MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 5: 83–97
DOI 10.3897/mbmg.5.64735

Research Article

Evaluating the performance of DNA metabarcoding for assessment of zooplankton communities in Western Lake Superior using multiple markers

Christy Meredith[1], **Joel Hoffman**[2], **Anett Trebitz**[2], **Erik Pilgrim**[3], **Sarah Okum**[3], **John Martinson**[4], **Ellen S. Cameron**[5]

1 *Montana Department of Environmental Quality, 1520 E. 6th Avenue, Helena, Montana, 59601, USA*
2 *U. S. Environmental Protection Agency, Office of Research and Development, Great Lakes Toxicology and Ecology Division, 6201 Congdon Blvd, Duluth, Minnesota, 55804, USA*
3 *U. S. Environmental Protection Agency Office of Research and Development, Watershed and Ecosystem Characterization Division, 26 West Martin Luther King Dr., Cincinnati, Ohio, 45268, USA*
4 *U. S. Environmental Protection Agency, Office of Research and Development, Great Lakes Toxicology and Ecology Division, 26 Martin Luther King Dr., Cincinnati, OH 45268, USA*
5 *Department of Biology, University of Waterloo, 200 University Ave. W, Waterloo, N2L 3G1, Ontario, Canada*

Corresponding author: Christy Meredith (Christy.Meredith@mt.gov)

Academic editor: Michael T. Monaghan | **Received** 21 February 2021 | **Accepted** 17 June 2021 | **Published** 21 July 2021

Abstract

For DNA metabarcoding to attain its potential as a community assessment tool, we need to better understand its performance versus traditional morphological identification and work to address any remaining performance gaps in incorporating DNA metabarcoding into community assessments. Using fragments of the 18S nuclear and 16S mitochondrial rRNA genes and two fragments of the mitochondrial COI marker, we examined the use of DNA metabarcoding and traditional morphological identification for understanding the diversity and composition of crustacean zooplankton at 42 sites across western Lake Superior. We identified 51 zooplankton taxa (genus or species, depending on the finest resolution of the taxon across all identification methods), of which 17 were identified using only morphological traits, 13 using only DNA and 21 using both methods. The taxa found using only DNA metabarcoding included four species and one genus-level identification not previously known to occur in Lake Superior, the presence of which still needs to be confirmed. A substantial portion of taxa that were identified to genus or species by morphological identification, but not identified using DNA metabarcoding, had zero ("no record") or < 2 ("underrepresented records") reference barcodes in the BOLD or NCBI databases (63% for COI, 80% for 16S, 74% for 18S). The two COI marker fragments identified the most genus- and species-level taxa, whereas 18S was the only marker whose family-level percent sequence abundance patterns showed high correlation to composition patterns from morphological identification, based on a NMDS analysis of Bray-Curtis similarities. Multiple replicates were collected at a subset of sites and an occupancy analysis was performed, which indicated that rare taxa were more likely to be detected using DNA metabarcoding than traditional morphology. Our results support that DNA metabarcoding can augment morphological identification for estimating zooplankton diversity and composition of zooplankton over space and time, but may require use of multiple markers. Further addition of taxa to reference DNA databases will improve our ability to use DNA metabarcoding to identify zooplankton and other invertebrates in aquatic surveys.

Key Words

zooplankton, Great Lakes, Lake Superior, identification, metabarcoding

PENSOFT.

Introduction

Quantifying biodiversity is an essential part of aquatic biomonitoring. Biodiversity information is often used to prioritise conservation efforts or to characterise the health of aquatic systems along gradients of natural conditions and anthropogenic impacts (Poole et al. 2004; Naidoo et al. 2008; Abell et al. 2008). With targeted biomonitoring efforts, we can detect changes in the composition of native communities, as well as potential invasive species before they become well-established (Vander Zanden 2008; Hoffman et al. 2011). However, logistical and financial constraints often limit these capabilities with respect to spatial, temporal and taxonomic resolution (Vos et al. 2000).

DNA metabarcoding has the potential to reduce some of these constraints by improving the efficiency and accuracy of aquatic surveys (Baird and Hajibabaei 2012). The general cost of processing DNA data has declined by over 100-fold in the last ten years (Reuter et al. 2015). The costs of DNA metabarcoding can be particularly low once laboratory and processing protocols are well-established. For instance, Zaiko et al. (2015) estimated that the costs of identifying plankton in ship ballast water with DNA metabarcoding were approximately 50% those of costs of identification using traditional morphological-identification techniques. Another benefit is that DNA metabarcoding can be used to find rare or cryptic taxa, damaged specimens, hard-to-identify early life-stages and eggs missed by morphological identification (Lindeque et al. 2013; Chain et al. 2016). It is also well-suited to identifying communities of very small organisms in water samples, such as plankton or bacteria, which are challenging to identify morphologically (Zaiko et al. 2015; Brown et al. 2016). A number of studies have highlighted the potential benefits of DNA metabarcoding for understanding the diversity of zooplankton (Chain et al. 2016; Zhang et al. 2018). DNA metabarcoding can also improve early detection of aquatic invasive species, including zooplankton (Brown et al. 2016).

Given the large geographic area in need of monitoring and potential limitations of morphological identification, DNA metabarcoding is a possible tool to help track changes in composition of zooplankton communities of the Laurentian Great Lakes (hereafter "Great Lakes"). Shifts in zooplankton composition have occurred due to the introductions of dreissenid mussels, land-use changes altering nutrients and lower trophic webs and the introduction of invasive zooplankton, such as the spiny water flea (*Bythotrephes longimanus*) (Barbiero et al. 2019; Yan et al. 2011). Estimating zooplankton composition is an established component of biological monitoring within the Great Lakes (Burlakova et al. 2018), but to date, it is accomplished almost entirely via morphological identification.

Some of the typical challenges of using DNA metabarcoding technology are pronounced for zooplankton taxa. For example, the choice of DNA marker has a large influence on the taxa and number of sequences detected (e.g. Clarke et al. 2017; Zhan et al. 2014), often resulting in an inability to both make species-level identifications and accurately estimate community composition patterns with the same marker. In addition, online barcode databases for zooplankton are incomplete. For instance, over 40% of crustacean zooplankton and 60% of rotifers known from the Great Lakes did not have records in the BOLD reference library at the species-level for the COI marker as of 2014 (Trebitz et al. 2015). Understanding the influence of marker and primer choice and reference library completeness on our ability to detect changes in species diversity and composition use is critical for incorporating DNA metabarcoding into zooplankton sampling efforts on the Great Lakes.

In this paper, we compare the ability of DNA metabarcoding for profiling zooplankton of Lake Superior against morphological identification. We did so for three DNA marker regions: fragments of the 18S nuclear and 16S mitochondrial rRNA genes and two fragments of the mitochondrial COI gene (hereafter referred to as 18S, 16S, COI-F230 and COI-BE). The questions which we asked were as follows:

1) Compared to taxa identified using traditional identification, what genus- or species-level taxa were identified using each of the DNA markers?
2) What percentage of these morphologically identified taxa currently has reference barcodes in online DNA libraries, resulting in the ability to assign taxonomic labels to genetic sequences?
3) Using an occupancy modelling approach, how does the ability to detect a taxa when present differ for DNA metabarcoding versus morphological identification (for taxa present in DNA libraries)?
4) What is the ability of the DNA markers for detecting overall shifts in percent biomass along broad taxonomic groups, which is a common use of zooplankton data in the Great Lakes Region?

Our study differs from many other studies applying DNA metabarcoding to zooplankton in that we examined all three markers typically used for DNA metabarcoding studies of aquatic taxa, with a focus specifically on crustaceous zooplankton. We also incorporated an occupancy modelling approach for detection of rare taxa.

Methods

Field collection

We used data from two separate sampling efforts in June of 2016, both conducted in Lake Superior, which is characterised by Precambrian geology located at the southern end of the Canadian Shield. The first sampling effort was conducted aboard the U.S. Environmental Protection Agency (EPA) research vessel the R/V Lake Guardian in conjunction with the Environmental Sea Grant June 2016

Teacher Cruise. During the cruise, teachers were assisted by EPA scientists in the collection of limnological field data. A total of 12 study sites were selected non-randomly to span a range of depths across three major sampling areas (Fig. 1). The second sampling effort consisted of a random sampling of 30 sites within the Saint Louis River Estuary and the nearshore area of Duluth, Minnesota, aboard the research vessel the EPA R/V Lake Explorer II. For this second sampling effort (Fig. 1), each site was sampled 2–3 times during a period of four days. The purpose of this repeated sampling was to use an occupancy modelling approach to explore the detectability of individual crustaceous zooplankton taxa using morphological identification versus DNA metabarcoding (Fiske and Chandler 2011). We subsequently refer to these two sampling efforts as the Western Superior-Guardian and the Duluth-LEII sampling efforts, respectively.

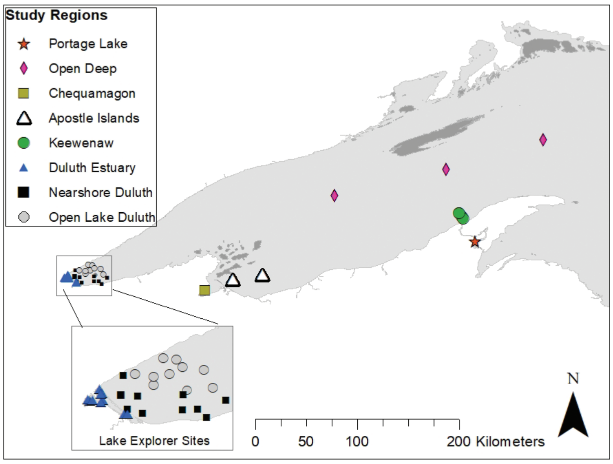

Figure 1. Sites sampled for zooplankton within study regions on Lake Superior as part of the Western Superior-Guardian and Duluth-LEII (e.g. "Lake Explorer Sites") research cruises. Multiple replicates were taken at the Lake Explorer sites for the occupancy modelling analysis.

A sampling event at a site consisted of two deployments of a standard 153-µm zooplankton net (WILDCO; Yulee FL, USA), towed vertically from a depth of 2 m above the bottom to the surface. We filtered the contents of each tow through a 153-µm mesh and washed them into a single plastic storage bottle containing 95% non-denatured alcohol (hereafter ethanol). Once all samples were collected, we used a Standard Folsom Plankton Splitter (WILDCO; Yulee FL, USA) to separate each sample from a site into two parts: one for morphological identification and one for DNA-based identification. Split samples were also stored in ethanol.

Morphological identification

Taxonomists identified crustaceous zooplankton to the lowest taxonomic resolution possible, using the EPA's Standard Operating Procedure for Zooplankton Analysis (GLNPO 2016), which includes enumerating up to four subsamples of 200 to 400 zooplankton individuals each. The identification is made on the basis of morphological traits using standard taxonomic keys for the Great Lakes. Up to ten individuals of each species were separated out for voucher specimens, with the goal of using Sanger sequencing techniques to generate additional DNA reference sequences. However, these attempts were unsuccessful due to difficulties in obtaining barcode-worthy tissue which has also been observed in other studies (Yang et al. 2017b). We excluded rotifer taxa, as these are not effectively sampled with a 153-µm mesh net.

DNA-based identification

In the lab, individual samples were condensed into 50-ml vials by decanting against a dried, bleached 153-µm mesh sieve. Condensed samples were stored in ethanol, then dried in their tubes in a vacuum desiccator until dry, which was approximately 4 days. Samples were digested using the Qiagen DNeasy Blood & Tissue Kit (Qiagen; Germantown MD, USA) using increased volumes of ATL buffer and proteinase K in the same ratio as the kit protocol. After incubation, a volume of 400 µl of liquid digest (approximately 50% of the material) was then transferred to a new centrifuge tube before adding equal volumes of AL Buffer and ethanol. Two spins were required to filter the entire extraction solution. Extracts were eluted with 100 µl of elution buffer from the Qiagen kit and stored at 4 °C. One primer set was used for each of the targeted regions of the 18S and 16S markers and two primer sets were used for the COI marker (Table 1). The first set of primers to evaluate COI targets the F230 fragment (COI-F230), which is ~ 230 bp in length and is found at the 5' end of the standard barcoding region. The second set of primers to evaluate COI (COI-BE) targets the BE fragment, which is ~ 314 bp in length and is found towards the 3' end of the standard barcoding region, with no overlap with the F230 fragment. Both COI fragments were designed for metabarcoding of

Table 1. Description of primers used and annealing temperatures for first PCR.

Marker	Primer name	Orientation	Reference	Sequence	Annealing temperature
COI-F230	F	Forward	Gibson et al. (2015)	GGTCAACAAATCATAAAGATATTGG	46 °C
COI-F230	230_R	Reverse	Gibson et al. (2015)	CTTATRTTRTTTATICGIGGRAAIGC	46 °C
COI-BE	B	Forward	Gibson et al. (2015)	CCIGAYATRGCITTYCCICG	46 °C
COI-BE	R5	Reverse	Gibson et al. (2015)	GTRATIGCICCIGCIARIAC	46 °C
18S	SSU_F04	Forward	Blaxter et al. (1998)	GCTTGTCTCAAAGATTAAGCC	50 °C
18S	SSU_R22	Reverse	Blaxter et al. (1998)	GCCTGCTGCCTTCCTTGGA	50 °C
16S	16Sar	Forward	Palumbi et al. (1996)	CGCCTGTTTATCAAAAACAT	50 °C
16S	16Sbr	Reverse	Palumbi et al. (1996)	GCCGGTCTGAACTCAGATCACGT	50 °C

invertebrate samples (Gibson et al. 2015). The primers to evaluate 18S target approximately 300+ bp of the V2-V3 region of the nuclear small subunit (Blaxter et al. 1998) and have been shown to be useful for metabarcoding applied to aquatic samples (Fonseca et al. 2010, Lindeque et al. 2013). The primers used in the evaluation of 16S were designed for invertebrates (Palumbi 1996) and amplify a short region of DNA (200+ bp) from the 3' end of the mitochondrial RNA (Table 1).

The laboratory procedures were performed separately for the regions of the 18S rRNA gene, 16S rRNA gene and the two fragments of the COI marker gene. Each primer set contained an upstream and downstream adapter that bound to index primers in the dual-indexing PCR step. The first round of PCR contained 2 µl DNA template, 2 µl 10× PCR buffer, 0.6 µl $MgCl_2$ (25 mM), 2 µl dNTPs (10 mM), 0.5 µl each of the forward and reverse primers per marker/primer combination (10 mM), 4 µl 1× BSA, 0.1 µl Taq polymerase (5 U/µl; Qiagen) and 9.9 µl ultrapure water. PCR was performed under cycling conditions, consisting of an initial 2.5-min denaturing step at 94 °C, followed by 35 cycles of 30 s at 94 °C, 1 min at the annealing temperatures in Table 1 pertaining to each marker, 1 min at 72 °C and a final elongation step of 10 min at 72 °C. Agarose gel electrophoresis was used to confirm the absence of visible amplification products in all field blanks and to confirm the presences of amplification products in test samples. Successful PCR products were purified with Qiagen QIAquick 96 PCR Purification kit. DNA templates that did not amplify for PCR were cleaned with the Zymo Research One Step PCR Inhibitor Removal Kit (Zymo; Irvine CA, USA) and re-run.

Dual-indexing PCR for DNA sequencing multiplexing was then run with primers containing the proprietary sequences necessary for a run on the Illumina MiSeq (Illumina, Inc; San Diego CA, USA) and index sequences for identification of each sample. These included a forward or reverse index and the upstream adapters from the initial PCR (Table 1). Thermal cycler conditions for dual-indexing PCR were 95 °C for 3 minutes, then 8 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds; this was followed by a final extension step of 72 °C for 5 minutes then a 4 °C soak. Dual-indexed amplicons were cleaned with AMPure XP beads (Beckman Coulter; Brea CA, USA), quantified and then normalised to the lowest nanomolar concentration in Qiagen EB Buffer. Normalised Index PCR plates were then pooled into a single sample by combining 3 µl from each well into a 1.5 ml micro-centrifuge tube. Pooled amplicon libraries were sent for sequencing on the MiSeq platform. Amplicons were sequenced using a 2 × 300 600-cycle Illumina MiSeq sequencing kit according to manufacturer's protocols. Index sequences were given to the MiSeq before the run and were used by MiSeq software to assign sequences to each individual sample.

Sequence data was processed using scripts in USEARCH v.9.2 64-bit (Edgar 2010) on demultiplexed reads generated from the MiSeq runs. Forward and reverse reads were merged and PCR primer sequences removed with Cudadapt v.1.14 (Martin 2011). Target lengths were 310, 220, 300 and 240 bp for the COI-BE, COI-F230, 16S and 18S markers, respectively. Sequences with an expected error rate greater than 1% were excluded. Remaining sequences were de-replicated and unique sequences were identified; those with < 4 observations in the total dataset were removed as possible sequencing artifacts. The remaining sequences were screened to remove chimeras and clustered into Operational Taxonomic Units (OTUs) of 97% or greater similarity, after which all quality-filtered reads were mapped to these OTUs. OTUs represented by only one or two sequences in the entire dataset for a given marker were removed from the final analysis. The final number of reads mapped to OTUs was 3.7, 6.8, 4.5 and 6.7 million sequences for the COI-BE, COI-F230 and targeted regions of the 16S and 18S genes, respectively.

Taxonomic identities were assigned to OTUs, based on sequence similarity to reference sequences in the BOLD and NCBI databases using BLAST (16S and 18S) and BOLD identification engine (COI). We also explored the use of PR2 and SILVA curated rRNA gene databases for species assignment for 18S (SILVA, PR2) and 16S (SILVA) using the DADA2 package in R (Guillou et al. 2013; Quast et al. 2013; Callahan et al. 2016), but found that over 50% percent of sequences that had matches in NCBI and BOLD did not have a match in these other databases, even for taxa identified at the family level or higher. These are high-quality curated databases, but had limited matches, which we attribute to the fact that PR2 largely focuses on protists and SILVA has only nearly full length sequences. Since the goals of our research were detection of rare taxa and analysis of percent biomass of broad-scale taxonomic groups, we chose to use NCBI and BOLD for taxonomic assignment. Potential implausible occurrences are elaborated in the results and discussion. Genus-level identifications of 90% or better and species-level of identifications of 97% or better were used to generate the final taxonomic list and in occupancy modelling. If a taxon matched multiple OTUs with the same percent similarity at the species-level, the taxa were assigned at the genus level. Ultimately, all genus-level zooplankton matches had a 90% or better similarity to online databases and were included in the analysis, while some species-level identifications matched at < 97% and were assigned to the genus-level. All zooplankton OTUs matched > 85% threshold and were included in the analysis of percent biomass of broad taxonomic groups, which occurred at the family or higher level.

We explored the use of rarefied and normalised data for use in the analysis (see supplemental information https://doi.org/10.17605/OSF.IO/ABNGX). While rarefying as a normalisation technique has been previously criticised (McMurdie and Holmes 2014), more recent research has indicated that repeatedly rarefying is a potentially appropriate normalisation technique (Cameron et al., submitted manuscript, https://doi.org/10.1101/2020.09.09.290049). We observed no loss of OTUs for any method, nor any notable effect on our analysis of broad-scale composition patterns when the transformed number of reads was used instead of proportion of sequences. However, we did observe a loss of rare OTUs at some sites for the COI marker. Given the desire to retain rare taxa for our occupancy analysis (which used combined data and did not investigate marker performance), we chose to use non-transformed data for our analysis.

Analysis

Combining data from both sampling efforts (i.e. Western Superior-Guardian and Duluth-LEII), we determined the number of sites at which each taxon was identified using morphological identification and with each of the three marker genes (including both fragments for COI) at the genus level and at the highest resolution at which the taxa were identified for each approach. For the Duluth-LEII sampling effort, 2–3 replicates were collected at each site and we combined data from all replicates to determine if the taxon were present at a site. We recognise that more effort was employed per sample than from the Western Superior-Guardian (given that enumerating organisms in additional samples may yield additional taxa), but these additional replicates were consistent across identification methods. We depict overlaps and unique detections amongst genera and lowest-level taxa detected using Venn diagrams created with the package vennDiagram in R (Chen 2011). Again, lowest-level may be at the genus- or species-level depending on the finest resolution detected across identification approaches. If a taxon were absent using a particular marker, but appeared in morphological identification or with a different marker, we determined if the potential reason for this non-detection was that reference sequences were under-represented in GenBank and BOLD databases (defined as having only one or two entries, respectively) or absent entirely (defined as having zero entries).

In a number of cases, morphological identification yielded a particular species, while DNA metabarcoding identified a closely-related different species. If a literature search revealed either a recent change in taxonomy or a disagreement as to the taxonomic classification, we considered the two species to be the same "taxon" in our analysis. This generally occurred because DNA metabarcoding yielded an updated taxonomic name, but this usage had not been employed by local taxonomists. We did this for the following taxa: *Acanthocyclops americanus*/*vernalis* (Dodson 1994), *Bosmina longirostris*/*liederi* (Kotov et al. 2009), *Eurytemora affinis*/*carolleeae* (Vasquez et al. 2016), *Chydorus sphaericus*/*brevilabris* (Belyaeva and Taylor 2009), *Eucyclops agilis*/*serrulatus* (Alekseev et al. 2006) and *Holopedium gibberum*/*glacialis* (Rowe et al. 2007). In addition, *Mysis relicta* has been split into multiple species groups that are highly similar in morphology (Audzijonyte and Vainola 2005). As OTUs assigned to *Mysis* matched multiples of species at the same high similarity, these were assigned to the *Mysis* genus.

We explored the use of DNA metabarcoding to quantitatively characterise relative zooplankton biomass by comparing estimates of percent biomass using morphological identification to percent sequence abundance for the four DNA approaches. We aggregated zooplankton to the following family or order-level categories for this analysis: bosminids, diaptomids, daphnids, cyclopoids, harpacticoids, other cladocerans (non-daphnids or bosminids), other calanoids (non-diaptomids) and mysids. To estimate biomass of each group for the morphological identification data, we multiplied abundance of each family or order category in a sample by the average biomass (in micrograms) across zooplankton species found in that family or order in Yurista et al. (2009). We recognise that a more rigorous approach, that considers biomass of individual species and/or length-biomass relationships, might yield more accurate results. However, neither biomass nor length-biomass relationships were available for all species in our dataset and lengths of the taxonomically-identified individuals were not measured. For this analysis, we were more interested in analysing broad patterns in biomass between sites, which was possible given that sequence abundance between taxonomic groups differed widely amongst broad-scale geographic zones.

We used stacked bar charts to visualise the broad changes in species composition by zone. Zone-level values were obtained by dividing the number of sequences from each broad-scale taxonomic group in a sample by the total number of sequences in the sample and aggregating by site and again by geographic zone. We also performed Bray-Curtis analysis (Bray 1957) on zone-level data to characterise differences in percent composition of broad-scale taxonomic groups. To visually compare broad-scale differences in composition, the vegan package (Oksanen et al. 2019) was used to perform a Non-Metric Dimensional Scaling (NMDS) analysis on the Bray-Curtis matrix. Each point in the resulting NMDS plots represents the composition of a zone and points are comparable across plots given that data from each marker were included in the same Bray-Curtis analysis. To further compare composition across sites, an additional Bray-Curtis analysis was similarly performed on site-level data. Spearman-Rank correlation was used to determine the correlation between NMDS axes resulting from a NMDS analysis performed on the site-level data.

By using multiple replicates, occupancy modelling allows for the determination of the probability of detecting a given taxon at a site, if it is present. For instance, if a taxon is detected at one out of three replicates at a site with a given method, the probability of detection with that method is approximately 33%. All field replicates were split into a subsample for DNA metabarcoding and a subsample for morphological identification. We used occupancy modelling to compare the probability of detection of each zooplankton taxon from DNA metabarcoding to the probability of detection using morphological identification. We confined this portion of the analysis to the Duluth/LEII sampling event, where multiple replicates were taken. As a result, not all taxa identified in the study were represented in this occupancy modelling analysis. We ran this analysis using the taxonomic resolution provided by morphological identification, even if it were coarser than that produced by metabarcoding. For a taxon to be considered present according to DNA metabarcoding, we required that it be identified in two or more field replicates using one identification method (e.g. COI-F230, COI-BE, 18S, 16S or morphological identification) or in one replicate by two or more methods. This reduced the likelihood that the detection probability for DNA would be inflated due to false positives. We performed the occupancy modelling using the unmarked package (Fiske and Chandler 2011) in R.

Results

The COI -F230R marker fragment generated the largest number of total OTUs and the 18S marker generated the smallest number of total OTUs (range 87 to 451; a factor of 4; Table 2). The COI-F230R fragment also generated the greatest number of crustaceous zooplankton OTUs, while the 16S marker generated the least number of crustaceous zooplankton OTUs (range 32 to 92; a factor of 3; Table 2). The difference amongst markers in the number of zooplankton taxa assigned to a genus or species was smaller still (factor of 2), with the two fragments of the COI finding more genera and species than the 16S or 18S.

Notably, none of the DNA markers, individually, found as many zooplankton genera or species as did morphological identification (Table 2). We identified 51 unique zooplankton taxa (Appendix 1), of which 17 were identified using only morphological traits, 13 using only DNA metabarcoding and 21 using both approaches. Again, performance varied by marker. The percent of these 51 lowest-resolution taxa, identified using each marker, was 41%, 35%, 22% and 16% for COI-F230R, COI-BE, 16S and 18S, respectively (Table 2; Fig.2). For taxa not found using DNA, approximately 63% (COI), 80% (16S) and 74% (18S) had no records or under-represented records (one or two reference sequences) in online databases. A total of 12 of the 17 taxa, found via morphological identification only, were not well-represented in online databases for at least one marker (Appendix 1; see "UR" or NR") (Fig. 3; Appendix 1). However, the other five taxa, found only via morphological identification, did have adequate barcodes in GenBank or BOLD. Three of these taxa were considered morphological identification-only records because they were found at a species resolution using morphological identification versus only at the genus-level using DNA, with no species of that genus identified using DNA metabarcoding. These included the genera of *Diaphanosoma* (COI-F230, COI-BE and 16S), *Eubosmina* (16S) and *Tropocyclops* (COI-F230). The other two taxa that were found using morphological identification, but not using DNA, despite being present in online databases for at least one marker, were *Mesocyclops edax* and *Eurycercus lamellatus.* No other members of these two genera were identified using metabarcoding.

Table 2. Summary of number of OTUs generated for each DNA approach (16S, 18S and 2 fragments for COI) and number of crustacean zooplankton genera, species and lowest-level taxa (e.g species for most taxa, but genus if no species were identified for that DNA approach) using each DNA approach and morphological identification.

	DNA	COI-BE	COI-F230	16S	18S	Morph ID
No. OTUs obtained	–	359	451	315	87	–
No. zooplankton OTUs	–	66	92	32	38	–
No. zooplankton genera	27	19	18	11	11	32
No. zooplankton species	29	13	18	12	7	33
No. lowest-level taxa	34	16	20	13	8	37

Of the taxa found only using DNA metabarcoding, one taxon was a species-level identification made using DNA metabarcoding (*Ceriodaphnia dubia*), while only a genus-level identification was made using morphological identification. The other 12 lowest-level taxa that were found only using DNA metabarcoding were *Daphnia longiremis* (identified with 16S and COI-BE), *Daphnia cucullata*, *Skistodiaptomus pallidus* and *Skistodiaptomus reighardi* (COI-F230R); *Pleuroxus* sp. and *Macrothrix* sp. (COI-BE); and *Daphnia pulex*, *Daphnia ambigua*, *Daphnia dentifera*, *Eubosmina longispina*, *Calanus* sp. and *Hemidiaptomus ingens* (16S). *Pleuroxus* sp., *Calanus* sp. and *D. dentifera* were identified at only one site, whereas the other taxa were identified at multiple sites.

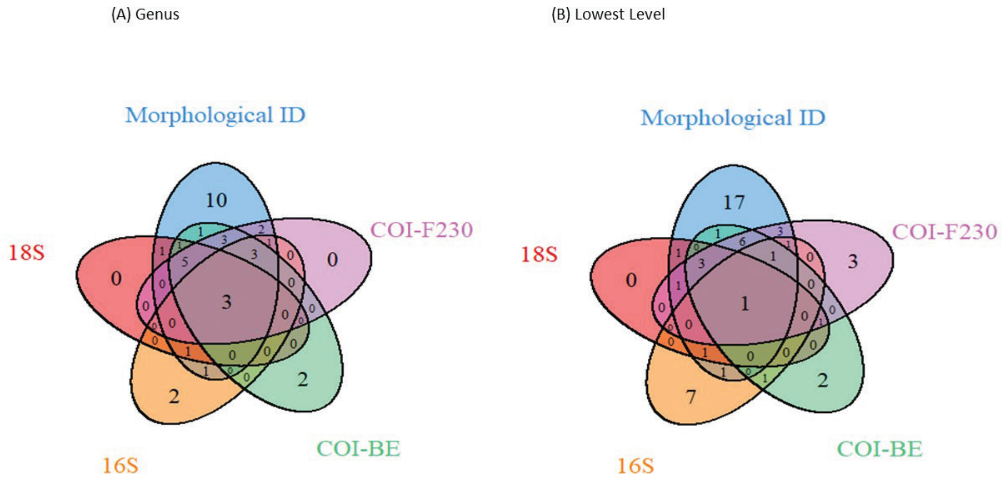

Figure 2. Venn Diagrams illustrating overlap in genus-level and lowest level (mostly species, but genus-level if no species were identified) taxa counts using each identification approach.

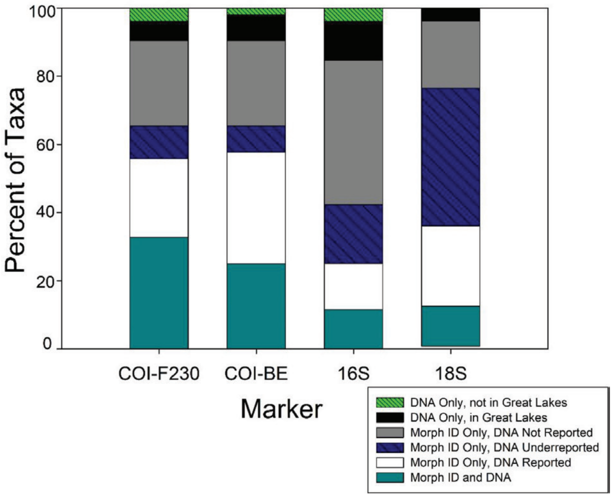

Figure 3. Percent of lowest level taxonomic records for each marker that were identified by both morphological identification and DNA, morphological identification only and DNA only. For morphological identification-only records, the label indicates the percentage of taxa that were reported (> 2 reference barcodes), under-reported (< 2 reference barcodes) or not reported (0 reference barcodes) in DNA libraries. For DNA-only records, the green bar indicates the percentage of taxa that has not been previously reported for the Great Lakes.

Several of the taxa, identified only by DNA metabarcoding, are ones for which the current Great Lakes fauna inventory (Trebitz et al. 2019; https://www.glerl.noaa.gov/data/waterlife/index.html) reports no prior morphologically-identified presences in Lake Superior at all – namely for *D. cucullata*, *C. dubia*, *H. ingens*, *Calanus* sp. and *S. pallidus*. Morphological identification did find the genera *Daphnia* and *Skistodiaptomus*, but without further identifying characteristics would have presumed they were one of the native species from these genera. While *Ceriodaphnia* appears regularly in surveys, this taxon is typically only identified to genus. *Daphnia cucullata*, *C. dubia* and *S. pallidus* matched sequences in BOLD or NCBI with nearly 100 percent similarity and 100 percent overlap in taxonomic coverage. The barcode assigned to *Calanus* sp. matched multiple species in the genus *Calanus* at 100 percent similarity, but all exhibited only 33 percent overlap with the reference barcode (e-value = 4e-43). The barcode assigned to *H. ingens* matched with 100% similarity, but had only 56% taxonomic overlap with the reference barcode (e-value = 0.004) and, considering that 16S did not perform well at identifying diaptomids, this identification is the most speculative. In addition to these taxa not previously identified for Lake Superior, *D. dentifera* is rare for the Great Lakes (Kerfoot et al. 2004) and was found at only one site. The barcode sequence assigned to *D. dentifera* showed high percent similarity (100%) to this species, with 86% taxonomic overlap with the reference barcode. However, several other taxa, including *Daphnia longispina* and *Daphnia galeata*, showed lesser levels of similarity, but higher levels of overlap with reference sequences.

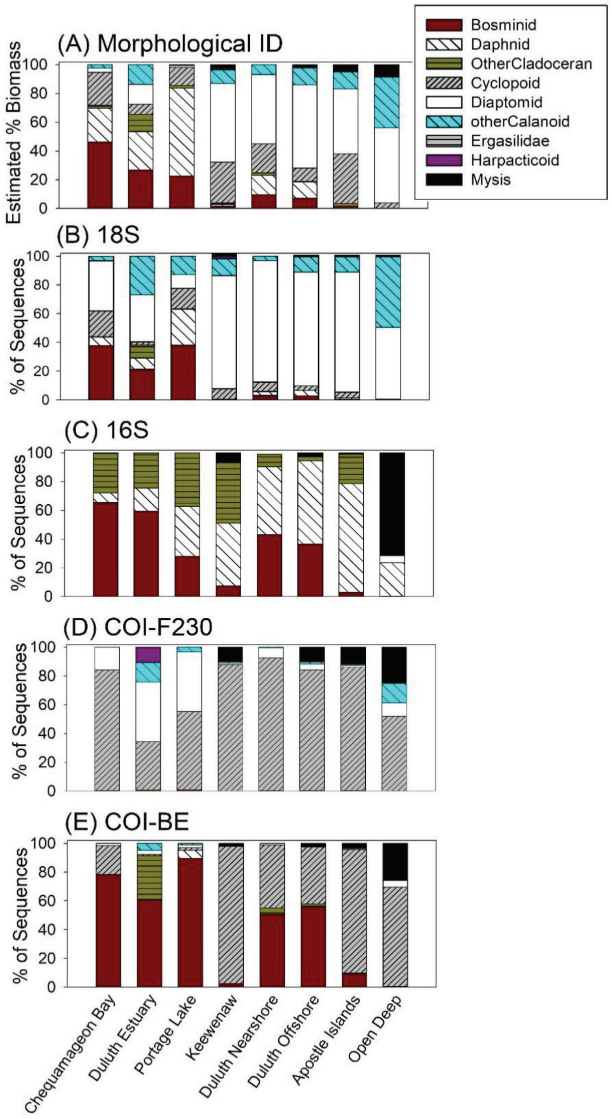

Figure 4. Estimated percent biomass identified for each major taxonomic group by: (A) morphological identification and percent of sequences in each taxonomic group for each DNA marker, including (B) 18S, (C) 16S, (D) COI with F230 primer and (E) COI with BE primer. Geographic zones from Figure 1 are presented from left to right in general order of increasing depth.

Table 3. Spearman correlation matrix comparing: (A) NMDS ordination axis 1 scores and (B) NMDS ordination axis 2 scores computed from Bray-Curtis similarities across site-level estimated percent biomass of organisms (for morphological identification) or percent abundance of sequences (for the DNA approaches).

Method	Morph ID	COI-BE	COI-F230	18S	16S
A) NMDS axis 1					
Morph identification	1.00				
COI-BE	0.86	1.00			
COI-F230	0.36	0.42	1.00		
18S	0.95	0.88	0.41	1.00	
16S	-0.035	-0.12	0.070	-.054	1.00
B) NMDS axis2					
Morph identification	1.00				
COI-BE	-0.22	1.00			
COI-F230	-0.014	0.72	1.00		
18S	0.45	-0.61	-0.38	1.00	
16S	0.30	-0.56	-0.43	0.63	1.00

Comparing relative biomass to relative sequence abundance within broad taxonomic groups, the metabarcoding methods varied greatly in their ability to characterise zooplankton composition (Figs 4, 5). Most notably, the two COI marker fragments greatly over-represented cyclopoids relative to morphological identification, while under-representing many other taxa. However, the two COI fragments differed overall in the taxa they targeted. The COI-F230 best captured percent biomass patterns in cyclopoids, other (non-diaptomid) calanoids and mysids. The COI-BE best captured percent biomass patterns in cyclopoids, other cladocerans (non-daphnids and bosminids), bosminids and mysids. Neither was able to capture trends in the percent biomass of daphnids. The 16S marker over-represented bosminids and generally did not identify other taxonomic groups. The 18S slightly over-represent-

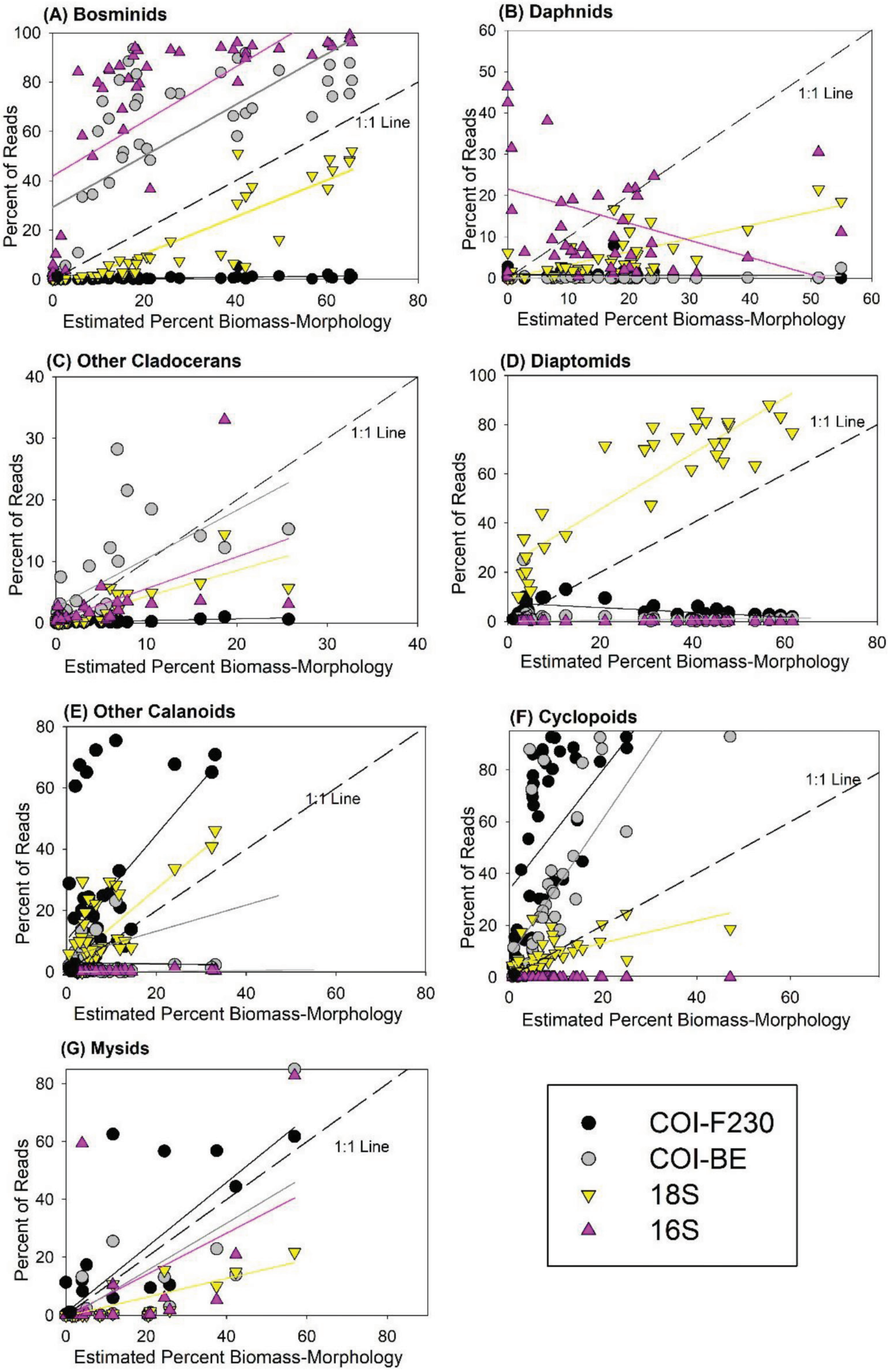

Figure 5. Relationship between estimated percent biomass using morphological identification and percent of reads for each major taxonomic group for the four DNA approaches.

ed other calanoids and did not fully capture cyclopoids compared to their estimated percent biomass (trendline compared to 1:1 line), but, in general, the trends across taxonomic groups approximated patterns in percent biomass obtained using morphological identification.

The correlations amongst NMDS axes illustrated general trends in relative composition across sites (Table 3 and Fig. 6). For morphological identification, 18S and COI-BE, the NMDS axis 1 was negatively correlated with the proportional abundance of the cladoceran groups (bosminids, daphnids and other cladocerans) and positively correlated with the proportion of diaptomid calanoids. For morphological identification, 18S and 16S, the NMDS axis 2 was strongly positively correlated with both the proportion of cladocerans and the proportion of cyclopoids. The 18S marker was the only marker which approximated morphological identification for both NMDS axes (Fig. 5; Table 3)

Across taxa, the detectability (estimated proportion of replicates in which a taxon was found using at least one marker) was considerably higher (0.55 +/- 0.37 SD) with DNA metabarcoding than the proportion of replicates where the taxon was found using traditional morphological identification (0.33 +/- 0.32 SD) (Fig. 7). Several taxa had too few replicates at occupied sites to accurately estimate detectability for either morphological identification or DNA, including *D. dentifera*, D. *longiremis*, *D. pulex*, *E. lamellatus*. *H. ingens*, *Calanus* sp. and *S. reighardi*.

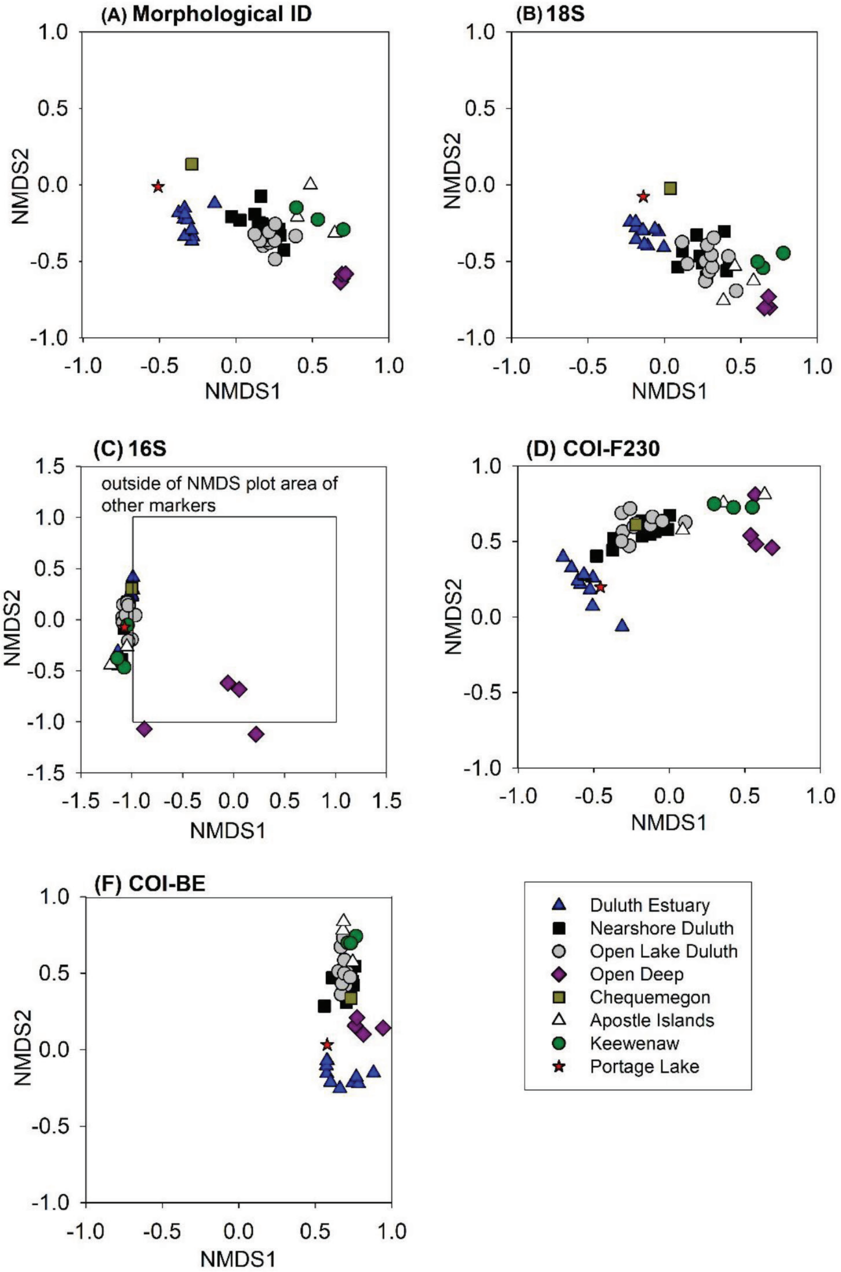

Figure 6. NMDS plots representing the differences in species composition amongst geographic zones for: (A) morphological identification, (B) 18S, (C) 16S, (D) COI with F230 primer and (E) COI with BE primer. The NMDS analysis was based on percent estimated biomass for morphological identification and percent sequence abundance for the DNA markers.

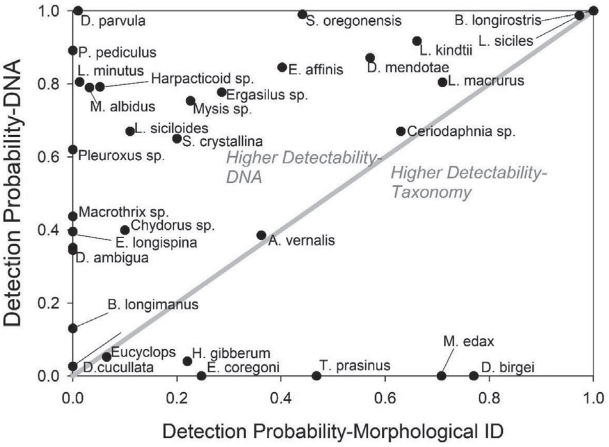

Figure 7. Detection probabilities across replicate Duluth/Lake Explorer samples using DNA metabarcoding (across all 4 DNA methods combined) versus morphological identification. Only taxa with DNA barcodes in at least one online database are included. The 1:1 line illustrates the relationship if detection probabilities were equal. Taxa with higher detectability using morphological identification plot to the right of the line while taxa with higher detectability with DNA identification plot to the left.

A taxon was considered to have zero detectability if it were not present with a given method (e.g. morphological identification or DNA), but was found with the other method, even though detectability could not actually be measured due to zero occupancy. Several individual taxa had higher detectability using morphological identification than metabarcoding. For some taxa, low DNA detectability occurred because DNA identified the taxa at the genus-level compared to identification at the species-level for morphological identification. These included *D. birgei*, *T. prasinus*, *E. coregoni* and *H. gibberum*. While species-level records for *Mesocyclops* are found in online databases, no *Mesocyclops* were found using DNA metabarcoding. *Eucyclops* also had higher detectability using morphological identification than DNA metabarcoding, although detectability was low for both methods. All other taxa had a higher detectability using DNA metabarcoding than morphological identification. Notably, invasive *B. longimanus* was only identified using DNA-based metabarcoding; despite being present at a high rate in the Lake Superior-Guardian samples, this species was not detected in Duluth-LEII samples (which were used for occupancy analysis) with morphological identification. This finding illustrates the enhanced ability to find difficult-to-identify or rare taxa using DNA metabarcoding for identification.

Discussion

Our goal was to evaluate the current capability of DNA metabarcoding versus morphological identification for the characterisation of genus- and species-level zooplankton diversity and broad-scale patterns in relative biomass of dominant taxonomic groups in Lake Superior. Our research highlights the importance of using multiple markers for the detection of rare crustaceous zooplankton taxa at the species-level as well as the potential usefulness of the 18S marker with our chosen primer for monitoring broad shifts in composition in Lake Superior.

Similar to other research (Chain et al. 2016; Yang et al. 2017a), we found that DNA-based and morphological identification together yielded a greater diversity of zooplankton than either technique independently. We also found that more taxa were identified because we used multiple DNA markers. A recent study of mock zooplankton communities found that using multiple DNA markers improved species detection rates by 11–30% (Zhang et al. 2018). In our study, we found that 35% of taxa detected using DNA metabarcoding were only detected with one marker. Despite the better performance of COI for high-resolution identification, numerous taxa that were rare in samples, including invasive *B. longimanus* and *H. gibberum*/*glacialis*, were better detected using 18S or 16S, further highlighting the importance of multiple markers for the detection of rare taxa.

The fact that we found 13 taxa with DNA metabarcoding that were not detected by morphological identification may, in part, be attributed to the differing “depth” to which these techniques delve into a sample. The GLNPO method for morphological identification of crustaceous zooplankton fully enumerates only a subsample of up to 400 individuals, while scanning additional sample fractions for larger and rarer taxa (GLNPO 2016). Across all our samples combined, we estimate that 21% of zooplankton individuals collected were actually morphologically identified. DNA metabarcoding also examines only a fraction of the sample (in this case ~ 50%), but that fraction is taken after the DNA is extracted, at which point all species should theoretically be present, if the sample is well mixed. In addition, the success of DNA-based identification does not depend on organism life stage, whereas morphological identification often can assign only coarse-level taxonomy to immature life stages. Some of the taxa that were only found in DNA are known to be present in Lake Superior (Trebitz et al. 2019) and, thus, could have been missed by the morphological identification process due to their being too immature to fully identify. This inability may risk failing to discern a new species from one already present (e.g. amongst *Skistodiaptomus* species). *Skistodiaptomus reighardi* is present in Lake Superior, while *S. pallidus* is present in other Great Lakes, but has not yet been found in Lake Superior (NOAA and USEPA 2019; https://www.glerl.noaa.gov/data/waterlife/index.html).

We recognise that our results yielded four suspect genera or species identifications. *D. cucullata* and *H. ingens* are native to Northern Europe/Asia and North Africa, respectively and have not previously been found in the Great Lakes (personal communication, Joseph Connolly, 19 June 2019). The genus *Calanus* represents one of the most prolific zooplankton in the North Atlantic Ocean (Choquet et al. 2017), which has also not been identified for the Great Lakes. It is possible that all or one of these could have been transported to Lake Superior on shipping vessels. Finally, *C. dubia* is a highly prolific species found in much of the world and potentially in Lake Superior, but this species may have been missed because only genus-level identifications of *Ceriodaphnia* are performed by Great Lakes taxonomists (personal communication, Heidi Schaefer; 25 June 2018). A review of matches to barcode libraries indicated that at least two of these (*D. cucullata* and *C. dubia*) have a high likelihood of being positive identifications. Suspect identifications should be further investigated before concluding they are correct. In the future, this could be accomplished by having taxonomists search the remainder of the sample for the suspect taxon or taxa. In some cases, especially in the case of potential invaders that are of-concern, additional field samples in the location of the suspect taxon may also be warranted. False positives are possible with DNA-based methods as a result of sequencing errors, contamination, primer bias and choices made in bioinformatics processing or the presence of closely-related taxa with a similar DNA signature (Coissac et al. 2012). At numerous steps in both the laboratory and post-processing workflow, there are risks of increasing the likelihood of either false positives (i.e. detecting a species that is not present) or false negatives (i.e. not detecting a species that is present). Future laboratory studies of mock samples may be used to determine potential rates of false positives and false negatives and how laboratory replicates, which we did not use in this study, may help identify them.

Overall, the absence of many zooplankton taxa from reference DNA sequence databases no doubt resulted in more taxa being identified using morphological identification than DNA metabarcoding. The success of other similar studies assessing zooplankton using DNA metabarcoding may be attributed to the development of local reference databases (Bucklin et al. 2016; Zhang et al. 2018; Yang et al. 2017b). We likewise collected zooplankton voucher specimens with the goal of adding reference barcodes to NCBI and BOLD, but unfortunately, obtained few specimens of some of the rare species that were used up during unsuccessful Sanger sequencing efforts. It is possible that some of the newer High Throughput Sequencing (HTS) approaches for developing reference databases will help in conquering these hurdles (Yang et al 2017a; Beninde 2020). Efforts to establish DNA metabarcoding as a focal method for characterising zooplankton communities on the Great Lakes may need to await build-out of barcode reference databases, an activity which is currently in progress.

Despite the limitations of not having a local reference database, our findings yield valuable information about the detection capability of different taxa using the selected DNA markers and primers. Results are consistent with a number of other studies showing better performance of COI compared to 18S and 16S for estimating species-level diversity of zooplankton (Tang et al. 2012; Clarke et al. 2017). Due to the lower genetic variability in the 18S target region, the 18S marker typically does not yield the deeper taxonomic resolution of the COI marker (Tang et al. 2012). We also specifically had an inability to fully identify individual cladoceran species using 18S, which is consistent with other findings (Brown et al. 2015; Clarke et al. 2017). More generally, we had less success in resolving high-resolution taxa using 18S than some other studies have found (Clarke et al. 2017; Zhang et al. 2018). This may be because our primer focused on the V2-V3 genomic region, which is not typical of zooplankton studies using 18S (but see Lindeque et al. 2013). Most taxonomic reads were distributed amongst a relatively low amount of OTUs. Therefore, clustering choices may have also resulted in fewer species-level assignments. Brown et al. (2015) advocates that zooplankton assignment to OTUs be done without clustering and that different thresholds be set for specific taxonomic groups of zooplankton. Various de-noising methods may also be employed to help with this issue (Prodan et al. 2020), which we may investigate in the future.

Our findings are also similar to other research showing that COI markers were not able to reproduce the relative abundances obtained using morphological identification, likely due to primer amplification bias (Pinol et al. 2015; Bucklin et al. 2016). Like other zooplankton metabarcoding studies, our COI marker-variants targeted the COI standard barcode region (Clarke et al. 2017; Yang et al. 2017b; Zhang et al. 2018). We found a disproportionate amount of cyclopoid copepods, which were over-amplified by our primer. However, findings published after the initiation of our study have shown the success of the COI marker at both identifying the majority of zooplankton species and closely approximating relative abundance. These studies used degenerate primers targeting the Leray fragment (Clarke et al. 2017; Yang et al. 2017b). The challenge of using degenerate primers is that they must be closely tailored to the taxa of interest. Nonetheless, these recent findings indicate the potential to determine both diversity and composition with the COI marker.

Very few studies have examined the use of the 16S marker for zooplankton, which is considered to be between 18S and COI with regard to both avoiding taxonomic bias, but retaining taxonomic resolution (Clarke et al. 2017). In our study, the low number of taxa in barcode libraries potentially contributed to a lack of well-resolved identity assignments for 16S. However, even the broad-scale taxonomic groups of zooplankton have some representation in online databases at the family and higher level, while only non-daphnid cladocerans were well detected with our 16S method. Another study, using different and longer 16S sequences, captured primarily calanoids (Clarke et al. 2017). These findings suggest that our chosen primer was not adequate for detecting a full range of diversity in zooplankton; however, it did result in detection of four less common cladoceran species not identified by the other markers.

Although we do not have an absolute “truth” to compare in our study, it is useful to compare our results to the traditional GLNPO morphological identification because this has been used historically for zooplankton assessment on the Great Lakes. Recently, a study using long-term data illustrated broad-scale trends in calanoids, daphnids and cyclopoids in the Great Lakes over the last 20 years (Barbiero et al. 2019). Our results suggest that these long-term trends in taxonomic groups would be adequately captured with the 18S marker and our chosen primer. While there were some differences in correlation between percent sequence abundance using 18S and percent biomass from morphological identification, this may be partly due to errors in our assumptions regarding biomass assigned to the different taxonomic groups and could be improved by further incorporating size distribution data. In addition to the potential to document trends in composition, our occupancy modelling approach highlighted the benefits of DNA metabarcoding for detection of rare taxa. DNA metabarcoding has similarly been used to detect rare zooplankton taxa in ballast water (Rey et al. 2019). The species *L. siciloides* and *C. sphaericus*, which we detected using only DNA metabarcoding, may serve as important early-warning indicators of water quality change (Barbiero et al. 2001). We also found that DNA metabarcoding was more successful at detecting *Bythotrephes*, an invasive species that has a large impact on the ecosystems of the Great Lakes (Yan et al. 2011).

Zooplankton data are widely used for assessing water quality changes, trophic interactions and new species invasions on the Great Lakes. Often, the potential future use of the data is often not known a priori (Barbiero et al. 2019; Pawlowski and Sierszen 2020). At present, DNA-based zooplankton data are not ready to replace or significantly augment morphologically-based identification as the basis for addressing such ecology questions due to a combination of still-evolving laboratory practices, primer and marker capabilities and currently expanding barcode reference libraries. However, our findings highlight that DNA metabarcoding can improve our characterisation of zooplankton communities over what is possible with morphological identification alone, specifically with regards to potential broad-scale taxonomic changes and detection of rare taxa, including potential new invaders. Improvements will continue to be made as our efforts incorporate ever-increasing knowledge regarding primer choice and laboratory best management practices, as well as development of regionally-appropriate reference databases.

Data accessibility

Taxonomic abundance data, OTU data and R code for the DNA versus taxonomic comparison and occupancy analysis can be found in a dedicated Open Source Framework site at https://doi.org/10.17605/OSF.IO/ABNGX. Raw sequence reads have been archived on NCBI with the Accession number PRJNA728961.

Acknowledgements

Funds from the Great Lakes Restoration Initiative supported the position of author C. Meredith and S. Okum. We are grateful for the comments and suggestions provided by two anonymous reviewers, which greatly improved the manuscript. We thank the staff of the Lake Guardian as well as participants and instructors of the 2016 Center for Great Lakes Literacy Teacher Cruise, with special mention to the zooplankton team of Nancy Hummel, Jacob Peterson, Deborah Campbell, Ann Quinn, and Ashlee Giordano. Additionally, we thank Captain Sam Miller and crew of the 2016 R/V Lake Explorer zooplankton trips, as well as Jill Scharold, Amelia Flannery, Leah Mohn, John Barge and Tim Corry for field and laboratory assistance. The Balcer Lab (Superior, Wisconsin) performed the morphological identification. All DNA preparation and analysis occurred at the EPA's Office of Research and Development in Cincinnati, OH.

References

Abell R, Thieme M, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N, Balderas S, Bussing W, Stiassny M, Skelton P, Allen G, Unmack P, Naseka A, Ng R, Sindorf N, Robertson J, Armijo E, Higgins J, Heibel T, Wikramanayake E, Olson D, López H, Reis R, Lundberg J, Sabaj Pérez M, Petry P (2008) Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation, BioScience 58(5): 403–414. https://doi.org/10.1641/B580507

Barbiero RP, Little RE, Tuchman ML (2001) Results from the U.S. EPA's Biological Open Water Surveillance Program of the Laurentian Great Lakes: III. Crustacean Zooplankton International Association of Great Lakes Research 2 (27): 167–184. https://doi.org/10.1016/S0380-1330(01)70628-4

Alekseev V, Dumont HJ, Pensaert J, Baribwegure D, Vanfleteren JR (2006) A redescription of *Eucyclops serrulatus* (Fischer, 1851) (Crustacea: Copepoda: Cyclopoida) and some related taxa, with a phylogeny of the *E. serrulatus*-group. Zoologica Scripta 35(2): 123–147. https://doi.org/10.1111/j.1463-6409.2006.00223.x

Audzijonytė A, Väinölä R (2005) Diversity and distributions of circumpolar fresh- and brackish-water *Mysis* (Crustacea: Mysida): descriptions of *M. relicta* Lovén, 1862, *M. salemaai* n. sp., *M. segerstralei* n. sp. and *M. diluviana* n. sp., based on molecular and morphological characters. Hydrobiologia 544: 89–141. https://doi.org/10.1007/s10750-004-8337-7

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation sequencing. Molecular Ecology 21: 2039–2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x

Barbiero RP, Rudstam LG, Watkins JM, Lesht BM (2019) A cross-lake comparison of crustacean zooplankton communities in the Laurentian Great Lakes, 1997–2016. Journal of Great Lakes Research 45: 672–690. https://doi.org/10.1016/j.jglr.2019.03.012

Belyaeva M, Taylor D (2009) Cryptic species within the *Chydorus sphaericus* species complex (Crustacea: Cladocera) revealed by molecular markers and sexual stage morphology. Molecular Phylogenetics and Evolution 50: 534–546. https://doi.org/10.1016/j.ympev.2008.11.007

Beninde J, Möst M, Meyer A (2020) Optimized and affordable high-throughput sequencing workflow for preserved and non-preserved small zooplankton specimens. Molecular Ecology Resources 20: 1632–1646. https://doi.org/10.1111/1755-0998.13228

Blaxter M, Ley D, Garey J, Liu L, Scheldeman P, Vierstraete A, Vanfleteren J, Mackey L, Dorris M, Frisse M, Vida J, Thomas W (1998) A molecular evolutionary framework for the phylum Nematoda. Nature 392: 71–75. https://doi.org/10.1038/32160

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of Southern Wisconsin. Ecological Monographs 27: 325–349. https://doi.org/10.2307/1942268

Brown E, Chain F, Crease T, MacIsaac H, Cristescu M (2015) Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities. Ecology and Evolution 5: 2234–2251. ttps://doi.org/10.1002/ece3.1485

Brown E, Chain F, Zhan A, MacIsaac H, Cristescu M (2016) Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. Diversity and Distributions 22: 1045–1059. https://doi.org/10.1111/ddi.12465

Bucklin A, Lindeque PK, Rodriguez-Ezpeleta N, Albaina A, Lehtiniemi M (2016) Metabarcoding of marine zooplankton: prospects, progress and pitfalls. Journal of Plankton Research 338: 393–400. https://doi.org/10.1093/plankt/fbw023

Burlakova LE, Hinchey EK, Karatayev AY, Rudstam LG (2018) US EPA Great Lakes National Program Office monitoring of the Laurentian Great Lakes: Insights from 40 years of data collection. Journal of Great Lakes Research 44: 535–538. https://doi.org/10.1016/j.jglr.2018.05.017

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13: 581–583. https://doi.org/10.1038/nmeth.3869

Chain F, Brown EA, MacIsaach HJ, Cristescu ME (2016) Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. Biodiversity Research 22: 493–504. https://doi.org/10.1111/ddi.12427

Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12. https://doi.org/10.1186/1471-2105-12-35

Choquet M, Hatlebakk M, Dhanasiri AKS, Kosobokova K, Smolina I, Søreide JE, Svensen C, Melle W, Kwaśniewski S, Eiane K, Daase M, Tverberg V, Skreslet S, Bucklin A, Hoarau G (2017) Genetics redraws pelagic biogeography of Calanus 13. https://doi.org/10.1098/rsbl.2017.0588

Clarke L, Beard J, Swadling K, Deagle B (2017) Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution 7: 873–88. https://doi.org/10.1002/ece3.2667

Coissac E, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. Molecular Ecology 21: 1834–1847. https://doi.org/10.1111/j.1365-294X.2012.05550.x

Dodson S (1994) Morphological analysis of Wisconsin (U.S.A.) species of the *Acanthocyclops vernalis* group. Journal of Crustacean Biology 14: 113–131. https://doi.org/10.1163/193724094X00515

Edgar R (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Fiske I, Chandler R (2011) Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. Journal of Statistical Software 43: 1–23. https://doi.org/10.18637/jss.v043.i10

Fonseca V, Carvalho G, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lambshead PJD, Kelley Thomas W, Creer S (2020) Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nature Communications 1: e98. https://doi.org/10.1038/ncomms1095

GLNPO (Great Lakes National Program Office) (2016) Standard operating procedure for zooplankton analysis. U. S. EPA Great Lakes National Program Office, method LG403. Revision 07.

Gibson J, Shokralla S, Curry C, Baird D, Monk W, King I, Hajibabaei M (2015) Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. PLoS ONE 10: e0138432. https://doi.org/10.1371/journal.pone.0138432

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 JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé 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 Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41: D597–D604. https://doi.org/10.1093/nar/gks1160

Hoffman J, Kelly J, Trebitz A, Peterson G, West C (2011) Effort and potential efficiencies for aquatic non-native species early detection. Canadian Journal of Fisheries and Aquatic Sciences 68: 2064–2079. https://doi.org/10.1139/f2011-117

Kerfoot W, Ma X, Lorence C, Weider L (2004) Toward Resurrection Ecology: Daphnia mendotae and D. Retrocurva in the coastal region of Lake Superior: Among the first successful outside invaders? Journal of Great Lakes Research 30: 285–299. https://doi.org/10.1016/S0380-1330(04)70392-5

Kotov A, Seiji S, Taylor D (2009) Revision of the genus *Bosmina Baird*, 1845 (Cladocera: Bosminidae) based on evidence from male morphological characters and molecular phylogenies. Zoological Journal of the Linnean Society 156: 1–51. https://doi.org/10.1111/j.1096-3642.2008.00475.x

Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, 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

McMurdie PJ, Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Computational Biology. https://doi.org/10.1371/journal.pcbi.1003531

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12. https://doi.org/10.14806/ej.17.1.200

Naidoo R, Balmford A, Constanza R, Fisher B, Green R, Lehner G, Malcolm T, Ricketts T (2008) Global mapping of ecosystem services and conservation priorities. Proceedings of the National Academy of Sciences 105: 9495–9500. https://doi.org/10.1073/pnas.0707823105

NOAA and USEPA (2019) Great Lakes Waterlife. https://www.glerl.noaa.gov/data/waterlife/index.html [Accessed on 01/27/2021]

Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, Henry M, Stevens H, Szoecs E, Wagner H (2019) vegan: Community Ecology Package. R package version 2.5-4. https://CRAN.R-project.org/package=vegan

Palumbi SR (1996) What can molecular genetics contribute to marine biogeography? An urchin's tale. Journal of Experimental Marine Biology and Ecology 203: 75–92. https://doi.org/10.1016/0022-0981(96)02571-3

Pawlowski MB, Sierszen ME (2020) A lake-wide approach for large lake zooplankton monitoring: Results from the 2006–2016 Lake Superior Cooperative Science and Monitoring Initiative Surveys. Journal of Great Lakes Research 46: 1015–2017. https://doi.org/10.1016/j.jglr.2020.05.005

Piñol J, Mir G, Gomez-Polo P, Agustí N (2005) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources 15: 819–830. https://doi.org/10.1111/1755-0998.12355

Poole K, Dowling J (2004) Relationship of declining mussel biodiversity to stream-reach and watershed characteristics in an agricultural landscape. Journal of the North American Benthological Society 23: 114–125. https://doi.org/10.1899/0887-3593(2004)023<0114:RODMBT>2.0.CO;2

Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E (2020) Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS ONE 15(1): e0227434. https://doi.org/10.1371/journal.pone.0227434

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. https://doi.org/10.1093/nar/gks1219

Reuter J, Spacek D, Snyder M (2015) High-Throughput Sequencing Technologies. Molecular Cell 58: 586–597. https://doi.org/10.1016/j.molcel.2015.05.004

Rey A, Carney KJ, Quinones LE, Pagenkopp Lohan KM, Ruiz GM, Basurko O, Rodriguez-Ezpeleta N (2019) Environmental DNA metabarcoding: a promising tool for ballast water monitoring. Environmental Science and Technology 53(20): 11849–11859. https://doi.org/10.1021/acs.est.9b01855

Rowe C, Adamowicz S, Herbert P (2007) Three new cryptic species of the freshwater zooplankton genus *Holopedium* (Crustacea: Branchiopoda: Ctenopoda) revealed by genetic methods. Zootaxa 1656: 1–49. https://doi.org/10.11646/zootaxa.1656.1.1

Tang C, Leasi F, Obertegger U, Kieneke A, Barraclough T, Fontaneto D (2012) The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences 109: 16208–16212. https://doi.org/10.1073/pnas.1209160109

Trebitz A, Hoffman J, Grant G, Billehus T, Pilgrim E (2015) Potential for DNA-based identification of Great Lakes fauna: match and mismatch between taxonomic inventories and DNA barcode libraries. Scientific Reports 5: e12162. https://doi.org/10.1038/srep12162

Trebitz A, Sykes M, Barge J (2019) A reference inventory for aquatic fauna of the Laurentian Great Lakes. Journal of Great Lakes Research 45: 1036–1046. https://doi.org/10.1016/j.jglr.2019.10.004

Vander Zanden M, Olden J (2008) A management framework for preventing the secondary spread of aquatic invasive species. Canadian Journal of Fisheries and Aquatic Sciences 65: 1512–1522. https://doi.org/10.1139/F08-099

Vos P, Meelis E, Ter Keurs W (2000) A framework for the design of ecological monitoring programs as a tool for environmental and nature management. Environmental Monitoring and Assessment 61: 317–344. https://doi.org/10.1023/A:1006139412372

Vasquez A, Hudson P, Fujimoto M, Keeler K, Armenio P, Ram J (2016) *Eurytemora carolleeae* in the Laurentian Great Lakes revealed by phylogenetic and morphological analysis. Journal of Great Lakes Research: 802–811. https://doi.org/10.1016/j.jglr.2016.04.001

Yan N, Leung B, Lewis M, Peacor S (2011) The spread, establishment and impacts of the spiny water flea, *Bythotrephes longimanus*, in temperate North America: a synopsis of the special issue. Biological Invasions 13: e2423. https://doi.org/10.1007/s10530-011-0069-9

Yang J, Zhang X, Xie Y, Song C, Zhang Y, Yu H, Burton A (2017) Zooplankton community profiling in a eutrophic freshwater ecosystem-Lake Tai basin by DNA Metabarcoding. Scientific Reports 7: e1773. https://doi.org/10.1038/s41598-017-01808-y

Yang J, Zhang X, Zhang W, Sun J, Xie Y, Zhang Y, Burton Jr GA, Yu H (2017) Indigenous species barcode database improves the identification of zooplankton. PLoS ONE 12(10): e0185697. https://doi.org/10.1371/journal.pone.0185697

Yurista P, Kelly J, Miller S (2009) Lake Superior zooplankton biomass: Alternate estimates from a probability-based net survey and spatially extensive LOPC surveys. Journal of Great Lakes Research 35: 337–346. https://doi.org/10.1016/j.jglr.2009.03.004

Zaiko A, Martinez J, Schmidt-Petersen J, Ribicic D, Samuiloviene A, Garcia-Vazquez E (2015) Metabarcoding approach for the ballast water surveillance-An advantageous solution or an awkward challenge. Marine Pollution Bulletin 92: 25–34. https://doi.org/10.1016/j.marpolbul.2015.01.008

Zhan A, Bailey S, Heath D, MacIsaac H (2014) Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. Molecular Ecology Resources 14: 1049–1059. https://doi.org/10.1111/1755-0998.12254

Zhang G, Chain F, Abbott C, Cristescu (2018). Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications 11: 1901–1914. https://doi.org/10.1111/eva.12694

Appendix 1

Number of sites where each taxon was detected using morphological identification and each marker. If a species-level identification occurred at a site, the observation was counted at both the genus- and species-levels.

	Taxonomy	COI-230	COI-BE	16S	18S
Cladocera					
Acroperus harpae	1	UR	UR	UR	UR
Bosmina longirostris/liederi	40	31	43	37	UR
Bythotrephes longimanus	6	0	0	11	8
Ceriodaphnia *	20	12	19	0	19
[Ceriodaphnia dubia]	0	12	14	UR	14
Chydorus sphaericus/brevilabrus	3	2	1	UR	3
*Daphnia**	36	39	24	43	38
Daphnia ambigua	0	0	0	2	UR
[Daphnia cucullata]	0	2	0	0	UR
Daphnia dentifera	0	0	0	1	0
Daphnia longiremis	0	UR	2	1	UR
Daphnia (galeata) mendotae	20	39	16	41	NR
Daphnia parvula	1	UR	UR	34	UR
Daphnia pulex	0	0	0	3	0
Daphnia retrocurva	35	NR	NR	NR	NR
*Diaphanosoma**	31	20	35	42	NR

	Taxonomy	COI-230	COI-BE	16S	18S
Diaphanosoma birgei	31	0	NR	NR	NR
Eubosmina	3	NR	NR	9	0
Eubosmina coregoni	3	UR	UR	0	UR
Eubosmina longispina	0	0	0	9	0
Eurycercus lamellatus	1	0	0	UR	UR
*Holopedium**	17	1	10	26	0
Holopedium gibberum/glacialis	17	1	1	UR	UR
Ilyocryptus acutiofrons	1	NR	NR	UR	UR
Kurzia latissima	1	NR	NR	NR	UR
Latona setifera	3	NR	NR	NR	UR
Leptodora kindtii	28	0	34	0	0
Macrothrix	0	0	4	UR	UR
Monospilus dispar	1	NR	NR	NR	NR
Polyphemus	1	1	0	38	0
Polyphemus pediculus	1	1	0	38	0
Pleuroxus	0	0	1	UR	0
Sida	4	2	12	11	10
Sida crystallina	4	2	12	11	10
Cyclopoida					
*Acanthocyclops**	8	12	27	NR	0
Acanthocyclops vernalis/americanus	8	8	0	NR	0
*Cyclops**	43	43	43	NR	43
Cyclops (*Diacyclops*) *thomasi*	43	NR	NR	NR	NR
Ergasilus	6	7	10	NR	0
Eucyclops	1	UR	UR	NR	3
Eucyclops agilis/serrulatus	1	UR	UR	NR	3
*Macrocyclops**	1	2	1	0	10
Macrocyclops albidus	1	2	0	0	10
*Mesocyclops**	26	0	0	NR	0
Mesocyclops americanus	2	NR	NR	NR	NR
Mesocyclops edax	20	0	0	NR	0
Paracyclops chiltoni	1	NR	NR	NR	NR
*Tropocyclops**	7	1	0	UR	0
Tropocyclops prasinus	7	NR	NR	NR	NR
Calanoida					
Calanus	0	0	0	1	0
Epischura lacustris	35	NR	NR	NR	NR
*Eurytemora**	22	41	33	NR	33
Eurytemora affinis/carolleeae	15	41	33	NR	39
[Hemidiaptomus ingens]	0	NR	NR	6	UR
*Leptodiaptomus**	42	43	43	NR	0
Leptodiaptomus ashlandi	3	NR	NR	NR	UR
Leptodiaptomus minutus	2	35	23	NR	UR
Leptodiaptomus sicilis	41	43	39	NR	UR
Leptodiaptomus siciloides	6	16	0	NR	UR
Skistodiaptomus	17	43	34	0	0
Skistodiaptomus oregonensis	17	43	34	0	0
Skistodiaptomus pallidus[1]	0	9	0	0	0
Skistodiaptomus reighardi	0	3	0	UR	UR
Limnocalanus macrurus	36	38	14	NR	UR
Senecella calanoides	3	NR	NR	NR	NR
Harpacticoida	3	32	0	NR	0
Mysida					
*Mysis**	19	35	17	11	15
Mysis relicta complex	19	34	14	11	UR

[1] This Skisotdiaptomus is known from the lower Great Lakes, but NOT Lake Superior

*indicates that some observations occurred at the genus-level only

[] indicates never before found in the Great Lakes

NR = Not reported in online databases

UR = Under-reported in online databases (< 3 records)