Freshwater diatom biomonitoring through benthic kick-net metabarcoding

Short title: Freshwater diatom biomonitoring through metabarcoding

Victoria **Carley** Maitland[1], Chloe Victoria Robinson[1], Teresita M. Porter[1,2], Mehrdad Hajibabaei[1]*.

[1]Centre for Biodiversity Genomics & Department of Integrative Biology, University of Guelph, Guelph, Ontario, N1G 2W1

[2]Great Lakes Forestry Centre, Natural Resources Canada, 1219 Queen Street East, Sault Ste. Marie, ON Canada

* Corresponding author

Email: mhajibab@uoguelph.ca

Abstract

Biomonitoring is an essential tool for assessing ecological conditions and informing management strategies. The application of DNA metabarcoding and high throughput sequencing has improved data quantity and resolution for biomonitoring of taxa such as macroinvertebrates, yet, there remains the need to optimise these methods for other taxonomic groups. Diatoms have a longstanding history in freshwater biomonitoring as bioindicators of water quality status. However, periphyton scraping, a common diatom sampling practice, is time-consuming and thus costly in terms of labour. This study examined whether the benthic kick-net technique used for macroinvertebrate biomonitoring could be applied to bulk-sample diatoms for metabarcoding. To test this approach, we collected samples using both conventional microhabitat periphyton scraping and bulk-tissue kick-net methodologies in parallel from replicated sites with different habitat status (good/fair). We found there was no significant difference in community assemblages between conventional periphyton scraping and kick-net methodologies, but there was significant difference between diatom communities depending on site quality (*P* = 0.029). These results show the diatom taxonomic coverage achieved through DNA metabarcoding of kick-net is suitable for ecological biomonitoring applications. The shift to a more robust sampling approach and capturing diatoms and macroinvertebrates in a single sampling event has the potential to significantly improve efficiency of biomonitoring programmes.

Key words: Biomonitoring, metabarcoding, periphyton, diatom, benthos, biodiversity, bioindicator, water quality, rbcL cpDNA, kick-net

Introduction

As climate change and other anthropogenic impacts continue to alter the environment, there is an increasing need for comprehensive ecological assessment. Rapid and robust biomonitoring is essential for informing management plans and mitigating further environmental degradation [1–3]. Freshwater biomonitoring typically involves sampling a range of aquatic taxa, with particular focus on biological indicator taxa, to assess environmental conditions based on diversity, richness, structure and function of the existing communities [3–5].

Traditionally, biomonitoring data is generated through morphological taxonomic classifications, however there has been a recent shift towards DNA-based identification using metabarcoding [6] coupled with high throughput sequencing [7]. In aquatic systems such as wadable streams, a combination of bulk-tissue benthic sampling using kick-net methodology with DNA metabarcoding, facilitates rapid data collection whilst maintaining data integrity [8–10]. The metabarcoding approach has been employed for numerous biomonitoring studies involving macroinvertebrates [11,12] for assessing freshwater health [5,10,13].

In addition to benthic macroinvertebrates, diatoms (members of Bacillariophyta) are also ideal biomonitoring target taxa for assessing freshwater system conditions [14–16]. These single-celled algae have a short generation time which allows for rapid responses to physical, chemical and biological changes in the environment [14,15,17]. Similar to macroinvertebrates, the high diversity and ubiquity of diatoms is used to create biotic indices that can accurately report freshwater quality [16,18,19]. Studies have shown that diatoms respond more readily to the presence of heavy metal pollutants compared to macroinvertebrates, which are generally more sensitive to shifts in hydrological conditions [17,20–22]. Monitoring only one of these taxonomic groups to assess overall ecosystem health could potentially cause gaps in knowledge that could subvert subsequent management strategies. Hence, diatoms are being used in a number of national and regional biomonitoring programmes.

Current methods for diatom sampling are time-consuming and laborious, which could hamper widespread use of diatoms for extensive freshwater biomonitoring [23,24]. The conventional diatom collection method involves the scraping of periphyton (a combination of algae, cyanobacteria, microbes, and detritus) from numerous substrates within littoral habitats [23–26]. These samples are then fixed and visualised using light microscopy [27–30]. From here, microscopy standards and keys are followed [29–31] to enable identification of diatoms to different taxonomic ranks. Within recent years, there has been the shift towards DNA metabarcoding-based identification of diatoms [15,16,32,33]. This involves the manual homogenized of periphyton scrapings into single samples, which are then processed via standard diatom metabarcoding procedures [34,35]. Alternative sampling methods, such as collection through the benthic kick-net technique, have not been tested for diatom biomonitoring applicability, however it is expected that this technique would drastically reduce time spent collecting samples. The ability to study diatom and macroinvertebrate assemblages from a single sample would allow biomonitoring programs to achieve an intensive appraisal of freshwater conditions. In a rapidly changing world, streamlining current methodology to obtain as much data in as little time as possible is crucial.

Because DNA-based analysis of environmental samples such as contents of a kick-net sample can provide a broad spectrum of organisms in the habitat sampled, we hypothesized that kick-net metabarcoding will provide diatom biodiversity comparable to commonly used scraping method. Specifically, we aimed to 1) investigate the feasibility of kick-net sampling for capturing community assemblages of freshwater diatoms versus conventional periphyton scraping using a high throughput sequencing coupled metabarcoding approach and 2) compare diatom community assemblages across a known habitat quality scale (Good and Fair) using both conventional and kick-net sampling to investigate presence of diatom indicator groups.

Methods

Field Sampling

Samples were collected in November 2019 from Grand River tributaries across four study sites in Waterloo, Ontario (Fig. 1). Status and location data were provided by Dougan & Associates based on a 2018 benthos biomonitoring project for the City of Waterloo (S1 Table). The four selected sites were a subset of the sites from this project and were chosen based on accessibility and habitat quality. Hilsenhoff Biotic Index ranges (weighted by species) informed the habitat quality scale [36] which categorized sites into ‘Good’ (4.51-5.50) and ‘Fair’ (5.51-6.50).

Collection occurred in riffles, starting with a benthic kick-net sample, followed by subsequent periphyton scrapings of microhabitats representative of the reach (S2 Table). Periphyton scraping refers to the sampling of sediment, rock, macrophytes and leaf litter. Three replicates of each sampling type were collected at each site. Kick-net collection followed the Canadian Aquatic Biomonitoring Network [CABIN] protocol [37]. Effort was standardized to three minutes. The sampler moved up stream in a zig-zag pattern to encompass all microhabitats within the reach. Periphyton scraping samples were comprised of five specimens per microhabitat type to account for variability within the microhabitat [23]. Negative controls, consisting of molecular grade water, were collected prior to the collection of each rock sample (n= 9) to ensure the toothbrushes used for scraping biofilms from rocks had been adequately sterilised (S3 Table). All other samples were collected using manufacture-sealed sterile equipment. All samples were collected in 1L sample jars and placed in a cooler to transport back to the lab. Upon arrival at the lab, samples (n=45) were preserved using 100% ethanol and stored in a -20°C freezer until processing.

Sample Validation and Extraction

To account for potential false negatives [38], diatom presence in the samples was confirmed using microscopy. A small amount of ethanol used to preserve the samples was placed on a slide and observed under a compound microscope at 100X magnification. Visual inspection confirmed the presence of diatoms in each sample type (S1 Fig.), however no taxonomic information was taken as morphological identification was beyond the scope of this study.

Once diatom presence was validated, samples were homogenized using standard blenders decontaminated by washing with ELIMINase® (VWR, Canada) then rinsing with deionized water before treating with UV light for 30 minutes. Homogenate was subsequently transferred to 50 mL Falcon tubes, where one tube was set aside and centrifuged at 2400 rpm for two minutes. Supernatant was removed and residual pellets were incubated at 70 ºC until fully dried. Next, approximately 300 mg dried tissue was subsampled into PowerBead tubes and DNA extractions were completed using the DNeasy Power Soil kit (Qiagen, CA) following the manufacturer's protocol. The only exception being that 50 µL of buffer C6 (TE) was used for final elution. Negative controls containing no tissue were also included with each batch of extractions. All negative controls failed to amplify and therefore were not sequenced.

DNA Amplification, Library Preparation and Sequencing

Amplification targeted the 312 base pair long region of the chloroplast gene ribulose bisphosphate carboxylase large chain (rbcL) using five diatom specific primers. Following the methods of Rivera et al. [39], forward primers Diat_rbcL_708F_1 (5'-AGGTGAAG- TAAAAGGTTCWTACTTAAA-3'), Diat_rbcL_708F_2 (5'-AGGT-GAAGTTAAAGGTTCWTAYTTAAA-3') and Diat_rbcL_708F_3 (5'-AGGTGAAAC-TAAAGGTTCWTACTTAAA-3') were combined in an equimolar mix. Two reverse primers, Diat_rbcL_R3_1 (5'-CCTTCTAATTTACC- WACWACTG-3') and Diat_rbcL_R3_2 (5'-CCTTCTAATTTACCWA-CAACAG-3'), were also combined and used for amplification. Each reaction used the following reagents: 17.5 µL HyPure^{TM}

molecular biology grade water, 2.5 µL 10X reaction buffer (200 mM Tris-HCl, 500 mM KCl, pH 8.4), 1 µL $MgCl_2$ (50 mM), 05. µL dNTPs mix (10 mM), 0.5 µL of both forward (10 mM) and reserve (10 mM) equimolar mixes, 0.5 µL Invitrogen's Platinum Taq polymerase (5 U) and 2 µL of DNA. Final reaction volume totaled 25 µL.

PCR protocol largely followed Rivera et al. [39] with minor adjustments. Instead of thirty cycles of denaturation at 95°C for 45 seconds, annealing at 55°C for 45 seconds and extension at 72°C for 45 seconds [39], this study increased the number of cycles to thirty-five. PCR amplification was also performed in two-steps, with the second PCR using 2 µL of amplicons from the first PCR instead of DNA, and Illumina-tailed primers. All PCRs were completed in Eppendorf Mastercycler ep gradient S thermal cycler. Successful amplification was confirmed using 1.5% agarose gel electrophoresis before purifying second PCR amplicons with the MinElute Purification kit (Qiagen). The next step was quantifying purified samples with a QuantIT PicoGreen daDNA assay kit and using these values to normalize all samples to 3 ng/µL. Samples were then indexed and pooled before purifying with AMpure magnetic beads. QuantIT PicoGreen daDNA assay kit was once again used to quantify the library and Bioanalyzer was used to determine fragment length. The library was diluted to 4 nM and 10% PhiX was added before being sequenced using Illumina MiSeq with a V3 MiSeq sequencing kit (300 X 2; MS-102-2003).

Bioinformatic Processing

Illumina MiSeq paired-end reads were processed using the SCVURL rbcL metabarcode pipeline-1.0.2 pipeline available from https://github.com/terrimporter/SCVURL_rbcL_metabarcode_pipeline . SCVURL is an automated snakemake [40] bioinformatic pipeline that runs in a conda [41] environment. SeqPrep v1.3.2 [42] was used to pair raw reads requiring a minimum Phred score of 20 to ensure 99% base-calling accuracy. CUTADAPT v2.6 was used to trim primers from sequences, leaving a minimum fragment length of at least 150 base pairs [43]. Global exact sequence variant (ESV) [44] analysis was performed on the primer-trimmed reads. Reads were dereplicated using the 'derep_fulllength' command with the 'sizein' and 'sizeout' options of VSEARCH v2.14.1 [45]. VSEARCH was also used to denoise the data using the unoise3 algorithm [46]. These steps were taken to remove sequences with errors, chimeric sequences, PhiX carry-over and rare reads (singletons or doubletons) [47]. ESVs were classified using the rbcL diatom Classifier available from https://github.com/terrimporter/rbcLdiatomClassifier. Reference rbcL sequences were downloaded from the INRA diatom project [48]and reformatted to train the naive Bayesian classifier to make rapid, accurate taxonomic assignments [49]. This method makes assignments to the species rank and produces a statistical measure of confidence for each taxon up to the domain rank to help reduce false positive taxonomic assignments. We used 0.60 cutoff at the family rank (99% accuracy) and 0.20 cutoff at the genus rank (95% accuracy). The accuracy of the method assumes that target taxa are present in the reference database.

Statistical Analysis

RStudio was used to analyze the data [50]. To account for variable reads within the library each sample was normalized to the 15th percentile using the ‘rrarefy’ function in the vegan package [51,52].

ESV richness across the various sampling and status categories was calculated to assess differences between the methods and sites. A non-metric multi-dimensional (NMDS) analysis on Sorensen dissimilarities (binary Bray-Curtis) was conducted using the vegan ‘metaMDS’ function to determine if sampling method or site status created variation in community structure [5]. A scree plot was run using the ‘dimcheckMDS’ command from the goeveg package to determine the number of dimensions (k=2) to use with vegan metaMDS function[53]. Shephard’s curve and goodness of fit calculations were calculated using the vegan ‘stressplot’ and ‘goodness’ functions. The vegan ‘vegdist’ command was used to build a Sorensen dissimilarity matrix. We checked for heterogeneous distribution of dissimilarities using the ‘betadisper’ function. We used the ‘adonis’ function to perform a permutational analysis of variance (PERMANOVA). PERMANOVA was performed on conventional sampling methods (periphyton scraping) and kick-net methods, as well as site status to test for significant interactions between the categories [54].

To maintain a balanced design during statistical testing, we pooled all periphyton sampling into one sample type (conventional) and maintained kick-net samples as a separate sample type. The Jaccard index was calculated to assess the overall similarities between the sites, collection methods and site status. Nestedness and turnover of between kick-net and conventional samples were calculated using R package betapart function ‘beta-pair’ [55] followed by vegan function ‘betadisper’. The number of diatom family ESVs detected from kick-net or pooled conventional samples was also plotted. A dendrogram of diatom families detected was plotted using RAWGraphs (app.rawgraphs.io) and color-coded to show the samples the families were detected in [56]. Lastly, the frequency of ESVs detected from diatom families was visualized using a heatmap generated using geom_tile (ggplot) in R, plotting individual sample types for each site, split into two plots according to site status.

Results

After bioinformatic processing, we generated 4,272 ESVs (2,166,157 reads). After taxonomic filtering (removal of non-diatom phyla), a total of 3,940 diatom ESVs (2,125,984 reads) were retained for data analysis. Read coverage per sample after normalisation (15th percentile cut-off) was 37,735.

Since the rarefaction curves plateau, this indicated that the sequencing depth was sufficient to capture the ESV diversity in our PCRs (S2 Fig.). In terms of the top 10 orders identified, the order Naviculales represented 30.6% of ESVs (30% of reads) and Bacillariales represented 18.6% of ESVs (15.4% of reads; S3 Fig.).

Taxonomic Coverage

In terms of taxonomic assignment, we identified a total of 1 phyla (Bacillariophyta), 4 classes, 23 orders, 44 families and 77 genera at the 95% correct assignment level. ESV richness varied across different sampling methods (Fig. 2). Mean overall ESV richness was used to calculate alpha diversity which displayed very similar values for all sampling methods across the four sites (S4 Table). Averaged across sites, kick-net samples produced the lowest mean ESV richness (225 ± 85), with sediment samples producing the highest ESV richness (317 ± 92).

Through investigating diatom families, a majority of families detected were present in all microhabitats and kick-net samples (Fig. 3). Two families (Coscinodiscaceae and Orthoseriaceae) were solely present in leaf litter samples and two families (Entomoneidaceae and Diadesmidaceae) were present only in sediment samples (Fig. 3).

In terms of diatom genera, some of the confidently identified genera represented by more than 2 sequence variants, identified from kick-net and conventional samples, included: *Nitzschia* (Bacillariales), *Polypedilum* (Chironomidae), *Navicula* (Naviculales), *Amphora* (Thalassiophysales) and *Ulnaria* (Licmophorales; Fig. 4).

Diatom Diversity by Method and Site Status

NMDS plots showed that replicates clustered close together for site and status, with overlap observed between sampling methods and replicates (Fig. 5). When pooling conventional periphyton samples (i.e. macrophyte, leaf litter, rock, and sediment) at each site, there remained overlap between kick-net and conventional samples and samples also remained clustered by site and status (S4 Fig). PERMANOVA of the pooled samples, shows that analyzing data from kick-net or conventional samples (method) explains 13% of the variation in Bray Curtis dissimilarities (p-value = 0.776), sampling site (site) explains 58% of the variation (p-value = 0.009) and habitat quality status (status) explains 22% of the variation observed (p-value = 0.029; S5 Table). The Jaccard index for kick-net compared with conventional samples is 0.53, indicating samples are 53% similar, whereas the Jaccard index for fair compared to good site quality status samples is 0.20, indicating samples are only 20% similar. In terms of beta diversities of communities aggregated by the treatments of "kick-net" and "conventional", there was no significant difference between turnover. For beta diversities of communities aggregated by site status, there was a significant difference between nestedness (*P* < 0.05) but not for turnover (*P* = 0.06). Fair samples appear to be significantly nested within good samples. These results further indicate that site status has a significant effect on the sampled community composition whereas conventional versus kick-net sampling methods do not.

For individual sample types (i.e. kick-net, macrophyte, leaf litter, rock, and sediment), the heatmap shows that kick-net samples are largely representative of the diversity of families detected within each conventional periphyton sampling method (Fig.6.). In some cases, kick-net samples failed to detect diatom families which were present in conventional periphyton samples (e.g. Sellaphoraceae and Diadesmidaceae in Clair15) and conversely, kick-net samples also detected families which were not detected in conventional periphyton samples (e.g. Eunotiaceae and Neidiaceae in Clair12; Fig. 6). Similar assemblages of diatoms communities were detected across both fair and good quality sites, with the main difference observed between fair and good sites being the number of reads produced for families such as Thalassiosiraceae which was detected with a high number of reads (1000+) in fair sites and a lower number of reads (10-100) in good sites (Fig. 6).

Discussion

The demand for high-quality, reproducible ecological data is increasing in conjunction with the degradation of ecosystems globally [57]. There is a need to further streamline existing biomonitoring methodologies without sacrificing the quality of data produced [4,7,54,58]. With diatom assemblages providing a unique insight into the water quality status of lentic and lotic systems, fast-tracking diatom data collection for ecological assessments is a priority [39]. We have demonstrated that kick-net methodology with DNA metabarcoding provides sufficient taxonomic coverage to potentially be utilised as a for assessing diatom biodiversity in freshwater systems.

Kick-net sampling technique, whereby a zig-zag path is taken across the reach, provided sufficient representation of existing diatom community assemblages within site-specific microhabitats. Samples derived from the kick-net technique were highly comparable with conventional samples in terms of diatom taxa detected, despite the kick-net approach being more passive compared to direct periphyton scraping. Specific diatom taxa are known to have ecological preferences for different freshwater microhabitats [59,60]. For watershed-level health estimates, it is beneficial to be able to efficiently detect the diversity of diatom taxa present without directly sampling each microhabitat within a reach. We have demonstrated that kick-net methodology can sufficiently capture the existing diatom biodiversity, ground truthed by comparing assemblages detected with periphyton scrapings.

Ultimately, the detection of bioindicator species is a key variable to consider when comparing biomonitoring methods, as these taxa are pivotal for detecting subtle differences in freshwater health [3,5,14]. Naviculaceae contains diatom species sensitive to herbicide exposure, which is a family we observed in all sites and with all collection methods [61]. Additionally, the bioindicator family Stephanodiscaceae, (a known tolerant taxon) [62], has a higher read abundance in 'Fair' sites compared to 'Good' in both conventional and kick-net sample types. Despite the direct sampling approach of periphyton rock scraping, this methodology failed to detect this family at one of the sites where kick-net samples were successful at detecting this benthic family. Rock scrapings are commonly used as the sole collection method for diatoms [14,39,63,64], which suggests that the kick-net approach facilitates the detection of taxa which otherwise may be missed from conventional sampling.

Conclusion

Overall, this study found that benthic kick-net methodology enables a robust and detailed assessment of freshwater diatom communities. This methodology is a scalable option for generating a holistic insight into the health of freshwater systems. The high similarity of diatom taxa detected between methods and significant differences between diatom communities detected in sites of differing habitat quality, demonstrates that this rapid method can provide accurate, fine-resolution taxonomic results. Future research should examine the duo-analyses approach of macroinvertebrate and diatom communities from a single kick-net sample, to determine reproducibility of multi-taxa targeting with this method. Additionally, future studies should consider exploring the use of multiple markers (i.e. rbcL cpDNA versus 18S rRNA gene), to address level of taxonomic resolution that can be obtained with these markers commonly used for diatom DNA barcoding.

Supporting Information

S1 Table. Information on study sites, including GPS coordinates and site status.

S2 Table. Outline of collections methods used in this study. Samples for periphyton scraping were taken from a depth no greater than 1m (King et al. 2006).

S3 Table. Summary table of decontamination and sterilisation procedures undertaken for the equipment in this study.

S4 Table. Mean ESV values (replicates pooled) for each sample type across the four sites. Based on normalised data.

S5 Table. rbcL exact sequence variants (ESVs) are not significantly different between sampling methods (kick-net versus conventional periphyton sampling). No significant beta dispersion was detected within groups (method, site, status). Only significant difference detected was rbcL ESVs between sites and status. Summary of

PERMANOVA results based on a Sorensen dissimilarity matrix of rbcL ESVs. Significant p-values are bolded.

S1 Fig. Example of confirmation of diatom presence from preservative of kick-net sample. Image: CBG Photography Group.

S2 Fig. All samples show that ESV sampling reached saturation. Samples were color-coded by site or method as shown in the legend. The vertical dashed line indicates the 15th percentile of sampling read depth, which is the number of reads that would be used in any future analysis based on normalized data.

S3 Fig. Naviculales is the most abundant diatom order detected. Results for the top 10 orders are shown with respect to proportion of ESVs and reads recovered. Based on raw unnormalized data.

S4 Fig. Non-metric multi-dimensional scaling plots of microhabitat samples pooled show clustering by due to site and status. Specifically, a) depicts overlap between the binary Bray Curtis (Sorensen) dissimilarities between different sampling approaches, b) sample site clustering c) clustering based on habitat quality status (stress = 0.012, R2 = 0.98). Based on rarefied data.

Acknowledgements

We would like to extend thanks to Michael Wright for providing advice on laboratory procedures, Josip Rudar for assistance with bioinformatics and Genevieve Johnson for help collecting samples. We would also like to acknowledge Jessica Robinson, Jaclyn McKeown, Allison Brown and Monica Young from the Centre for Biodiversity Genomics’ Collections department for assisting with imaging and use of equipment. This study is funded by the Government of Canada through Genome Canada and Ontario Genomics. Teresita Porter was funded by the Government of Canada through the Genomics Research and Development Initiative (GRDI) Metagenomics-Based Ecosystem Biomonitoring Project (Ecobiomics).

Author Contributions and Competing Interests

V.C.M. collected samples, conducted molecular and genomic analyses and wrote the manuscript with help from all authors, C.V.R and M.H. designed the study, C.V.R contributed to sampling, bioinformatic processing and statistical analyses, T.M.P trained the classifier, contributed to bioinformatic processing and advised on data analysis. All authors helped to write/edit the manuscript. The authors have declared that no competing interests exist.

References

1. Bayramoglu S, Chakir R, Lungarska A. Impacts of Land Use and Climate Change on Freshwater Ecosystems in France. Environ Monit Assess. 2020;25: 147–172. doi:10.1007/bf02837416
2. Karaouzas I, Smeti E, Vourka A, Vardakas L, Mentzafou A, Tornés E, et al. Assessing the ecological effects of water stress and pollution in a temporary river - Implications for water management. Sci Total Environ. 2018;618: 1591–1604. doi:10.1016/j.scitotenv.2017.09.323
3. Lefrançois E, Apothéloz-Perret-Gentil L, Blancher P, Botreau S, Chardon C,

Crepin L, et al. Development and implementation of eco-genomic tools for aquatic ecosystem biomonitoring: the SYNAQUA French-Swiss program. Environ Sci Pollut Res. 2018;25: 33858–33866. doi:doi.org/10.1007/s11356-018-2172-2

4. Keck F, Vasselon V, Tapolczai K, Rimet F, Bouchez A. Freshwater biomonitoring in the Information Age. Front Ecol Environ. 2017;15: 266–274. doi:10.1002/fee.1490

5. Hajibabaei M, Porter TM, Wright M, Rudar J. COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS One. 2019;14: 1–18. doi:10.1371/journal.pone.0220953

6. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;21: 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x

7. Baird DJ, Hajibabaei M. Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol Ecol. 2012;21: 2039–2044. doi:10.1111/j.1365-294X.2012.05519.x

8. Borisko JP, Kilgour BW, Stanfield LW, Jones FC. An Evaluation of Rapid Bioassessment Protocols for Stream Benthic Invertebrates in Southern Ontario , Canada. Water Qual Res J Canada. 2007;42: 184–193.

9. Brua RB, Culp JM, Benoy GA. Comparison of benthic macroinvertebrate communities by two methods: Kick- and U-net sampling. Hydrobiologia. 2011;658: 293–302. doi:10.1007/s10750-010-0499-x

10. Hajibabaei M, Porter TM, Robinson C V., Baird DJ, Shokralla S, Wright MTG. Watered-down biodiversity? A comparison of metabarcoding results from DNA

extracted from matched water and bulk tissue biomonitoring samples. PLoS One. 2019;14: 1–16. doi:10.1371/journal.pone.0225409

11. Emilson CE, Thompson DG, Venier LA, Porter TM, Swystun T, Chartrand D, et al. DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. Sci Rep. 2017;7: 1–11. doi:10.1038/s41598-017-13157-x

12. Leese F, Altermatt F, Bouchez A, Ekrem T, Hering D, Meissner K, et al. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Res Ideas Outcomes. 2016;2: e11321. doi:10.3897/rio.2.e11321

13. Gibson JF, Shokralla S, Curry C, Baird DJ, Monk WA, King I, et al. Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. PLoS Genet. 2015;10: 1–15. doi:10.5061/dryad.vm72v

14. Chonova T, Kurmayer R, Rimet F, Labanowski J, Vasselon V, Keck F, et al. Benthic Diatom Communities in an Alpine River Impacted by Waste Water Treatment Effluents as Revealed Using DNA Metabarcoding. Front Microbiol. 2019;10: 1–17. doi:10.3389/fmicb.2019.00653

15. Rivera SF, Vasselon V, Jacquet S, Bouchez A, Ariztegui D, Rimet F. Metabarcoding of lake benthic diatoms□: from structure assemblages to ecological assessment. Hydrobiologia. 2018;807: 37–51. doi:10.1007/s10750-017-3381-2

16. Vasselon V, Rimet F, Tapolczai K, Bouchez A. Assessing ecological status with diatoms DNA metabarcoding□: Scaling-up on a WFD monitoring network (

Mayotte island , France). Ecol Indic. 2017;82: 1–12. doi:10.1016/j.ecolind.2017.06.024

17. Pandey LK, Bergey EA, Lyu J, Park J, Choi S, Lee H, et al. The use of diatoms in ecotoxicology and bioassessment□: Insights , advances and challenges. Water Res. 2017;118: 39–58. doi:10.1016/j.watres.2017.01.062

18. Bailet B, Bouchez A, Franc A, Frigerio J-M, Keck F, Karjalainen S-M, et al. Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. Metabarcoding and Metagenomics. 2019;3: 21–35. doi:10.3897/mbmg.3.34002

19. Visco JA, Apothéloz-Perret-Gentil L, Cordonier A, Esling P, Pillet L, Pawlowski J. Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. Environ Sci Technol. 2015;49: 7597–7605. doi:10.1021/es506158m

20. Blanco S, Bécares E. Chemosphere Are biotic indices sensitive to river toxicants□? A comparison of metrics based on diatoms and macro-invertebrates. Chemosphere. 2010;79: 18–25. doi:10.1016/j.chemosphere.2010.01.059

21. Muñoz I, Sabater S. Integrating chemical and biological status assessment: Assembling lines of evidence for the evaliuation of river ecosystem risk. Acta Biol Colomb. 2014;19: 25–33.

22. Sharifinia M, Mahmoudifard A, Gholami K, Namin JI, Ramezanpour Z. Benthic diatom and macroinvertebrate assemblages, a key for evaluation of river health and pollution in the Shahrood River, Iran. Limnology. 2016;17: 95–109. doi:10.1007/s10201-015-0464-5

23. King L, Clarke G, Bennion H, Kelly M, Yallop M. Recommendations for sampling littoral diatoms in lakes for ecological status assessments. J Appl Phycol. 2006;18: 15–25. doi:10.1007/s10811-005-9009-3

24. Aloi JE. Review of Recent Freshwater Periphyton Field Methods. Can J Fish Aquat Sci. 1990;47: 656–670.

25. Smucker NJ, Vis ML. Contributions to diatom diversity and distributional patterns in streams□: implications for conservation. Biodivers Conserv. 2011;20: 643–661. doi:10.1007/s10531-010-9972-0

26. Weilhoefer CL, Pan Y. A comparison of diatom assemblages generated by two sampling protocols. J North Am Benthol Soc. 2007;26: 308–318.

27. Rimet F, Bouchez A. Biomonitoring river diatoms: Implications of taxonomic resolution. Ecol Indic. 2012;15: 92–99. doi:10.1016/j.ecolind.2011.09.014

28. Winter JG, Duthie HC. Stream epilithic, epipelic and epiphytic diatoms: Habitat fidelity and use in biomonitoring. Aquat Ecol. 2000;34: 345–353. doi:10.1023/A:1011461727835

29. Álvarez-Blanco I, Cejudo-Figueiras C, Bécares E, Blanco S. Spatiotemporal changes in diatom ecological profiles: Implications for biomonitoring. Limnology. 2011;12: 157–168. doi:10.1007/s10201-010-0333-1

30. Smucker NJ, Vis ML. Diatom biomonitoring of streams: Reliability of reference sites and the response of metrics to environmental variations across temporal scales. Ecol Indic. 2011;11: 1647–1657. doi:10.1016/j.ecolind.2011.04.011

31. Lange-Bertalot H, Krammer K. Bacillariaceae, Epithemiaceae, Surirellaceae. J. Kramer; 1987.

32. Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshw Sci. 2017;36: 162–177. doi:10.1086/690649

33. Kermarrec L, Franc A, Rimet F, Chaumeil P, Frigerio JM, Humbert JF, et al. A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshw Sci. 2014;33: 349–363. doi:10.1086/675079

34. Rimet F, Vasselon V, Keszte BA, Bouchez A. Do we similarly assess diversity with microscopy and high-throughput sequencing? Case of microalgae in lakes. Org Divers Evol. 2018;18: 51–62. doi:10.1007/s13127-018-0359-5

35. Tapolczai K, Keck F, Bouchez A, Rimet F, Kahlert M, Vasselon V. Diatom DNA Metabarcoding for Biomonitoring: Strategies to Avoid Major Taxonomical and Bioinformatical Biases Limiting Molecular Indices Capacities. Front Ecol Evol. 2019;7: 1–15. doi:10.3389/fevo.2019.00409

36. Gazedam E, Gharabaghi B, Jones FC, Whiteley. Evaluation of the Qualitative Habitat Evaluation Index as a Planning and Design Tool for Restoration of Rural Ontario Waterways. Can Water Resour J. 2011;36: 149–158.

37. Environment Canada. Canadian aquatic Biomonitoring Network -Field Manual: Wadeable Streams. 2012.

38. Guillera-Arroita G, Lahoz-Monfort J., Rooyen AR Van, Weeks AR, Tingley R. Dealing with false-positive and false-negative errors about species occurrence at multiple levels. Methods Ecol Evol. 2017;8: 1081–1091. doi:10.1111/2041-210X.12743

39. Rivera SF, Vasselon V, Jacquet S, Bouchez A, Ariztegui D, Rimet F.

Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment. Hydrobiologia. 2018;807: 37–51. doi:10.1007/s10750-017-3381-2

40. Köster J, Rahmann S. Snakemake - A scalable bioinformatics workflow engine. Bioinformatics. 2012;28: 2520–2522. doi:10.1093/bioinformatics/bty350

41. Anaconda. Conda. 2016.

42. Seqprep.

43. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: 10–12.

44. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11: 2639–2643. doi:10.1038/ismej.2017.119

45. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: A versatile open source tool for metagenomics. PeerJ. 2016;2016: 1–22. doi:10.7717/peerj.2584

46. Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv. 2016. doi:10.1101/081257

47. Reeder J, Knight R. The “rare biosphere”: A reality check. Nat Methods. 2009;6: 636–637. doi:10.1038/nmeth0909-636

48. Rimet F, Chaumeil P, Keck F, Kermarrec L, Vasselon V, Kahlert M, et al. R-Syst□:: diatom□: an open-access and curated barcode database for diatoms and freshwater monitoring. Database. 2016;2016: 1–21. doi:10.1093/database/baw016

49. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ

Microbiol. 2007;73: 5261 LP – 5267. doi:10.1128/AEM.00062-07

50. Team Rs. RStudio: Integrated Development for R. In: RStudio, Inc., Boston, MA. 2016.

51. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome. 2017;5: 1–18. doi:10.1186/s40168-017-0237-y

52. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-2. 2018.

53. Goral F, Schellenberg J. goeveg: Functions for Community Data and Ordinations. 2018.

54. Robinson C V, Porter TM, Wright MTG, Hajibabaei M. Propylene glycol-based antifreeze as an effective preservative for DNA metabarcoding of benthic arthropods. bioRxiv. 2020; 2020.02.28.970475. doi:10.1101/2020.02.28.970475

55. Baselga A, Orme CDL. betapart: An R package for the study of beta diversity. Methods Ecol Evol. 2012;3: 808–812. doi:10.1111/j.2041-210X.2012.00224.x

56. Mauri M, Elli T, Caviglia G, Uboldi G, Azzi M. RAWGraphs: A visualisation platform to create open outputs. Proceedings of the 12th Biannual Conference on Italian Sigchi Chapter. New York, NY, USA: Association for Computing Machinery; 2017. pp. 1–5. doi:10.1145/3125571.3125585

57. Buss DF, Carlisle DM, Chon TS, Culp J, Harding JS, Keizer-Vlek HE, et al. Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environ Monit Assess. 2015;187. doi:10.1007/s10661-014-4132-8

58. Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring□: Do DNA extraction methods matter□? Freshw Sci. 2017;36: 162–177. doi:10.1086/690649.

59. Castro E, Siqueira T, Melo AS, Bini LM, Landeiro VL, Schneck F. Compositional uniqueness of diatoms and insects in subtropical streams is weakly correlated with riffle position and environmental uniqueness. Hydrobiologia. 2019;842: 219–232. doi:10.1007/s10750-019-04037-8

60. Yang X, Lv H, Li W, Guo M, Zhang X. Effect of water motion and microhabitat preferences on spatio-temporal variation of epiphytic communities: a case study in an artificial rocky reef system, Laoshan Bay, China. Environ Sci Pollut Res. 2018;25: 12896–12908. doi:10.1007/s11356-018-1349-z

61. Debenest T, Pinelli E, Coste M, Silvestre J, Mazzella N, Madigou C, et al. Sensitivity of freshwater periphytic diatoms to agricultural herbicides. Aquat Toxicol. 2009;93: 11–17. doi:10.1016/j.aquatox.2009.02.014

62. Reavie ED, Cai M. Consideration of species-specific diatom indicators of anthropogenic stress in the Great Lakes. PLoS One. 2019;14: 1–15. doi:10.1371/journal.pone.0210927

63. Jakovljević OS, Popović SS, Vidaković DP, Stojanović KZ, Krizmanić J. The application of benthic diatoms in water quality assessment (Mlava River, Serbia). Acta Bot Croat. 2016;75: 199–205. doi:10.1515/botcro-2016-0032

64. Srivastava P, Grover S, Verma J, Khan AS. Applicability and efficacy of diatom indices in water quality evaluation of the Chambal River in Central India. Environ Sci Pollut Res. 2017;24: 25955–25976. doi:10.1007/s11356-017-0166-0

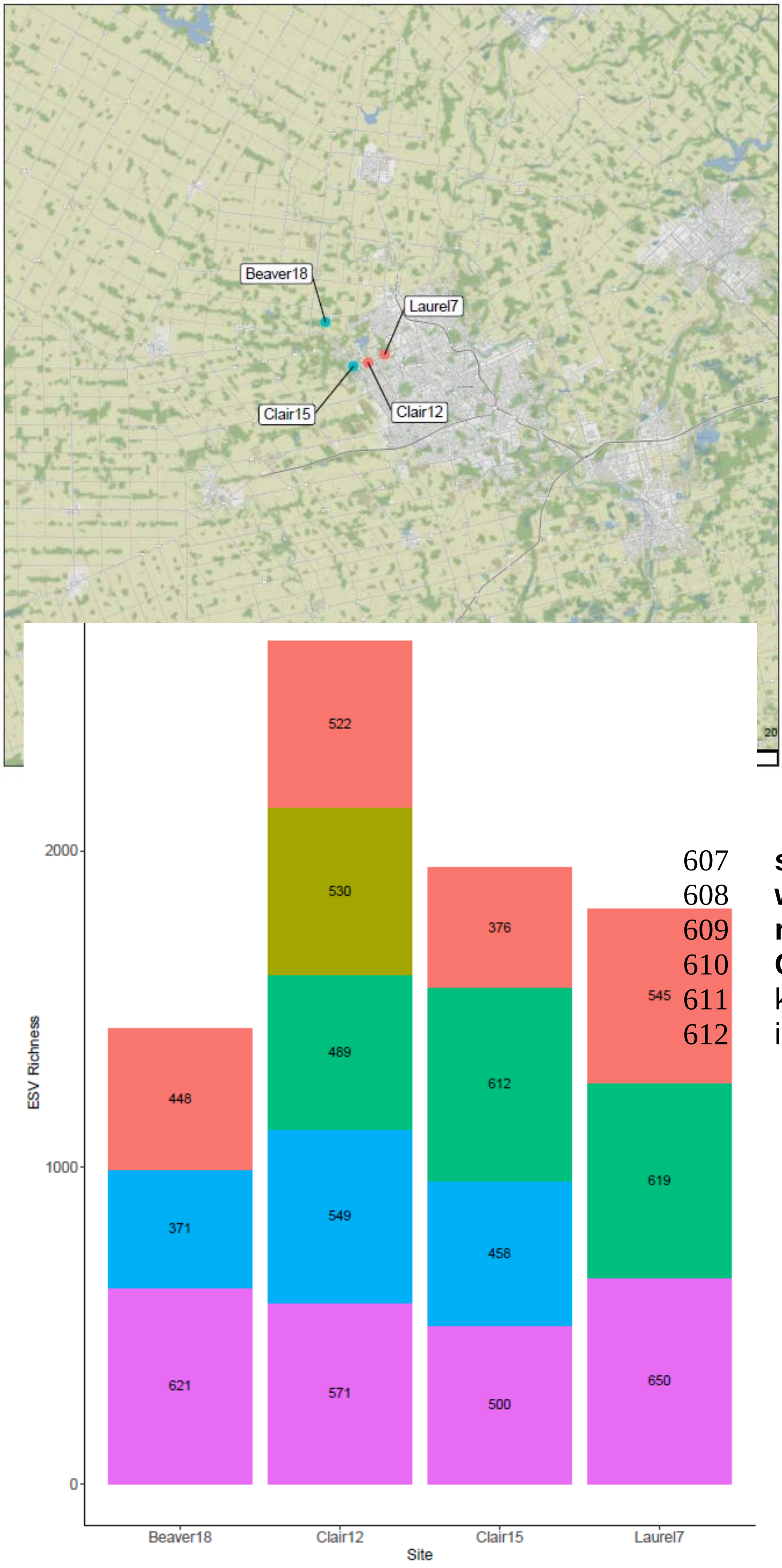


Fig. 1. Map of sample sites located within the Waterloo region (Ontario, Canada) Scale bar shown in km, site habitat status indicated in legend.

Fig. 2. ESV richness varies across different sample types. Methods refer to the different sampling approaches analyzed (i.e. Kick-net, Macrophyte, Leaf Litter, Rock and Sediment). Replicates are pooled. Based on rarefied data.

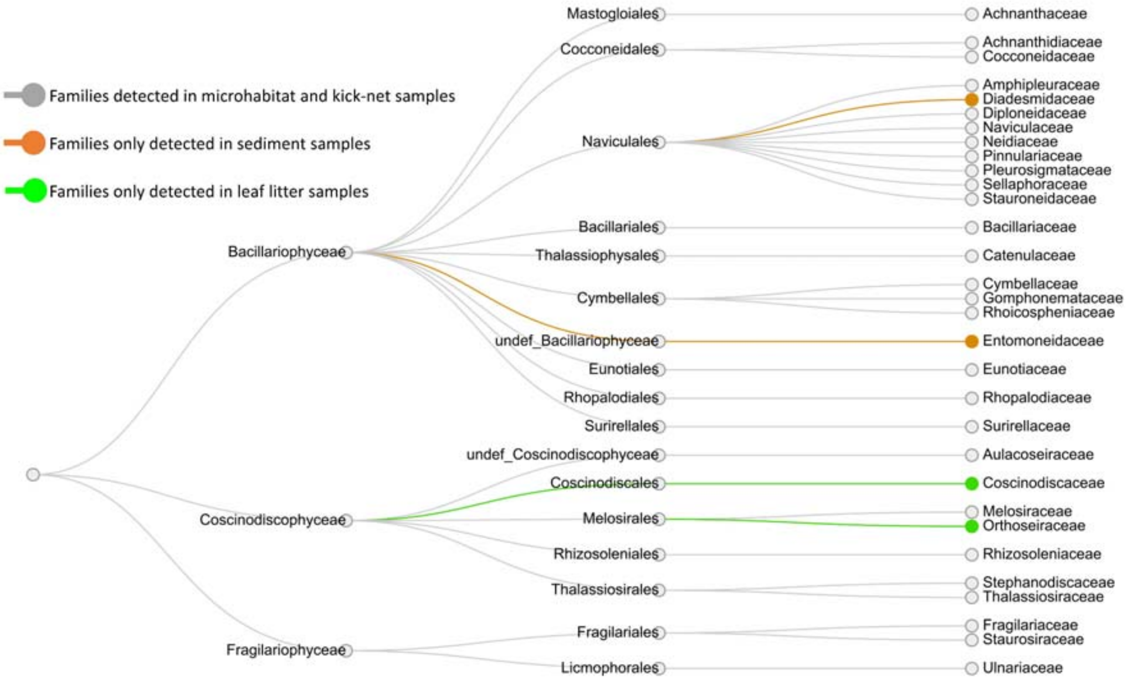
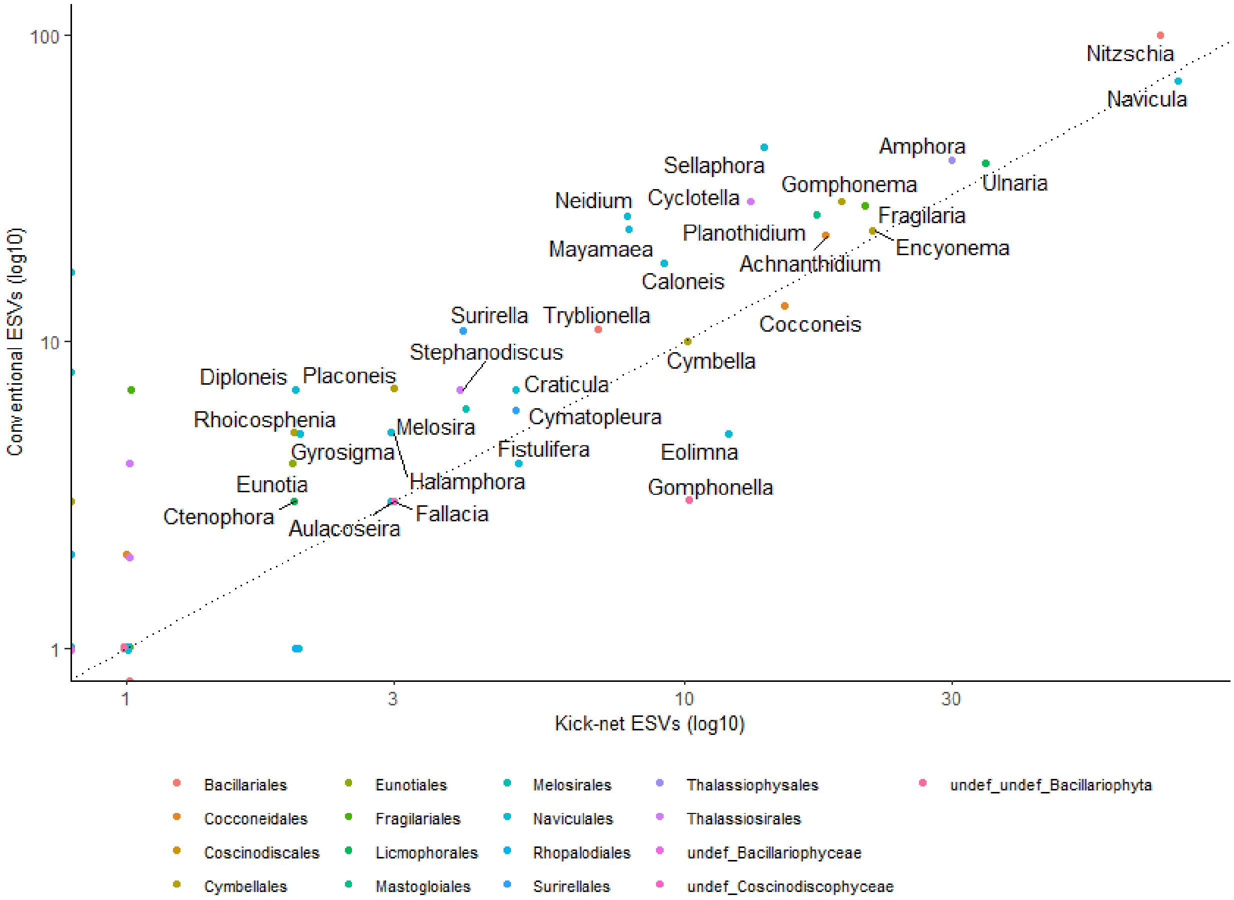


Fig. 3. A majority of diatom families were detected in both microhabitat and kick-net samples.



637

638 **Fig. 4. Number of ESVs detected from genera detected from kick-net versus**
639 **conventionally sampled diatoms are similar.** The points are color-coded for the
640 orders detected in this study. A 1:1 correspondence line (dotted) is also shown. A log10
641 scale is shown on each axis to improve the spread of points with small values. Based
642 on rarefied data.

643

644
645
646
647
648

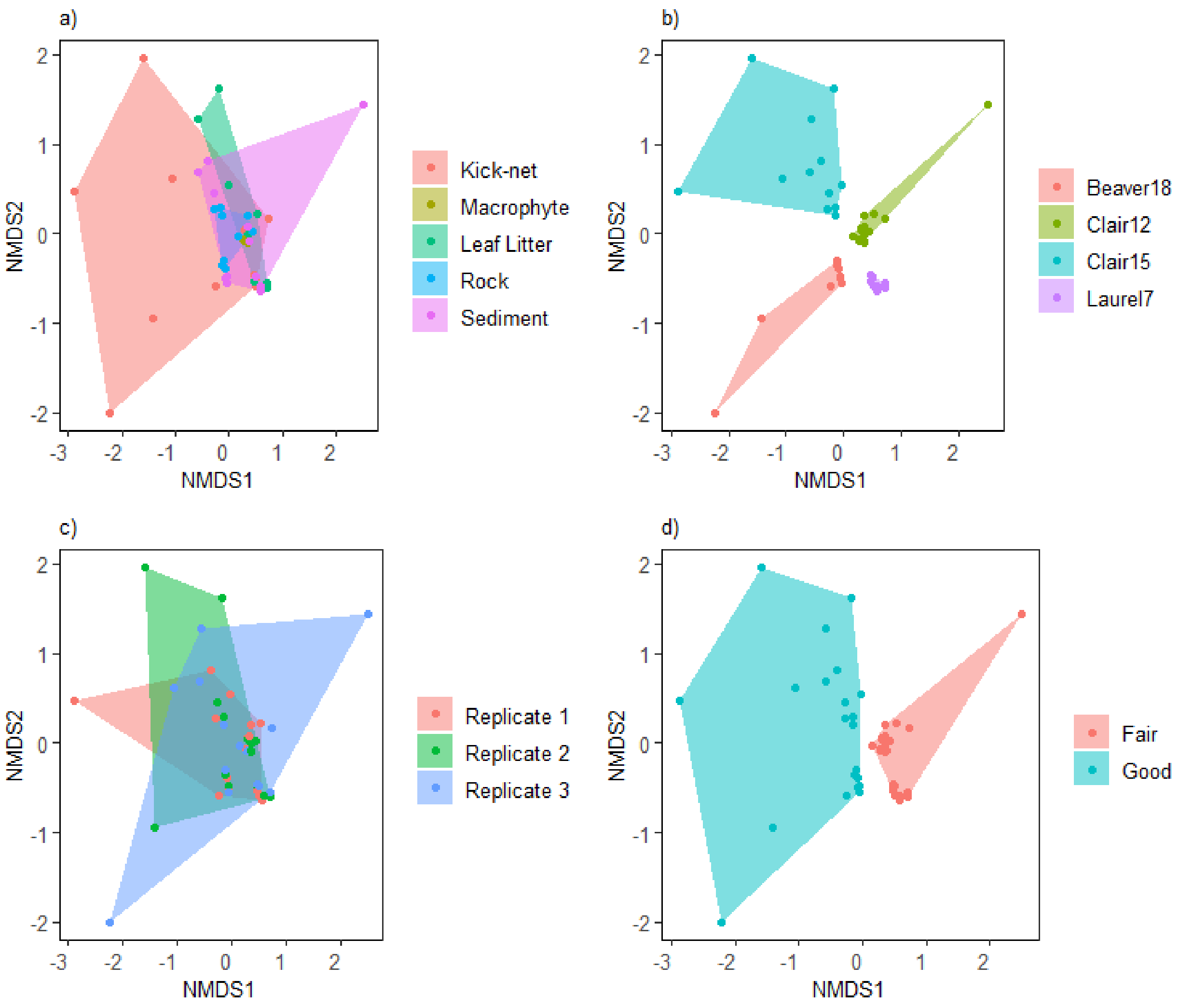


Fig. 5. Non-metric multi-dimensional scaling plots show clustering mainly due to site and status. Specifically, a) binary Bray Curtis (Sorensen) dissimilarities overlapping across different sampling approaches, b) clustering by site, c) overlap between replicates, and d) clustering based on habitat quality status (stress = 0.111, R^2 = 0.98). Based on rarefied data.

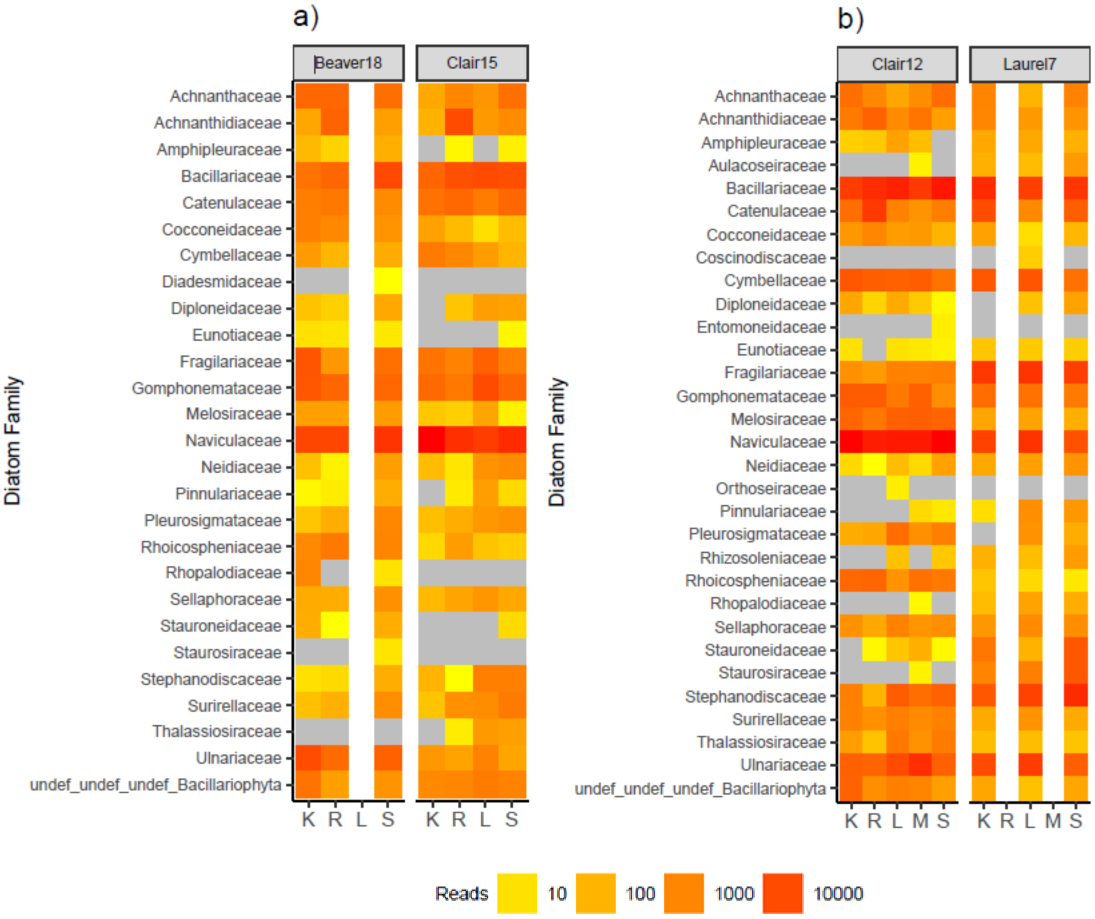


Fig. 6. Samples detect similar diatom families across sampling methods and site status. Only ESVs taxonomically assigned to families with high confidence (bootstrap support >= 0.60 for 95% accuracy) are included. Part a) shows sites with a ‘good’ quality status b) sites with a ‘fair’ quality status. Sampling methods: K = kick-net; R = rock scraping; L = leaf litter; M = macrophyte; S = sediment. Empty lanes indicate the corresponding microhabitat was not present at the site. For each site, three replicates for each sampling method are pooled. Based on normalized data.

Data Availability

Raw sequences will be available from NCBI SRA on acceptance. The SCVURL rbcL metabarcode pipeline-1.0.2 is available from https://github.com/terrimporter/SCVURL_rbcL_metabarcode_pipeline and the rbcLdiatomClassifier v1 we used is available on GitHub at https://github.com/terrimporter/rbcLdiatomClassifier.