J Paleolimnol
DOI 10.1007/s10933-016-9926-y

ORIGINAL PAPER

Sedimentary DNA versus morphology in the analysis of diatom-environment relationships

Katharina Dulias · Kathleen R. Stoof-Leichsenring · Luidmila A. Pestryakova · Ulrike Herzschuh

Received: 19 April 2016 / Accepted: 19 October 2016

Abstract The Arctic treeline ecotone is characterised by a steep vegetation gradient from arctic tundra to northern taiga forests, which is thought to influence the water chemistry of thermokarst lakes in this region. Environmentally sensitive diatoms respond to such ecological changes in terms of variation in diatom diversity and richness, which so far has only been documented by microscopic surveys. We applied next-generation sequencing to analyse the diatom composition of lake sediment DNA extracted from 32 lakes across the treeline in the Khatanga region, Siberia, using a short fragment of the rbcL chloroplast gene as a genetic barcode. We compared diatom richness and diversity obtained from the genetic approach with diatom counts from traditional microscopic analysis. Both datasets were employed to investigate diversity and relationships with environmental variables, using ordination methods. After effective filtering of the raw data, the two methods gave similar results for diatom richness and composition at the genus level (DNA 12 taxa; morphology 19 taxa), even though there was a much higher absolute number of sequences obtained per genetic sample (median 50,278), compared with microscopic counts (median 426). Dissolved organic carbon explained the highest percentage of variance in both datasets (14.2 % DNA; 18.7 % morphology), reflecting the compositional turnover of diatom assemblages along the tundra-taiga transition. Differences between the two approaches are mostly a consequence of the filtering process of genetic data and limitations of genetic references in the database, which restricted the determination of genetically identified sequence types to the genus level. The morphological approach, however, allowed identifications mostly to species level, which permits better ecological interpretation of the diatom data. Nevertheless, because of a rapidly increasing reference database, the genetic approach with sediment DNA will, in the future, enable reliable

Electronic supplementary material The online version of this article (doi:10.1007/s10933-016-9926-y) contains supplementary material, which is available to authorized users.

K. Dulias · K. R. Stoof-Leichsenring (✉) · U. Herzschuh
Periglacial Research, Alfred Wegener Institute Helmholtz
Centre for Polar and Marine Research, Telegrafenberg
43A, 14473 Potsdam, Germany
e-mail: Kathleen.Stoof-Leichsenring@awi.de

L. A. Pestryakova
Department of Geography and Biology, North-Eastern
Federal University of Yakutsk, Belinskogo 58, Yakutsk,
Russia 67700

U. Herzschuh
Institute of Earth and Environmental Science, University
of Potsdam, Karl- Liebknecht-Strasse 24-25,
14476 Potsdam-Golm, Germany

Present Address:
K. Dulias
Department of Biological Sciences, School of Applied
Sciences, University of Huddersfield, Queensgate,
Huddersfield HD1 3DH, UK

Published online: 15 November 2016

investigations of diatom composition from lake sediments that will have potential applications in both paleoecology and environmental monitoring.

Keywords Environmental DNA · Diatoms · Metabarcoding · Lake sediments · Arctic thermokarst lakes · Treeline ecotone

Introduction

In addition to macroscopic and microscopic visible fossils, molecular remains, including DNA, are preserved in lake sediments (Anderson-Carpenter et al. 2011; Thomsen and Willerslev 2015). Sedimentary DNA can be used to characterise recent and past biodiversity and thus is suitable to reveal information about past environmental change (Jørgensen et al. 2012; Parducci et al. 2012; Pedersen et al. 2013). A metabarcoding approach to environmental DNA and subsequent DNA sequencing is commonly applied to characterise the composition of communities in environmental samples (Taberlet et al. 2012a; Pedersen et al. 2015). The obtained DNA sequences are compared with those of known specimens recorded in a database (Taberlet et al. 2012b). If the database is incomplete, algorithms are used to find the best taxonomic assignment for the query sequences within a certain threshold of sequence dissimilarity (Boyer et al. 2016; Rimet et al. 2016). Metabarcoding, combined with new high-throughput, next-generation sequencing (NGS) techniques such as Illumina amplicon sequencing, potentially provides a sufficiently high number of sequences compared to cloning and Sanger sequencing, to reveal the community composition adequately (Shokralla et al. 2012; Taberlet et al. 2012a). Other advantages of NGS-based metabarcoding are the potential for automated sample handling and standardized laboratory protocols that allow for better comparison between different studies (Hajibabaei et al. 2011; Thomsen and Willerslev 2015).

Diatoms are one of the most commonly used organism groups for environmental reconstructions, mainly because their silicified frustules are well preserved in lake sediments. They are very diverse and because of their sensitivity, diatom assemblages track a variety of environmental variables (Battarbee et al. 2001). The Arctic is known for its particularly strong reaction to global warming over a variety of timescales (ACIA 2004), with boreal treeline areas often shown to be subject to environmental transition (MacDonald et al. 2008). Siberia possesses the largest forest-tundra ecotone belt in the world (Frost and Epstein 2014) and vegetation in the southern Taymyr Peninsula has been found to reflect climate change on millennial, centennial and decadal timescales (Klemm et al. 2015; Niemeyer et al. 2015). Tundra-to-forest transition in Siberia, in space and time, is reflected in the physical and chemical characteristics of water in the numerous thermokarst lakes, and thus affects diatom assemblages (Duff et al. 1999; Laing and Smol 2000; Rühland et al. 2003; Herzschuh et al. 2013). Most studies in Siberia that used diatoms as environmental indicators relied solely on morphological analysis of diatom frustules (Laing et al. 1999b; Cherapanova et al. 2007; Biskaborn et al. 2012; Pestryakova et al. 2012). The studies by Stoof-Leichsenring et al. (2014, 2015) were the first studies that connected morphological and genetic analysis of Siberian diatoms. The majority of diatom species found in Siberia are very small because of the harsh conditions, and hence it is very difficult to identify them by their morphology (Biskaborn et al. 2012). Therefore, it was thought that a genetic approach, such as metabarcoding, might more easily reveal greater diatom diversity, including cryptic lineages. A particularly suitable diatom gene to study for environmental applications is the large subunit of the ribulose-1,5-bisphosphate-carboxylase/oxygenase (rbcL) gene (Stoof-Leichsenring et al. 2014). This gene can be used not only to differentiate species, but because it is very variable within specific groups of diatoms, it provides information about putative intraspecific variation (Medlin et al. 2012; Evans et al. 2007). For example, Stoof-Leichsenring et al. (2015) showed a spatial and temporal change in *Staurosira* lineages along the Arctic treeline and associated these changes with the surrounding vegetation and/or related environmental variables.

We analysed sub-fossil sedimentary diatom assemblages from thermokarst lakes in the Siberian Arctic that represent a latitudinal transect from the northern taiga to the southern tundra on the southern part of the Taymyr Peninsula (Krasnoyarsk District, Russia). This study combined a microscopic approach with metabarcoding of sedimentary diatom assemblages. Both the morphological and the genetic approaches aim to identify the diatom composition of a lake community, which can then be related to environmental variables. Our study addressed the following three questions. (1) To what extent are morphologically and genetically identified diatom assemblages similar with respect to richness and composition? (2) Are the morphologically and genetically identified diatom assemblages correlated with the same environmental variables? (3) What are the implications of using ancient sedimentary DNA of diatoms as a proxy for past environmental change?

Materials and methods

Sampling and collection of environmental data

Thermokarst lakes of the Siberian lowlands are typically small, shallow and oligotrophic (Herzschuh et al. 2013). Mean annual, July, and January temperatures are about −12 °C, 13 °C, and −32 °C, respectively (Khatanga meteorological station 71.98°N, 102.47°E, 1929–2010, https://www.ncdc.noaa.gov/cdo-web/datasets). Mean annual precipitation is ~250 mm, about half of it falling between June and August. In summer 2013, lacustrine surface sediments and water samples were collected from 32 lakes located along a latitudinal transect that crosses the boreal treeline (Fig. 1). Water and sediment samples were transported to the Alfred Wegener Institute (AWI), Potsdam, Germany and stored in the dark at about 8 °C until further processing. Surface sediments (13-TY-01 to 13-TY-32) were sampled with an Ekman-Birge bottom sampler. To minimize cross-contamination between the lakes, sampling equipment was washed before use and sampling was performed on different days. For most lakes, the uppermost centimetre of surface sediment was sampled in sterile, 150-ml plastic bottles using new, sterile plastic spoons for each sample. For some lakes (13-TY-03, -09, -15, -19, -26, -27, -29), however, only bulk samples that integrate the uppermost 5 cm of surface sediment were sampled in sterile Whirl–pak® bags. Water samples were collected from 0.5 m below the water surface. Dissolved organic carbon (DOC) and contents of Cl^-, $SO4^{2-}$, Ca^{2+}, K^+, Mg^{2+}, Na^+ and SiO_2 were analysed using a Shimadzu TOC-VCPH (DOC), a DIONEX-DX320 (anions) and a Perkin Elmer Optima 8300DV (cations) in the AWI laboratory. Morphological (lake depth, lake area), physical (Secchi depth) and chemical (pH, conductivity) variables were recorded for each site. Morphological, physical and chemical variables of the sampled lakes are summarized in Electronic Supplementary Material (ESM) Table S1.

Genetic diatom approach

Extractions were performed using the PowerMax®-Soil DNA Isolation Kit (MoBio Laboratories, California). Samples were processed in an isolation laboratory using a dedicated UV-Hood for extraction of environmental DNA. The risk of cross-contamination was reduced by handling each bottle separately and by cleaning the working area with DNA-ExitusPlus™ between each sample preparation. An extraction blank, used as a control for chemical contamination, was included for each extraction batch. A sterile spatula was used to transfer about 5 g of sediment into the Bead tube, containing bead solution, C1 buffer and 400 µl of 20 mg ml^{-1} proteinase K (VWR International). Solutions were vortexed for 2 min at maximum speed and placed over night at 56 °C in a shaking incubator and processed the next day. Further extraction steps were carried out according to the manufacturer's instruction.

PCR amplifications were conducted using diatom-specific primers that amplify a short DNA fragment (76 bp without primer sequence) of the rbcL gene (Stoof-Leichsenring et al. 2012). We double-checked the specificity of the primers to diatom sequences against the complete EMBL Nucleotide Sequence Database (release embl_127, August 2016), using the EcoPCR package (Ficetola et al. 2010) as explained in Stoof-Leichsenring et al. (2012). According to multiplexing purposes for subsequent Illumina amplicon sequencing, forward and reverse primers were modified on the 5′ end with unique 8 bp tags that differed from each other in at least 5 base pairs following Binladen et al. (2007), and were preceded by a random primer suffix NNN to improve cluster generation on the sequencer (Coissac 2012; De Barba et al. 2014). Unique primer combinations were used for all sediment samples as well as for the extraction blank and PCR negative control (NTC) amplifications. The following reagents were added to the PCR reactions: Primers (forward: 5′ NNN(8 bp tag)AACAGGTGAAGTTAAAGGTTCATAYTT 3′,

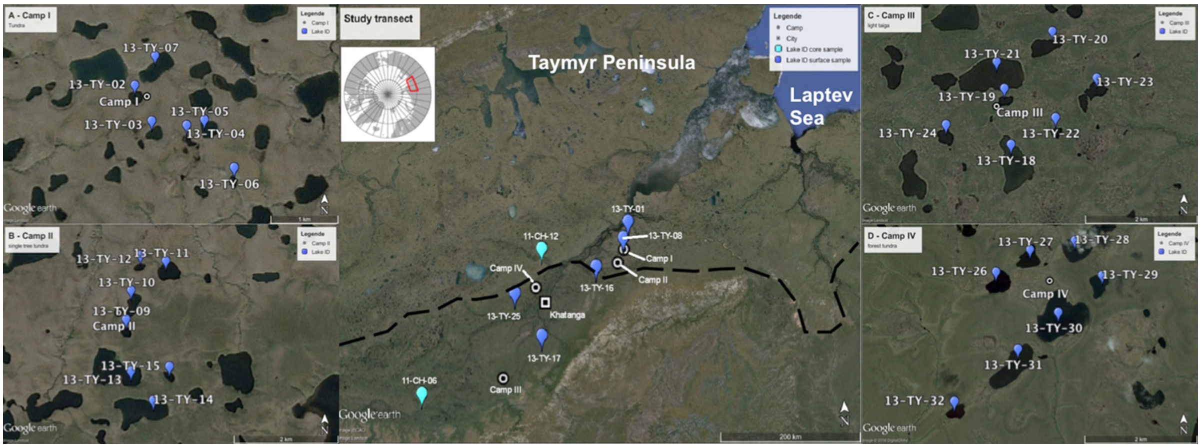


Fig. 1 Maps showing the locations of the sampled lakes. **a** camp I: tundra, **b** camp II: single tree tundra, **c** camp III: light taiga, and **d** camp IV: forest tundra. The *black dashed line* indicates the current position of the treeline. Google Earth Image © 2016 DigitalGlobe

reverse: 5′ NNN(8 bp tag)TGTAACCCATAACT AAATCGATCAT 3′) each at a final concentration of 0.4 µM, 0.1 mM dNTPS, 2 mM MgSO4 (Invitrogen/Life Corp., Carlsbad, CA, USA), 0.8 µg BSA (VWR International), 10× Platinum® Taq DNA Polymerase High Fidelity PCR buffer (Invitrogen/Life Corp., Carlsbad, CA, USA), 1.25 U Platinum® Taq DNA Polymerase High Fidelity (Invitrogen/Life Corp., Carlsbad, CA, USA) and 2 µl of DNA template solution. The PCR set-up was conducted under a dedicated UV hood only used for PCR set-ups. Subsequently, PCR reactions were transferred to a Biometra thermo cycler (Jena Analytik, Germany), which is located in a Post-PCR laboratory, physically separated from the DNA Isolation and PCR set-up laboratory. The following reaction profile was run: initial denaturation for 5 min at 94 °C, 50 cycles at 94, 49 and 68 °C each for 30 s, and a final elongation at 72 °C for 5 min. Replications for each amplification were performed using identical primer tag combinations and PCR conditions. PCR replications were performed on different days.

Prior to amplicon sequencing PCR replications, including amplification of samples, extraction blanks and PCR NTCs (no template control) were pooled and subsequently purified using a MinElute PCR Purification Kit (Qiagen, Hilden, Germany), following the supplied protocol. DNA of the pooled and purified PCR products was prepared for DNA quantification using the Qubit® dsDNA BR Assay Kit (Invitrogen/Life Corp., Carlsbad, CA, USA). Each sample solution was measured twice with the Qubit® 2.0 Fluorometer (Invitrogen/Life Corp., Carlsbad, CA, USA) and the volume of each sample for equimolar pooling of all samples calculated. A maximum of 10 µl of a PCR product was used if the concentration of DNA was too low, which was only the case for extraction blanks and PCR NTCs. Library preparation and parallel high-throughput paired-end (2 × 125 bp) amplicon sequencing were performed on the Illumina HiSeq 2000 platform (Illumina Inc.), facilitated by the Fasteris SA sequencing service (Switzerland).

The paired-end sequencing run resulted in two data files containing the forward and the reverse reads, which were analysed using OBITools, a package providing several single programs for DNA sequence analysis (available at http://metabarcoding.org/obitools). The OBITools pipeline consists of the following bioinformatic steps, also explained in more detail in Boyer et al. (2016). Our raw data, consisting of two single data files, were first merged to a single file using the algorithm illuminapairedend. Subsequently, the obigrep command discards reads that were not merged and the ngsfilter algorithm then assigns each sequence to the corresponding primer tag combination. Non-assigned sequences, having a different primer tag combination, are saved in an additional file and can be checked for the percentage of erroneously tagged sequences. Furthermore, the program obiclean was used to filter the assigned sequences from PCR and sequencing errors, as it classifies the sequences into head, internal and singleton based on the count and sequence similarity within one sample. Finally, the validated sequence types are taxonomically assigned using ecotag. This algorithm assigns the query sequence to a taxon based on sequence similarity by using a group-specific reference library, which was created using the program ecoPCR, which extracts artificially targeted sequences using an in silico-PCR (Ficetola et al. 2015) with the specific rbcL diatom primers (see above) from the EMBL database (release 124, July 2015). Ecotag first searches for the primary and secondary reference(s) that show highest similarity to the query sequence and then assigns it (them) to the most recent common ancestor of the primary and secondary reference sequence(s). This process is highly dependent on the depth of the reference database, which is also crucial for the taxonomic assignation. To increase the number of sequences in the diatom-specific library, five mismatches between primer and artificially targeted sequences of the EMBL entries were allowed. After using the OBITools pipeline the final results were entered into a spreadsheet and were further filtered manually. First, all sequences with less <96 % match to entries in the diatom reference library were removed to guarantee good taxonomic assignation, which we based on the average of intra-genus variation in the rbcL gene (Kermarrec et al. 2014). Second, we subtracted the diatom sequence counts found in the extraction blanks or NTCs from the relevant samples belonging to each extraction and/or PCR batch. Then we removed all sequences that did not have the exact length of 76 bp to eliminate erroneous sequence types, because there is no length variation in the rbcL marker. We then removed all non-diatom sequences. All sequence types with a count lower than 1000 were ignored because we wanted to include only sequence types that are highly abundant in the dataset and that account for at least 0.03 % of the total data. Only sequences identified to at least genus level were kept for further analyses and grouped according to their taxonomic assignment. Finally, we checked our sequences for the typical error base sequence GGC in Illumina amplicon sequencing (Nakamura et al. 2011), but this sequence did not occur as an error in our sequences. As the morphology-based approach always led to at least a genus level identification, the two approaches are thus comparable on the analysed taxonomic level. Sequence types that occurred at a minimum of 0.5 % and in at least three samples were kept in the dataset for further statistical analyses.

For detailed qualitative comparison of both approaches we used all sequence types, including those with a count less than 1000, which were identified to genus level and all morphological data: this dataset, however, was not used for statistical analyses. The final, genetically identified sequence types and their abundances and taxonomic identity are deposited in PANGEA (https://doi.pangaea.de/10.1594/PANGAEA.867329).

Diatom morphology approach

Calcareous and organic components of the sediment samples (approximately 0.3 g per sample) were removed by heating with hydrochloric acid (10 %) and hydrogen peroxide (30 %). Cleaned samples were mounted on microscope slides using Naphrax®. Valves were counted (500 valves per sample) using a Zeiss microscope at 100× magnification. Diatom determination was conducted using the literature cited in Pestryakova et al. (2012). Each identified species or genus was clustered into morphologically identified taxa, grouping species from the same genera into one taxonomic group.

Statistical analyses

The richness of the samples in each dataset was investigated using rarefaction analysis (Oksanen et al. 2015), which estimates the potential species richness for a given number of subsamples. The number of subsamples (i.e. counts in the morphologic approach and sequence reads in the genetic approach) is defined by the minimum sample size in the dataset used. To increase the signal-to-noise ratio, only abundant species were kept in the dataset. Genera that occurred at a minimum of 0.5 % and in at least three samples were kept in for further statistical analyses. All environmental variables (maximum depth, Secchi depth, DOC, conductivity, HCO_3–, Cl^-, $SO4^{2-}$, Ca^{2+}, K^+, Mg^{2+}, Na^+ and SiO_2), except pH, were log-transformed. Spearman's rank correlation test was used to assess the relationship among environmental variables to identify groups of highly correlated variables. Groups of highly correlated variables were depth (maximum and Secchi depth), conductivity

(HCO_3–, conductivity, Ca^{2+} and Mg^{2+}), and DOC and ion content ($SO4^{2-}$, DOC and SiO_2). The final morphological and genetic datasets were Hellinger-transformed prior to applying ordination methods (Legendre and Gallagher 2001). An initial detrended correspondence analysis (DCA) suggested that ordination methods that assume linear relations between diatoms and environmental variables in our datasets are appropriate (ter Braak and Šmilauer 2002). Accordingly, principal component analysis was used to portray the main structure in the datasets and redundancy analysis (RDA) was used to investigate relations between environmental variables and diatom composition. Finally, the following sets of RDAs were run for both the morphologically and the genetically derived datasets. Those variables in each group that when considered alone in an initial RDA explained the highest variance in the diatom data, and had a statistically significant relationship with the diatom dataset ($p < 0.03$ for the genetic approach and $p < 0.005$ for the morphological approach), were included in variance partitioning to extract the uniquely explained variation for each single environmental variable. Explained variations are given as adjusted r^2 values.

Statistical analyses were carried out in the "vegan" package (Oksanen et al. 2015) in R version 3.1.2 (R Core Team, 2014). The stratigraphic plots were implemented in C2 version 1.7.6 (Juggins 2007). CorelDRAW® Graphics Suite ×6 version (16.4.0.1280) was used to modify and merge graphics from R outputs.

Results

Genetic-based diatom richness, composition and diatom-environment relationships

The primer specificity analysis indicates high specificity of the primers to diatom sequences, as 90.4 % (total: 2032 entries; diatom: 1838 entries) of the in silico amplified sequences belong to diatoms. The genetic raw data counted 4,667,815 reads, consisting of 5141 sequence types (ESM Table S2). Sequence reads identified in extraction blanks and PCR negative controls were always <0.03 % within one batch, with one exception in which we identified a diatom sequence type with 17,738 reads (0.5 %) that does not occur in the samples of the extraction batch (ESM Table S3). This phenomenon might be explained by tag-switching (Carlsen et al. 2012). After stepwise filtering (ESM Table S2) of the raw data, until only diatom sequence types that were identified to genus level remained, results of genetically and morphologically identified genera (Fig. 2; ESM Table S4) were compared. This comparison included rare sequence types (<1000 counts) and indicated that 50 % (32 genera) of the morphologically identified genera were also detected by the genetic approach, whereas 50 % (32 genera) were solely identified by microscopic analysis. Ten genera (24 % of the genetically identified genera) were only retrieved by the genetic approach.

After further filtering the genetic data of diatom genus-level sequences that have a count higher than 1000, the genetic data comprised 154 sequence types (total number of reads: 3,190,116), of which 126 sequence types (total number of reads: 2,990,056) were assigned to genus or lower taxonomic level (Fig. 3). Thirty genera (33 sequence types), 27 species (84 sequence types) and 4 subspecies (9 sequence types) were recognized, of which 12 genera (107 sequence types, consisting of taxa solely identified to genus and lower taxonomic levels) and 13 species (70 sequence types) passed all thresholds (correction for rare taxa) and were used for further statistical analysis. Because of the low number of sequence counts (147), sample 13-TY-25 was excluded from further analysis. Despite the equimolar pooling of PCR products for multiplex sequencing (i.e. samples are tagged and pooled with respect to DNA quantity in each sample, which allows the sequencing of different samples in one NGS run), the total number of sequence reads varied substantially among samples, i.e. between 11,337 and 491,273, with a median of 50,278. Rarefaction results of genera (ESM Fig. S1) suggest that for only seven of 31 samples was a sufficient count total reached that would likely retrieve all occurring genera. The rarefaction curve of the genera shows that the richness of samples probably varies between three and ten genera based on the minimum sample size of 11,337. With respect to total counts, *Staurosira*/*Staurosirella*, *Sellaphora* and *Pinnularia* are most often recorded (Fig. 4). All diatom assemblages were dominated by *Staurosira*/*Staurosirella*. The first and second principal component axes explain 27 % and 22 %, respectively, of the variance in the genus-level, genetic-based diatom assemblages. Tundra sites,

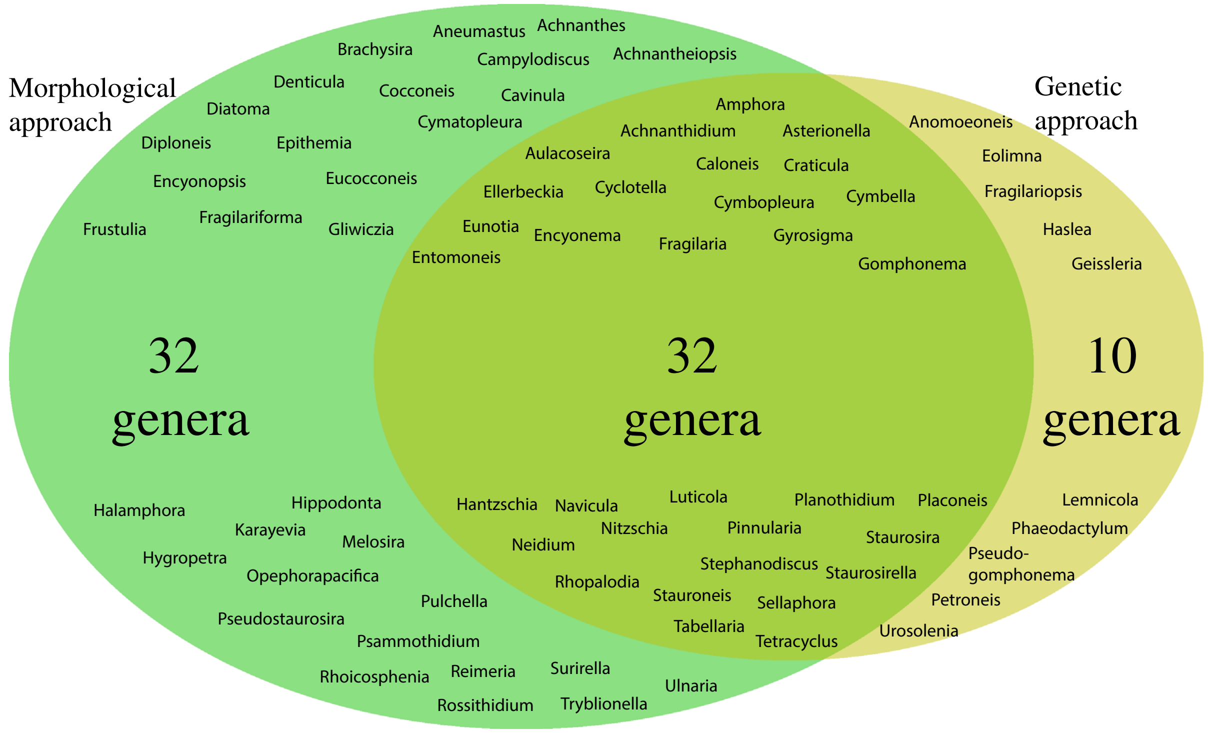


Fig. 2 Venn diagram of identified genera retrieved by the morphological (64 genera) and the genetic (42 genera) approach and shared genera (32 genera). The data also includes rare genera with less than 1000 counts

typically characterised by low DOC and high conductivity, form a cluster in the lower part of the biplot (Fig. 5a), whereas forest lakes are concentrated in the upper part of the plot. When included as sole variables, DOC, conductivity, Ca^{2+}, pH, and maximum depth (Table 1) explain a significant portion ($p < 0.05$) of the variance in the genus-level, genetic-based diatom assemblages. Of these, DOC, pH and maximum depth also uniquely explain a significant portion, with DOC explaining the dataset best. RDA results for the genetically retrieved species data are presented in ESM Table S5.

Morphology-based diatom richness, composition and diatom-environment relationships

The morphological approach identified 225 taxa (total number of counted valves: 17,036), with all taxa identified to genus or lower taxonomic level (Fig. 3). Nineteen genera, 196 species and 10 subspecies were identified, of which 19 genera (including 43 morphologically identified taxa) and 29 species passed all thresholds and were used for further analyses. No valves could be extracted from sample 13-TY-29, thus this sample was excluded from all further analyses. The total number of valves counted per sample varied from 236 to 566, with a median of 426 valves. Rarefaction results of genera (ESM Fig. S2) suggest that for all 31 samples a sufficient count total was reached capable of retrieving all occurring genera. The rarefaction curve of the genera shows that the sample richness varied between nine and 15 genera (assuming a total of 236 counted valves). With respect to taxa counts, *Staurosira*, *Staurosirella* and *Pseudostaurosira* were most often recorded (Fig. 6). All diatom assemblages were dominated by *Staurosira*/*Staurosirella*. The first and second principal component axes explain 26 % and 18 %, respectively, of the variance in the genus-level morphology-based diatom assemblages. Tundra sites form a cluster in the lower left part of the biplot (Fig. 5b), whereas forest lakes are mostly concentrated in the upper right part of the biplot. When included as sole variables, DOC, conductivity, maximum depth, Ca^{2+} and silica (Table 2) explain a significant portion ($p < 0.005$) of the variance in the genus-level morphology-based

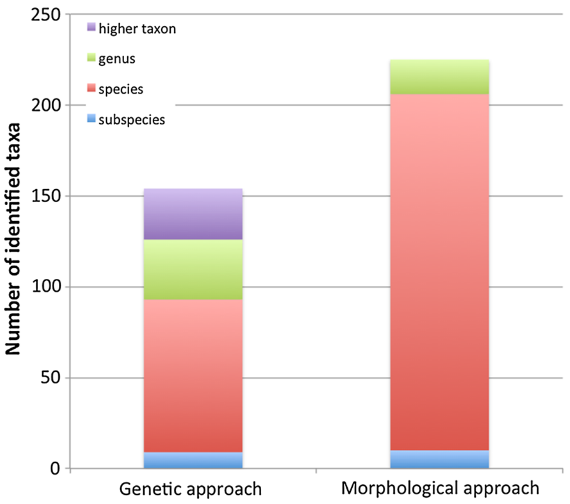


Fig. 3 Number of diatom taxa identified with the genetic and morphology approaches, after effective filtering steps (e.g. 96 % sequence similarity at least to EMBL database entries, extraction blanks/PCR NTC, exact length of 76 bp, and more than 1000 counts per sequence type). Genetically identified taxa include 4 subspecies (9 sequence types), 27 species (84 sequence types), 30 genera (33 sequence types) and 23 sequence types identified to higher taxonomic levels. The morphological approach identifies 10 subspecies, 169 species and 19 genera. The taxonomic assignment of all taxa identified is indicated by the colour code

diatom assemblages. Of those variables, only DOC explains a unique portion. RDA results for the morphologically retrieved species data are presented in ESM Table S6.

Discussion

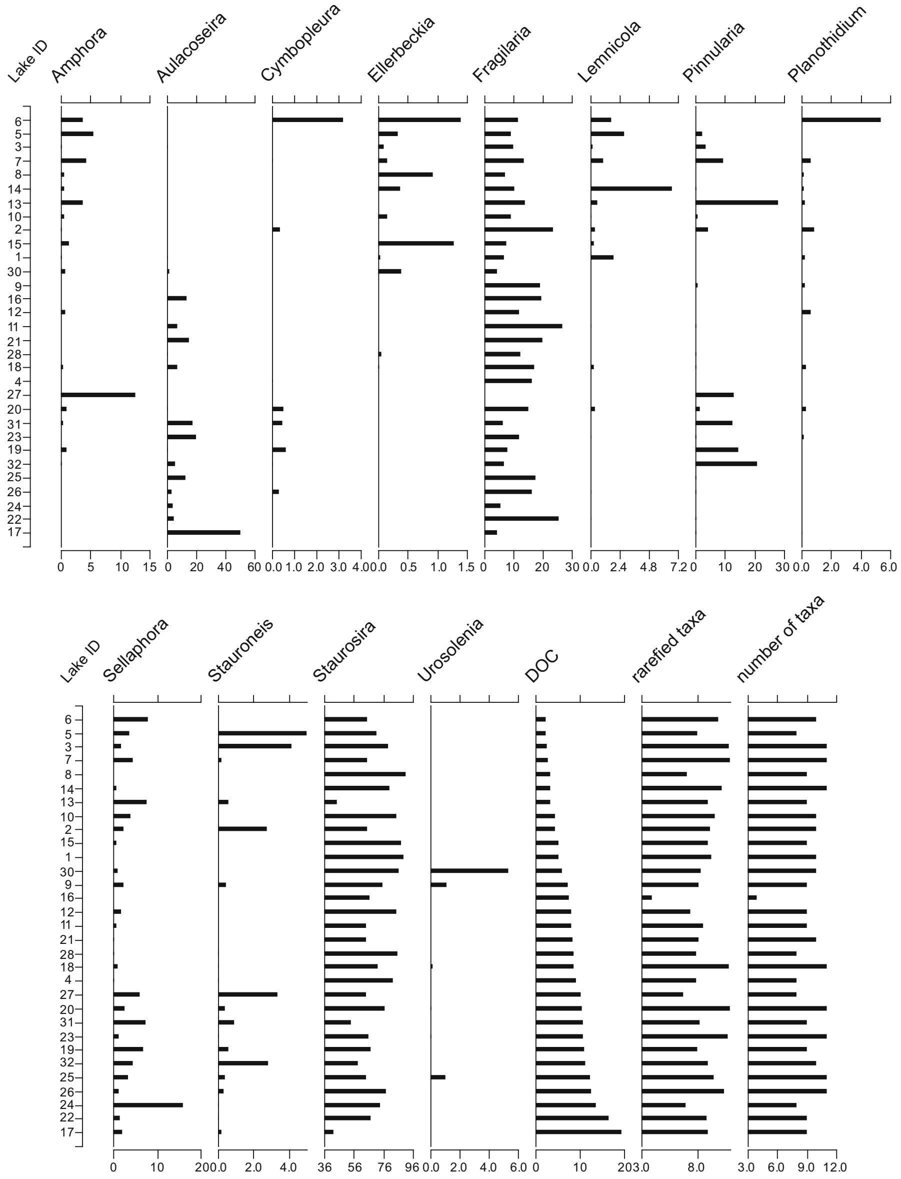
Comparison of the genetic and morphological approaches in recording diatom assemblages

Across the entire dataset, and including rare sequence types (Table S4), 33 genera were detected with each of the approaches and account for about 99 % of the sequence reads and 44 % of the microscopic counts, respectively. Overall, the genetic and morphological diatom approaches (final datasets) yield fairly similar results in the overall number of recorded taxa (154 vs. 225), with strong similarity between the genus-level diatom compositions of the individual sediment samples. This result is achieved despite the large discrepancy between the absolute number of counts: the genetic approach is based on almost three million sequence reads (after diverse filter steps), whereas only about 17,000 valves were counted from the 31 samples. The minimum adequate sample size needed to represent the entire taxonomic diversity differs significantly between the two methods. A possible explanation for this discrepancy might be the incompleteness of the reference database, in contrast to the extremely high number of counts in the genetic data. Nevertheless, the number of morphologically identified species is high, even though the sample sizes are much smaller, albeit more consistent. The filtering of sequence data reduced the final number of taxa substantially, and deviations from single filter steps produce different results. For example, if sequences with a lower sequence count than 1000 were not removed, 1886 taxa will be retrieved, exceeding the number of microscopically identified taxa. Despite the number of identified taxa, the relative abundance of selected genera between the two approaches is in remarkable agreement, e.g. with respect to the dominance of *Staurosira*-like taxa or the relative abundance of *Amphora, Aulacoseira* and *Sellaphora* in nearly all samples.

Reasons for qualitative and quantitative differences between the two approaches fall into two main areas: (1) the fossil recording processes (taphonomy) (Flower 1993) of DNA and diatom valves, and (2) differences in the diversity retrieved from the record. With respect to taphonomy, differences originate mainly from the number of available units of analysis in a single diatom (i.e. two valves vs. chloroplast-DNA copies). The number of chloroplast DNA copies results mainly from the number of chloroplasts per cell and the number of chloroplast genomes per chloroplast, which vary across diatom taxa (Bedoshvili et al. 2009). Other aspects regarding taphonomy are the preservation potential of DNA (Epp et al. 2015; Pedersen et al. 2015) versus diatom valves (Ryves et al. 2006) and the different representations of DNA versus valves in the various lake habitats, for example caused by different transport characteristics and adsorption of DNA to sediments (Torti et al. 2015; Turner et al. 2015).

In our study, one genus retrieved with the genetic-based approach, but not found in the morphological survey was *Urosolenia*. This genus belongs to the centric diatoms. Frustules are long, thin and slightly silicified. They are known to be fragile and the standard preparation for morphological analysis

Fig. 4 Percentages of diatom sequence types identified to genus level ordered according to DOC (mg L^{-1}) content of the lakes. Rarefied taxa and number of taxa indicate the retrieved genetic diversity



typically destroys the frustules (Rott et al. 2006). Therefore, this genus will only be retrieved using the molecular approach, whereas it is lost from the morphological record. Another example is *Pinnularia*. Although this genus is recorded by both methods, it is strongly underrepresented with the morphological approach compared to the genetic approach, which may stem from mechanical destruction of the long frustules in the sediment. On the other hand, genera identified by means of light microscopy, but that are only rarely found using the genetic-based approach, are *Cyclotella* and *Tabellaria*, diatom taxa that inhabit the water column. It is probable that DNA from planktonic diatom genera is underrepresented in recent sediments relative to DNA of bottom-dwelling taxa. This is because of the well-preserved DNA of living benthic diatom cells, such as *Staurosira* and *Pinnularia* types, in contrast to the more decayed DNA from dead planktonic cells.

The fact that valves of different taxa differ in their preservation potential may impact DNA degradation, but the extent of DNA retrieval during DNA

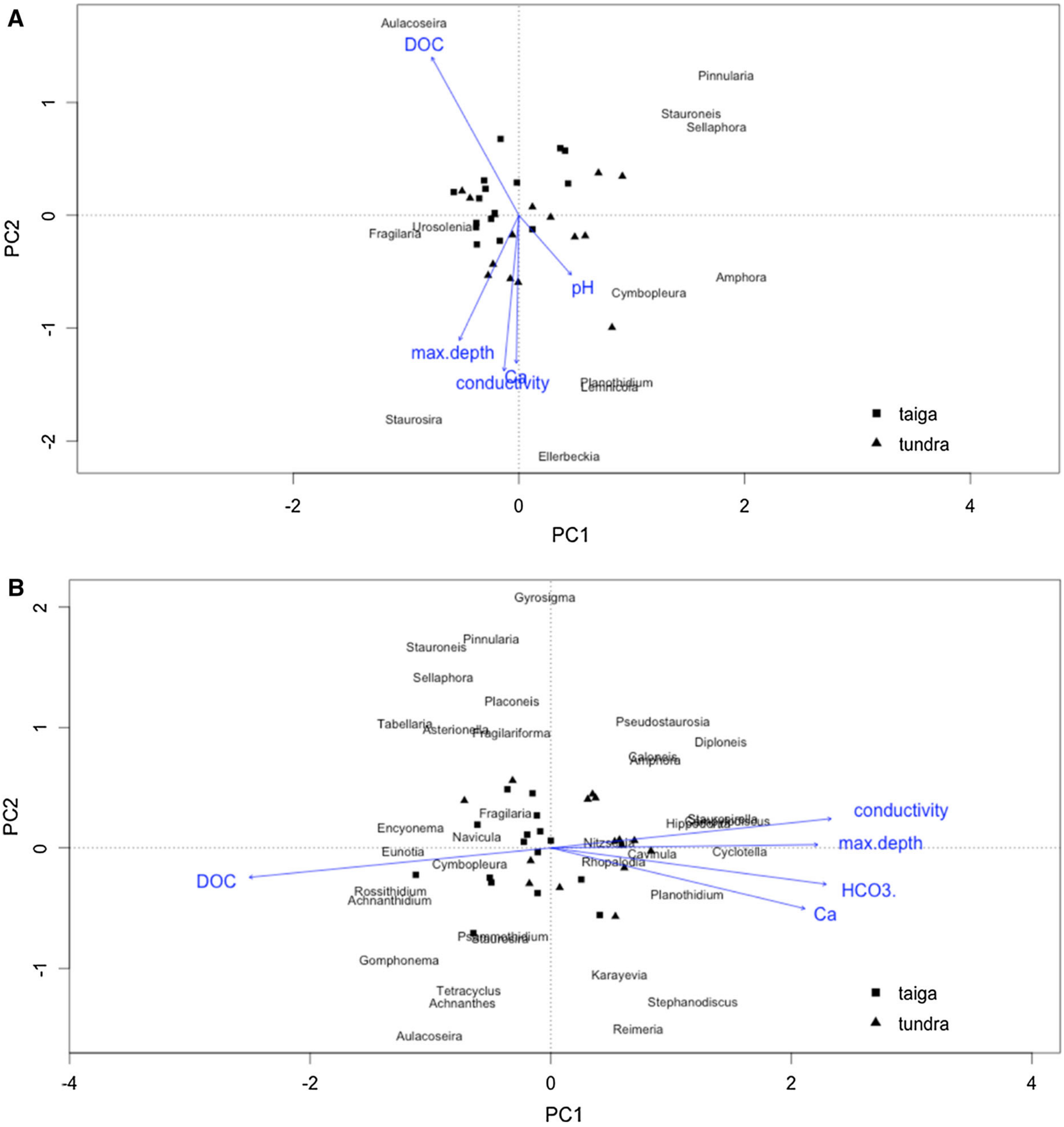


Fig. 5 Principal Component Analysis of the genetically retrieved genera (**a**) and the morphologically retrieved genera (**b**)

extraction, and the success of PCR amplification (primer specificity) may be of greater relevance with respect to whether a taxon is recorded by the sedimentary DNA approach or not. With regard to record retrieval, the specificity of the marker versus morphological characteristics leads to differences in the records of the two applied methods, as do the availability and quality of references in a genetic database and the knowledge of the diatom analyst. Although we used a diatom-specific primer pair (90.4 % of in silico amplified sequences are diatoms), differences in the composition and quantity of retrieved taxa are still likely. Those differences are produced by variations in the specificity of the primer-binding region in taxa not present in the reference database, the efficiency of PCR amplifications and

Table 1 Morphological and chemical variables as a proportion of the genetic sequence data for genera

	Single proportion		Unique proportion	
	Adjusted R^2	*P* value	Adjusted R^2	*P* value
Morphological and chemical variables				
Max. depth	0.052	0.030	0.040	0.040
pH	0.101	0.005	0.045	0.040
Conductivity	0.124	0.010	0.027	0.090
Ca^{2+}	0.068	0.025	0.022	0.110
DOC	0.142	0.005	0.107	0.005

	Proportion	*P* value				
	Adjusted R^2	Max. depth	pH	Conductivity	Ca^{2+}	DOC
All significant variables	0.3186023	0.005	0.005	0.04	0.12	0.005

dilution effects during library preparation (Pawluczyk et al. 2015; Schirmer et al. 2015). The diversities gained through metabarcoding studies are mostly reliant on the genetic marker, but so far there is no consensus on an appropriate diatom barcode, although several markers have been tested and used (Hamsher et al. 2011; MacGillivary and Kaczmarska 2011). Generally, markers were developed for different genomic regions, i.e. ribosomal, mitochondrial and chloroplast markers (e.g. 18S rRNA, cox1 and rbcL) (Evans et al. 2007; Medinger et al. 2010; Theriot et al. 2010; Hamsher et al. 2011; Zimmermann et al. 2011; Epp et al. 2012; Kermarrec et al. 2013; Stoof-Leichsenring et al. 2014). In our study, a short and variable sequence of 76 bp of the chloroplast rbcL gene (rbcL_76 amplicon) was used because of its specificity to the amplification of diatom DNA from sedimentary records, and especially ancient sedimentary DNA. In previous studies, this marker was used successfully on sediments from different habitats, such as fresh and saline African lakes (Stoof-Leichsenring et al. 2012), freshwater lakes in Siberia (Stoof-Leichsenring et al. 2014, 2015) and a high arctic lake in Greenland (Epp et al. 2015). Compared to ribosomal or mitochondrial markers, the specificity of the rbcL chloroplast marker to photosynthetic organisms strongly reduces the amplification of non-specific products, such as DNA from heterotrophic bacteria and fungi, which dominate sedimentary DNA samples.

The genetic and the morphological datasets largely agree as to which is the dominant taxon—*Staurosira* (genetic approach) and *Staurosira*-like diatoms (morphological approach)—in most samples, which corroborates earlier findings about the composition of Siberian diatom communities (Rühland 2001; Biskaborn et al. 2012; Pestryakova et al. 2012) and those from permafrost regions in North America and northern Europe (Weckström et al. 1997; Laing and Smol 2000; Rühland et al. 2003). However, the genetic approach assigned *Fragilaria construens* and *Staurosira elliptica* as the taxa, whereas the morphological species identified were mostly *Staurosira venter*, *Staurosirella pinnata*, *Pseudostaurosira pseudoconstruens* and *P. brevistriata*. Of course, the assignment of DNA sequences to a taxonomic unit is dependent on the correctness of the related reference descriptions in the database and the completeness of the available references. That only two species were identified by the genetic approach, despite there being 37 different *Staurosira*-like sequences identified genetically and five *Staurosira*-like diatoms in the morphologically derived taxa list, originates primarily from the extremely limited number of references included in the EMBL database. The large diversity of *Staurosira*-like diatoms was already reported by Stoof-Leichsenring et al. (2014, 2015) with respect to changes in haplotype appearances across tundra–northern taiga transects, and is corroborated by the results of our study. The taxonomy of *Staurosira*-like diatom taxa, as revealed by morphological analyses is still in revision (Williams and Round 1987; Paull et al. 2008; Stoof-Leichsenring et al. 2014). For example, one confusing reference description is *S. elliptica* (accession number HQ828193) and its likely synonym *S. pinnata*, used in morphological analyses. Accordingly, the morphological characteristics used to

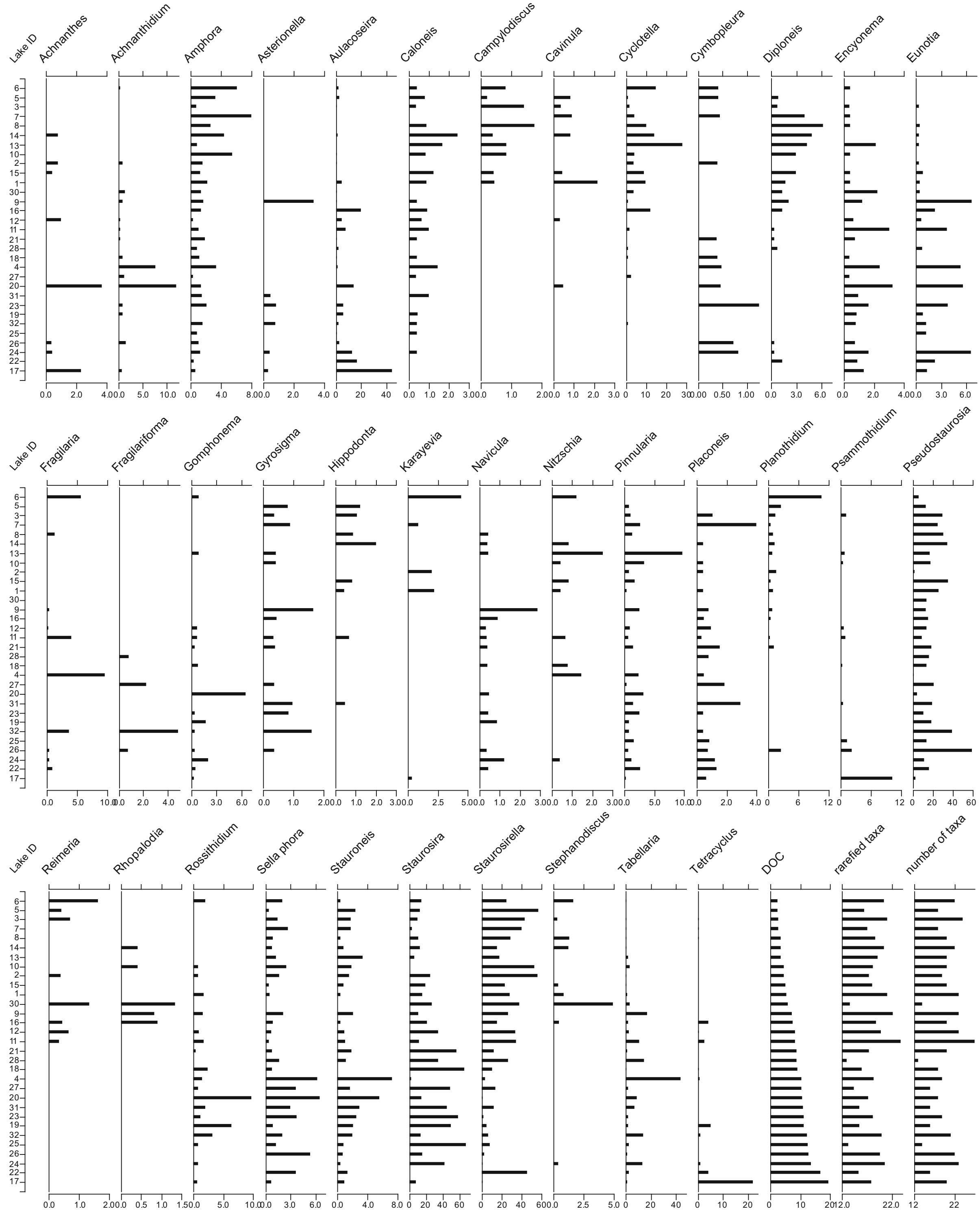


Fig. 6 Morphologically identified diatom genera ordered according to DOC (mg L^{-1}) content of the lakes. Rarefied taxa and number of taxa indicate the retrieved morphological diversity

Table 2 Morphological and chemical variables as a proportion of the morphological data for genera

	Single proportion		Unique proportion	
	Adjusted R^2	*P* value	Adjusted R^2	*P* value
Morphological and chemical variables				
Max. depth	0.124	0.005	0.018	0.115
Conductivity	0.162	0.005	0.000	0.430
Ca^{2+}	0.109	0.005	−0.015	0.850
SiO_2	0.061	0.005	0.018	0.170
DOC	0.187	0.005	0.039	0.025

	P value					
	Adjusted R^2	Max. depth	Conductivity	Ca^{2+}	DOC	SiO_2
All significant variables	0.2345334	0.005	0.005	0.355	0.035	0.26

identify the respective reference material in the genetic database are very ambiguous or not provided at all. Our results suggest that with respect to *Staurosira*-like diatoms, the genetic approach can record the diatom composition (i.e. the composition of related sequence types identified by unique genetic code) more reliably than the classical morphological approach.

Relationships between genetically and morphologically derived diatom compositions and environmental variables

Diatom compositions in Siberia are known to reflect a variety of environmental variables, but are particularly sensitive to lakewater chemical characteristics (Biskaborn et al. 2012; Pestryakova et al. 2012; Herzschuh et al. 2013). Both the genetically and morphologically derived diatom compositions are best explained by DOC, and the percentage of explained variance is rather similar (14 and 19 %, respectively).

DOC is known to represent a key driver of diatom communities in circumpolar, treeline lakes (Laing and Smol 2000; Rühland et al. 2003) because DOC concentration limits the depth of the photic zone and alters the light composition at greater depths (Duff et al. 1999). Generally, an increase in DOC is observed from tundra to taiga lakes across the treeline ecotone (Pienitz and Smol 1993; Duff et al. 1999; Laing et al. 1999a).

Similar to the study of a large transect of lakes in northern Yakutsk by Pestryakova et al. (2012), water depth was identified as an environmental driver of diatom composition of subordinate importance. In shallow lakes, lake depth partially exerts an impact similar to DOC, in that it regulates light availability at the lake bottom, whereas in deeper lakes it regulates the extent and timing of stratification (Boehrer and Schultze 2008). Furthermore, shallow lakes have a better nutrient supply, because they warm easily and do not stratify (Pestryakova et al. 2012), but diatom communities in shallow lakes need to withstand regular disturbance when the lake freezes to the bottom in winter (Battarbee et al. 2001).

Our diatom dataset did not yield a strong unique relationship with alkalinity/conductivity, as observed by Herzschuh et al. (2013), but our geographic and alkalinity/conductivity gradients are much shorter. In Herzschuh et al. (2013), a decreasing trend from tundra in the northern Siberian lowlands to typical taiga in central Yakutia was observed (total gradient for alkalinity 16–281 mg L^{-1} and conductivity 39–666 µS cm^{-1}). In contrast, our dataset had an alkalinity range of 14.2–69.3 mg L^{-1} and conductivity range of 33–180 µS cm^{-1}. The higher ion content in tundra lakes, located in the northern part of the transect, compared with lakes further south, is probably related to their proximity to the Arctic Ocean and the impact of sea spray, as suggested by a study from the Pechora River, Taymyr region, and Lena River region (Laing and Smol 2000).

Conclusions

Our analysis revealed that, with respect to the total number of identified taxa and genus-level diatom assemblage composition, the genetically and morphologically retrieved diatom assemblages from surface sediments of thermokarst lakes in northern Siberia are very similar. Nevertheless, ecological interpretation of diatom spectra based on known habitat preferences of certain species is, at present, still more reliable when based on morphologically retrieved diatom assemblages than when based on the genetic approach. In the future, with an increase in the availability of references in genetic databases, it will be possible to identify much broader diversity at the species level. The genetic approach is already complementary to the morphological approach in consistently identifying small benthic diatoms such as *Staurosira* or fragile diatoms such as *Urosolenia*.

We found that both the morphologically and genetically retrieved diatom assemblages are best explained by DOC concentrations in the lakes, which further emphasizes that both methods capture similar diatom signals. Genetically retrieved diatom assemblages can therefore be used in a similar way, i.e. using similar multivariate numerical techniques, to morphologically retrieved assemblages, to identify past and ongoing environmental changes, and to create transfer functions after calibrating diatom-DNA data with modern datasets. Our results suggest that the applied genetic approach may be useful for inferring recent and past lake diatom compositions. This offers the opportunity to use large metabarcoding approaches in addition to, or even instead of, the classical cell-counting (morphological) method. Furthermore, the genetic approach provides an opportunity to analyse diatom spectra according to standardized protocols.

Acknowledgments We thank Cathy Jenks for English editing of the manuscript. This study was funded by the Helmholtz Association Strategy Fund, and by the Ministry for Education and Science of the Russian Federation (project 5.184.2014/K).

References

ACIA (2004) Arctic climate impact assessment—scientific report. Cambridge University Press, Cambridge

Anderson-Carpenter LL, McLachlan JS, Jackson ST, Kuch M, Lumibao CY, Poinar HN (2011) Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. BMC Evol Biol 11:30

Battarbee RW, Jones VJ, Flower RJ, Cameron NG, Bennion H, Carvalho L, Juggins S (2001) Diatoms. In: Smol JP, Birks HJB, Last WM (eds) Tracking environmental change using lake sediments, vol 3. Terrestrial, algal and siliceous indicators. Kluwer Academic Publisher, New York, pp 155–202

Bedoshvili YD, Popkova TP, Likhoshway YV (2009) Chloroplast structure of diatoms of different classes. Cell and Tissue Biol 3:297–310

Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE 2:e197

Biskaborn BK, Herzschuh U, Bolshiyanov D, Savelieva L, Diekmann B (2012) Environmental variability in northeastern Siberia during the last ~13,300 yr inferred from lake diatoms and sediment-geochemical parameters. Palaeogeogr Palaeoclimatol Palaeoecol 329:22–36

Boehrer B, Schultze M (2008) Stratification of lakes. Rev Geophys 46:1–27

Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2016) OBITools: a Unix-inspired software package for DNA metabarcoding. Mol Ecol Res 16:176–182

Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T, Kauserud H (2012) Don't make a mista(g)ke: Is tag switching an overlooked source of error in amplicon pyrosequencing studies? Fungal Ecol 5:747–749

Cherapanova MV, Snyder JA, Brigham-Grette J (2007) Diatom stratigraphy of the last 250 ka at Lake El'gygytgyn, northeast Siberia. J Paleolimnol 37:155–162

Coissac E (2012) OligoTag: a program for designing sets of tags for next-generation sequencing of multiplexed samples. In: Pompanon F, Bonin A (eds) Data production and analysis in population genomics. Humana Press, New York, pp 13–31

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol Ecol Res 14:306–323

Duff KE, Laing TE, Smol JP, Lean DRS (1999) Limnological characteristics of lakes located across arctic treeline in northern Russia. Hydrobiologia 391:205–222

Epp LS, Boessenkool S, Bellemain EP, Haile J, Esposito A, Riaz T, Erséus C, Gusarov VI, Edwards ME, Johnson A, Stenøien HK, Hassel K, Kauserud H, Yoccoz NG, Bråthen KA, Willerslev E, Taberlet P, Coissac E, Brochmann C (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. Mol Ecol 21:1821–1833

Epp LS, Gussarova G, Boessenkool S, Olsen J, Haile J, Schrøder-Nielsen A, Ludikova A, Hassel K, Stenøien HK, Funder S, Willerslev E, Kjær K, Brochmann C (2015) Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. Quat Sci Rev 117:152–163

Evans KM, Wortley AH, Mann DG (2007) An assessment of potential diatom "Barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). Protist 158:349–364

Ficetola GF, Coissac E, Zundel S, Rias T, Shehzad W, Bessière J, Taberlet P, Pompanon F (2010) An in silico approach for the evaluation of DNA barcodes. BMC Genet 11:434

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Rayé G, Taberlet P (2015) Replication levels, false presences, and the estimation of presence/absence from eDNA metabarcoding data. Mol Ecol Res 15:543–556

Flower RJ (1993) Diatom preservation: experiments and observations on dissolution and breakage in modern and fossil material. Hydrobiologia 269:473–484

Frost GV, Epstein HE (2014) Tall shrub and tree expansion in Siberian tundra ecotones since the 1960s. Glob Change Biol 20:1264–1277

Hajibabaei M, Shokralla S, Zhou X, Singer GA, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6:e17497

Hamsher SE, Evans KM, Mann DG, Poulíčková A, Saunders GW (2011) Barcoding diatoms: exploring alternatives to COI-5P. Protist 162:405–422

Herzschuh U, Pestryakova LA, Savelieva LA, Heinecke L, Böhmer T, Biskaborn BK, Andreev A, Ramisch A, Shinneman ALC, Birks HJB (2013) Siberian larch forests and the ion content of thaw lakes form a geochemically functional entity. Nat Commun 4:2408

Jørgensen T, Haile J, Möller P, Andreev A, Boessenkool S, Rassmussen M, Kinast F, Coissac E, Taberlet P, Brochmann C, Biegelow NH, Andersen K, Orlando L, Gilbert MT, Willerslev E (2012) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Mol Ecol 21:1989–2003

Juggins S (2007) C2 Version 1.5 user guide. Software for ecological and palaeoecological data analysis and visualisation, Newcastle University, Newcastle upon Tyne, UK

Kermarrec L, Franc A, Rimet F, Chaumeil P, Humbert JF, Bouchez A (2013) Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. Mol Ecol Res 13:607–619

Kermarrec L, Franc A, Rimet F, Chaumeil P, Frigerio JM, Humbert JF, Bouchez A (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshw Sci 33:349–363

Klemm J, Herzschuh U, Pestryakova LA (2015) Vegetation, climate and lake changes over the last 7000 years at the boreal treeline in north-central Siberia. Quat Sci Rev. doi:10.1016/j.quascirev.2015.08.015

Laing TE, Smol JP (2000) Factors influencing diatom distributions in circumpolar treeline. J Phycol 36:1035–1048

Laing TE, Pienitz R, Smol JP (1999a) Freshwater diatom assemblages from 23 lakes located near Norilsk, Siberia: a comparison with assemblages from other circumpolar treeline regions. Diatom Res 14:285–305

Laing TE, Rühland KM, Smol JP (1999b) Past environmental and climatic changes related to tree-line shifts inferred from fossil diatoms from a lake near the Lena River Delta, Siberia. Holocene 9:547–557

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280

MacDonald GM, Kremenetski KV, Beilman DW (2008) Climate change and the northern Russian treeline zone. Philos Trans R Soc Lond B Biol Sci 363:2285–2299

MacGillivary ML, Kaczmarska I (2011) Survey of the efficacy of a short fragment of the *rbc*L gene as a supplemental DNA barcode for diatoms. J Eukaryot Microbiol 58: 529–536

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwälder B, Schlötterer C, Boenigk J (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol Ecol 19:32–40

Medlin L, Yang I, Sato S (2012) Evolution of the diatoms. VII. Four gene phylogeny assesses the validity of selected araphid genera. Nova Hedwig 141:505–513

Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-Ul-Amin M, Ogasawara N, Kanaya S (2011) Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res 39:e90

Niemeyer B, Herzschuh U, Pestryakova LA (2015) Vegetation and lake changes on the southern Taymyr peninsula, northern Siberia, during the last 300 years inferred from pollen and Pediastrum green algae records. Holocene 25:596–606. doi:10.1177/0959683614565954

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2015) vegan: community ecology package

Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, Haile J, Kjaer KH, Larsen NK, Drouzas AD, Willerslev E (2012) Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. Mol Ecol 22:3511–3524

Paull TM, Hamilton PB, Gajweski K, LeBlanc M (2008) Numerical analysis of small Arctic diatoms (Bacillariophyceae) representing the *Staurosira* and *Staurosirella* species complexes. Phycologia 47:213–224

Pawluczyk M, Weiss J, Links MG, Egana Aranguren M, Wilkinson MD, Egea-Cortines M (2015) Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. Anal Bioanal Chem 407:1841–1848

Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, Holm J, Funder S, Willerslev E, Kjaer KH (2013) A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quat Sci Rev 75:161–168

Pedersen MW, Overballe-Petersen S, Ermini L, Der Sarkissian C, Haile J, Hellstrom M, Spens J, Thomsen PF, Bohmann K, Cappellini E, Bærholm Schnell I, Wales NA, Carøe C, Campos PF, Schmidt AMZ, Gilbert MTP, Hansen AJ, Orlando L, Willerslev E (2015) Ancient and modern environmental DNA. Philos Trans R Soc Lond B Biol Sci 370:20130383

Pestryakova LA, Herzschuh U, Wetterich S, Ulrich M (2012) Present-day variability and Holocene dynamics of permafrost-affected lakes in central Yakutia (Eastern Siberia) inferred from diatom records. Quat Sci Rev 51:56–70

Pienitz R, Smol JP (1993) Diatom assemblages and their relationship to environmental variables in lakes from the boreal forest-tundra ecotone near Yellowknife, Northwest Territories, Canada. Hydrobiologia 269–270:391–404

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

Rott E, Kling H, McGregor G (2006) Studies on the diatom *Urosolenia* round & crawford (Rhizosoleniophycideae) Part 1. New and re-classified species from subtropical and tropical freshwaters. Diatom Res 21:105–124

Rühland KM (2001) Diatom assemblage shifts relative to changes in environmental and climatic conditions in the circumpolar treeline regions of the Canadian and Siberian Arctic. Queen's University, Kingston

Rühland KM, Smol JP, Pienitz R (2003) Ecology and spatial distributions of surface-sediment diatoms from 77 lakes in the subarctic Canadian treeline region. Can J Bot 81:57–73

Ryves DB, Batterbee RW, Juggings S, Fritz SH, Anderson NJ (2006) Physical and chemical predictors of diatom dissolution in freshwater and saline lake sediments in North America and West Greenland. Limnol Oceanogr 51:1355–1368

Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res 43:e37

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21:1794–1805

Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedemann R (2012) Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. Mol Ecol 21:1918–1930

Stoof-Leichsenring KR, Bernhardt N, Pestryakova LA, Epp LS, Herzschuh U, Tiedemann R (2014) A combined paleolimnological/genetic analysis of diatoms reveals divergent evolutionary lineages of *Staurosira* and *Staurosirella* (Bacillariophyta) in Siberian lake sediments along a latitudinal transect. J Paleolimnol 52:77–93

Stoof-Leichsenring KR, Herzschuh U, Pestryakova LA, Klemm J, Epp LS, Tiedemann R (2015) Genetic data from algae sedimentary DNA reflect the influence of environment over geography. Sci Rep 5:12924

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012a) Environmental DNA. Mol Ecol 21:1789–1793

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012b) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21:2045–2050

ter Braak CJF, Šmilauer P (2002) CANOCO reference manual and CanodDraw for windows user's guide: software for canonical community ordination (version 4.5.), Microcomputer Power, New York

Theriot EC, Ashworth M, Ruck E, Nakov T, Jansen RK (2010) A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research. Plant Ecol Evol 143:278–296

Thomsen PF, Willerslev E (2015) Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. Biol Cons 183:4–18

Torti A, Lever MA, Jorgensen BB (2015) Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Mar Genomics 24:185–196

Turner CR, Uy KL, Everhart RC (2015) Fish environmental DNA is more concentrated in aquatic sediments than surface water. Biol Cons 183:93–102

Weckström J, Korhola A, Blom T (1997) Diatoms as quantitative indicators of pH and water temperature in subarctic Fennoscandian lakes. Hydrobiologia 347:171–184

Williams DM, Round FE (1987) Revision of the genus *Fragilaria*. Diatom Res 2:267–288

Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. Org Divers Evol 11:173–192