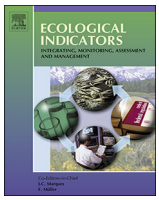
Ecological Indicators 113 (2020) 106225

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



18S-V9 DNA metabarcoding detects the effect of water-quality impairment on stream biofilm eukaryotic assemblages

Alison D. Minerovic[a,*], Marina G. Potapova[a], Christopher M. Sales[b], Jacob R. Price[c], Mihaela D. Enache[d]

[a] Academy of Natural Sciences, Drexel University, Philadelphia 19103, PA, USA
[b] Civil, Architectural, and Environmental Engineering, Drexel University, Philadelphia, PA 19104, USA
[c] Department of Soil Science, College of Agricultural & Life Sciences, University of Wisconsin – Madison, Madison, WI 53703, USA
[d] New Jersey Department of Environmental Protection, Division of Science and Research Trenton, NJ 08625, USA

ARTICLE INFO

Keywords:
Amplicon sequencing
Benthos
Bioassessment
DNA metabarcoding
Eukaryotes
Streams

ABSTRACT

DNA metabarcoding is rapidly expanding as a new approach to biodiversity assessments and biomonitoring and is especially valuable for characterizing microbial communities in aquatic habitats. When applied to eukaryotic organisms, metabarcoding is usually targeting specific taxonomic groups, such as macroinvertebrates, fungi, diatoms, or other protists. The goal of this study was to explore the potential use of metabarcoding of entire biofilm eukaryotic assemblages for the purpose of stream biomonitoring. We sampled 14 stream sites in New Jersey, USA along an impairment gradient and characterized rock biofilm assemblages using Illumina Mi-Seq sequencing of the V9 hypervariable region of 18S rDNA following the Earth Microbiome Project (EMP) protocol. We also enumerated diatoms from the same samples to compare DNA metabarcoding results with morphological assessments. Among the 5866 unique rDNA sequence variants, the fungal and holozoan sequences were the most diverse, while diatom sequences were the most abundant in most sites. Among-site variability of assemblage composition was significantly higher than within-site variability of field and lab replicates, which indicates an acceptable level of reproducibility of the method. Different taxonomic groups of eukaryotes exhibited similar, but not identical patterns of assemblage variation in response to underlying environmental gradients. Both morphological and metabarcoding approaches recovered strong relationships between diatom assemblage composition and water quality impairment. Several other groups of eukaryotes, such as fungi, peronosporomycetes, green algae, and holozoans had only slightly weaker response to water quality impairment than diatoms. These findings suggest that molecular characterization of biofilm eukaryotic assemblages can be an effective tool for monitoring stream biota and its responses to disturbance even if the taxonomic assignments of sequences are only partially resolved.

1. Introduction

Biological monitoring of rivers and streams in the United States most frequently involves assessment of fish, benthic macroinvertebrate and algal assemblages (Barbour et al., 1999). Traditional methods of stream bioassessment are based on morphological identification of organisms, which is often costly and time-consuming. Many small soft-bodied organisms, especially protists, are difficult to preserve or have few morphological characters that are observable using conventional microscopy and, therefore, are not used in biomonitoring. Even identification of relatively character-rich protistan groups, such as diatoms, is hampered by the presence of cryptic and pseudocryptic species (e.g., Vanormelingen et al., 2013), morphological plasticity, and an unknown degree of geographic variability in species morphology. Ultimately, this leads to considerable uncertainties in environmental inferences and lower confidence of these biological indicators.

DNA metabarcoding allows for automated identification of organisms from environmental samples and is thus less affected by human bias and is potentially more cost-effective than conventional methods of bioassessment (Ji et al., 2013; Pawlowski et al., 2018; Taberlet et al., 2012). This approach employs high-throughput next-generation sequencing (NGS) of short DNA fragments from mixed-species samples. The obtained sequences are then matched to reference sequences with known taxonomic assignments in genetic databases (Keck et al., 2017),

* Corresponding author.
E-mail address: adm354@drexel.edu (A.D. Minerovic).

https://doi.org/10.1016/j.ecolind.2020.106225
Received 23 August 2019; Received in revised form 11 February 2020; Accepted 14 February 2020
Available online 28 February 2020
1470-160X/

although taxonomy-blind approaches have also been explored (Apothéloz-Perret-Gentil et al., 2017; Cowart et al., 2015). In the last few years, DNA metabarcoding has been successfully used to evaluate microbial communities, including eukaryotes, in various environments including oceanic and lacustrine plankton (de Vargas et al., 2015; Grattepanche et al., 2016; Lindeque et al., 2013; Rachik et al., 2018; Xu et al., 2017; Filker et al., 2016), and marine and freshwater benthos (Bik et al., 2012; Chariton et al., 2010; Volant et al., 2016). Several research groups have been developing methods for metabarcoding of freshwater diatoms (Kermarrec et al., 2013, 2014; Kelly et al., 2018; Rivera et al., 2018; Rimet et al., 2018; Visco et al., 2015; Zimmermann et al., 2011, 2015) and benthic macroinvertebrates (Gibson et al., 2015; Stein et al., 2014; Elbrecht et al., 2017) for the purposes of routine biomonitoring of inland waters. Metabarcoding is known for its high sensitivity and ability to detect the presence of rare or low-abundance species (Zhan et al., 2013), while increased taxonomic resolution of metabarcoding-based assessments has been shown to improve the ability to detect shifts in assemblages caused by environmental perturbations (Stein et al., 2014).

A variety of marker genes and gene fragments have been explored for metabarcoding of specific eukaryotic taxonomic groups. The mitochondrial COI gene was initially promoted as a universal barcoding marker (Meusnier et al., 2008) and is used now for metabarcoding of metazoans (Elbrecht et al., 2017; Hajibabaei et al., 2011). The nuclear ribosomal ITS region is used for fungi (Schoch et al., 2012; Blaalid et al., 2013), and plants (Wang et al., 2014). Several potential markers have been explored for metabarcoding diatoms, with the V4 hypervariable region of the 18S rDNA gene and partial plastid rbcL gene identified as the most efficient for discriminating diatom taxa at a level similar to that achieved by morphological methods, and has been recommended for routine biomonitoring in European fresh waters (Zimmermann et al., 2011; Kermarrec et al., 2013; Kelly et al., 2018).

The V9 region of the 18S rDNA gene can be amplified with a single set of primers in a wide range of eukaryotes (Amaral-Zettler et al., 2009) and has been applied to metabarcoding studies of oceanic and estuarine plankton (Abad et al., 2016; de Vargas et al., 2015; Malviya et al., 2016). It has been successfully used in tracking seasonal dynamics and studying spatial variability of marine plankton (Brannock et al., 2016) and therefore is potentially useful for biomonitoring. One of the advantages of 18S-V9 metabarcoding is the availability of a relatively simple one-step-PCR amplicon library preparation method (Earth Microbiome Project, Thompson et al., 2017; Gilbert et al., 2010; Caporaso et al., 2012), which greatly reduces the cost of sequencing when carried out on a large scale. On the other hand, this marker was found to have insufficient taxonomic resolution to discriminate among species or even genera and therefore has been dismissed as a potential biomonitoring tool (Tang et al., 2012).

The need to find genetic markers able to identify organisms at the species level arises from the idea of using biotic metrics and indices developed based on the knowledge of ecological attributes of morphotaxa (Hering et al., 2018). This approach has been promoted because of the need to maintain continuity of ambient monitoring programs and, therefore, apply the same metrics that were developed prior to metabarcoding era. For example, Kelly et al. (2018) modified the Trophic Diatom Index (TDI) routinely used in the UK to be used with a subset of diatom taxa that can be identified to species level using metabarcoding with the rbcL gene fragment. Visco et al. (2015) applied metabarcoding with 18S-V4 region to identify diatoms for calculating a version of the DI-CH diatom index adopted in Switzerland, while Vasselon et al. (2017), Bailet et al (2019) and Mortágua et al (2019) employed rbcL or 18S-V4 diatom metabarcoding for calculating the Specific Pollution-sensitivity index (SPI or IPS) commonly used in France and several other European countries. Using metabarcoding for species-level identification requires a well-curated reference database, which ideally would contain sequences of all species that could be found in the geographic area of interest. Although several reference databases for specific taxonomic groups of organisms already exist, they vary in their completeness and taxonomic extent. For instance, diatom reference databases that require considerable resources and time for their development only represent a portion of the taxa most common to European freshwaters (Rimet et al., 2015, 2019; Zimmermann et al., 2014; Kelly et al., 2018).

If we assume, however, that new metrics can be developed by associating DNA sequences with specific environmental conditions, metabarcoding may be used directly to generate bioassessment data without the need to identify to species level (Hering et al., 2018; Apothéloz-Perret-Gentil et al., 2017). Although such a 'taxonomy-free' approach would not be appropriate for estimating species richness, it may represent a cost-effective biomonitoring tool suitable for large-scale assessments (Cordier et al., 2019; Pawlowski et al., 2016, 2018). The goal of this study was to test the hypothesis that 18S-V9 molecular signatures of eukaryotic assemblages in stream biofilms are sufficiently informative for detecting the effects of water quality impairment on these assemblages. If biological assemblages characterized by this marker show response to impairment, we expect it to be a good candidate for use in a large-scale biomonitoring.

2. Methods

2.1. Field sampling

Fourteen stream sites throughout northern New Jersey, USA, were sampled in this study (Fig. 1).

Sites were selected based on historical water quality data available from the Water Quality Portal (https://www.waterqualitydata.us/), US Geological Survey Surface-Water Data for New Jersey (https://waterdata.usgs.gov/nj/nwis/sw) and the Delaware River Watershed Initiative (DRWI) (https://4states1source.org/), representing a gradient of "Good" to "Poor" water quality in terms of nutrients, conductivity, and in-stream and riparian habitat conditions (Ponader et al., 2007) (Table 1).

Samples of epilithic biofilms were collected between June and August 2017 (Table 1). At each site, three field replicates were collected using the 'top-rock' scraping method (Moulton et al., 2002) and suspended in ambient stream water in 250 mL bottles. Rocks were selected to represent in-stream habitat variability, where present (e.g. shallow vs. deep water; fast- vs. slow-moving water; less vs. greater canopy cover). Water temperature, pH, and conductivity were measured in the field with an ExStick II meter. Nutrient concentration data reported in Table 1 were retrieved from the Water Quality Portal (https://www.

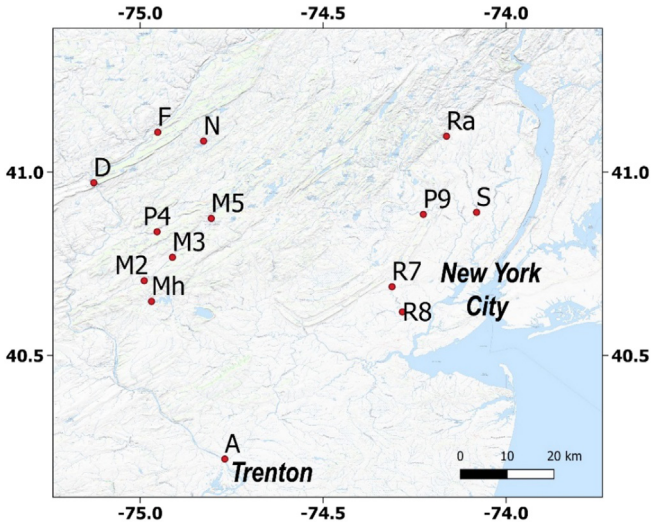


Fig. 1. Map of sampling locations in New Jersey, USA.

Table 1
Locations and water quality characteristics of sampling sites.

Site name	Code	Latitude	Longitude	Date sampled	EC, μS/cm	pH	TN, mg/L	TP, mg/L	Impairment category
Dunnfield Creek	D	40.97106	−75.12669	3-Jun-17	39	6.8	0.1	0.01	Good
Flat Brook	F	41.10888	−74.95195	3-Jun-17	238	7.2	0.2	0.01	Good
Neldons Brook	N	41.08489	−74.82656	3-Jun-17	209	7.2	0.3	0.01	Good
Musconetcong River, Lower	M2	40.70397	−74.98880	21-Jun-17	610	8.4	2.7	0.02	Fair
Musconetcong River at Point Mountain	M3	40.76749	−74.91166	21-Jun-17	640	8.4	1.9	0.03	Fair
Mulhockaway Creek	Mh	40.64750	−74.96888	21-Jun-17	336	8.3	1.0	0.03	Fair
Pequest River	P4	40.83719	−74.95371	21-Jun-17	645	8.9	1.4	0.04	Fair
Musconetcong River at Stephens State Park	M5	40.87370	−74.80565	21-Jun-17	708	8.8	1.0	0.03	Poor
Assunpink Creek	A	40.21722	−74.76861	1-Aug-17	494	7.2	5.0	0.5	Poor
Passaic River	P9	40.88472	−74.22611	1-Aug-17	595	8.1	2.9	0.2	Poor
Rahway River near Springfield	R7	40.68750	−74.31167	1-Aug-17	913	7.6	1.3	0.12	Poor
Rahway River at Rahway	R8	40.61889	−74.28333	1-Aug-17	746	7.9	1.3	0.1	Poor
Ramapo River	Ra	41.09806	−74.16278	1-Aug-17	845	7.6	1.7	0.1	Poor
Saddle River	S	40.89028	−74.08056	1-Aug-17	1169	8.2	5.1	0.9	Poor

waterqualitydata.us/) using approximate sampling dates.

2.2. Laboratory procedures

The molecular library was prepared using a modified version of the Earth Microbiome Project (EMP) 18S Illumina Amplicon Protocol (Amaral-Zettler et al., 2009; Caporaso et al., 2012; Stoeck et al., 2010). Fifty-milliliter aliquots of each of the 42 field samples were transferred to 50 mL conical-bottom tubes, centrifuged for 5 min, and supernatant stream water was aspirated using a vacuum pump until all stream water was removed. Genomic DNA was extracted from each field sample, in duplicate, for a total of 84 lab samples, using the Qiagen /MoBio PowerSoil DNA Isolation Kit following manufacturer's instructions. Pellets were frozen at −20 °C following DNA extraction. DNA concentration was quantified using a QuBit 2.0 fluorometer prior to amplification with a QuBit dsDNA high sensitivity assay kit (Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA). PCR reaction mixtures for DNA amplification were prepared using the Illumina_Euk_1391f forward primer and a unique barcoded Illumina_Euk_br reverse primer for each of the 84 lab samples. The mammal blocking primer was not used in this study. DNA extraction aliquots were not amplified in triplicate as recommended by the EMP protocol to reduce the risk of contamination during either the amplification or pooling steps. PCR products were purified using a Qiagen QIAquick Purification kit following manufacturer's instructions and quantified using a QuBit 2.0 fluorometer. Equal masses of each sample were pooled and submitted along with the forward, reverse, and index sequencing primers for paired-end (2 × 300 bp) sequencing on the Illumina MiSeq platform at the QB3 Vincent J. Coates Genomics Sequencing Lab, UC Berkley. The sequence data are available at the NCBI short read repository under the accession number PRJNA603149.

For morphological identification of diatoms, 0.2–0.5 g biomass subsamples from each of the 42 field samples were re-suspended in deionized water and digested in 30% H_2O_2 to remove organic material. Digested samples were centrifuged at 4000 RPM for 10 min, supernatant was decanted, and pellets were re-suspended in deionized water for five rinsing cycles. Cleaned material was dried on 22 mm^2 cover slips and affixed as permanent microscope slides using Naphrax mounting medium (Brunel Microscopes, UK). Slides were examined with a Zeiss AxioImager A1 light microscope (LM) equipped with an AxioScopeMRm digital camera using DIC under oil immersion at 1000-1600x total magnification. At least 400 valves were identified and enumerated to the lowest taxonomic level using standard floras (Hofmann et al., 2013; Krammer & Lange-Bertalot 1986, 1988, 1991a,b; Spaulding et al., 2018).

2.3. Data analysis

Amplicon sequence variants were detected in R (R Core Team, 2018) using the tutorial 'A DADA2 workflow for Big Data: Paired-end' (Callahan et al., 2016a; 2016b). Primers and adapters were trimmed from the demultiplexed paired-end sequences (2x300 bp reads) using Cutadapt v. 1.18 (Martin, 2011). The resulting 18S-V9 fragments were on average 127 bp long. Reads were then checked for quality, denoised, truncated and merged using the DADA2 workflow. Sequence truncation was based on the visual inspection of forward and reverse read quality plots. Forward reads were truncated at 82 bp and reverse reads at 98 bp. Chimeras were identified and removed, and the resulting amplicon sequences were used in the downstream analyses, without binning them into OTUs based on similarity at certain threshold. The use of exact sequence variants (ESV) instead of traditional OTUs that represent clusters of similar sequences has been recently advocated by Callahan et al. (2017) as this approach better captures biological variation in the data and allows direct comparisons among data sets. Exact sequence variants have been shown to successfully reveal ecological patterns of the communities while increasing taxonomic resolution of the data (Glassman and Martiny, 2018). Exact sequence variants are called "OTUs" here for convenience as this term is more commonly used in metabarcoding literature. Taxonomic assignment of exact sequence variants ("OTUs") was done using the RDP Naïve Bayes Classifier algorithm (Wang et al., 2007) and aligned to the SILVA reference database v. 128 (Pruesse et al., 2007), using subsets of OTUs clustered at 97% identity with consensus taxonomy at ≥80% bootstrap support. OTUs were identified to genus or lowest possible taxonomic level and confirmed by subsequent GenBank BLAST searches, using less stringent thresholds of ≥80% identity and e-values $\leq 1e^{-5}$. For OTUs identified at 'Phylum' level, GenBank BLAST searches were carried out to further verify their taxonomic identities, using the same threshold for percent identity and e-value (Zhang et al., 2000).

For downstream analysis of eukaryotic OTUs, we used the R package 'phyloseq' (McMurdie and Holmes, 2013) to combine the resulting taxonomy table with the respective matrix of OTUs by sample and associated field data. To estimate recovered OTU richness, we constructed rarefaction curves showing unique OTU accumulation with sampling size for each sampling site using the 'vegan' R package (Oksanen et al., 2018).

We assigned OTUs to major taxonomic groups and constructed barplots to explore diversity and abundance within and among sites using the 'graphics' R package (R Core Team, 2018). We then subsetted our OTU matrix by each eukaryote group. To investigate distributional patterns of various biotic components of the biofilm assemblage along environmental gradients, we constructed non-constrained Nonmetric Multidimensional Scaling (NMDS) ordinations of each Hellinger-transformed OTU and diatom morphological data set with fitted

environmental variables using the R package 'vegan' (Oksanen et al., 2018) and visualized sample-environment relationships for all data sets using 'ggplot2' (Wickham, 2016). Environmental variables, such as conductivity and nutrient concentrations were log-transformed prior to the analyses. The NMDS plots of diatom OTUs and morphospecies (Fig. 4b, 4d) were constructed by using the 'plot' function in the 'graphics' package (R Core Team, 2018). We used the 'ordiselect' function from the 'goeveg' R package (Goral and Schellenberg, 2018) to display only those with the best species-environment correlation. The OTUs displayed represented the 15% best environmental fit; the morphospecies displayed represented the 25% best environmental fit. We carried out Procrustes analyses between each pair of ordinations using symmetric rotation using 'vegan' (Oksanen et al., 2018) to estimate similarity of response to environment among groups of organisms and between molecularly and morphologically characterized diatom assemblages.

For all OTU groups and diatom morphological data, we conducted Permutational Multivariate Analyses of Variance (PERMANOVA) using the R package 'vegan' (Oksanen et al., 2018) to compare within- and among-site variability in assemblage composition and thus to evaluate reproducibility of the 18S-V9 metabarcoding. To test the response of each taxonomic group and molecularly and morphologically characterized diatom assemblages to impairment, we evaluated the CCA model performance constrained by an impairment vector (Table 4) using the R package 'vegan' (Oksanen et al., 2018). Finally, we identified which OTUs were associated with good and poor water-quality sites using an Indicator Species Analysis (Dufrene and Legendre 1997) with R package 'indicspecies' (De Caceres and Legendre, 2009).

3. Results

3.1. Assemblage composition

DNA metabarcoding of 18S-V9 revealed highly diverse and variable taxonomic composition of eukaryotes in each of our study sites. For the 84 samples, a total of 8,049,665 raw reads were obtained from the sequencing facility. A total of 6,995,104 reads passed the quality control and bioinformatic pipeline and were subsequently assigned membership to one of the 5,866 OTUs (Supplementary Material 1). 3831 of the 5866 OTUs (representing approximately 37% of total abundance) could only be assigned to the level of "Eukaryota" using the SILVA reference database. A subsequent BLAST search of all OTUs revealed that 837 did not meet the e-value and/or percent identity thresholds and were removed from further analysis. These unassigned OTUs represented 16% of total abundance and 14.2% of OTU richness. From the remaining 5,029 OTUs (those possessing an assigned taxonomy), 972 were matched to GenBank sequences at 100% identity. The total number of sequences per sample ranged from 21,832 to 126,036, with an average of 76,117 reads per sample. The total number of OTUs per sample ranged from 98 to 856 with an average of 380 sequence variants per sample. Rarefaction curves of OTUs for individual sites (Fig. 2) show that eukaryotic diversity was well-sampled as plateaus were reached for all sites and that OTU richness varied considerably among sites.

The lowest richness was observed at a shaded Good-quality site, Neldons Brook, where the curve plateaued at about 400 OTUs and the highest at two heavily impacted sites, Assunpink Creek and Musconetcong River at Stephens State Park where the estimated number of OTUs was about 1400 OTUs at each site.

In 10 out of the 14 sites, diatoms were determined to be the most abundant phylum (Fig. 3), although they were represented by only 343 unique OTUs (Supplementary material 1).

Among diatoms, the OTUs present at all 14 sites corresponded to *Gomphonema* spp. (OTU_0055, _0165) and *Mayamaea* (OTU_0131). The most abundant diatom OTUs among all sites were identified as *Rhoicosphenia* cf. *abbreviata* and various *Navicula* species. Another group of abundant diatom OTUs were identified as *Paralia sol*, a marine planktonic species and therefore definitely not the actual diatom corresponding to these sequence variants. These "*Paralia*" sequences were found only in the Musconetcong River. Microscopy revealed a total of 234 diatom taxa (Supplementary material 2). Diatoms with the highest number of occurrences were *Cocconeis placentula* (41 out of 42 samples), *Achnanthdium minutissimum* (40 samples), *Amphora pediculus* (36 samples), *Sellaphora atomus* (36 samples), *Rhoicosphenia abbreviata* (32), *Achnanthidium rivulare* (31 samples), *Navicula gregaria* (30 samples) and *Nitzschia amphibia* (30 samples). A large-celled diatom present only in Musconetcong River at low abundance was *Ellerbeckia arenaria*; therefore, it is possible that sequences of "*Paralia sol*" belonged to this species.

The two other abundant groups of 18S-V9 OTUs belonged to Holozoans (animals) and Chlorophyta (green algae). Although Holozoans were quite diverse within the samples (693 unique OTUs), they were dominant at only two sites (Fig. 3). The most abundant and common among Holozoan sequence variants were those of various rotifers, nematodes and arthropods. The holozoan OTUs found in all 14 sites belonged to nematodes (OTU_1319, _2060), rotifers (OTU_1402) and parasitic flatworms (OTU_2593, _3684).

Among green algae, a *Chlamydomonas* (OTU_0555) and a *Scenedesmus* (OTU_0599) were found in all 14 sites. The most abundant green algal OTUs included filamentous *Chaetophora* (OTU_0488, found in two good water-quality sites only), *Rhizoclonium* (OTU_0465) and *Cladophora* (OTU_0468). Fungi were very diverse (924 unique OTUs), but not abundant among sites (Fig. 3). Another major group was Peronosporomycetes (water molds), represented by 159 OTUs.

Besides diatoms and water molds, several groups of organisms that belonged to the Stramenopiles-Alveolates-Rhizaria (SAR) clade were very diverse and common in studied samples. Among these 1370 OTUs, which are grouped here in a category called "SAR_other" (Fig. 3), the most common (present in all 14 sites) were OTU_1600 assigned to *Rhogostoma*, a Rhizarian fish parasite, OTU_2088 *Holosticha*, a ciliate, and OTU_4724, a representative of parasitic coccidians Eimeriidae. The most abundant "SAR_other" sequences were OTU_1954, _1955 assigned to the genus of Xanthophytes (yellow-green algae) *Characiopsis* and OTU_2012 assigned to the *Pseudoellipsoidon*, a genus of Eustigmatophyte algae.

3.2. Among- and within site variation

There was a fair amount of variation in the proportions and absolute numbers of reads from major clades among field and laboratory replicate samples (Fig. 3). In some sites, such as Dunnfield Creek, where benthic assemblages were visibly heterogeneous in the field with visible patches of filamentous green algae and chrysophyte alga *Hydrurus*, the variation among replicates was particularly pronounced. At the same time, PERMANOVA results indicated that among-site variability of the entire eukaryotic assemblage and of its components significantly exceeded within-site variability (Table 2).

The highest R^2 values for diatoms, characterized both molecularly and morphologically, showed that they were distributed more evenly among sample replicates within sites compared to green algae, holozoans, fungi and water molds.

This pattern is also clearly seen in the NMDS ordinations (Figs. 4 and 5A, C) that visualize among and within-site variation in various subsets of data. In both molecular (Fig. 5A) and morphological (Fig. 5C) diatom datasets, there is tighter clustering of replicate samples within sites than in the other datasets (Fig. 4B-D). However, samples from Neldons Brook, where holozoans were the dominant group (Fig. 3B), were distinctly more tightly clustered relative to all other sites (Fig. 4C).

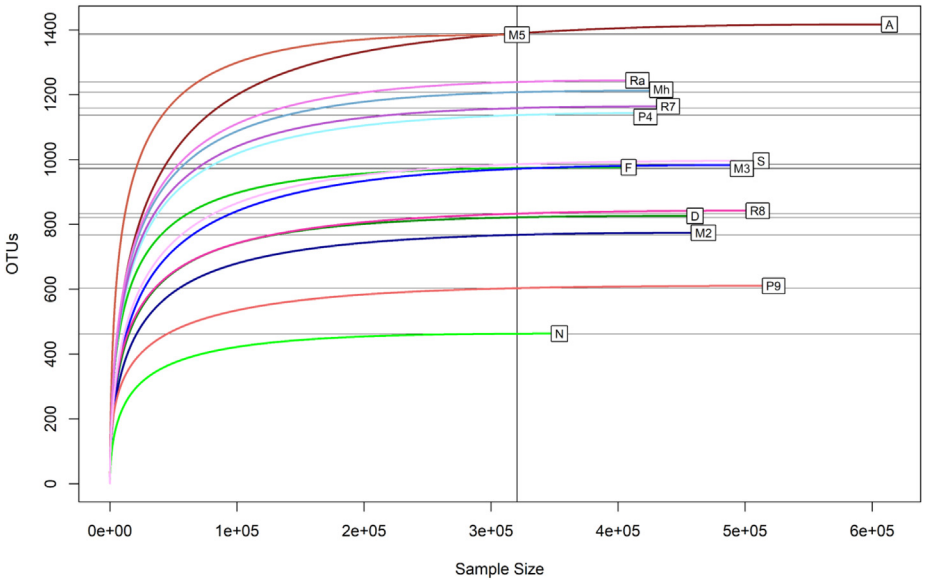


Fig. 2. Rarefaction curves of OTUs for individual sites. "Good" quality sites are represented by shades of green; "Fair" quality sites are shades of blue; "Poor" quality sites are shades of red. The vertical line represents the minimum expected species richness among all sites. Horizontal lines represent expected species richness in random subsamples of each site.

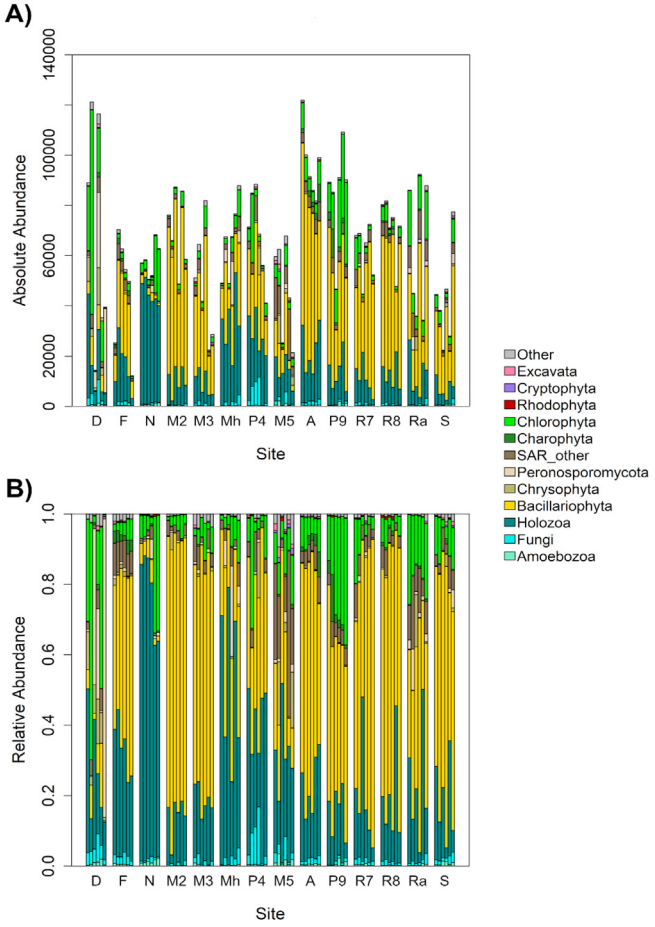


Fig. 3. Distribution of OTUs by major group. A: absolute abundance and B: relative abundance of OTUs. Each bar represents one lab replicate; each site has two lab replicates of three field replicates.

3.3. Distributional patterns of different groups of organisms

Nonmetric Multidimensional Scaling (NMDS) sample plots indicated that various components of benthic assemblages had somewhat similar, but far from identical, responses to underlying environmental gradients. Both algal groups included in the ordinations, green algae and diatoms, demonstrated much better separation of good water-quality sites from impaired sites (Fig. 4B, 5A, C), while holozoan (Fig. 4C) and water mold (Fig. 4D) assemblages were not as distinct among the high-quality sites. Long vectors for conductivity in all ordination plots indicated that ionic strength was strongly associated with the variability of all tested biofilm assemblages, while pH was apparently more strongly related to variation in green algae (Fig. 4B) than in other biofilm components. Procrustes symmetric rotations among ordinations (Table 3) showed that distributional patterns of fungi and water molds were the most similar and that there was also a strong correlation between diatom assemblage variation using molecular and morphological approaches. Since all correlations were significant ($P < 0.05$), there was a considerable similarity in the response to the environment among all assemblage components.

Table 2
Results of the PERMANOVA test for the significance of differences of biological assemblage composition among sites. Individual datasets represent relative abundance of 18S-V9 OTUs or diatom valves in the diatom count dataset. R^2 values show percent variation in distances explained by grouping samples by site.

Dataset	Df	Sums Of Squares	Mean Square	F	R^2	p-value
All eukaryotes	13	20.34	1.56	10.31	0.66	< 0.001
Bacillariophyta	13	21.87	1.68	18.85	0.78	< 0.001
Chlorophyta	13	20.56	1.58	13.38	0.71	< 0.001
Fungi	13	21.96	1.69	8.85	0.62	< 0.001
Holozoa	13	15.99	1.23	5.29	0.50	< 0.001
Peronosporomycota	13	18.68	1.44	6.04	0.53	< 0.001
Diatom count data	13	7.39	0.57	8.11	0.79	< 0.001

3.4. Response to impairment

Nonmetric Multidimensional Scaling sample plots revealed a clear response of all molecular and diatom morphospecies assemblages to water quality impairment. The separation of "Good" sites from the rest was especially clear in diatom (Fig. 5A, C) and green algal (Fig. 4B) datasets. Diatoms also strongly separated "Poor" from "Fair" sites (Fig. 5A, C). The holozoans demonstrated a good separation of "Poor" sites from the rest, while "Good" and "Fair" categories were not well separated from each other (Fig. 4C).

Most diatom OTUs with the highest fit to measured environmental

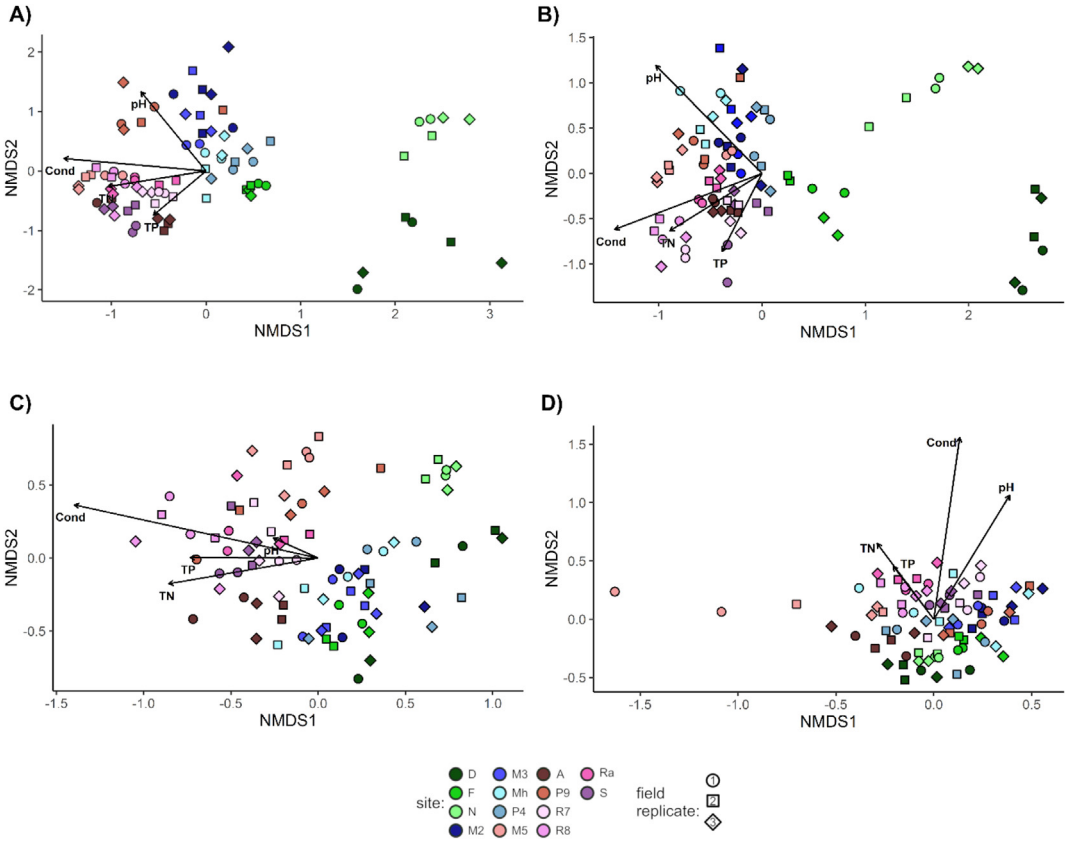


Fig. 4. NMDS ordination sample plots of OTU datasets with fitted vectors of environmental variables. A: all eukaryotes, B: Chlorophyta (green algae), C: Holozoa, D: Peronosporomycetes (water molds). "Good" quality sites are represented by shades of green; "Fair" quality sites are shades of blue; "Poor" quality sites are shades of red. Cond = Conductivity (μS/cm); TN = Total Nitrogen; TP = Total Phosphorus.

variables shown in the NMDS plot (Fig. 5B) could not be assigned with certainty to specific morphotaxa. The NMDS species plot of morphologically identified diatoms (Fig. 5D) shows that taxa with the best fit to measured water-quality parameters included indicators of high (*Halamphora coffeaformis, Surirella brebissonii* var. *kuetzingii*) and low (*Psammothidium subatomoides, Diatoma mesodon, Eunotia implicata, E. rhomboidea*) nutrient content. As expected, these indicators of low and high nutrients occupied opposite ends of the ordination space along the water-quality impairment gradient.

Canonical Correspondence Analysis (CCA) confirmed that the relationship between impairment and OTU assemblage was significant for all tested subsets of data and for the whole eukaryotic dataset (Table 4).

The F-values (Table 4) demonstrate that the top performing assemblages were diatoms (characterized by both metabarcoding and by morphology) and green algae, and, to a lesser extent, all eukaryotes together.

Indicator species analysis identified several 18S-V9 OTUs associated with good or poor groups of sites. The OTUs with the highest indicator values (IV > 0.98, P-value < 0.005) for the highly impaired sites belonged to diatoms: *Rhoicosphenia* (OTU_0026), *Surirella* (OTU_0303), *Cocconeis* (OTU_0081) and *Eolimna* (OTU_0176). As there were fewer high-quality sites, there were fewer OTUs with very high indicator values for those sites. Those with IV > 0.90 (P-value < 0.05) were a chironomid (possibly *Cricotopus*, OTU_2050), a nematode (possibly *Tridentulus,* OTU_2040), a green alga (OTU_0695) and two diatoms: *Achnanthidium* (OTU_0099) and *Sellaphora* (OTU_0128).

4. Discussion

4.1. What information on biotic diversity can be obtained from 18S-V9 metabarcoding data?

We found a considerable number of unique eukaryotic sequence variants (5866), but a direct comparison of the total number of OTUs with other metabarcoding surveys is difficult because of the differences in sequencing platforms, markers, laboratory and bioinformatics procedures. However, despite our relatively small sample size, the eukaryotic diversity we found is comparable to similar recent studies that used metabarcoding as a biomonitoring tool, particularly in sites impaired by anthropogenic pollution. Bricheux et al. (2013) found diatoms predominant in terms of abundance, and holozoans, diatoms, and chlorophytes accounted for the majority of eukaryotic diversity in biofilms from a single river site by using 454 pyrosequencing technology. The number of OTUs in their study was 4264, but it was based on multiple 18S and 16S markers and clustering sequences into OTUs at 97% similarity, while the current study used only one marker, the Illumina MiSeq sequencing platform, and retained ESVs rather than clustering sequences into OTUs. Another study that included sampling at four sites on a French river found 456 eukaryotic OTUs, obtained by clustering 18S V1-V2 sequences at 95% identity, using 454 pyrosequencing (Zancarini et al., 2017). Volant et al. (2016) detected a total of 323 eukaryotic OTUs in a creek affected by acid mine drainage using a single 18S V2-V3 marker, 454 pyrosequencing and clustering unique sequence variation at 97% similarity. Even though binning ESVs into traditional OTUs in these studies decreased the apparent alpha-diversity, it still appears as Illumina MiSeq sequencing is revealing higher eukaryotic diversity in stream biofilms compared to 454 pyrosequencing. Directly comparing cluster-based OTU counts, such as

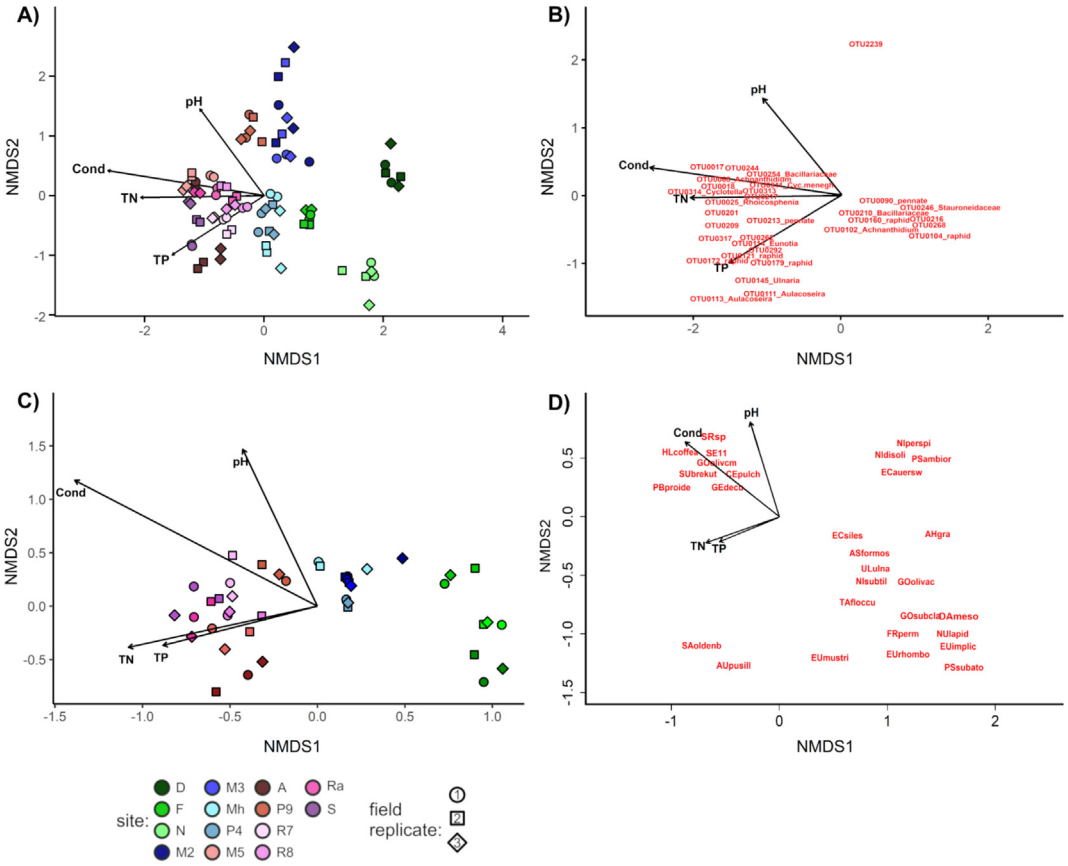


Fig. 5. NMDS ordinations of diatom molecular (A and B) and morphological (C and D) datasets with fitted vectors of environmental variables. Plots on the left (A and C) show position of individual samples in the ordination space. "Good" quality sites are represented by shades of green; "Fair" quality sites are shades of blue; "Poor" quality sites are shades of red. Plots on the right show positions of 18S-V9 OTUs (B) and morphotaxa (D). Only OTUs and taxa with the highest fit to environmental variables are shown; the short codes of taxa are listed in Supplementary Material files. Cond = Conductivity (μS/cm); TN = Total Nitrogen; TP = Total Phosphorus.

those found in previous investigations, with the ESV counts determined in the current study is difficult due to the nature of how the OTU and ESV sequences are identified and how read sequences are assigned membership (Callahan et al., 2017). Most similarity-based methods rely on sequence identity, i.e. the base-pair to base-pair correspondence between two or more read and/or reference sequences while the exact sequence variant approach used by 'dada2' incorporates read quality into its statistical model to determine the most likely parent (biological) read for each sequenced (observed) read (Callahan et al., 2016a). This improved approach is effective and sensitive enough that it can detect biological sequence variants that differ by only a single nucleotide (Callahan et al., 2016a, 2017, 2019). In addition to properly representing the true (biological) community structure within samples, refining the detail within OTU tables by using ESVs may improve our ability to detect the sources of fine scale variation, identify more specific or accurate indicator species or groups of taxa, better reproducibility of results, and allow investigators to simply combine the results

Table 4
The effect of impairment on the composition of biological assemblages tested by the Canonical Correspondence Analysis (CCA). The "impairment" factor was used as a single constraining variable. PVE1 (%) = percent variance explained by the first canonical axis. Assemblage composition is expressed as relative abundance of 18S-V9 OTUs in the molecular datasets of specific taxonomic groups or diatom valves in the diatom count dataset.

Dataset	PVE1 (%)	F-ratio	p-value
All eukaryotes	5.64	4.91	0.01
Bacillariophyta	9.84	8.94	0.01
Chlorophyta	8.04	7.17	0.01
Fungi	4.09	3.49	0.01
Holozoa	4.35	3.73	0.01
Peronosporomycota	3.41	2.89	0.01
Diatom count data	11.81	5.35	0.01

Table 3
Procrustes correlations among two-dimensional NMDS ordinations of 18S-V9 OTU datasets of major taxonomic groups and between diatom 18S-V9 OTUs and morphological count datasets. All correlations were significantly similar (p-value < 0.001), indicating similarity of responses of various biological assemblages to underlying environmental gradients.

	Bacillario-phyta	Chloro-phyta	Fungi	Holozoa	Peronosporo-mycota	Diatom count data
All eukaryotes	0.78	0.91	0.69	0.62	0.58	
Bacillariophyta		0.64	0.59	0.55	0.52	0.71
Chlorophyta			0.65	0.61	0.46	
Fungi				0.61	0.76	
Holozoa					0.53	

from different studies, provided that the same primer pairs are used (Callahan et al., 2017).

Like findings of most metabarcoding surveys of river biofilms, diatom OTUs were the most abundant eukaryotic OTUs in most of our sites. We found 343 unique diatom OTUs, ranging from 4 to 110 unique OTUs per sample, with an average of 52 unique OTUs per sample. In contrast, we identified a total of 234 morphospecies, with 13–65 unique morphospecies per sample with an average of 41 unique morphospecies per sample. In recent studies focused solely on investigating diatom communities, rbcL is largely preferred over 18S markers, as it was found to detect higher diversity (e.g. Keck et al., 2017; Mortágua et al., 2019; Rimet et al., 2018). However, these studies found similar ratios of morphologically versus molecularly characterized diatom diversity to the ratio that we found with 18S-V9. This may be partially due to clustering rbcL exact sequence variants into OTUs, while we did not do this because of the lower taxonomic resolution power of 18S-V9 marker compared to rbcL and other markers (Tang et al., 2012).

As in many other metabarcoding studies involving assemblages of protists (e.g. de Vargas et al., 2015), a large portion of sequences in our dataset could only be ascribed with certainty to higher-level taxa, while 14.2% of OTUs could only be identified as eukaryotes; this is not a limitation of molecular approaches themselves, but of the taxonomic extent and completeness of the taxonomic reference databases that are available for use. Less than one-sixth of the OTUs (972) had a perfect 100% match to GenBank sequences. It is difficult to tell with certainty whether even these perfect matches correspond to the assigned low-level taxa (species or genera) as other closely related taxa may have identical 18S-V9 sequences, therefore all taxonomic assignments to lower taxonomic levels listed in Supplementary Material 1 should be considered provisional. Some of the OTUs with 100% identity matched to sequences of organisms with ecological or geographical ranges of distribution clearly incompatible with New Jersey rivers. The examples include marine organisms, such as diatoms *Tursiocola* and *Licmosphenia*, dinoflagellate *Scrippsiella* and ciliates *Cyclidium* and *Parallelostrombidium*. The low proportion of sequences with assigned taxonomy is by no means a limitation of 18S-V9 marker, but rather reflects the current lack of coverage of many eukaryotic lineages in the existing reference databases. A number of reference libraries are being specifically developed for the purposes of barcoding as it becomes a common tool for biodiversity research and biomonitoring (e.g., Carew et al., 2017; Guillou et al., 2013; Nilsson et al., 2019; Rimet et al., 2015, 2019; Zhou et al., 2013). Unlike NCBI GenBank, such databases are typically well-curated, but often either cover limited number of taxa or markers other than 18S gene or are focused on particular habitats or geographic areas. Here, we used SILVA and GenBank for the sake of fair comparison between lineages, but further exploration of our data is certainly possible with better curated reference databases targeting specific taxonomic or ecophysiological groups of organisms.

4.2. Reproducibility

In this study, metabarcoding was determined to be a reliable method of detecting a unique molecular fingerprint for each of our study sites, as evidenced by significant among-site variability of OTU assemblages among lab and field replicates using PERMANOVA (Table 2). Among individual sites, all tested assemblages (all eukaryotes together, diatoms, holozoans, water molds, green algae and fungi) were significantly different from one another. This indicated adequate quantitative assessment of the abundance and diversity of taxa, as well as an appropriate level of reproducibility using the sampling scheme employed in this study (three field and two lab replicates per site). A considerable amount of random variation in the detection of OTUs, or in the estimates of their abundance, has been reported in several metabarcoding studies (Leray and Knowlton, 2017; Zhan et al., 2014; Zhou et al., 2011; Wolf, 2018). Likewise, our results demonstrate large variation among field and lab replicates (Fig. 2), but among-site differences still strongly responded to effects of water-quality impairment. The importance of multiple replicates obtained at different stages of metabarcoding has been previously emphasized (Leray and Knowlton, 2017) and a thorough assessment of reproducibility and the number of required replicates should be a necessary step in the development of standard biomonitoring procedures involving metabarcoding.

The results of our PERMANOVA analysis (Table 2) indicate that diatoms and green algae were more equally distributed among replicates than other eukaryotes, while holozoans had the highest variability. This was expected as holozoans are larger multicellular organisms and sampling three rocks per site is, of course, insufficient to thoroughly characterize their assemblages. Many larger holozoans are not functional members of biofilms, so excluding their sequences from the assessment of biofilm assemblages may be an option in the future biomonitoring efforts as was done, for example, by Volant et al. (2016).

4.3. Detecting impairment

The results of this study indicate that water-quality impairment had a significant effect on the structure of eukaryotic biofilm assemblages characterized by 18S-V9 DNA metabarcoding. All tested components of the assemblages (diatoms, green algae, holozoans, fungi and water molds) showed statistically significant association with impairment, but their specific responses to the environmental variation differed in strength and direction. Diatom and green algae OTUs showed a stronger relationship to impairment than other tested groups of eukaryotes (Table 4). Diatoms were especially effective in detecting impairment with both molecularly and morphologically characterized assemblages, clearly separating clusters of minimally, moderately and heavily impacted sites (Fig. 5A, C). This strong relationship may be at least partially attributed to the way the impairment was characterized in this study. Freshwater algae are known to be especially sensitive to water-quality characteristics, such as conductivity, pH and nutrients (Stevenson, 2014) that we used to quantify impairment, while other organisms are likely more sensitive to other factors. For example, fungi and fungi-like protists such as water molds may better respond to the amount and quality of dissolved and particulate organic carbon (Bai et al., 2018), while macroinvertebrates are primarily influenced by water oxygen content, temperature, and physical characteristics of their habitat (De Pauw et al., 2006; Resh, 2008). If we characterized impairment using another suite of environmental characteristics, we could probably observe a stronger response from non-algal groups to other stressors.

The relative strength of the response of various taxonomic groups of organisms to human impacts also depends on the environmental context. Diatoms are extremely sensitive to variations in the ionic composition of fresh water (Potapova and Charles, 2003) and therefore respond especially strongly to an impairment that increases water mineral content as is the case with New Jersey streams. In other types of environment different groups may be the most strongly associated with impairment. For example, a similar comparison of the distribution of various protestant groups characterized by 18S-V9 metabarcoding in marine sediments revealed ciliates as the most powerful indicators of the impairment (Stoeck et al., 2018), while diatoms and chrysophytes had a weaker response.

4.4. Potential use of 18S-V9 DNA metabarcoding for biomonitoring

An obvious advantage of 18S-V9 metabarcoding is its potential use for simultaneously characterizing multiple groups of eukaryotic organisms in a cost-effective way (Hadziavdic et al., 2014). Combining this ability of sampling and analyzing assemblage structure of various groups of organisms with knowledge of their responses to specific stressors may greatly enhance bioassessments. A larger-scale study covering sites impaired in different ways and a standardization of the field and laboratory procedures would be necessary to develop

metabarcoding-based metrics and indices for practical use. However, even our pilot project was sufficient to demonstrate the power of this approach to detect the effects of impairment on biota.

Although chloroplast genes are often suggested as metabarcoding markers for riverine benthos, their use is limited to the photosynthetic portion of the overall assemblage. For example, Kelly et al. (2018) advocated for using a fragment of rbcL gene for amplicon sequencing of benthic diatoms, arguing that a protein-coding gene is free from intraspecific variability and its species-discriminating ability closely matches that of a microscopist identifying diatoms. Several studies found higher diversity of diatoms recovered by rbcL markers as compared to markers using other gene fragments (Evans et al., 2007; Hamsher et al., 2011; Kermarrec et al., 2013). Another marker, the Universal Plastid Amplicon (UPA), targeting a fragment of the 23S chloroplast gene, was developed by Sherwood and Presting (2007) and used for metabarcoding of airborne algae in Hawaii (Sherwood et al., 2017) and assessing the effects of acid mine drainage on stream algae in Ohio (Wolf, 2018).

Ribosomal DNA markers can be used for amplifying multiple groups of organisms. They have been extensively used for profiling prokaryotic communities and can also be used to investigate eukaryotes (e.g. Pawlowski et al., 2012; de Vargas et al., 2015; Volant et al., 2016). However, in the majority of metabarcoding studies of eukaryotic assemblages, 18S markers targeting individual phylogenetic lineages have been employed. For example, fungal-specific ITS primers (Blaalid et al., 2013) and diatom specific 18S-V4 primers (Zimmermann et al., 2011, 2015) are widely used, and reference libraries were constructed to support their application to metabarcoding (Zimmermann et al., 2014). Several studies compared the effectiveness of various hypervariable regions of the 18S gene to characterize biotic diversity and community patterns. For example, comparisons of the V4 and V9 regions revealed an approximately similar performance in characterizing marine protistan assemblages (Piredda et al., 2017; Tragin et al., 2018), while other authors recommended either longer V4 region as a preferred marker for detecting eukaryotic diversity (Pawlowski et al., 2012) or the V9 region for its ability to better capture diversity and community structure of photosynthetic eukaryotes (Bradley et al., 2016).

Although the taxonomic resolution of the 18S-V9 marker is lower than that of the longer markers, we still obtained more diatom molecular OTUs (343) than morphologically identified morphospecies (234). Even assuming that some of these OTUs may represent the same biological species, this taxonomic resolution does not seem to be much reduced in comparison to morphology. Moreover, there is no guarantee that morphologically defined entities correspond to biological species. Both molecular and morphological approaches provide unique, but not ideal representation of the structure of biological assemblages and may be considered complementary measures. Taxonomic assignments are still provisional in most metabarcoding studies, including ours, as sequence reference databases are sparsely populated (Cowart et al., 2015). Despite the rigorous quality assurance used to create reference databases such as R-syst::diatom/Diat.barcode (Rimet et al., 2015, 2019), they are ultimately based on a morphological species concept. It is unclear, for example, as to whether rDNA reference databases address the issues of cryptic and pseudocryptic species and infraspecfic or intragenomic variation. We did not intend to find congruency between molecular and morphological methods, but rather explore metabarcoding as a complementary approach to water quality monitoring. We suggest further investigating a "taxonomy-free" approach in order to incorporate more OTUs in an index calculation (Apothéloz-Perret-Gentil et al., 2017), using OTU assemblages as "fingerprints" to characterize and detect changes in water quality. Some of the diatom OTUs that were identified in this study as indicative of highly impaired sites belonged to genera mostly associated with poor water quality (*Rhoicosphenia, Surirella, Eolimna*), while others, such as species of the genus *Cocconeis* have very wide tolerance to water quality characteristics. Morphologically defined taxa with wide ecological niches may represent cryptic or semi-cryptic species complexes and therefore it could be advantageous use molecular OTUs rather than morphotaxa for constructing water-quality metrics.

One problem with matching 18S-V9 sequences to GenBank entries often seems to be due to missing approximately 20–30 nucleotides at the 3′ end of accessioned 18S sequences. Obtaining sequences of even the most common microbial organisms in each geographic region is a daunting task and will take an indefinitely long time unless there is considerable funding available. But, if the primary data (raw unprocessed reads or processed exact sequence variants) are deposited in public data repositories, they can be matched to biological taxa whenever correspondence between taxa and sequences is established in the future. The results of our study agree with conclusions of Pawlowski et al. (2016, 2018) in that precise taxonomic assignments are not absolutely necessary for studying patterns of biotic assemblages and for using metabarcoding-based biomonitoring.

The most important aspect of biomonitoring is observing a strong response of indicators to impairment using a comparatively cost-effective method. The EMP one-step PCR protocol provides molecular signatures of eukaryotic assemblages at a fraction of the cost of two-step PCR methods commonly used with other genetic markers. The per-sample cost of metabarcoding progressively decreases as the number of samples pooled in a single library increases (Stein et al., 2014). Using the EMP one-step library preparation method opens opportunities for dramatically increasing the number of samples collected in the field and scaling up of the monitoring programs in a standardized fashion. Taxonomists' efforts could be redirected from routine processing of a large number of samples to high-quality morphological identification of selected organisms or enumeration of small fraction of samples for establishing correspondence between molecular signatures of the assemblages and morphotaxa.

One shortcoming of using metabarcoding for biomonitoring is the gross lack of correspondence between numbers of organisms and the number of reads in the metabarcoding data. This is usually a consequence of larger organisms having more gene copies compared to smaller ones (Elbrecht et al., 2017) or variation in copy numbers among phylogenetic lineages (Angly et al., 2014). A number of standardization methods have been suggested to overcome this problem. For example, Vasselon et al. (2017) related gene copy numbers to diatom cell biovolume using qPCR and applied the resulting correction factor to biofilm metabarcoding data to improve estimation of cell numbers. Several tools have been developed based on genome sequencing to correct for variation in 16S gene copy number in prokaryotes (Louca et al., 2018). However, the correspondence between gene copy numbers is only important in biomonitoring if metabarcoding data are identified and used in metric and index calculations in the same way as morphological data. If metabarcoding data are related to environmental gradients without reference to taxonomy as in this study, and the response to impairment is obvious without any standardization by cell/organism size or gene copy numbers, this step is not essential for monitoring purposes. One way the PCR bias may negatively influence the accuracy of metabarcoding-based assessment is the probability of missing small-sized organisms. The drawback of morphology-based assessments is however an opposite tendency to overlook larger but sparsely species, which emphasizes complementarity of the two approaches.

In conclusion, this study found a strong response of the eukaryotic assemblages of stream biofilms characterized by 18S-V9 DNA metabarcoding to water-quality impairment. Given the efficiency of the EMP metabarcoding protocol, this approach may be recommended for monitoring stream eukaryotic biota. While we are not suggesting a complete replacement of morphology-based biomonitoring with DNA metabarcoding, it should be possible to integrate two approaches to produce cost-effective and informative assessments.

CRediT authorship contribution statement

Alison D. Minerovic: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Marina G. Potapova:** Project administration, Supervision, Visualization, Data Curation, Investigation, Formal analysis, Validation, Software, Conceptualization. **Christopher M. Sales:** Resources. **Jacob R. Price:** Software, Resources. **Mihaela D. Enache:** Resources, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding for this study was provided by the Division of Science and Research, New Jersey Department of Environmental Protection through Agency Order number 48106000731.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2020.106225.

References

Abad, D., Albaina, A., Aguirre, M., Laza-Martínez, A., Uriarte, I., Iriarte, A., et al., 2016. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Mar. Biol. 163 (7), 1–13.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., Huse, S.M., 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subuinit ribosomal RNA genes. PLoS ONE 4 (7), e6372. https://doi.org/10.1371/journal.pone.0006372.

Angly, F.E., Dennis, P.G., Skarshewski, A., Vanwonterghem, I., Hugenholtz, P., Tyson, G.W., 2014. CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. Microbiome 2 (1), 11. https://doi.org/10.1186/2049-2618-2-11.

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., Pawlowski, J., 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Mol. Ecol. Resour. 17, 1231–1242. https://doi.org/10.1111/1755-0998.12668.

Bai, Y., Qiaojuan, W., Kailingli, L., Zhiyu, J., Chen, Z., Jiuhui, Q., 2018. Fungal community as a bioindicator to reflect anthropogenic activities in a river ecosystem. Front. Microbiol. 9, 3152. https://doi.org/10.3389/fmicb.2018.03152.

Bailet, B., Bouchez, A., Franc, A., Frigerio, J.-M., Keck, F., Karjalainen, S.-M., Rimet, F., Schneider, S., Kahlert, M., 2019. Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. Metabarcoding and Metagenomics 3, 21–25.

Barbour, M.T., Gerritsen, J., Snyder, B.D., Stribling, J.B., 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, second edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Bik, H.M., Sung, W., De Ley, P., Baldwin, J.G., Sharma, J., Rocha-Olivares, A., Thomas, W.K., 2012. Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. Mol. Ecol. 21 (5), 1048–1059.

Blaalid, R., Kumar, S., Nilsson, R.H., Abarenkov, K., Kirk, P.M. & Kauserud, H. 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. Molecular Ecology Resources 13(2): 218-24. pmid: 23350562.

Bradley, I.M., Pinto, A.J., Guest, J.S., 2016. Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. Appl. Environ. Microbiol. 82, 5878–5891. https://doi.org/10.1128/AEM.01630-16.

Brannock, P.M., Ortmann, A.C., Moss, A.G., Halanych, K.M., 2016. Metabarcoding reveals environmental factors influencing spatio-temporal variation in pelagic micro-eukaryotes. Mol. Ecol. 25, 3593–3604.

Bricheux, G., Morin, L., Le Moal, G., Coffee, G., Balestrino, D., Charbonnel, N., Bohatier, J., Forestier, C., 2013. Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river. Microbiol. Open 2, 402–414. https://doi.org/10.1002/mbo3.80.

Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., et al., 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. gkz569. Nucl. Acids Res. https://doi.org/10.1093/nar/gkz569.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639–2643. https://doi.org/10.1038/ismej.2017.119.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., Holmes, S.P., 2016a. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583.

Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P., 2016b. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. F1000Research 5: 1492.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6 (8), 1621–1624.

Carew, M.E., Nichols, S.J., Batovska, J., St Clair, R., Murphy, N.P., Blacket, M.J., Shackleton, M.E., 2017. A DNA barcode database of Australia's freshwater macroinvertebrate fauna. Mar. Freshw. Res. 68, 1788–1802.

Chariton, A.A., Court, L.N., Hartley, D.M., Colloff, M.J., Hardy, C.M., 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. Front. Ecol. Environ. 8, 233–238.

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., Pawlowski, J., 2019. Embracing environmental genomics and machine learning for routine biomonitoring. Trends Microbiol. 27(5), 387–397. doi: 10.1016/j.tim.2018.10.01.

Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PLoS ONE 10 (2), e0117562. https://doi.org/10.1371/journal. pone.0117562.

De Caceres, M., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90 (12), 3566–3574.

De Pauw, N., Gabriels, W., Goethals, P.L.M., 2006. River monitoring and assessment methods based on macroinvertebrates. In: Ziglio, G., Siligardi, M., Flaim, G. (Eds.), Biological Monitoring of Rivers. Wiley, Chichester, pp. 113–134.

De Vargas, C.S., Audic, A.N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., et al., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science 348, 1261605a.

Dufrene, M., Legendre, F., 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol. Monogr. 67 (3), 345–366.

Elbrecht, V., Peinert, B., Leese, F., 2017. Sorting things out: assessing effects of unequal specimen biomass on DNA metabarcoding. Ecol. Evol. 7, 6918–6926.

Evans, K.M., Wortley, A., Mann, D.G., 2007. An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). Protist 158, 349–364.

Filker, S., Sommaruga, R., Vila, I., Stoeck, T., 2016. Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. Mol. Ecol. 25, 2286–2301.

Gibson, J.F., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I. & Hajibabaei, M. 2015. Large-scale biomonitoring of remote and threatened ecosystems via High-Throughput Sequencing. PLoS ONE 10(10): e0138432. doi: 1371/journal.pone.0138432.

Gilbert, J.A., Meyer, F., Jansson, J., Gordon, J., Pace, N., Tiedje, J., et al., 2010. The Earth Microbiome Project: meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6, 2010. Standards in Genomic Sciences 3: 249–253.

Glassman, S.I., Martiny, B.H., 2018. Broadscale ecological patterns are robust to use of Exact Sequence Variants versus Operational Taxonomic Units. mSphere 3: 1-5. doi: 10.1128/mSphere.00148-18.

Goral, F., Schellenberg, J., 2018. Goeveg: Functions for community data and ordinations. R package version 0.4.2. https://CRAN.R-project.org/package = goeveg.

Grattepanche, J., Santoferrara, L.F., McManus, G.B., Katz, L., 2016. Unexpected biodiversity of ciliates in marine samples from below the photic zone. Mol. Ecol. 25, 3987–4000.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahe, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small subunit rRNA sequences with curated taxonomy. Nucl. Acids Res. D597–D604.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., Troedsson, C., 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PLoS ONE 9 (2), e87624. https://doi.org/10.1371/journal.pone.0087624.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. Environmental barcoding: a Next-Generation Sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6 (4), e17497. https://doi.org/10.1371/journal.pone.0017497.

Hamsher, S.E., Evans, K.M., Mann, D.G., Poulickova, A., Saunders, G.W., 2011. Barcoding diatoms: exploring alternatives to COI-5P. Protist 162, 405–422.

Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., Kelly, M., 2018. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Resour. 138, 192–205. https://doi.org/10.1016/j.watres.2018.03.003.

Hofmann, G., Lange-Bertalot, H., Werum, M., 2013. Diatomeen im Süßwasser-Benthos von Mitteleuropa: 2 Corrected Edition Koeltz Scientific Books, Königstein, 908 pp.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., et al., 2013.

Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecol. Lett. 16 (10), 1245–1257.
Keck, F., Vasselon, V., Tapolczai, K., Rimet, F., Bouchez, A., 2017. Freshwater biomonitoring in the information age. Front. Ecol. Environ. 15 (5), 266–274.
Kelly, M., Boonham, N., Juggins, S., Kille, P., Mann, D., Pass, D., Sapp, M., Sato, S., Glover, R., 2018. A DNA-based diatom metabarcoding approach for Water Framework Directive classification of rivers. Environmental Agency Report SC140024/R. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684493/A_DNA_based_metabarcoding_approach_to_assess_diatom_communities_in_rivers_-_report.pdf.
Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J.M., Humbert, J.F., Bouchez, A., 2014. A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshwater Sci. 33, 349–363.
Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Humbert, J.F., Bouchez, A., 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. Mol. Ecol. Resour. 1, 607–619.
Krammer, K., Lange-Bertalot, H., 1991a. Bacillariophyceae. 3. Teil: Centrales, Fragilariaceae, Eunotiaceae. in Ettl, H., Gerloff, J., Heynig, H. and Mollenhauer, D. (eds) Süsswasserflora von Mitteleuropa, Band 2/3. Gustav Fischer Verlag: Stuttgart, Jena. 576 pp.
Krammer, K., Lange-Bertalot, H., 1991b. Bacillariophyceae. 4. Teil: Achnanthaceae, Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema, Gesamtliteraturverzeichnis Teil 1-4. in Ettl, H., Gärtner, G., Gerloff, J., Heynig, H. and Mollenhauer, D. (eds) Süsswasserflora von Mitteleuropa, Band 2/4. Gustav Fischer Verlag: Stuttgart, Jena. 437 pp.
Krammer, K., Lange-Bertalot, H., 1988. Bacillariophyceae. 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae. in Ettl, H., Gerloff, J., Heynig, H. and Mollenhauer, D. (eds) Süsswasserflora von Mitteleuropa, Band 2/2. VEB Gustav Fischer Verlag: Jena. 596 pp.
Krammer, K., Lange-Bertalot, H., 1986. Bacillariophyceae. 1. Teil: Naviculaceae In: Ettl, H., J. Gerloff, H. Heynig and D. Mollenhauer (eds.) Süsswasserflora von Mitteleuropa, Band 2/1. Gustav Fisher Verlag, Jena. 876 pp.
Leray, M., Knowlton, N., 2017. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ 5, e3006. https://doi.org/10.7717/peerj.3006.
Lindeque, P.K., Parry, H.E., Harmer, R.A., Somerfield, P.J., Atkinson, A., 2013. Next Generation Sequencing reveals the hidden diversity of zooplankton assemblages. PLoS ONE 8 (11), e81327. https://doi.org/10.1371/journal.pone.0081327.
Louca, S., Doebeli, M., Parfrey, L.W., 2018. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome 6, 41. https://doi.org/10.1186/s40168-018-0420-9.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17 (10), 10–12.
Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A., Bowler, C., 2016. Insights into global diatom distribution and diversity in the world's ocean. Proc. Natl. Acad. Sci. 113 (11), e1516–e1525.
McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8 (4), e61217.
Meusnier, I., Singer, G.A., Landry, J.F., Hickey, D.A., Hebert, P.D.N., et al., 2008. A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9, 214.
Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., João Feio, M., Almeida, S.F.P., 2019. Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. Ecol. Ind. 106, 105470. https://doi.org/10.1016/j.ecolind.2019.105470.
Moulton, S.R., Kennen, J.G., Goldstein, R.M. & Hambrook, J.A. 2002. Revised protocols for sampling algal, invertebrate and fish communities as part of the National Water Quality Assessment Program. Open-File Report 02-150. U.S. Geological Survey, Reston, VA. 75 pp.
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2018. Vegan: Community ecology package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan.
Nilsson, R.H., Larsson, K.H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., Saar, I., Kõljalg, U., Abarenkov, K., 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucl. Acids Res. 47 (D1), D259–D264. https://doi.org/10.1093/nar/gky1022.
Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., et al., 2012. CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol. 10 (11), e1001419. https://doi.org/10.1371/journal.pbio.1001419.
Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., et al., 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637–638, 1295–1310.
Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., Esling, P., 2016. Protist metabarcoding and environmental biomonitoring: time for change. Eur. J. Protistol. 55A, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003.
Piredda, R., Tomasino, M.P., D'Erchia, A.M., Manzari, C., Pesole, G., Montresor, M., et al., 2017. Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long-Term Ecological Research site. FEMS Microbiol. Ecol. 93, fiw200. https://doi.org/10.1093/femsec/fiw200.
Ponader, K.C., Charles, D.F., Belton, T.J., 2007. Diatom-based TP and TN inference models and indices for monitoring nutrient enrichment of New Jersey streams. Ecol. Ind. 7, 79–93.
Potapova, M., Charles, D.F., 2003. Distribution of benthic diatoms in U.S. rivers in relation to conductivity and ionic composition. Freshw. Biol. 48, 1311–1328. https://doi.org/10.1046/j.1365-2427.2003.01080.x.
Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucl. Acids Res. 35, 7188–7196.
R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rachik, S., Christaki, U., Li, L.L., Genitsaris, S., Breton, E., Monchy, S., 2018. Diversity and potential activity patterns of planktonic eukaryotic microbes in a mesoeutrophic coastal area (eastern English Channel). PLoS ONE 13 (5), e0196987.
Resh, V.H., 2008. Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. Environ. Monit. Assess. 138 (1–3), 131–138.
Rimet, F., Chaumeil, P., Keck, F., Kermarrec, L., Vasselon, V., Kahlert, M., Franc, A., Bouchez, A., 2015. R-Syst::diatom: a barcode database for diatoms and freshwater biomonitoring – data sources and curation procedure. INRA Report, 14 p. http://www.rsyst.inra.fr/.
Rimet, F., Gusev, E., Kahlert, M., Kelly, M.G., Kulikovskiy, M., Maltsev, Y., Mann, D.G., Pfannkuchen, M., Trobajo, R., Vasselon, V., Zimmermann, J., Bouchez, A., 2019. Diat.barcode, an open-access curated barcode library for diatoms. Sci. Rep. 9, 15116. https://doi.org/10.1038/s41598-019-51500-6.
Rimet, F., Vasselon, V., Keszte, B., Bouchez, A., 2018. Do we similarly assess diversity with microscopy and high-throughput sequencing? Case of microalgae in lakes. Org. Divers. Evol. 18, 51.
Rivera, S., Vasselon, V., Jacquet, S., Ariztegui, D., Bouchez, A., Rimet, F., 2018. Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment. Hydrobiologia 807, 37–51.
Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., et al. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci., 109(16): 6241-6246. pmid:22454494; PubMed Central PMCID: PMC3341068.
Sherwood, A.R., Dittbern, M.N., Johnston, E.T., Conklin, K.Y., 2017. A metabarcoding comparison of windward and leeward airborne algal diversity across the Ko'Olau Mountain Range on the island of O'Ahu, Hawai'i. J. Phycol. 53, 437–445.
Sherwood, A.R., Presting, G.G., 2007. Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. J. Phycol. 43, 605–608.
Spaulding, S.A., Bishop, I.W., Edlund, M.B., Lee, S., Potapova, M. 2018. Diatoms of North America. https://diatoms.org/.
Stein, E.D., Martinez, M.C., Stiles, S., Miller, P.E., Zakharov, E.V., 2014. Is DNA Barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? PLoS ONE 9 (4), e95525. https://doi.org/10.1371/journal.pone.0095525.
Stevenson, J., 2014. Ecological assessments with algae: a review and synthesis. J. Phycol. 50 (3), 437–461.
Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., Richards, T.A., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. 19 (Suppl. 1), 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x.
Stoeck, T., Kochems, R., Forster, D., Lejzerowicz, F., Pawlowski, J., 2018. Metabarcoding of benthic ciliate communities shows high potential for environmental monitoring in salmon aquaculture. Ecol. Ind. 85, 153–164. https://doi.org/10.1016/j.ecolind.2017.10.041.
Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willersley, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21 (8), 2045–2050.
Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., Fontaneto, D., 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proc. Natl. Acad. Sci. 109 (40), 16208–16212.
Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Jansson, J.K., Gilbert, J.A., Knight, R., The Earth Microbiome Project Consortium, 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457–463.
Tragin, M., Zingone, A., Vaulot, D., 2018. Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. Environ. Microbiol. 20 (2), 506–520.
Vanormelingen, P., Evans, K.M., Chepurnov, V.A., Vyverman, W., Mann, D.G., 2013. Molecular species discovery in the diatom *Sellaphora* and its congruence with mating trials. Fottea 13 (2), 133–148.
Vasselon, V., Rimet, F., Tapolczai, K., Bouchez, A., 2017. Assessing ecological status with diatoms DNA metabarcoding: scaling-up on a WFD monitoring network (Mayotte island, France). Ecol. Ind. 82, 1–12.
Visco, J.A., Apothéloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillet, L., Pawlowski, J., 2015. Environmental monitoring: inferring the Diatom Index from Next-Generation Sequencing data. Environ. Sci. Technol. 49 (13), 7597–7605.
Volant, A., Héry, M., Desoeuvre, A., Casiot, C., Morin, G., Bertin, P.N., et al., 2016. Spatial distribution of eukaryotic communities using high-throughput sequencing along a pollution gradient in the arsenic-rich creek sediments of Carnoulès Mine, France. Microb. Ecol. 72, 608–620.
Wang, J., Wang, F., Chu, L., Wang, H., Zhong, Z., et al., 2014. High genetic diversity and novelty in eukaryotic plankton assemblages inhabiting saline lakes in the Qaidam Basin. PLoS ONE 9 (11), e112812.
Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ.

Microbiol. 73 (16), 5261–5267.
Wickham, H., 2016. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
Wolf, D., 2018. Multi-marker metabarcoding assessment of biodiversity within stream biofilm communities along an acid mine drainage recovery gradient. Electronic Thesis or Dissertation. Ohio University. https://etd.ohiolink.edu/.
Xu, D., Li, R., Hu, C., Sun, P., Jiao, N., Warren, A., 2017. Microbial eukaryote diversity and activity in the water column of the South China Sea based on DNA and RNA high throughput sequencing. Front. Microbiol. 8, 1121.
Zancarini, A., Echenique-Subiabre, I., Debroas, D., Taïb, N., Quiblier, C., Humbert, J., 2017. Deciphering biodiversity and interactions between bacteria and micro-eukaryotes within epilithic biofilms from the Loue River, France. Scientific Reports 7, 4344.
Zhan, A., Hulak, M., Sylvester, F., Huang, X., Adebayo, A.A., Abbott, C.L., Adamowicz, S.J., Heath, D.D., et al., 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods Ecol. Evol. 4, 558–565.
Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7 (1–2), 203–214.
Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y., Tu, Q., Xie, J., Van Nostrand, J.D., He, Z., Yang, Y., 2011. Reproducibility and quantitation of amplicon sequencing-based detection. ISME J. 5, 1303–1313.
Zimmermann, J., Abarca, N., Enke, N., Skibbe, O., Kusber, W.-H., Jahn, R., 2014. Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. PLoS ONE 9, e108793.
Zimmermann, J., Glöckner, G., Jahn, R., Enke, N., Gemeinholzer, B., 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Mol. Ecol. Resour. 15, 526–542.
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.