

Cyanobacterial Harmful Algal Bloom Dynamics and Associated Microbial Communities in Clear Lake, California

Alle Lie¹, David Niknejad², Isha Kalra², Brittany Stewart², Kylie
Langlois¹, Eric Webb², David A. Caron², Jayme Smith¹

¹*Southern California Coastal Water Research Project, Costa Mesa, CA*

²*University of Southern California, Los Angeles, CA*

March 2025

ACKNOWLEDGEMENTS

This report was produced with funding from, and is a grant deliverable to, the Central Valley Regional Water Quality Control Board, in fulfillment of the agreement “Clear Lake Cyanobacteria DNA Sequencing” (Agreement No. 19-078-270-3).

The authors thank the Big Valley Band of Pomo Indians and their Environmental Protection Office for sampling assistance. The authors also thank the following individuals for assistance: Christopher Denniston and Gaurav Sukhatme (assistance and data processing for drone sampling and robotics), Amanda Tinoco, Anjali Bhatnagar, Ania Webb, and Gerid Ollison (assistance with field sampling).

Prepared for:

As a Deliverable for the Central Valley Regional Water Quality Control Board "Clear Lake Cyanobacteria DNA Sequencing"

(Agreement No. 19-078-270-3)

This report should be cited as:

Lie, A., D. Niknejad, I. Kalra, B. Stewart, K. Langlois, E. Webb, D.A. Caron, and J. Smith. 2025. Cyanobacterial Harmful Algal Bloom Dynamics and Associated Microbial Communities in Clear Lake. Southern California Coastal Water Research Project. Costa Mesa, CA.

Keywords:

Harmful algal blooms; cyanotoxins; cyanobacteria; recreation; lakes; reservoirs

EXECUTIVE SUMMARY

Clear Lake is a large, freshwater lake in Lake County, California, U.S.A. that experiences annual cyanobacterial harmful algal blooms (cyanoHABs) that impact multiple beneficial uses of water. Monitoring of cyanotoxins in the lake has documented extremely high concentrations of cyanotoxins, specifically microcystins (MC), for which it has recently been listed on the 303(d) list of impaired waters. Given the clear impacts of cyanoHABs in the lake, a major goal is to understand the environmental factors that may be stimulating their chronic reoccurrence in the lake, to understand what management options are possible.

To date, careful study of the cyanobacterial and associated microbial and algal communities has been limited, constituting a key knowledge gap limiting the effective management of cyanoHABs in the lake. The study of these communities via environmental DNA and RNA via molecular 'omics approaches can inform mitigation strategies and support the development of models for predicting cyanoHABs. This current study is a continuation (Phase 2) of a larger effort to investigate the environmental drivers of cyanoHABs and cyanotoxin production in Clear Lake. This current study focused on a key recommendation of the project Phase 1 findings (reported in Florea et al. (2022) and Kalra et al. (accepted)) to understand cyanobacterial diversity, identify which taxa can produce cyanotoxins, and determine the conditions under which toxin is produced. Towards this goal, we applied a variety of molecular 'omics approaches to characterize the cyanobacterial, bacterial, and eukaryotic algal communities within the lake.

Key Findings

This study focused on understanding the differences in environmental and community composition of lake conditions in relations to the chl-a and MC concentrations. Chl-a and MC thresholds were used in study to divide the lake in distinct conditions (categories). The chl-a total daily maximum loads (TMDL) of 73 $\mu\text{g/L}$ target established for Clear Lake was chosen as a policy-driven threshold for chl-a. Using this threshold will allow a better understanding of environmental conditions (such as nutrient availability) and community composition when the chl-a concentrations were above or below the chl-a TMDL target. The threshold for MC presence was the detection limit for MC measurement at 0.15 $\mu\text{g/L}$. Therefore, samples in this study were divided into four lake conditions: 1) Above chl-a TMDL target & MC present; 2) Above chl-a TMDL target & MC absent; 3) Below chl-a TMDL target & MC present; and 4) Below chl-a TMDL target & MC absent.

Key findings of this study and the relationship to the findings of Phase 1 include the following:

- 1. Multiple cyanobacterial taxa in Clear Lake have the ability to product microcystins:** While *Microcystis* was identified as a MC producer in Florea et al. (2022) and Kalra et al. (accepted), the results from this study further suggests that *Limnoraphis* and *Planktothrix* may also be producers of MC in Clear Lake.
- 2. Distinct cyanobacterial and microbial taxa were associated with lake conditions:** DNA metabarcoding showed that certain taxa had strong association with specific lake conditions. Since most of the samples sequenced for cyanobacterial communities in this study were from 2020, when *Microcystis* was not a dominant part of the cyanobacterial community, *Microcystis* did not constitute a notable proportion of the cyanobacterial community in most samples (Figure 8 Kalra et al., accepted). Instead, the cyanobacteria *Limnoraphis* primarily dominated communities when MC was present while *Dolichospermum* was abundant in samples without detectable levels of MC.
- 3. Diversity and composition of the microbial community varied significantly with chl-a and MC concentrations:** The diversity of the prokaryotic community was significantly lower in the presence of MC. On the other hand, the diversity of the eukaryotic community was not affected by the presence of MC but was instead significantly higher when chl-a concentrations were above the target TMDL level. Multivariate statistical analyses indicated both chl-a and MC concentrations significantly explained the variations in the composition of prokaryotic and eukaryotic communities. While MC is produced by certain cyanobacteria, its presence can also shape microbial community structure through potential allelopathic interactions and grazer deterrence effects on eukaryotes.
- 4. Nutrients played an important role in driving lake conditions:** TN concentration was strongly and positively correlated to total MC concentrations in both this study and that of Florea et al. (2022) and Kalra et al. (accepted). As might be expected with the analysis of the total nutrients, which includes phytoplankton, TN and total phosphorus (TP) had positive correlation to chl-a concentrations across studies. Higher nutrients (TN) were associated with higher MC. Multivariate statistical analyses also revealed that TN, TP, and TN:TP ratios were significantly different across lake conditions, and samples above chl-a TMDL target & MC present tended to have higher TN and TP in 2021. These results reaffirmed the importance of nutrients and chl-a in driving high concentrations of MC.
- 5. TP concentrations exerted significant control on microbial community diversity:** Multivariate statistical analyses revealed that TP was a significant explanatory factor of both prokaryotic (including cyanobacteria) and eukaryotic microbial communities. The multivariate analysis results suggested that eukaryotic communities associated with Above chl-a TMDL target & MC present conditions were compositionally distinct, with a higher proportion of ciliates and other heterotrophic protists that tended to occur with higher TP

concentrations. While this pattern is based on multivariate ordination and does not necessarily imply a statistically significant relationship on its own, statistical analysis of the full model confirmed that TP had a significant influence on microbial community composition. The results of this current study also align with the findings of Kalra et al. (accepted) that found TP concentrations were a driver of microbial community diversity.

Recommendations for future work

The findings in this study reinforce the complexity of cyanoHAB dynamics in Clear Lake, emphasizing the importance of phosphorus and nitrogen availability, microbial interactions, and potential adaptive responses of the eukaryotic community to long-term cyanotoxin exposure. These insights provide a foundation for refining monitoring strategies and developing targeted mitigation efforts for Clear Lake. Recommendations for future work include:

- **Cyanobacterial HAB monitoring efforts should account for the presence of multiple genera of microcystin producers:** The results of this study identified several cyanobacteria genera positively associated with total MC concentrations. Monitoring efforts should not focus on a single cyanobacterial taxon (e.g., *Microcystis*) and instead should be more broadly focused. This will ensure rapid and early detection of blooms with MC production potential.
- **Bloom management efforts in Clear Lake should consider the influence of multiple bloom drivers:** Results of this study indicated that both TN and TP were strong drivers of lake conditions, including blooms with chl-a levels exceeding the current TMDL target and the presence of MC. TP has long been the focus of bloom controls, however the results of this study have pointed to the potential role of nitrogen in both the formation of high levels of cyanobacterial biomass and the regulation of toxin production. Future management efforts should include nitrogen in addition to phosphorus.
- **Apply the molecular methods piloted in this study to explore additional cyanobacterial bloom drivers:** This study was conducted in years that were identified as dry (2020) and critically dry (2021) hydrologically. In addition to nutrients, a recent historical analysis of Clear Lake suggested that hydrologic factors such as precipitation, lake level, and lake discharge rates may play an important role in bloom formation and toxin production. Thus, this study did not capture the full gradient of hydrologic conditions at Clear Lake. Future study should consider these factors to determine if management of these factors may also support water quality goals for Clear Lake. The molecular ‘omics methods employed in this study provide a useful framework for this work.

TABLE OF CONTENTS

Acknowledgements	i
Executive Summary	ii
Table of Contents	v
Table of Figures	vi
Table of Tables	x
Introduction	1
Methods	3
Study Site Description	3
Discrete Sample Collection	5
Discrete Laboratory Analyses	5
Nucleic acid Analyses	6
Statistical Analyses	10
Results and Discussion	11
Environmental parameters associated with lake conditions	11
Prokaryotic community composition associated with lake conditions	17
Eukaryotic community composition associated with lake conditions	24
Environmental drivers of prokaryotic and eukaryotic communities	29
Identification of MC producers via metagenomic sequencing	33
Insights from metatranscriptomic analyses	35
Recommendations	44
References	46

TABLE OF FIGURES

Figure 1. Map of Clear Lake indicating the ten sampling sites. 4

Figure 2. Division of samples into four lake conditions based on chlorophyll *a* and total microcystins concentrations. The target chl-*a* TMDL at Clear Lake is 73 µg/L while the presence of toxin is defined as total MC ≥ 0.15 µg/L. Bars indicate chl-*a* concentrations while alphabets above the bars indicate the total MC level: “P” = “Present” (≥ 0.15 and < 0.8 µg/L); “C” = “Caution” (≥ 0.8 and < 6 µg/L); “W” = “Warning” (≥ 6 and < 20 µg/L); “D” = “Danger” (≥ 20 µg/L). The lack of alphabet above the bar indicates no total MC was detected (< 0.15 µg/L). Red dashed line indicates a chl-*a* concentration of 73 µg/L. Color of the bars represent the lake condition..... 12

Figure 3. Principal coordinates analysis of environmental parameters (TN, TP, and TN:TP) of data from 2020 and 2021. Samples (points) are colored by lake conditions. ‘Above chl-*a* TMDL target & MC absent’ samples (green) are in the shape of squares for easy differentiation with the blue ‘Below chl-*a* TMDL target & MC present’ samples. Vectors represent environmental variables correlated with the ordination, with arrow direction indicating the gradient and arrow length representing the strength of correlation. The embedded table shows the correlation coefficient and *p* value for each environmental parameter in relation to the principal coordinates (PC). 15

Figure 4. Principal coordinates analysis of environmental parameters (TN, TP, and TN:TP) of data from 2020. Samples (points) are colored by lake conditions. ‘Above chl-*a* TMDL target & MC absent’ samples (green) are in the shape of squares for easy differentiation with the blue ‘Below chl-*a* TMDL target & MC present’ samples. Vectors represent environmental variables correlated with the ordination, with arrow direction indicating the gradient and arrow length representing the strength of correlation. The embedded table shows the correlation coefficient and *p* value for each environmental parameter in relation to the principal coordinates (PC). 16

Figure 5. Principal coordinates analysis of environmental parameters (TN, TP, and TN:TP) of data from 2021. Samples (points) are colored by lake conditions. ‘Above chl-*a* TMDL target & MC absent’ samples (green) are in the shape of squares for easy differentiation with the blue ‘Below chl-*a* TMDL target & MC present’ samples. Vectors represent environmental variables correlated with the ordination, with arrow direction indicating the gradient and arrow length representing the strength of correlation. The embedded table shows the correlation coefficient and *p* value for each environmental parameter in relation to the principal coordinates (PC). 17

Figure 6. Shannon diversity index of the prokaryotic community across A) different lake conditions, B) in the presence or absence of MC (≥ 0.15 µg/L), and C) whether chl-*a* was above or below the TMDL target (≥ 73 µg/L chl-*a*)..... 18

Figure 7. Composition of the prokaryotic community (based on ASV reads) across different lake conditions. Sample names follow the format of

Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 19

Figure 8. Composition of the prokaryotic community (bottom panel) displayed for each sample, arranged in ascending order of chl-a concentrations (top panel; i.e. samples on the left have lowest chl-a and samples on the right have highest chl-a). Red dashed line in the top panel indicates the chl-a TMDL target value of 73 µg/L, the threshold for bloom categorization in this study. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 20

Figure 9. Composition of the prokaryotic community (bottom panel) displayed for each sample, arranged in ascending order of total MC concentrations (top panel; i.e. samples on the left have lowest MC and samples on the right have highest MC). Dashed lines in the top panel indicate various levels of the California Cyanobacteria and Harmful Algal Bloom Network (CCHAB) microcystins trigger level guidelines. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 21

Figure 10. Composition of the cyanobacteria community (based on ASV reads) across different lake conditions. Reads for other bacteria or chloroplasts were excluded. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample. 22

Figure 11. Percentage of reads contribution identified as several cyanobacteria taxa across different lake conditions. 23

Figure 12. Composition of the cyanobacteria community (bottom panel) displayed for each sample, arranged in ascending order of total microcystins (MC) concentrations (top panel; i.e. samples on the left have lowest MC and samples on the right have highest MC). Reads for other bacteria or chloroplasts were excluded. Dashed lines in the top panel indicate various levels of the California Cyanobacteria and Harmful Algal Bloom Network (CCHAB) microcystins trigger level guidelines. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 24

Figure 13. Shannon diversity index of the eukaryotic community across A) different lake conditions, B) in the presence or absence of MC (≥ 0.15 µg/L), and C), or whether chl-a concentrations were above/below the TMDL target (73 µg/L chl-a). 25

Figure 14. Shannon diversity index of the eukaryotic community displayed for each sample, arranged in ascending order of chl-a concentrations (top panel; i.e. samples on the left have lowest chl-a and samples on the right have highest chl-a). Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 26

Figure 15. Composition of the eukaryotic community (based on ASV reads) across different lake conditions. Sample names follow the format of

Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 27

Figure 16. Composition of the eukaryotic community (bottom panel) displayed for each sample, arranged in ascending order of total microcystins (MC) concentrations (top panel; i.e. samples on the left have lowest MC and samples on the right have highest MC). Dashed lines in the top panel indicate various levels of the California Cyanobacteria and Harmful Algal Bloom Network (CCHAB) microcystins trigger level guidelines. Colors of the bars in the top panel indicate the year of the sample. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample. 28

Figure 17. Composition of the eukaryotic community (bottom panel) displayed for each sample, arranged in ascending order of chl-a concentrations (top panel; i.e. samples on the left have lowest chl-a and samples on the right have highest chl-a). Red dashed line in the top panel indicates the TMDL value of 73 µg/L, the threshold for bloom categorization in this study. Colors of the bars in the top panel indicates the year of the sample. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample. 29

Figure 18. Canonical correlation analysis of the prokaryotic community and environmental parameters (temperature, DO, total MC, chl-a, TN, TP, TN:TP). Samples (points) are colored by lake conditions. Green 'Above chl-a TMDL target & MC absent' samples have square shapes to distinguish them from the blue 'Below chl-a TMDL target & MC absent' samples. Right panel is a zoomed in portion of the portion framed by the red square in the left panel. Some sample names are indicated for outlier data points. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample. 30

Figure 19. Canonical correlation analysis of the cyanobacterial community and environmental parameters (temperature, DO, total MC, chl-a, TN, TP, TN:TP). Samples (points) are colored by lake conditions. Green 'Above chl-a TMDL target & MC absent' samples have square shapes to distinguish them from the blue 'Below chl-a TMDL target & MC absent' samples. 31

Figure 20. Canonical correlation analysis of the eukaryotic community and environmental parameters (temperature, DO, total MC, chl-a, TN, TP, TN:TP). Samples (points) are colored by lake conditions. Green 'Above chl-a TMDL target & MC absent' samples have square shapes to distinguish them from the blue 'Below chl-a TMDL target & MC absent' samples. Right panel is a zoomed in portion of the portion framed by the red square in the left panel. Some sample names are indicated for outlier data points. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample. 32

Figure 21. The distribution of the three categories (BP: Biological Process; CC: Cellular Component; MF: Molecular Function) of gene ontology (GO) terms for mapped prokaryotic RNA across different lake conditions. Sample names follow the format of

Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 36

Figure 22. The distribution of the three categories (BP: Biological Process; CC: Cellular Component; MF: Molecular Function) of gene ontology (GO) terms for mapped eukaryotic RNA across different lake conditions. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 37

Figure 23. Percentage of prokaryotic RNA associated with photosynthesis proteins. Colors of the bar graphs indicate sample year. Y-axis shows the percentage of RNA reads that were identified as genes related to photosynthesis based on the GO termed assigned to the gene. X-axis shows the sample names, which follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 38

Figure 24. Percentage of eukaryotic RNA associated with photosynthesis proteins. Colors of the bar graphs indicate sample year. Y-axis shows the percentage of RNA reads that were identified as genes related to photosynthesis based on the GO termed assigned to the gene. X-axis shows the sample names, which follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample..... 39

TABLE OF TABLES

Table 1. Geographic coordinates of sampling stations.....	4
Table 2. Number of samples in each lake condition selected for different types of nucleic acid analyses.	7
Table 3. Primer sequences used for metabarcoding sequencing in Phase 2.....	9
Table 4. Spearman’s correlation results between chl-a and total MC, and measured physicochemical parameters including temperature, dissolved oxygen (DO), total nitrogen (TN), total phosphorus (TP), and the ratio of TN:TP. Asterisks indicate statistical significance (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).....	13
Table 5. Details on genomes assembled from metagenomic samples and indication of the presence of microcystin synthetase (<i>mcy</i>) gene cluster.	33

INTRODUCTION

Clear Lake is a culturally, economically, and recreationally important freshwater lake in California that has been experiencing recurring cyanobacterial harmful algal blooms (cyanoHABs) for over a decade. The lake provides a variety of beneficial uses for recreation, agricultural irrigation, drinking water, and habitat for fish and wildlife. Notably, it is important for Tribal Beneficial Uses, that include uses of the water that support tribal cultural, spiritual, and ceremonial uses and Tribal subsistence fishing. Clear Lake has experienced cyanoHABs associated with extreme concentrations of cyanotoxins, exceeding 16,920 µg/L, which is more than 840 times the California 'danger' (20 µg/L) recreational guidelines for microcystins (MC; Smith et al., 2023; California Cyanobacterial and Harmful Algal Bloom Network, 2016).

CyanoHABs and the associated cyanotoxins in the water pose a threat to residents and visitors to the lake, as well as Tribal cultural and subsistence fishing uses that may overlap in time and place with toxic events. MCs have recently been reported at levels exceeding US EPA health guidance levels in multiple self-supplied water systems around the lake (Stanton et al., 2023). At least one dog death has been linked to the presence of cyanotoxins in the lake (Moore et al., 2016), and adverse effects of these toxins on aquatic organisms is suspected but presently not well understood (Mehinto et al., 2021). Recent concerns in other geographic areas in the United States over potential exposure through aerosolization of toxins has also raised human health concerns (Backer et al., 2008; Facciponte et al., 2018; Olson et al., 2020). Finally, the decrease in water quality and aesthetics has threatened tourism and fishing related revenue to the community. Due to the adverse impacts to beneficial uses by cyanoHABs, Clear Lake was added to the Clean Water Act Section 303(d) List of Impaired Water Bodies due to nutrient impairments. More recently, Clear Lake has also been listed on the 303(d) list due to impairments caused by MCs.

Given the clear impacts in the lake caused by cyanoHABs, a major goal is to understand the environmental factors that may be stimulating the chronic reoccurrence of cyanoHABs in the lake. A major management need is to better link the chemical and physical processes in the lake to the cyanobacterial bloom dynamics to understand potential mitigation options. A variety of toxigenic cyanobacterial taxa have been reported in the lake over the last decade, but to date careful study of the cyanobacterial and associated microbial and algal communities has been rare (Smith et al., 2023). Quantifying the interactions between the chemical, physical and biological factors that stimulate cyanoHABs and cyanotoxin production will be necessary to properly design effective lake mitigation and management strategies to combat them. Approaches such as metabarcoding, metatranscriptomics, and metagenomics are highly effective for studying cyanobacteria because they provide comprehensive insights into their

genetic diversity, metabolic potential, and adaptive responses to environmental changes. Integration of molecular 'omics 'approaches into such studies can inform mitigation strategies and support the development of models for predicting cyanoHABs.

The overall purpose of the study was to identify the environmental drivers (i.e., lake chemistry and physics) leading to cyanobacterial bloom development and the production of cyanotoxins. This study builds upon the work of Florea et al. (2022) and Kalra et al. (accepted), which reported on the first phase of the environmental drivers study conducted in Clear Lake in 2020 and 2021. This study is a continuation of that work and is focused on investigating the environmental factors and community composition of both the prokaryotic and eukaryotic communities that are associated with different lake conditions. The diversity of cyanobacteria and other taxa was explored via DNA metabarcoding, while metabolic activities through metatranscriptomic reads and metagenomic assemblage. The goals of this molecular investigation of the microbial community of Clear Lake were to 1) more clearly identify the specific taxa that are producing cyanotoxins (focusing on MC); 2) understand the differences in diversity and community composition of the cyanobacterial and co-occurring microbial assemblages across differing lake conditions; and 3) examine the extent of abiotic factors in driving microbial community structure and cyanoHAB occurrence patterns within Clear Lake.

METHODS

Study Site Description

Clear Lake has a distinctive basin morphology with three major Arms (Figure 1). The Upper Arm (western lobe) is the oldest and largest basin of the lake at roughly 127 km². The Oaks Arm (the northeastern lobe) and Lower Arm (the southeastern lobe) are smaller at 37.2 and 12.5 km², respectively. Clear Lake is a shallow lake, with maximal depths of approximately 18 meters in the Oaks and Lower Arms, and 12 meters in the Upper Arm (Richerson et al., 1994).

Paleoecological studies have indicated it has been a shallow, productive system since the last ice age (Bradbury, 1988). Lake depth is affected by seasonal rain and runoff, human use which includes agricultural irrigation and drinking water, and is controlled by a dam at the southeastern end of the Lower Arm. Clear Lake is polymictic throughout the year due to its depth, large surface area, and the strong winds that result in frequent and thorough water column mixing. The lake is naturally eutrophic and is among the oldest lakes in North America, formed through volcanism, seismic activity, and erosion (Richerson et al., 1994).

A total of thirteen lake-wide surveys were conducted in 2020 and 2021 during the first phase of the study to characterize the microbial community, cyanotoxin concentrations, and corresponding physio-chemical parameters at Clear Lake. The 2020 survey focused on capturing cyanobacterial dynamics at a higher temporal resolution through sampling approximately every 3-4 days during the month of August. Boat surveys were conducted in 2020 on August 5, August 8, August 11, August 14, August 18, August 21, August 25, and August 28. The 2021 survey focused on characterizing seasonal dynamics across the bloom season and sampling was conducted at a coarser monthly temporal interval. Surveys in 2021 were conducted on June 17, July 13, August 10, September 20, and October 28. Both years consisted of ten sampling stations per lake survey, distributed amongst the three arms of the lake (Figure 1, Table 1). All sampling stations were conducted in open water (i.e., away from the immediate proximity of the shore).

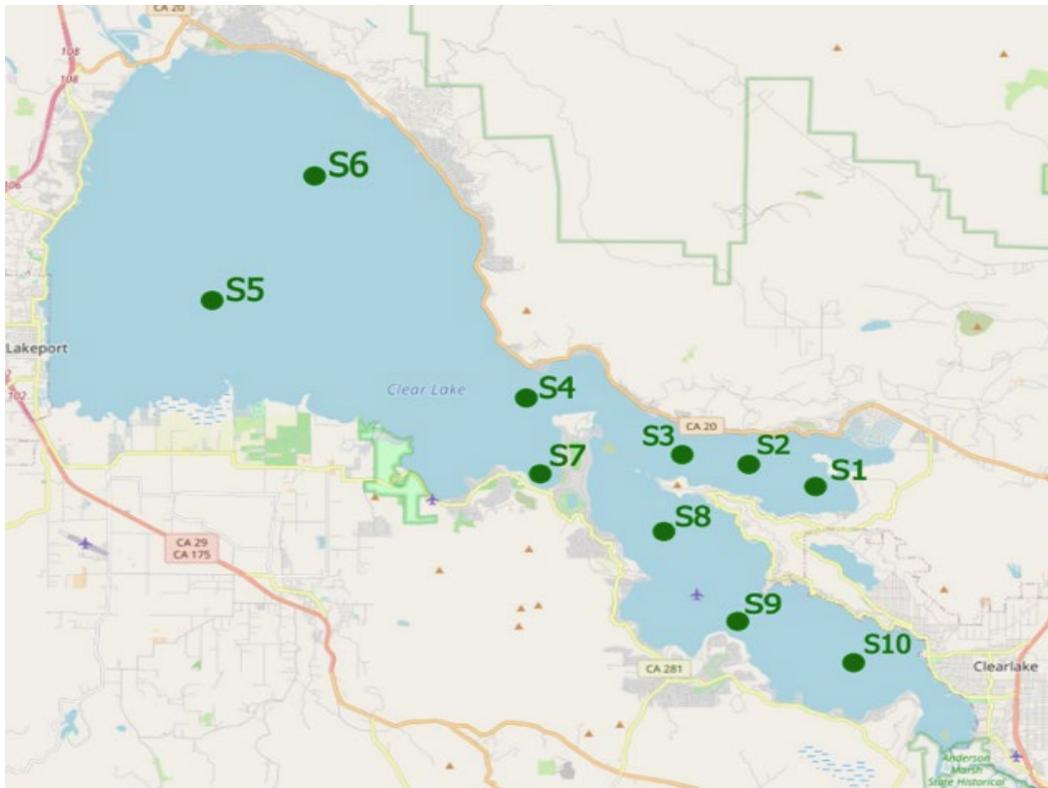


Figure 1. Map of Clear Lake indicating the ten sampling sites.

Table 1. Geographic coordinates of sampling stations.

Station ID	Lake Arm	Latitude	Longitude
Station 1 (S01)	Oaks	39°0.352'	-122°41.137'
Station 2 (S02)	Oaks	39°0.747'	-122°42.347'
Station 3 (S03)	Oaks	39°0.881'	-122°43.455'
Station 4 (S04)	Upper	39°1.767'	-122°46.092'
Station 5 (S05)	Upper	39°3.410'	-122°51.746'
Station 6 (S06)	Upper	39°5.405'	-122°50.067'
Station 7 (S07)	Upper	39°0.593'	-122°45.871'
Station 8 (S08)	Lower	38°59.703'	-122°43.780'

Station ID	Lake Arm	Latitude	Longitude
Station 9 (S09)	Lower	38°58.234'	-122°42.349'
Station 10 (S10)	Lower	38°57.582'	-122°40.353'

Discrete Sample Collection

Whole (i.e., unfiltered) water samples were collected at each sampling location using a clean plastic bucket at or near the water surface. Filled bottles were kept cool and dark while in the field. Additional processing of water samples was conducted at an onsite field laboratory within 12 hours of sample collection. Water samples were used for molecular analyses, cyanotoxins, chlorophyll-*a*, and nutrients. Molecular samples were flash frozen in liquid nitrogen and stored long term at -80°C. Nutrient, cyanotoxin, and chlorophyll-*a* samples were frozen in the field and later stored at -20°C upon return to the lab.

Vertical profiles of water temperature, dissolved oxygen (DO), pH, and conductivity were conducted at each station using an RBR Concerto (<https://rbr-global.com/>). Operation and maintenance of water quality meters followed the manufacturer's recommendations. Profiles were successfully collected at most stations each year, except for the survey conducted in October 2021, when the instrument failed.

Discrete Laboratory Analyses

Microcystins analyses from Water Samples

Total MC was analyzed via ADDA ELISA test kits (Abraxis, Part No. 520011, Warminster, PA). The assay detects all microcystin and nodularin variants with the ADDA side group in bulk and does not provide data for specific congeners of the toxin class. Samples were lysed via a freeze-thaw cycle three times to ensure cell disruption. The extract was then filtered and analyzed according to the manufacturer's instructions. Samples with concentrations higher than the standard curve were serially diluted with kit-provided dilution buffer until sample concentration was within the working range of the kit.

Chlorophyll-*a*

25 mL of whole water for chlorophyll-*a* (chl-*a*) analysis were concentrated via gentle filtration onto glass fiber filters (Sterlitech, grade F, Kent, WA). Filters were extracted in 100% acetone at -20°C in the dark for 24 hours to ensure thorough extraction. Sample extracts were analyzed

fluorometrically via the non-acidification method using a Trilogy Turner Designs fluorometer (Turner Designs, Sunnyvale, CA). Duplicate filters were collected at all stations and the average chl-a concentration of the two filters is reported.

Nutrients

Total nitrogen (TN) and total phosphorus (TP) samples were collected by aliquoting whole water into a single 60 mL HDPE jar for analyses. Samples were kept on ice, and frozen at -20°C immediately upon return to the lab. Samples were analyzed colorimetrically following EPA Methods 353.2 for TN and 365.1 for TP at the University of Maryland Center for Environmental Science Nutrient Analytical Services Laboratory. Ratios of TN:TP were obtained by calculation.

Nucleic acid Analyses

This study incorporated several types of nucleic acid analyses. Metabarcoding combines DNA barcoding with high-throughput sequencing to analyze the biodiversity and composition of communities. The V4 hypervariable region of the small subunit rRNA genes (16S for prokaryotes and 18S for eukaryotes) was amplified and sequenced in this study to investigate the microbial community. Metatranscriptomic sequencing examines the complete set of RNA transcripts in an environmental sample to identify active genes and active members of the community. Metagenomic sequencing, on the other hand, studies the collective genomic content of microorganisms in an environmental sample, providing insights into their gene diversity and metabolic capabilities, as well as serving as a framework for mapping RNA sequences from metatranscriptomic analysis.

Sample selection

A subset of samples was selected for the different types of nucleic acid analyses. This subset is distinct from the subset of samples for which sequencing results were reported in Florea et al. (2022) and Kalra et al. (accepted). Samples were divided into lake conditions based on their chl-a and MC concentrations. The TMDL chl-a target of 73 µg/L was chosen as the chl-a threshold in this study due to its relevance to policy (Central Valley Regional Water Quality Control Board, 2007). This TMDL threshold has been adopted by the Central Valley Regional Water Quality Control Board and approved by the California State Water Resources Control Board and United States Environmental Protection Agency (USEPA). The presence or absence of MC was determined using the method detection limit of the ADDA ELISA test kits (≥ 0.15 µg/L). Therefore, samples were divided into four lake conditions: 1) Above chl-a TMDL target & MC present; 2) Above chl-a TMDL target & MC absent; 3) Below chl-a TMDL target & MC present; and 4) Below chl-a TMDL target & MC absent. Representative samples from each category were selected for nucleic acid analyses. A total of 117 samples were sequenced for metabarcoding

analyses, 86 for metatranscriptomic analyses, and 7 for metagenomic analyses (Table 2). Samples without detectable MC were specifically selected for metagenomic analysis to complement the primarily MC-present samples (26 of 27) analyzed in Florea et al. (2022) and to provide baseline genomes from an environment without MC.

Table 2. Number of samples in each lake condition selected for different types of nucleic acid analyses.

Table 2A. Number of samples in each lake condition selected for metabarcoding sequencing using universal primers.

	Above chl-a TMDL target ($\geq 73 \mu\text{g/L}$)	Below chl-a TMDL target ($< 73 \mu\text{g/L}$)
MC present ($\geq 0.15 \mu\text{g/L}$)	4	20
MC not present ($< 0.15 \mu\text{g/L}$)	3	19

Table 2B. Number of samples in each lake condition selected for metabarcoding sequencing using 18S primers.

	Above chl-a TMDL target ($\geq 73 \mu\text{g/L}$)	Below chl-a TMDL target ($< 73 \mu\text{g/L}$)
MC present ($\geq 0.15 \mu\text{g/L}$)	25	23
MC not present ($< 0.15 \mu\text{g/L}$)	3	20

Table 2C. Number of samples in each lake condition selected for prokaryotic metatranscriptomic sequencing.

	Above chl-a TMDL target ($\geq 73 \mu\text{g/L}$)	Below chl-a TMDL target ($< 73 \mu\text{g/L}$)
MC present ($\geq 0.15 \mu\text{g/L}$)	3	20
MC not present ($< 0.15 \mu\text{g/L}$)	1	19

Table 2D. Number of samples in each lake condition selected for eukaryotic metatranscriptomic sequencing.

	Above chl-a TMDL target ($\geq 73 \mu\text{g/L}$)	Below chl-a TMDL target ($< 73 \mu\text{g/L}$)
MC present ($\geq 0.15 \mu\text{g/L}$)	3	20
MC not present ($< 0.15 \mu\text{g/L}$)	1	19

Table 2E. Number of samples in each lake condition selected for metagenomic sequencing in Phase 2.

	Above chl-a TMDL target ($\geq 73 \mu\text{g/L}$)	Below chl-a TMDL target ($< 73 \mu\text{g/L}$)
MC present ($\geq 0.15 \mu\text{g/L}$)	0	0
MC not present ($< 0.15 \mu\text{g/L}$)	1	6

Nucleic Acid Extraction

DNA samples for metabarcoding and metagenomics were extracted using the Qiagen DNEasy PowerBiofilm Kit. Cyanobacteria from Clear Lake are difficult to extract using the kit alone because of the high biomass of cyanobacteria and the difficulty in breaking open colonial cyanobacteria embedded in mucilage. Several protocol optimization experiments resulted in the addition of a few steps to the Qiagen protocol. The first two Qiagen solutions were added to the samples as the protocol states. The samples were then rapidly freeze-thawed in liquid nitrogen to enhance lysing of cells. 25 μL of proteinase K was then added to each sample and incubated at 55°C for a minimum of four hours or overnight. After the proteinase K incubation, the Qiagen kit protocol was followed according to the manufacturer's instructions. Extracted DNA were quantified with NanoDrop UV-Vis spectroscopy and Qubit Spectrofluorometry. NanoDrop was used to determine nucleic acid quality by confirming a 260/280 ratio between 1.8 and 2. The prokaryotic community was investigated by amplifying the 16S V4 – V5 region using the 515F/926R primer set (Yeh et al., 2021; Table 3), while the 18S V4 region of eukaryotic organisms was amplified using the TAREuk454FWD1/TAREukREV3 primer set (Stoeck et al., 2010; Table 3). While the 515Y/926R primer set universally binds to both 16S and 18S rRNA genes, this primer set was only used for the study of the prokaryotic community since the primers don't amplify full, overlapping regions for the 18S rRNA gene. This limitation makes bioinformatic processing of 18S sequences less effective. To address this, primers specific to the

18S V4 region were used to investigate the eukaryotic community. Extracted and amplified DNA samples were sequenced on the Illumina MiSeq and NovaSeq platforms using the 2x250 bp chemistry.

Table 3. Primer sequences used for metabarcoding sequencing in Phase 2.

Target gene	Primer name	Direction	Primer sequence
16S rRNA gene	515F	Forward	GTGYCAGCMGCCGCGGTAA
16S rRNA gene	926R	Reverse	CCGYCAATTYMTTTRAGTTT
18S rRNA gene	TAReuk454FWD1	Forward	CCAGCASCYGCGGTAATTCC
18S rRNA gene	TAReukREV3	Reverse	ACTTTCGTTCTTGATYRA

RNA samples for metatranscriptomic analysis were extracted using the Qiagen RNeasy Mini Kit following the manufacturer’s instructions, including DNA digestion with DNase. PolyA RNA selection was performed on the samples to enrich for eukaryotic RNA for the eukaryotic RNA sample. The extracted RNA samples were then sequenced on an Illumina NovaSeq platform with 2x150 bp chemistry.

Metabarcoding data processing

The 16S and 18S metabarcoding sequences (amplicons) were processed and analyzed separately, but both were processed using the software Qiime 2 (Amplicon Distribution 2024.10) using a modification of the Caron Lab 18S rRNA tag-sequencing pipeline (https://github.com/shu251/V4_tagsequencing_18Sdiversity_qiime2). Briefly, both sets of amplicons were preliminary trimmed using trimmomatic (with a min length requirement of 249 bp). Cutadapt was used to remove primers and any untrimmed reads were discarded. Then dada2 was used to call amplicon sequence variants (ASVs) and remove chimera sequences (artificial sequences formed when two or more distinct DNA sequences are erroneously joined together). ASVs are unique amplicon sequences that represent a high-resolution taxonomic unit, providing a more precise view of microbial diversity. Each ASV sequence was then assigned a taxonomic identification using BLAST (0.97 similarity cutoff) against a reference database (CyanoSeq V.1.3 merged with Silva 138 for 16S; PR² V4.13 for 18S). ASVs with low occurrences (< 10 reads) across the dataset were removed.

Metatranscriptomic data processing

Seq2fun (Liu et al., 2021) was used to process metatranscriptomic data and provide an overview of the number of RNA sequences (reads) associated with certain functions. This tool first performs quality control of raw reads and then joins overlapped pair-end reads. The tool then translates the RNA reads into protein sequences for alignment to a protein database. The prokaryotic metatranscriptomic data was aligned to the Seq2fun pre-built algae database, while the eukaryotic metatranscriptomic data was aligned to a database that combined the Seq2fun pre-built algae, protists, alveolates, apicomplexans, stramenopiles, amoebozoa, and euglenozoa databases.

Metagenomic data processing

Metagenomic data was assembled and specifically analyzed for the presence of the MC synthetase gene cluster (*mcy*), responsible for microcystin production, in order to identify producers of MC. While the *mcy* gene cluster is most commonly studied in *Microcystis*, it is also present in other MC-producing cyanobacteria, such as *Planktothrix* and *Dolichospermum* (Christiansen et al., 2003; Li et al., 2016). The *mcy* gene cluster from the cyanobacterial type strain *Microcystis aeruginosa* PCC7806 was used as a reference due to its well-characterized genomic organization. Assemblies were generated using metaSPAdes (Nurk et al., 2017) for high-quality assembly of metagenomic sequences. The resulting contigs were binned using the automatic binning tool MaxBin2 (Wu et al., 2016), which allowed for the identification of distinct microbial genomes. Therefore, a bin represents a collection of contigs that likely originate from a single microbial genome. To assess the quality of the bins, CheckM (Parks et al., 2015) was employed to evaluate completeness, contamination, and to provide insight into the lineage of the bins. Completeness refers to the percentage of the target organism's genome that is present in the assembled bin, with higher completeness indicating a more accurate representation of the organism's genome. Contamination refers to the presence of sequences from non-target organisms within the bin, with lower contamination suggesting a more pure and accurate bin.

Statistical Analyses

The multivariate analysis Permutational Analysis of Variance (PERMANOVA) was used to test for significant differences in environmental parameters or community composition across lake conditions (number of permutations = 999). Principal Coordinates Analysis (PCoA) was used to visualize differences in samples based on environmental parameters and Spearman's correlation was used to calculate the correlation of each environmental parameter with the principal coordinate axes. Canonical Correspondence Analysis (CCA) was used to visualize differences in samples based on community composition and environmental parameters. All

statistical analyses were performed in R (Version 4.4.2). Spearman's correlation, Kruskal Wallis, and Mann Whitney tests were done using base R. Dunn's posthoc for Kruskal Wallis test was done using the FSA package (0.9.6). Multivariate analyses were performed using the vegan package (2.6-8).

RESULTS

Environmental parameters associated with lake conditions

There was a wide range of chl-a and total MC concentrations across the samples. Chl-a ranged from 12.68 – 736.74 µg/L, while total MC ranged from non-detect (< 0.15 µg/L) – 190 µg/L. The highest total MC concentrations were detected in 2021, which coincided with the highest ever observed MC levels in Clear Lake (160,378 µg/L) by the Big Valley Band of Pomo Indian's Monitoring program since 2014 (Big Valley Band of Pomo Indians, 2024). Detailed descriptions and distributions of environmental variables can be found in Florea et al. (2022). Each sampling date and site was categorized into one of four lake conditions according to the extracted chl-a and total MC concentrations. This resulted in 49 Above chl-a TMDL target & MC present samples, 3 Above chl-a TMDL target & MC absent samples, 56 Below chl-a TMDL target & MC present samples, and 22 Below chl-a TMDL target & MC absent samples (Figure 2).

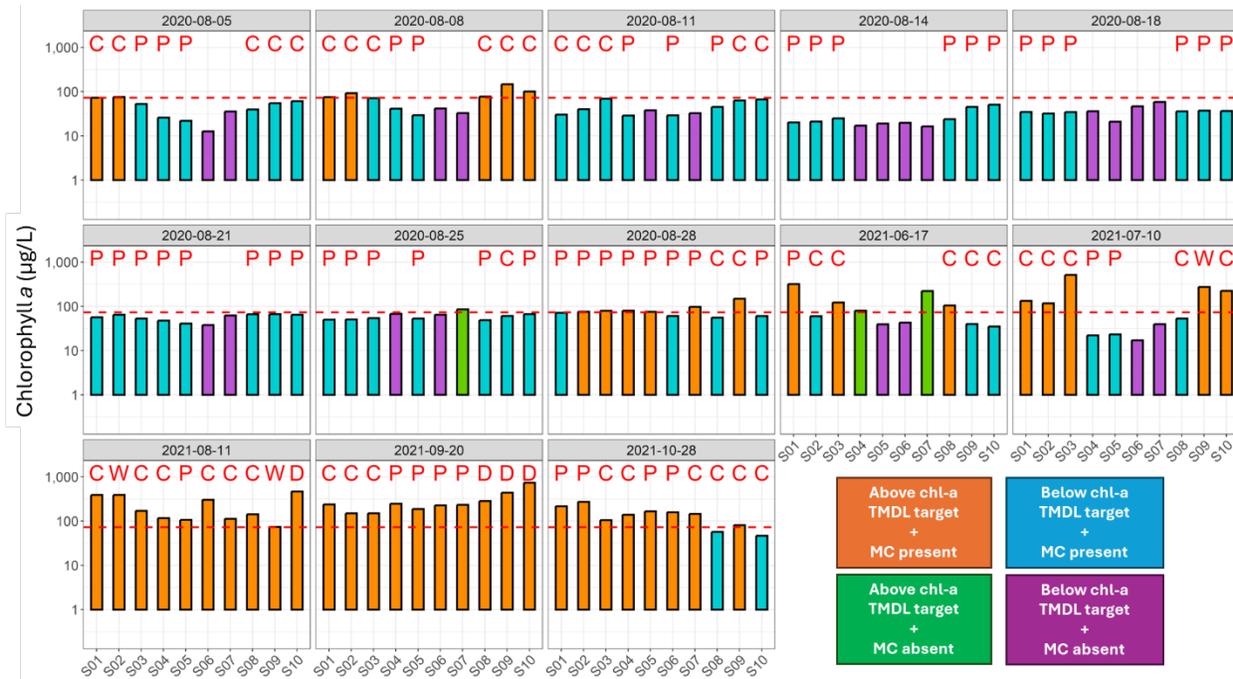


Figure 2. Division of samples into four lake conditions based on chlorophyll a and total microcystins concentrations. The target chl-a TMDL at Clear Lake is 73 µg/L while the presence of toxin is defined as total MC ≥ 0.15 µg/L. Bars indicate chl-a concentrations while alphabets above the bars indicate the total MC level: “P” = “Present” (≥ 0.15 and < 0.8 µg/L); “C” = “Caution” (≥ 0.8 and < 6 µg/L); “W” = “Warning” (≥ 6 and < 20 µg/L); “D” = “Danger” (≥ 20 µg/L). The lack of alphabet above the bar indicates no total MC was detected (< 0.15 µg/L). Red dashed line indicates a chl-a concentration of 73 µg/L. Color of the bars represent the lake condition.

Nutrients, particularly TN, were strongly associated with both chl-a and total MC concentrations (Table 4). Temperature was generally negatively correlated with chl-a concentrations, but also negatively correlated with total MC in the 2020 dataset. TN was strongly and positively correlated with chl-a and total MC concentrations in both years. TP was only positively correlated with chl-a in 2021, as well as having a negative correlation to total MC concentration in 2020.

Table 4. Spearman’s correlation results between chl-a and total MC, and measured physicochemical parameters including temperature, dissolved oxygen (DO), total nitrogen (TN), total phosphorus (TP), and the ratio of TN:TP. Asterisks indicate statistical significance (*: $p < 0.05$; **: $p < 0.01$; *: $p < 0.001$)**

Table 4A. Spearman’s correlation results between chl-a and total MC, and physicochemical parameters using the entire dataset (2020 + 2021).

	Chl-a	Total MC
Temperature	-0.38***	-0.10
DO	-0.34***	-0.07
TN	0.69***	0.66***
TP	0.28**	-0.12
TN:TP	-0.04	0.37***

Table 4B. Spearman’s correlation results between chl-a and total MC, and physicochemical parameters using data from 2020.

	Chl-a	Total MC
Temperature	-0.66***	-0.29*
DO	-0.20	0.16
TN	0.43***	0.66***
TP	-0.05	-0.65***
TN:TP	0.11	0.69***

Table 4C. Spearman’s correlation results between chl-a and total MC, and physicochemical parameters using data from 2021.

	Chl-a	Total MC
Temperature	-0.11	0.11
DO	-0.29	-0.06
TN	0.59***	0.44**
TP	0.39**	0.09
TN:TP	-0.31*	0.04

Multivariate analyses were performed to identify whether the entire suite of environmental parameters (temperature, DO, TN, TP, and TN:TP) were significantly different between lake conditions. Results from PERMANOVA indicated the suite of environmental parameters was significantly different between different lake conditions ($p < 0.01$). However, the partial R^2 value of the analysis was low (0.14), which indicated that only 11% of the variation between the lake conditions can be explained by the suite of environmental parameters. Differences in environmental parameters across lake conditions were visualized using PCoA and environmental parameters (TN, TP, and TN:TP) were overlaid onto the ordination plot as vectors to indicate their correlation with the principal coordinate (PC) axes. Temperature and DO were not included in the PCoA as there were missing data due to sonde failure, and the removal of samples with missing data led to non-significant differences between lake conditions (PERMANOVA; $p > 0.05$). The lack of statistical significance when the dataset was reduced, as compared to the complete dataset showing statistical significance, may have been due to the reduced sample size and lower statistical power. PCoA with all (2020 and 2021) data included showed a wide spread of the Above chl-a TMDL target & MC present (orange circles) and Below chl-a TMDL target & MC present (blue circles) samples across both PC1 (x-axis) and PC2 (y-axis; Figure 3). This indicated that samples in these two lake conditions exhibited a broad range of TN, TP, and TN:TP values. Below chl-a TMDL target & MC absent samples (purple circles) tended to cluster on the right side of the plot, suggesting an association with higher TP, which was strongly correlated with PC1 ($\rho = 0.93$; $p < 0.001$). When analyzing data separated by year, results from 2020 were similar to the full dataset (Figure 4), with Above chl-a TMDL target & MC present and Below chl-a TMDL target & MC present samples widely spread across both axes. Below chl-a TMDL target & MC absent samples showed stronger clustering on the right side of the plot, aligning with higher TP values. In contrast, results from 2021 showed a different pattern (Figure 5). Above chl-a TMDL target & MC absent (green squares) and Below

chl-a TMDL target & MC absent (purple circles) clustered on the left side of the plot, associated with higher TN:TP values. Meanwhile, Above chl-a TMDL target & MC present samples tended to cluster on the right side, aligning with both higher TN and TP values.

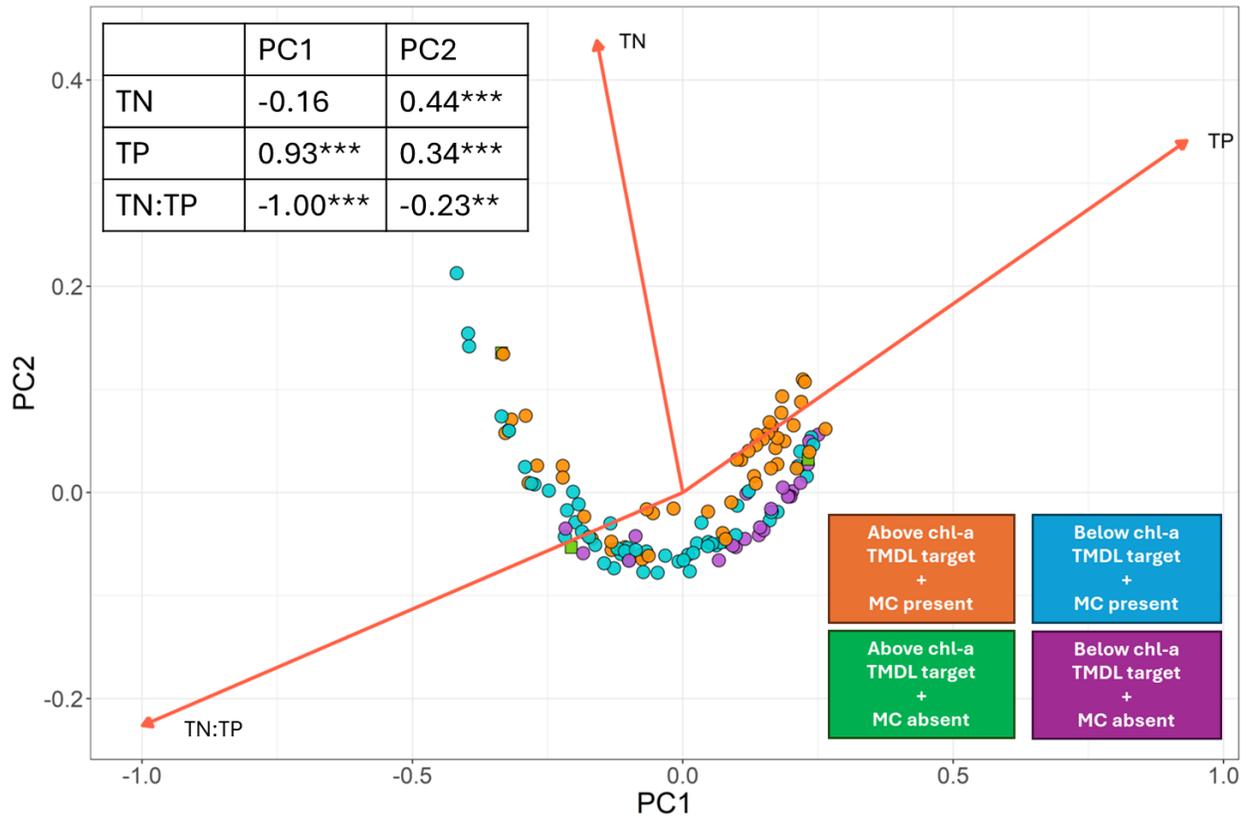


Figure 3. Principal coordinates analysis of environmental parameters (TN, TP, and TN:TP) of data from 2020 and 2021. Samples (points) are colored by lake conditions. ‘Above chl-a TMDL target & MC absent’ samples (green) are in the shape of squares for easy differentiation with the blue ‘Below chl-a TMDL target & MC present’ samples. Vectors represent environmental variables correlated with the ordination, with arrow direction indicating the gradient and arrow length representing the strength of correlation. The embedded table shows the correlation coefficient and *p* value for each environmental parameter in relation to the principal coordinates (PC).

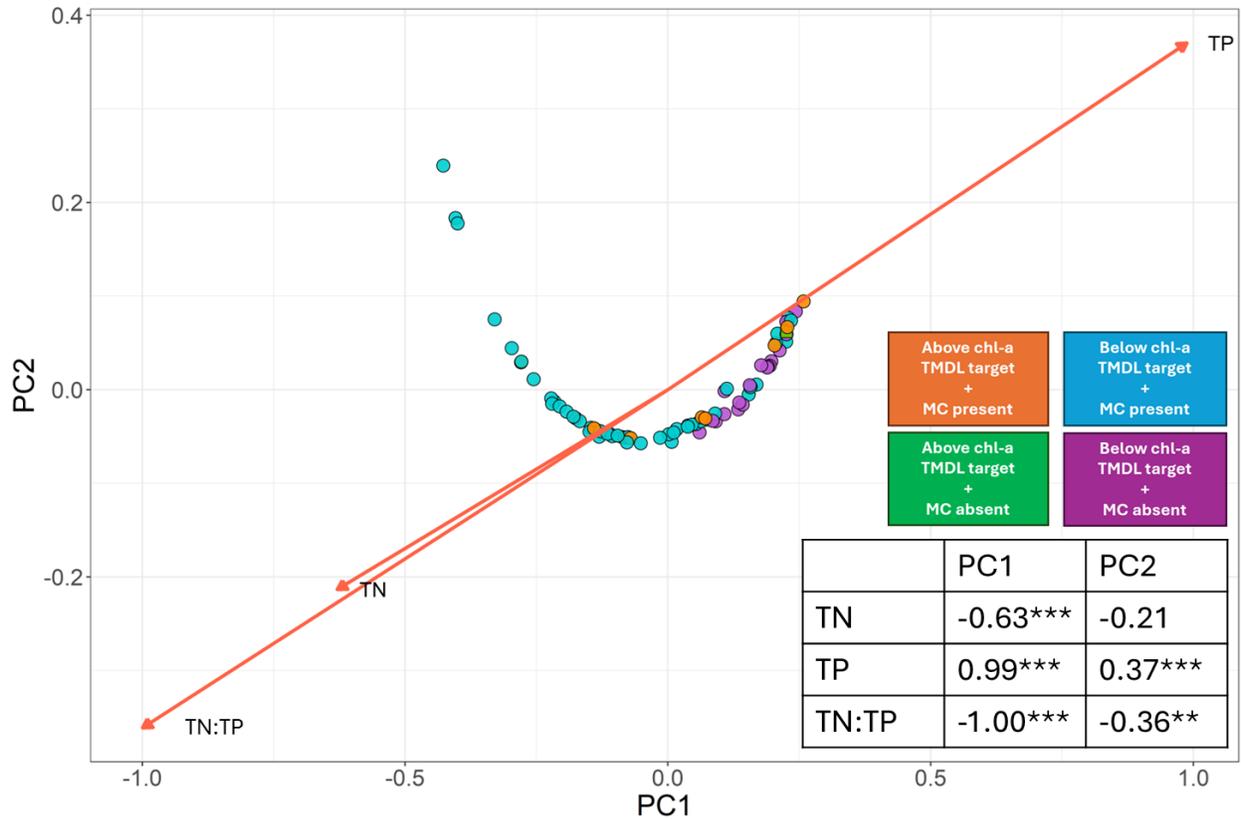


Figure 4. Principal coordinates analysis of environmental parameters (TN, TP, and TN:TP) of data from 2020. Samples (points) are colored by lake conditions. ‘Above chl-a TMDL target & MC absent’ samples (green) are in the shape of squares for easy differentiation with the blue ‘Below chl-a TMDL target & MC present’ samples. Vectors represent environmental variables correlated with the ordination, with arrow direction indicating the gradient and arrow length representing the strength of correlation. The embedded table shows the correlation coefficient and *p* value for each environmental parameter in relation to the principal coordinates (PC).

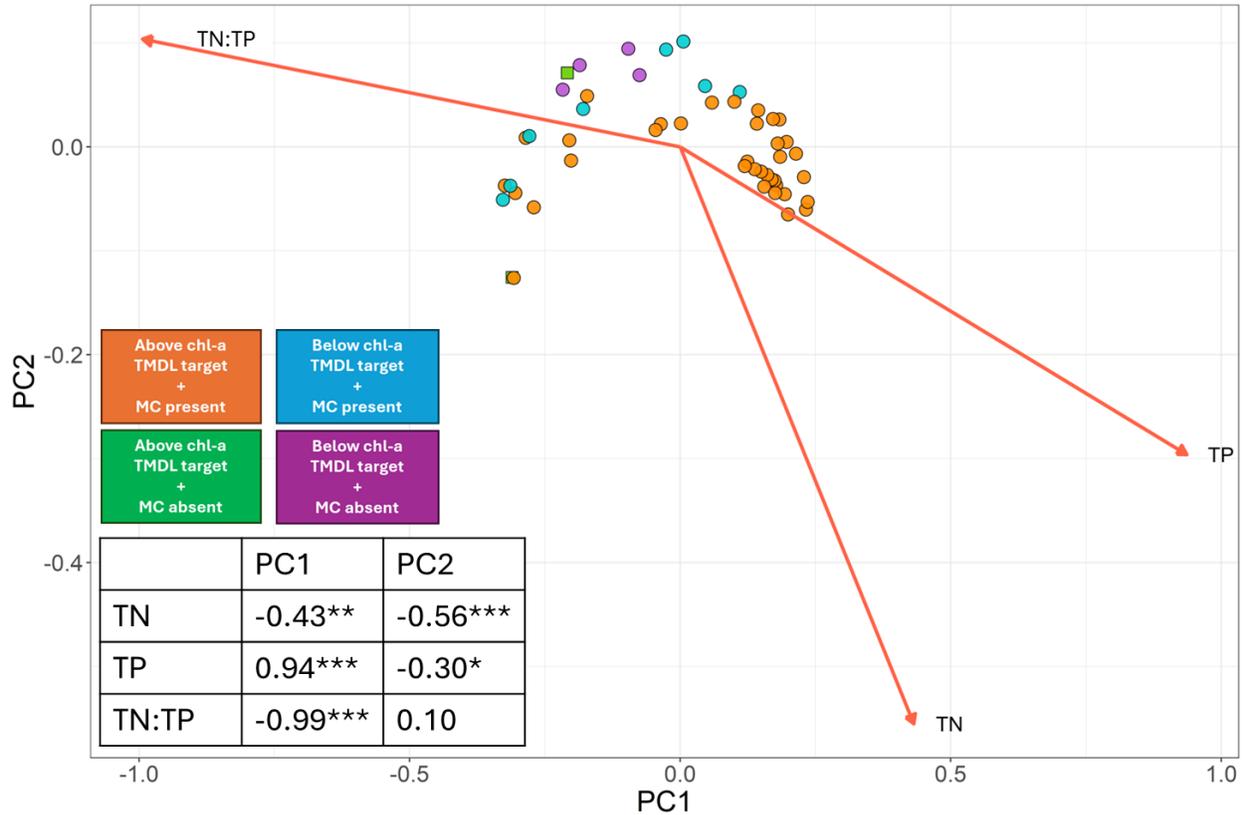


Figure 5. Principal coordinates analysis of environmental parameters (TN, TP, and TN:TP) of data from 2021. Samples (points) are colored by lake conditions. ‘Above chl-a TMDL target & MC absent’ samples (green) are in the shape of squares for easy differentiation with the blue ‘Below chl-a TMDL target & MC present’ samples. Vectors represent environmental variables correlated with the ordination, with arrow direction indicating the gradient and arrow length representing the strength of correlation. The embedded table shows the correlation coefficient and p value for each environmental parameter in relation to the principal coordinates (PC).

Prokaryotic community composition associated with lake conditions

The composition and diversity of the prokaryotic community, which includes both cyanobacteria and other types of bacteria, were significantly different between different lake conditions. PERMANOVA based on the percent contribution of ASVs for each sample indicated there was a significant difference between lake conditions ($p < 0.001$), and Shannon’s diversity indices calculated with ASV abundances were significantly higher in Below chl-a TMDL target & MC absent samples than Above chl-a TMDL target & MC present and Below chl-a TMDL target & MC present samples (Kruskal-Wallis with Dunn’s post hoc test; $p < 0.05$; Figure 6A).

Shannon’s diversity index measures both the richness (number of species) and the evenness (distribution of individuals among species) in a community. Higher values indicate greater

diversity, where both species richness and evenness are higher, while lower values suggest less diversity, with fewer species or uneven distribution among species. In general, Shannon's diversity index values above 3 are often considered as an indication of high diversity, while values below 1 suggest low diversity (Magurran, 2004). However, abundances based on ASVs may inflate the index value due to the higher number of taxa represented by ASVs. Nevertheless, Shannon's diversity index is commonly used for estimating the diversity of metabarcoding data. In this study, the diversity of the prokaryotic community was significantly lower in the presence of MC (Mann-Whitney test; $p < 0.001$; Figure 6B), but not significantly different whether chl-a concentrations were above or below the TMDL target (Mann-Whitney test; $p > 0.05$; Figure 6C).

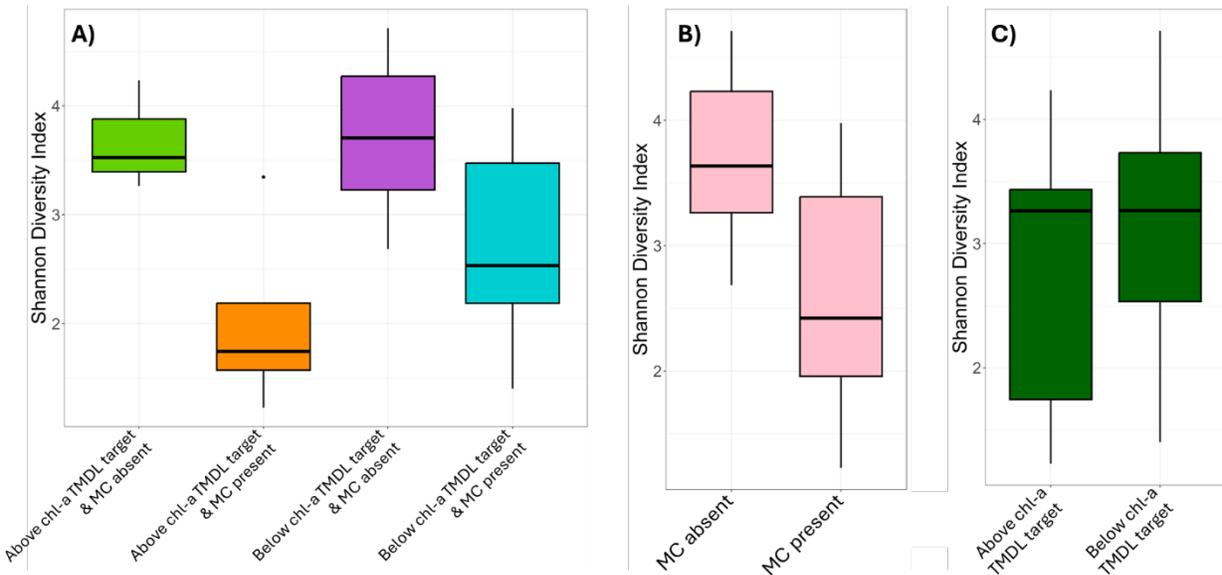


Figure 6. Shannon diversity index of the prokaryotic community across A) different lake conditions, B) in the presence or absence of MC ($\geq 0.15 \mu\text{g/L}$), and C) whether chl-a was above or below the TMDL target ($\geq 73 \mu\text{g/L}$ chl-a).

The prokaryotic community was comprised of varying proportions of cyanobacteria across samples and lake conditions (Figure 7). Reads identified as cyanobacteria were present in all samples, and the percentage ranged from 0.5 – 73%, with a mean of 23%. There was no significant difference in percentage of cyanobacteria reads between different lake conditions (Kruskal-Wallis test; $p > 0.05$), nor the presence/absence of a bloom (Figure 8) or toxins (Mann-Whitney test; $p > 0.05$; Figure 9). The relative abundance of cyanobacteria reads was also tested between the three arms of Clear Lake, but no significant difference was found (Kruskal-Wallis $p > 0.05$). Surprisingly, the sample from 2021 Sep 20 at S10 was highly dominated by non-cyanobacterial microbial taxa and only had 0.7% of reads classified as cyanobacteria

despite high levels of total MC (right most bar in Figure 9). A closer look at the non-cyanobacterial community in this sample indicated that the majority of reads (94%) were from the bacteria phylum Pseudomonadota (formerly known as Proteobacteria), with 75% of the reads as an ASV identified as a gammaproteobacteria from the family Sutterellaceae.

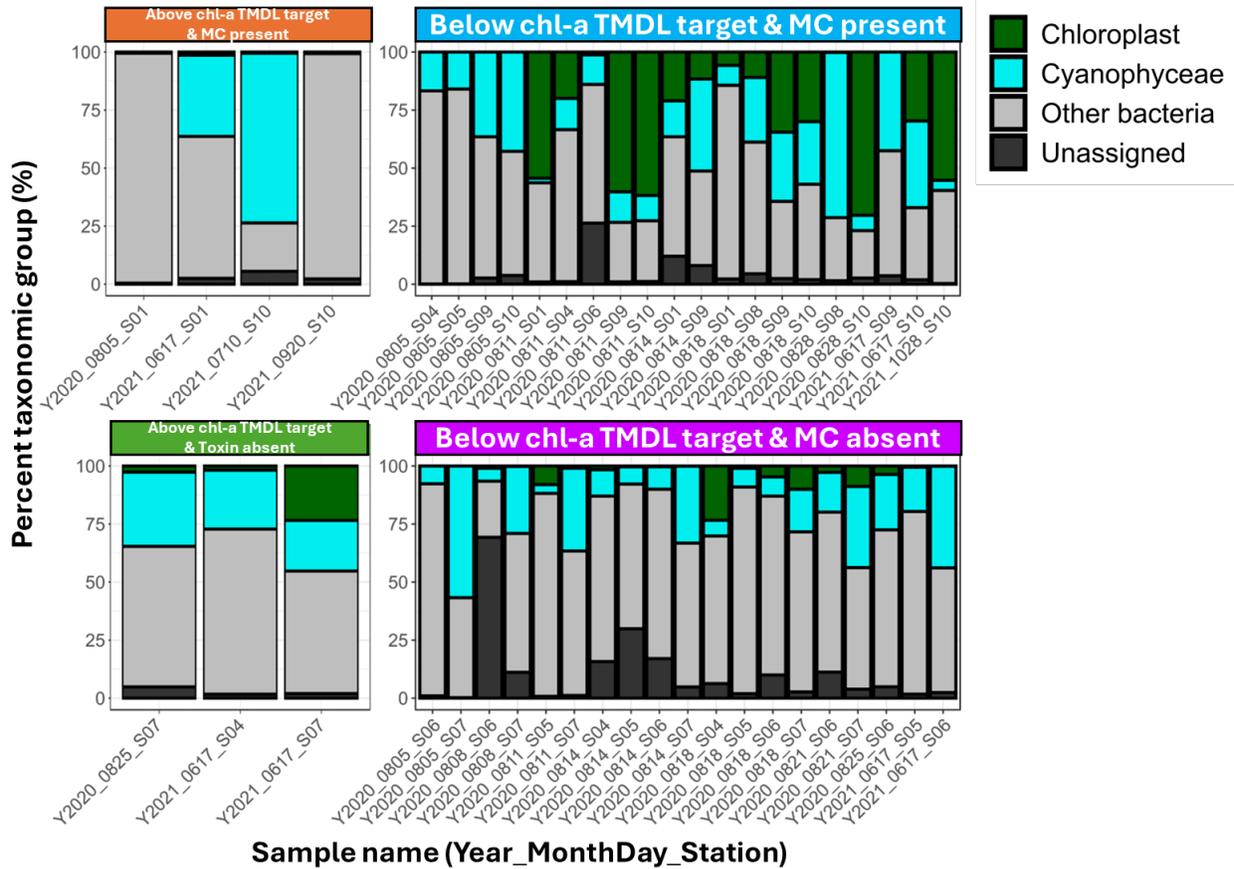


Figure 7. Composition of the prokaryotic community (based on ASV reads) across different lake conditions. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

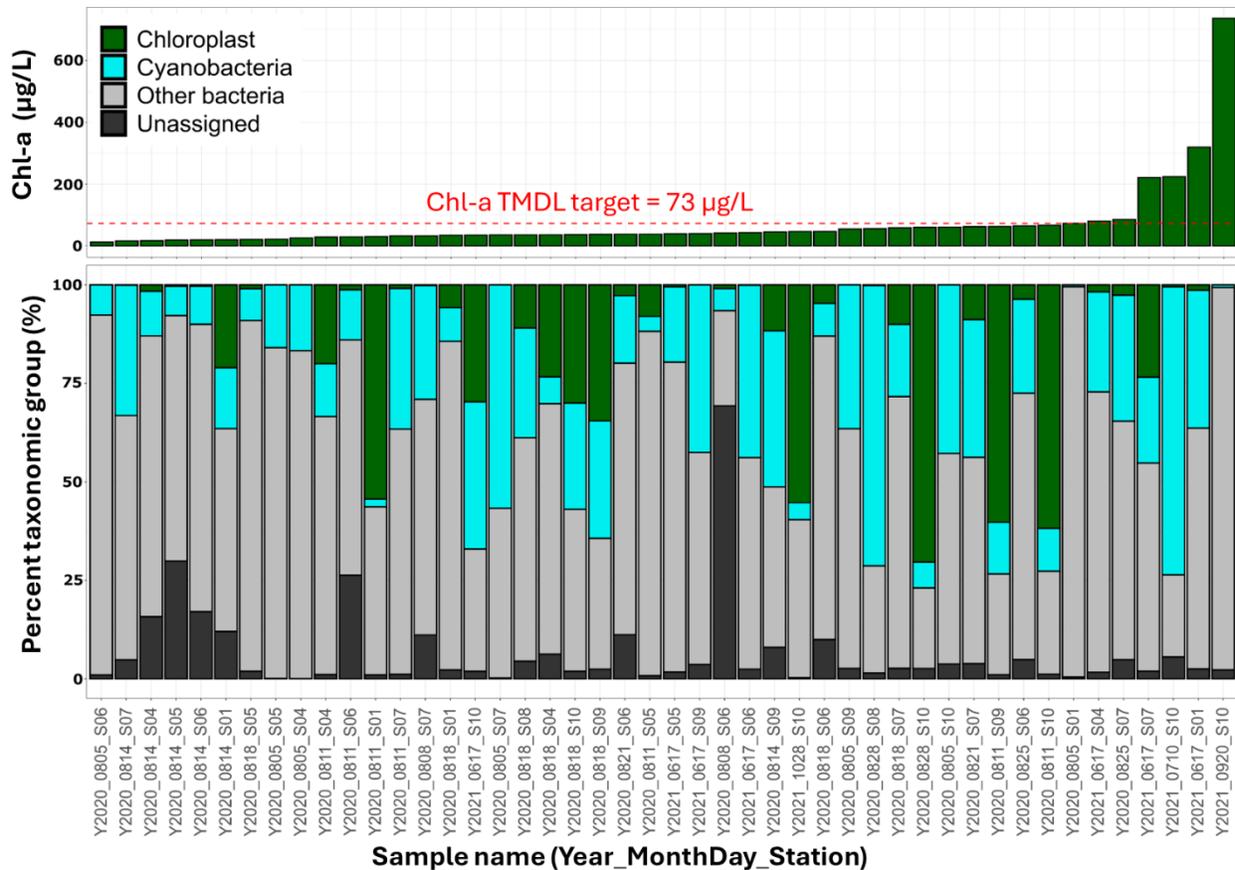


Figure 8. Composition of the prokaryotic community (bottom panel) displayed for each sample, arranged in ascending order of chl-a concentrations (top panel; i.e. samples on the left have lowest chl-a and samples on the right have highest chl-a). Red dashed line in the top panel indicates the chl-a TMDL target value of 73 µg/L, the threshold for bloom categorization in this study. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

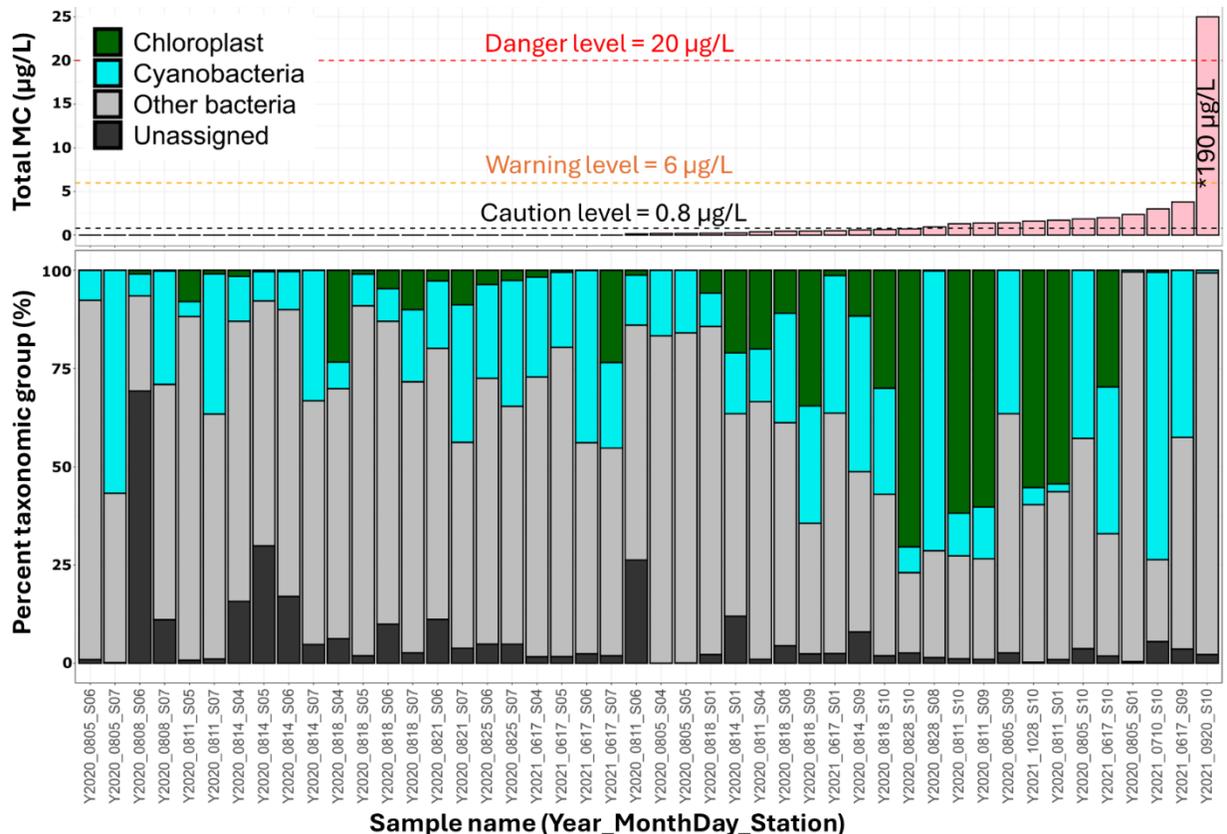


Figure 9. Composition of the prokaryotic community (bottom panel) displayed for each sample, arranged in ascending order of total MC concentrations (top panel; i.e. samples on the left have lowest MC and samples on the right have highest MC). Dashed lines in the top panel indicate various levels of the California Cyanobacteria and Harmful Algal Bloom Network (CCHAB) microcystins trigger level guidelines. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

The composition of cyanobacterial taxa differed significantly between lake conditions (PERMANOVA; $p < 0.001$; Figure 10). The cyanobacteria community was mainly dominated by *Limnoraphis* (formerly *Lyngbya*), *Dolichospermum*, *Aphanizomenon*, and *Vulcanococcus*. These genera align with the results from past reports as most of the samples sequenced for 16S metabarcoding were from 2020, when *Limnoraphis* was observed as the dominant species through microscopy observations (Figure 4 in Florea et al., 2022; Figure 5 in Kalra et al., accepted; labels as *Lyngbya*) and DNA metabarcoding (Figure 8 in Kalra et al., accepted). The results also revealed that the contribution of these dominant genera differed significantly between lake conditions (Figure 11). The percentage of *Limnoraphis* reads within the cyanobacteria community was significantly higher in samples where MC was detected, regardless of chl-a concentrations (Kruskal-Wallis and Dunn’s post hoc test; adjusted $p < 0.05$). On the other hand, the percentages of *Aphanizomenon* in Above chl-a TMDL target & MC

absent and Below chl-a TMDL target & MC absent samples were significantly higher than Below chl-a TMDL target & MC present samples. Similarly, the percentage of *Dolichospermum* in Below chl-a TMDL target & MC absent samples were significantly higher than Above chl-a TMDL target & MC present and Below chl-a TMDL target & MC present samples. *Aphanizomenon* was largely absent from samples with detectable MC, comprising less than 5% of those communities. *Dolichospermum* accounted for a notable portion (24 – 55%) of the community in three samples where MC was present, and was otherwise largely absent in samples with MC. *Vulcanococcus*, which is closely related to *Synechococcus*, also had higher contributions in Below chl-a TMDL target & MC absent samples than Below chl-a TMDL target & MC present samples. The contributions of all of these four dominant genera strongly differed between samples with or without MC (Mann-Whitney test; $p < 0.001$), but only the contribution *Dolichospermum* was distinct between samples identified as above and below the chl-a TMDL target (Mann-Whitney test; $p < 0.05$). There is no significant difference in the relative abundance of *Microcystis* between lake conditions, the presence/absence of MC, and chl-a levels above/below the TMDL target.

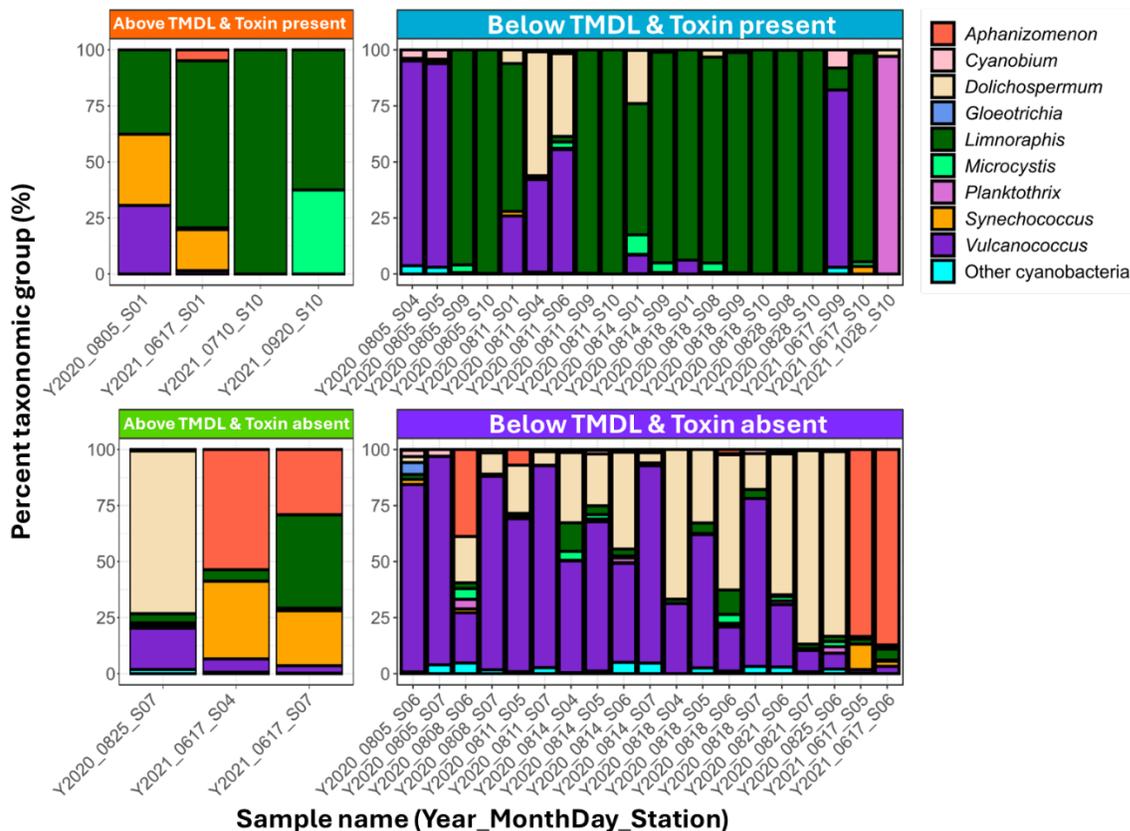


Figure 10. Composition of the cyanobacteria community (based on ASV reads) across different lake conditions. Reads for other bacteria or chloroplasts were excluded. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

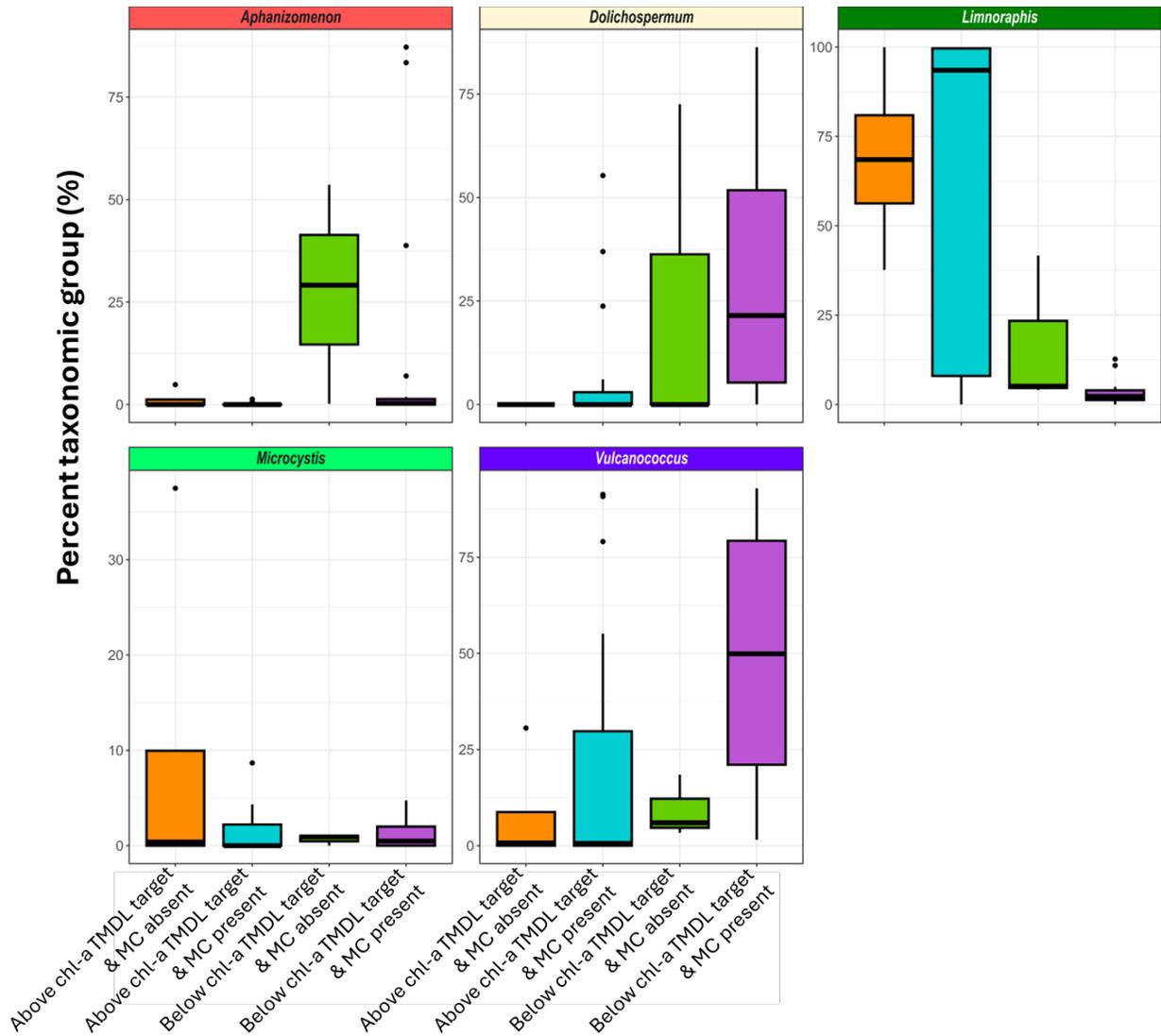


Figure 11. Percentage of reads contribution identified as several cyanobacteria taxa across different lake conditions.

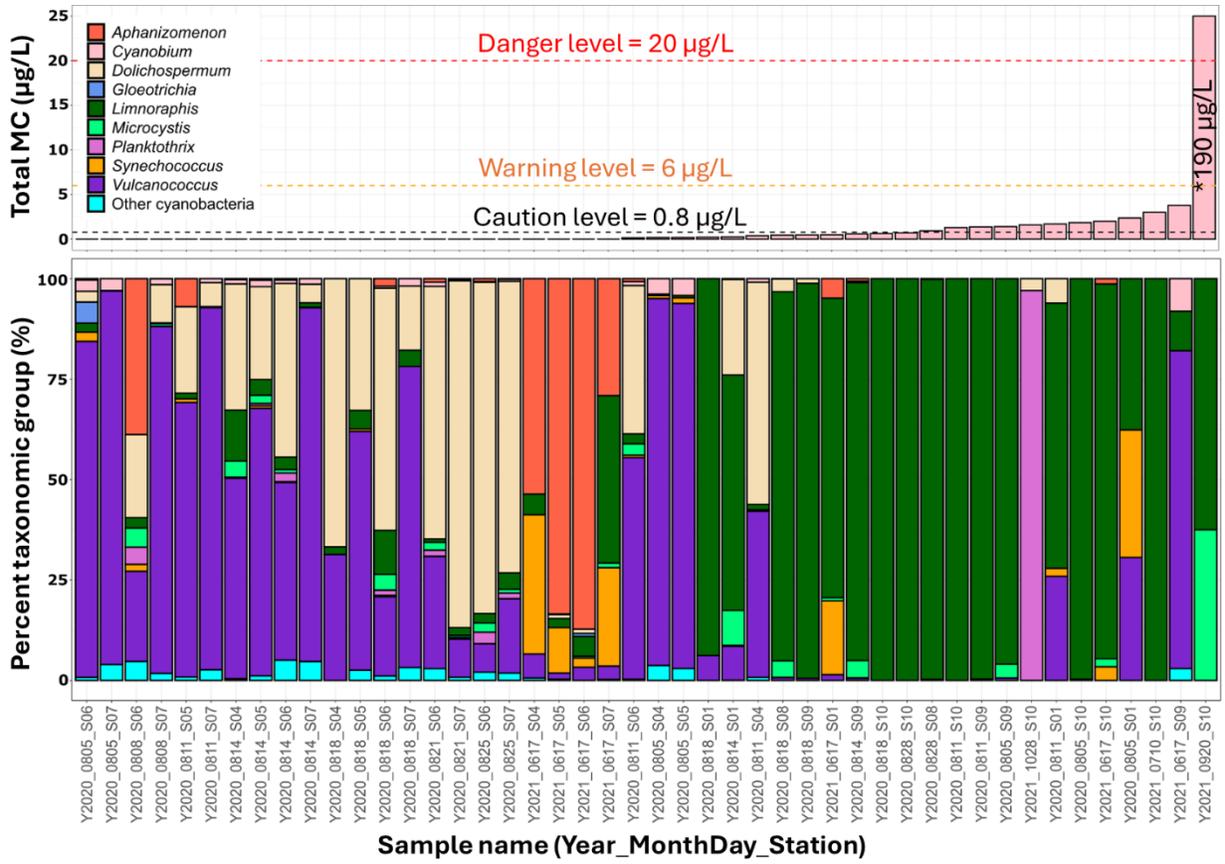


Figure 12. Composition of the cyanobacteria community (bottom panel) displayed for each sample, arranged in ascending order of total microcystins (MC) concentrations (top panel; i.e. samples on the left have lowest MC and samples on the right have highest MC). Reads for other bacteria or chloroplasts were excluded. Dashed lines in the top panel indicate various levels of the California Cyanobacteria and Harmful Algal Bloom Network (CCHAB) microcystins trigger level guidelines. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

Eukaryotic community composition associated with lake conditions

The composition and diversity of the eukaryotic community were also significantly different between different lake conditions. PERMANOVA based on the percent contribution of ASVs for each sample showed significant differences across lake conditions ($p < 0.001$). Shannon's diversity indices calculated with ASV abundances indicated significant differences across lake conditions as well, but in contrast to the prokaryotic community, the eukaryotic community had significant higher diversity in Above chl-a TMDL target & MC present samples than Below chl-a TMDL target & MC present and Below chl-a TMDL target & MC absent samples (Figure 13). While the prokaryotic community had significantly higher diversity in samples with no MC

present and did not differ significantly in terms of chl-a levels, the eukaryotic community had significantly higher diversity in the presence of chl-a levels above the TMDL target (Figure 13C), and did not differ significantly in terms of the presence of toxins (Figure 13B). A plot of the eukaryotic community Shannon's diversity index sorted by chl-a concentration in this study (Figure 14) indicated a plateau or even a slight decrease in diversity at \sim chl-a > 200 $\mu\text{g/L}$, but Spearman's correlation indicated significant correlation between the diversity index and chl-a concentrations ($\rho = 0.59$, $p < 0.001$).

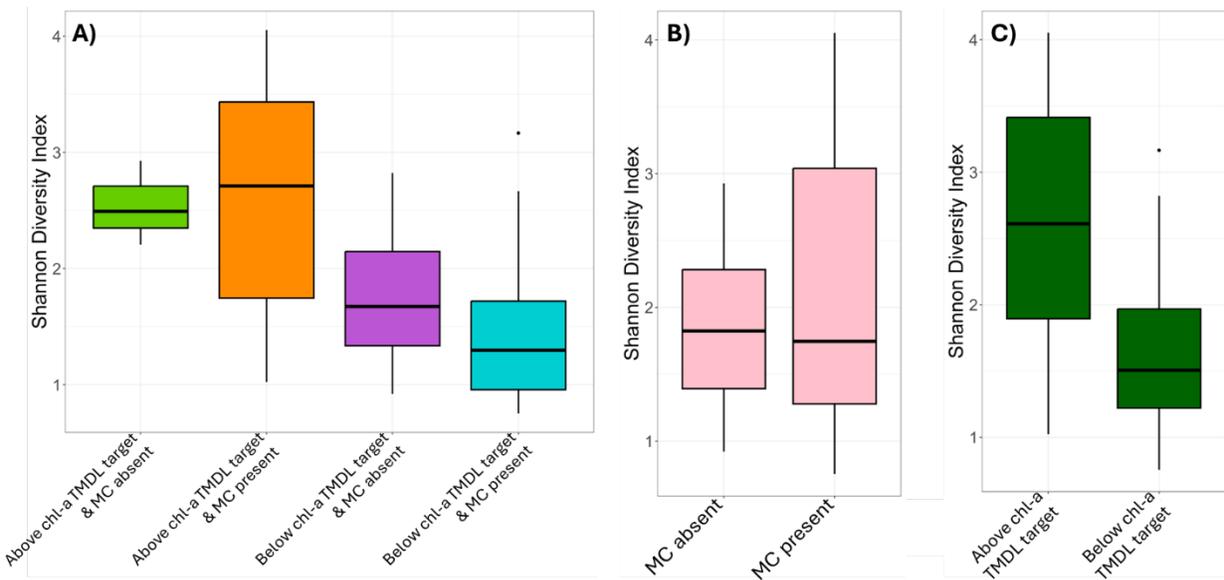


Figure 13. Shannon diversity index of the eukaryotic community across A) different lake conditions, B) in the presence or absence of MC ($\geq 0.15 \mu\text{g/L}$), and C), or whether chl-a concentrations were above/below the TMDL target (73 $\mu\text{g/L}$ chl-a).

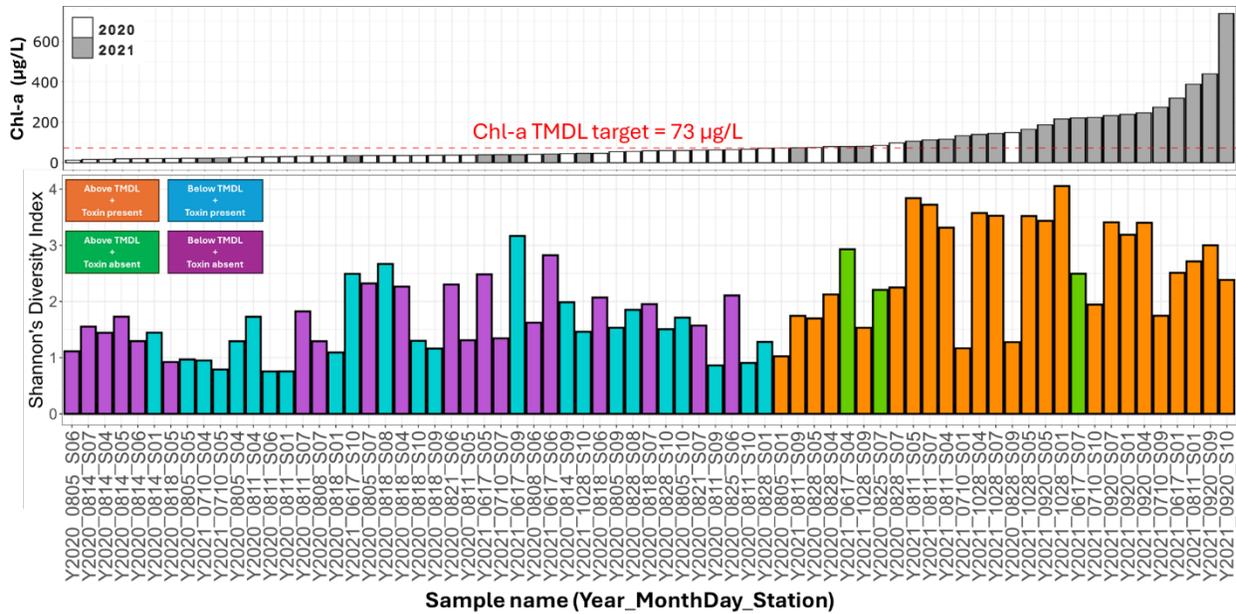


Figure 14. Shannon diversity index of the eukaryotic community displayed for each sample, arranged in ascending order of chl-a concentrations (top panel; i.e. samples on the left have lowest chl-a and samples on the right have highest chl-a). Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

The eukaryotic community was mostly dominated by chlorophytes (chlorophyta) and diatoms (bacillariophyta) (Figure 14). A list of eukaryotic genera with notable relative abundance ($\geq 5\%$) are listed in Table S2. There was a higher percentage of diatoms in Below chl-a TMDL target & MC present samples that coincided with high percentage of chloroplast reads in the prokaryotic community (Figure 7). Cryptophytes (cryptophyta) also contributed a notable portion of the eukaryotic community, especially in samples with no MC present. Members of parasitic groups including Apicomplexa and Perkinsea also took up notable portions in some samples (a max of 10%). Apicomplexa is a large phylum of mainly parasites that include notable human disease parasites such as Plasmodium (causing malaria) and Toxoplasma (causing toxoplasmosis). Perkinsea is a group of intracellular parasites with a broad range of hosts such as bivalves and fish.

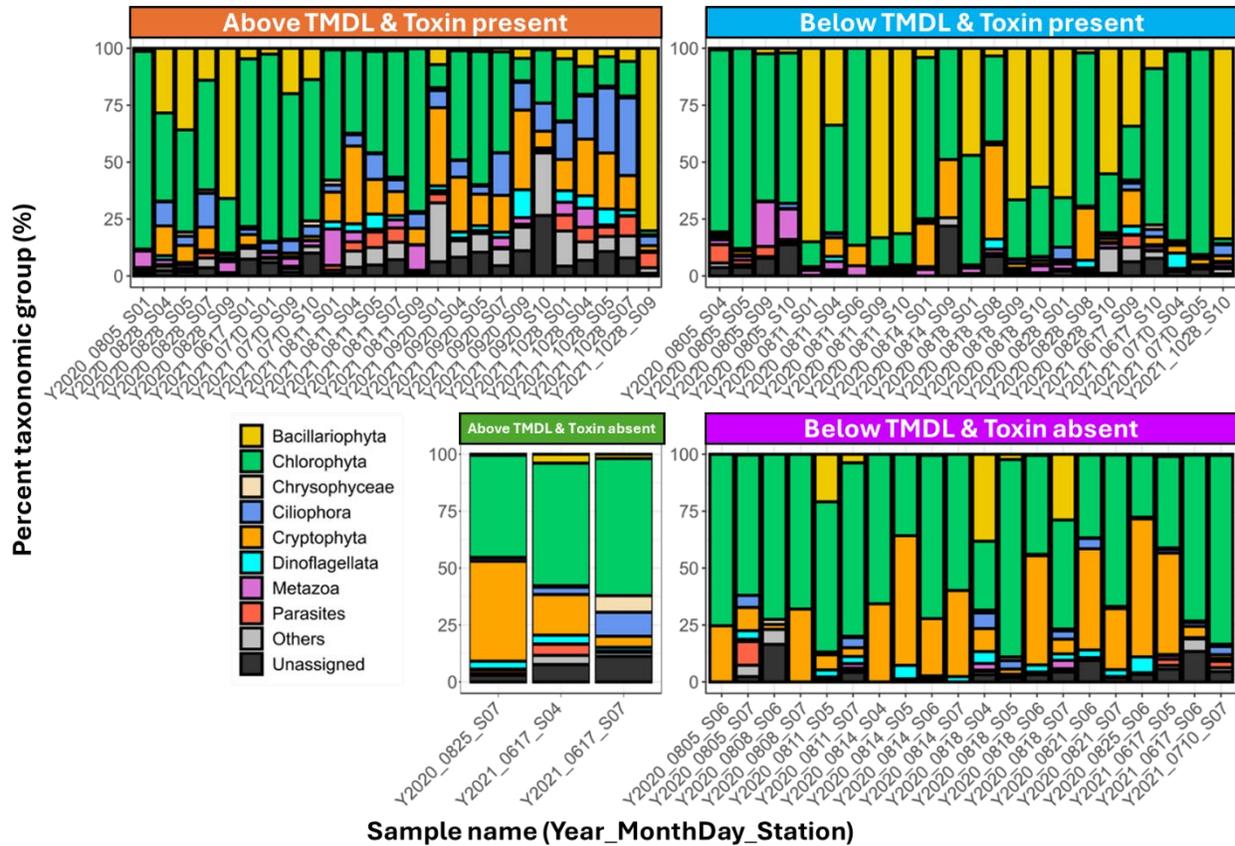


Figure 15. Composition of the eukaryotic community (based on ASV reads) across different lake conditions. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

Arrangement of samples by MC concentrations provided an even clearer visualization of the presence of diatoms in samples with MC present and the presence of cryptophytes in samples with no MC present (Figure 15). Indeed, the relative abundance of diatoms was significantly higher in samples with MC present, while the relative abundance of cryptophytes was significantly higher in samples without MC (Mann-Whitney test, $p < 0.01$). The percentages of both diatom and cryptophyte reads were also significantly correlated to total MC concentration (diatom: $\rho = 0.47$, $p < 0.001$; cryptophyte: $\rho = -0.36$, $p < 0.01$).

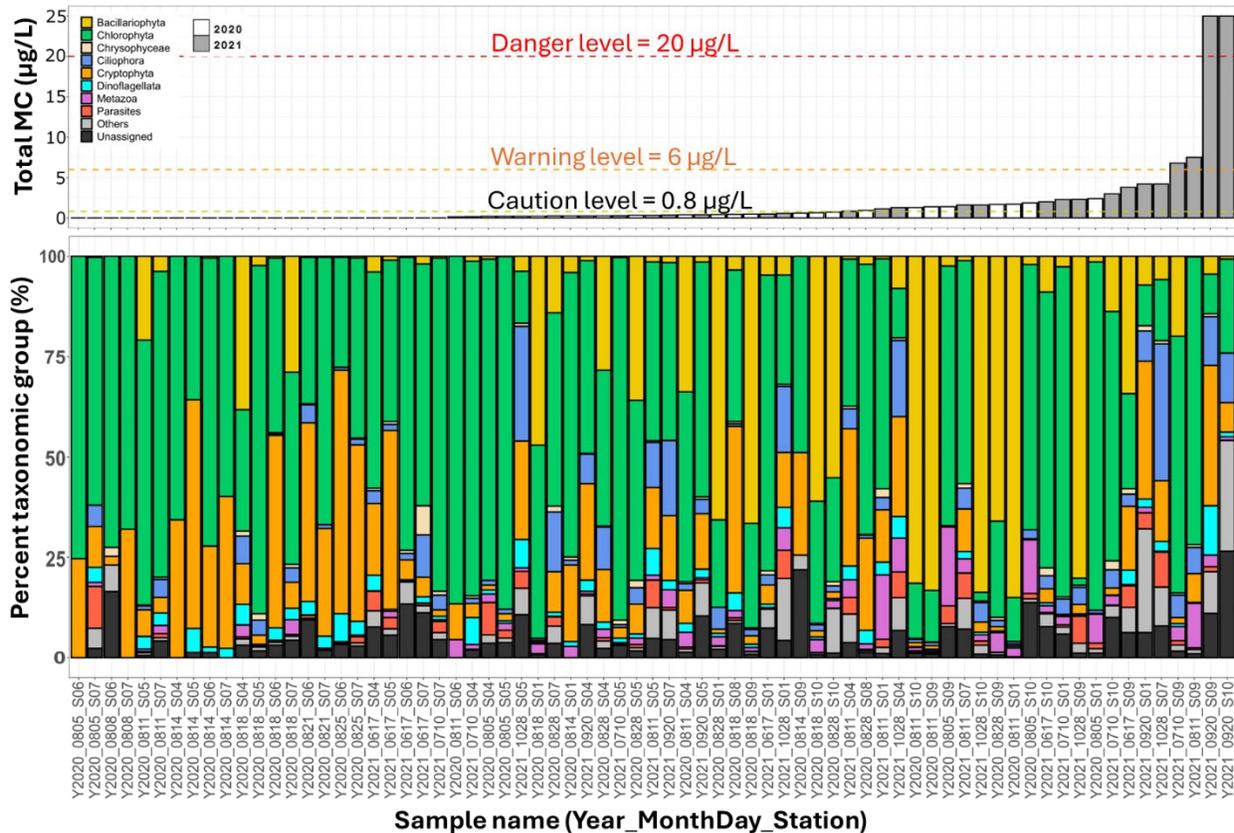


Figure 16. Composition of the eukaryotic community (bottom panel) displayed for each sample, arranged in ascending order of total microcystins (MC) concentrations (top panel; i.e. samples on the left have lowest MC and samples on the right have highest MC). Dashed lines in the top panel indicate various levels of the California Cyanobacteria and Harmful Algal Bloom Network (CCHAB) microcystins trigger level guidelines. Colors of the bars in the top panel indicate the year of the sample. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

The higher diversity of the eukaryotic community observed at higher chl-a concentrations included a variety of heterotrophic protists such as ciliates (blue bars), katablepharid (Placed in ‘Others’, 21 % in sample from 2021 Sep 20 at S01; 6th bar from the right in Figure 16), and cercozoans (Placed in ‘Others’, 25% in sample from 2021 Sep 20 at S10; 1st bar from the right in Figure 16). Therefore, the higher diversity observed in Lake conditions may be due to the higher availability of prey (including bacteria and algal cells). A portion of reads (0 - 20% per sample) were identified as ‘Metazoa’, but it should be noted that the reference database used for taxonomic assignment is PR2, a database focused on protists. Therefore, it is possible that a number of ‘Unassigned’ reads could map to metazoan grazers. Metazoans identified in this study mainly consisted of rotifers and *Daphnia*, which are bacterial or algal cell grazers. Large metazoan consumers such as copepods and cladocerans (e.g. *Daphnia*) have been documented

to have positive associations with cyanobacterial (including *Microcystis*) blooms (Sun et al., 2012).

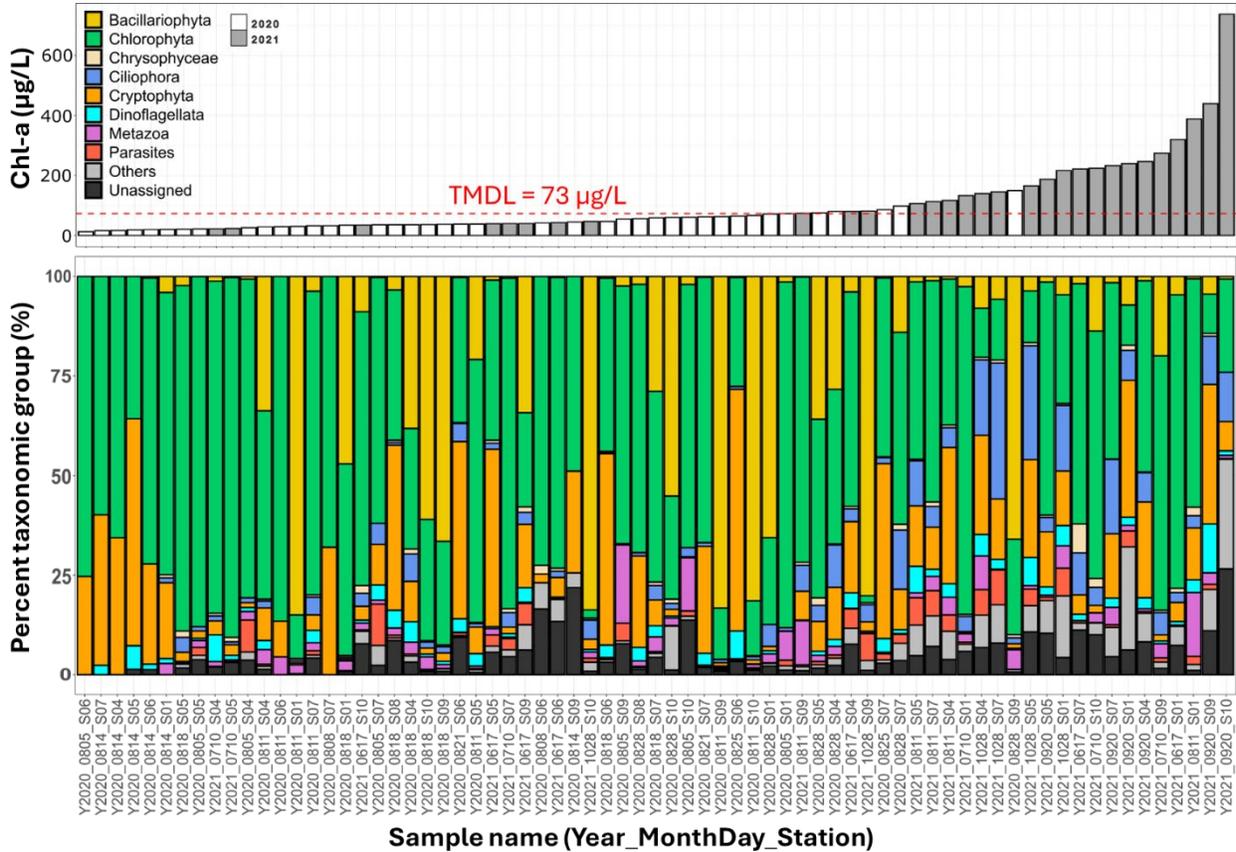


Figure 17. Composition of the eukaryotic community (bottom panel) displayed for each sample, arranged in ascending order of chl-a concentrations (top panel; i.e. samples on the left have lowest chl-a and samples on the right have highest chl-a). Red dashed line in the top panel indicates the TMDL value of 73 µg/L, the threshold for bloom categorization in this study. Colors of the bars in the top panel indicates the year of the sample. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

Environmental drivers of prokaryotic and eukaryotic communities

CCA revealed the full prokaryotic community generally clustered based on lake conditions (Figure 17). Below chl-a TMDL target & MC absent samples, in particular, mostly clustered at the top of the plot in the direction of TP and Temperature, indicating the community in these samples (based on ASV composition) were similar and tended to be associated with higher TP and temperature (Top right corner of the plot). Below chl-a TMDL target & MC present samples

tended to be spread along on the opposite end of the CCA 2 axis (y-axis) in the direction of higher TN:TP, TN, and DO. Two Below chl-a TMDL target & MC present samples from 2020 Aug 05 Lower Arm (S09 and S10) had distinct communities from the rest of the samples, and had positive associations with total MC, chl-a, TN, and TN:TP concentrations. The sample with highest total MC and chl-a concentrations was completely separated from the rest of the samples at the far-left portion of the plot. PERMANOVA analysis of the CCA results indicated that environmental parameters overall explained the community variability ($p < 0.001$), but only total MC ($p < 0.001$), chl-a ($p < 0.05$), and TP ($p < 0.001$) were significant drivers of community variability.

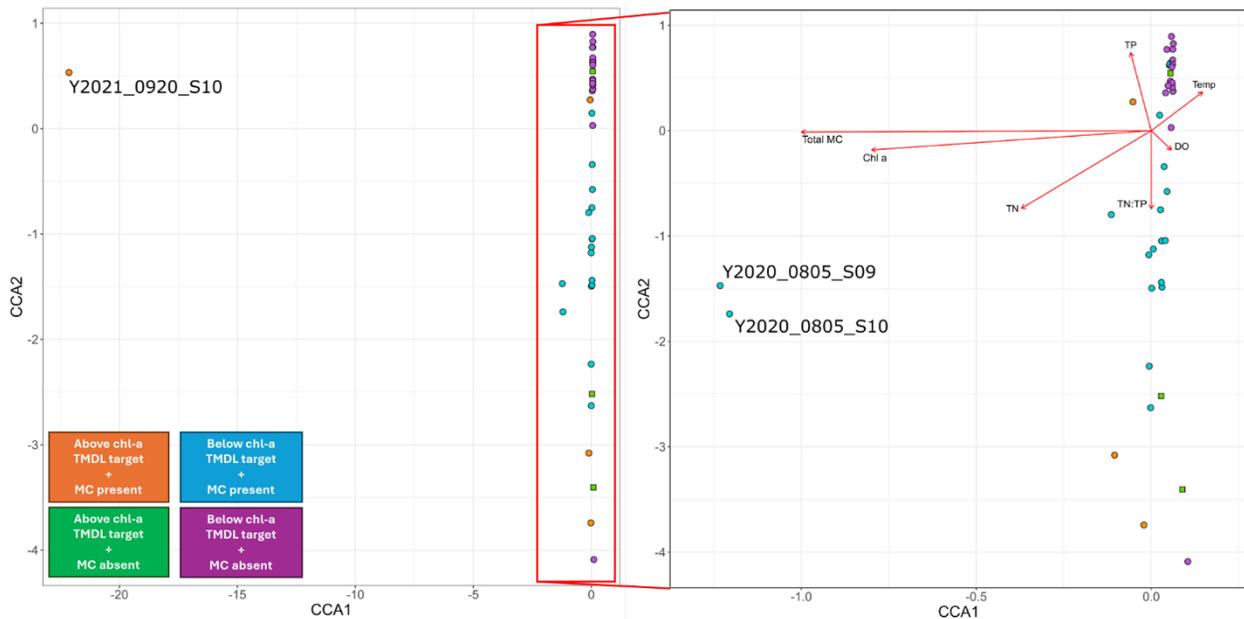


Figure 18. Canonical correlation analysis of the prokaryotic community and environmental parameters (temperature, DO, total MC, chl-a, TN, TP, TN:TP). Samples (points) are colored by lake conditions. Green ‘Above chl-a TMDL target & MC absent’ samples have square shapes to distinguish them from the blue ‘Below chl-a TMDL target & MC absent’ samples. Right panel is a zoomed in portion of the portion framed by the red square in the left panel. Some sample names are indicated for outlier data points. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

The cyanobacteria community also generally clustered based on lake conditions (Figure 18), especially for Below chl-a TMDL target & MC absent samples that were also positively associated with TP (Right side of Figure 18). Communities with MC present tended to cluster towards the left side of the plot, with positive association to TN, TN:TP, and total MC. PERMANOVA analysis of the CCA results indicated that environmental parameters overall

explained the community variability ($p < 0.05$), but only TP was a significant driver of variability in the cyanobacteria community ($p < 0.001$).

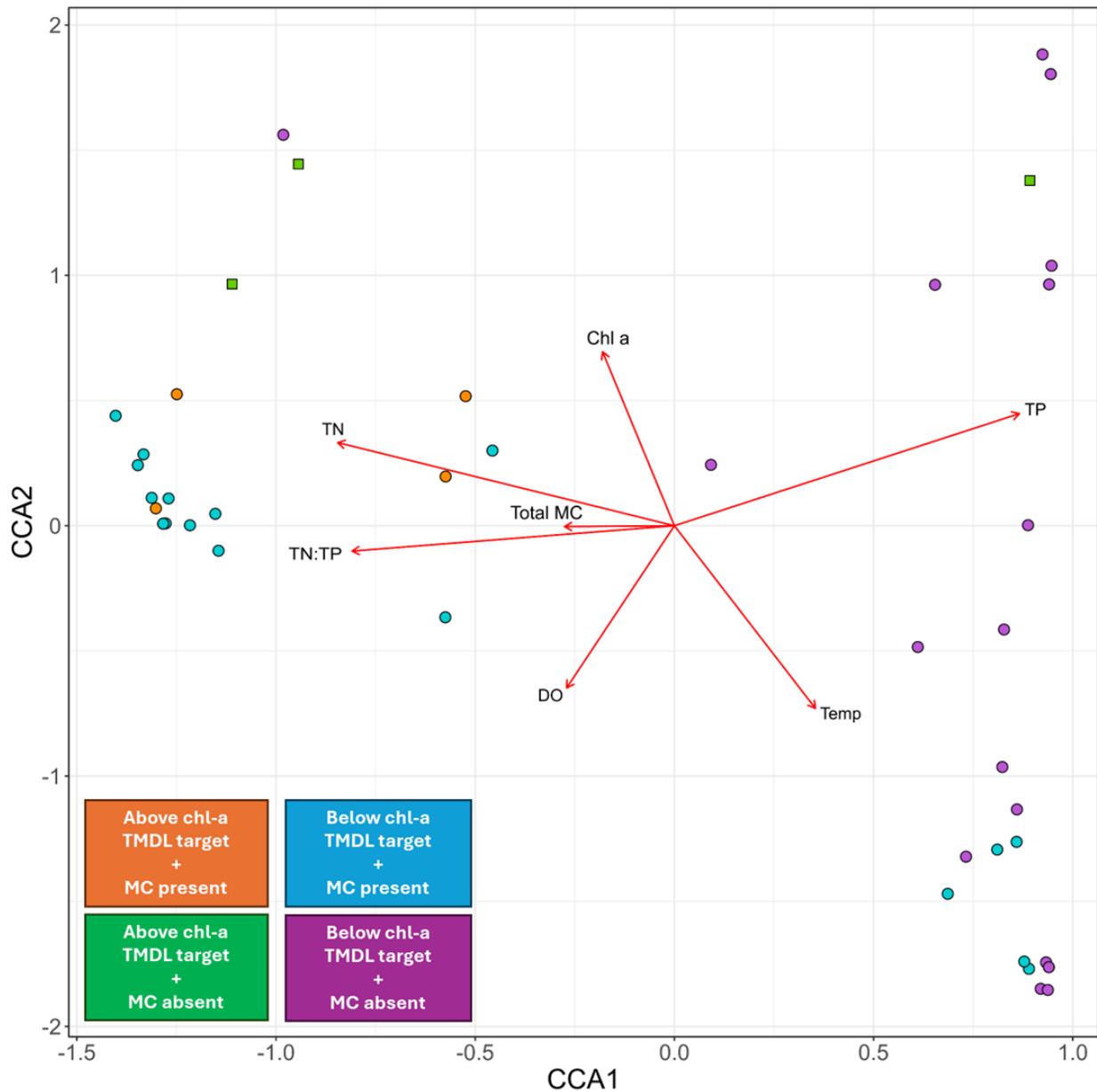


Figure 19. Canonical correlation analysis of the cyanobacterial community and environmental parameters (temperature, DO, total MC, chl-a, TN, TP, TN:TP). Samples (points) are colored by lake conditions. Green ‘Above chl-a TMDL target & MC absent’ samples have square shapes to distinguish them from the blue ‘Below chl-a TMDL target & MC absent’ samples.

There was a wider spread of eukaryotic communities from the same lake conditions in the CCA (Figure 19), so the eukaryotic communities did not cluster as tightly as that of the prokaryotic

communities based on lake conditions. Communities from September and August 2021 were particularly distinct from the majority of samples as they were placed far away from the rest of the samples. These samples were all Above chl-a TMDL target & MC present samples and samples from the 2021 Sep 20 Lower Arm (S09 and S10) were positively associated with TN, total MC, and chl-a concentrations, while the remaining samples from September and August 2021 were positively associated with TP. PERMANOVA analysis of the CCA results once again confirmed that environmental parameters had significant influence on the community variability ($p < 0.001$), with Temperature ($p < 0.001$), TP ($p < 0.001$), Total MC ($p < 0.001$), and Chl-a ($p < 0.001$) all driving community variability significantly.

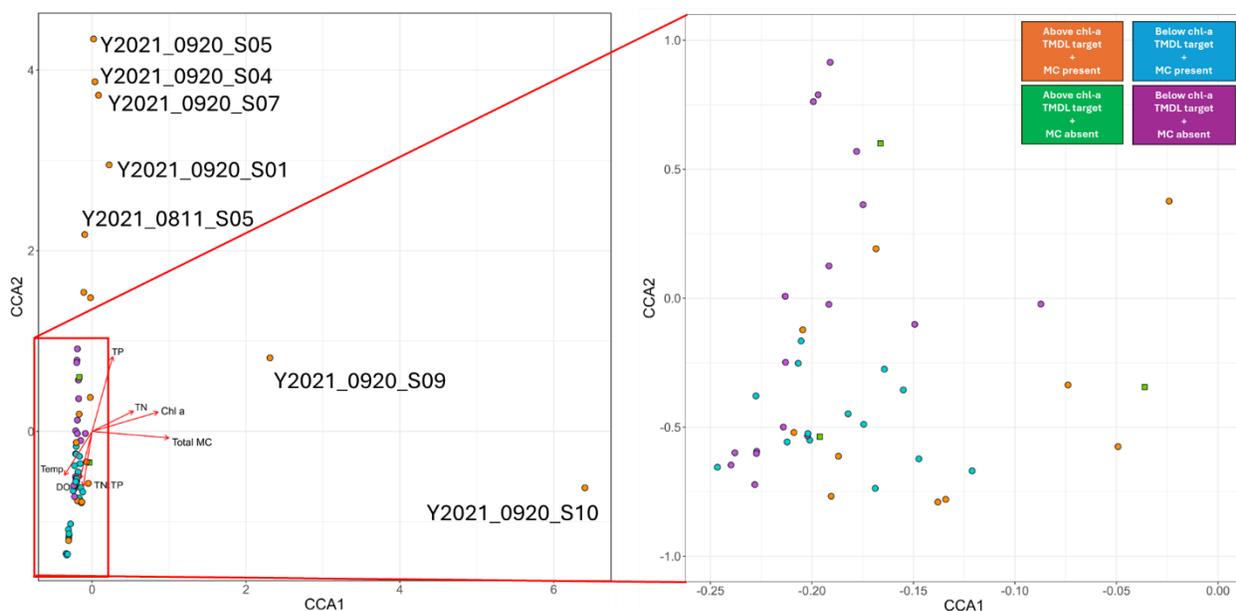


Figure 20. Canonical correlation analysis of the eukaryotic community and environmental parameters (temperature, DO, total MC, chl-a, TN, TP, TN:TP). Samples (points) are colored by lake conditions. Green ‘Above chl-a TMDL target & MC absent’ samples have square shapes to distinguish them from the blue ‘Below chl-a TMDL target & MC absent’ samples. Right panel is a zoomed in portion of the portion framed by the red square in the left panel. Some sample names are indicated for outlier data points. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

CCA results and PERMANOVA tests indicated that TP was a significant driver of both prokaryotic (including cyanobacteria) and eukaryotic communities. Interestingly, TP was generally associated with prokaryotic communities in Below chl-a TMDL target & MC absent samples but was also linked to eukaryotic communities in Above chl-a TMDL target & MC present samples from August and September 2021. Notably, TP concentrations in September 2021 were among

the highest recorded in this study (0.4771 – 0.5706 mg/L; Fig. 12D in Florea et al., 2022), and CCA results showed that these samples had distinct eukaryotic communities, including a higher composition of ciliates and other bacterivores (e.g., cercozoans and katablepharids, as previously discussed).

Insights from metagenomic analyses

While the MC concentrations in samples from 2021 were generally higher (Florea et al., 2022), full *mcy* gene clusters were only found in samples from 2020. A total of 16 bins of prokaryotic organisms (13 cyanobacteria, 2 Pseudomonadota; 1 Planctomycetota) were assembled from samples collected on 2020 Aug 28 at S01, S04, S05, S07, and S09 (Table 5). A bin refers to a collection of DNA sequences that are bioinformatically grouped together and likely originate from a single microbial genome. Completeness refers to how much of the microbe’s genome is present in the bin, estimated by comparing it to reference genomes from similar organisms. Contamination refers to estimated presence of DNA from multiple organisms within the same bin. All bins had > 50% completeness, with all but two exhibiting <10% contamination. Among the cyanobacterial bin, four were classified as *Microcystis panniformis*, with the highest quality bin reaching 90% completion and 7% contamination. Additionally, several high-quality bins belonging to the genus *Limnoraphis* were identified. Notably, two *mcy* gene clusters were detected in a *Microcystis* bin and a *Limnoraphis* bin.

Table 5. Details on genomes assembled from metagenomic samples and indication of the presence of microcystin synthetase (*mcy*) gene cluster. Completeness refers to how much of the microbe’s genome is present in the bin and contamination refers to estimated presence of DNA from multiple organisms within the same bin.

Sample Name	Classification	Completeness (%)	Contamination (%)	Presence of <i>mcy</i> ?
Y2020_0828_S01	<i>Dolichospermum circinale</i>	90	8	no
Y2020_0828_S01	<i>Microcystis panniformis</i>	90	7	yes
Y2020_0828_S04	Fonsibacter (Pseudomonadota)	99	5	no
Y2020_0828_S04	Pirellulales (Planctomycetota)	80	2	no
Y2020_0828_S04	Burkholderiaceae (Pseudomonadota)	98	23	no

Sample Name	Classification	Completeness (%)	Contamination (%)	Presence of <i>mcy</i> ?
Y2020_0828_S05	<i>Microcystis panniformis</i>	81	2	no
Y2020_0828_S05	<i>Limnoraphis robusta</i>	98	12	no
Y2020_0828_S05	<i>Planktothrix agardhii</i>	96	2	no
Y2020_0828_S05	<i>Planktothrix</i>	75	6	no
Y2020_0828_S07	<i>Dolichospermum circinale</i>	85	3	no
Y2020_0828_S07	<i>Microcystis panniformis</i>	84	1	no
Y2020_0828_S07	<i>Limnoraphis robusta</i>	99	4	yes
Y2020_0828_S07	<i>Planktothrix agardhii</i>	96	8	no
Y2020_0828_S09	<i>Limnoraphis robusta</i>	99	1	no
Y2020_0828_S09	<i>Microcystis panniformis</i>	51	0	no
Y2020_0828_S09	<i>Dolichospermum</i>	60	6	no

Additional metagenomic sequencing of samples without MC produced an additional 209 bins with > 50% completeness and < 20% contamination, 16 of which were classified as cyanobacteria (Table 6). Similar to results from the 2020 Aug 25 samples, bins classified as *Limnoraphis* spp., *Microcystis panniformis*, and *Dolichospermum* spp. were common in the samples. No *mcy* genes were found amongst any of these bins, congruent with the lack of MC detection in these samples.

Table 6. Details on genomes assembled from metagenomic samples without MC detected and indication of the presence of microcystin synthetase (*mcy*) gene cluster. Completeness refers to how much of the microbe's genome is present in the bin and contamination refers to estimated presence of DNA from multiple organisms within the same bin.

Sample Name	Classification	Completeness	Contamination	Presence of <i>mcy</i> ?
Y2020_0814_S04	<i>Limnoraphis robusta</i>	96	7	no
Y2020_0814_S04	<i>Microcystis panniformis</i>	53	3	no
Y2020_0814_S04	<i>Dolichospermum circinale</i>	83	12	no

Sample Name	Classification	Completeness	Contamination	Presence of <i>mcy</i> ?
Y2020_0814_S05	<i>Limnoraphis robusta</i>	58	8	no
Y2020_0814_S05	<i>Dolichospermum</i>	76	16	no
Y2020_0814_S06	<i>Dolichospermum circinale</i>	91	6	no
Y2020_0814_S07	<i>Sphaerospermopsis</i>	58	15	no
Y2020_0825_S04	<i>Limnoraphis</i>	60	13	no
Y2020_0825_S04	<i>Dolichospermum</i>	100	3	no
Y2020_0825_S06	<i>Microcystis panniformis</i>	90	2	no
Y2020_0825_S06	<i>Dolichospermum</i>	100	1	no
Y2021_0617_S04	<i>Limnoraphis robusta</i>	53	9	no
Y2021_0617_S04	<i>Dolichospermum flosaquae</i>	100	2	no
Y2021_0617_S04	<i>Cyanobium</i>	100	1	no
Y2021_0617_S04_Rep 2	<i>Dolichospermum flosaquae</i>	100	2	no
Y2021_0617_S04_Rep 2	<i>Cyanobium</i> sp947458155	94	0	no

Insights from metatranscriptomic analyses

The profiles of annotated RNA reads from metatranscriptomic data did not show significant differences between lake conditions for both prokaryotic and eukaryotic communities (Figure 21 & Figure 22). The assignment of RNA reads to protein reference databases was low (prokaryotic RNA: 2.39 – 15.10%; eukaryotic RNA: 0.23 – 5.17%). Each mapped RNA read was assigned Gene Ontology (GO) term(s) to describe the function of the gene. GO terms are categorized into three main groups: 1) Cellular Components (CC), which include genes coding for proteins that are part of the cell or its extracellular environment (e.g. cell membrane); 2) Molecular Function (MF), which include genes coding for proteins involved in molecular activities (e.g. binding and oxidoreductase activities); and 3) Biological Process (BP), which include genes coding for proteins involved in biological processes essential to cellular function (e.g. toxin production, nitrogen fixation, and photosynthesis). The distributions of the three GO terms category differed significantly between lake conditions for both prokaryotic and eukaryotic RNA (PERMANOVA, $p < 0.05$), indicating that overall functional composition varied by lake condition. However, Kruskal-Wallis tests of individual GO terms did not reveal significant differences ($p > 0.05$), suggesting that the observed PERMANOVA differences are subtle and spread across multiple GO terms rather than driven by a dominant function. This suggests small but widespread shifts in functional composition between lake conditions for

both prokaryotic and eukaryotic genes, rather than major changes in specific gene functions. The proportion of RNA reads associated with photosynthesis-related proteins did not differ significantly across lake conditions for either prokaryotic or eukaryotic communities (Figure 23 & Figure 24; Kruskal-Wallis, $p > 0.05$).

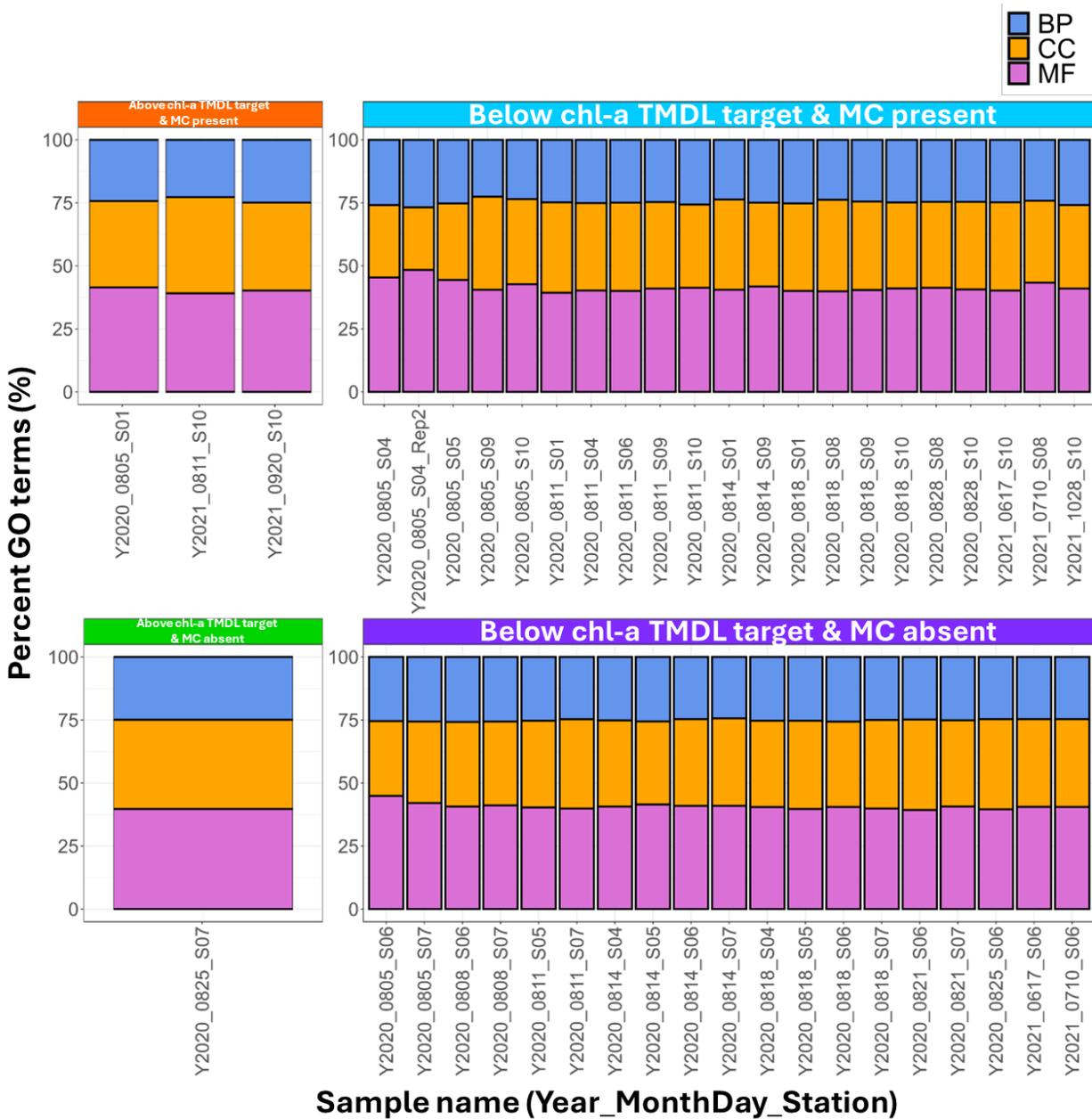


Figure 21. The distribution of the three categories (BP: Biological Process; CC: Cellular Component; MF: Molecular Function) of gene ontology (GO) terms for mapped prokaryotic RNA across different lake conditions. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

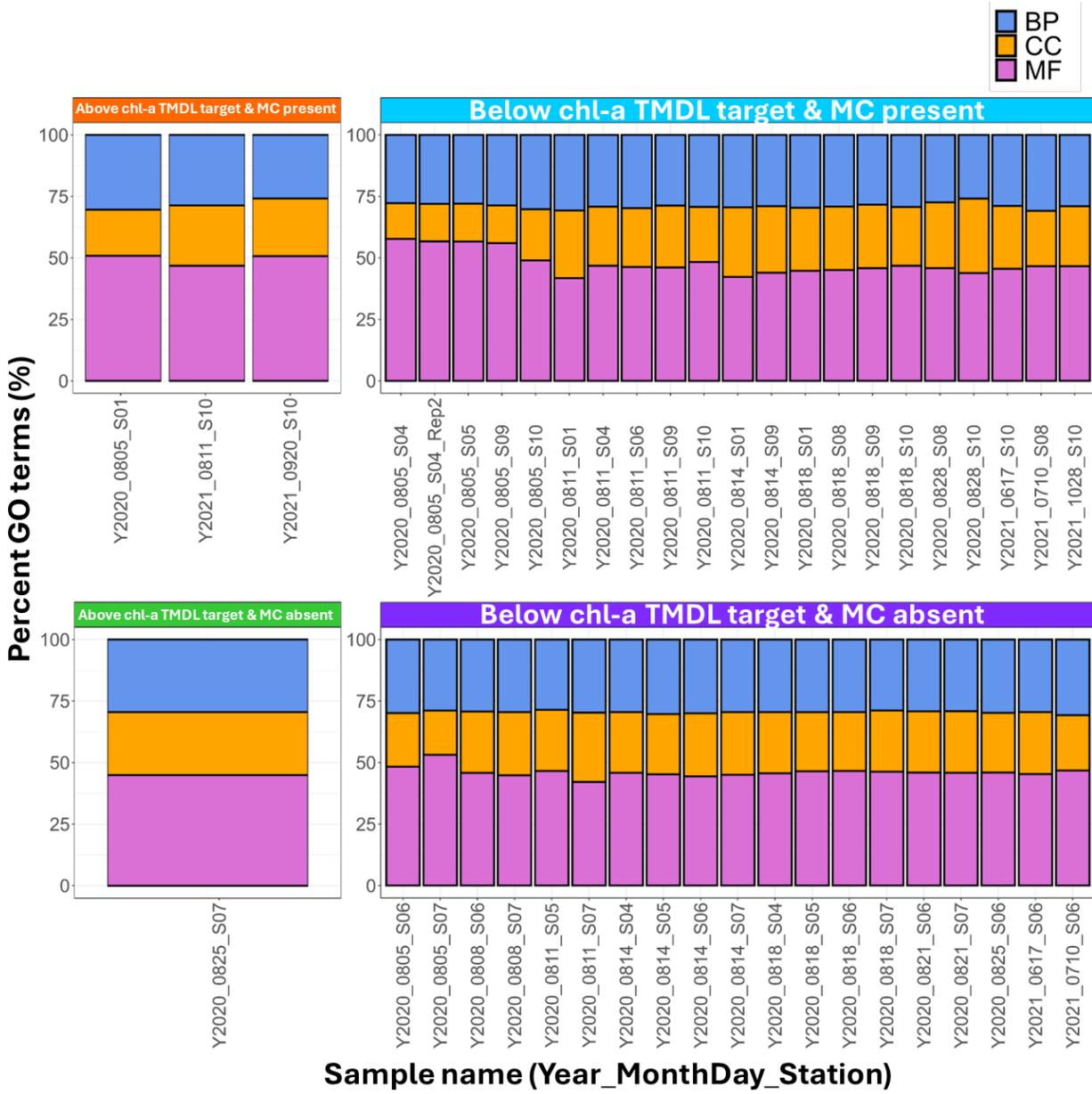


Figure 22. The distribution of the three categories (BP: Biological Process; CC: Cellular Component; MF: Molecular Function) of gene ontology (GO) terms for mapped eukaryotic RNA across different lake conditions. Sample names follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

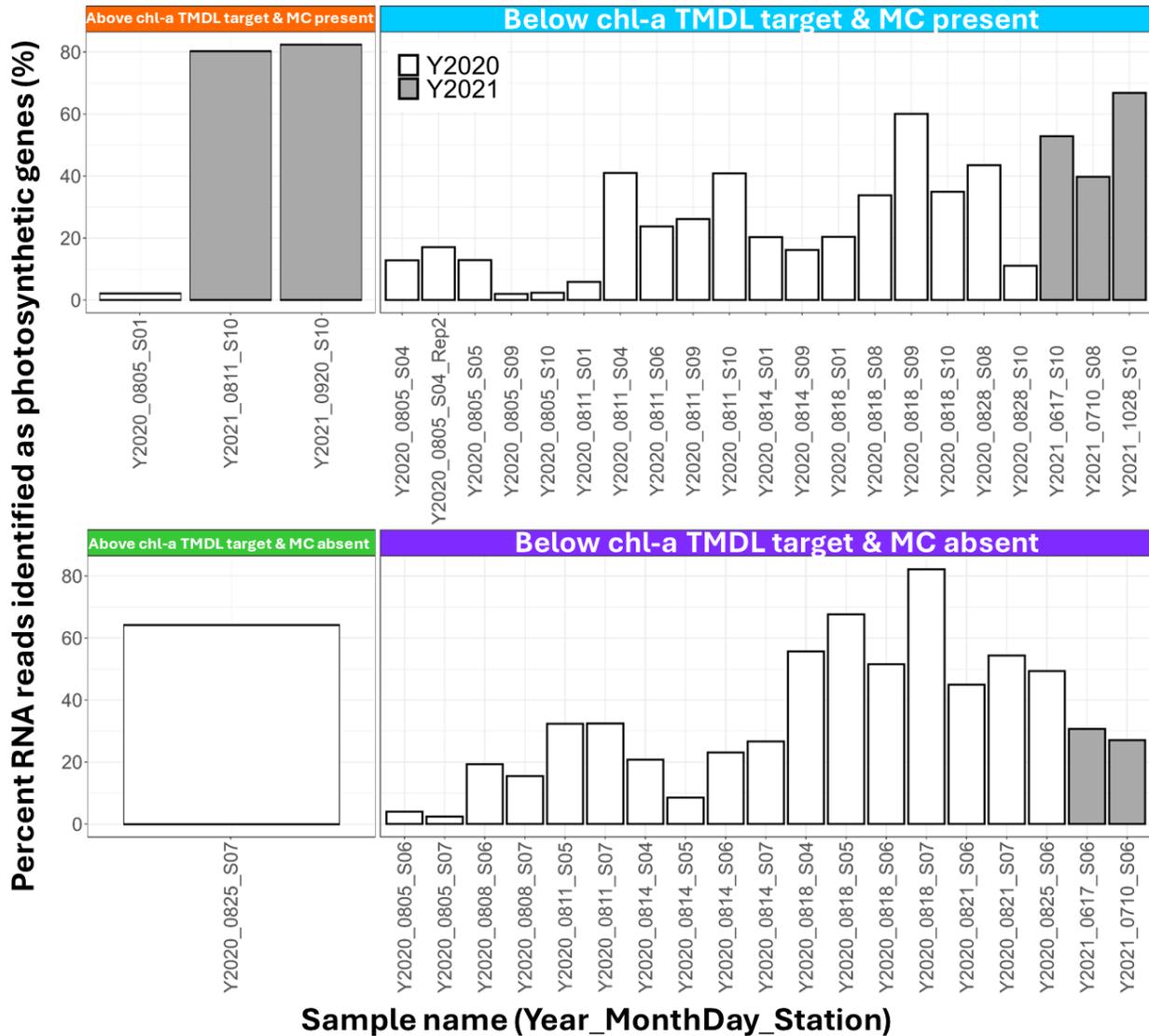


Figure 23. Percentage of prokaryotic RNA associated with photosynthesis proteins. Colors of the bar graphs indicate sample year. Y-axis shows the percentage of RNA reads that were identified as genes related to photosynthesis based on the GO termed assigned to the gene. X-axis shows the sample names, which follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

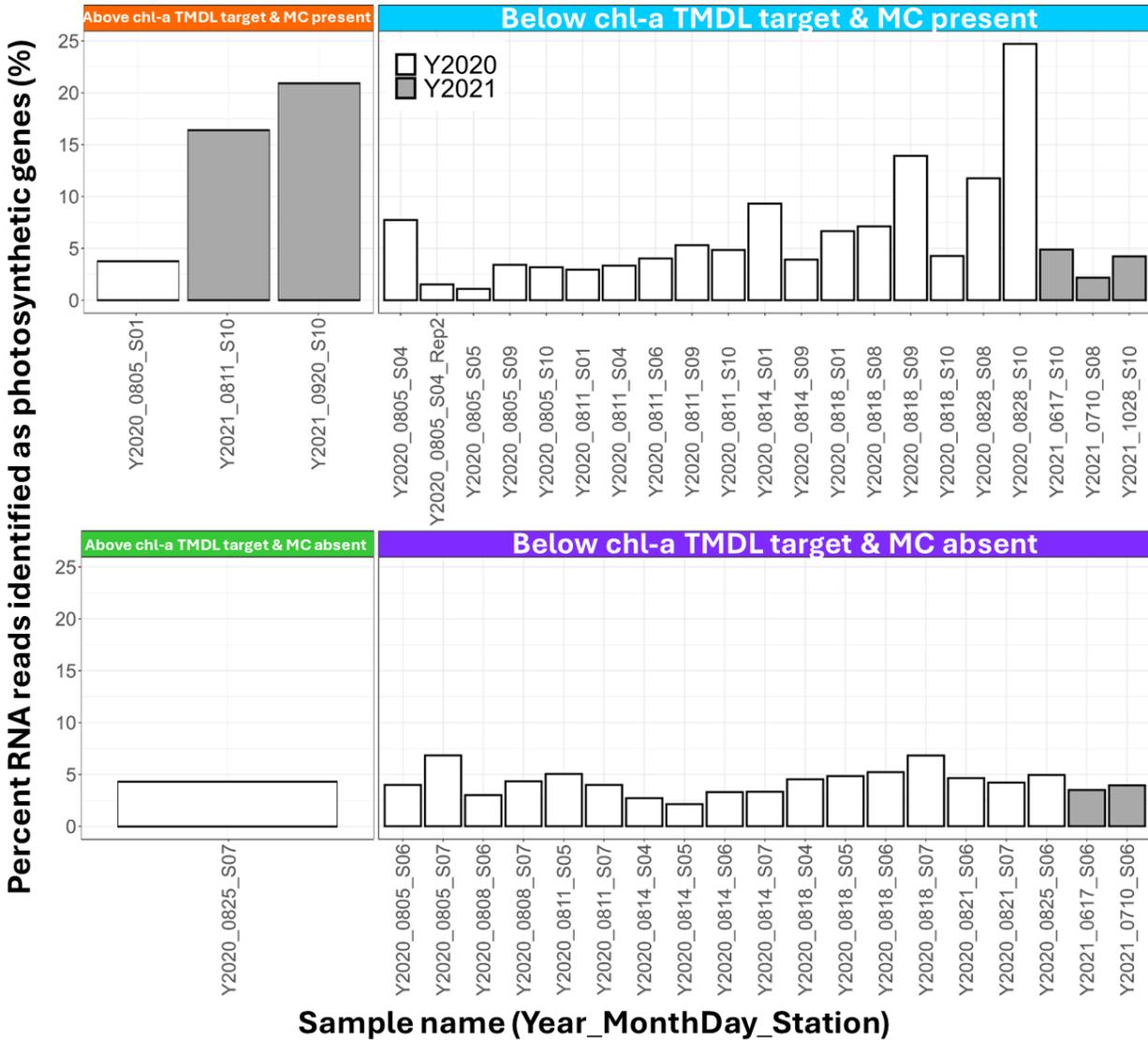


Figure 24. Percentage of eukaryotic RNA associated with photosynthesis proteins. Colors of the bar graphs indicate sample year. Y-axis shows the percentage of RNA reads that were identified as genes related to photosynthesis based on the GO termed assigned to the gene. X-axis shows the sample names, which follow the format of Y<yyyy>_<mmdd>_S<station number>. Refer to Table S1 for more details on each sample.

DISCUSSION

Potential producers of MC at Clear Lake

The results of this study suggested that *Limnoraphis* may be a potential producer of MC at Clear Lake. *Limnoraphis*/*Lyngbya* is a genus that produces a number of toxic compounds (Thuan et

al., 2019), and *Lyngbya*-like cyanobacteria isolated from southern California has been shown to produce microcystin-LR (Izaguirre et al., 2007), but no direct report of *Limnoraphis*/*Lyngbya* as microcystin producer is currently available. However, metagenomic data in this study revealed a *mcy* gene in an assembled *Limnoraphis* genome. There is a possibility that the automatic binning program maxbin2 erroneously assigned the *mcy* gene cluster to the *Limnoraphis* bin. As mentioned previously, there is no direct report of *Limnoraphis* as an MC producer, and a coverage difference was observed between the *mcy* operon and the rest of the *Limnoraphis* genome. Further verification, such as nucleotide composition analysis or long-read sequencing, will be conducted to confirm whether the *Limnoraphis* strain at Clear Lake possessed a *mcy* gene cluster. Such confirmation would not only clarify the MC producing taxa in Clear Lake, but also provide the first evidence of *Limnoraphis* as a putative MC producer.

Besides *Limnoraphis*, *Planktothrix* could also be a producer of MC at Clear Lake. *Planktothrix* is a known producer of MC (Christiansen et al., 2003) and one sample (from 2021 Oct 28 at S10; 8th bar from the right in Figure 12) had notably high contributions of *Planktothrix* in the presence of MC. In synthesis with results from Kalra et al. (accepted), these results suggested that the *Limnoraphis* and *Planktothrix* may be producers of MC at low levels in Clear Lake, but the presence of high levels of MC (> 100 µg/L) may be due to production by *Microcystis*. This is suggested by the higher contribution of *Microcystis* in the one sample with high (190 µg/L) MC concentration, as well as the significant correlation of *Microcystis* and MC concentrations when 2021 samples were included in Kalra et al., (accepted). Interestingly, while both *Aphanizomenon* and *Dolichospermum* are potential producers of MC (Cirés and Ballot, 2016; Li et al., 2016), they tended to be associated with samples without MC in this study.

Microcystis is a widely accepted producer of MC in Clear Lake, but *Microcystis* was not present in most of the samples sequenced for 16S metabarcoding in this phase of the study. *Microcystis* was observed in some samples, but did not constitute a significant portion of the cyanobacterial community except in the sample from 2021 Sep 20 at S10 (with MC concentrations at 190 µg/L; Figure 10). Most of the 16S samples for which metabarcoding analyses were conducted in this study were from 2020, and 2021 samples were from June 2021 and July – October (Station 10). *Microcystis* constituted a much more significant portion of the cyanobacterial community beginning August 2021, especially at the Lower Arm (Kalra et al. accepted). Therefore, *Microcystis* was a much more significant component of Above chl-a TMDL target & MC present samples in the later months of 2021 that were not included in this study (Figure 4 in Kalra et al., accepted). While the relative abundance of *Microcystis* was not significantly related to MC presence/absence in this study, care must be taken when interpreting this result. Kalra et al. (accepted) included a set of 2021 samples in their analysis and found that the relative abundance of *Microcystis* correlated positively with MC concentrations (Figure 10 in Kalra et al., accepted).

Other members of the prokaryotic community at Clear Lake

Pseudomonadota (formerly proteobacteria) had high relative abundance in several samples. A family of gammaproteobacteria, Sutterellaceae, was particularly dominant in the sample with the highest MC concentration (Y2021_0921_S10). This family of bacteria has also been documented as a key indicator of *Microcystis* bloom in the Daechung Reservoir in South Korea (Le et al., 2023) and in the Sulejow Reservoir in Central Poland (Mankiewicz-Boczek and Font-Nájera, 2022). Members of Pseudomonadota (specifically Alphaproteobacteria and Gammaproteobacteria) are often reported to be associated with *Microcystis* blooms (Mou et al., 2013; Shia et al., 2010; Zuo et al., 2021), especially within the bacteria community attached to *Microcystis* colonies (Mankiewicz-Boczek and Font-Nájera, 2022; Wu et al., 2019). All Pseudomonadota in this sample from 2021 Sep 20 at S10 were either Alphaproteobacteria or Gammaproteobacteria, with the majority (93% of all reads in the sample) being Gammaproteobacteria. Gammaproteobacteria had been reported to be highly associated with declining *Microcystis* blooms (Zheng et al., 2008). Several genera of Pseudomonadota, such as *Pseudomonas*, *Sphingomonas*, and *Sphingophxis* have been identified as capable of degrading various microcystin congeners (Massey and Yang, 2020), but none of these genera were present in this sample. It should also be noted that S10 is located at the far end of the Lower Arm (Figure 1), where blooms tend to accumulate due to northwesterly wind and hydrology patterns (Florea et al., 2022; Smith et al., 2023). Given the high relative abundance of gammaproteobacteria in this sample and the general association of this taxon with declining *Microcystis*, another possible explanation of the high MC concentration despite low relative abundance of cyanobacteria may be MC transport. The MC detected at this site may have been dissolved MC released from lysed *Microcystis* cells elsewhere in the lake, which were subsequently transported to this region.

Despite the observation of high cyanobacterial cell abundance in samples via microscopy in Florea et al. (2022), the relative abundance of cyanobacteria varied notably among samples in this study. Many samples from this study showed a higher proportion of non-cyanobacterial bacteria based on 16S metabarcoding reads. This discrepancy is likely due to certain bacterial and cyanobacterial taxa (e.g., picocyanobacteria) being more abundant in these samples while remaining undetected in microscopy observations due to their smaller size. It should also be noted that the percentage of reads does not necessarily reflect actual relative abundance of the taxonomic groups as there is potential for gene copy number bias. Cyanobacteria typically have two copies of the 16S rRNA gene, while some groups of bacteria (e.g. Gammaproteobacteria) have a mean number of ~six copies (Větrovský and Baldrian, 2013). Additionally, amplification and sequencing biases toward certain bacterial groups could further underestimate the relative

abundance of cyanobacteria in the metabarcoding results compared to the previously reported microscopy results.

Environmental factors associated with chl-a, MC, and the microbial community

Nutrients played a major role in the accumulation of chl-a and MC concentrations, as well as community composition. A recent historical analysis of water quality trends in Clear Lake found that surface water temperatures have increased only minimally over the past 70 years, suggesting that rising temperatures may not be a primary driver of the increased frequency of blooms in the last decade (Smith et al., 2023). This aligns with the findings of the present study that temperature did not correlate to chl-a or MC concentrations. On the other hand, TN was found to be strongly and positively correlated with both chl-a and MC concentrations. While the relationship between chl-a and TN might be expected (particularly given that the algal cells comprising the chl-a contribute to TN), the relationship to MC does suggest the nitrogen plays a role in MC production. While phosphorus (P) has long been considered the primary limiting nutrient for cyanobacterial growth in freshwater systems (Schindler, 1974; Schindler et al., 2008), growing evidence suggests that nitrogen (N) also plays a critical role in bloom formation and cyanotoxin regulation (Dolman et al., 2012; Gobler et al., 2016; Paerl et al., 2016). Different cyanobacterial taxa can exhibit varied responses to nitrogen and phosphorus availability (Dolman et al., 2012), which may explain the differing relationships between nutrients, chl-a, and total MC across years. The cyanobacterial community was dominated by *Dolichospermum* and *Lyngbya* in 2020, whereas *Microcystis* and *Cyanobium* contributed more substantially to the community in 2021 (Kalra et al., accepted). Nutrient availability can both shape or be influenced by community composition. Different cyanobacteria may have different nutrient requirements or preferences, but specific cyanobacteria, such as nitrogen fixers, could also influence nutrient availability by introducing nitrogen into the system. Therefore, differences in relationship between nutrients and lake conditions across years, as observed through both univariate correlation and multivariate analyses, may be closely linked to the cyanobacteria community.

Differences observed between years could also be due to drought, as 2020 was identified as dry by the California Department of Water Resources Water Year Hydrologic Classification Index for the larger Sacramento and San Joaquin Valley region, and 2021 was identified as critically dry (California Department of Water Resources, 2025). Drought conditions can significantly impact nutrient dynamics in aquatic ecosystems and influence phytoplankton communities, including cyanobacteria. Reduced freshwater inflow during droughts often leads to higher nutrient concentrations due to decreased dilution, potentially favoring cyanobacterial blooms (Brasil et

al., 2016). Additionally, increased water temperatures during droughts can also enhance cyanobacterial growth (Paerl and Huisman, 2009).

Overall, these analyses agreed with the univariate analyses from Florea et al. (2022), confirming strong associations between TN, TP, and TN:TP ratios with chl-a and total MC concentrations. However, the environmental parameters explored here could not fully explain the differences between lake conditions. While this analysis considered all lake conditions together, it is important to note that differences between lake arms, as documented in Florea et al. (2022), may introduce additional variability, particularly given the unique size, circulation, and bathymetry characteristics of each lake arm (Smith et al., 2023).

The diversity of the prokaryotic community at Clear Lake was positively associated with MC presence while the diversity of the eukaryotic community was positively associated with chl-a concentrations above the TMDL target. The lower diversity of the prokaryotic community in the presence of MC is not surprising, given that one of the hypothesized function of MC is allelopathy for competition reduction (Wei et al., 2024). Reduced bacterial community evenness and diversity has been observed in the presence of toxic *Microcystis* blooms in several other studies (Mankiewicz-Boczek and Font-Nájera, 2022; Zhang et al., 2019). While the presence of MC was not associated with a difference in the diversity of the eukaryotic community in this study, previous studies have reported that eukaryotic communities can be significantly influenced by the presence of MC which led to decrease in diversity and shifts in community composition (Chen et al., 2010; Zhang et al., 2019). It is possible that the persistence of MC in Clear Lake and prolonged exposure of the eukaryotic community to the toxin had led to the development of resistance to MC. Studies supporting this theory for protistan organisms is limited, but there is suggestion on pathways that are able to degrade MC in the mixotrophic chrysophyte *Ochromonas* (Zhang et al., 2018). There are also a number of studies on rapid development of tolerance to toxic *Microcystis* by cladocerans (Guo and Xie, 2006; Jiang et al., 2016). There is also documentation of higher eukaryotic community diversity during lower-density cyanobacterial blooms of 0.34 and 2.80 mg/L of cyanobacterial biomass (Xu et al., 2022). The authors did not measure chl-a in their study, so we are not able to compare the intensity their low-density bloom directly to the present study. A high-level bloom (6.27 mg/L of cyanobacterial biomass) in their study did lead to reduced eukaryotic community diversity. Similarly, results from this study suggested a plateau or even slight decrease in Spearman's correlation at high chl-a concentrations (> 200 µg/L; Figure 14).

TP and TN played an important role in shaping the microbial community. Florea et al. (2022) suggested that alternating periods of calm and well-mixed water in Clear Lake could lead to anoxic conditions and release of phosphorus at the sediment surface (internal loading), so the authors proposed that phosphorus fluxes from internal loading may be a major driver of

cyanobacterial community dynamics. The results from multivariate analyses in this study further support that phosphorus was a strong driver of overall microbial community structure. In contrast, TN was not identified as a significant driver of either the prokaryotic or eukaryotic communities based on the multivariate analyses. While TN was strongly correlated with chl-a and MC (Table 4), its lack of influence on microbial community composition suggests that bulk nitrogen availability does not directly shape microbial community structure. This may be due to multiple factors: 1) TN includes various nitrogen forms, not all of which are bioavailable to microbes, making it a broad indicator of nutrient enrichment rather than a direct driver of community composition; 2) microbial community structure was influenced more strongly by other factors, such as TP (as indicated by multivariate analyses); 3) the common presence (and sometimes dominance) of nitrogen-fixing cyanobacteria, such as *Limnoraphis* and *Dolichospermum* (Figure 10), may supply sufficient nitrogen to the microbial community, reducing the overall influence of external nitrogen sources. Indeed, Kalra et al. (accepted) noted a seasonal succession of cyanobacterial taxa in 2021, in which nitrogen fixers such as *Dolichospermum* preceded non-nitrogen fixers such as *Microcystis* and *Planktothrix*. Nevertheless, TN was significantly correlated to the relative abundance of specific cyanobacteria, such as *Cyanobium*, *Vulcanococcus*, *Dolichospermum*, and *Limnoraphis* (Spearman's correlation, $p < 0.001$). These results reaffirmed the influence of TN on certain cyanobacterial taxa, even if it did not significantly influence overall microbial community composition.

RECOMMENDATIONS

The findings in this study reinforce the complexity of cyanoHAB dynamics in Clear Lake, emphasizing the importance of phosphorus and nitrogen availability, microbial interactions, and potential adaptive responses of the eukaryotic community to long-term cyanotoxin exposure. These insights provide a foundation for refining monitoring strategies and developing targeted mitigation efforts for Clear Lake. Recommendations for future work include:

- 1. Cyanobacterial HAB monitoring efforts should account for the presence of multiple genera of microcystin producers:** The results of this study identified several cyanobacteria genera positively associated with total MC concentrations, implying there are multiple producers of MC in Clear Lake. Monitoring efforts should not focus on a single cyanobacterial taxon (e.g., *Microcystis*) and instead should be more broadly focused. For example, current Clear Lake Cyanotoxin Monitoring Program by the Big Valley Band of Pomo Indians utilize qPCR analyses to track increases in *mcy* genes and guide toxin analysis efforts (Big Valley Band of Pomo Indians, 2024). The primers used for monitoring programs at Clear Lake should be evaluated to ensure they include *mcy*

genes identified in the metagenomic analysis, as well as potential MC producers identified in this study. This will ensure rapid and early detection of blooms with MC production potential.

- 2. Bloom management efforts in Clear Lake should consider the influence of multiple bloom drivers:** Results of this study indicated that TN and TP were strong drivers of lake conditions, including blooms with chl-a levels exceeding the current TMDL target as well as the presence of MC. TP has long been the focus of bloom controls, however the results of this study and others (Gobler et al., 2016; Paerl et al., 2016) have pointed to the potential role of nitrogen in both the formation of high levels of cyanobacterial biomass and the regulation of toxin production by toxigenic taxa. This study indicated that TP plays a role in bloom formation and shaping microbial communities. And while TN did not significantly explain the variation observed in the microbial communities as a whole, it was strongly correlated with chl-a and MC concentrations and was correlated to the relative abundance of specific cyanobacteria. Together, these findings suggest both nutrients play an important role in the accumulation of cyanobacterial biomass, community composition, and MC presence. Future management efforts should include nitrogen in addition to phosphorus.
- 3. Apply the molecular methods piloted in this study to explore additional cyanobacterial bloom drivers:** This study was conducted in years that were identified as dry (2020) and critically dry (2021) by the California Department of Water Resources Water Year Hydrologic Classification Index for the larger Sacramento and San Joaquin Valley region. In addition to nutrients, a recent historical analysis of Clear Lake suggested that hydrologic factors such as precipitation, lake level and lake discharge rates may play an important role in bloom formation and toxin production through a variety of pathways (Smith et al., 2023). Thus, this study did not capture the full gradient of hydrologic conditions at Clear Lake. Future study should consider these factors to determine if management of these factors may also support water quality goals for Clear Lake. The molecular methods (metabarcoding, metatranscriptomics, and metagenomics) employed in this study provides a useful framework for this work. For example, the quantitative assessment of the cyanobacterial community allowed for the identification of multiple MC producers and this approach could be applied to identify potential shifts in producers in response to environmental changes in the lake. The metagenomic sequencing efforts also resulted in a useful future resource in the development of a custom reference library for any future prokaryotic metatranscriptome analyses.

REFERENCES

- Backer, L.C., Carmichael, W., Kirkpatrick, B., Williams, C., Irvin, M., Zhou, Y., Johnson, T.B., Nierenberg, K., Hill, V.R., Kieszak, S.M., Cheng, Y.-S., 2008. Recreational Exposure to Low Concentrations of Microcystins During an Algal Bloom in a Small Lake. *Mar. Drugs* 6, 389–406. <https://doi.org/10.3390/md6020389>
- Big Valley Band of Pomo Indians, 2024. Clear Lake Cyanobacteria and Cyanotoxins Monitoring Program.
- Bradbury, J.P., 1988. Diatom biostratigraphy and the paleolimnology of Clear Lake, Lake County, California, in: Sims, J.D. (Ed.), *Late Quaternary Climate, Tectonism, Sedimentation in Clear Lake, Northern California Coasts*. Geological Society of America, Boulder, CO.
- Brasil, J., Attayde, J.L., Vasconcelos, F.R., Dantas, D.D.F., Huszar, V.L.M., 2016. Drought-induced water-level reduction favors cyanobacteria blooms in tropical shallow lakes. *Hydrobiologia* 770, 145–164. <https://doi.org/10.1007/s10750-015-2578-5>
- California Cyanobacterial and Harmful Algal Bloom Network, 2016. Appendix to the CCHAB Preliminary Changes to the Statewide Voluntary Guidance on CyanoHABs in Recreational Waters, January 2016. California Cyanobacterial and Harmful Algal Bloom Network.
- California Department of Water Resources, 2025. CDEC Water Year Type Dataset.
- Central Valley Regional Water Quality Control Board, 2007. Amendment to the Water Quality Control Plan for the Sacramento River and San Joaquin River Basins for the Control of Nutrients in Clear Lake. Central Valley Regional Water Quality Control Board, Rancho Cordova, CA.
- Chen, M., Chen, F., Xing, P., Li, H., Wu, Q.L., 2010. Microbial eukaryotic community in response to *Microcystis* spp. bloom, as assessed by an enclosure experiment in Lake Taihu, China. *FEMS Microbiol. Ecol.* 74, 19–31. <https://doi.org/10.1111/j.1574-6941.2010.00923.x>
- Christiansen, G., Fastner, J., Erhard, M., Börner, T., Dittmann, E., 2003. Microcystin Biosynthesis in *Planktothrix*: Genes, Evolution, and Manipulation. *J. Bacteriol.* 185, 564–572. <https://doi.org/10.1128/jb.185.2.564-572.2003>
- Cirés, S., Ballot, A., 2016. A review of the phylogeny, ecology and toxin production of bloom-forming *Aphanizomenon* spp. and related species within the Nostocales (cyanobacteria). *Glob. Expans. Harmful Cyanobacterial Blooms Divers. Ecol. Causes Controls* 54, 21–43. <https://doi.org/10.1016/j.hal.2015.09.007>

- Dolman, A.M., Rücker, J., Pick, F.R., Fastner, J., Rohrlack, T., Mischke, U., Wiedner, C., 2012. Cyanobacteria and Cyanotoxins: The Influence of Nitrogen versus Phosphorus. *PLOS ONE* 7, e38757. <https://doi.org/10.1371/journal.pone.0038757>
- Facciponte, D.N., Bough, M.W., Seidler, D., Carroll, J.L., Ashare, A., Andrew, A.S., Tsongalis, G.J., Vaickus, L.J., Henegan, P.L., Butt, T.H., Stommel, E.W., 2018. Identifying aerosolized cyanobacteria in the human respiratory tract: A proposed mechanism for cyanotoxin-associated diseases. *Sci. Total Environ.* 645, 1003–1013. <https://doi.org/10.1016/j.scitotenv.2018.07.226>
- Florea, K., Stewart, B., Webb, E., Caron, D.A., Smith, J., 2022. Environmental Drivers of Cyanobacterial Harmful Algal Blooms and Cyanotoxins in Clear Lake: 2020-2021. Technical Report 1261. Southern California Coastal Water Research Project, Costa Mesa, CA.
- Gobler, C.J., Burkholder, J.M., Davis, T.W., Harke, M.J., Johengen, T., Stow, C.A., Van de Waal, D.B., 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. *Glob. Expans. Harmful Cyanobacterial Blooms Divers. Ecol. Causes Controls* 54, 87–97. <https://doi.org/10.1016/j.hal.2016.01.010>
- Guo, N., Xie, P., 2006. Development of tolerance against toxic *Microcystis aeruginosa* in three cladocerans and the ecological implications. *Environ. Pollut.* 143, 513–518. <https://doi.org/10.1016/j.envpol.2005.11.044>
- Izaguirre, G., Jungblut, A., Neilan, B., 2007. Benthic cyanobacteria (Oscillatoriaceae) that produce microcystin-LR, isolated from four reservoirs in Southern California. *Water Res.* 41, 492–8. <https://doi.org/10.1016/j.watres.2006.10.012>
- Jiang, X., Gao, H., Zhang, L., Liang, H., Zhu, X., 2016. Rapid evolution of tolerance to toxic *Microcystis* in two cladoceran grazers. *Sci. Rep.* 6, 25319. <https://doi.org/10.1038/srep25319>
- Kalra, I., Stewart, B.P., Florea, K.M., Smith, J., Webb, E.A., Caron, D.A., accepted. Temporal and spatial dynamics of harmful algal bloom-associated microbial communities in eutrophic Clear Lake, California. *Appl. Environ. Microbiol.*
- Le, V.V., Kang, M., Ko, S.-R., Jeong, S., Park, C.-Y., Lee, J.J., Choi, I.-C., Oh, H.-M., Ahn, C.-Y., 2023. Dynamic response of bacterial communities to *Microcystis* blooms: A three-year study. *Sci. Total Environ.* 902, 165888. <https://doi.org/10.1016/j.scitotenv.2023.165888>
- Li, X., Dreher, T.W., Li, R., 2016. An overview of diversity, occurrence, genetics and toxin production of bloom-forming *Dolichospermum* (*Anabaena*) species. *Glob. Expans. Harmful Cyanobacterial Blooms Divers. Ecol. Causes Controls* 54, 54–68. <https://doi.org/10.1016/j.hal.2015.10.015>

- Liu, P., Ewald, J., Galvez, J.H., Head, J., Crump, D., Bourque, G., Basu, N., Xia, J., 2021. Ultrafast functional profiling of RNA-seq data for nonmodel organisms. *Genome Res* 31, 713–720. <https://doi.org/10.1101/gr.269894.120>
- Magurran, A.E., 2004. *Measuring Biological Diversity*. Blackwell Publishing.
- Mankiewicz-Boczek, J., Font-Nájera, A., 2022. Temporal and functional interrelationships between bacterioplankton communities and the development of a toxigenic *Microcystis* bloom in a lowland European reservoir. *Sci. Rep.* 12, 19332. <https://doi.org/10.1038/s41598-022-23671-2>
- Massey, I.Y., Yang, F., 2020. A Mini Review on Microcystins and Bacterial Degradation. *Toxins* 12. <https://doi.org/10.3390/toxins12040268>
- Mehinto, A.C., Smith, J., Wenger, E., Stanton, B., Linville, R., Brooks, B.W., Sutula, M.A., Howard, M.D.A., 2021. Synthesis of ecotoxicological studies on cyanotoxins in freshwater habitats – Evaluating the basis for developing thresholds protective of aquatic life in the United States. *Sci. Total Environ.* 795, 148864. <https://doi.org/10.1016/j.scitotenv.2021.148864>
- Moore, C.E., Juan, J., Lin, Y., Gaskill, C.L., Puschner, B., 2016. Comparison of Protein Phosphatase Inhibition Assay with LC-MS/MS for Diagnosis of Microcystin Toxicosis in Veterinary Cases. *Mar. Drugs* 14. <https://doi.org/10.3390/md14030054>
- Mou, X., Lu, X., Jacob, J., Sun, S., Heath, R., 2013. Metagenomic Identification of Bacterioplankton Taxa and Pathways Involved in Microcystin Degradation in Lake Erie. *PLOS ONE* 8, e61890. <https://doi.org/10.1371/journal.pone.0061890>
- Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27, 824–834. <https://doi.org/10.1101/gr.213959.116>
- Olson, N.E., Cooke, M.E., Shi, J.H., Birbeck, J.A., Westrick, J.A., Ault, A.P., 2020. Harmful Algal Bloom Toxins in Aerosol Generated from Inland Lake Water. *Environ. Sci. Technol.* 54, 4769–4780. <https://doi.org/10.1021/acs.est.9b07727>
- Paerl, H.W., Huisman, J., 2009. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environ. Microbiol. Rep.* 1, 27–37. <https://doi.org/10.1111/j.1758-2229.2008.00004.x>
- Paerl, H.W., Scott, J.T., McCarthy, M.J., Newell, S.E., Gardner, W.S., Havens, K.E., Hoffman, D.K., Wilhelm, S.W., Wurtsbaugh, W.A., 2016. It takes two to tango: when and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environ. Sci. Technol.* 50, 10805–10813.

- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25, 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Richerson, P., Suchanek, T., Why, Stephen J., 1994. The causes and control of algal blooms in Clear Lake. Clean Lakes Diagnostic/Feasibility Study For Clear Lake, California. Lake County Flood Control and Water Conservation District, California State Water Resources Control Board, United States Environmental Protection Agency.
- Schindler, D.W., 1974. Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management. *Science* 184, 897–899. <https://doi.org/10.1126/science.184.4139.897>
- Schindler, D.W., Hecky, R.E., Findlay, D.L., Stainton, M.P., Parker, B.R., Paterson, M.J., Beaty, K.G., Lyng, M., Kasian, S.E.M., 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proc. Natl. Acad. Sci.* 105, 11254–11258. <https://doi.org/10.1073/pnas.0805108105>
- Shia, L., Cai, Y., Wang, X., Li, P., Yu, Y., Kong, F., 2010. Community Structure of Bacteria Associated with Microcystis Colonies from Cyanobacterial Blooms. *J. Freshw. Ecol.* 25, 193–203. <https://doi.org/10.1080/02705060.2010.9665068>
- Smith, J., Eggleston, E., Howard, M.D.A., Ryan, S., Gichuki, J., Kennedy, K., Tyler, A., Beck, M., Huie, S., Caron, D.A., 2023. Historic and recent trends of cyanobacterial harmful algal blooms and environmental conditions in Clear Lake, California: A 70-year perspective. *Elem Sci Anth* 11, 00115. <https://doi.org/10.1525/elementa.2022.00115>
- Stanton, B., Little, A., Miller, L., Solomon, G., Ryan, S., Paulukonis, S., Cajina, S., 2023. Microcystins at the tap: A closer look at unregulated drinking water contaminants. *AWWA Water Sci.* 5, e1337. <https://doi.org/10.1002/aws2.1337>
- Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., Richards, T.A., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19, 21–31. <https://doi.org/10.1111/j.1365-294X.2009.04480.x>
- Sun, X., Tao, M., Qin, B., Qi, M., Niu, Y., Zhang, J., Ma, Z., Xie, P., 2012. Large-scale field evidence on the enhancement of small-sized cladocerans by Microcystis blooms in Lake Taihu, China. *J. Plankton Res.* 34, 853–863. <https://doi.org/10.1093/plankt/fbs047>
- Thuan, N.H., An, T.T., Shrestha, A., Canh, N.X., Sohng, J.K., Dhakal, D., 2019. Recent Advances in Exploration and Biotechnological Production of Bioactive Compounds in Three Cyanobacterial Genera: Nostoc, Lyngbya, and Microcystis. *Front. Chem.* 7.

- Větrovský, T., Baldrian, P., 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE* 8, e57923. <https://doi.org/10.1371/journal.pone.0057923>
- Wei, N., Hu, C., Dittmann, E., Song, L., Gan, N., 2024. The biological functions of microcystins. *Water Res.* 262, 122119. <https://doi.org/10.1016/j.watres.2024.122119>
- Wu, Q., Zhang, Y., Li, Y., Li, J., Zhang, X., Li, P., 2019. Comparison of community composition between *Microcystis* colony-attached and free-living bacteria, and among bacteria attached with *Microcystis* colonies of various sizes in culture. *Aquat. Ecol.* 53, 465–481. <https://doi.org/10.1007/s10452-019-09702-7>
- Wu, Y.-W., Simmons, B.A., Singer, S.W., 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Xu, H., Liu, W., Zhang, S., Wei, J., Li, Y., Pei, H., 2022. Cyanobacterial bloom intensities determine planktonic eukaryote community structure and stability. *Sci. Total Environ.* 838, 156637. <https://doi.org/10.1016/j.scitotenv.2022.156637>
- Yeh, Y.-C., McNichol, J., Needham, D.M., Fichot, E.B., Berdjeb, L., Fuhrman, J.A., 2021. Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities, and estimation of sequencing bias against 18S. *Env. Microbiol* 23, 3240–3250. <https://doi.org/10.1111/1462-2920.15553>
- Zhang, L., Lyu, K., Wang, N., Gu, L., Sun, Y., Zhu, X., Wang, J., Huang, Y., Yang, Z., 2018. Transcriptomic Analysis Reveals the Pathways Associated with Resisting and Degrading Microcystin in *Ochromonas*. *Environ. Sci. Technol.* 52, 11102–11113. <https://doi.org/10.1021/acs.est.8b03106>
- Zhang, M., Lu, T., Paerl, H.W., Chen, Y., Zhang, Z., Zhou, Z., Qian, H., 2019. Feedback Regulation between Aquatic Microorganisms and the Bloom-Forming Cyanobacterium *Microcystis aeruginosa*. *Appl. Environ. Microbiol.* 85. <https://doi.org/10.1128/AEM.01362-19>
- Zheng, X., Xiao, L., Ren, J., Yang, L., 2008. The Effect of a *Microcystis aeruginosa* Bloom on the Bacterioplankton Community Composition of Lake Xuanwa. *J. Freshw. Ecol.* 23, 297–304. <https://doi.org/10.1080/02705060.2008.9664202>
- Zuo, J., Hu, L., Shen, W., Zeng, J., Li, L., Song, L., Gan, N., 2021. The involvement of α -proteobacteria *Phenylobacterium* in maintaining the dominance of toxic *Microcystis* blooms in Lake Taihu, China. *Environ. Microbiol.* 23, 1066–1078. <https://doi.org/10.1111/1462-2920.15301>

APPENDIX A. SUPPLEMENTARY TABLE

Table S1. Summary of all samples from 2020–2021, including the date and location of sampling, lake condition, and the type of sequencing performed. ‘N’ indicates that sequencing was not performed, while ‘Y’ indicates that sequencing was performed. ‘MetaG’ refers to metagenomic sequencing, ‘metaB’ refers to metabarcoding, ‘Prok metaT’ indicates prokaryote metatranscriptomics, and ‘Euk metaT’ indicates eukaryote metatranscriptomics. Lake conditions are abbreviated, ‘Above TMDL & MC’ indicates chl-a above TMDL target and presence of MC, ‘Above TMDL & No MC’ indicates chl-a above TMDL target and absence of MC, ‘Below TMDL & MC’ indicates chl-a below TMDL target and presence of MC, and ‘Below TMDL & No MC’ indicates chl-a below TMDL target and absence of MC.

Sample Name	Date	Location	Lake condition	MetaG	16S metaB	18S metaB	Prok metaT	Euk metaT
Y2020_0805_S01	8/5/2020	S01	Above TMDL & MC	N	Y	Y	Y	Y
Y2020_0805_S04	8/5/2020	S04	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0805_S05	8/5/2020	S05	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0805_S06	8/5/2020	S06	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0805_S07	8/5/2020	S07	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0805_S09	8/5/2020	S09	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0805_S10	8/5/2020	S10	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0808_S06	8/8/2020	S06	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0808_S07	8/8/2020	S07	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0811_S01	8/11/2020	S01	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0811_S04	8/11/2020	S04	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0811_S05	8/11/2020	S05	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0811_S06	8/11/2020	S06	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0811_S07	8/11/2020	S07	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0811_S09	8/11/2020	S09	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0811_S10	8/11/2020	S10	Below TMDL & MC	N	Y	Y	Y	Y

Sample Name	Date	Location	Lake condition	Meta G	16S meta B	18S meta B	Prok meta T	Euk meta T
Y2020_0814_S01	8/14/2020	S01	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0814_S04	8/14/2020	S04	Below TMDL & No MC	Y	Y	Y	Y	Y
Y2020_0814_S05	8/14/2020	S05	Below TMDL & No MC	Y	Y	Y	Y	Y
Y2020_0814_S06	8/14/2020	S06	Below TMDL & No MC	Y	Y	Y	Y	Y
Y2020_0814_S07	8/14/2020	S07	Below TMDL & No MC	Y	Y	Y	Y	Y
Y2020_0814_S09	8/14/2020	S09	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0818_S01	8/18/2020	S01	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0818_S04	8/18/2020	S04	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0818_S05	8/18/2020	S05	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0818_S06	8/18/2020	S06	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0818_S07	8/18/2020	S07	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0818_S08	8/18/2020	S08	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0818_S09	8/18/2020	S09	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0818_S10	8/18/2020	S10	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0821_S06	8/21/2020	S06	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0821_S07	8/21/2020	S07	Below TMDL & No MC	N	Y	Y	Y	Y
Y2020_0825_S04	8/25/2020	S04	Below TMDL & No MC	Y	N	N	N	N
Y2020_0825_S06	8/25/2020	S06	Below TMDL & No MC	Y	Y	Y	Y	Y
Y2020_0825_S07	8/25/2020	S07	Above TMDL & No MC	N	Y	Y	Y	Y
Y2020_0828_S01	8/28/2020	S01	Below TMDL & MC	N	N	Y	N	N
Y2020_0828_S04	8/28/2020	S04	Above TMDL & MC	N	N	Y	N	N
Y2020_0828_S05	8/28/2020	S05	Above TMDL & MC	N	N	Y	N	N
Y2020_0828_S07	8/28/2020	S07	Above TMDL & MC	N	N	Y	N	N

Sample Name	Date	Location	Lake condition	Meta G	16S meta B	18S meta B	Prok meta T	Euk meta T
Y2020_0828_S08	8/28/2020	S08	Below TMDL & MC	N	Y	Y	Y	Y
Y2020_0828_S09	8/28/2020	S09	Above TMDL & MC	N	N	Y	N	N
Y2020_0828_S10	8/28/2020	S10	Below TMDL & MC	N	Y	Y	Y	Y
Y2021_0617_S01	6/17/2021	S01	Above TMDL & MC	N	Y	Y	N	N
Y2021_0617_S04	6/17/2021	S04	Above TMDL & No MC	Y	Y	Y	N	N
Y2021_0617_S05	6/17/2021	S05	Below TMDL & No MC	N	Y	Y	N	N
Y2021_0617_S06	6/17/2021	S06	Below TMDL & No MC	N	Y	Y	Y	Y
Y2021_0617_S07	6/17/2021	S07	Above TMDL & No MC	N	Y	Y	N	N
Y2021_0617_S09	6/17/2021	S09	Below TMDL & MC	N	Y	Y	N	N
Y2021_0617_S10	6/17/2021	S10	Below TMDL & MC	N	Y	Y	Y	Y
Y2021_0710_S01	7/10/2021	S01	Above TMDL & MC	N	N	Y	N	N
Y2021_0710_S04	7/10/2021	S04	Below TMDL & MC	N	N	Y	N	N
Y2021_0710_S05	7/10/2021	S05	Below TMDL & MC	N	N	Y	N	N
Y2021_0710_S06	7/10/2021	S06	Below TMDL & No MC	N	N	N	Y	Y
Y2021_0710_S07	7/10/2021	S07	Below TMDL & No MC	N	N	Y	N	N
Y2021_0710_S08	7/10/2021	S08	Below TMDL & MC	N	N	N	Y	Y
Y2021_0710_S09	7/10/2021	S09	Above TMDL & MC	N	N	Y	N	N
Y2021_0710_S10	7/10/2021	S10	Above TMDL & MC	N	Y	Y	N	N
Y2021_0811_S01	8/11/2021	S01	Above TMDL & MC	N	N	Y	N	N
Y2021_0811_S04	8/11/2021	S04	Above TMDL & MC	N	N	Y	N	N
Y2021_0811_S05	8/11/2021	S05	Above TMDL & MC	N	N	Y	N	N
Y2021_0811_S07	8/11/2021	S07	Above TMDL & MC	N	N	Y	N	N
Y2021_0811_S09	8/11/2021	S09	Above TMDL & MC	N	N	Y	N	N

Sample Name	Date	Location	Lake condition	Meta G	16S meta B	18S meta B	Prok meta T	Euk meta T
Y2021_0811_S10	8/11/2021	S10	Above TMDL & MC	N	N	N	Y	Y
Y2021_0920_S01	9/20/2021	S01	Above TMDL & MC	N	N	Y	N	N
Y2021_0920_S04	9/20/2021	S04	Above TMDL & MC	N	N	Y	N	N
Y2021_0920_S05	9/20/2021	S05	Above TMDL & MC	N	N	Y	N	N
Y2021_0920_S07	9/20/2021	S07	Above TMDL & MC	N	N	Y	N	N
Y2021_0920_S09	9/20/2021	S09	Above TMDL & MC	N	N	Y	N	N
Y2021_0920_S10	9/20/2021	S10	Above TMDL & MC	N	Y	Y	Y	Y
Y2021_1028_S01	10/28/2021	S01	Above TMDL & MC	N	N	Y	N	N
Y2021_1028_S04	10/28/2021	S04	Above TMDL & MC	N	N	Y	N	N
Y2021_1028_S05	10/28/2021	S05	Above TMDL & MC	N	N	Y	N	N
Y2021_1028_S07	10/28/2021	S07	Above TMDL & MC	N	N	Y	N	N
Y2021_1028_S09	10/28/2021	S09	Above TMDL & MC	N	N	Y	N	N
Y2021_1028_S10	10/28/2021	S10	Below TMDL & MC	N	Y	Y	Y	Y

Table S2. List of genera with $\geq 5\%$ relative abundance in samples sequenced for 18S metabarcoding. The 'Taxon' column corresponds to the taxa presented in the figures of this study.

Sample	Date	Site	Taxon	Genus	Percent
Y2020_0805_S01	2020_0805	S01	Chlorophyta	<i>Neglectella</i>	78
Y2020_0805_S01	2020_0805	S01	Chlorophyta	<i>Sphaeropleales_XX</i>	8
Y2020_0805_S04	2020_0805	S04	Chlorophyta	<i>Neglectella</i>	77
Y2020_0805_S04	2020_0805	S04	Parasites	<i>Chrompodellids_CHR1_XX</i>	7
Y2020_0805_S05	2020_0805	S05	Chlorophyta	<i>Neglectella</i>	84
Y2020_0805_S06	2020_0805	S06	Chlorophyta	<i>Neglectella</i>	59

Sample	Date	Site	Taxon	Genus	Percent
Y2020_0805_S06	2020_0805	S06	Cryptophyta	<i>Cryptomonas</i>	25
Y2020_0805_S06	2020_0805	S06	Chlorophyta	<i>Atractomorpha</i>	16
Y2020_0805_S07	2020_0805	S07	Chlorophyta	<i>Neglectella</i>	50
Y2020_0805_S07	2020_0805	S07	Cryptophyta	<i>Cryptomonas</i>	7
Y2020_0805_S07	2020_0805	S07	Parasites	<i>Chrompodellids_CHR1_XX</i>	6
Y2020_0805_S07	2020_0805	S07	Chlorophyta	<i>Pandorina</i>	5
Y2020_0805_S09	2020_0805	S09	Chlorophyta	<i>Neglectella</i>	60
Y2020_0805_S09	2020_0805	S09	Metazoa	<i>Brachionus</i>	16
Y2020_0805_S10	2020_0805	S10	Chlorophyta	<i>Neglectella</i>	62
Y2020_0805_S10	2020_0805	S10	Metazoa	<i>Conochilus</i>	6
Y2020_0808_S06	2020_0808	S06	Chlorophyta	<i>Sphaeropleales_XX</i>	43
Y2020_0808_S06	2020_0808	S06	Chlorophyta	<i>Neglectella</i>	24
Y2020_0808_S06	2020_0808	S06	Others	<i>Heterophrys</i>	7
Y2020_0808_S06	2020_0808	S06	Chlorophyta	<i>Ankyra</i>	5
Y2020_0808_S07	2020_0808	S07	Chlorophyta	<i>Neglectella</i>	44
Y2020_0808_S07	2020_0808	S07	Cryptophyta	<i>Cryptomonas</i>	32
Y2020_0808_S07	2020_0808	S07	Chlorophyta	<i>Pandorina</i>	24
Y2020_0811_S01	2020_0811	S01	Bacillariophyta	<i>Aulacoseira</i>	85
Y2020_0811_S01	2020_0811	S01	Chlorophyta	<i>Neglectella</i>	10
Y2020_0811_S04	2020_0811	S04	Chlorophyta	<i>Neglectella</i>	46
Y2020_0811_S04	2020_0811	S04	Bacillariophyta	<i>Aulacoseira</i>	34
Y2020_0811_S05	2020_0811	S05	Chlorophyta	<i>Neglectella</i>	65
Y2020_0811_S05	2020_0811	S05	Bacillariophyta	<i>Aulacoseira</i>	21
Y2020_0811_S06	2020_0811	S06	Chlorophyta	<i>Neglectella</i>	80

Sample	Date	Site	Taxon	Genus	Percent
Y2020_0811_S06	2020_0811	S06	Cryptophyta	<i>Cryptomonas</i>	9
Y2020_0811_S07	2020_0811	S07	Chlorophyta	<i>Neglectella</i>	65
Y2020_0811_S07	2020_0811	S07	Chlorophyta	<i>Pandorina</i>	8
Y2020_0811_S09	2020_0811	S09	Bacillariophyta	<i>Aulacoseira</i>	83
Y2020_0811_S09	2020_0811	S09	Chlorophyta	<i>Neglectella</i>	12
Y2020_0811_S10	2020_0811	S10	Bacillariophyta	<i>Aulacoseira</i>	81
Y2020_0811_S10	2020_0811	S10	Chlorophyta	<i>Neglectella</i>	13
Y2020_0814_S01	2020_0814	S01	Chlorophyta	<i>Neglectella</i>	65
Y2020_0814_S01	2020_0814	S01	Cryptophyta	<i>Cryptomonas</i>	16
Y2020_0814_S04	2020_0814	S04	Chlorophyta	<i>Neglectella</i>	56
Y2020_0814_S04	2020_0814	S04	Cryptophyta	<i>Cryptomonas</i>	29
Y2020_0814_S04	2020_0814	S04	Chlorophyta	<i>Pandorina</i>	6
Y2020_0814_S04	2020_0814	S04	Cryptophyta	<i>Komma</i>	5
Y2020_0814_S05	2020_0814	S05	Cryptophyta	<i>Cryptomonas</i>	57
Y2020_0814_S05	2020_0814	S05	Chlorophyta	<i>Neglectella</i>	26
Y2020_0814_S05	2020_0814	S05	Dinoflagellata	<i>Suessiaceae_X</i>	6
Y2020_0814_S06	2020_0814	S06	Chlorophyta	<i>Neglectella</i>	67
Y2020_0814_S06	2020_0814	S06	Cryptophyta	<i>Cryptomonas</i>	24
Y2020_0814_S07	2020_0814	S07	Chlorophyta	<i>Neglectella</i>	46
Y2020_0814_S07	2020_0814	S07	Cryptophyta	<i>Cryptomonas</i>	31
Y2020_0814_S07	2020_0814	S07	Chlorophyta	<i>Sphaeropleales_XX</i>	13
Y2020_0814_S07	2020_0814	S07	Cryptophyta	<i>Komma</i>	7
Y2020_0814_S09	2020_0814	S09	Chlorophyta	<i>Neglectella</i>	26
Y2020_0814_S09	2020_0814	S09	Chlorophyta	<i>Sphaeropleales_XX</i>	23

Sample	Date	Site	Taxon	Genus	Percent
Y2020_0814_S09	2020_0814	S09	Cryptophyta	<i>Cryptomonas</i>	22
Y2020_0818_S01	2020_0818	S01	Chlorophyta	<i>Neglectella</i>	47
Y2020_0818_S01	2020_0818	S01	Bacillariophyta	<i>Aulacoseira</i>	47
Y2020_0818_S04	2020_0818	S04	Bacillariophyta	<i>Aulacoseira</i>	38
Y2020_0818_S04	2020_0818	S04	Chlorophyta	<i>Neglectella</i>	30
Y2020_0818_S04	2020_0818	S04	Cryptophyta	<i>Cryptomonas</i>	7
Y2020_0818_S05	2020_0818	S05	Chlorophyta	<i>Neglectella</i>	85
Y2020_0818_S06	2020_0818	S06	Cryptophyta	<i>Cryptomonas</i>	46
Y2020_0818_S06	2020_0818	S06	Chlorophyta	<i>Neglectella</i>	36
Y2020_0818_S06	2020_0818	S06	Chlorophyta	<i>Pandorina</i>	6
Y2020_0818_S07	2020_0818	S07	Chlorophyta	<i>Neglectella</i>	47
Y2020_0818_S07	2020_0818	S07	Bacillariophyta	<i>Aulacoseira</i>	29
Y2020_0818_S08	2020_0818	S08	Cryptophyta	<i>Cryptomonas</i>	27
Y2020_0818_S08	2020_0818	S08	Chlorophyta	<i>Neglectella</i>	20
Y2020_0818_S08	2020_0818	S08	Chlorophyta	<i>Sphaeropleales_XX</i>	14
Y2020_0818_S08	2020_0818	S08	Cryptophyta	<i>Komma</i>	12
Y2020_0818_S09	2020_0818	S09	Bacillariophyta	<i>Aulacoseira</i>	67
Y2020_0818_S09	2020_0818	S09	Chlorophyta	<i>Neglectella</i>	25
Y2020_0818_S10	2020_0818	S10	Bacillariophyta	<i>Aulacoseira</i>	61
Y2020_0818_S10	2020_0818	S10	Chlorophyta	<i>Neglectella</i>	29
Y2020_0821_S06	2020_0821	S06	Cryptophyta	<i>Cryptomonas</i>	41
Y2020_0821_S06	2020_0821	S06	Chlorophyta	<i>Neglectella</i>	30
Y2020_0821_S07	2020_0821	S07	Chlorophyta	<i>Neglectella</i>	60
Y2020_0821_S07	2020_0821	S07	Cryptophyta	<i>Cryptomonas</i>	26

Sample	Date	Site	Taxon	Genus	Percent
Y2020_0825_S06	2020_0825	S06	Cryptophyta	<i>Cryptomonas</i>	58
Y2020_0825_S06	2020_0825	S06	Chlorophyta	<i>Neglectella</i>	23
Y2020_0825_S07	2020_0825	S07	Cryptophyta	<i>Cryptomonas</i>	43
Y2020_0825_S07	2020_0825	S07	Chlorophyta	<i>Neglectella</i>	37
Y2020_0825_S07	2020_0825	S07	Chlorophyta	<i>Pandorina</i>	7
Y2020_0828_S01	2020_0828	S01	Bacillariophyta	<i>Aulacoseira</i>	66
Y2020_0828_S01	2020_0828	S01	Chlorophyta	<i>Neglectella</i>	21
Y2020_0828_S04	2020_0828	S04	Chlorophyta	<i>Neglectella</i>	38
Y2020_0828_S04	2020_0828	S04	Bacillariophyta	<i>Aulacoseira</i>	28
Y2020_0828_S04	2020_0828	S04	Ciliophora	<i>Rimostrombidium_A</i>	8
Y2020_0828_S04	2020_0828	S04	Cryptophyta	<i>Cryptomonadales_XX</i>	6
Y2020_0828_S04	2020_0828	S04	Cryptophyta	<i>Cryptomonas</i>	5
Y2020_0828_S05	2020_0828	S05	Chlorophyta	<i>Neglectella</i>	44
Y2020_0828_S05	2020_0828	S05	Bacillariophyta	<i>Aulacoseira</i>	36
Y2020_0828_S07	2020_0828	S07	Chlorophyta	<i>Neglectella</i>	47
Y2020_0828_S07	2020_0828	S07	Bacillariophyta	<i>Aulacoseira</i>	14
Y2020_0828_S07	2020_0828	S07	Ciliophora	<i>Rimostrombidium_A</i>	12
Y2020_0828_S08	2020_0828	S08	Chlorophyta	<i>Neglectella</i>	56
Y2020_0828_S08	2020_0828	S08	Cryptophyta	<i>Cryptomonas</i>	18
Y2020_0828_S08	2020_0828	S08	Chlorophyta	<i>Sphaeropleales_XX</i>	6
Y2020_0828_S09	2020_0828	S09	Bacillariophyta	<i>Aulacoseira</i>	66
Y2020_0828_S09	2020_0828	S09	Chlorophyta	<i>Neglectella</i>	23
Y2020_0828_S10	2020_0828	S10	Bacillariophyta	<i>Aulacoseira</i>	55
Y2020_0828_S10	2020_0828	S10	Chlorophyta	<i>Neglectella</i>	25

Sample	Date	Site	Taxon	Genus	Percent
Y2020_0828_S10	2020_0828	S10	Others	<i>Catenophlyctis</i>	11
Y2021_0617_S01	2021_0617	S01	Chlorophyta	<i>Neglectella</i>	37
Y2021_0617_S01	2021_0617	S01	Chlorophyta	<i>Sphaeropleales_XX</i>	25
Y2021_0617_S01	2021_0617	S01	Chlorophyta	<i>Atractomorpha</i>	6
Y2021_0617_S04	2021_0617	S04	Chlorophyta	<i>Neglectella</i>	37
Y2021_0617_S04	2021_0617	S04	Cryptophyta	<i>Cryptomonadales_XX</i>	13
Y2021_0617_S04	2021_0617	S04	Chlorophyta	<i>Sphaeropleales_XX</i>	9
Y2021_0617_S05	2021_0617	S05	Cryptophyta	<i>Cryptomonadales_XX</i>	32
Y2021_0617_S05	2021_0617	S05	Chlorophyta	<i>Neglectella</i>	31
Y2021_0617_S05	2021_0617	S05	Cryptophyta	<i>Cryptomonas</i>	12
Y2021_0617_S05	2021_0617	S05	Chlorophyta	<i>Sphaeropleales_XX</i>	6
Y2021_0617_S06	2021_0617	S06	Chlorophyta	<i>Sphaeropleales_XX</i>	29
Y2021_0617_S06	2021_0617	S06	Chlorophyta	<i>Neglectella</i>	20
Y2021_0617_S06	2021_0617	S06	Chlorophyta	<i>Atractomorpha</i>	18
Y2021_0617_S07	2021_0617	S07	Chlorophyta	<i>Neglectella</i>	46
Y2021_0617_S07	2021_0617	S07	Chlorophyta	<i>Sphaeropleales_XX</i>	8
Y2021_0617_S07	2021_0617	S07	Chrysophyceae	<i>Poteriospumella</i>	7
Y2021_0617_S09	2021_0617	S09	Bacillariophyta	<i>Aulacoseira</i>	32
Y2021_0617_S09	2021_0617	S09	Chlorophyta	<i>Neglectella</i>	16
Y2021_0617_S09	2021_0617	S09	Cryptophyta	<i>Cryptomonas</i>	8
Y2021_0617_S10	2021_0617	S10	Chlorophyta	<i>Neglectella</i>	46
Y2021_0617_S10	2021_0617	S10	Chlorophyta	<i>Sphaeropleales_XX</i>	15
Y2021_0617_S10	2021_0617	S10	Bacillariophyta	<i>Aulacoseira</i>	8
Y2021_0710_S01	2021_0710	S01	Chlorophyta	<i>Neglectella</i>	80

Sample	Date	Site	Taxon	Genus	Percent
Y2021_0710_S04	2021_0710	S04	Chlorophyta	<i>Neglectella</i>	81
Y2021_0710_S04	2021_0710	S04	Dinoflagellata	<i>Ceratium</i>	7
Y2021_0710_S05	2021_0710	S05	Chlorophyta	<i>Neglectella</i>	87
Y2021_0710_S07	2021_0710	S07	Chlorophyta	<i>Neglectella</i>	78
Y2021_0710_S09	2021_0710	S09	Chlorophyta	<i>Neglectella</i>	55
Y2021_0710_S09	2021_0710	S09	Bacillariophyta	<i>Aulacoseira</i>	19
Y2021_0710_S09	2021_0710	S09	Chlorophyta	<i>Sphaeropleales_XX</i>	8
Y2021_0710_S10	2021_0710	S10	Chlorophyta	<i>Neglectella</i>	57
Y2021_0710_S10	2021_0710	S10	Bacillariophyta	<i>Aulacoseira</i>	11
Y2021_0710_S10	2021_0710	S10	Chlorophyta	<i>Sphaeropleales_XX</i>	5
Y2021_0811_S01	2021_0811	S01	Chlorophyta	<i>Neglectella</i>	36
Y2021_0811_S01	2021_0811	S01	Cryptophyta	<i>Cryptomonas</i>	12
Y2021_0811_S01	2021_0811	S01	Metazoa	<i>Daphnia</i>	11
Y2021_0811_S01	2021_0811	S01	Chlorophyta	<i>Hamakko</i>	8
Y2021_0811_S04	2021_0811	S04	Cryptophyta	<i>Cryptomonas</i>	30
Y2021_0811_S04	2021_0811	S04	Chlorophyta	<i>Neglectella</i>	11
Y2021_0811_S04	2021_0811	S04	Chlorophyta	<i>Chlamydomonas</i>	10
Y2021_0811_S04	2021_0811	S04	Others	<i>Katablepharidales_XX</i>	6
Y2021_0811_S05	2021_0811	S05	Chlorophyta	<i>Chlamydomonas</i>	17
Y2021_0811_S05	2021_0811	S05	Cryptophyta	<i>Cryptomonas</i>	10
Y2021_0811_S05	2021_0811	S05	Ciliophora	<i>Spathidium_1</i>	7
Y2021_0811_S05	2021_0811	S05	Parasites	<i>Aphamonas</i>	6
Y2021_0811_S05	2021_0811	S05	Others	<i>Katablepharidales_XX</i>	6
Y2021_0811_S07	2021_0811	S07	Chlorophyta	<i>Chlamydomonas</i>	17

Sample	Date	Site	Taxon	Genus	Percent
Y2021_0811_S07	2021_0811	S07	Chlorophyta	<i>Neglectella</i>	16
Y2021_0811_S07	2021_0811	S07	Cryptophyta	<i>Cryptomonas</i>	7
Y2021_0811_S07	2021_0811	S07	Parasites	<i>Aphamonas</i>	6
Y2021_0811_S09	2021_0811	S09	Chlorophyta	<i>Neglectella</i>	61
Y2021_0811_S09	2021_0811	S09	Chlorophyta	<i>Sphaeropleales_XX</i>	10
Y2021_0811_S09	2021_0811	S09	Cryptophyta	<i>Cryptomonas</i>	7
Y2021_0811_S09	2021_0811	S09	Metazoa	<i>Brachionus</i>	6
Y2021_0920_S01	2021_0920	S01	Cryptophyta	<i>Cryptomonas</i>	33
Y2021_0920_S01	2021_0920	S01	Others	<i>Katablepharidales_XX</i>	21
Y2021_0920_S01	2021_0920	S01	Bacillariophyta	<i>Aulacoseira</i>	5
Y2021_0920_S04	2021_0920	S04	Chlorophyta	<i>Parachlorella</i>	21
Y2021_0920_S04	2021_0920	S04	Cryptophyta	<i>Cryptomonas</i>	12
Y2021_0920_S04	2021_0920	S04	Cryptophyta	<i>Cryptomonadales_XX</i>	11
Y2021_0920_S04	2021_0920	S04	Chlorophyta	<i>Micractinium</i>	7
Y2021_0920_S04	2021_0920	S04	Chlorophyta	<i>Desmodesmus</i>	5
Y2021_0920_S05	2021_0920	S05	Chlorophyta	<i>Parachlorella</i>	23
Y2021_0920_S05	2021_0920	S05	Chlorophyta	<i>Micractinium</i>	12
Y2021_0920_S05	2021_0920	S05	Cryptophyta	<i>Cryptomonadales_XX</i>	9
Y2021_0920_S07	2021_0920	S07	Chlorophyta	<i>Parachlorella</i>	20
Y2021_0920_S07	2021_0920	S07	Ciliophora	<i>Rimostrombidium_A</i>	11
Y2021_0920_S07	2021_0920	S07	Cryptophyta	<i>Cryptomonadales_XX</i>	9
Y2021_0920_S07	2021_0920	S07	Chlorophyta	<i>Micractinium</i>	7
Y2021_0920_S07	2021_0920	S07	Cryptophyta	<i>Cryptomonas</i>	7
Y2021_0920_S07	2021_0920	S07	Others	<i>Katablepharidales_XX</i>	5

Sample	Date	Site	Taxon	Genus	Percent
Y2021_0920_S09	2021_0920	S09	Cryptophyta	<i>Cryptomonas</i>	32
Y2021_0920_S09	2021_0920	S09	Dinoflagellata	<i>Suessiaceae_X</i>	12
Y2021_0920_S09	2021_0920	S09	Chlorophyta	<i>Neglectella</i>	6
Y2021_0920_S09	2021_0920	S09	Others	<i>Paracercomonas</i>	5
Y2021_0920_S10	2021_0920	S10	Others	<i>Paracercomonas</i>	25
Y2021_0920_S10	2021_0920	S10	Chlorophyta	<i>Neglectella</i>	20
Y2021_0920_S10	2021_0920	S10	Cryptophyta	<i>Cryptomonas</i>	7
Y2021_0920_S10	2021_0920	S10	Ciliophora	<i>Vorticella</i>	6
Y2021_1028_S01	2021_1028	S01	Others	<i>Katablepharidales_XX</i>	9
Y2021_1028_S01	2021_1028	S01	Ciliophora	<i>Rimostrombidium_A</i>	7
Y2021_1028_S01	2021_1028	S01	Cryptophyta	<i>Cryptomonas</i>	7
Y2021_1028_S01	2021_1028	S01	Chlorophyta	<i>Micractinium</i>	6
Y2021_1028_S01	2021_1028	S01	Cryptophyta	<i>Cryptomonadales_XX</i>	6
Y2021_1028_S01	2021_1028	S01	Chlorophyta	<i>Parachlorella</i>	5
Y2021_1028_S04	2021_1028	S04	Cryptophyta	<i>Cryptomonas</i>	23
Y2021_1028_S04	2021_1028	S04	Ciliophora	<i>Tintinnidium</i>	12
Y2021_1028_S04	2021_1028	S04	Metazoa	<i>Daphnia</i>	8
Y2021_1028_S04	2021_1028	S04	Bacillariophyta	<i>Aulacoseira</i>	6
Y2021_1028_S05	2021_1028	S05	Cryptophyta	<i>Cryptomonas</i>	21
Y2021_1028_S05	2021_1028	S05	Ciliophora	<i>Tintinnidium</i>	20
Y2021_1028_S05	2021_1028	S05	Dinoflagellata	<i>Suessiaceae_X</i>	5
Y2021_1028_S07	2021_1028	S07	Ciliophora	<i>Tintinnidium</i>	27
Y2021_1028_S07	2021_1028	S07	Cryptophyta	<i>Cryptomonas</i>	14
Y2021_1028_S09	2021_1028	S09	Bacillariophyta	<i>Aulacoseira</i>	78

Sample	Date	Site	Taxon	Genus	Percent
Y2021_1028_S09	2021_1028	S09	Parasites	<i>Perkinsida_XXX</i>	6
Y2021_1028_S10	2021_1028	S10	Bacillariophyta	<i>Aulacoseira</i>	81