https://archive.ourworldindata.org/20260304-094028/grapher/number-of-entries-in-biological-sequence-databases.html> [archived on March 4, 2026]
- European Parliament and Council of the European Union (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Union L327, 73 pp.
- Gelis MMN, Barry-Martinet R, Briand J-F, Chonova T, Cochero J, Gassiole G, Kahlert M, Karthick B, Keck F, Kelly M, Kochoska H, Mann D, Nayak P, Pfannkuchen M, Servulo T, Trobajo R, Vasselon V, Vidakovic D, Viollaz L, Wetzel C, Zimmermann J, Rimet F (2025) Assigning taxonomy and traits to DNA sequences of river diatoms in a region with limited taxonomic knowledge using the updated and annotated reference library Diat.barcode. *International Journal of Limnology* 61: 10. <https://doi.org/10.1051/limn/2025009>
- GGBN (2011+) [continuously updated] The GGBN Data Portal. <https://www.ggbn.org> [accessed 26th of July 2025]
- Gonçalves S, Kahlert M, Almeida SFP, Figueira E (2018) Assessing Cu impacts on freshwater diatoms: biochemical and metabolomic responses of *Tabellaria flocculosa* (Roth) Kützing. *Science of The Total Environment* 625: 1234–1246. <https://doi.org/10.1016/j.scitotenv.2017.12.320>
- Groom Q, Hyam R, Güntsch A (2017) Stable identifiers for collection specimens. *Nature* 546: 33–33. <https://doi.org/10.1038/546033d>
- Gueidan C, Li L (2022) A long-read amplicon approach to scaling up the metabarcoding of lichen herbarium specimens. *MycKeys* 86: 195–212. <https://doi.org/10.3897/mycokeys.86.77431>
- Guillard RRL, Lorenzen CJ (1972) Yellow-green algae with chlorophyllide C^{1,2}. *Journal of Phycology* 8: 10–14. <https://doi.org/10.1111/j.1529-8817.1972.tb03995.x>
- Hamilton PB, Lefebvre K, Bull R (2015) Single cell PCR amplification of diatoms using fresh and preserved samples. *Frontiers in Microbiology* 6: 16. <https://doi.org/10.3389/fmicb.2015.01084>
- Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research* 138: 192–205. <https://doi.org/10.1016/j.watres.2018.03.003>
- Heudre D, Wetzel EC, Lange-Bertalot H, Van de Vijver B, Moreau L, Ector L (2021) A review of *Tabellaria* species from freshwater environments in Europe. *Fottea* 21: 180–205. <https://doi.org/10.5507/fot.2021.005>
- Hofmann G, Werum M, Lange-Bertalot H (2011) Diatomeen im Süßwasser-Benthos von Mitteleuropa. Bestimmungsflora Kieselalgen für die ökologische Praxis. Über 700 der häufigsten Arten und ihre Ökologie. A.R.G. Gantner Verlag K.G., Rugell, 908 pp.
- Jahn R, Kusber W-H, Skibbe O, Zimmermann J, Van AT, Buczkó K, Abarca N (2019) *Gomphonella olivacea* (Bacillariophyceae) – a new phylogenetic position for a well-known taxon, its typification, new species and combinations. *Plant Ecology and Evolution* 152(2): 219–247. <https://doi.org/10.5091/plecevo.2019.1603>
- Jahn R, Abarca N, Kusber W-H, Skibbe O, Zimmermann J, Mora D (2020) Integrative taxonomic description of two new species of the *Cocconeis placentula* group (Bacil-

- lariophyceae) from Korea based on unialgal strains. *Algae* 35: 303–324. <https://doi.org/10.4490/algae.2020.35.8.1>
- Johnson RK, Wiederholm T, Rosenberg DM (1993) Freshwater biomonitoring using individuals organisms, populations, and species assemblages of benthic macroinvertebrates. In: Rosenberg DM, Resh VH (Eds) *Freshwater Biomonitoring and Benthic Invertebrates*. Chapman & Hall, New York, USA, 40–158. <https://doi.org/10.5860/choice.30-6159>
- Jupke JF, Birk S, Apostolou A, Aroviita J, Baattrup-Pedersen A, Baláži P, Barešová L, Blanco S, Borrego-Ramos M, van Dam H, Dimitriou E, Feld CK, Ferreira MT, Gecheva G, Gomà J, Hanžek N, Haslev IM, Isheva T, Jamoneau A, Jyrkänkallio-Mikkola J, Kahlert M, Karaouzas I, Karjalainen SM, Olenici A, Panek P, Paril P, Peeters ETHM, Polášek M, Pont D, Pumpuťte A, Sandin L, Sochuliaková L, Soininen J, Stanković I, Straka M, Šušnjara M, Sutela T, Tison-Rosebery J, Udovič MG, Verhofstad M, Žitinić P, Schäfer RB (2023) European river typologies fail to capture diatom, fish, and macrophyte community composition. *Science of The Total Environment* 896: 165081. <https://doi.org/10.1016/j.scitotenv.2023.165081>
- Kahlert M, Albert R-L, Anttila E-L, Bengtsson R, Bigler C, Eskola T, Galman V, Gottschalk S, Herlitz E, Jarlman A, Kasperoviciene J, Kokocinski M, Luup H, Miettinen J, Paunksnyte I, Piirsoo K, Quintana I, Raunio J, Sandell B, Simola H, Sundberg I, Vilbaste S, Weckstrom 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: 471–482. <https://doi.org/10.1007/s10811-008-9394-5>
- Kahlert M, Jarlman A, Sundberg I, Herlitz E (2018) *Taxalista – kiselalger i svenska söt-vatten*. https://www.slu.se/contentassets/14e5c0b0cd354b7a9cbcd85181d44670/kiselalger_mall_20180405.xlsx [accessed 6th of November 2025]
- Kahlert M, Kelly MG, Mann DG, Rimet F, Sato S, Bouchez A, Keck F (2019) Connecting the morphological and molecular species concepts to facilitate species identification within the genus *Fragilaria* (Bacillariophyta). *Journal of Phycology* 55: 948–970. <https://doi.org/10.1111/jpy.12886>
- Kahlert M, Aarnio S, Anijalg U, Gälman V, Karjalainen SM, Pajunen V, Teittinen A, Verweij G, Vilmi A (2021a) 6th Nordic-Baltic diatom intercalibration/harmonization exercise 2020. Diatom exercise in times of pandemics. Rapport / Sveriges lantbruksuniversitet, Institutionen för vatten och miljö. Institutionen för vatten och miljö, Uppsala, 6 pp.
- Kahlert M, Bailet B, Chonova T, Karjalainen SM, Schneider SC, Tapolczai K (2021b) 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, Rühland KM, Lavoie I, Keck F, Saulnier-Talbot E, Bogan D, Brua RB, Campeau S, Christoffersen KS, Culp JM, Karjalainen SM, Lento J, Schneider SC, Shaftel R, Smol JP (2022) Biodiversity patterns of Arctic diatom assemblages in lakes and streams: Current reference conditions and historical context for biomonitoring. *Freshwater Biology* 67: 116–140. <https://doi.org/10.1111/fwb.13490>
- Keck F, Vasselon V, Tapolczai K, Rimet F, Bouchez A (2017) Freshwater biomonitoring in the Information Age. *Frontiers in Ecology and the Environment* 15: 266–274. <https://doi.org/10.1002/fee.1490>
- Kelly M, Boonham N, Juggins S, Kille P, Mann D, Pass D, Sapp M, Sato S, Glover R (2018) A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. Bristol, 157 pp.
- Kelly MG, Juggins S, Mann DG, 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 MG, Mann DG, Taylor JD, Juggins S, Walsh K, Pitt J-A, Read DS (2024) Maximising environmental pressure-response relationship signals from diatom-based metabarcoding in rivers. *Science of The Total Environment* 914: 169445. <https://doi.org/10.1016/j.scitotenv.2023.169445>
- Kollár J, Pinseel E, Vanormelingen P, Pouličková A, Souffreau C, Dvůrák P, Vyverman W (2019) A polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). *Journal of Phycology* 55: 365–379. <https://doi.org/10.1111/jpy.12825>
- Kollár J, Pinseel E, Vyverman W, Pouličková A (2021) A time-calibrated multi-gene phylogeny provides insights into the evolution, taxonomy and DNA barcoding of the *Pinnularia gibba* group (Bacillariophyta). *Fottea* 21: 62–72. <https://doi.org/10.5507/fot.2020.017>
- Kollár J, Kopalová K, Kohler JT (2025) Diatom studies three decades into the molecular age: a bibliometric analysis reveals genetic underexploration of diatoms compared to other taxa. *Fottea* 25: 1–11. <https://doi.org/10.5507/fot.2024.009>
- Kollár J, Van de Vijver B, Kulichová J, Hejduková E, Viso R, Kohler TJ, Kahlert M, Pinseel E, Pouličková A, Zimmermann J, Neustupa J, Vyverman W, Kopalová K (2026) A polyphasic taxonomy of the pseudocryptic diatom species complex *Pinnularia acidicola*, including description of new species. *European Journal of Phycology*, 1–25. <https://doi.org/10.1080/09670262.2026.2626682>
- Krammer K (2000) Diatoms of the European Inland Waters and Comparable Habitats. Vol. 1. The genus *Pinnularia*. A.R.G. Gantner Verlag K.G, Ruggell, 703 pp.
- Krammer K, Lange-Bertalot H (2004) Bacillariophyceae. 3. Teil: Centrales, Fragilariaceae, Eunotiaceae. Rev. ed. 2004. With new supplement (= p. 580–599). Reprint 2008. Vol. 2/3, Gustav Fischer Verlag, Stuttgart New York, 599 pp.
- Kulichová J, Fialová M (2016) Correspondence between morphology and ecology: morphological variation of the *Frustulia crassinervia-saxonica* species complex (Bacillariophyta) reflects the ombro-minerotrophic gradient. *Cryptogamie, Algologie* 37: 15–28. <https://doi.org/10.7872/crya/v37.iss1.2016.15>
- Kulikovskiy MS, Kociolek JP, Solak CN, Kuznetsova IV, Glushchenko AM (2023) New and interesting taxa from the diatom genus *Gomphonema* Ehrenberg in shallow, nearshore sites on the eastern coast of Lake Baikal. *Plants* 12: 1835. <https://doi.org/10.3390/plants12091835>
- Lange-Bertalot H (2001) Diatoms of the European Inland Waters and Comparable Habitats. Vol. 2. *Navicula* sensu stricto. 10 Genera Separated from *Navicula* sensu lato. *Frustulia*. A.R.G. Gantner Verlag K.G, Ruggell, 526 pp.
- Lange-Bertalot H, Metzeltin D (1996) Indicators of Oligotrophy. Vol. 2, Koeltz Scientific Books, 390 pp.
- Lange-Bertalot H, Ulrich S (2014) Contributions to the taxonomy of needle-shaped *Fragilaria* and *Ulnaria* species. *Lauterbornia* 78: 1–73.
- Lange-Bertalot H, Bąk M, Witkowski A, Tagliaventi N (2011) Diatoms of the European Inland Waters and Comparable Habitats. Vol. 6. *Eunotia* and some related genera. A.R.G. Gantner Verlag K.G, Ruggell, 747 pp.
- Lindegårdh M, Carstensen J, Drakare S, Johnson R, Nyström Sandman A, Söderpalm A, Wikström SA (2016) Ecological Assessment of Swedish Water Bodies; development, harmonisation and integration of biological indicators. Final report of the research programme WATERS. Deliverable 1.1-4, WATERS report no 2016:10. Havsmiljöinstitutet, Sweden, 169 pp.

- Lindström K (1991) Nutrient requirements of the dinoflagellate *Peridinium gatunense*. *Journal of Phycology* 27: 207–219. <https://doi.org/10.1111/j.0022-3646.1991.00207.x>
- Mann DG, Vanormelingen P (2013) An inordinate fondness? The number, distributions, and origins of diatom species. *Journal of Eukaryotic Microbiology* 60: 414–420. <https://doi.org/10.1111/jeu.12047>
- Mann DG, Sato S, Trobajo R, Vanormelingen P, Souffreau C (2010) DNA barcoding for species identification and discovery in diatoms. *Cryptogamie Algologie* 31: 557–577.
- Masouras A, Karaouzas I, Dimitriou E, Tsirtsis G, Smeti E (2021) Benthic diatoms in river biomonitoring—present and future perspectives within the Water Framework Directive. *Water* 13(4): 478. <https://doi.org/10.3390/w13040478>
- Miljödata-MVM (2025) Swedish University of Agricultural Sciences (SLU), National data host lakes and watercourses, and national data host agricultural land. <http://miljodata.slu.se/mvm/> [accessed 22th of September 2025]
- Mohamad H, Mora D, Skibbe O, Abarca N, Deutschmeyer V, Enke N, Kusber W-H, Zimmermann J, Jahn R (2022) Morphological variability and genetic marker stability of 16 monoclonal pennate diatom strains under medium-term culture. *Diatom Research* 37: 307–328. <https://doi.org/10.1080/0269249X.2022.2141346>
- Mora D, Abarca N, Proft S, Grau JH, 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: 448–464. <https://doi.org/10.1086/704827>
- Morales EA, Siver PA, Trainor FR (2001) Identification of diatoms (Bacillariophyceae) during ecological assessments: comparison between light microscopy and scanning electron microscopy techniques. *Proceedings of the Academy of Natural Sciences of Philadelphia* 151: 95–103, 109. [https://doi.org/10.1635/0097-3157\(2001\)151\[0095:IOBDE\]2.0.CO;2](https://doi.org/10.1635/0097-3157(2001)151[0095:IOBDE]2.0.CO;2)
- Müller A, Güntsch A, Jahn R, Kohlbecker A, Kusber W-H, Zimmermann J (2022) Phyco-Bank: Repository for algal novelties. *Biodiversity Information Science and Standards* 6: e90885. <https://doi.org/10.3897/biss.6.90885>
- Pawlowski J, Kelly-Quinn M, Altermatt F, Apotheloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Feio MJ, Filipe AF, Fornaroli R, Graf W, Herder J, van der Hoorn B, Iwan Jones J, Sagova-Mareckova M, Moritz C, Barquin J, Piggott JJ, Pinna M, Rimet F, Rinkevich B, Sousa-Santos C, Specchia V, Trobajo R, Vasselon V, Vitecek S, Zimmerman J, Weigand A, Leese F, Kahlert M (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of The Total Environment* 637–638: 1295–1310. <https://doi.org/10.1016/j.scitotenv.2018.05.002>
- Pérez-Burillo J, Mann DG, 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 DG (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>
- Pinseel E, Kulichová J, Scharfen V, Urbánková P, Van de Vijver B, Vyverman W (2019) Extensive cryptic diversity in the terrestrial diatom *Pinnularia borealis* (Bacillariophyceae). *Protist* 170: 121–140. <https://doi.org/10.1016/j.protis.2018.10.001>

- Pinseel E, Janssens SB, Verleyen E, Vanormelingen P, Kohler TJ, Biersma EM, Sabbe K, Van de Vijver B, Vyverman W (2020) Global radiation in a rare biosphere soil diatom. *Nature Communications* 11: 2382. <https://doi.org/10.1038/s41467-020-16181-0>
- Reichardt E (1999) Zur Revision der Gattung *Gomphonema*: Die Arten um *G. affine/insigne*, *G. angustatum/micropus*, *G. acuminatum* sowie gomphonemoide Diatomeen aus dem Oberoligozän in Böhmen. Vol. 8, Koeltz Scientific Books.
- Rimet F (2021) A ready-to-use database for DADA2: Diat.barcode_rbcL_263bp_DADA2 based on Diat.barcode v10. <https://doi.org/10.15454/D7FO2T>
- Rimet F, Abarca N, Bouchez A, Kusber W-H, Jahn R, Kahlert M, Keck F, Kelly M, Mann D, Piuze A, Trobajo R, Tapolczai K, Vasselon V, Zimmermann J (2018) The potential of High-Throughput Sequencing (HTS) of natural samples as a source of primary taxonomic information for reference libraries of diatom barcodes. *Fottea* 18: 37–54. <https://doi.org/10.5507/fot.2017.013>
- Rimet F, Gusev E, Kahlert M, Kelly MG, Kulikovskiy M, Maltsev Y, Mann DG, Pfannkuchen M, Trobajo R, Vasselon V, Zimmermann J, Bouchez A (2019) Diat.barcode, an open-access curated barcode library for diatoms. *Scientific Reports* 9: 15116. <https://doi.org/10.1038/s41598-019-51500-6>
- Rimet F, Aylagas E, Borja A, Bouchez A, Canino A, Chauvin C, Chonova T, Ciampor Jr F, Costa FO, Ferrari BJD, Gastineau R, Goulon C, Gugger M, Holzmann M, Jahn R, Kahlert M, Kusber W-H, Laplace-Treytore C, Leese F, Leliaert F, Mann DG, Marchand F, Méléder V, Pawlowski J, Rasconi S, Rivera S, Rougerie R, Schweizer M, Trobajo R, Vasselon V, Vivien R, Weigand A, Witkowski A, Zimmermann J, Ekrem T (2021) Metadata standards and practical guidelines for specimen and DNA curation when building barcode reference libraries for aquatic life. *Metabarcoding and Metagenomics* 5: e58056. <https://doi.org/10.3897/mbmg.5.58056>
- Rose DT, Cox EJ (2014) What constitutes *Gomphonema parvulum*? Long-term culture studies show that some varieties of *G. parvulum* belong with other *Gomphonema* species. *Plant Ecology and Evolution* 147: 366–373. <https://doi.org/10.5091/plecevo.2014.1059>
- Schimani K, Abarca N, Skibbe O, Mohamad H, Jahn R, Kusber W-H, Campana GL, Zimmermann J (2023) Exploring benthic diatom diversity in the West Antarctic Peninsula: insights from a morphological and molecular approach. *Metabarcoding and Metagenomics* 7: e110194. <https://doi.org/10.3897/mbmg.7.110194>
- Simons AL, Theroux S, Osborne M, Nuzhdin S, Mazar R, Steele J (2023) Zeta diversity patterns in metabarcoded lotic algal assemblages as a tool for bioassessment. *Ecological Applications* 33(3): e2812. <https://doi.org/10.1002/eap.2812>
- Smol JP, Stoermer EF (2010) *The Diatoms: Applications for the Environmental and Earth Sciences*. Cambridge University Press, [xviii +] 667 pp. <https://doi.org/10.1017/CBO9780511763175>
- Solak CN, Gastineau R, Lemieux C, Turmel M, Gorecka E, Trobajo R, Rybak M, Yilmaz E, Witkowski A (2021) *Nitzschia anatoliensis* sp. nov., a cryptic diatom species from the highly alkaline Van Lake (Turkey). *PeerJ* 9: e12220. <https://doi.org/10.7717/peerj.12220>
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stosch HAV, Fecher K (1979) Internal thecae of *Eunotia soleirolii* (Bacillariophyceae) - development, structure and function as resting spores. *Journal of Phycology* 15: 233–243. <https://doi.org/10.1111/j.1529-8817.1979.tb02631.x>

- Strayer DL, Dudgeon D (2010) Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society* 29: 344–358. <https://doi.org/10.1899/08-171.1>
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30: 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tapolczai K, Selmečzy GB, Szabó B, B-Béres V, Keck F, Bouchez A, Rimet F, Padisák J (2021) The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. *Ecological Indicators* 122: 107322. <https://doi.org/10.1016/j.ecolind.2020.107322>
- The Belgian Coordinated Collections of Microorganisms (BCCM) consortium (2025) The BCCM/DCG Diatoms Collection <https://bccm.belspo.be/about-DCG> [accessed 6th of November 2025]
- Trobajo R, Mann DG (2019) A rapid cleaning method for diatoms. *Diatom Research* 34: 115–124. <https://doi.org/10.1080/0269249X.2019.1637785>
- Turland N, Wiersema J, Barrie F, Gandhi K, Gravendyck J, Greuter W, Hawksworth D, Herendeen P, Kloppe R, Knapp S, Kusber W-H, Li D-Z, May T, Monro A, Prado J, Price M, Smith G, Zamora JC (2025) International Code of Nomenclature for algae, fungi, and plants (Madrid Code). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226839479.001.0001>
- Urbanková P, Vesela J (2013) DNA-barcoding: A case study in the diatom genus *Frustulia* (Bacillariophyceae). *Nova Hedwigia*: 147–162.
- Urbánková P, Scharfen V, Kulichová J (2016) Molecular and automated identification of the diatom genus *Frustulia* in northern Europe. *Diatom Research* 31: 217–229. <https://doi.org/10.1080/0269249x.2016.1224780>
- Van de Vijver B, Williams DM (2022) Four new *Fragilaria* (Fragilariaceae, Bacillariophyta) species in Europe with coarse striae. *Nova Hedwigia* 115: 317–347. https://doi.org/10.1127/nova_hedwigia/2022/0722
- Van de Vijver B, Schuster TM, Kusber W-H, Hamilton PB, Wetzel CE, Ector L (2021a) Revision of European *Brachysira* species (Brachysiraceae, Bacillariophyta): I. The *Brachysira microcephala* - *B. neoexilis* enigma. *Botany Letters* 168: 467–484. <https://doi.org/10.1080/23818107.2021.1909499>
- Van de Vijver B, Williams MD, Kelly M, Jarlman A, Wetzel EC, Ector L (2021b) Analysis of some species resembling *Fragilaria capucina* (Fragilariaceae, Bacillariophyta). *Fottea* 21: 128–151. <https://doi.org/10.5507/fot.2021.001>
- Van de Vijver B, Schuster TM, Kusber W-H, Kennedy B, Hamilton PB, Albert R-L, Ballings P, Wetzel CE, Ector L (2022a) Revision of European *Brachysira* species (Brachysiraceae, Bacillariophyta): III. Species formerly included in the *Brachysira seriata*-complex. *Botany Letters* 169: 83–105. <https://doi.org/10.1080/23818107.2021.1941250>
- Van de Vijver B, Williams D, Kusber W-H, Cantonati M, Hamilton P, Wetzel C, Ector L (2022b) *Fragilaria radians* (Kützinger) D.M.Williams et Round, the correct name for *F. gracilis* (Fragilariaceae, Bacillariophyta): a critical analysis of this species complex in Europe. *Fottea* 22: 256–291. <https://doi.org/10.5507/fot.2022.006>
- Van de Vijver B, Williams D, Schuster T, Kusber W-H, Cantonati M, Wetzel C, Ector L (2022c) Analysis of the *Fragilaria rumpens* complex (Fragilariaceae, Bacillariophyta) with the description of two new species. *Journal of the Czech Phycological Society* 22: 93–121. <https://doi.org/10.5507/fot.2021.018>

- Van de Vijver B, Ector L, Wetzel C (2023a) Taxonomic diatom workshop *Planothidium*. https://www.nvkd-diatom.nl/_files/ugd/f21ff7_84c46f3e76c24398aff448f25912d98b.pdf [accessed 8th of August 2023]
- Van de Vijver B, Schuster MT, Jónsson SG, Hansen I, Williams MD, Kusber W-H, Wetzel EC, Ector L (2023b) A critical analysis of the *Fragilaria vaucheriae* complex (Bacillariophyta) in Europe. *Fottea* 23: 62–96. <https://doi.org/10.5507/fot.2022.013>
- Vanormelingen P, Chepurinov VA, Mann DG, Sabbe K, Vyverman W (2008) Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of *Eunotia bilunaris sensu lato* (Bacillariophyta). *Protist* 159: 73–90. <https://doi.org/10.1016/j.protis.2007.08.004>
- Vaulot D, Bilcke G, Chaerle P, Falcatore A, Gourvil P, Lomas MW, Probert I, Vyverman W (2023) Culturing Diatoms. In: Goessling JW, Serôdio J, Lavaud J (Eds) *Diatom Photosynthesis: From Primary Production to High Value Molecules*. Wiley-Scrivener. <https://doi.org/10.1002/9781119842156.ch14>
- Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wyler SA, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Dijkstra K-DB, Eisendle U, Freyhof J, Gadawski P, Graf W, Haegerbaeumer A, van der Hoorn BB, Japoshvili B, Keresztes L, Keskin E, Leese F, Macher JN, Mamos T, Paz G, Pešić V, Pfannkuchen DM, Pfannkuchen MA, Price BW, Rinkevich B, Teixeira MAL, Várbíró G, 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>
- 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: 173–192. <https://doi.org/10.1007/s13127-011-0050-6>
- 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: e108793. <https://doi.org/10.1371/journal.pone.0108793>
- Zou S, Bao Y, Wu X, Wang C (2021) DNA barcoding diatoms from China with multiple genes. *Frontiers in Marine Science* 8: 698331. <https://doi.org/10.3389/fmars.2021.698331>

Supplementary material 1

Suppl. figure S1

Authors: Maria Kahlert, Demetrio Mora, Wolf-Henning Kusber, Nélida Abarca, Jonas Zimmermann

Data type: pdf

Explanation note: **fig. S1:** Preliminary phylogeny of the studied strains of freshwater diatoms (Bacillariophyceae) from Sweden (FRESHBAR project), on a ≥990 bp stretch of the *rbcL* barcode. Bootstrap values are given for each node. Scale bar: number of substitutions per site.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.10.186778.suppl1>

Supplementary material 2

Suppl. figure S2

Authors: Maria Kahlert, Demetrio Mora, Wolf-Henning Kusber, Nélida Abarca, Jonas Zimmermann

Data type: pdf

Explanation note: **fig. S2:** Preliminary phylogeny of the studied strains of freshwater diatoms (Bacillariophyceae) from Sweden (FRESHBAR project) added to the available multiple alignment of the entire Diat.barcode reference library (v10), on the 263 bp *rbcL* short barcode stretch. Bootstrap values are given for each node. Scale bar: number of substitutions per site.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.10.186778.suppl2>

Supplementary material 3

Clonal cultures/strains of the FRESHBAR project

Authors: Maria Kahlert, Demetrio Mora, Wolf-Henning Kusber, Nélida Abarca, Jonas Zimmermann

Data type: xlsx

Explanation note: **table S1.** Clonal cultures/strains of the FRESHBAR project: strain name, voucher number, DNA Bank number, taxon name, origin, date, collector, GenBank accession numbers, and additional identification comments.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.10.186778.suppl3>

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Artificial Intelligence (AI) use

The authors accept full responsibility for the content of the manuscript, including the disclosure of any use of AI.

No AI tools were used in the preparation of this manuscript.

Funding

FRESHBAR was funded by the Swedish Environmental Protection Agency in collaboration with the Swedish Agency for Marine and Water Management (Research Grant Dnr. 18/171). The project also received support from the SYNTHESIS+ project <http://www.synthesys.info/>, which is financed by the European Community Research Infrastructure Action under the H2020 Integrating Activities Programme, Project number 823827, and from SLU's environmental monitoring and assessment program "Lakes and Watercourses." The isolation and cultivation work completed by Demetrio Mora was funded by a Short-Term Scientific Mission through the COST Action DNAqua-Net (CA15219).

Author contributions

The following authors made major contributions to (i) the conception or design of the study, (ii) the acquisition, analysis, or interpretation of the data, and (iii) the writing of the manuscript: MK: Conceptualization—Lead; data acquisition, analysis, and interpretation—Lead; writing—original draft—Lead; writing—review and editing—Lead. DM: Data acquisition—Equal; data analysis—Support; writing—review and editing—Equal. WHK: Data acquisition and analysis—Equal; writing—review and editing—Equal. NA: Data acquisition and analysis—Equal; writing—review and editing—Support. JZ: Conceptualization—Support; data acquisition and analysis—Equal; writing—review and editing—Support.

Author ORCIDs

M. Kahlert  <https://orcid.org/0000-0001-9643-4281>
D. Mora  <https://orcid.org/0000-0003-3866-0488>
W.-H. Kusber  <https://orcid.org/0000-0003-4543-5764>
N. Abarca  <https://orcid.org/0000-0001-8897-160X>
J. Zimmermann  <https://orcid.org/0000-0002-0522-0569>

Data availability

All data and material are shared under the FAIR data sharing principles (Findable, Accessible, Interoperable, and Reusable) to enable their use as a resource for conducting research, teaching, and environmental monitoring.

Vouchers (raw material, microscopical slides, and SEM stubs) and data of physical objects are stored in the Herbarium Berolinense (B) and the collection management system for diatoms, "Micro Algae SpecimenTool" (BGBM 2016+), of Herbarium Berolinense, available at https://diatoms.bo.berlin/algae_bgbm/sptool/sptool.php, under the Collection Code = [FRESHBAR].

All strains of the FRESHBAR project are stored under their strain number as an identifier, together with a subset of environmental data (e.g., geo-reference, collector, sampling date, locality name, and identifier, as given in Table 1) of the Swedish gathering. Each strain is linked to the accession number in the Herbarium Berolinense (B). This accession number is unique within the Herbarium Berolinense and is represented by a worldwide unique stable identifier, as agreed by CETAF (Groom et al. 2017). Example: B 40 0046697, found under <https://herbarium.bgbm.org/object/B400046697>.

The DNA is stored freeze-dried in the DNA biobank of the Botanic Garden and Botanical Museum, Freie Universität Berlin, and the associated metadata and identifiers are published in the portal of the Global Genome Biodiversity Network (GGBN 2011+) following their data standard. Available at https://www.ggbn.org/ggbn_portal/; reference numbers: Suppl. material 3: table S1.

Sequence data have been deposited in the NCBI BioProject database under accession number PRJNA1454281; therefore, they are also accessible via other database members of the International Nucleotide Sequence Database Collaboration (INSDC).

Images of living and prepared strains are stored with open access (license: CC0) in the e-archive of the Swedish University of Agricultural Sciences (SLU), available with a stable handle as: Kahlert M, Mora D, Kusber W-H, Abarca N, Zimmermann J (2025) Kahlert images Freshbar. Swedish University of Agricultural Sciences. <https://hdl.handle.net/20.500.12703/6968>

The compiled documentation of the images is published as Kahlert M, Mora D, Kusber W-H, Abarca N, Zimmermann J (2026) Data (images) for: New barcodes of freshwater diatoms from Sweden (Version 1) [Data set]. Swedish University of Agricultural Sciences. <https://researchdata.se/en/catalogue/dataset/2025-373>

The 36 clonal cultures sent to the BCCM/DCG diatom culture collection of the Belgian Co-ordinated Collection of Micro-organisms (<https://bccm.belspo.be/>) are available under the accession numbers DCG 1032, 1035–1045, 1093–1101, and 1211–1225 (Suppl. material 3: table S1).

Environmental data of the sampled sites are given in Table 1. Additional data can be found in or requested from Miljödata-MVM, the national database for lakes and watercourses, at <https://miljodata.slu.se/mvm/>.