Hydrobiologia
DOI 10.1007/s10750-017-3381-2

PRIMARY RESEARCH PAPER

Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment

S. F. Rivera · V. Vasselon · S. Jacquet · A. Bouchez · D. Ariztegui · F. Rimet

Received: 21 May 2017 / Revised: 7 September 2017 / Accepted: 16 September 2017

Abstract Benthic diatoms are relevant indicators of the ecological status of the littoral zone of lakes. Their use as bio-indicators is based on their morphological identification at species level using microscopy which is time consuming, requires taxonomic expertise, and is consequently expensive. To overcome these limitations, a molecular approach for diatom identification has been tested with success in rivers. DNA metabarcoding enables species identification from a standardized DNA barcode and high-throughput sequencing (HTS), using DNA reference library. The suitability of the morphological and molecular approaches to assess the diatom community structure and the ecological status of the littoral zone of the largest deep lake in France (Lake Bourget) was compared. 66 sites were sampled in August 2015 along the shoreline, all around the lake. The composition of diatom assemblages was similar with both morphological and molecular approaches, and diatom assemblages were structured by the same environmental factors. However, the ecological status of Lake Bourget differed significantly among approaches since floristic inventories to species level also differed significantly. The main source of this difference was the incompleteness of the DNA reference library. Nevertheless, in a near future, when this constraint will be solved, the use of DNA metabarcoding for biomonitoring purposes seems promising.

Keywords Algae · Benthic biomonitoring · Eutrophication · High-throughput sequencing · Lake Bourget · Pollution

Electronic supplementary material The online version of this article (doi:10.1007/s10750-017-3381-2) contains supplementary material, which is available to authorized users.

Handling editor: Judit Padisák

S. F. Rivera (✉)
Institute for Environmental Sciences, University of Geneva, 66 Boulevard Carl-Vogt, 1205 Geneva, Switzerland
e-mail: sinzianaflorina@gmail.com

V. Vasselon · S. Jacquet · A. Bouchez · F. Rimet
CARRTEL, INRA, Université de Savoie Mont Blanc, 75 Av. de Corzent, 74200 Thonon les bains Cedex, France

D. Ariztegui
Section of Earth & Environmental Sciences, University of Geneva, Rue des Maraîchers 13, 1205 Geneva, Switzerland

Introduction

Among the biological indicators required for monitoring the ecological status of lakes, phytoplankton has been used for decades as a proxy of the trophic state (Vadeboncoeur et al., 2002; Brucet et al., 2013). It has already been proved that pelagic phytoplankton provides a representative measure of the ecological status of lakes and adequate information for authorities

Published online: 10 October 2017

concerning their preservation and management (Birk et al., 2012; Jacquet et al., 2014). However, at present there is an increasing interest in the use of benthic algae as ecological indicators (Cantonati & Lowe, 2014) since they constitute important components of the littoral zone of lakes (Dokulil, 2003) and are able to provide an early warning system of the anthropogenic pressures along their shorelines which cannot be detected through pelagic indicators (Bielczyńska, 2015; Rimet et al., 2016a, b). Among benthic algae, diatoms—siliceous, unicellular algae of the phylum Bacillariophyta—are the major constituents of phytobenthos in terms of biomass (Stevenson, 1998) and specific diversity (Mann & Vanormelingen, 2013). Diatoms are widespread distributed in almost all aquatic habitats (Potapova & Charles, 2003) and are good indicators of the ecological status of aquatic ecosystems because of their short generation time (Round et al., 1990) and their sensitivity to nutrient content and other physical and chemical parameters (pH, conductivity, temperature, dissolved oxygen, etc.) (Stevenson & Pan, 1999; Bere & Tundisi, 2010). The use of benthic diatoms as river quality indicators really started in the 1950s (Rimet, 2012) and several biological indices based on the ecology of benthic diatoms and their abundance have been developed for this purpose (Rimet et al., 2005). Prior to 2007, the assessment of lake pollution through benthic diatoms was mainly carried out with diatom indices originally developed for rivers (Blanco et al., 2004; Bolla et al., 2010; Cellamare et al., 2012; Rimet et al., 2016a, b). From that year, several diatom indices were specifically developed for lakes (Hofmann, 1994; Ács, 2007; Sgro et al., 2007; Stenger-Kovács et al., 2007; Marchetto et al., 2013; Stevenson et al., 2013; Bennion et al., 2014) as well as a special protocol for sampling (King et al., 2006). Since 2000, benthic diatoms are required by legislation (typical the Water Frame European Directive) to assess lake water quality.

Although diatoms are reliable indicators of water quality, their use in monitoring programs requires unambiguous identification at species level which relies on the morphological characteristics of their silica cell walls. This is time consuming and requires specialized taxonomic knowledge, especially when dealing with closely related taxa (Kahlert et al., 2009; Kermarrec et al., 2014). To overcome these problems, an alternative approach has been developed in recent years for the study of environmental samples. Metabarcoding (Taberlet et al., 2012) uses molecular techniques at community level by combining DNA barcoding (Hebert et al., 2003) with high-throughput sequencing (HTS). DNA barcoding allows an accurate identification of diatom taxa to species level from a short standardized DNA fragment while HTS allows the sequencing of millions of DNA fragments from many samples simultaneously. Sequencing data is then used to obtain an accurate identification of diatom taxa to species level by comparison to a DNA reference library at a higher throughput than with the morphological approach. Several studies have already shown the potential of metabarcoding for the assessment of the ecological status of rivers using benthic diatom communities (Kermarrec et al., 2014; Visco et al., 2015; Zimmermann et al., 2015; Pawlowski et al., 2016), but to the best of our knowledge, no studies have been carried out for this purpose in lakes yet.

The aim of this study was to test the hypothesis that morphological and molecular approaches using littoral benthic diatoms can provide comparable results in the assessment of the ecological status of lakes. In that goal, 66 samples were collected in August 2015 along the entire shoreline of Lake Bourget (France) to determine the ecological status along its littoral zone based on benthic diatoms. Samples were analyzed using both morphological and molecular approaches and results were compared to answer the following questions: (i) Do molecular and morphological methods provide comparable results regarding the structure and composition of benthic diatom assemblages? (ii) Do the environmental factors affecting the structure of diatom assemblages are the same for both approaches? (iii) Do diatom indices calculated on the basis of morphological and molecular data provide the same ecological evaluation of Lake Bourget shoreline? (iv) If the assessment results are different, which reasons can explain these differences?

Materials and methods

Study area

Lake Bourget is located in the eastern part of France (45°44′N, 05°51′W), on the edge of the Alps (Fig. 1). Situated at an altitude of 231 m a.s.l., it has a surface of

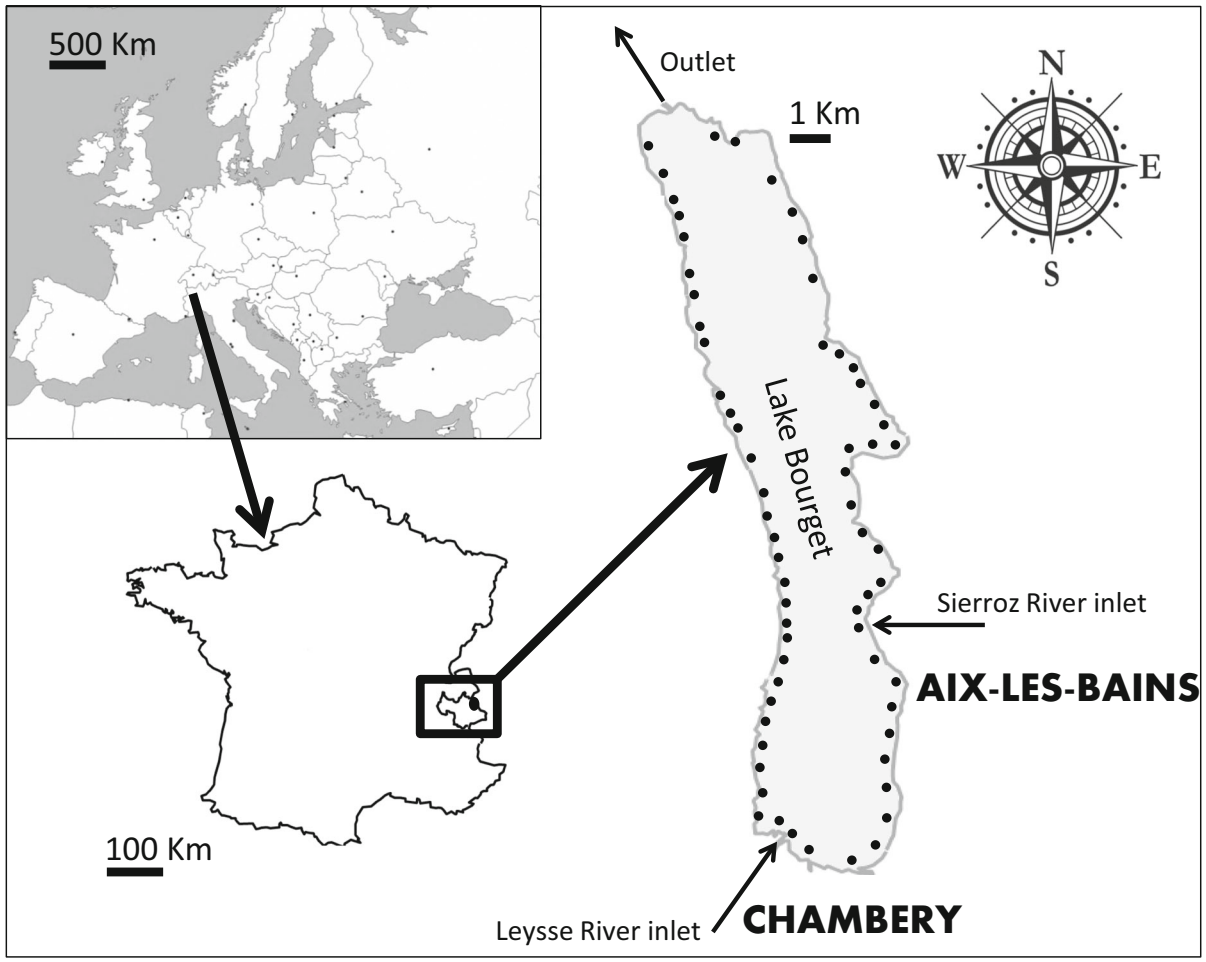

Fig. 1 Location of Lake Bourget in France. Black dots in lake outline depict sampling stations

45 km^2, a mean depth of 80 m and it is the largest natural (deep) lake in France. The littoral area of the lake is characterized by a marked spatial variability. The south eastern coast is highly urbanized in contrast to the western coast which is scarcely urbanized, rugged and forested (Balvay et al., 2012). The main tributaries of the lake, i.e., Leysse and Sierroz rivers, cross the cities of Chambéry (south, 58,000 inhabitants) and Aix-les-Bains (east, 30,000 inhabitants), and are responsible for about 80% of the water inflow being an important source of nutrients to the lake (Bryhn et al., 2010; Meunier & Jacquet, 2015; Jacquet et al., 2016).

Lake Bourget suffered eutrophication during the last century until restoration programs launched in the 1970s allowed its recovery (Jacquet et al., 2012, 2014). It is now considered as oligo-mesotrophic and is in a good ecological status. However, its ecological status is solely based on pluriannual bi-monthly sampling of a single station located at the deepest point of the lake, 1.5 km away from each bank, and therefore restricted to its pelagic zone. Information about the quality of its shoreline is scarce.

Sampling

66 samples of benthic littoral biofilms were collected along the entire shoreline of Lake Bourget in August 2015 (Fig. 1). Sampling stations were located according to their accessibility at a distance of approximately 800 m between each other. At each station, five stones situated at 60–70 cm depth were collected randomly following the recommendations of King et al. (2006).

The epilithon of the upper surface of the stones was then scrapped off and removed using a tooth brush following the European standard EN 13946 (Afnor, 2003). The brushed area of the stones (at least 100 cm^2) and the toothbrush were cleaned on a plastic tray containing 90% ethanol. Finally, the suspension was collected in a tube and fixed with a minimum of 70% ethanol to ensure a good fixation of the samples (CEN, 2015) compatible with DNA extraction. Fixed samples were then taken to the laboratory for morphological analysis and DNA extraction. At each sampling station, three environmental factors were evaluated: land use, typology of the coast and underwater slope.

Morphological analysis

Samples were cleaned with 40% H_2O_2 and HCl according to the European standard EN 13946 (Afnor, 2003). After repeated rinsing and decantation with distilled water, air-dried aliquots were mounted on permanent glass slides using Naphrax®. At least 400 valves were identified and counted under the light microscope at a magnification of 1000X using a Zeiss Axio Imager A1 ® microscope. Identification to species level was done based on European floras such as Krammer & Lange-Bertalot (1986, 1988, 1991), Reichardt (1997), Lange-Bertalot (2001), and Hofmann et al. (2011) and according to the European standard EN 14407 (Afnor, 2004). A list of the taxa and their relative abundances was produced for each of the samples.

Molecular analysis

DNA contained in the samples was isolated using the GenElute™-LPA (Sigma-Aldrich) method according to Chonova et al. (2016). The gene marker *rbc*L (312 base pairs fragment) was amplified by PCR using an equimolar mix of the three diatom-specific primers Diat_*rbc*L_708F_1 (5′-AGGTGAAGTAAAAGGTTCWTACTTAAA-3′) (Bruder & Medlin, 2007), Diat_rbcL_708F_2 (5′-AGGTGAAGTTAAAGGTTCWTAYTTAAA-3′) and Diat_rbcL_708F_3 (5′-AGGTGAAACTAAAGGTTCWTACTTAAA-3′) as forward primers, combined with an equimolar mix of the two primers Diat_rbcL_R3_1 (5′-CCTTCTAATTTACCWACWACTG-3′) (Stoof-Leichsenring et al., 2012), and Diat_rbcL_R3_2 (5′-CCTTCTAATTTACCWACAACAG-3′) as reverse primers. PCR reaction was performed in a thermal cycler with a 25 µL reaction mixture containing 1 µL of extracted DNA, 0.75 U of TaKaRa LA Taq® polymerase (TaKaRa Bio, Sugatsu, Japan), 2.5 µL of 10X Buffer, 1.25 µL of 10 µM of each primer, 1.25 µL of 10 g L^{-1} bovine serum albumin (BSA), 2 µL of 2.5 mM deoxynucleotide (dNTP) and 15.6 µL of H_2O (molecular biology grade). PCR conditions were: initial denaturation of DNA at 95°C for 15 min followed by 30 cycles of denaturation at 95°C for 45 s, annealing at 55°C for 45 s and extension at 72°C for 45 s. After PCR, the amplification of the *rbc*L barcode was confirmed by agarose gel electrophoresis stained with ethidium bromide and visualized with ultraviolet light (Lee et al., 2012).

For all environmental samples, amplicons were first purified with Agencourt AMPure beads (Beckman–Coulter, Brea, California) following the manufacturer's instructions except for the beads/DNA ratio, which was adjusted to 1.5:1. The quality and quantity of purified amplicons was then assessed using a 2200 TapeStation (Agilent Technologies, Santa Clara, California) with D1000 screen tape and reagents. 66 amplicons corresponding at each sampling station were used to prepare 66 DNA libraries for HTS with the PGM Ion Torrent technology using the NEBNext® Fast DNA Library Prep set for Ion Torrent™ (BioLabs, Ipswich, Massachusetts). Libraries were prepared following the manufacturer protocol for end repair, PCR amplification of adapter ligated DNA (7 cycles), and cleaning steps.

Ligation of library adapters to purified amplicons was conducted with 2 µL of P1 adapter (NEB kit) and 2 µL of A–X tag adapter provided in Ion Express™ Barcode adapters (Life Technologies) using one different tag per amplicon. Quality, size and concentration of the libraries were verified using the 2200 TapeStation (Agilent Technologies, Santa Clara, California) with D1000 High Sensitivity screen tape and reagents. After that, each library was diluted to 100 pM and combined into a single mix which was subsequently sequenced by the Plateforme Genome Transcriptome of Bordeaux (PGTB). Sequencing was performed using 1 Ion 318™Chip Kit V2 (Life Technologies) on a PGM Ion Torrent machine.

Sequence data processing was performed starting from 66 unique fastq files resulting from the

sequencing process and provided by the PGTB according to Vasselon et al. (2017a, b). For each fastq file, DNA reads were filtered by length and quality using Mothur software (Schloss et al., 2009), according to the following criteria: minimum length = 250 bp, Phred quality score > 23 over a moving window of 25 bp, maximum 1 mismatch in forward primer sequence, homopolymers < 8 bp. In addition, any sequences containing ambiguous base calls were removed, as well as any reads that did not perfectly match with the *rbc*L barcode. The 66 resulting files were then combined and analyzed as a whole. Denoising of the sequencing error was performed by creating read clusters allowing one nucleotide difference between DNA reads using the *pre.cluster* command. After that, the Uchime algorithm was used to detect and remove chimeric DNA sequences. Then, taxonomic assignment of DNA sequences was performed using the naïve Bayesian method (Wang et al., 2007) with a confidence score threshold of 85% and the Rsyst::diatom database (Rimet et al., 2016a, b) as a reference library (version updated in January 2015). Only DNA sequences belonging to diatoms (Bacillariophyta) were kept for further analysis.

Subsequently, a similarity distance matrix was generated using the align.seqs command. Using this distance matrix, sequences belonging to closely related groups were clustered in operational taxonomical units (OTUs) using the farthest neighbor algorithm at a 95% similarity level. OTUs containing one-single sequence (singletons) were removed and all samples were normalized to the smallest read abundance obtained among the 66 libraries for further analysis (this normalization was done to obtain comparable samples in terms of numbers of OTUs). A list of the OTUs and their relative abundances, based on read abundances per OTU, was produced for each of the samples. Molecular taxa lists were then created by assigning taxonomy to the OTUs using the *classify.otu* command with a stringent consensus confidence threshold (> 80%) (Schloss et al., 2009). A list of taxa and their relative abundances, based on read abundances, was produced for each of the samples.

Statistical analyses

The structure of diatom assemblages derived from both morphological and molecular approaches was compared using a Mantel test. The test was performed between the morphological and the molecular OTU data in the statistical software PC-ORD v. 5.

For both, morphological and molecular data, the influence of each environmental factor on the structure of diatom assemblages was explored using non parametric ANOVA, Multi-response Permutation Procedure (MRPP) (McCune et al., 2002). MRPP was performed with 9999 random permutations using Sorensen distance and rank transformation in the statistical software PC-ORD v. 5. To determine if the selected environmental factors affect diatom assemblages in the same way, the MRPP statistics derived from both methods were compared: Test statistic T, indicating the separation between groups established whiting factors, and the agreement statistic A, indicating within-group homogeneity.

Three diatom-based indices were calculated from both morphological and molecular inventories to assess the ecological status of Lake Bourget. The first one is the "Indice de Polluosensibilité Spécifique''—IPS—(Cemagref, 1982), which is an index widely used in Europe for river quality assessment. Despite this index was not developed to characterize the ecological status of lakes, it was applied to the case of Lake Bourget due to the availability of ecological preferenda for a large number of diatom taxa (2595 taxa). (Cemagref, 1997; Besse-Lototskaya et al., 2011; Schmidt-Kloiber & Hering, 2015). The second one is the "Indice per valutazione della qualità delle acque lacustri italiane a partire dalle diatomee epifitiche ed epilitiche"—EPI-L—(Marchetto et al., 2013), developed in Italy for lake water quality assessment, in particular, for the assessment of peri-alpine lakes; it has already been applied in Lake Geneva with success (Rimet et al., 2016a, b). This index provides trophic and indicator values for 109 diatom taxa (Marchetto et al., 2013). The third one is the "Diatom quality index "—S—(Sgro et al., 2007), developed in the United States for the assessment of the ecological status of the Great lakes, and provides optimal and tolerance values for 402 diatom taxa. For the molecular inventories, prior to the indices calculation, the numbers of reads obtained for each species were converted into relative abundances of OTUs per

sample (as well as it was done for the microscopic counts). The IPS index was calculated in the software OMNIDIA and the EPI-L index was calculated in an Excel spreadsheet proposed by the author (http://www.ise.cnr.it/wfd). For the S index, optimal and tolerance values of diatom species were recovered from Sgro et al. (2007) and the index was calculated manually in Excel. For each index, a correlation test between the morphological and molecular indices scores was performed.

Morphological versus molecular inventories

The morphological and molecular inventories were compared at family, genus and species levels through Venn diagrams using the interactive tool Venny v. 2.1 (Oliveros, 2007). We checked the availability of the species detected in the morphological inventories in the R-Syst::diatom reference library (Rimet et al., 2016a, b). In addition, a comparative table of the five dominant species detected with both approaches was elaborated.

Results

Morphological results

A total of 120 diatom species belonging to 40 genera were identified by light microscopy (LM) among all samples. The number of species per sample ranged between 22 and 46, with an average of 30 species per sample. The dominant taxa were *Encyonopsis subminuta* Krammer & Reichardt (average abundance in all the samples: 29.6%), *Achnanthidium minutissimum* (Kützing) Czarnecki (7.7%), *Cyclotella costei* Druart & Straub (7.5%), *Navicula cryptotenelloides* Lange-Bertalot (7.2%) and *Amphora pediculus* (Kützing) Grunow ex A.Schmidt (6.4%). All these taxa have been reported in the literature as indicators of good ecological conditions (Lecointe et al., 1993; Van Dam et al., 1994; Hofmann et al., 2011).

Molecular results

Clustering of sequences at 95% level resulted in 2174 OTUs. The number of OTUs per sample ranged between 164 and 542, with an average of 335 OTUs per sample. Taxonomical assignation of the OTUs using R-Syst::diatom database resulted in 35 diatom genera containing 61 species. From the 2174 OTUs, only 1531 could be assigned to family level, of which 1270 were identified to genus level. From these 1270 OTUs, only 514 were identified to species level. 498 OTUs could not be assigned to any diatom family, genus nor species according to R-Syst::diatom database and remained "unclassified". The number of species per sample ranged between 9 and 47, with an average of 24 species per sample.

The comparison of the structure of diatom assemblages by means of a Mantel test revealed a highly significant relationship between the morphological and the molecular OTU data ($r = 0.45$; P value < 0.0001) (Fig. 2).

The comparison between the MRPP statistics (T and A statistics) obtained for each environmental factor with both approaches indicated that the structure of diatom assemblages in both cases was structured primarily by the underwater slope, second by the land use and finally by the typology of the coast (Fig. 3).

The ecological status of Lake Bourget was assessed through three diatom indices. Diatom indices scores obtained on the basis of morphological and molecular inventories indicated differences in the ecological status of the littoral zone of Lake Bourget between indices and between approaches (detailed results are given in Supplementary data 1). The results of the correlation test performed between the indices scores

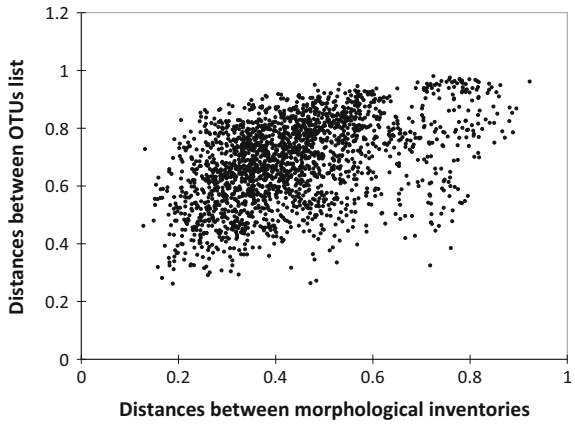

Fig. 2 Plot of the results of the Mantel test comparing the morphological and the molecular OTU data. Pairwise species distance for the 66 samples is plotted on the *X*-axis. Pairwise OTU distance for the 66 samples is plotted on the *Y*-axis. $r = 0.45$; P value < 0.0001

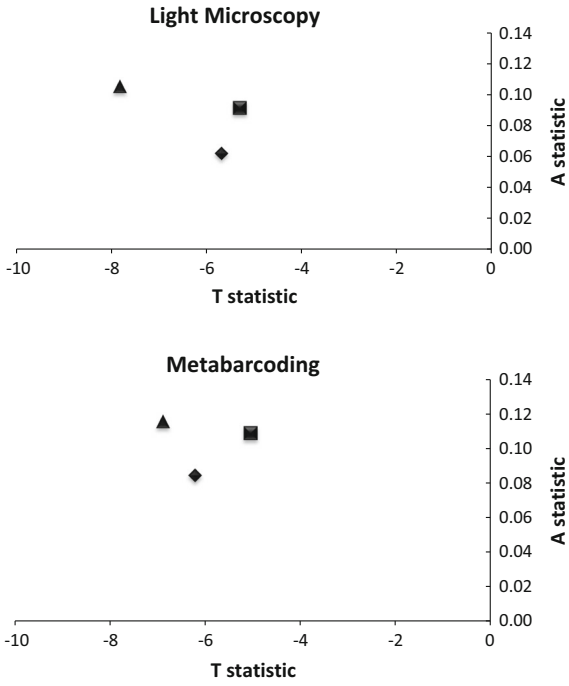

Fig. 3 Plot of MRRP statistics elaborated with the morphological (Light Microscopy) and the molecular OTU (Metabarcoding) diatom communities. *Y*-axis: Agreement statistic *A* describing within-group homogeneity. The closer A is to 1, the identity of the items inside that group is higher. *X*-axis: Test statistic describing the separation between groups. The more negative is *T*, the differences between the groups are higher. *Triangle*: Underwater slope; *P* value: < 0.0001, *square*: Typology of the coast; *P* value < 0.0001, *diamond*: Land use; *P* value < 0.0001

derived from both morphological and molecular inventories are presented in Table 1. No significant correlation was observed in the case of the IPS ($R^2 = 0.0042$; *P* value > 0.05) and the EPI-L ($R^2 = 0.0278$; *P* value > 0.05) indices. S index, for its part, showed significant correlation (*P* value < 0.010). However, Pearson correlation coefficient is very low.

Sources of divergence between the morphological and molecular data

The qualitative comparison between morphological and molecular inventories through Venn diagrams revealed important differences between lists at family, genus and particularly at species level (Fig. 4). A total of 25 taxa were detected at family level from which only 16 taxa (representing 64%) were shared by both methods. At genus level, from a total 55 taxa, only 29 (52.7%) were shared by both methods. At species level, from a total of 166 taxa only 26 (15.7%) were shared by both methods (the detailed comparison is given in Supplementary data 2).

The comparison of the morphological taxa list with the R-Syst::diatom database showed that from the 120 species identified with LM, only 39 species were present in the database (Fig. 5). From the 81 species that were not present in the R-syst:: diatom database, ten corresponded to species with average abundances higher than 1% (*Encyonopsis subminuta* 29.6%, *Amphora indistincta* 5.3%, *Gomphonema bavaricum* 3.2%, *Achnanthidium straubianum* 2.8%, *Fragilaria tenera* 2.4%, *Cymbella neoleptoceros* 2%, *Gomphonema elegans* 1.7%, *Cymbella lange-bertalotii* 1.3%, *Navicula uthermoehlii* 1.2%, and *Denticula tenuis* 1.1%).

In addition, the focus on the five most abundant species detected with both methods presented in Table 2, revealed that the most abundant species

Table 1 Characteristics of the three diatom indices applied to assess the ecological quality of the littoral zone of Lake Bourget including the results of the correlation test performed between each index scores inferred from both morphological and molecular inventories

Diatom index	Ecosystem	Number of species used by the index	% of species detected by LM and used for the index calculation	% of species detected by metabarcoding and used for the index calculation	R^2	*P* value
IPS	River	2595	100	100	0.0042	NS
EPI-L	Lake	109	38	24	0.0278	NS
S	Lake	409	49	33	0.1342	0.0024

NS: *P* value > 0.5

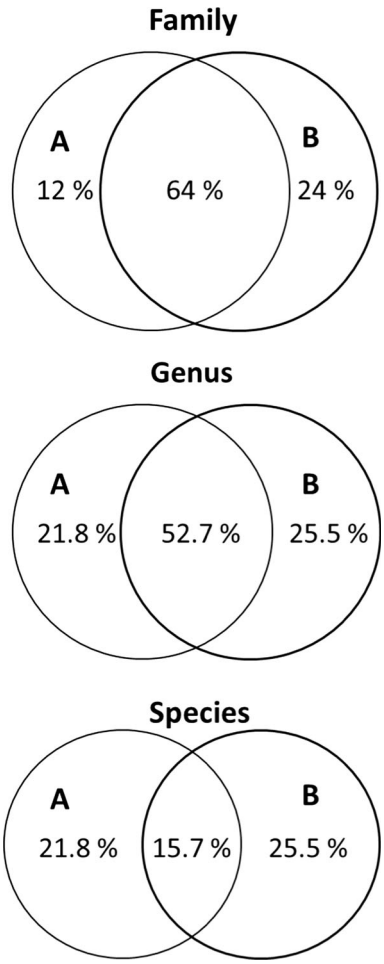

Fig. 4 Venn diagrams showing the percentage of similarity between morphological (**A**) and molecular (**B**) taxa lists at family, genus and species level

detected by LM are different from those detected through the molecular approach.

Discussion

Diatom assemblages structure obtained with both methods are comparable and they are structured by the same environmental factors.

It was already shown that the structure of benthic macrofauna communities based on microscopic counts or on OTUs relative abundances were similar (Lejzerowicz et al. 2015). Until now, comparisons between molecular and morphological approaches for diatom communities were only carried out at species level (Kermarrec et al., 2014; Zimmermann et al., 2015; Visco et al., 2015), but not directly at molecular OTU level. In our study, we demonstrate a high similarity in the structure of diatom assemblages between morphological and molecular approaches. This is true only when molecular assemblages are expressed in terms of OTUs abundances, but not when molecular assemblages are expressed in terms of species after taxonomic assignation of OTUs with the reference library, given its incompleteness (see "Incompleteness of R-Syst::diatom reference library" of this section).

The composition and structure of diatom assemblages can be affected by several factors such as water chemistry, geology as well as topological and geomorphological characteristics of the aquatic system studied (Stevenson & Pan, 1999; Tornés et al., 2007; Smol & Stoermer, 2010). In addition, Cantonati & Lowe (2014) pointed out that distribution of benthic diatom assemblages in lakes is determined by substratum (i.e., rocks, macrophytes, sediments, etc.), depth distribution gradient, physical disturbances (i.e., wave action), irradiance, nutrient content as well as biotic interactions (i.e., grazing). In our study, we could only assess the impact of the three measured parameters: underwater slope, land use and coast typology. The lack of physical and chemical measurements of water at the sampling sites is a limit in our study since these parameters may be determining in understanding changes in diatom assemblages.

Nevertheless, our results show that benthic diatom assemblages in Lake Bourget are affected by the same kind of environmental parameters, whether diatom assemblages are obtained with microscopy or with metabarcoding. They are affected primarily by underwater slope, then by land use and finally by coast typology. The effect of the underwater slope on the structure and composition of diatom assemblages has already been observed by Maruyama et al. (2015) who reported significant differences in littoral diatoms assemblages from Lake Malawi (Africa) among the slopes of the rocks where biofilms were collected. The influence of the land use on benthic diatom assemblages has also been documented in the literature in rivers, but not yet in lakes. These studies show evident differences in community structure and composition between relatively undisturbed and disturbed sites (Potapova & Charles, 2003; DeNicola et al., 2004; Pan

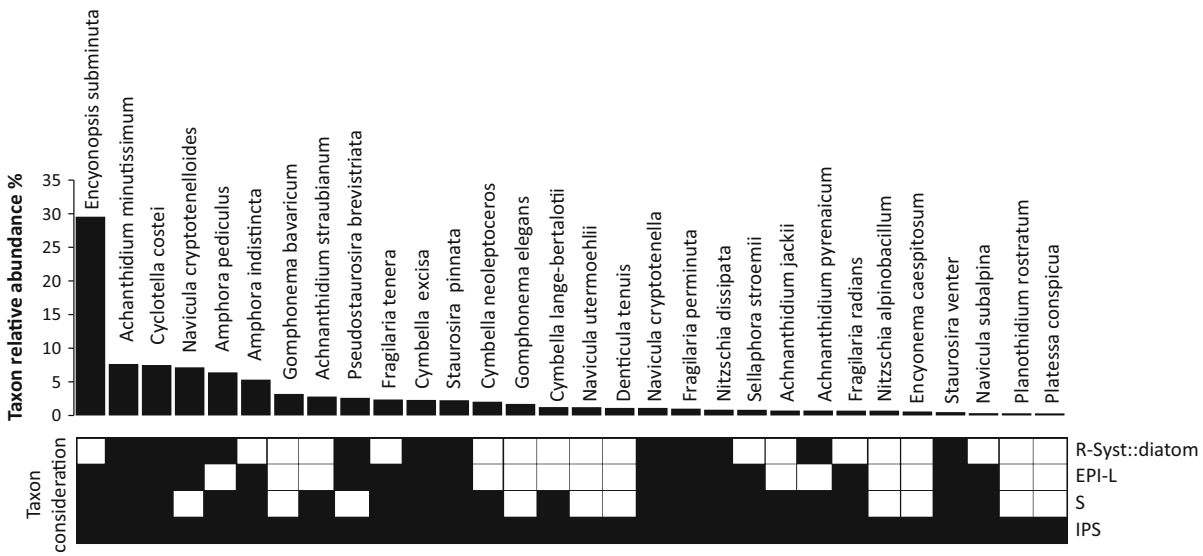

Fig. 5 Average percentages of the 30 most abundant taxa detected by LM in the samples. The bottom table indicates if the taxa were taken into account in the indices calculation (EPI-L, S, IPS) and if they were barcoded in R-Syst::diatom version of January 2015 (black squares)

Table 2 Comparison of the five most common taxa detected with the morphological (left) and the molecular (right) approach

Taxon (microscopy)	Abundance (%)	Taxon (metabarcoding)	Abundance (%)
Encyonopsis subminuta Krammer & Reichardt	29.6	*Amphora unclassified*	30.3
Achnanthidium minutissimum (Kützing) Czarnecki	7.7	*Unclassified*	24.2
Cyclotella costei Druart & Straub	7.5	*Cymbella unclassified*	9.4
Navicula cryptotenelloides Lange-Bertalot	7.2	*Navicula cryptotenella* Lange-Bertalot	7.7
Amphora pediculus (Kützing) Grunow	6.4	*Cymbella excisa* Kützing	4.0

et al., 2004; Vilmi et al., 2016) as we observed in Lake Bourget.

Concerning coast typology, we identified significant changes in the composition of littoral diatom assemblages with respect to their proximity to rivers inlets; this was already pointed out by Rimet et al. (2016a, b) since water chemistry in such sites is directly influenced by high nutrient and organic matter concentrations coming from rivers. Spitale et al. (2014) also observed an impact of shoreline urbanization on phytobenthos at different scales; in particular, they show that artificial shores show a different composition than natural ones. This corroborates our observations in Lake Bourget since sites with artificial ripraps show significantly different molecular and morphological diatom compositions than sites with marshes and beaches.

The ecological status of Lake Bourget was assessed with three diatom indices based on morphological and molecular approaches. Results of the ecological assessments, however, differed between these different diatom indices and between these two approaches. These differences can be explained by five main reasons: (a) use of non-adapted diatom indices, (b) incompleteness of the reference library, (c) the presence of dead frustules and of eDNA (extracellular DNA) in the samples, (d) differences in the estimation of species abundances between approaches (e) system inherent biasing factors of the HTS procedure (e.g., extraction methods, choice of HTS technology, sequencing errors, etc.).

Our discussion will be focused on the first four points (a–d). For the last point (e), our data do not enable an in-depth discussion regarding such bias.

Hereafter, we present important references concerning the major points related to this issue (e). The impact of the choice of the DNA extraction method (Deiner et al., 2015, Vasselon et al., 2017a, b), the importance of the choice of a targeted gene for DNA barcode (Kermarrec et al., 2013; Valentini et al., 2016), the choice of the set of primers for DNA metabarcoding (Elbrecht & Leese 2015), the choice of the PCR amplification protocol (Kebschull & Zador, 2015), the choice of the sequencing technology (Quail et al., 2012), the choice of the bioinformatics data processing (Schmidt et al., 2015), and the importance of the variation of cell biomass among taxa (Thomas et al., 2016).

a) Use of non-adapted diatom indices

The observed discrepancy between indices is related to the fact that the indices used differed in the ecosystems for which they were developed for and in the number of taxa they take into account for their calculation. IPS index was specifically developed for the assessment of rivers; its application to lakes is limited since ecological preferenda of benthic diatom taxa can change from lotic to lentic systems (Bennion et al., 2014). Moreover, IPS index was developed taking into account a large range of chemical parameters (conductivity, nutrients, organic matter, pH) so its determination provides an overall assessment of aquatic system quality (Prygiel & Coste, 1993) which is adapted to river quality assessment, but not to lake quality assessment. In contrast, the other two indices (EPI-L and S indices) were developed for lakes and are based only on total phosphorus concentrations, thus providing an assessment based only on the nutrient content.

Concerning the EPI-L index, even though this index was specifically developed for the ecological assessment of Italian lakes, it includes indicator values for only 109 benthic diatom taxa among which we detected only 45 by LM and 24 by metabarcording in Lake Bourget. This is clearly not enough to have a robust assessment. In the case of the S index, ecological preferenda are available for a much higher number of diatom taxa (402) so there was a better correspondence between the morphological and molecular inventories (59 and 39 taxa, respectively) than with the EPI-L index.

In short, IPS has the advantage to take into account a large number of species (all diatom taxa detected to genus or species level were used for the index calculation for both approaches) but its disadvantage is not to be adapted to lakes, while the EPI-L has exactly the opposite advantage/disadvantage. The S index is more suitable than the EPI-L in our case study.

b) Incompleteness of the reference library

The discrepancy observed between each diatom index inferred with both morphological and molecular approaches can be largely explained by the incompleteness of the R-Syst::diatom reference library (Rimet et al., 2016a, b). Indeed, a large number of diatom species identified in LM could not be detected through metabarcoding because their sequences were not present in the database. Hence, morphological and molecular inventories differed significantly. Differences between inventories were more evident for dominant taxa such as *Encyonopsis subminuta* which was not present in the database and was yet the prevailing species detected in microscopy. *E. subminuta* is an indicator of good ecological quality (Van Dam et al., 1994) and this explains the higher indices scores obtained with the morphological approach compared to the molecular one since predominant taxa and their ecological preferences are the drivers of diatom indices scores (Bigler et al., 2010).

The incompleteness of the reference library and its subsequent impact in the differences observed between approaches was also reflected by the presence of dominant related taxa such as *Navicula cryptotenella* and *Navicula cryptotenelloides.* While *N. cryptotenelloides* was a dominant species in the morphological inventory, *N. cryptotenella* was detected in abundance with metabarcording. Since these taxa are closely related, it is possible that sequences identified as *N. cryptotenella* in the R-Syst::diatom database correspond to the microscopic identifications of *N. cryptotenelloides*. Misidentifications of closely related taxa under the microscope have been observed frequently (e.g., Kahlert et al., 2009) and decrease the accuracy of diatom indexes (Besse-Lototskaya et al., 2006). This situation also highlights the need of curation with help of experts in taxonomy to verify the identity of the organisms before entering their sequences in the

reference library, as proposed by Rimet et al. (2016a, b).

The presence of a large number of sequences that could not be assigned to a taxonomy level (24%) is another important source of divergence between morphological and molecular indices scores since *unclassified* taxa were not included in the indices calculation. On the other hand, it is possible that dominant species from the morphological inventory, and not present in the R-Syst::diatom database, are among these *unclassified* taxa. This is the case of *Encyonopsis subminuta*, *Amphora indistincta*, *Gomphonema bavaricum* and other abundant species detected only by microscopy (see Fig. 5; Supplementary data 2).

Another situation underlying the incompleteness of the reference library can be observed in the molecular inventories where the dominant species is a taxon from the genus *Amphora* which could not be detected at the species level: *Amphora unclassified* (see Table 2). In contrast, *A. indistincta* and *A. copulata* were detected in microscopy, but their sequences are absent in R-Syst::diatom database. It is possible that the *A. unclassified* detected in metabarcoding corresponds to these species. The same was observed for *Cymbella unclassified*, the third most abundant species detected by metabarcoding. Since *C. neoleptoceros* and *C. lange-bertalotii* were detected in microscopy but their sequences are absent from R-Syst::diatom database, it is possible that *C. unclassified* corresponds to these species. However, it could be also possible that the unclassified OTUs are belonging to other or even new taxa.

The incompleteness of reference libraries has already been shown to be one of the main limitation of metabarcoding in diatom identification for monitoring studies (Mann et al., 2010; Kermarrec et al., 2014; Zimmermann et al., 2015; Pawlowski et al., 2016; Rimet et al., 2016a, b). In this sense, it is clear that reference libraries need to be more comprehensive and curated to detect a larger number of taxa at species level. Despite the efforts carried out until now to complete reference libraries, they are still not enough representing the large biodiversity of diatoms (more than 100.000 species according to Mann & Vanormelingen (2013)) and much work still has to be done.

c) Presence of dead frustules and eDNA (extracellular DNA) in the samples

Another source of divergence between morphological and molecular inventories is related to the presence of the planktonic species *Cyclotella costei*, which was abundant in microscopic inventories (7.5%), but was not detected with metabarcoding despite that its sequence is available in R-Syst::diatom database. Our hypothesis is that some dead planktonic frustules are present in the sampled biofilms due to sedimentation, and further collected by scrapping the upper surface of the stones. This may explain why frustules of *C. costei* were detected under the microscope but could not be detected by metabarcoding since no DNA was present. The persistence of diatom frustules from dead cells as a source of divergence between morphological and molecular inventories has also been observed by Kermarrec et al. (2014) while studying river samples. These authors recommended avoiding the use of planktonic species for river biomonitoring purposes.

Nevertheless, we can also offer a vice versa hypothesis. There is probably persisting free-floating eDNA (extracellular DNA) which could also influence the results of the metabarcoding approach. eDNA is a promising area often used to detect rare, endangered or invasive species in aquatic ecosystems (e.g., Valentini et al., 2016). In our case, this point is difficult to demonstrate but it certainly has an impact on the comparability of both methods.

d) Differences in the estimation of species abundances

Another factor responsible of the differences observed between the scores of diatom indices inferred with both methods could be the divergence between species abundances estimated with morphological (number of valves) and molecular (number of DNA reads) methodologies. Since *rbc*L gene is contained in the chloroplasts, the number and the size of plastids per cell may influence the number of genomes, and therefore the number of reads. The number of plastids per cell varies in diatoms (Round et al., 1990) depending on the genus. Vasselon et al. (2017a, b) recently suggested that bigger diatom cells had higher number of *rbc*L copies and this influences the number of reads for a given species in metabarcoding analyses.

Since small-sized species have smaller plastids than bigger species, numbers of reads are lower for smallest species regardless of their cell abundance in the sample. This was the situation for *Achnanthidium minutissimum* and *Amphora pediculus* in the samples of Lake Bourget. These small-sized species were often dominant in morphological inventories (abundances of 7.7 and 6.4%, respectively) but they were less abundant in molecular inventories (0.5 and 0.8%, respectively). These two species, indicators of good water quality, were underestimated in molecular inventories, explaining why morphological indices scores were higher than the molecular ones. At last, Vasselon et al. (submitted) suggest the development of a correction factor based on the variation of the number of plastids of diatoms cells according to their species to adjust these differences.

Conclusions and prospects for the future

Our results showed that morphological and molecular (OTU) inventories provided similar structure of littoral benthic diatom assemblages and that the selected environmental factors structuring morphological and molecular inventories were the same. This suggests that molecular inventories can be used to obtain an unambiguous characterization of diatom communities for ecological studies, as well as microscopic inventories.

Nevertheless, for monitoring purposes, the differences obtained between morphological and molecular approaches in the assessment of the ecological status of the littoral zone of Lake Bourget indicate that metabarcording seems promising insofar as the reference library will be more complete. Other constraints such as the presence of dead frustules and differences in the abundance estimation of species also need to be solved. There is also a need for a specific diatom index for lakes to take into account a larger number of lake benthic diatom species and their appropriate autecology.

Since molecular (OTU) data provide unambiguous characterization of the structure of diatom assemblages, the development of a biotic diatom index based on molecular OTU data is likely to be another convenient option. Apothéloz-Perret-Gentil et al. (2017) have recently showed the potential of a taxonomy-free diatom index to assess the ecological status of Swiss rivers. Nevertheless, it should be kept in mind that working only with OTU data will lead to a loss of all long-lasting and valuable information that is currently linked to taxonomy such as species autecology, life forms, cell-sizes and ecological guilds of diatoms taxa. In these sense, an integrated approach that combines morphological, molecular and ecological aspects should be considered for monitoring purposes.

Acknowledgements The data presented herein is part of a master thesis in environmental sciences presented at the University of Geneva by the first author. The first author is in debt to the Simon I. Patiño Foundation for awarding the scholarship to undertake a Master in Environmental Sciences at the University of Geneva. We thank Lea Féret and Victor Frossard for the sample collection. We also thank the CISALB (Comité Intersyndical pour l'Assainissement du Lac du Bourget) for financing part of the study. The authors also acknowledge the European COST network DNAqua-Net (CA15219) as a fruitful scientific discussion space on molecular approaches for biomonitoring.

References

Ács, É., 2007. Spatial and temporal change of epiphytic algae and their connection with the ecological condition of swallow Lake Velencei–To (Hungary). Acta Biologica Debrecina Oecologica Hungarica 17: 9–111.

Afnor, 2004. NF EN 14407. Qualité de l'eau-Guide pour l'identification et le dénombrement des échantillons de diatomées benthiques de rivières, et leur interprétation. Afnor: 1–13.

Afnor, 2003. NF EN 13946. Qualité de l'eau-Guide pour l'échantillonnage en routine et le prétraitement des diatomées benthiques de rivières. Afnor: 1–18.

Apothéloz-Perret-Gentil, L., A. Cordonier, F. Straub, J. Iseli, P. Esling & J. Pawlowski, 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Molecular Ecology Resources. doi:10.1111/1755-0998.12668.

Balvay, G., J.-C. Druart & S. Jacquet, 2012. Le lac du Bourget ses eaux et sa biologie. Versailles, Editions Quae: 150 pp.

Bennion, H., M. G. Kelly, S. Juggins, M. L. Yallop, A. Burgess, J. Jamieson & J. Krokowski, 2014. Assessment of ecological status in UK lakes using benthic diatoms. Freshwater Science 33: 639–654.

Bere, T. & J. G. Tundisi, 2010. Biological monitoring of lotic ecosystems: the role of diatoms. Brazilian Journal of Biology 70: 493–502.

Besse-Lototskaya, A., P.F.M. Verdonschot & J.A. Sinkeldam, 2006. Uncertainty in diatom assessment: sampling, identification and counting variation. Hydrobiologia 566: 247–260.

Besse-Lototskaya, A., P. F. Verdonschot, M. Coste & B. Van de Vijver, 2011. Evaluation of European diatom trophic indices. Ecological Indicators 11: 456–467.

Bielczyńska, A., 2015. Bioindication on the basis of benthic diatoms: advantages and disadvantages of the Polish phytobenthos lake assessment method (IOJ–the Diatom Index for Lakes)/Bioindykacja na podstawie okrzemek bentosowych: Mocne i s\labe strony polskiej metody oceny jezior na podstawie fitobentosu (IOJ–Indeks Okrzemkowy Jezior). Ochrona Srodowiska i Zasobów Naturalnych 26: 48–55.

Bigler, C., V. Gälman & I. Renberg, 2010. Numerical simulations suggest that counting sums and taxonomic resolution of diatom analyses to determine IPS pollution and ACID acidity indices can be reduced. Journal of Applied Phycology 22: 541–548.

Birk, S., W. Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, A. Solimini, W. Van de Bund, N. Zampoukas & D. Hering, 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.

Blanco, S., L. Ector & E. Bécares, 2004. Epiphytic diatoms as water quality indicators in Spanish shallow lakes. Vie Milieu 54: 71–80.

Bolla, B., G. Borics, K. Kiss, N. M. Reskóné, G. Várbíró & É. Ács, 2010. Recommendations for ecological status assessment of Lake Balaton (largest shallow lake of Central Europe), based on benthic diatom communities. Vie Milieu 60: 197–208.

Brucet, S., S. Poikane, A. Lyche-Solheim & S. Birk, 2013. Biological assessment of European lakes: ecological rationale and human impacts. Freshwater Biology 58: 1106–1115.

Bruder, K. & L. K. Medlin, 2007. Molecular assessment of phylogenetic relationships in selected species/genera in the naviculoid diatoms (Bacillariophyta). I. The genus Placoneis. Nova Hedwig. 85: 331–352.

Bryhn, A. C., C. Girel, G. Paolini & S. Jacquet, 2010. Predicting future effects from nutrient abatement and climate change on phosphorus concentrations in Lake Bourget, France. Ecological Modelling 221: 1440–1450.

Cantonati, M. & R. L. Lowe, 2014. Lake benthic algae: toward an understanding of their ecology. Freshwater Science 33: 475–486.

Cellamare, M., S. Morin, M. Coste & J. Haury, 2012. Ecological assessment of French Atlantic lakes based on phytoplankton, phytobenthos and macrophytes. Environmental Monitoring and Assessment 184: 4685–4708.

CEN. 2015. Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. CEN/TC 230/WG 23 - Aquatic Macrophytes and Algae, 8pp.

Cemagref, 1997. Qualité biologique des eaux du Rhône et essai d'estimation d'effets toxiques sur les communautés de diatomées récoltées à l'aide de substrats artificiels. Agence Eau Rhône-Méditerranée-Corse: 78.

Cemagref, 1982. Etude des méthodes biologiques quantitatives d'appréciation de la qualité des eaux. Rapport Division Qualité des Eaux Lyon. Agence financiè de Bassin Rhone-Méditerarée Corse^ ePierre-Bénite Pierre-Bénite.

Chonova, T., F. Keck, J. Labanowski, B. Montuelle, F. Rimet & A. Bouchez, 2016. Separate treatment of hospital and urban wastewaters: a real scale comparison of effluents and their effect on microbial communities. Science of the Total Environment 542: 965–975.

Deiner, K., J.-C. Walser, E. Mächler & F. Altermatt, 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation 183: 53–63.

DeNicola, D. M., E. de Eyto, A. Wemaere & K. Irvine, 2004. Using epilithic algal communities to asses trophic status in Irish Lakes. Journal of Phycology 40: 481–495.

Dokulil, M. T., 2003. Algae as ecological bio-indicators. Trace Metals and Other Contaminants in the Environment 6: 285–327.

Elbrecht, V. & F. Leese, 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. PLoS ONE 10: e0130324.

Hebert, P. D., A. Cywinska, S. L. Ball, et al., 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270: 313–321.

Hofmann, G., 1994. Aufwuchs-diatomeen in Seen und ihre Eignung als Indikatoren der Trophie. Bibliotheca Diatomologica 30, Cramer, Berlin: 241 pp.

Hofmann, G., M. Werum & H. Lange-Bertalot, 2011. Diatomeen im Sü\s swasser-Benthos von Mitteleuropa: Bestimmungsflora Kieselalgen für die ökologische Praxis: über 700 der häufigsten Arten und ihre Ökologie. ARG Gantner.

Jacquet, S., O. Anneville & I. Domaizon, 2012. Evolution de paramètres clés indicateurs de la qualité des eaux et du fonctionnement écologique des grands lacs péri-alpins (Léman, Annecy, Bourget): étude comparative de trajectoire de restauration post-eutrophisation. Archives des Sciences 65: 191–208.

Jacquet, S., I. Domaizon & O. Anneville, 2014. The need for ecological monitoring of freshwaters in a changing world: a case study of Lakes Annecy, Bourget, and Geneva. Environmental Monitoring and Assessment 186: 3455–3476.

Jacquet, S., D. Barbet, C. Barbier, S. Cachera, M. Colon, L. Espinat, J. Guillard, V. Hamelet, J.-C. Hustache, D. Lacroix, L. Laine, B. Leberre, J. Neasta, G. Paolini, M.-E. Perga, P. Perney & F. Rimet, 2016. Suivi environnemental des eaux du lac du Bourget pour l'année 2015. Rapport INRA-CISALB-CALB.

Kahlert, M., R.-L. Albert, E.-L. Anttila, R. Bengtsson, C. Bigler, T. Eskola, V. Gälman, S. Gottschalk, E. Herlitz, A. Jarlman, et al., 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.

Kebschull, J. M. & A. M. Zador, 2015. Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Research 43: e143–e143.

Kermarrec, L., A. Franc, F. Rimet, P. Chaumeil, J.-F. Humbert & A. Bouchez, 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities:

a test for freshwater diatoms. Molecular Ecology Resources 13: 607–619.
Kermarrec, L., A. Franc, F. Rimet, P. Chaumeil, J.-M. Frigerio, J.-F. Humbert & A. Bouchez, 2014. A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshwater Science 33: 349–363.
King, L., G. Clarke, H. Bennion, M. Kelly & M. Yallop, 2006. Recommendations for sampling littoral diatoms in lakes for ecological status assessments. Journal of Applied Phycology 18: 15–25.
Krammer, K. & H. Lange-Bertalot, 1986. Bacillariophyceae. 1. Teil: Naviculaceae. Süswasserflora Von Mitteleuropa. Gustav Fischer Verlag, Sttuttgart edn: 610 pp.
Krammer, K. & H. Lange-Bertalot, 1988. Süsswasserflora von Mitteleuropa. Bacillariophyceae. 2. Teil: Epithemiaceae, Bacillariaceae, Surirellaceae, vol 2/2. Gustav Fischer Verlag, Stuttgart edn: 610 pp.
Krammer, K. & H. Lange-Bertalot, 1991. Susswasserflora von Mitteleuropa. Bacillariophyceae 3. Teil: Centrales, Fragilariaceae, Eunotiaceae. Süs Swasserflora Von Mitteleuropa. Gustav Fischer Verlag, Sttuttgart edn: 598 pp.
Lange-Bertalot, H., 2001. Navicula sensu stricto 10 genera separated from Navicula sensu lato Frustulia. Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats 2: 526.
Lecointe, C., M. Coste & J. Prygiel, 1993. "Omnidia": software for taxonomy, calculation of diatom indices and inventories management. Hydrobiologia 269: 509–513.
Lee, P. Y., J. Costumbrado, C.-Y. Hsu & Y. H. Kim, 2012. Agarose gel electrophoresis for the separation of DNA fragments. JoVE Journal of Visualized Experiments 62: e3923.
Lejzerowicz, F., P. Esling, L. Pillet, T. A. Wilding, K. D. Black & J. Pawlowski, 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Scientific Reports 5: 13932.
Mann, D. G. & P. Vanormelingen, 2013. An inordinate fondness? The number, distributions, and origins of diatom species. Journal of Eukaryotic Microbiology 60: 414–420.
Mann, D. G., S. Sato, R. Trobajo, P. Vanormelingen & C. Souffreau, 2010. DNA barcoding for species identification and discovery in diatoms. Cryptogamie Algologie 31: 557–577.
Marchetto, A., Agostinelli, C., Alber, R., Beghi, A., Balsamo, S., Bracchi, S., Buzzi, F., Carena, E., Cavalieri, S., Cimoli, F., et al., 2013. Indice per valutazione della qualità delle acque lacustri italiane a partire dalle diatomee epifitiche ed epilitiche (EPI-L). Rep. CNR-ISE 2: 75–92.
Maruyama, A., K. Shinohara, M. Sakurai, T. Ohtsuka & B. Rusuwa, 2015. Microhabitat variations in diatom composition and stable isotope ratios of the epilithic algae in Lake Malawi. Hydrobiologia 748: 161–169.
McCune, B., J.B. Grace & D.L. Urban, 2002. Analysis of ecological communities. MjM software design Gleneden Beach, OR: 300 pp.
Meunier, A. & S. Jacquet, 2015. Do phages impact microbial dynamics, prokaryotic community structure and nutrient dynamics in Lake Bourget? Biology Open 4: 1528–1537.
Oliveros, J.C., 2007. Venny. An interactive tool for comparing lists with Venn's diagrams (WWW Document). Venny Interactions Tool Comparing Lists with Venns Diagram [available on internet at http://bioinfogp.cnb.csic.es/tools/venny/index.html].
Pan, Y., A. Herlihy, P. Kaufmann, J. Wigington, J. Van Sickle & T. Moser, 2004. Linkages among land-use, water quality, physical habitat conditions and lotic diatom assemblages: a multi-spatial scale assessment. Hydrobiologia 515: 59–73.
Pawlowski, J., F. Lejzerowicz, L. Apotheloz-Perret-Gentil, J. Visco & P. Esling, 2016. Protist metabarcoding and environmental biomonitoring: time for change. European Journal of Protistology 55: 12–25.
Potapova, M. & D. F. Charles, 2003. Distribution of benthic diatoms in US rivers in relation to conductivity and ionic composition. Freshwater Biology 48: 1311–1328.
Prygiel, J. & M. Coste, 1993. The assessment of water quality in the Artois-Picardie water basin (France) by the use of diatom indices. Hydrobiologia 269/270 (Dev. Hydrobiologia 90): 343–349.
Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow & Y. Gu, 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13: 341.
Reichardt, E., 1997. Taxonomische revision des artenkomplexes um Gomphonema pumilum (Bacillariophyceae). Nova Hedwigia 65: 99–130.
Rimet, F., 2012. Recent views on river pollution and diatoms. Hydrobiologia 683: 1–24.
Rimet, F., H.-M. Cauchie, L. Hoffmann & L. Ector, 2005. Response of diatom indices to simulated water quality improvements in a river. Journal of Applied Phycology 17: 119–128.
Rimet, F., A. Bouchez & K. Tapolczai, 2016a. Spatial heterogeneity of littoral benthic diatoms in a large lake: monitoring implications. Hydrobiologia 771: 179–193.
Rimet, F., Chaumeil, P., Keck, F., Kermarrec, L., Vasselon, V., Kahlert, M., Franc, A. & Bouchez, A., 2016b. R-Syst:: diatom: an open-access and curated barcode database for diatoms and freshwater monitoring. Database 2016: baw016.
Round, F. E., R. M. Crawford & D. G. Mann, 1990. Diatoms: Biology and Morphology of the Genera. Cambridge University Press, Cambridge.
Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, et al., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environment Microbiology 75: 7537–7541.
Schmidt, T. S., J. F. Matias Rodrigues & C. Mering, 2015. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. Environmental Microbiology 17: 1689–1706.
Schmidt-Kloiber, A. & D. Hering, 2015. www. freshwaterecology. info—an online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. Ecological Indicators 53: 271–282.
Sgro, G. V., E. D. Reavie, J. C. Kingston, A. R. Kireta, M. J. Ferguson, N. P. Danz & J. R. Johansen, 2007. A diatom

quality index from a diatom-based total phosphorus inference model. Environmental Bioindicators 2: 15–34.

Smol, J. P. & E. F. Stoermer, 2010. The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press, Cambridge.

Spitale, D., A. Scalfi & M. Cantonati, 2014. Urbanization effects on shoreline phytobenthos: a multiscale approach at lake extent. Aquatic Sciences 76: 17–28.

Stenger-Kovács, C., K. Buczko, E. Hajnal & J. Padisák, 2007. Epiphytic, littoral diatoms as bioindicators of shallow lake trophic status: Trophic Diatom Index for Lakes (TDIL) developed in Hungary. Hydrobiologia 589: 141–154.

Stevenson, R. J., 1998. Diatom indicators of stream and wetland stressors in a risk management framework. Environmental Monitoring and Assessment 51: 107–118.

Stevenson, R. J. & Y. Pan, 1999. Assessing environmental conditions in rivers and streams with diatoms. The Diatoms: Applications for the Environmental and Earth Sciences 1: 4.

Stevenson, R. J., J. T. Zalack & J. Wolin, 2013. A multimetric index of lake diatom condition based on surface-sediment assemblages. Freshwater science 32(3): 1005–1025.

Stoof-Leichsenring, K. R., L. S. Epp, M. H. Trauth & R. Tiedemann, 2012. Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. Molecular Ecology 21: 1918–1930.

Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann & E. Willerslev, 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21: 2045–2050.

Thomas, A. C., B. E. Deagle, J. P. Eveson, C. H. Harsch & A. W. Trites, 2016. Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. Molecular Ecology Resources 16: 714–726.

Tornés, E., J. Cambra, J. Gomà, M. Leira, R. Ortiz & S. Sabater, 2007. Indicator taxa of benthic diatom communities: a case study in Mediterranean streams. Annales de Limnologie-International Journal of Limnology. EDP Sciences 43: 1–11.

Vadeboncoeur, Y., M. J. Vander Zanden & D. M. Lodge, 2002. Putting the Lake Back Together: reintegrating Benthic Pathways into Lake Food Web Models: lake ecologists tend to focus their research on pelagic energy pathways, but, from algae to fish, benthic organisms form an integral part of lake food webs. BioScience 52: 44–54.

Valentini, A., P. Taberlet, C. Miaud, R. Civade, J. Herder, P. F. Thomsen, E. Bellemain, A. Besnard, E. Coissac, F. Boyer, et al., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25: 929–942.

Van Dam, H., A. Mertens & J. Sinkeldam, 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. Aquatic Ecology 28: 117–133.

Vasselon, V., I. Domaizon, F. Rimet, M. Kahlert & A. Bouchez, 2017a. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: do DNA extraction methods matter? Freshwater Science 36: 162–177.

Vasselon, V., I. Domaizon, F. Rimet, K. Tapolczai & A. Bouchez, 2017b. Optimization of Diatom DNA Metabarcoding for Freshwater Biomonitoring: Application to Mayotte Streams Monitoring Network. University of Essen, Germany.

Vilmi, A., S. M. Karjalainen, S. Hellsten & J. Heino, 2016. Bioassessment in a metacommunity context: are diatom communities structured solely by species sorting? Ecological Indicators 62: 86–94.

Visco, J. A., L. Apothéloz-Perret-Gentil, A. Cordonier, P. Esling, L. Pillet & J. Pawlowski, 2015. Environmental monitoring: inferring the diatom index from next-generation sequencing data. Environmental Science and Technology 49: 7597–7605.

Wang, Q., G. M. Garrity, J. M. Tiedje & J. R. Cole, 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environment Microbiology 73: 5261–5267.

Zimmermann, J., G. Glöckner, R. Jahn, N. Enke & B. Gemeinholzer, 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Molecular Ecology Resources 15: 526–542.