

Lawrence Berkeley National Laboratory

LBL Publications

Title

Time-series metagenomics reveals changing protistan ecology of a temperate dimictic lake.

Permalink

<https://escholarship.org/uc/item/5cj3x624>

Journal

Microbiome, 12(1)

Authors

Krinos, Arianna
Bowers, Robert
Rohwer, Robin
et al.

Publication Date

2024-07-20

DOI

10.1186/s40168-024-01831-y

Peer reviewed

RESEARCH

Open Access



Time-series metagenomics reveals changing protistan ecology of a temperate dimictic lake

Arianna I. Krinos^{1,2,3,4*}, Robert M. Bowers⁴, Robin R. Rohwer⁵, Katherine D. McMahon⁶, Tanja Woyke⁴ and Frederik Schulz^{4*}

Abstract

Background Protists, single-celled eukaryotic organisms, are critical to food web ecology, contributing to primary productivity and connecting small bacteria and archaea to higher trophic levels. Lake Mendota is a large, eutrophic natural lake that is a Long-Term Ecological Research site and among the world's best-studied freshwater systems. Metagenomic samples have been collected and shotgun sequenced from Lake Mendota for the last 20 years. Here, we analyze this comprehensive time series to infer changes to the structure and function of the protistan community and to hypothesize about their interactions with bacteria.

Results Based on small subunit rRNA genes extracted from the metagenomes and metagenome-assembled genomes of microeukaryotes, we identify shifts in the eukaryotic phytoplankton community over time, which we predict to be a consequence of reduced zooplankton grazing pressures after the invasion of a invasive predator (the spiny water flea) to the lake. The metagenomic data also reveal the presence of the spiny water flea and the zebra mussel, a second invasive species to Lake Mendota, prior to their visual identification during routine monitoring. Furthermore, we use species co-occurrence and co-abundance analysis to connect the protistan community with bacterial taxa. Correlation analysis suggests that protists and bacteria may interact or respond similarly to environmental conditions. Cryptophytes declined in the second decade of the timeseries, while many alveolate groups (e.g., ciliates and dinoflagellates) and diatoms increased in abundance, changes that have implications for food web efficiency in Lake Mendota.

Conclusions We demonstrate that metagenomic sequence-based community analysis can complement existing efforts to monitor protists in Lake Mendota based on microscopy-based count surveys. We observed patterns of seasonal abundance in microeukaryotes in Lake Mendota that corroborated expectations from other systems, including high abundance of cryptophytes in winter and diatoms in fall and spring, but with much higher resolution than previous surveys. Our study identified long-term changes in the abundance of eukaryotic microbes and provided context for the known establishment of an invasive species that catalyzes a trophic cascade involving protists. Our findings are important for decoding potential long-term consequences of human interventions, including invasive species introduction.

Keywords rRNA, Metagenomics, Community ecology, Protistan ecology

*Correspondence:

Arianna I. Krinos
akrinos@whoi.edu
Frederik Schulz
fschulz@lbl.gov

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Protists are ubiquitous unicellular eukaryotic microbes capable of widespread dispersal and found globally across most ecosystems, from diverse terrestrial soil environments [1] to all reaches of the ocean [2]. Protists also encompass vast taxonomic and physiological diversity [3]. Lake ecosystems are no exception, where protists are key players in the microbial loop and determinants of lake trophic status [4]. Like bacteria, protists are essential in the movement of nutrients and energy through an ecosystem [5–7] but are behaviorally and morphologically more complex [5]. They can display diverse trophic strategies (e.g., mixotrophy [8]) that can alter freshwater food webs [9]. Protists may also colonize macroalgae and assist in carbon transformation, vitamin B₁₂ synthesis, and even silica cycling and storage, functions which have recently been predicted metagenomically [10]. Despite the importance of protists in ecosystem functioning, much less effort has been placed on understanding their biogeography, temporal patterns, and ecology as compared to bacteria. This has been in part due to the visible importance of cyanobacterial blooms [11], and in part due to challenges of defining and differentiating protistan morphology and function [12] or deciphering their complex genomes [5, 12], leaving a large knowledge gap on their geographic extent and lifestyle [5, 13].

Lake Mendota (Dane County, Madison, WI), a temperate dimictic lake, is considered among the world's best-studied freshwater systems due to the longevity and intensiveness of its monitoring program [11]. Previous studies of Lake Mendota have investigated the seasonal dynamics of the bacterial community in this and nearby lakes [14–16], including using amplicon sequencing approaches [15], but eukaryotic algae and their trophic linkages to bacteria were less frequently considered, though visually identified and counted in microscopy-based surveys [17]. Despite evidence for strong network relationships between bacteria and eukaryotes in both lacustrine and oceanic ecosystems [18], heterotrophic protists are frequently excluded from long-term bio-monitoring programs [19], including those in Lake Mendota. Metagenomic sequencing data have the potential to expand the known diversity of lake protists [13], and a recent study showed that 18S rRNA gene data from metagenomes can accurately recapitulate diatom abundance data from sediments [20].

Two invasive species reached high abundances in Lake Mendota during the period from 2000 to 2020 considered in this study. The spiny water flea (*Bythotrephes longimanus*) was first detected in Lake Mendota in 2009 [21] but was shown via sediment cores to have been present more than 10 years prior to net tow identification [22, 23]. Zebra mussels (*Dreissena polymorpha*) were

detected in the lake in 2015 [24]. Zebra mussels consume phytoplankton, which can result in “benthification” due to enhanced nutrient and light availability in normally shaded benthic lake habitats [24]. New zooplankton predators in lake ecosystems can shunt nutrients from the surface to depth [25]. Trophic cascades can emerge when introduced invasive species eliminate a trophic level, allowing a different taxon to become overabundant [15]. Ongoing microscopic count observations in Lake Mendota confirmed an increase in diatoms (class *Bacillariophyta*) and chlorophytes (phylum *Chlorophyta*) and a decrease in cryptophytes (phylum *Cryptophyta*) but lacked taxonomic resolution or the inclusion of heterotrophic protists that cannot be easily captured by net tows or visually identified [26].

To explore protistan ecology and interactions in Lake Mendota, we analyzed the diversity, structure, and composition of the lake's protistan community over the course of 20 years. We have extracted 16S and 18S rRNA gene sequences and quantified the abundance of these sequences in the underlying raw read data. We enhanced the taxonomic resolution of Lake Mendota's eukaryotic microbial community and explored how it has changed in association with prokaryotic community members over a 20-year time series. Furthermore, we introduce the first curated eukaryotic metagenome-assembled genomes (MAGs) from the same long-term metagenomic time series. We show that some of these MAGs can be linked to 18S rRNA gene data and compare relative abundance of highly complete MAGs in the time series compares to using the 18S rRNA gene. Using enriched metagenomes (metagenomes sequenced from samples with exceptionally high concentrations of the target taxon) sequenced from samples with high abundance of two key metazoans in the lake, *Daphnia pulicaria* (water flea) and *Bythotrephes longimanus* (spiny water flea), we binned MAGs to track environmental abundances of these key metazoans. We tested whether the spiny water flea was present in low abundance in the lake prior to its visual identification in 2009. Our long-term metagenomic study of lake protists demonstrates the power of time series metagenomics to transform our knowledge of freshwater protists and their contribution to lake biodiversity and ecology.

Methods

Twenty years of metagenome data from Lake Mendota

Lake Mendota is a eutrophic, temperate lake that is heavily influenced by both urban and agricultural land use. The North Temperature Lakes Microbial Observatory (NTL-MO) has collected a time series of water samples starting in 2000 to study microbial communities in Lake Mendota (N 43° 06', W 89° 24'). The NTL-MO is coupled with the Long-Term Ecological Research site at Lake Mendota,

which collects core limnological data such as water temperature, water chemistry, and plankton microscopy counts (all data available at <http://lter.limnology.wisc.edu>). NTL-MO collected depth-integrated samples from the epilimnion (approximately 0–12 m depth), primarily during the ice-free period (approximately March–December) and filtered at 0.2 μm prior to storage at -80°C until further processing.

At the Joint Genome Institute (JGI), 323 of these samples collected from 2000 to 2018 were sequenced on the Illumina HiSeq 2500 platform. The workflow used to process these reads is described in JGI's metagenomic assembly workflow [27], and includes standard deduplication, filtering, and error correction prior to metagenomic assembly with *metaSPAdes* versions 3.13.0 ($n = 78$) and 3.14.1 ($n = 245$) [28]. The resulting 323 assembled metagenomes across the 20-year time series were processed and annotated using the Integrated Microbial Genomes and Microbiomes (IMG/M) processing pipeline [29].

rRNA gene discovery workflow

To identify small subunit (SSU) rRNA genes in the metagenomic contigs, we used *cmsearch* from the *Infernal* package (1.1.4) [30] on each assembly with the eukaryotic (RF01960) and bacterial (RF00177) covariance models from the *Rfam* database [31, 32]. The longest match per contig from each model was retained, and the 16S and 18S rRNA gene sequences were extracted. Identified SSU rRNA gene sequences were filtered by sequence length and clustered at 97% sequence similarity using *VSEARCH* [33]. An absolute minimum cutoff of 500 bp was used for 16S rRNA gene sequences, while a cutoff of 750 bp was used for 18S rRNA gene sequences. We filtered the sequence matches using the National Center for Biotechnology Information (NCBI)'s *BLAST* alignment tool [34] against an alignment database of the rRNA gene sequences from the RefSeq database available from the NCBI's ftp server [34–36]. An *e*-value cutoff of $1e-4$ was used, and a minimum alignment length of 100 bp was applied. After obtaining the initial *blastn* hit to the RefSeq rRNA gene database, we used the Entrez Direct command-line utilities to query relevant identifiers from the NCBI database [37, 38]. In order to minimize any issues inherent to rRNA discovery within assembled genomes, namely the fact that rRNA genes may not assemble consistently, in particular for less abundant organisms, we retained only clusters of 16S or 18S rRNA gene sequences which appeared in ten or more datasets out of the total of 323 selected assembled metagenomes for the purposes of the time series trends. Only within-contig matches to the PR2/Silva database that also aligned at the domain level to the appropriate rRNA

model (e.g., domain bacteria with RF00177 as the underlying model) were included, and contigs with a qualifying match to both domains were entirely excluded. This condition occurred three or fewer times per assembly, and was likely the consequence of a poor quality sequence to begin with, as these alignments also tended not to meet the alignment length or percentage identity cutoffs of 200 bp and 80%, respectively.

Subsequently, we constructed an 18S rRNA gene and 16S rRNA gene phylogenetic tree separately for eukaryotes and bacteria, respectively using randomly selected sequences from the combined curated database clustered at 85% sequence similarity to approximate order-level groupings [39] as well as the clustered operational taxonomic units (OTUs) from the rRNA gene discovery workflow. All sequences were aligned to the corresponding eukaryotic (RF01960) and bacterial (RF00177) [31, 32] covariance model. A combined alignment file was generated using the *cmalign* tool from the *Infernal* package [30] (1.1.4). *IQ-TREE* (2.1.4) was used to construct a phylogenetic tree using the combined alignment file [40], which was then visualized using *GGTree* in R (v. 4.1) [41, 42]. In the case of disagreement in the taxonomic annotation between the cluster representative (the longest sequence in the 97% sequence identity OTU grouping) and the consensus annotation of the cluster members, the placement in the tree was used to determine the more likely annotation of the sequence.

Once we had found a set of contigs containing an SSU rRNA gene, we quantified the abundance of these contigs in the raw sequence files, in order to track the relative abundance of taxa in the dataset over time. We aligned filtered raw reads from the JGI assembly pipeline to the subset of the assembled contigs that had a 16S or 18S rRNA gene recovered. We did this such that all samples were quantified against the same assembled contig set; however, it does result in much higher estimated abundance values than when the respective assembly for each set of raw reads is used. We compared this approach to quantification using only clustered rRNA genes and found a $> 66.9\%$ R^2 ($p < 0.01$) correlation coefficient between the two, with much of the difference being driven by the fact that many samples did not contain the 16S or 18S rRNA gene when only clustered genes extracted by each assembly were used. Alignment was done using the *Bowtie 2* aligner (2.4.4) [43], then BAM, sorted BAM, and SAM files were generated using *SAMtools* [6] (1.3), and duplicated reads were marked and removed using *Picard* [44] (2.26.9). Finally, *SAMtools* [6] (1.3) was used to obtain coverage estimates and a number of mapped raw reads per contig. This final number of mapped reads per contig was used with the contig length to calculate a per-contig version of

transcripts per million (TPM) [45], which has recently been adopted for metagenomic data [46]. We call this version of TPM “CPM_{contig}”. We calculated “CPM_{contig}” using the following formula:

$$\text{CPM}_{\text{contig}} = \frac{\frac{\text{mapped reads} \times 10^3}{\text{contig length}}}{\sum_{i=0}^n \frac{\text{mapped reads}_i \times 10^3}{\text{contig length}_i}} \times 10^6$$

where n is the total number of contigs in the assembly. The total sum of CPM_{contig} in a given assembly will hence always be equal to 10^6 .

Because of inconsistencies in 16S and 18S rRNA gene copy number between different taxa (e.g., [47–49]), we further normalized abundance estimates to Z -scores during the period of interest. While not making disparate samples directly comparable [45, 50, 51], Z -scores standardize measurements already subject to variance stabilization via CPM_{contig} calculation and simplifies interpretation of changes in abundance during the sampling period. Instead of using raw abundance, the Z -score of each abundance estimate was calculated as a relative measure of the change in the size of the community of the respective taxonomic group over time. We also took into account unbalanced sampling over the total sampling period by using the weighted mean and weighted standard deviation within climatological seasons (the ice period, spring, early summer, late summer, and fall) as defined by Rohwer et al. (2022; [15]). We took unbalanced sampling into account by dividing by the total number of samples observed during each season and year to avoid unduly biasing overall abundance estimates towards years that were sampled more densely.

$$Z\text{-score}_{\text{sample}} = \frac{\text{RPM}_{\text{contig}} \text{ for sample of OTU} - \text{weighted mean}(\text{RPM}_{\text{for samples of OTU}})}{\text{weighted stdev}(\text{samples of OTU})}$$

The Z -score metric constrains the range of values for multiple species over the 20 years of measurements [52] and partially addresses the limitations of TPM (thus, RPM_{contig}) for between-sample comparisons [45, 51].

We conducted weighted two-sample t -tests (using the R package `weights` version 1.0.4 [53]) with a Benjamini-Hochberg correction with a 0.05 α value as separated between the pre-invasion of spiny water flea (2009 and before) [21, 22] and post-invasion (after 2010) segments of the time series and their corresponding abundance estimates in genes per million to categorize each eukaryotic clustered 18S rRNA gene.

Identifying co-abundant community members in time

In order to identify frequently co-occurring organisms on the basis of the abundance of their 18S and 16S rRNA gene

sequences over the full 20-year time series, we used hierarchical clustering. We clustered the sequences with the `hclust` function from the `stats` package within R and created a cluster dendrogram using the Ward method and the Canberra distance function [42]. We used the `pvclust` function for uncertainty estimation and bootstrapping with 1000 bootstrap iterations [54] and an initial adjusted approximately unbiased bootstrap probability value (AU) of 0.80 for preliminary filtering of cluster results. Visualization was performed using the `dendextend` package [55].

We also conducted a broader co-abundance network analysis between all taxa using all-by-all Pearson correlations as computed using the `rcorr` function within package `Hmisc` [56]. To control for multiple testing and identify significant Pearson correlations between taxa, we used the Benjamini-Hochberg procedure with a false discovery rate of 0.25 and an alpha value of 0.05 [57]. In our network visualizations, we use a correlation coefficient cutoff of 0.5 to identify correlations that are both significant and strong. Network visualization was performed using the `network` and `igraph` packages within R version 4.1.0 [42, 58, 59]. The Louvain clustering algorithm as implemented in `igraph` was used to identify coherence communities within the network [59, 60].

Eukaryotic metagenome-assembled genomes (MAGs) from Mendota metagenomes highlight key groups of protists

Eukaryotic metagenome-assembled genome identification, taxonomic annotation, and phylogeny

Each individual assembled metagenome from discrete sampling dates was scanned for eukaryotic sequences

with EukRep [61] (0.6.7). Identified eukaryotic sequences were combined into bins using METABAT2 [62] (2.12.1), and gene prediction was performed using Prodigal (2.6.3) [63]. Completeness and contamination of MAGs was estimated using EukCC [64] (0.2).

We used the BUSCO (Benchmarking Universal Single-Copy Orthologs, [65]) genes found in 60% of MAGs of completeness > 40% as well as those same genes from a selection of reference genomes and transcriptomes (73 reference genomes sequenced and curated by the JGI and 35 eukaryotic transcriptomes from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; [66]) to align the highly complete MAGs and construct a phylogenetic tree. The highly complete MAGs each had more than 90% of the genes found in 60% of all highly complete MAGs.

The above workflow only retained four of the identified MAGs, so we constructed a secondary phylogenetic tree using MAGs with completeness $\geq 10\%$ that had 30% or more of the BUSCO genes that were found in 15% or more of the MAGs. This tree was constructed using this subset of BUSCO genes found with differing consistency $\geq 30\%$ among the MAGs.

In addition to extracting MAGs from the time series metagenomes, we also did so for two enriched metagenomes, one enriched with *Daphnia pulicaria* and one enriched with *Bythotrephes longimanus*, the spiny water flea. Enrichment was performed by physically separating each zooplankton group from a water sample and sequencing zooplankton-dominated material. The raw read sequencing data for these two enriched metagenomes is available online at the NCBI's Sequence Read Archive (SRA) under accession number *to be added*. We used the same workflow for these MAGs, including completeness estimation with EukCC as described in the "Methods".

Using the extracted BUSCO proteins for each of the MAGs and a number of reference genomes curated by the JGI [67], we used trimal [68] to reduce spurious sequences, Clustal Omega [69] (1.2.4) to perform multiple sequence alignment, and FastTree (2.1.10) and RAxML (8.2.12) [70] for phylogenetic analysis.

For taxonomic annotation of the MAGs, we compared two approaches. First, we used the suggested NCBI database and utility provided by EukCC ([64]; 0.2) in concert with completeness estimates. Second, we used Eukulele [71] with a custom database comprised of the MMETSP [66, 72], MarRef [73], and a selection of reference genomes of eukaryotes from the JGI's Genome Portal [67] to annotate the proteins extracted from the contigs binned into each MAG.

Eukaryotic MAG abundance quantification

The abundance of the top 51 most complete MAGs was quantified using the output of the same workflow as for the rRNA gene sequences to which the filtered raw reads from the JGI metaSPAdes [28, 74] assemblies were mapped. In brief, raw reads were aligned to contigs from full metagenomic assembly using the Bowtie 2 (2.4.4) [43], then BAM, sorted BAM, and SAM files were generated using SAMtools [6] (1.3) with duplicates marked and removed using Picard [44] (2.26.9). Contigs from the MAGs were searched against the full metagenomic assembly to identify MAG representatives in each sample using mmseqs2 search with a sensitivity parameter of 0.70 [75]. Matches were filtered for those with $e\text{-value} \leq 10^{-50}$ and percentage identity of 0.90 or above. The sum of the abundance of those clustered contigs was taken for each sample to be the per-sample MAG

abundance. SAMtools (1.3) was used to obtain coverage estimates and a number of mapped raw reads per contig [6]. This procedure was used in place of mapping the MAGs directly to ensure that an appropriate value for $\text{CPM}_{\text{contig}}$ could be established using the full set of contigs extracted from each sample. This avoided bias related to the sample that each MAG was originally extracted from (since all MAGs were treated using the same procedure) and the total number of MAGs tested. Bias related to the total size of each MAG was unavoidable. Using a sequence search between binned contigs and all the contigs minimizes bias because quantification is done only based on assembled sequences, but bins with larger numbers of contigs will inevitably have more opportunities for a sequence match to be made that meets the selection criteria.

Phylogenetic analysis for individual clades

Phylogenetic analysis for individual clades was performed similarly to the general phylogenetic analysis for eukaryotes and bacteria. In brief, using clustering with VSEARCH [33], alignment with Infernal [30], and phylogenetic analysis using IQ-TREE [40] (2.1.4) and plotting in ggTree [41].

Putative trends of bacterial symbionts of microeukaryotes

Several bacterial taxa are known endo- or ectosymbionts of eukaryotic microorganisms. We identified representatives of these taxa from the bootstrapped hierarchical clustering procedure described in the "Identifying co-abundant community members in time" section. In order to explore the bacterial taxa which fell into clusters with defining temporal trends, we calculated the correlation between the Z-scores (as calculated over the full time series) of the putative endosymbiont clades and all eukaryotic OTUs, selecting OTUs with Pearson correlation coefficients of at least 0.5 as explained by the co-abundance network analysis, and which co-occurred in at least 10 distinct samples with the putative endosymbiont OTU. We also explored the number of samples in common between each bacterial taxon and co-clustering eukaryotic taxa, identifying the proportion of the time that these bacterial taxa were present independently of their eukaryotic counterparts.

Results

Microbial communities in Lake Mendota show seasonal partitioning

rRNA gene discovery from 323 assembled Lake Mendota metagenomes recovered 606 clusters of 18S rRNA gene sequences and 2,223 clusters of 16S rRNA gene sequences (at 97% sequence similarity). Sequences were then further filtered to include only those with

ten or more occurrences in the dataset and high quality blastn matches, resulting in 109 18S rRNA gene sequences (Fig. 1) and 608 16S rRNA gene sequences (Additional file 1: Figure S5). The average length of filtered 18S rRNA gene sequences was 1834 ± 363 bp, while the average length of 16S sequences was 1298 ± 299 bp. The eukaryotic OTUs spanned major branches

of eukaryotic life, including *Metazoa* (including invertebrate arthropods like *Daphnia* and *Bythotrephes longimanus*, the water flea and spiny water flea, respectively), *Apicomplexa* (parasites), *Cryptophyceae* (cryptophyte algae), and *Ochrophyta* (a group that includes diatoms-class *Bacillariophyta*-and other algae). Some of these sequences, including many cryptophytes and arthropods,

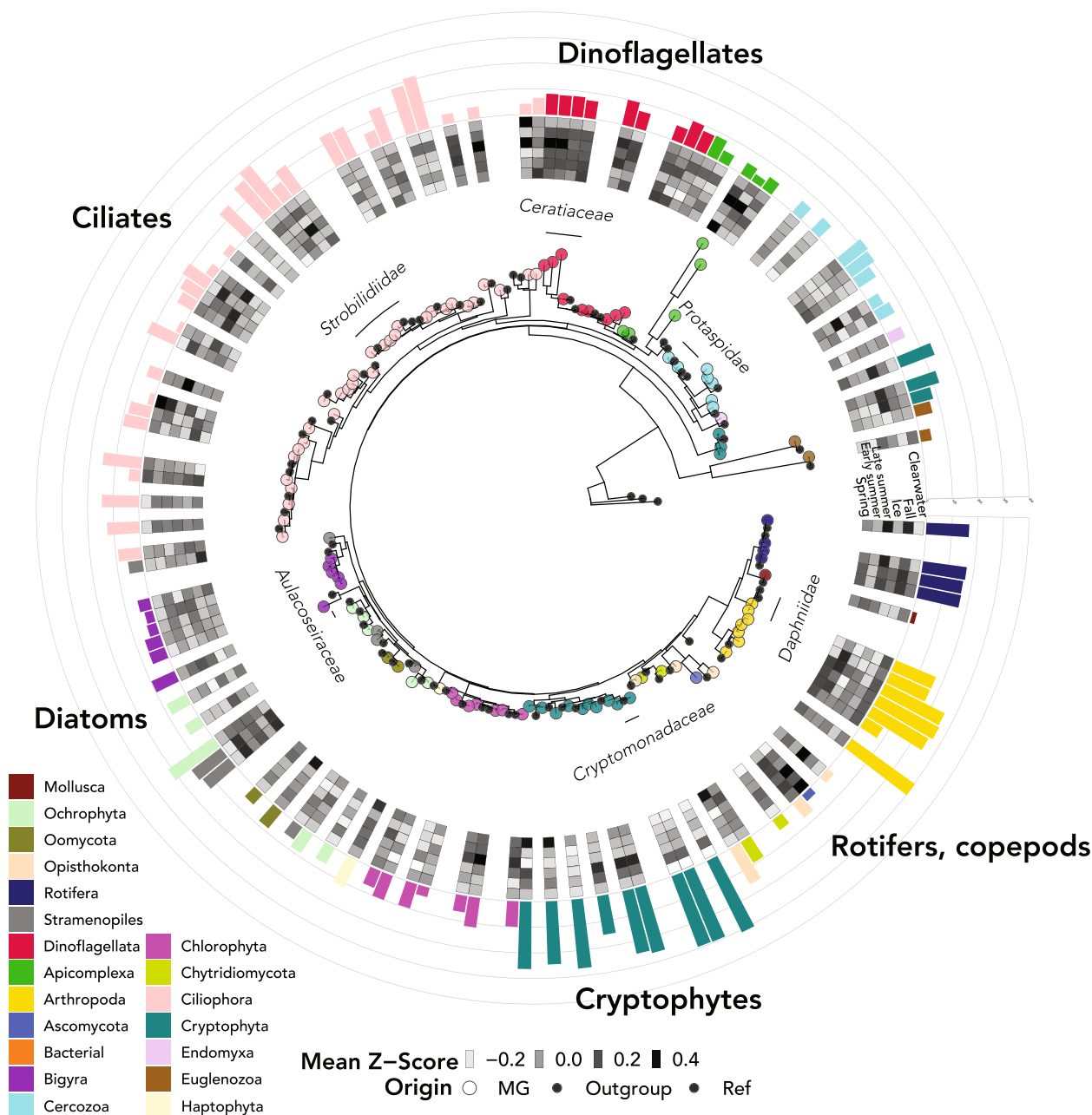


Fig. 1 Phylogenetic tree of 18S rRNA gene sequences extracted from Lake Mendota metagenomes over the 20-year time series. Colored, filled circular points on the tree indicate novel sequences extracted from the Mendota metagenomes, while smaller black circles denote a representative set of previously published sequences. The outer heatmap represents the mean Z-score of extracted sequence during each of the seasons of the year, and bars show the number of samples the 18S rRNA gene was found in

were independently assembled and recovered from the majority of the Lake Mendota metagenomes (Fig. 1). As expected based on the Plankton Ecology Group model [76, 77], some eukaryotic taxa also showed strong seasonal patterns, such as the tendency of cryptophytes to be most relatively abundant in winter and the tendency of diatoms (labeled Ochrophyta, class *Bacillariophyta*) such as family *Stephanodiscaceae* to be abundant in fall and spring ([78], Figs. 1 and 2). The majority of micro-eukaryotic taxa were most relatively abundant in spring, the ice-on period, and fall (Fig. 2A). However, arthropods, including nearly all zooplankton taxa, were most

abundant in the clearwater period (mid-spring to early-summer, when phytoplankton abundances decline; Fig. 2A), coinciding with the lower number of distinct SSU rRNA genes identified (alpha diversity; Fig. 2B and C). Both bacteria and eukaryotes had highest taxonomic diversity in the fall and winter according to SSU rRNA gene recovery (Fig. 2B and C). Bacteria had universally higher alpha diversity than eukaryotes, but eukaryotic diversity was highest relative to bacterial diversity in the month of April during the spring bloom (Fig. 2B). Copepods had higher relative abundance in winter, rotifers had higher relative abundance in summer and fall, and

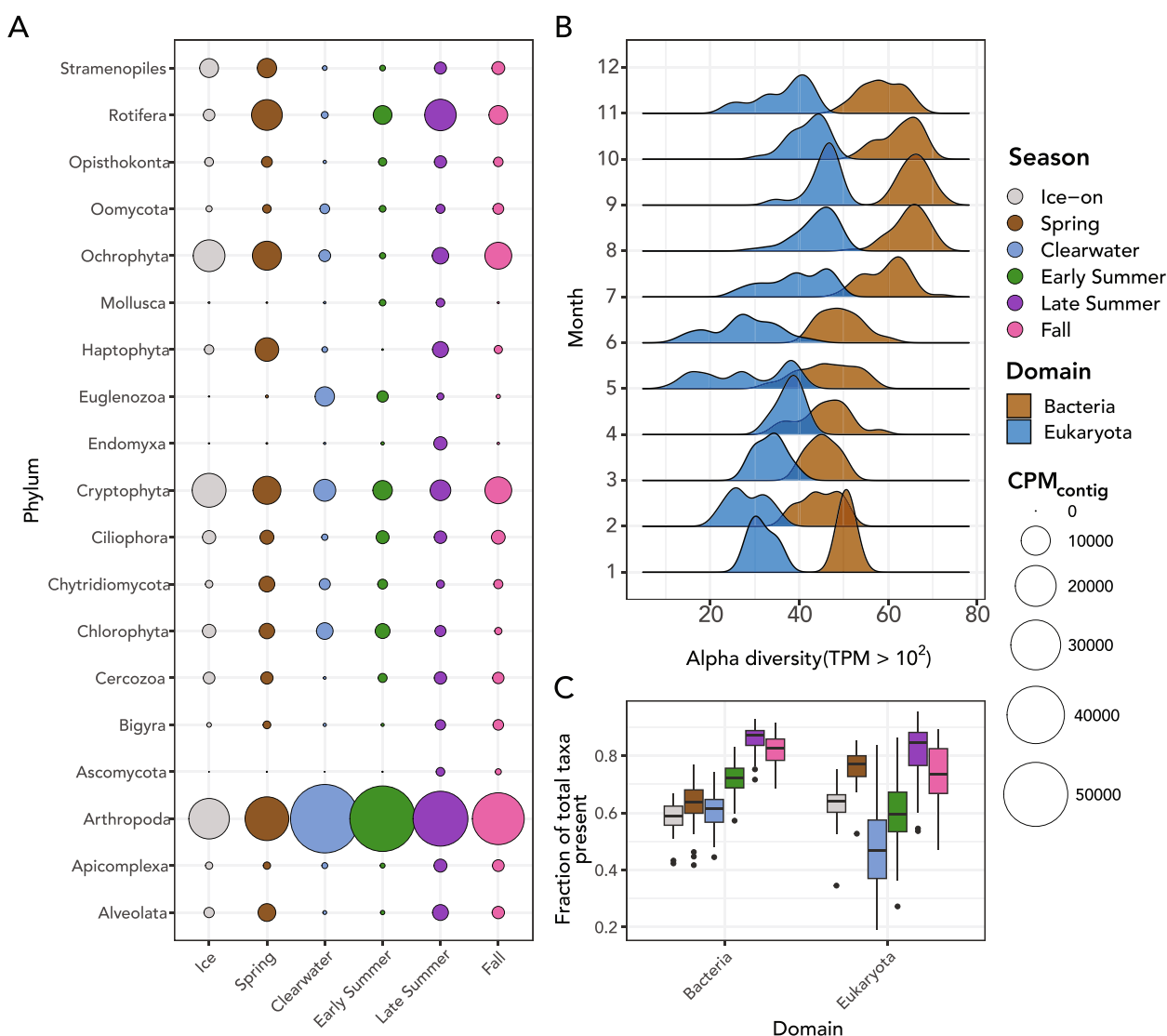


Fig. 2 Evidence of seasonal variability of SSU rRNA genes. **A** The abundance of major taxa of protists and metazoans in Lake Mendota by season. Circles are colored by the season of the year and sized according to total abundance via CPM_{contig} for that month. **B** Alpha diversity as assessed by extracting OTUs with TPM of 100 or greater. Distributions of alpha diversity are shown for both eukaryotes and bacteria for each month of the year in the time series. **C** Proportion of total observed taxa present in each season of the dataset for (left) bacteria and (right) eukaryotes

cladocerans had higher relative abundance in spring and summer (Fig. 5).

Changes in community composition: an increase in *Alveolata* and a decrease in cryptophytes

In the second half of the time series (2010–2019), the relative abundances of several taxa significantly increased following the invasion of *Bythotrephes longimanus* (spiny water flea) (Fig. 3) [21]. These taxa included 28 alveolates, 7 dinoflagellates of class *Dinophyceae*, 14 ciliates of class *Ciliophora*, and 3 candidate Apicomplexan parasites (Figs. 3 and 4). Ciliate classes *Nassophorea*, *Litostomatea*, and *Heterotrichea* and dinoflagellate class *Dinophyceae* had statistically significant increases in 2010–2019 relative to 2000–2009 (Fig. 4B, D). *Nassophorea*, *Litostomatea*, and *Dinophyceae* also showed higher relative alpha diversity during this period (Fig. 4E), meaning that more of the total OTUs belonging to each class were recovered with sufficient abundance in the second half of the time series.

Within metazoans, the abundance of an OTU assigned to the Zebra mussel (genus *Dreissena*; $p = 8.2e^{-4}$) significantly increased after being low or zero abundance for the majority of the time series (which correlated with its documented irruption in Lake Mendota in 2015; Fig. 5; [24]). Other OTUs with significant increases included four copepods of order Cyclopoida (2 with $p < 2e^{-5}$, 1 with $p < 3e^{-3}$, 1 with $p \approx 0.01$), a rotifer of order *Flosculariaceae* ($p < 2e^{-5}$) and two rotifers of order *Ploima* ($p < 2e^{-5}$) (Figs. 3 and 5).

The remainder of the OTUs that showed significant increases in abundances were two Stramenopiles (order Hyphochytriaceae), two Chlorophytes (order Chlamydomonadales), and a Haptophyte (order Prymnesiales).

In contrast, after 2010, significant decreases in abundance were observed for the eukaryotic phyla Cryptophyta, Ciliophora, and Cercozoa. This included two unclassified Cercozoans ($p < 0.005$), two ciliates of class *Spirotrichea*, and five cryptophytes. This was 41.7% of all cryptophyte OTUs recovered (12 total) and 100% of the cryptophyte OTUs annotated to belong to taxonomic order *Cryptomonadales* ($n=2$, $p < 0.001$).

Putative relationships between eukaryotic and bacterial taxa

Putative relationships between eukaryotic and bacterial taxa were predicted using both hierarchical clustering and network analysis of the rRNA gene sequences (Fig. 6). Initially focusing on bacteria that could be potential endosymbionts of protists, we identified six bacterial OTUs whose annotated taxonomies were affiliated with known endosymbionts of microbial

eukaryotes (Additional file 1: Figure S10; complete list of correlations and hierarchical clustering modules available in Additional file 1: Tables S1 and S2). The abundances of two endosymbiotic genera, a *Caedimonas* OTU from order *Holosporales* and a *Rickettsia* OTU from order *Rickettsiales* [79], significantly increased ($p < 0.01$) after the spiny water flea invasion and correlated with eukaryotic OTUs, including multiple ciliate (phylum Ciliophora) OTUs (*Rickettsiales* $r = 0.57$; *Holosporales* $r = 0.60$) (Fig. 6). Two ciliates formed a module with a *Sphingobacteriaceae* bacterium (*Bacteroidetes*; AU p -value < 0.01). One diatom (phylum *Ochrophyta*; class *Bacillariophyta*) OTU was significantly correlated with an OTU belonging to the family *Anaplasmataceae* (order *Rickettsiales*; $r = 0.72$). We also identified bacterial taxa that were associated with eukaryotes likely due to local environmental conditions. One eukaryotic OTU annotated as Chytridiomycota was significantly correlated with a *Planctomycetes* (*Isosphaerales*) OTU ($r = 0.66$). A diatom OTU was significantly correlated with an OTU of cyanobacterium *Chamaesiphonaceae* (*Synechococcales*; $r = 0.87$) and a cyanobacterial OTU of *Microcystaceae* (*Chroococcales*; $r = 0.55$). A dinoflagellate (*Peridinales*) was correlated with a bacterium of order *Saprospirales* ($r = 0.54$) and another bacterium of order *Sphingomonadales* ($r = 0.67$). One cluster of ciliate *Nassulida* contained bacteria of orders *Desulfuromonadales*, *Synechococcales*, and *Physcisphaerales* (AU p -value < 0.01). Other relationships may have been associated with direct trophic relationships, for example the correlation between a dinoflagellate (*Peridinales*) with a cercozoan (*Cryomonadida*; $p < 0.001$, $r = 0.58$) and a ciliate (*Oligotrichia*; $p < 0.001$, $r = 0.60$). Furthermore, we identified bacterial OTUs known to be potential constituents of protistan microbiota or members of biofilms [80, 81]. These included an OTU of *Borreliaceae* (order *Spirochaetales*) that significantly correlated to a diatom ($r = 0.75$), and one of cyanobacterium *Chamaesiphonaceae* (*Synechococcales*), also correlated to a diatom with the strongest correlation coefficient value recorded in the dataset ($r = 0.87$; Fig. 12; Additional file 1: Figure S6). Chlorophytes (green algae) were correlated with two *Comamonadaceae* (*Burkholderiales*) OTUs ($r = 0.50$ and $r = 0.51$), a *Rhodobacteraceae* (*Rhodobacterales*) OTU ($r = 0.53$), a *Chthoniobacteraceae* (*Chthoniobacterales*) OTU ($r = 0.52$, 0.53), a *Flavobacteriaceae* (*Flavobacteriales*) OTU ($r = 0.51$, 0.52), and a *Sphingosinellaceae* (*Sphingomonadales*) OTU ($r = 0.54$) (Fig. 12). Some of these bacterial OTUs were also contained in a large hierarchical clustering module which contained several eukaryotes: two chlorophytes, a chrysophyte, a dinoflagellate,

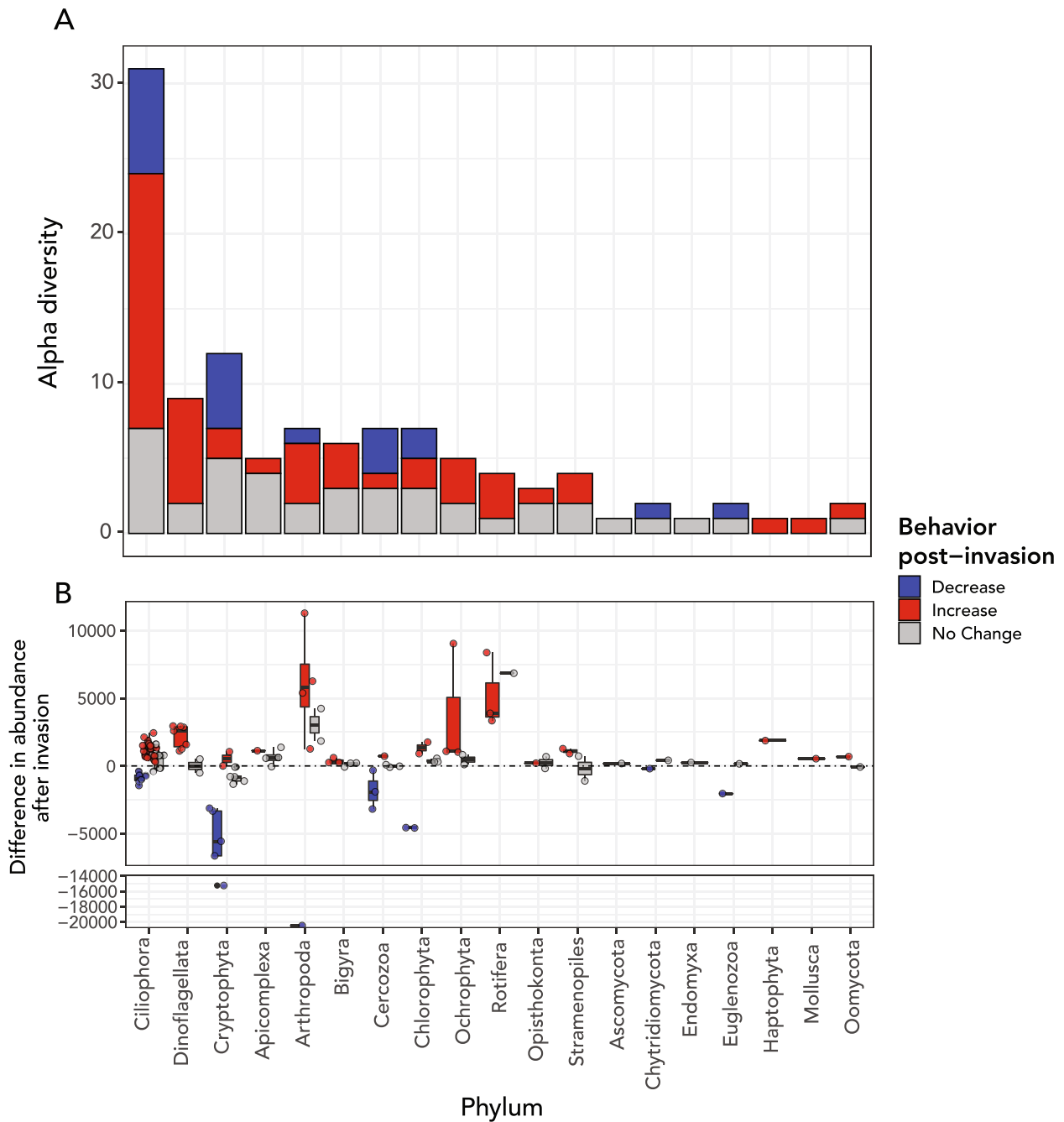


Fig. 3 Shift in diversity and abundance of microbial eukaryotes and other relevant eukaryotic taxa after the spiny water flea invasion. **A** Phylum breakdown of OTUs that did or did not show statistically significant increases in CPM_{contig} before and after 2010. The majority of statistically significant increases were from OTUs annotated as clade Alveolata (including phyla Ciliophora, Apicomplexa, and Dinoflagellata). **B** Difference in mean CPM_{contig} after the invasion for OTUs and mean CPM_{contig} before the invasion of the spiny water flea, grouped and colored by the statistical significance of the change in abundance

an apicomplexan, three ciliates, and two cercozoans (Fig. 12; Additional file 1: Table S2). Nine bacteria were contained in the same module as this cluster of

eukaryotes, including two alphaproteobacteria (*Sphingomonadales* and *Rhodospirillales*) and two planctomycetes (Additional file 1: Table S2).

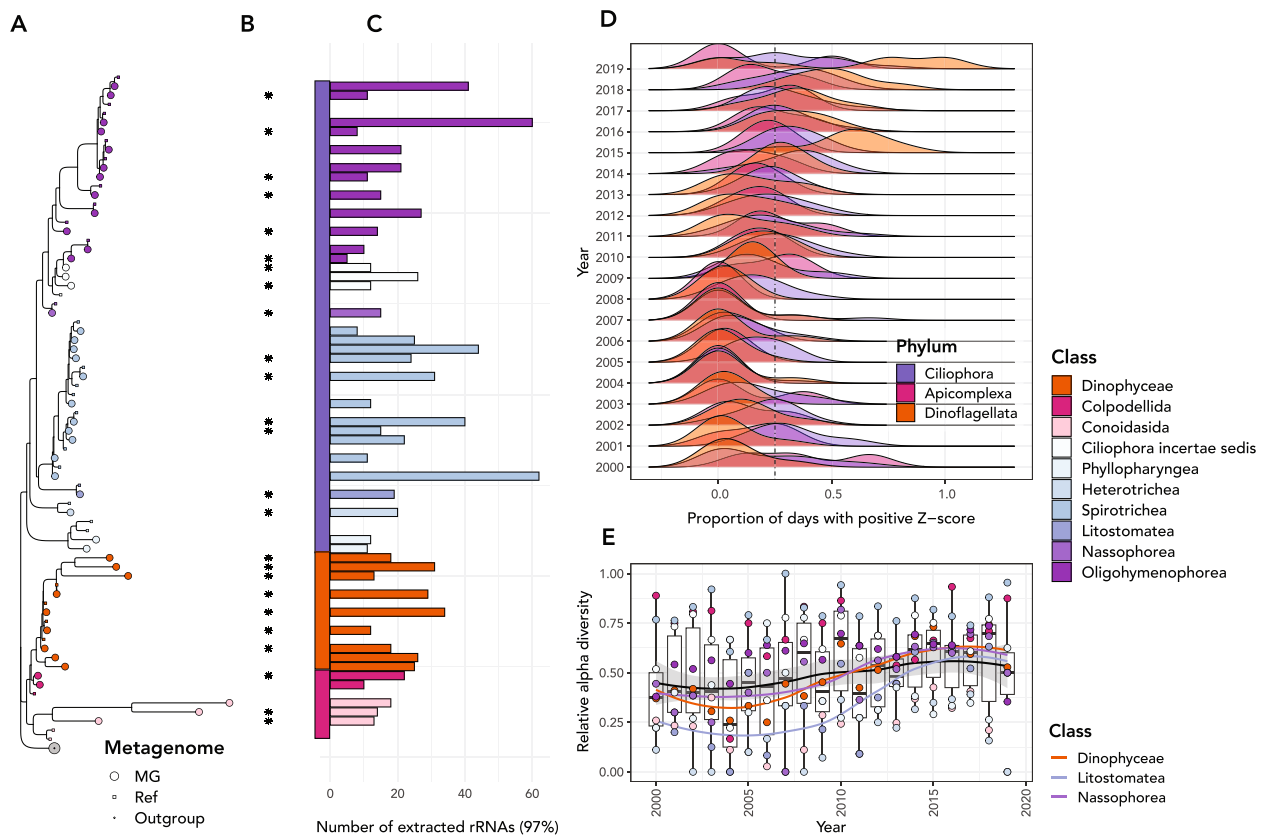


Fig. 4 Abundances of Alveolates in the Lake Mendota time series. **A** Phylogenetic tree showing all alveolate (clade Alveolata) sequences (square tip points) alongside their NCBI RefSeq closest blast matches (small circles), highlighted by their taxonomic class. The green circle at the bottom of the tree is a clade of *Ochrophyta* used as an outgroup. **B** Extracted 18S rRNA gene sequences are labeled with a “*” if they changed significantly in abundance according to the procedure described in the text. **C** The number of 18S rRNA genes from all samples that fell into the cluster (97% sequence identity) in the phylogenetic tree. Ciliates of class *Spirotrichea* tended to be most highly abundant in the dataset. **D** Distribution of days that had positive Z-scores in abundance for each Alveolate phylum for each year of the dataset. In the later years of the dataset, dinoflagellates in particular had a higher proportion of total sampling days that had elevated abundance relative to the mean. **E** Alpha diversity (number of OTUs) present relative to total extracted for each alveolate class in each year of the dataset. Classes *Nassophorea*, *Litostomatea*, and *Heterotrichea* were universally more commonly present in the latter half of the dataset (2010–2019) and in particular in 2012–2018. Lines drawn on top of the boxplots and point show the overall trend (black) and colored trends in relative alpha diversity for the three taxonomic classes for which alpha diversity trends were significant

Eukaryotic MAGs match several abundant taxa in 18S rRNA gene time series

A total of 4511 eukaryotic MAGs were recovered across all samples, of which the majority were small and incomplete. Only 154 were sufficiently large for EukCC [64] to calculate a completeness estimate. Fourteen MAGs had greater than 20% completeness, 5 had greater than > 50% completeness, and 2 had above > 75% completeness. Leveraging EukCC completeness estimates, the 4 MAGs that were at least 40% complete had 28 BUSCO genes from which a 90% subset could be found in at least 60% of the 40% complete MAGs. Ten MAGs met the second threshold criterion (see the “Methods”) and were included in the phylogenetic tree alignment (Fig. 7). In practice, the 10 MAGs which met this second shared BUSCO

criterion had completeness of 17% or greater. A total of 86 BUSCO genes were used for the alignment.

We found a chlorophyte (green algal) MAG which had both an associated 18S rRNA gene and was annotated as a chlorophyte and aligned within the chlorophyte clade in the MAG phylogenetic tree (Fig. 7A). While its 18S rRNA was part of a cluster that only appeared 6 times in the dataset and hence was excluded from the time series analysis, it had 96.5% identity and a 1785-nucleotide alignment length to a chlorophyte of genus *Monomastix* via NCBI RefSeq sequence (accession number AB491653). The vast majority of eukaryotic MAGs did not have a co-binned 18S rRNA gene sequence from the SSU rRNA discovery procedure. The most complete MAGs extracted belonged to phylum Haptophyta

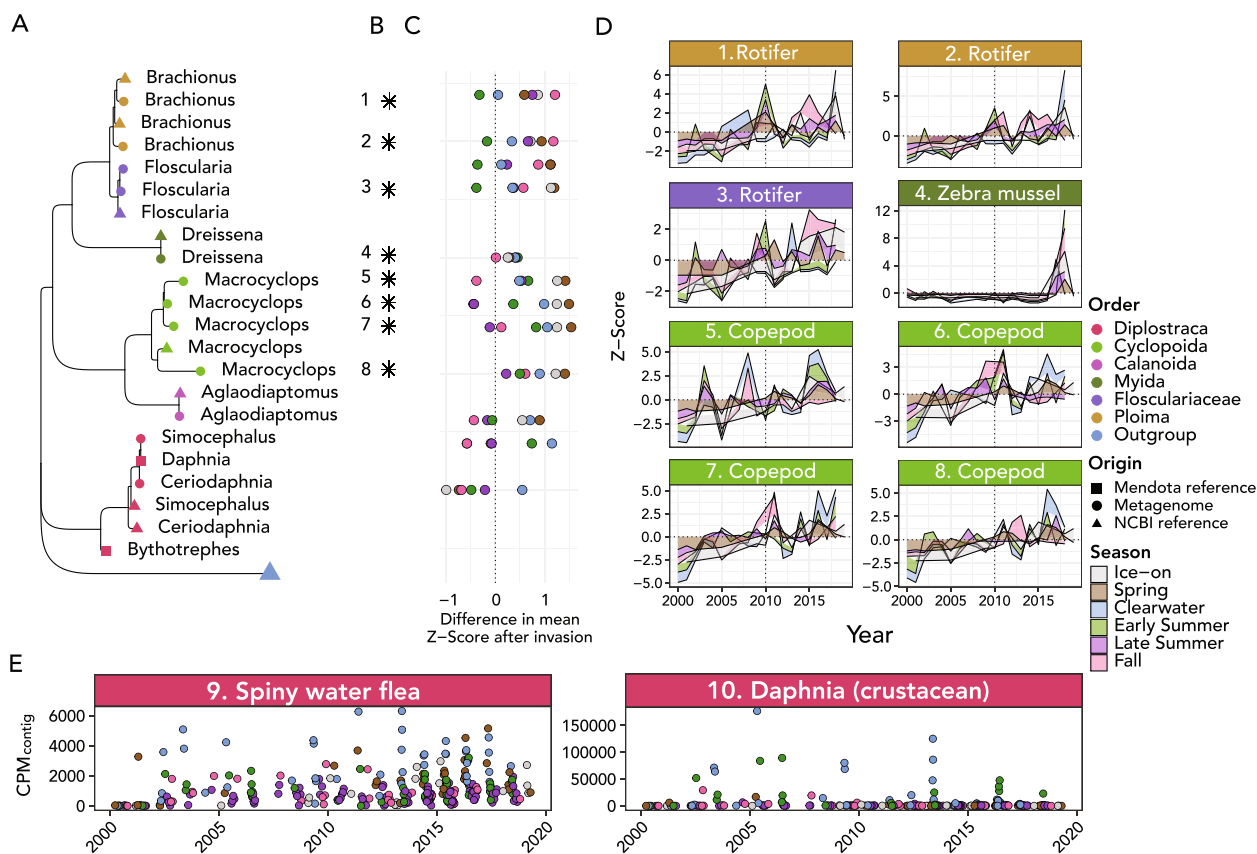


Fig. 5 Changing community profile of metazoans in Lake Mendota. **A** Phylogenetic tree showing genus-level taxonomic assignment of each metazoan 97% OTU. Shapes indicate sample origin and are colored based on order-level taxonomic assignments of the respective genus. Squares correspond to references extracted from companion metagenomes for *Daphnia* (water flea) and *Bythotrephes* (spiny water flea); references for each generated using enriched metagenomes from Lake Mendota). Tree is rooted at the phylum *Ochrophyta*. **B** Stars (“*”) denote significant abundance increases of rRNA clustered OTUs after 2010. **C** Difference in average Z-score of abundance of metagenomic rRNA before and after the spiny water flea invasion by year and season. The x-axis corresponds to Z-score, and points are colored by season. **D** Time series seasonal Z-scores for taxa numbered in panel (B). **E** Mapped abundances of each of the 18S rRNA genes extracted from the companion metagenomes for *Bythotrephes* and *Daphnia*

(MAG IDs ME-TYMEFLIES-MAG 32,36,37,38,42,43,45, and 46; Fig. 7) and appeared to likely belong to genus *Chrysochromulina*, a documented haptophyte genus in lake ecosystems [82]. In the most complete haptophyte MAG, 29.5% of contigs in the MAG showed consensus for *Chrysochromulinaceae*, as opposed to 52.3% for order *Prymnesiales*, 87.5% for class *Prymnesiophyceae*, and 90.2% for phylum Haptophyta.

MAG abundance quantification shows similar patterns to 18S rRNA gene survey

The abundances of the top 20 MAGs of highest completeness per EukCC increased significantly after the spiny water flea invasion ($p < 0.001$), showing the same pattern as that observed for the 18S rRNA sequence data. The highest overall abundance was associated with MAGs of highest completeness, though all of the MAGs revealed the same trend of increasing abundance.

Putative chlorophyte MAGs included ME-TYMEFLIES-MAG 33, 47 (70.0% Eukulele consensus; 31.4% for class *Mamiellophyceae*; 16.2% for order *Mamiellales*), 48 (86.5% consensus for phylum *Chlorophyta*; 39.4% consensus for class *Trebouxiophyceae*; 26.6% for order *Chlorellales*), and 50 (71.5% Eukulele consensus for phylum *Chlorophyta*; 32.6% for class *Mamiellophyceae*). The only putative cryptophyte MAG was ME-TYMEFLIES-MAG 34 (Eukulele consensus 79.5% for phylum *Cryptophyta*; 78.5% for class *Cryptophyceae*; 55.7% for order *Cryptomonadales*). Putative ciliate MAGs included ME-TYMEFLIES-MAG 40 (35.5% consensus) and 44 (37.7% consensus).

Putative ochrophyte MAGs included ME-TYMEFLIES-MAG 35 (23.3% consensus), 39 (14.3% consensus), 41 (20.3% consensus), and 49 (90.6% consensus for Ochrophyta; 88.3% consensus for class *Bacillariophyta*; 21.6% consensus for order *Thalassionemales*), though it should be noted that all of these MAGs apart from

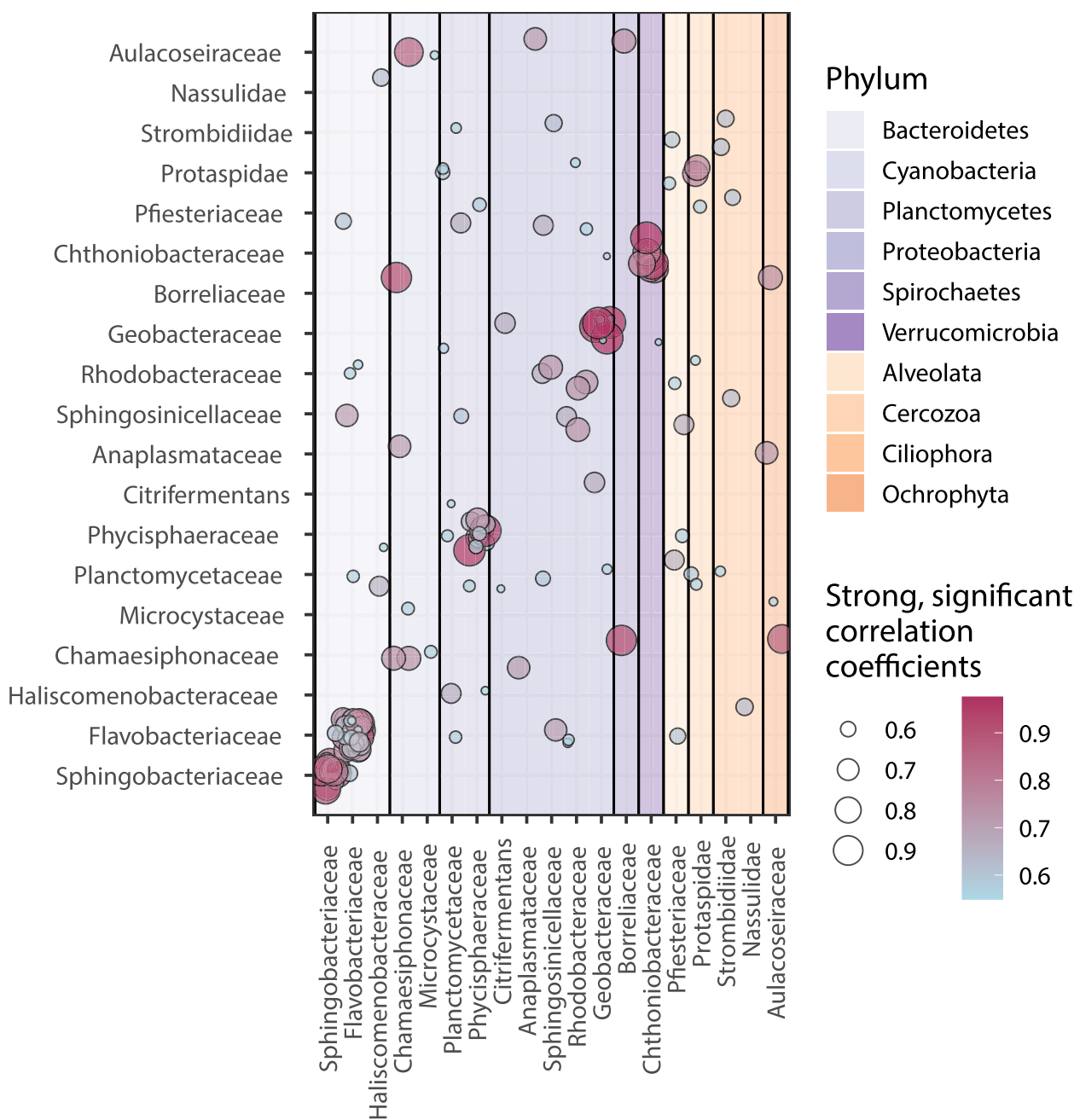


Fig. 6 Network analysis over 20-year time series of metagenomes suggests stable connections between eukaryotes and bacteria. Circles correspond to significant, strong correlations between taxa listed on the x- and y-axes and are both colored and sized according to the correlation coefficient. Only correlations with a significant Benjamini-Hochberg-corrected *p*-value and a correlation coefficient of greater than or equal to 0.50 are included visually. The color corresponds to whether the x-axis partner is bacterial or eukaryotic, and shading indicates specific phylum. Self-correlations are excluded when the small subunit rRNA gene sequences were the same

MAG 49 had conflicting annotations between dinoflagellates and ochrophytes, and only MAG 49 was complete enough to be included in the phylogenetic tree constructed using BUSCO genes (Fig. 7). In MAG 49, the strong signal of an exceptionally large diatom bloom

near the end of the time series was reproduced by the contigs that clustered with the MAG, although the annotation provided by the 18S rRNA gene sequence was order *Aulacoseirales*, which is another taxonomic order of pennate diatom. MAG 39 also showed a

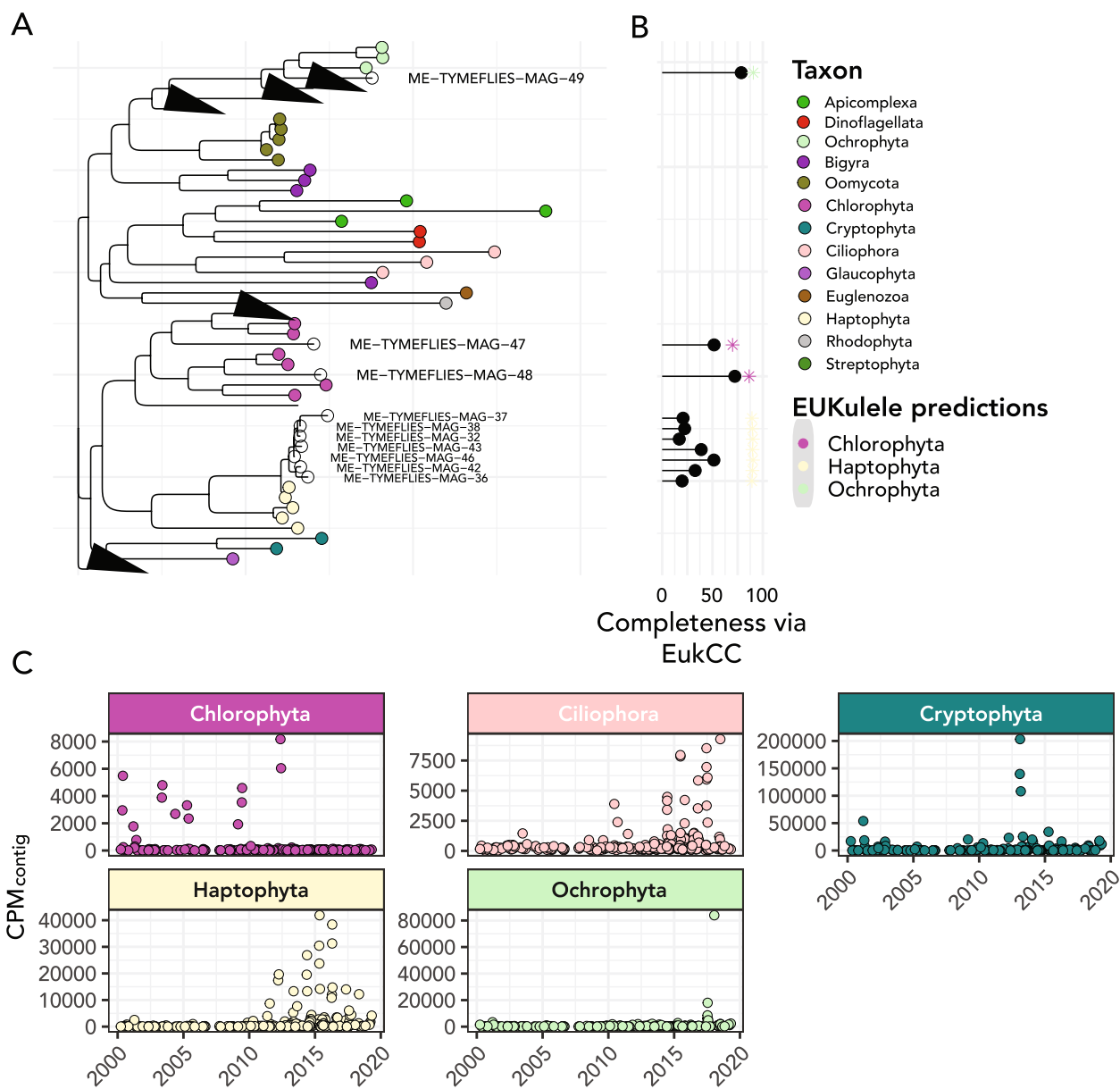


Fig. 7 Metagenome-assembled genomes recovered from Lake Mendota span a wide range of eukaryotic taxa. **A** Phylogenetic tree of extracted MAGs using concatenated alignment of BUSCO genes alongside reference genomes from the corresponding phyla. Only MAGs which contained a sufficient number of BUSCO genes (see the “Methods”) were included. Unfilled and labeled circles correspond to MAGs extracted from the Lake Mendota time series. **B** Lollipop plot showing completeness of each MAG via EukCC (connected point) and the percentage consensus of the phylum-level taxonomic annotation provided by Eukulele (colored asterisk (*)). **C** Abundance of the most abundant metagenome-assembled genomes extracted from within each taxonomic group across the metagenomic time series, expressed in CPM_{contig}; each point is an estimated abundance for one MAG at one timepoint

substantial, though less than 10% of the CPM maximum of MAG 49, increase in abundance during the same time period.

Furthermore, we used two auxiliary datasets, one of which was enriched in cladoceran *Daphnia pulex* and the other in spiny water flea *Bythotrephes longimanus*, as

a comparison. We extracted 18S rRNA gene sequences from these datasets and binned MAGs. EukCC estimated the completeness of the *Daphnia pulex* MAG at 32.4% with 8.8% contamination, and the *Bythotrephes longimanus* MAG at 38.2% completeness and 8.8% estimated contamination (Fig. 5).

Discussion

Our results show that 18S rRNA genes can be recovered from metagenomic time series data and used to interpret temporal patterns. These patterns combined with even coarse taxonomic information provided ecological insights into protists, a diverse and abundant group of organisms frequently ignored by freshwater microbial ecologists. We identified eukaryotic lineages that significantly increased over the course of the time series and conducted a network analysis to predict potential associations between protists and bacteria. Major changes in abundance were observed in Alveolates (including a selection of dinoflagellates, ciliates, and Apicomplexans; no phytoplankton of clade *Alveolata* are cataloged in the LTER dataset for Lake Mendota), some diatoms, as well as some classes of metazoans (*Opisthokonta*), including genus *Dreissena* (Zebra mussel) and several copepods, and decreases in several OTUs annotated to be cryptophytes. These changes occurred following the invasion of the spiny water flea, *Bythotrephes longimanus*, near the end of 2009, which significantly changed the ecology of Lake Mendota and likely contributed to these shifts in protistan community structure [83]. Increases in eukaryotic phytoplankton may be attributable to a decrease in grazing by *Daphnia* predators due to *Bythotrephes* invasion, as previously established in the literature [21, 22, 83, 84].

Recent decrease in abundance of Cryptophytes in Lake Mendota

One particular example of marked change in Lake Mendota eukaryotic communities is the decrease of abundance and diversity of cryptophytes, which have been shown to quickly respond to changes in other similar temperate lakes [85]. In the NTL-LTER phytoplankton time series, only two genera of cryptophytes are documented [86], whereas our results suggest the presence of at least seven genera and multiple putative cryptophyte MAGs. The lack of attention to cryptophyte diversity in Lake Mendota in previous studies could be because of methodological limitations in the microscopy used for the NTL-LTER measurements, as cryptophytes are smaller than many eukaryotic algae, and their size may also be variable in situ [87].

Cryptophyte abundance may change in response to other ecological shifts in Lake Mendota. Cryptophytes are known to be adapted to low-light conditions [88], so higher light intensity or smaller blooms of other organisms may reduce the available cryptophyte niche. The presence of zebra mussels (*Dreissena polymorpha*), which are predicted to increase in abundance in response to increases in spiny water fleas [24], and which irrupted into Lake Mendota in 2015 [89], may also result in

reduced cryptophyte abundances, as was observed in Lake Oneida in the late 1980s [85]. To our knowledge, no other study has investigated recent changes in cryptophyte abundances in relation to spiny water flea and zebra mussel invasions. Taken together, our findings indicate that there has been a decline in cryptophyte diversity and abundance in Lake Mendota in recent (2010–2019) years, which may also have increased the available niche space for other eukaryotic organisms.

Metagenome-assembled genomes reveal only the dominant fraction of Lake Mendota's protistan diversity

Our results demonstrate that it is possible to link mostly complete eukaryotic MAGs to an 18S rRNA gene sequence and to use this to clarify the overall representation of assembled eukaryotic contigs in a mixed environmental metagenomic assembly.

The number of MAGs recovered from Lake Mendota metagenomes is on par with other recent studies from similar sample types [90], though the complete MAGs constituted a small fraction of the overall eukaryotic microbial diversity that we identified via 18S rRNA gene discovery. Our observation that the vast majority of MAGs did not contain discovered 18S rRNA gene sequences is in line with previous observations that rRNA genes do not tend to co-bin with MAGs [91]. In some cases, 18S rRNA genes may have been filtered from the dataset, e.g., a single rRNA gene that appeared to be *Bythotrephes longimanus*, which was subsequently excluded due to not being present in multiple samples.

We have shown that MAGs and 18S rRNA gene sequences provide complementary insights into protistan diversity within a single metagenomic time series and have the forthcoming potential to efficiently link multiple sources of -omic data. In particular, MAGs from samples enriched with a single organism will be a powerful tool for maximizing inference from time-series metagenomic surveys, because once MAGs are identified from the enriched samples, the putative genomes can be quantified across the entire time series. Indeed, for the two enriched metagenomes collected from *Daphnia pulex* and *Bythotrephes longimanus*, although the extracted MAGs had relatively low completeness, this approach was far superior to the community-level metagenomes in allowing a MAG to be binned in the first place. Because count data suggest high metazoan abundances [92], we suspect detecting their genome sequences may be limited technically by assembly or by only capturing environmental DNA rather than individuals during sampling (for example, gelatinous blooms of *B. longimanus* clogging filters).

Abundance quantification both via the 18S rRNA gene extracted from the time series metagenomes and the

extracted MAG from the enriched metagenome strongly suggest that *Bythotrephes longimanus* may have been present in the lake in low numbers before 2009 (Fig. 5), which has been previously discussed in the literature [21, 93]. Past work in lake sediment cores suggests that the spiny water flea was present in low abundances in Lake Mendota since at least 1999 [22, 23] as a “sleeper population” until conditions were right for the organism to become dominant. Our analysis of this long-term time series provides additional evidence that suggest that the spiny water flea could be detected genetically in the lake prior to 2009 (Fig. 5). Without collecting metagenomic time series in a similar way to the established method of long-term sediment coring, this powerful retrospective insight into the spiny water flea invasion and its effects would not have been possible. Enriched metagenomes and MAG binning for other organisms could yield more accurate estimation of their composition in the lake and enable insights about the functional potential of taxa without sequenced genomes.

Predicted eukaryote-prokaryote associations are common in Lake Mendota

Our analysis predicted a wide range of association between diverse bacteria and eukaryotes, in particular chlorophytes, diatoms, chytrid fungi, and ciliates.

Chlorophyte (green algal) OTUs were associated with OTUs of several bacteria identified by other studies to be chlorophyte microbiome constituents, including families *Comamonadaceae*, *Rhodobacteraceae*, and *Chthoniobacteraceae* [94, 95] and *Flavobacteriaceae* and order *Sphingomonadales* (Alphaproteobacteria) [96]. Chlorophytes also correlated with sulfur bacteria (family *Chromatiaceae*) which are known to be abundant in lakes [97], and cyanobacteria of order *Oscillatoriales* (family *Microcoleaceae*), and two families of potentially pathogenic Gammaproteobacteria, *Moraxellaceae* and *Aeromonadaceae* [98, 99]. Chlorophytes are known to thrive under high nutrient concentrations early in the season, in particular high nitrate levels and a deep mixed layer [100–102], and thus may increase in abundance, alongside cyanobacteria, with high levels of nutrient pollution [101, 102].

A diatom (phylum *Ochrophyta*, class *Bacillariophyta*) of order *Aulacoseirales* was significantly correlated with several other bacterial and eukaryotic OTUs: one ciliate, a bacterium of order *Spirochaetales* which is known to be present among some diatom microbiota [95], cyanobacteria of orders *Chroococcales* and *Synechococcales*, one OTU of order *Verrucomicrobiales*, and an Alphaproteobacterium of order *Rickettsiales*. Several associations, in particular with cyanobacteria of orders *Chroococcales* and *Synechococcales*, may be related to

their shade-tolerance during dense blooms of diatoms and other eukaryotes [103, 104]. A second diatom was correlated with *Synechococcales* (*Chamaesiphonaceae*); not only was this a distinct OTU as indicated by our clustering approach, it also was one of the strongest correlations we recorded in the dataset. This may be related to their co-habitation of biofilms, of which *Chamaesiphonaceae* is a major component [80], and diatom frustules were found embedded in biofilms present in other lakes [105, 106].

The analysis also suggested an association of Chytridiomycota and planctomycetes of the order *Isosphaerales*; both taxa degrade organic matter in lake ecosystems [107–109] and thus may clean up organic matter in the lake. Chytridiomycota and *Isosphaerales* were most abundant in April, which might be associated with the influx of organic matter observed with spring runoff during the ice-off period in lakes [110]. Due to uneven sampling during the winter period at the beginning of the time series, it remains unknown whether earlier ice-off due to increasing temperatures is changing the prevalence of Chytridiomycota and *Planctomycetes*, though there is some evidence of their increased abundance (Additional file 1: Figure S4). Interestingly, some other *Planctomycetes* showed correlations only with individual eukaryotic OTUs or distinct taxonomic groups, which may indicate that they depend on the presence of eukaryotic taxa as specialized hosts [111]. *Planctomycetes* can use the nutrients generated from algal or cyanobacterial blooms as a food source [111], which might explain its strong correlation with a dinoflagellate, potentially supporting the hypothesis that specific bloom types offer substrates for these bacteria [111, 112].

A broad diversity of ciliates formed a high number of clusters with both bacteria and other eukaryotes. These results are particularly informative because many species of ciliates, apicomplexans, and cryptophytes can be quite challenging to delineate based on morphological features, despite their abundance in the lake system [113, 114]. Among the putative interactions, there was one coherent module that consisted of two ciliates (families *Vorticellidae* and *Acinetidae*) and a *Bacteroidetes* OTU of family *Sphingobacteriaceae*. This strong association may be related to an endosymbiotic relationship between the bacterium and one or both of these ciliates, as a *Sphingobacteria* has previously been shown to have an endosymbiotic relationship with *Ichthyophthirius multifiliis* [115], an organism in the same taxonomic class as one of these ciliates (*Oligohymenophorea*). Some bacteria can associate with ciliates if they are resistant to digestion by the ciliate [116]. We identified at least one of these bacteria, family *Phycisphaeraceae*, that clustered with a ciliate OTU, highlighting yet another route by

which eukaryotes may be associated with bacteria in this system.

Long-term time series facilitated the detection of invasive species responses and putative associations, which provides a foundation for future testable hypotheses

The prediction of protistan relationships in long-term time series metagenomic data from Lake Mendota supports and extends our understanding of the roles played by the major taxonomic groups in lake ecosystems and their effects on food web structure [117]. We identified putative interactions between protists and bacteria on the basis of their consistently correlated abundances, which opens the door for exploration of novel interactions between these taxa, including lifestyles like endo- or ectosymbioses [118] or farming (whereby organisms like amoebas may tote and proliferate their bacterial food) [119]. Our results provide evidence for the presence and abundance of such interactions and enable new perspectives on their ecology. Furthermore, our metagenomic identification of a possible “sleeper population” of spiny water flea aligns with observations from previous research [21, 23], provides further genetic contextualization for this important invasion event, and highlights the importance of long-term ecosystem surveillance using metagenomics.

Conclusion

Protists are critical to lake ecology and biogeochemistry. Relatively little is known about the protistan community dynamics at Lake Mendota, despite it being among the world’s best-studied freshwater systems. Here, we leveraged a 20-year time series of metagenomes from Lake Mendota to explore patterns in the abundance of phytoplankton and heterotrophic protists in the lake and identified co-occurring bacteria from the same time series. We found dramatic shifts in the protistan community in association with the introduction of two distinct invasive species to the lake. We also identified the major groups of protist present and abundant in Lake Mendota and extracted metagenome-assembled genomes for a few photosynthetic protists. Our results reveal how the dynamics of the protistan community have changed in the past 20 years and highlight the power of metagenomic time series in enabling retrospective analysis of diverse taxonomic groups. Our work further highlights the impact of human interventions in freshwater systems and provides new insights into lake protistan ecology.

Abbreviations

rRNA Ribosomal RNA gene
MAG Metagenome-assembled genome

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-024-01831-y>.

Additional file 1: Supplementary Figures and Tables [120–152].

Acknowledgements

We gratefully acknowledge the many individuals who assisted with data collection and sequencing over the course of this project. We gratefully acknowledge Heather Maughan for valuable professional feedback on the revision of the manuscript text.

Authors’ contributions

AIK, FS, TW, and RMB designed the co-occurrence study. AIK and FS refined methods for rRNA gene extraction. AIK processed and analyzed data and generated figures and statistical analyses. KDM oversaw original data collection, provided feedback and ecological insight, and helped refine text. AIK wrote the paper with editing and input from all authors.

Funding

The work conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0020347, under which AIK was supported. AIK completed the work during a practicum experience at the Joint Genome Institute. RRR acknowledges support from a U.S. National Science Foundation Postdoctoral Research Fellowship in Biology award number 2011002. KDM acknowledges support from the National Institute of Food and Agriculture, U.S. Department of Agriculture, Hatch Projects WIS01516, WIS01789, and WIS03004; U.S. National Science Foundation North Temperate Lakes Long-Term Ecological Research site, NTL-LTER, award numbers DEB-9632853, DEB-0217533 (Stephen Carpenter) and DEB-0822700, DEB-1440297 (Emily Stanley); U.S. National Science Foundation Microbial Observatories program, award numbers MCB-9977903 (Eric Triplett) and DEB-0702395 (KDM); and a U.S. National Science Foundation INSPIRE award, DEB-1344254 (KDM).

Availability of data and materials

All sequencing data and assembly products are available online through the Joint Genome Institute Genome Portal under proposal identification number 504350; Additional file 1: Table S1 indicates specific numbers for each of the meta genomic samples. Metagenome-assembled genomes and 18S rRNA gene sequences are accessible from the Open Science framework via link https://osf.io/9epa8/?view_only=152af26e11894ac0bcdfe542e02c6ab1. All analysis code is provided in a GitHub repository accessible via the link <https://github.com/akrinos/2022-krinos-mendota18s>. All other data are provided in supplementary tables.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ²Department of Earth, Atmospheric, and Planetary Science, Massachusetts Institute of Technology, Cambridge, MA, USA. ³MIT-WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Cambridge, Woods Hole, MA, USA. ⁴Joint Genome Institute, Lawrence Berkeley National

Laboratory, Berkeley, CA, USA. ⁵Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA. ⁶Department of Bacteriology, University of Wisconsin at Madison, Madison, WI, USA.

Received: 16 December 2023 Accepted: 6 May 2024

Published online: 20 July 2024

References

- Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, et al. Global biogeography of highly diverse protistan communities in soil. *ISME J*. 2013;7(3):652–9.
- Sherr BF, Sherr EB, Caron DA, Vulot D, Worden AZ. Oceanic protists. *Oceanography*. 2007;20(2):130–4.
- Singer D, Seppely CV, Lentendu G, Dunthorn M, Bass D, Belbahri L, et al. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ Int*. 2021;146:106262.
- Chrost RJ, Tomasz A, Kalinowska K, Skowronska A. Abundance and structure of microbial loop components (bacteria and protists) in lakes of different trophic status. *J Microbiol Biotechnol*. 2009;19(9):858–68.
- Keeling PJ, Del Campo J. Marine protists are not just big bacteria. *Curr Biol*. 2017;27(11):R541–9.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
- Azam F. Microbial control of oceanic carbon flux: the plot thickens. *Science*. 1998;280(5364):694–6.
- Jones RL. Mixotrophy in planktonic protists: an overview. *Freshw Biol*. 2000;45(2):219–26.
- Stoecker DK. Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. *Eur J Protistol*. 1998;34(3):281–90.
- Graham LE, Knack JJ, Graham ME, Graham JM, Zulkifly S. A metagenome for lacustrine Cladophora (Cladophorales) reveals remarkable diversity of eukaryotic epibionts and genes relevant to materials cycling. *J Phycol*. 2015;51(3):408–18.
- Brock TD. *A eutrophic lake: Lake Mendota, Wisconsin*, vol. 55. New York: Springer Science & Business Media; 2012.
- Caron DA, Hu SK. Are we overestimating protistan diversity in nature? *Trends Microbiol*. 2019;27(3):197–205.
- Lepère C, Domaizon I, Taïb N, Mangot JF, Bronner G, Boucher D, et al. Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. *FEMS Microbiol Ecol*. 2013;85(1):85–94.
- Newton RJ, McMahon KD. Seasonal differences in bacterial community composition following nutrient additions in a eutrophic lake. *Environ Microbiol*. 2011;13(4):887–99.
- Rohwer RR, Hale RJ, Vander Zanden MJ, Miller TR, McMahon KD. Species invasions shift microbial phenology in a two-decade freshwater time series. *Proc National Acad Sci*. 2023;120(11):e2211796120.
- Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD. Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol Oceanogr*. 2007;52(2):487–94.
- Kent AD, Jones S, Yannarell A, Graham JM, Lauster G, Kratz T, et al. Annual patterns in bacterioplankton community variability in a humic lake. *Microb Ecol*. 2004;48(4):550–60.
- Sun P, Huang X, Wang Y, Huang B. Protistan-bacterial microbiota exhibit stronger species sorting and greater network connectivity offshore than nearshore across a coast-to-basin continuum. *mSystems*. 2021;6(5):e00100–21.
- Keck F, Millet L, Debroas D, Etienne D, Galop D, Rius D, et al. Assessing the response of micro-eukaryotic diversity to the Great Acceleration using lake sedimentary DNA. *Nat Commun*. 2020;11(1):1–8.
- Gauthier J, Walsh D, Selbie DT, Bourgeois A, Griffiths K, Domaizon I, et al. Evaluating the congruence between DNA-based and morphological taxonomic approaches in water and sediment trap samples: analyses of a 36-month time series from a temperate monomictic lake. *Limnol Oceanogr*. 2021;66(8):3020–39.
- Walsh JR, Carpenter SR, Vander Zanden MJ. Invasive species triggers a massive loss of ecosystem services through a trophic cascade. *Proc Natl Acad Sci*. 2016;113(15):4081–5.
- Walsh JR, Munoz SE, Vander Zanden MJ. Outbreak of an undetected invasive species triggered by a climate anomaly. *Ecosphere*. 2016;7(12):e01628.
- Spear MJ, Walsh JR, Ricciardi A, Zanden MJV. The invasion ecology of sleeper populations: prevalence, persistence, and abrupt shifts. *BioScience*. 2021;71(4):357–69.
- Spear MJ, Wakker PA, Shannon TP, Lowe RL, Burlakova LE, Karatayev AY, et al. Early changes in the benthic community of a eutrophic lake following zebra mussel (*Dreissena polymorpha*) invasion. *Inland Waters*. 2022;12(2):311–29.
- Vanni MJ. Invasive mussels regulate nutrient cycling in the largest freshwater ecosystem on Earth. *Proc Natl Acad Sci*. 2021;118(8):e2100275118.
- Rohwer RR, Ladwig R, Dugan HA, Hanson PC, Walsh JR, Vander Zanden MJ. The aftermath of a trophic cascade: Increased anoxia following species invasion of a eutrophic lake. *bioRxiv*. 2023;2023–01.
- Clum A, Huntemann M, Bushnell B, Foster B, Foster B, Roux S, et al. DOE JGI metagenome workflow. *mSystems*. 2021;6(3):e00804–20.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27(5):824–34.
- Chen IMA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system v. 7: content updates and new features. *Nucleic Acids Res*. 2023;51(D1):D723–32.
- Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29(22):2933–5.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res*. 2003;31(1):439–41.
- Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res*. 2021;49(D1):D192–200.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584.
- Madden T. The BLAST sequence analysis tool. *NCBI Handb*. 2013;2:425–36.
- O’Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733–45.
- Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2007;35(suppl_1):D61–5.
- Kans J. Entrez Direct: E-utilities on the UNIX command line. In: Entrez Programming Utilities Help [Internet]. Washington, D.C.: National Center for Biotechnology Information (US); 2021.
- Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*. 2010;39(suppl_1):D52–7.
- Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*. 2014;12(9):635–45.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2017;8(1):28–36.
- R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357.
- Broad Institute. Picard. 2022. <http://broadinstitute.github.io/picard/>. Accessed 1 Nov 2022.
- Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*. 2012;131(4):281–5.

46. Puente-Sánchez F, García-García N, Tamames J. SQMtools: automated processing and visual analysis of omics data with R and anvio. *BMC Bioinformatics*. 2020;21(1):1–11.
47. Gong W, Marchetti A. Estimation of 18S gene copy number in marine eukaryotic plankton using a next-generation sequencing approach. *Front Mar Sci*. 2019;6:219.
48. Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol*. 2012;8(10):e1002743.
49. Herrera M, Vallor A, Gelfond J, Patterson T, Wickes B. Strain-dependent variation in 18S ribosomal DNA copy numbers in *Aspergillus fumigatus*. *J Clin Microbiol*. 2009;47(5):1325–32.
50. Cockrum C, Kaneshiro KR, Rechtsteiner A, Tabuchi TM, Strome S. A primer for generating and using transcriptome data and gene sets. *Development*. 2020;147(24):dev193854.
51. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*. 2020;26(8):903–9.
52. Coenen AR, Hu SK, Luo E, Muratore D, Weitz JS. A primer for microbiome time-series analysis. *Front Genet*. 2020;11:310.
53. Pasek J, Pasek MJ, Hmisc D. Package 'weights'. Weighting and Weighted Statistics R package version 085. 2016. <https://cran.r-project.org/web/packages/weights/weights.pdf>. Accessed 15 Mar 2021.
54. Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006;22(12):1540–2.
55. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*. 2015;31(22):3718–20.
56. Harrell Jr FE, Harrell Jr MFE. Package 'hmisc'. CRAN2018. 2019;2019:235–6.
57. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)*. 1995;57(1):289–300.
58. Butts CT. network: a package for managing relational data in R. *J Stat Softw*. 2008;24:1–36.
59. Csardi G, Nepusz T, et al. The igraph software package for complex network research. *InterJ Comput Complex Syst*. 2006;1695(5):1–9.
60. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp*. 2008;2008(10):P10008.
61. West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res*. 2018;28(4):569–80.
62. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 2019;7:e7359.
63. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):1–11.
64. Saary P, Mitchell AL, Finn RD. Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. *Genome Biol*. 2020;21(1):1–21.
65. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
66. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*. 2014;12(6):e1001889.
67. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The Genome Portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res*. 2014;42(D1):D26–31.
68. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3.
69. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinforma*. 2014;48(1):3–13.
70. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
71. Krinos AI, Hu SK, Cohen NR, Alexander H. EUKulele: Taxonomic annotation of the unsung eukaryotic microbes. *J Open Source Softw*. 2021;6(57):2817. <https://doi.org/10.21105/joss.02817>.
72. Caron DA, Alexander H, Allen AE, Archibald JM, Armbrust EV, Bachy C, et al. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat Rev Microbiol*. 2017;15(1):6–20.
73. Klemetsen T, Raknes IA, Fu J, Agafonov A, Balasundaram SV, Tartari G, et al. The MAR databases: development and implementation of databases specific for marine metagenomics. *Nucleic Acids Res*. 2018;46(D1):D692–9.
74. Vosloo S, Huo L, Anderson CL, Dai Z, Sevillano M, Pinto A. Evaluating de novo assembly and binning strategies for time series drinking water metagenomes. *Microbiol Spectr*. 2021;9(3):e01434-21.
75. Mirdita M, Steinegger M, Söding J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*. 2019;35(16):2856–8.
76. Sommer U, Gliwicz ZM, Lampert W, Duncan A. The PEG-model of seasonal succession of planktonic events in freshwaters. *Archiv Hydrobiol*. 1986;106(4):433–71.
77. Sommer U, Adrian R, De Senerpont Domis L, Elser JJ, Gaedke U, Ibelings B, et al. Beyond the Plankton Ecology Group (PEG) model: mechanisms driving plankton succession. *Ann Rev Ecol Evol Syst*. 2012;43:429–48.
78. Carey CC, Hanson PC, Lathrop RC, St Amand AL. Using wavelet analyses to examine variability in phytoplankton seasonal succession and annual periodicity. *J Plankton Res*. 2016;38(1):27–40.
79. Lanzoni O, Sabaneyeva E, Modeo L, Castelli M, Lebedeva N, Verni F, et al. Diversity and environmental distribution of the cosmopolitan endosymbiont "Candidatus Megaira". *Nat Sci Reports*. 2019;9(1):1–13.
80. Kurmayer R, Christiansen G, Holzinger A, Rott E. Single colony genetic analysis of epilithic stream algae of the genus *Chamaesiphon* spp. *Hydrobiologia*. 2018;811(1):61–75.
81. Labrenz M, Banfield J. Sulfate-reducing bacteria-dominated biofilms that precipitate ZnS in a subsurface circumneutral-pH mine drainage system. *Microb Ecol*. 2004;47:205–17.
82. Edvardsen B, Paasche E. Bloom dynamics and physiology of *Prymnesium* and *Chrysochromulina*. *NATO ASI Ser G Ecol Sci*. 1998;41:193–208.
83. Walsh JR, Lathrop RC, Vander Zanden MJ. Invasive invertebrate predator, *Bythotrephes longimanus*, reverses trophic cascade in a north-temperate lake. *Limnol Oceanogr*. 2017;62(6):2498–509.
84. Walsh JR, Lathrop RC, Vander Zanden MJ. Uncoupling indicators of water quality due to the invasive zooplankter, *Bythotrephes longimanus*. *Limnol Oceanogr*. 2018;63(3):1313–27.
85. Idrisi N, Mills EL, Rudstam LG. Long-term phytoplankton community dynamics: Oneida Lake (1975–2011). Oneida Lake: long term dynamics of a managed ecosystem and its fisheries. Bethesda: American Fisheries Society; 2016. pp. 139–59.
86. Magnuson J, Carpenter S, Stanley E. North Temperate Lakes LTER: Phytoplankton - Madison Lakes Area 1995 - current ver 29. Environmental Data Initiative; 2020. <https://doi.org/10.6073/pasta/f7550858af209a778ca3f8717ed31ed8>.
87. Unrein F, Gasol JM, Not F, Forn I, Massana R. Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. *ISME J*. 2014;8(1):164–76.
88. Arvola L, Ojala A, Barbosa F, Heaney SI. Migration behaviour of three cryptophytes in relation to environmental gradients: an experimental approach. *Br Phycol J*. 1991;26(4):361–73.
89. Matsuzaki SIS, Lathrop RC, Carpenter SR, Walsh JR, Vander Zanden MJ, Gahler MR, et al. Climate and food web effects on the spring clear-water phase in two north-temperate eutrophic lakes. *Limnol Oceanogr*. 2021;66(1):30–46.
90. McHugh AJ, Yap M, Crispie F, Feehily C, Hill C, Cotter PD. Microbiome-based environmental monitoring of a dairy processing facility highlights the challenges associated with low microbial-load samples. *NPJ Sci Food*. 2021;5(1):1–13.
91. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2(11):1533–42.
92. Magnuson J, Carpenter S, Stanley E. North Temperate Lakes LTER: Zooplankton - Madison Lakes Area 1997 - current ver 31. Environmental Data Initiative; 2019. <https://doi.org/10.6073/pasta/8b265c0300252c87805f26f4e1e174aa4>.

93. Walsh JR, Spear MJ, Shannon TP, Krysan PJ, Vander Zanden MJ. Using eDNA, sediment microfossils, and zooplankton nets to detect invasive spiny water flea (*Bythotrephes longimanus*). *Biol Invasions*. 2019;21(2):377–89.
94. Kublanovskaya A, Solovchenko A, Fedorenko T, Chekanov K, Lobakova E. Natural communities of carotenogenic chlorophyte *Haematococcus lacustris* and bacteria from the White Sea coastal rock ponds. *Microb Ecol*. 2020;79(4):785–800.
95. Loftus SE, Hunt DE, Johnson ZI. Reused cultivation water from a self-inhibiting alga does not inhibit other algae but alters their microbiomes. *Algal Res*. 2020;51:102067.
96. Ramanan R, Kang Z, Kim BH, Cho DH, Jin L, Oh HM, et al. Phycosphere bacterial diversity in green algae reveals an apparent similarity across habitats. *Algal Res*. 2015;8:140–4.
97. Esteve I, Mir J, Gaju N, McKhann HI, Margulis L. Green endosymbiont of Coleps from Lake Cisó identified as *Chlorella vulgaris*. *Symbiosis*. 1988;6:197–210.
98. Selivanova E, Khlopko YA, Gogoleva N, Plotnikov A. Detection of potentially pathogenic bacteria in the brackish rivers flowing into the Elton Lake by high-throughput sequencing. *J Microbiol Epidemiol Immunobiol*. 2018;95(4):87–95.
99. Fernández-Bravo A, Figueras MJ. An update on the genus *Aeromonas*: taxonomy, epidemiology, and pathogenicity. *Microorganisms*. 2020;8(1):129.
100. Lewis WM. Phytoplankton succession in lake Valencia, Venezuela. In: *Seasonality of Freshwater Phytoplankton*. New York: Springer; 1986. pp. 189–203.
101. Jensen J, Jeppesen E, Ollrik K, Kristensen P. Impact of nutrients and physical factors on the shift from cyanobacterial to chlorophyte dominance in shallow Danish lakes. *Can J Fish Aquat Sci*. 1994;51(8):1692–9.
102. Bogard MJ, Vogt RJ, Hayes NM, Leavitt PR. Unabated nitrogen pollution favors growth of toxic cyanobacteria over chlorophytes in most hypereutrophic lakes. *Environ Sci Technol*. 2020;54(6):3219–27.
103. Cantonati M, Rott E, Spitale D, Angeli N, Komárek J. Are benthic algae related to spring types? *Freshwat Sci*. 2012;31(2):481–98.
104. Li Q, Xiao J, Ou T, Han M, Wang J, Chen J, et al. Impact of water level fluctuations on the development of phytoplankton in a large subtropical reservoir: implications for the management of cyanobacteria. *Environ Sci Pollut Res*. 2018;25(2):1306–18.
105. Sorokovikova EG, Belykh OI, Gladkikh AS, Kotsar OV, Tikhonova IV, Timoshkin OA, et al. Diversity of cyanobacterial species and phylotypes in biofilms from the littoral zone of Lake Baikal. *J Microbiol*. 2013;51(6):757–65.
106. Parfenova V, Gladkikh A, Belykh O. Comparative analysis of biodiversity in the planktonic and biofilm bacterial communities in Lake Baikal. *Microbiology*. 2013;82(1):91–101.
107. Liu Y, Shen Y, Cheng C, Yuan W, Gao H, Guo P. Analysis of the influence paths of land use and landscape pattern on organic matter decomposition in river ecosystems: Focusing on microbial groups. *Sci Total Environ*. 2022;817:152999.
108. Jobard M, Rasconi S, Solinhac L, Cauchie HM, Sime-Ngando T. Molecular and morphological diversity of fungi and the associated functions in three European nearby lakes. *Environ Microbiol*. 2012;14(9):2480–94.
109. Cai HY, Yan Zs, Wang AJ, Krumholz LR, Jiang HL. Analysis of the attached microbial community on mucilaginous cyanobacterial aggregates in the eutrophic Lake Taihu reveals the importance of *Planctomycetes*. *Microb Ecol*. 2013;66(1):73–83.
110. Crump BC, Kling GW, Bahr M, Hobbie JE. Bacterioplankton community shifts in an arctic lake correlate with seasonal changes in organic matter source. *Appl Environ Microbiol*. 2003;69(4):2253–68.
111. Kaboré OD, Godreuil S, Drancourt M. *Planctomycetes* as host-associated bacteria: a perspective that holds promise for their future isolations, by mimicking their native environmental niches in clinical microbiology laboratories. *Front Cell Infect Microbiol*. 2020;10:519301.
112. Pizzetti I, Fuchs BM, Gerdts G, Wichels A, Wiltshire KH, Amann R. Temporal variability of coastal *Planctomycetes* clades at Kabeltonne station, North Sea. *Appl Environ Microbiol*. 2011;77(14):5009–17.
113. Lepère C, Domaizon I, Debroas D. Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Appl Environ Microbiol*. 2008;74(10):2940–9.
114. Lefèvre E, Bardot C, Noël C, Carrias JF, Viscogliosi E, Amblard C, et al. Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environ Microbiol*. 2007;9(1):61–71.
115. Sun H, Noe J, Barber J, Coyne R, Cassidy-Hanley D, Clark T, et al. Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis*. *Appl Environ Microbiol*. 2009;75(23):7445–52.
116. Gong J, Qing Y, Zou S, Fu R, Su L, Zhang X, et al. Protist-bacteria associations: Gammaproteobacteria and Alphaproteobacteria are prevalent as digestion-resistant bacteria in ciliated protozoa. *Front Microbiol*. 2016;7:498.
117. Blanchet FG, Cazelles K, Gravel D. Co-occurrence is not evidence of ecological interactions. *Ecol Lett*. 2020;23(7):1050–63.
118. Husnik F, Tashyreva D, Boscaro V, George EE, Lukeš J, Keeling PJ. Bacterial and archaeal symbioses with protists. *Curr Biol*. 2021;31(13):R862–77.
119. Haselkorn TS, DiSalvo S, Miller JW, Bashir U, Brock DA, Queller DC, et al. The specificity of *Burkholderia* symbionts in the social amoeba farming symbiosis: Prevalence, species, genetic and phenotypic diversity. *Mol Ecol*. 2019;28(4):847–62.
120. Hou F, Zhang H, Xie W, Zhou X, Zhu X, Zhang D. Co-occurrence patterns and assembly processes of microeukaryotic communities in an early-spring diatom bloom. *Sci Total Environ*. 2020;711:134624.
121. Hansson LA, Annadotter H, Bergman E, Hamrin SF, Jeppesen E, Kairesalo T, et al. Biomanipulation as an application of food-chain theory: constraints, synthesis, and recommendations for temperate lakes. *Ecosystems*. 1998;1(6):558–74.
122. Carpenter SR, Kitchell JF. Consumer control of lake productivity. *BioScience*. 1988;38(11):764–9.
123. Vanni MJ, Temte J, Allen Y, Dodds R, Howard PJ, Leavitt PR, et al. Herbivory, nutrients, and phytoplankton dynamics in Lake Mendota, 1987–89. In: *Food Web Management*. New York: Springer; 1992. pp. 243–73.
124. Princiotta SD, Hendricks SP, White DS. Production of cyanotoxins by *Microcystis aeruginosa* mediates interactions with the mixotrophic flagellate *Cryptomonas*. *Toxins*. 2019;11(4):223.
125. Jones AC, Hambright KD, Caron DA. Ecological patterns among bacteria and microbial eukaryotes derived from network analyses in a low-salinity lake. *Microb Ecol*. 2018;75(4):917–29.
126. Thomas MK, Fontana S, Reyes M, Kehoe M, Pomati F. The predictability of a lake phytoplankton community, over time-scales of hours to years. *Ecol Lett*. 2018;21(5):619–28.
127. Capo E, Debroas D, Arnaud F, Perga ME, Chardon C, Domaizon I. Tracking a century of changes in microbial eukaryotic diversity in lakes driven by nutrient enrichment and climate warming. *Environ Microbiol*. 2017;19(7):2873–92.
128. Cottingham KL, Weathers KC, Ewing HA, Greer ML, Carey CC. Predicting the effects of climate change on freshwater cyanobacterial blooms requires consideration of the complete cyanobacterial life cycle. *J Plankton Res*. 2021;43(1):10–9.
129. Ladwig R, Hanson PC, Dugan HA, Carey CC, Zhang Y, Shu L, et al. Lake thermal structure drives inter-annual variability in summer anoxia dynamics in a eutrophic lake over 37 years. *Hydro Earth Syst Sci Discuss*. 2020;25(2):1–45.
130. Farrell KJ, Ward NK, Krinos AI, Hanson PC, Daneshmand V, Figueiredo RJ, et al. Ecosystem-scale nutrient cycling responses to increasing air temperatures vary with lake trophic state. *Ecol Model*. 2020;430:109134.
131. Carpenter SR, Arani BM, Hanson PC, Scheffer M, Stanley EH, Van Nes E. Stochastic dynamics of Cyanobacteria in long-term high-frequency observations of a eutrophic lake. *Limnol Oceanogr Lett*. 2020;5(5):331–6.
132. Kara EL, Hanson PC, Hu YH, Winslow L, McMahon KD. A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA. *ISME J*. 2013;7(3):680–4.
133. Linz AM, Aylward FO, Bertilsson S, McMahon KD. Time-series metatranscriptomes reveal conserved patterns between phototrophic and heterotrophic microbes in diverse freshwater systems. *Limnol Oceanogr*. 2020;65:5101–12.
134. Garcia SL, Stevens SL, Cray B, Martinez-Garcia M, Stepanauskas R, Woyke T, et al. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *ISME J*. 2018;12(3):742–55.

135. He S, Stevens SL, Chan LK, Bertilsson S, Glavina del Rio T, Tringe SG, et al. Ecophysiology of freshwater *Verrucomicrobia* inferred from metagenome-assembled genomes. *mSphere*. 2017;2(5):e00277–17.
136. Roux S, Chan LK, Egan R, Malmstrom RR, McMahon KD, Sullivan MB. Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nat Commun*. 2017;8(1):1–12.
137. Beversdorf LJ, Chaston SD, Miller TR, McMahon KD. Microcystin mcyA and mcyE gene abundances are not appropriate indicators of microcystin concentrations in lakes. *PLoS ONE*. 2015;10(5):e0125353.
138. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2012;41(D1):D590–6.
139. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and “all-species living tree project (LTP)” taxonomic frameworks. *Nucleic Acids Res*. 2014;42(D1):D643–8.
140. Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, et al. 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J Biotechnol*. 2017;261:169–76.
141. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res*. 2012;41(D1):D597–604.
142. Del Campo J, Kolisko M, Boscaro V, Santoferrara LF, Nenarokov S, Massana R, et al. EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biol*. 2018;16(9):e2005849.
143. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
144. Benson BJ, Kratz TK, Magnuson JJ, et al. Long-term dynamics of lakes in the landscape: long-term ecological research on north temperate lakes. Oxford University Press on Demand, New York, NY, USA: Oxford; 2006.
145. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261–72.
146. McKinney W, et al. pandas: a foundational Python library for data analysis and statistics. *Python High Perform Sci Comput*. 2011;14(9):1–9. Accessed 15 Jan 2021.
147. Python Language Reference, version 3.7. Python Software Foundation; 2020. <http://www.python.org>. Accessed 15 May 2024.
148. Hoffmann K, Hassenrück C, Salman-Carvalho V, Holtappels M, Bienhold C. Response of bacterial communities to different detritus compositions in Arctic deep-sea sediments. *Front Microbiol*. 2017;8:266.
149. Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P. Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders ‘Frankiales’ and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov. *Int J Syst Evol Microbiol*. 2014;64(Pt_11):3821–32.
150. Zhou Y, Lai YS, Eustance E, Xia S, Rittmann BE. Phosphate depletion controls lipid content and accumulation of heterotrophic bacteria during growth of *Synechocystis* sp. PCC 6803. *Appl Microbiol Biotechnol*. 2019;103(12):5007–14.
151. Siano R, Montresor M. Morphology, ultrastructure and feeding behaviour of *Protoperidinium vorax* sp. nov. (Dinophyceae, Peridiniales). *Eur J Phycol*. 2005;40(2):221–32.
152. Spilling K, Olli K, Lehtoranta J, Kremp A, Tedesco L, Tamelander T, et al. Shifting diatom-dinoflagellate dominance during spring bloom in the Baltic Sea and its potential effects on biogeochemical cycling. *Front Mar Sci*. 2018;5:327.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.