

## Diverse and variable community structure of picophytoplankton across the Laurentian Great Lakes

John Gale,<sup>1</sup> Carey Sweeney,<sup>1</sup> Sara Paver,<sup>2</sup> Maureen L. Coleman,<sup>2</sup> Anne W. Thompson <sup>1\*</sup>

<sup>1</sup>Department of Biology, Portland State University, Portland, Oregon, USA

<sup>2</sup>Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois, USA

### Abstract

The Laurentian Great Lakes provide economic support to millions of people, drive biogeochemical cycling, and are an important natural laboratory for characterizing the fundamental components of aquatic ecosystems. Small phytoplankton are important contributors to the food web in much of the Laurentian Great Lakes. Here, for the first time, we reveal and quantify eight phenotypically distinct picophytoplankton populations across the Lakes using a multilaser flow cytometry approach, which distinguishes cells based on their pigment phenotype. The distributions and diversity of picophytoplankton flow populations varied across lakes and depths, with Lake Erie standing out with the highest diversity. By sequencing sorted cells, we identified several distinct lineages of *Synechococcales* spanning Subclusters 5.2 and 5.3. Distinct genotypic clusters mapped to phenotypically similar flow populations, suggesting that there may not be a clear one-to-one mapping between genotypes and phenotypes. This suggests genome-level differentiation between lakes but some degree of phenotypic convergence in pigment characteristics. Our results demonstrate that ecological selection for locally adapted populations may outpace homogenization by physical transport in this interconnected system. Given the reliance of the Lakes on in situ primary production as a source for organic carbon, this work sets the foundation to test how the community structure of small primary producers corresponds to biogeochemical and food web functions of the Great Lakes and other freshwater systems.

The Laurentian Great Lakes provide important and unique perspectives for the study of aquatic microbial ecosystems. The Great Lakes are among the largest freshwater ecosystems on Earth, comprised of five relatively deep lakes, which span large environmental gradients. The health of their waters is central to the welfare and livelihood of 30 million people of two nations who live in the Great Lakes Basin (Dobiesz et al. 2010; Tyner and Boyer 2020). In addition, the lakes provide a natural laboratory for addressing fundamental scientific

questions about biogeochemical cycles, food web ecology, water quality, and microbial evolution. The Great Lakes also face numerous pressures due to global climate change, which have the potential to alter adversely their ecological, social, and economic value (Kehl 2018). To robustly predict the consequences of change to the Lakes system, research must advance understanding of the mechanisms that underlie the Great Lakes' microbial ecosystem.

The Great Lakes are more dependent on in situ primary production by photosynthetic microorganisms (i.e., phytoplankton) than terrestrial inputs of organic carbon due to their immense volume and small catchment area (Sterner 2010, 2021). Large cyanobacteria and eukaryotic phytoplankton are known to contribute to this primary production and their community dynamics have been studied extensively (Barbiero and Tuchman 2001; Twiss et al. 2012; Carrick et al. 2015; Barbiero et al. 2018; U.S. EPA 2019; Reavie et al. 2021). However, relatively little is known about the ecology of the small phytoplankton (i.e., picophytoplankton, < 3 μm in diameter) despite recognition of their importance in Great Lakes primary production. For example, early reports indicated that the picocyanobacterium *Synechococcus* performed 20–40% of primary production in Lakes Huron and Michigan (Fahnenstiel et al. 1986, 1991b) and in separate studies cyanobacteria

\*Correspondence: awt@pdx.edu

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Additional Supporting Information may be found in the online version of this article.

**Author Contribution Statement:** A.W.T. and M.L.C. involved in conceptualization. A.W.T. and M.L.C. involved in resources. A.W.T., J.G., C.S., M.L.C., and S.P. performed methodology. A.W.T., M.L.C., J.G., and S.P. did formal analysis. A.W.T., M.L.C., J.G., C.S., and S.P. involved in investigation. A.W.T., J.G., S.P., and M.L.C. wrote the manuscript. A.W.T., J.G., S.P., and M.L.C. did visualization. A.W.T. performed supervision. A.W.T. and M.L.C. did funding acquisition. Description of author contributions follows the “CRediT” taxonomy (Brand et al. 2015).

comprised 40% of the biomass of small microorganisms in Lake Ontario (Caron et al. 1985) and 10% of the primary productivity in Lakes Huron and Michigan (Fahnenstiel et al. 1991a). However, the dynamics, identity, and community structure of these small cells are poorly understood.

Several lines of evidence from small lakes and the Great Lakes suggest that the freshwater picophytoplankton are diverse and sensitive to environmental gradients and ecosystem change (Fahnenstiel and Carrick 1992; Carrick et al. 2015). Genomes of freshwater picocyanobacteria from around the world demonstrate great functional diversity (Cabello-Yeves et al. 2022), which suggests similar patterns in the Great Lakes. Amplicon sequencing across the Great Lakes has revealed extensive genotypic diversity of cyanobacteria, with multiple coexisting *Synechococcales* variants in each sample (Ivanikova et al. 2008; Paver et al. 2020). Moreover, some variants appear specialized to particular lakes or environmental regimes, suggesting genetic and likely phenotypic divergence that enables niche partitioning. In other systems, several *Synechococcus* pigment types have been shown to co-occur, which suggests niche partitioning across spectral niches in both oceanic and lake waters and is fundamental to photosynthesis and primary production of picocyanobacteria (Palenik 2001; Six et al. 2007; Haverkamp et al. 2009; Callieri 2017; Grébert et al. 2018; Cabello-Yeves et al. 2022). These studies strongly suggest that picophytoplankton would niche partition the Great Lakes, and in doing so, impact primary production, trophic structure, and water quality (Munawar et al. 1994).

A limitation of datasets from previous studies of picophytoplankton ecology of the Great Lakes and other smaller freshwater systems is their compositional nature, which is due to use of sequence-based and culture-based approaches. Thus, freshwater ecosystems, particularly in the Great Lakes, still lack a systematic and quantitative picture of how abundance, phenotypic, genomic, and functional diversity of picophytoplankton shift across space and time. Given the important role of picophytoplankton in primary production, detailed investigation of the picophytoplankton component of the microbial community is timely and critical for advancing understanding of the patterns of primary production that underlie the Lakes' social, biogeochemical, and economic value.

In this study we used multilaser flow cytometry coupled to whole genome sequencing to identify and enumerate several distinct picophytoplankton phenotypes ("flow phenotypes") from complex natural communities across the environmental gradients of the Laurentian Great Lakes. This approach revealed variety across the Lakes in their picocyanobacterial and picophytoplankton diversity and community structure based on phenotype and genotype. Future work will leverage this approach to examine the ecology of diverse small phytoplankton populations across seasons, fine-scale environmental gradients, and in the context of biogeochemical function and food web structure of the Great Lakes and other freshwater bodies.

## Materials and methods

### Sample collection

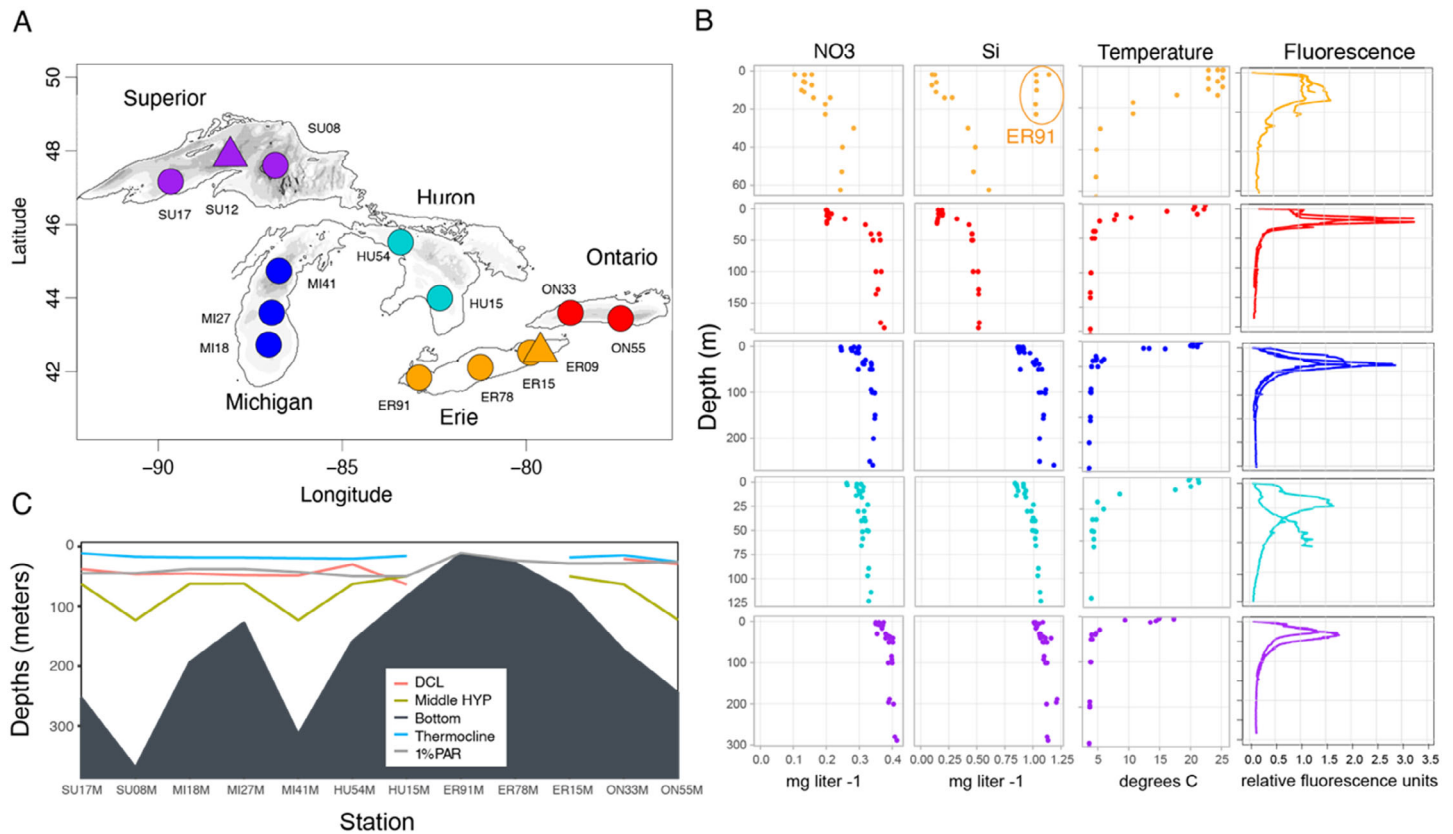
Water samples were collected across the five Laurentian Great Lakes (Fig. 1A) in August 2019 on board the R/V *Lake Guardian* during the U.S. Environmental Protection Agency's (EPA) Summer Survey cruise. Multiple stations and depths were sampled by CTD rosette, which also recorded temperature, chlorophyll fluorescence, dissolved oxygen, and turbidity during descent (Fig. 1B,C, Supporting Information Table S1). Depths included surface, the deep chlorophyll layer (DCL) when present, water below the DCL but above the bottom (i.e., mid-hypolimnion (HYP), depths ranged from 50 to 100 m deep), and water at or 10 m above the bottom ("bottom"). In cases where a DCL was not apparent (Erie only), the lower epilimnion or upper HYP were sampled, and these depths were considered with surface or HYP samples, in further analyses, respectively. Water samples for flow cytometry were preserved by fixation with 0.125% electron microscopy grade glutaraldehyde (Electron microscopy Sciences, Hatfield, PA) then flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until analysis. Additional water samples were collected for flow sorting of specific cells. Samples for flow sorting were from Superior Station (Sta.) SU12 (47.85611 N, 88.04194 W) and ER09 (42.538333 N, 79.616667 W), also in August 2019. This water was prefiltered with a  $2.7\ \mu\text{m}$  filter to remove larger particles then concentrated on a  $0.45\ \mu\text{m}$  filter from 1 L down to 10 mL. The concentrated samples were fixed and stored as above.

### Environmental data

The U.S. EPA's Great Lakes National Program Office analyzed water chemistry from each sample to measure total oxidized nitrogen, dissolved silicate, and chlorophyll *a* (Chl *a*) concentration, according to standard protocols. These data can be accessed through the Great Lakes Environmental Database system (GLENDa: available at <https://cdx.epa.gov/>) and is presented for each of our samples in Fig. 1B. Great Lakes bathymetry data were obtained from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI; Fig. 1A,C).

### Flow cytometry

A BD Influx cell sorter (BD Biosciences, San Jose, CA, USA) equipped with three lasers, an  $80\ \mu\text{m}$  diameter nozzle, and a small particle detector was used for flow cytometric analysis. We ran Biosure Sheath Solution (BioSure, Grass Valley, CA) as the sheath fluid. Weighing replicate sample tubes before and after sampling allowed us to calculate sample flow rates during instrument calibration each day. Sample flow rates were around  $0.5\ \mu\text{L}$  per second. The analyzed volume of each sample was then calculated from the measured flow rate of the instrument each day and the time each sample was run, resulting in counts per volume. Cytometer optics configurations are displayed in



**Fig. 1.** (A) Station map with bathymetry data for the Laurentian Great Lakes study system. Stations shown with triangles are sources of sorted cells (ER09 and SU12). (B) Summer 2019 environmental data over depth including oxidized nitrogen (NO<sub>2</sub> and NO<sub>3</sub>), silicate (Si), temperature, and fluorescence. Note different scales on the y-axes. Elevated Si concentrations at Sta. ER91 are circled in orange. (C) Bathymetry relative to sampling depths for each lake and station with DCL, hypolimnion (HYP), 1% surface PAR. PAR data are from Scofield et al. (2020).

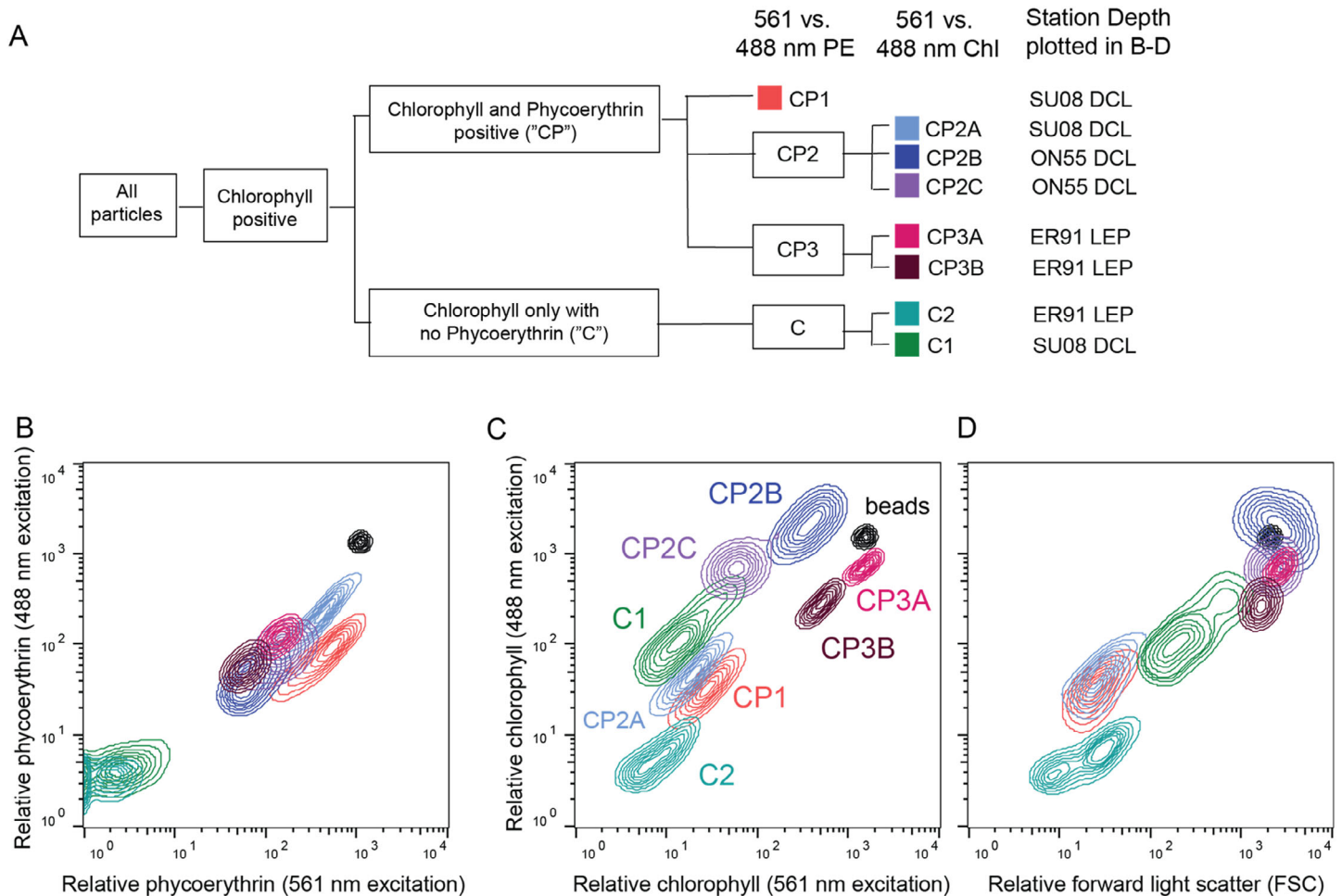
Supporting Information Table S2, following a previously published configuration (Thompson and van den Eng 2016). Data collection and cell sorting were triggered on forward light scatter (FSC) from the 488 nm laser using BD Software (BD Biosciences, San Jose, CA, USA). Chl *a* fluorescence (detected with 692/40 nm bandpass filters) and phycoerythrin fluorescence (detected with 572/27 nm bandpass filters) were collected for each event (i.e., cell) and each laser, except the 640 nm laser for which only Chl fluorescence was collected. UltraRainbow (Spherotech, Lake Forest, IL) fluorescent beads (3.8  $\mu$ m in diameter) provided laser alignment and internal reference for gating. Raw flow cytometry data files are available at [FlowRepository.org](https://flowrepository.org) (FR-FCM-Z2UB).

To identify populations genetically, cells from a subset of populations were flow sorted using the 1.0 Drop Pure sort setting in BD Software. From each population, between 5000 and 10,000 cells were sorted in replicates (Supporting Information Table S3) and stored at  $-80^{\circ}\text{C}$  prior to preparation for metagenomic sequencing via multiple displacement amplification (MDA) followed by Illumina sequencing (see details below) or *petB* PCR.

### Flow cytometry gating strategy

Gating for quantitative analysis, identification of cell populations, statistical tests for differences between populations, and flow cytometry graphics were completed in *FlowJo* (BD Biosciences, San Jose, CA, USA). To illustrate this strategy, we present a graphic representation of the gating hierarchy (Fig. 2A), gating of an example sample to show gates relative to populations (Supporting Information Fig. S1), a compilation of all detected populations relative to each other (Fig. 2B–D), dot plots of all detected particles colored by their membership to the population gates (Supporting Information Figs. S2–S6), dot plots of the concentrated samples used for cell sorting and the sort gates (Supporting Information Fig. S7), and examples of difficult-to-gate populations (Supporting Information Fig. S8).

The flow cytometry gates were created and named as follows. First, chlorophyll-positive cells were determined by bivariate plots of FSC and chlorophyll (excited by the 488 nm laser; Supporting Information Fig. 1A). Chlorophyll (C)-positive cells were those with chlorophyll fluorescence greater than noise. Next, the chlorophyll positive cells were distinguished as phycoerythrin (P)-positive (designated “CP”



**Fig. 2.** Picophytoplankton populations revealed by multilaser flow cytometry. **(A)** Hierarchical flow cytometry gating strategy, relationships between detected populations, and lake/station source for representative populations displayed in **(B–D)**. An example of gating on all events from one sample is presented in Supporting Information Fig. S1. Dot plots of all events from all samples, with defined populations indicated are presented for all samples in Supporting Information Figs. S2–S7. Panels B–D are different displays of overlays of unique populations from multiple samples (source indicated in A). **(B)** Relative phycoerythrin fluorescence of detected populations excited by yellow-green (561 nm, x-axis) and blue (488 nm, y-axis) lasers. **(C)** Relative chlorophyll fluorescence of populations excited by yellow-green (561 nm, x-axis) and blue (488 nm, y-axis) lasers. **(D)** FSC (x-axis) vs. relative chlorophyll fluorescence (488 nm excited, y-axis) for all detected populations relative to standard 3.8  $\mu\text{m}$  beads (black). Note that in some samples, CP1 cells formed subpopulations on bivariate plots of chlorophyll vs. phycoerythrin, which are illustrated in Supporting Information Figs. S7, S8.

populations) or phycoerythrin-negative (designated “C” populations) based on the presence of orange fluorescence at levels greater than noise in bivariate plots of phycoerythrin excited by the green (561 nm) and blue (488 nm) lasers (Supporting Information Fig. S1B). We note that the naming of flow cytometry populations is a challenging endeavor that has received concerted attention in marine systems (Thyssen et al. 2022). In the Great Lakes system, we have discovered many more flow cytometry populations than in typical marine systems. Thus, we chose a simple naming scheme (C vs. CP cells), that does not include assumptions on the taxonomy of cells or their size. As more research groups identify the breadth of these freshwater populations with flow cytometry, the community will need a collaborative and

holistic effort to name these new populations based on available data.

Distinct, non-overlapping populations of CP cells were then gated on bivariate plots of phycoerythrin excited by the blue (488 nm) and green (561 nm) lasers. We examined these populations for subpopulations using bivariate plots of chlorophyll excited by the blue (488 nm) and green (561 nm) lasers (Supporting Information Figs. S1C–S1E). Population counts were set to zero when a distinct cluster of cells in the gate area was not discerned. Population parameters were compared with Chi Squared tests  $T(x)$ , performed in *FlowJo* with *Compare Populations*, which is nonparametric with respect to the distribution of the cells or particles. Statistics are reported for each comparison in the results.



### Statistical analysis

We used R (version 1.3.959) for data analysis and graphical outputs. We used previously published scripts for maps, bathymetry, and environmental data display (Paver et al. 2020). Diversity analyses were done with *vegan* (Oksanen et al. 2019) with reference to microbial ecology best practices by GUSTAME (Buttigieg and Ramette 2014). Within non-metric multidimensional scaling (NMDS) ordination, environmental drivers analysis was done with *vegan::envfit()* (Oksanen et al. 2019) with strengths and statistical significances of each environmental driver proportional to the arrowheads' directions and vector lengths. Abundance and distribution of the flow populations was visualized with a heatmap using the R package *pheatmap* (version 1.0.12).

### Cloning and sequencing the marker gene *petB* from sorted cells

We cloned and sequenced the marker gene cytochrome b6 (*petB*), a subunit of the cytochrome b6-f complex, from the sorted populations, using the TOPO TA cloning for sequencing kit manufacturer's protocol (Invitrogen, Waltham, MA). The gene *petB* was chosen because it is a single copy gene that offers better resolution than 16S rRNA and has been used extensively as a marker for phylogenetic diversity (Farrant et al. 2016; Six et al. 2021). In addition, *petB* reflects the phylogeny of the *Synechococcus* core genome (Grébert et al. 2022). Primers were degenerate and designed to capture a range of picocyanobacteria. We took a nested PCR approach, first amplifying sorted cells with "external" *petB* primers then using this product in another round of PCR with "internal" *petB* primers (Supporting Information Table S4). Cells for *petB* analysis were not amplified using MDA, as below for the cells used to generate metagenomes. The yielded PCR products were quantified fluorometrically with the Qubit (Invitrogen) and the correct product size was confirmed with the Bioanalyzer (Agilent DNA 7500 kit) before cloning and sequencing. Sequences were identified by BLASTn against a database of freshwater picocyanobacteria whole genome sequences (Cabello-Yeves et al. 2022; Supporting Information Table S5). The phylogeny of the *petB* sequences was analyzed in the context of select sequences from known freshwater, marine, and brackish picocyanobacteria. We used Molecular Evolutionary Genetics Analysis (MEGA v10.1.8) (Stecher et al. 2020) to align sequences with ClustalW and construct maximum likelihood trees with the Hasegawa-Kishino-Yano model and 1000 bootstrap replicates. The phylogenetic tree was visualized and annotated with the Interactive Tree of Life (version 4; Letunic and Bork 2019).

### MDA and metagenome analysis

To confirm and further identify the diversity and phylogeny of the flow populations, sorted cells from a subset of flow populations were whole genome amplified with the REPLI-g Single Cell Kit (QIAGEN, Hilden, Germany). The manufacturer's

protocol was followed, with a 5-h amplification step. Amplified DNA (1 ng) was prepared for sequencing with the NexteraXT DNA library preparation kit (Illumina, San Diego, CA) and sequenced on the Illumina NextSeq 2000 (2 × 150 bp). We amplified this genomic material to ensure plenty of material for sequencing and for consistency across samples, where different numbers of cells were sorted.

Metagenomic analysis of sequences from sorted cells was performed with *anvi'o* version 7 (Eren et al. 2021). Paired-end reads were quality filtered using *iu-filter-quality-minoche* with default parameters. Assembly was performed with MEGAHIT with the minimum contig size set to 1000 bp (Li et al. 2015). Read normalization using *bbnorm* ([sourceforge.net/projects/bbmap/](https://sourceforge.net/projects/bbmap/)) did not substantially affect the assemblies. Reads were mapped to the assembled contigs with Bowtie 2 (Langmead and Salzberg 2012). Sequence data are available in NCBI's sequence read archive (PRJNA917635). Predicted gene sequences for single-copy bacterial core genes and ribosomal RNA genes (16S and 18S) were recovered from the contigs using the *anvi'o* programs *anvi-gen-contigs-database*, *anvi-run-hmms*, and *anvi-get-sequences-for-hmm-hits*. Recovered 16S and 18S rRNA gene sequences were searched by BLASTn against NCBI-Genbank nt database and a database of freshwater picocyanobacteria whole genome sequences (Cabello-Yeves et al. 2022). Assembled contigs were manually binned in *anvi'o*.

For cyanobacterial populations, a phylogenomic tree was constructed using 22 predicted ribosomal protein sequences, extracted from both the assembled sequence bins and reference genomes (Cabello-Yeves et al. 2022). Each protein was individually aligned using *muscle* v3.8 (Edgar 2004) and a maximum likelihood tree was estimated using *PhyML* v3.3.20190909 (Guindon et al. 2010). The 22 individual trees were combined using the function *mrp.supertree* in the R package *phytools* (Revell 2011). Four supertrees with identical parsimony scores were identified; their consensus is shown in Fig. 8.

Two sorted metagenomes did not yield good assemblies for cyanobacteria and appeared to contain many eukaryotic sequences. To analyze these metagenomes, contigs were classified as prokaryote or eukaryote using the program *Whokaryote* (Pronk and Medema 2022). Gene prediction for putative eukaryote contigs was carried out using *Augustus* (Hoff and Stanke 2013) trained on the picoeukaryote *Bathycoccus prasinus*. Prokaryote gene prediction was performed with *Prodigal* (Hyatt et al. 2010). All predicted eukaryote and prokaryote proteins were searched against the nr database using *BLASTp* to identify similar sequences.

## Results

### Environmental gradients

Twelve stations across the five Great Lakes were sampled for environmental data and small phytoplankton surveys (Fig. 1A). At nearly every station, the water column was

stratified during the sampling period in August 2019, as indicated by temperature profiles (Fig. 1B). The thermocline depth ranged from 10 to 30 m except in the shallowest Erie stations where no stratification was evident (Fig. 1C). Deep chlorophyll layers were evident between 29.9 and 51.2 m in most lakes but did not exist for Erie stations (Fig. 1B,C). Surface temperatures were warmest in Lake Erie and coolest in Lake Superior. Below the mixed layer, Erie temperatures were warmest ( $\sim 5^{\circ}\text{C}$ ), while other lakes converged near  $\sim 3^{\circ}\text{C}$ . Nitrate and silicate concentrations were depleted in the epilimnion relative to the HYP for all lakes. Upper lakes (Huron, Michigan, and Superior) displayed higher silicate concentrations than the lower lakes (Ontario and Erie), except for the western Erie Sta. (ER91M) that matched upper lake silicate levels. Surface nitrate concentrations were highest in Superior, comparable for Huron and Michigan, lower in Ontario, and lowest in Erie. These values were consistent with environmental conditions in the Great Lakes over the past two decades (Scofield et al. 2020).

### Identification of unique picophytoplankton pigment phenotypes

Hierarchical gating of flow cytometry data revealed eight distinct populations of small chlorophyll-containing cells across the five Laurentian Great Lakes. These chlorophyll-positive cells accounted for 0–50% of the total particles detected by forward scattered light in each sample. The positions of the unique populations in bivariate plots of pigment fluorescence and size are presented in Fig. 2 and Supporting Information Figs. S2–S6. We call these populations “flow phenotypes” as the populations contain cells with similar size, shape, and pigment fluorescence properties (i.e., their phenotype as detected by flow cytometry). These patterns are similar to “needles,” which we previously defined as groups of cells with similar ratios of chlorophyll (or phycoerythrin) fluorescence excited by different wavelengths of light (Thompson and van den Engh 2016). In some cases, subpopulations were easily defined within the flow phenotypes, but in other cases subpopulations were difficult to define due to overlapping distributions (Supporting Information Fig. S8). Detailed description of the flow cytometry and pigment characteristics of each population are as follows.

Two flow phenotypes contained chlorophyll but no phycoerythrin (designated “C” populations) (Fig. 2, Supporting Information Fig. 1B). These two populations (C1 and C2) displayed parallel elongated distributions in bivariate plots of chlorophyll excited by blue (488 nm) and green (561 nm) lasers (Fig. 2C, Supporting Information Fig. S1E). C1 chlorophyll fluorescence was about two orders of magnitude greater than C2 fluorescence, which was partially within the instrument noise in the chlorophyll detectors.

Six flow phenotypes contained chlorophyll and phycoerythrin (designated “CP” populations) (Fig. 2, Supporting Information Fig. S1B–S1D). CP1 was an elongated population when

visualized and gated on a bivariate plot of phycoerythrin excited by blue (488 nm) and green (561 nm) lasers (Fig. 2, Supporting Information Fig. S1B). Viewing this population on a bivariate plot of chlorophyll excited by blue and green lasers (Fig. 2C) did not reveal any subpopulations. However, we did see subpopulations within CP1 in bivariate plots of chlorophyll and phycoerythrin in some samples, but the overlap of these subpopulations prevented us from further definitive gating across the whole sample set. Examples of samples with these ambiguous subpopulations are presented in Supporting Information Fig. S8 and for Sta. ER09, which was sorted (Supporting Information Fig. S7). CP2, like CP1, was also an elongated population when visualized on a bivariate plot of phycoerythrin excited by blue and green lasers (Fig. 2B, Supporting Information Fig. S1B). For CP2, bivariate plots of chlorophyll excited by blue and green lasers revealed distinct subpopulations, which we designated CP2A, CP2B, and CP2C (Supporting Information Fig. S1C). Compared to subpopulation CP2A, the subpopulations CP2B and CP2C had lower phycoerythrin fluorescence (Fig. 2B) and higher chlorophyll fluorescence (Fig. 2C). CP3 (unlike CP1 and CP2) did not display an elongated shape on the phycoerythrin bivariate plot (Fig. 2B) but included two distinct adjacent subpopulations, designated CP3A and CP3B, when viewed in the chlorophyll bivariate plot (Fig. 2C). Overall, these patterns are similar to those seen using multilaser flow cytometry to analyze *Synechococcus* of different pigment types and picoeukaryotes of different pigment composition and/or genotype (Thompson and van den Engh 2016).

We used forward scatter normalized to standard beads to reveal that the flow phenotypes ranged in cell size (Fig. 2D). CP1, CP2A, and C2 cells were relatively small, about  $1\ \mu\text{m}$  in diameter, with no significant difference between them. However, C2 was bimodal in its forward scatter distribution, suggesting a distinct subpopulation of smaller cells within it, which may be a result of the population overlapping with noise in the chlorophyll channel, and thus may include small heterotrophic cells along with the chlorophyll-containing cells. The other populations (C1, CP2B, CP2C, CP3A, and CP3B), were larger, at  $\sim 2\text{--}3\ \mu\text{m}$  in diameter.

### Signs of photoacclimation to diminishing light with increasing depth

We chose two of the most abundant flow phenotypes in the dataset (CP1 and CP2A) to examine changes in fluorescence intensity (i.e., pigment concentration) with diminishing light over depth at one well-stratified station (SU08M). For both flow phenotypes, we detected increasing chlorophyll and phycoerythrin fluorescence with increasing depth (Fig. 3). For both flow phenotypes, cells from the surface had lower chlorophyll fluorescence than cells from the DCL and cells from 100 m ( $T(x) > 4$ , corresponding to  $p\text{-value} < 0.01$ ) (Supporting Information Table S6). For both flow phenotypes, cells from the DCL and 100 m had similar chlorophyll fluorescence

( $T(x) < 4$ , corresponding to  $p$ -value  $> 0.01$ ; Supporting Information Table S6). For both flow phenotypes, cells from all depths were significantly different from each other in their phycoerythrin fluorescence ( $T(x) > 4$ , corresponding to  $p$ -value  $< 0.01$ ; Supporting Information Table S6). This pattern of increasing pigment concentration with increasing depth is similar to well-documented patterns of phytoplankton photoacclimation to low light with increasing depth in stratified marine systems (Dusenberry et al. 2000). Cultivation of these cells under a range of light availabilities would help test this hypothesis.

### Microcolony formation in some flow phenotypes

We investigated whether the elongated shape of some flow phenotypes' cell distributions (e.g., CP1) could be the result of a continuum of particle sizes, yielding a large range of fluorescence intensities with the same ratio of chlorophyll or phycoerythrin fluorescence excited by 488 and 561 nm lasers. We tested whether brighter particles (i.e., higher chlorophyll fluorescence) had higher forward scatter signals (i.e., cell size) than dimmer particles (i.e., lower chlorophyll fluorescence; Supporting Information Fig. S9). We found that brighter particles were significantly larger than dimmer particles ( $T(x) > 4$ , corresponding to  $p$ -value  $< 0.01$ ). This result is consistent with previous studies that identified flow cytometry signatures of *Synechococcus* microcolony formation as larger particles with fluorescence signatures consistent with single cells (Callieri 2010).

### Abundance and richness varied across lakes and depth

To examine how picophytoplankton abundances varied within the Lakes system, we counted total picophytoplankton across stations from each lake and at several depths across the water column (Figs. 1A, 4). Erie picophytoplankton abundances were higher than all other lakes (near significant Kruskal-Wallis  $p$ -value = 0.063, Wilcoxon pairwise comparisons  $p$ -value  $< 0.05$ ). The high total picophytoplankton for Erie derives from high abundances at most depths, including bottom waters at western and central stations (Fig. 4C), which are within the photic zone (defined by 1% surface photosynthetically active radiation, PAR) due to Erie's shallow depth (Fig. 1C). No other between-lake differences were significant.

Picophytoplankton median abundance was also significantly different across depths (Kruskal-Wallis  $p$ -value = 0.00087). Pairwise comparisons showed significant differences between the DCL depths and underlying waters and near bottom waters ( $p$ -values  $< 0.01$ ). The DCL median picophytoplankton abundances exceeded the surface but was not significant. Overall, most picophytoplankton in our study came from the DCL (38.6% of total) and surface (27.9%) depths. The maximum total picophytoplankton abundance from a single sample was  $2.45 \times 10^5$  cells/mL from the Michigan Sta. MI18M DCL (Fig. 4A,B, black outliers), but the

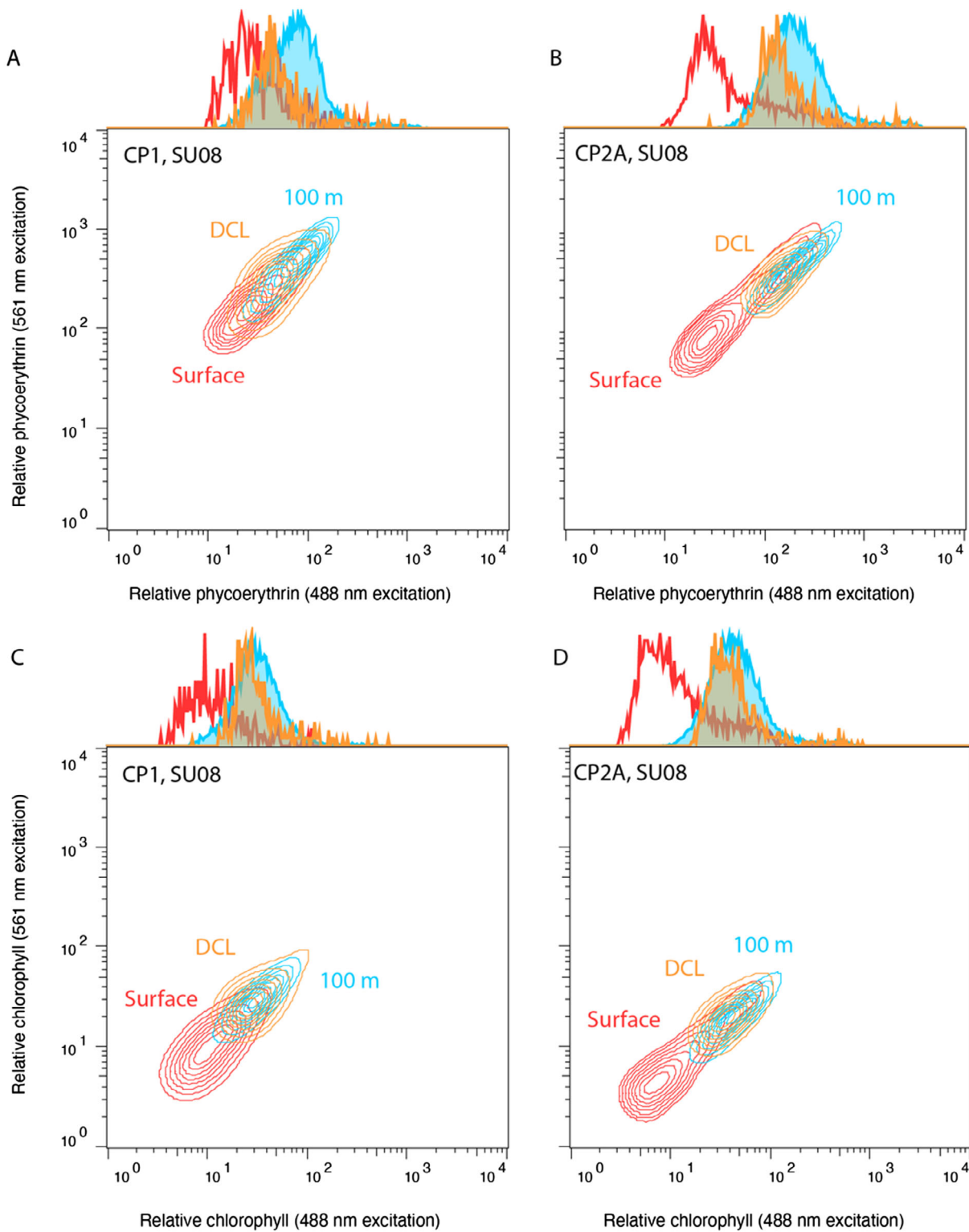
majority of high picophytoplankton counts were in Erie (Fig. 4C). Overall, picophytoplankton abundances were similar to previous works that quantified small chlorophyll-containing cells in the Great Lakes (Fahnenstiel et al. 1986, 1991b; Fahnenstiel and Carrick 1992; Ivanikova et al. 2007).

To test for differences in the community structure of the picophytoplankton flow phenotypes across the Lakes system, we examined median richness (alpha diversity) at each depth and station. Richness differed across depths (Kruskal-Wallis analysis  $p$ -value = 0.00017), with communities equally rich in surface and DCL samples but each of these exceeded richness of underlying waters ( $p$ -values  $< 0.01$ , Wilcoxon test; Fig. 4E). In contrast, median richness for each lake did not vary significantly (Kruskal-Wallis  $p$ -value = 0.75; Fig. 4D). However, richness of individual samples did vary. In the upper lakes, richness was uniform as most samples (16 of the 28) from Superior, Michigan, and Huron had the same three picophytoplankton flow phenotypes. In contrast, richness of individual samples from Erie and Ontario ranged from one to five flow phenotypes in Erie and from one to six flow phenotypes in Ontario. Total richness was highest in Ontario, when combining unique flow phenotypes from all depths (Fig. 4F).

### Dynamic picophytoplankton community structure across lakes and depths

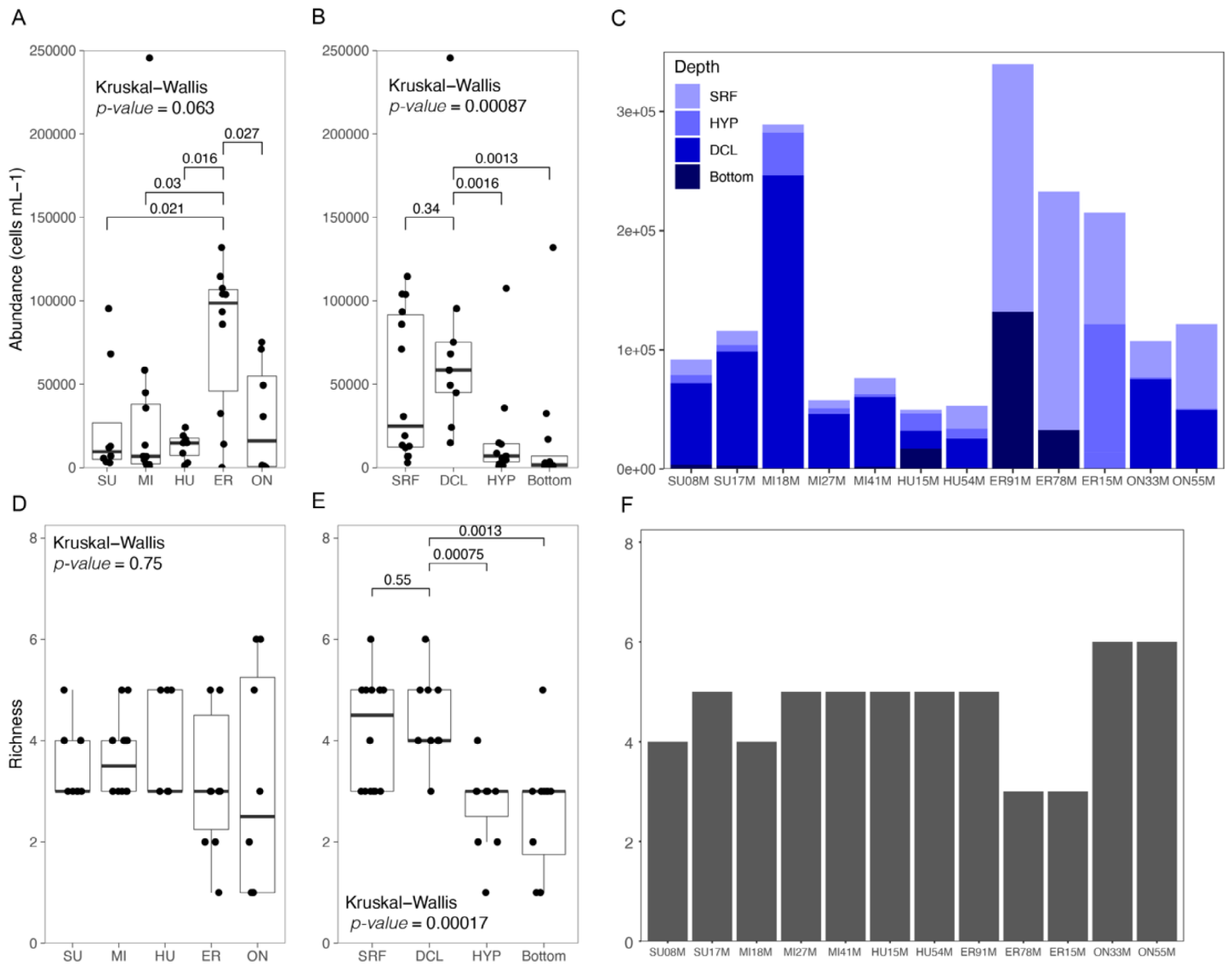
To further examine picophytoplankton community structure across lakes and depths we analyzed beta diversity with Bray-Curtis dissimilarity and NMDS visualization (Fig. 5). Beta diversity varied significantly by both depth (ANOSIM significance  $p$ -value = 0.002) and lake (ANOSIM significance  $p$ -value = 0.001) though the relationship strengths were modest (ANOSIM statistic  $R = 0.196$  for depth and 0.243 for lake). NMDS ordination showed that the Lake Erie samples grouped together separately from upper lake samples (Fig. 5, upper right vs. lower left respectively). Within the Lake Erie group, the western ER91M station samples formed a tight group compared to other samples from that lake. A single outlying Erie sample was ER15M bottom, which had a single low abundance flow phenotype. Lake Ontario samples aligned closely with upper lake samples, except for eastern Ontario ON55M surface, which grouped with Erie. Considering depth, NMDS revealed samples from the upper lakes and Ontario to cluster by depth (i.e., surface and DCL samples were more like each other than those from deeper in the water column). The same pattern was not apparent for Lake Erie as samples from multiple depths clustered together.

To identify environmental factors associated with the NMDS ordination patterns, we examined correlations between microbial community structure and environmental variables. Nitrite/Nitrate (NN) correlated most strongly to community structure ( $r^2 = 0.626$ ,  $p$ -value = 0.001) and vector direction indicated positive correlation with upper lake samples, which is consistent with relatively high nitrate/nitrite levels in the



**Fig. 3.** Photoacclimation of populations CP1 (A, C) and CP2A (B, D) with increasing depth shown by increasing relative phycoerythrin (A, B) and relative chlorophyll (C, D) fluorescence from surface to the DCL to 100 m depth. Each plot shows contour plots (10%) for the cell populations. Histograms are normalized to total cell count. Results from statistical comparison of the populations are presented as Chi Squared  $T(x)$  values, implemented in *FlowJo*, in Supporting Information Table S6.





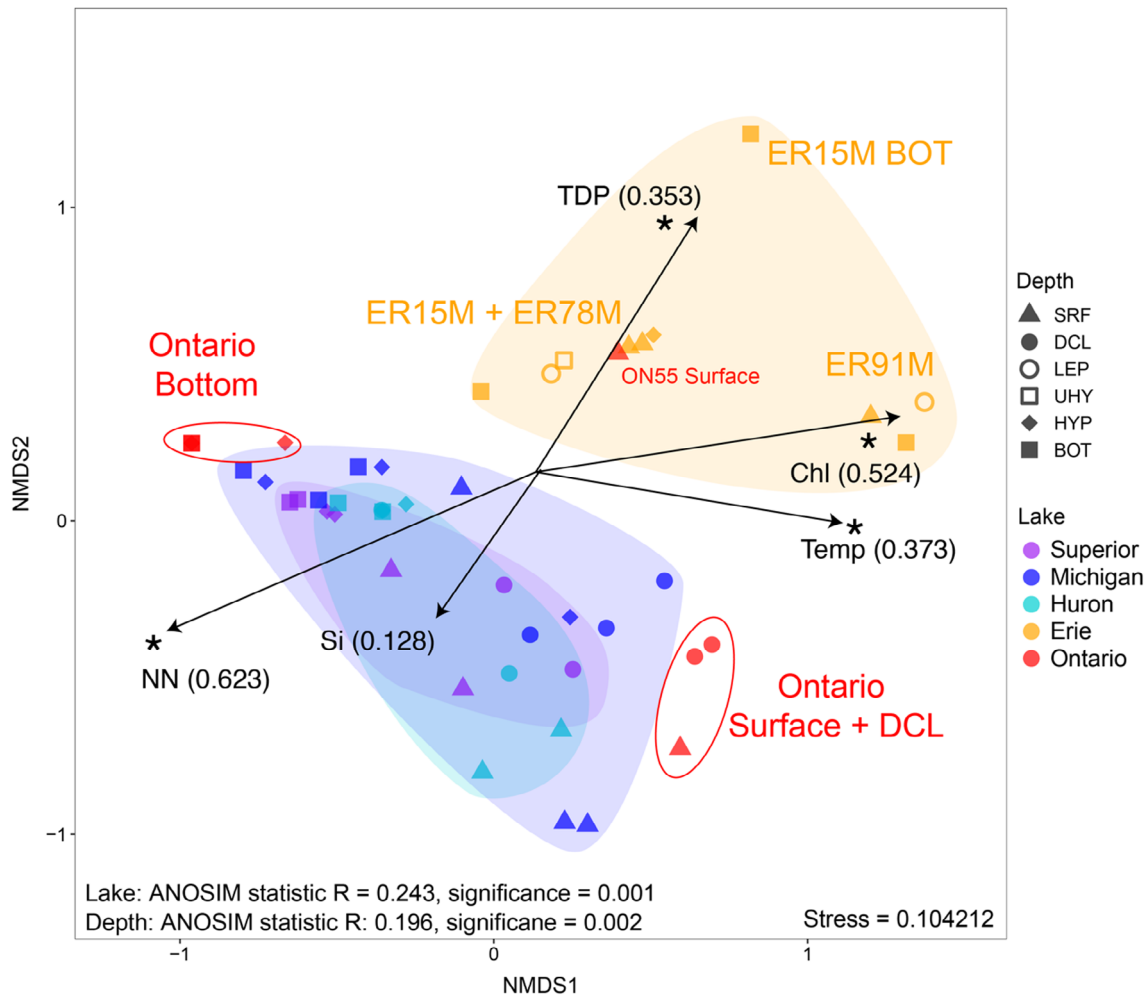
**Fig. 4.** Abundance and diversity of picophytoplankton (i.e., chlorophyll containing cells) in all samples, stations, and lakes. **(A)** Cell abundances for all depths and stations in each lake. **(B)** Cell abundances for each depth bin. **(C)** Cumulative abundances for each station, colored by depth (legend inset). **(D)** Alpha diversity analysis by lake. **(E)** Alpha diversity analysis by depth. **(F)** Cumulative richness (i.e., number of different flow phenotypes) for all depths from each station. Statistical strengths shown by Kruskal-Wallis overall *p*-value and Wilcoxon pairwise statistical *p*-values (brackets). Lower epilimnion data from Erie is included in surface bins. Upper HYP data from Erie are included in HYP bins. Abbreviations: surface (SRF), DCL, mid HYP, bottom or near bottom (bottom).

upper lakes (Fig. 1B). Temperature ( $r^2 = 0.373$ ;  $p$ -value = 0.00) and total dissolved phosphorus ( $r^2 = 0.353$ ,  $p$ -value = 0.0) were the second and third strongest correlations. Vector directions toward lower lake communities were consistent with warmer temperatures and phosphorus eutrophy in the lower lakes (Fig. 1B). Silicate ( $r^2 = 0.128$ ,  $p$ -value 0.05) was also significantly, but less strongly, correlated to the beta diversity pattern.

To investigate the overall distribution of each flow phenotype, we examined their abundances at each station and depth

(Fig. 6). We recognized four categories of flow populations: those present in almost all samples (“Ubiquitous”), those present mostly in upper lakes and Ontario, those present in the lower lakes only, and those present at Sta. ER91M only. For an alternate visualization of each population’s distribution across stations see Supporting Information Fig. S10.

Ubiquitous flow phenotypes C1 and CP1 were abundant in almost all samples from every lake. However, these two populations were distinct in the details of their abundance distributions across the lakes (Fig. 6). CP1 was most abundant in



**Fig. 5.** NMDS visualization of Bray–Curtis dissimilarity analysis based on the cell count of each flow population in each station and depth. The correlation of environmental parameters with the NMDS ordination are shown with arrows. The arrow length is scaled to the strength of the correlation (square root of the  $R^2$ , shown in parentheses) and significant correlations are denoted with asterisks (\*). Depths are distinguished by symbol shape. Lakes are distinguished by symbol color. Clusters of samples from each lake are marked by ellipses, which were manually drawn. ANOSIM statistics are noted for tests by lake and by depth. Stress of the NMDS ordination is reported on the plot. Notes: Three data points overlap at the position of the Ontario bottom (square) sample, including ON33M bottom and ON55M HYP and bottom samples. Abbreviations: SRF, surface; DCL, deep chlorophyll layer; UHY, upper hypolimnion (Erie only); LEP, lower epilimnion (Erie only); mid-HYP, mid-hypolimnion; BOT, bottom.

the lower lakes (72.5% of the total CP1 cells in the dataset) while C1 was more abundant in the upper lakes (80.5% of the total C1 cells in the dataset). C1 was detected in all samples except ER15M bottom, while CP1 was detected in all samples except depths below the DCL in Lake Ontario.

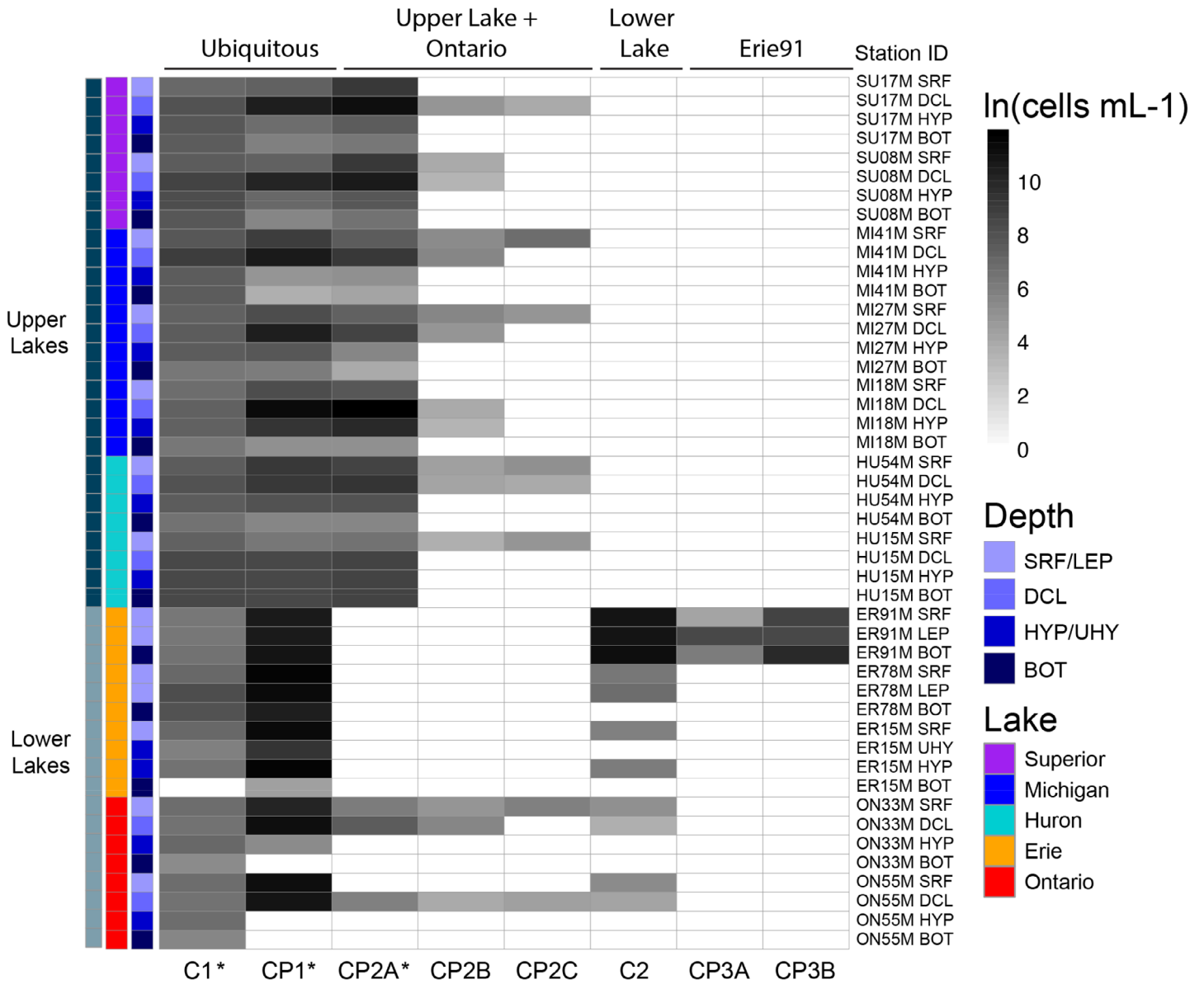
CP2A, CP2B, and CP2C were present at some stations in all lakes we surveyed except Erie (Fig. 6). However, as an exception, CP2C was found at one Erie station (ER09), which we used to sort that population for genomic analysis (see next section; Supporting Information Fig. S7, Table S3). All CP2 populations favored surface and DCL layers of the water column. CP2B and CP2C were almost exclusively found in surface and DCL samples (98.3% and 100%, respectively, of cells from the whole dataset). CP2A also favored surface and

DCL depths (88.8% of CP2A cells from the whole dataset) but was also present in low abundance at all depths in upper lakes. C2 was lower lake-specific, occurring only in Lakes Erie and Ontario.

Two flow populations, CP3A and CP3B, were ER91M-specific as they occurred in Erie only at its westernmost ER91M station.

#### Phylogenetic identity of selected flow phenotypes

Next, we connected the flow phenotypes to specific picophytoplankton genotypes through two sequencing approaches applied to sorted cells from selected flow phenotypes. Expecting to recover some picocyanobacterial sequences, we amplified and sequenced the target gene *petB*, which has been used as a

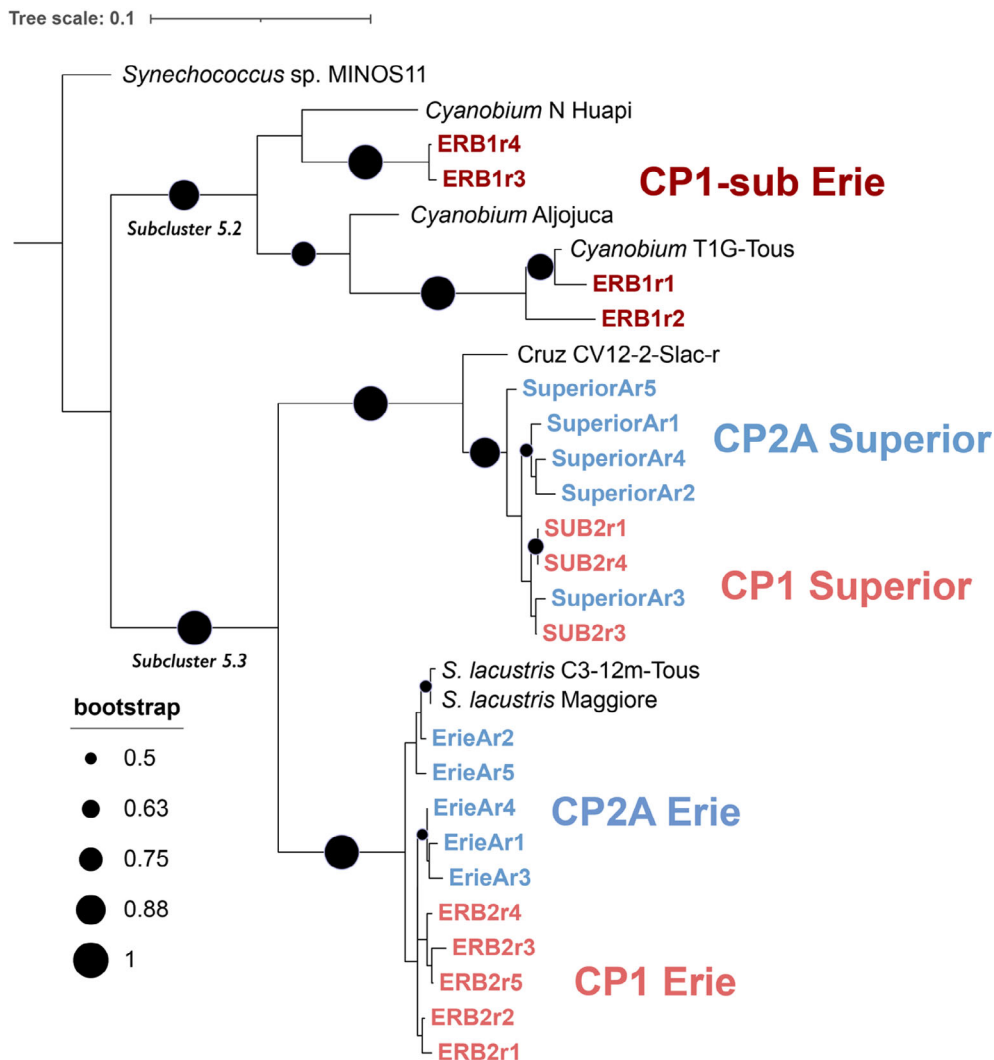


**Fig. 6.** Abundance of each picophytoplankton flow phenotype population in each sample. Population identifiers are along the x-axis. Asterisks (\*) indicate populations that were sorted and sequenced (Fig. 7). Population distribution categories are named along the top of the plot. Abbreviations: SRF, surface; DCL, deep chlorophyll layer; UHY, upper hypolimnion (Erie only); LEP, lower epilimnion (Erie only); mid-HYP, mid-hypolimnion; BOT, bottom. Note that CP2A cells were present in Erie at very low levels, which was below our limit of detection for calculating cell counts per volume for this sample set, but not below our limit for cell sorting and sequencing (Fig. 7, Supporting Information Fig. S5, Table S3).

phylogenetic marker for *Synechococcus*. To capture sequence diversity across domains and other functional genes, we also examined contigs assembled from metagenomes prepared from the sorted cells. Flow phenotypes C1, CP1, and CP2A from the DCL at Superior Sta. SU12 and C1, CP1, CP2A, and a CP1 sub-population (“CP1-sub”) at Erie Sta. ER09 were chosen for this analysis (Supporting Information Fig. S7).

Together these sequencing approaches suggest that the C1 populations in Erie and Superior are photosynthetic picoeukaryotes. Amplification of C1 sorted cells with

cyanobacteria-specific *petB* primers failed to yield a product. From the assembled metagenomes, we recovered 16S and 18S rRNA gene sequences that best matched eukaryotic phytoplankton and heterotrophic bacteria, but not cyanobacteria. We identified several predicted bacterial ribosomal proteins in the assemblies, but all had best matches to heterotrophic bacteria. To focus on the photosynthetic eukaryotes, we identified putative eukaryote contigs and predicted open reading frames on these contigs using the gene prediction tool Augustus (Hoff and Stanke 2013). Based on these predicted protein



**Fig. 7.** Maximum likelihood phylogenetic tree of *petB* gene sequences recovered from sorted cells from CP1, CP1-subpopulation (“CP1-sub,” Erie only), and CP2A flow populations (red and orange) in Erie and Superior samples. Recovered sequences are shown in the context of selected known picocyanobacteria including freshwater, brackish, and marine strains defined in Cabello-Yves et al. 2022. Bootstraps > 0.5 are shown with black circles placed on tree nodes. Circle size indicates bootstrap strength (see legend). Note that CP2A cells were present at low concentrations in Erie (Supporting Information Fig. 5), which were below the limit of detection for the cell per volume display in Fig. 6.

sequences, the most common phytoplankton identified as best hits were the green algae *Choricystis*, *Monomastix*, and *Chlorella* (Supporting Information Table S7). Therefore, we conclude that the C1 populations in both Erie and Superior are likely Chlorophyta, though their specific taxonomic identity warrants future work. This is consistent with the identification of “red-fluorescing phototrophic picoplankton” as *Chlorella*-like cells in previous work from Lake Huron (Fahnenstiel et al. 1991b).

CP1, CP1-subpopulation (ER only), and CP2A flow phenotypes yielded *petB* PCR products, which best matched known phycoerythrin-containing freshwater picocyanobacteria of the *Synechococcales* lineage (Supporting Information Table S5; Fig. 7). The specific identities of the *petB* sequences from each

flow phenotype were distinct from each other and sequences from the same flow phenotype were distinct between Superior and Erie samples. Erie CP1 and CP2A *petB* clones were most closely related to *petB* from freshwater *Synechococcus* isolates *Synechococcus lacustris* 12 m-Tous and *S. lacustris* Maggiore, affiliated with *Synechococcus* Subcluster 5.3 (Cabello-Yeves et al. 2022) (Fig. 7). Sequences from a targeted subpopulation of Erie CP1 were distinct from Erie CP1 and CP2A and best matched representatives of the freshwater picocyanobacterium *Cyanobium* sp. (Subcluster 5.2). The CP1-subpopulation *petB* sequences belonged to two distinct clusters.

Superior CP1 and CP2A cloned *petB* sequences were phylogenetically distinct from the Erie sequences, even for the same



flow phenotype. CP1 and CP2A Superior *petB* best matched *Synechococcus* isolate *Synechococcus* Cruz CV12-2-Slac-r (Cabello-Yeves et al. 2022), from Subcluster 5.3. Superior CP2A and CP1 *petB* sequences were largely distinct from each other, though one CP2A sequence clustered best with the CP1 sequences, which may be due to free-DNA contamination during sorting or an outlier CP1 cell sorted with the CP2A cells.

To further examine genomic diversity among the different flow phenotypes, we analyzed 22 individual single-copy core genes encoding ribosomal proteins from each sorted metagenome in comparison to reference picocyanobacteria genomes (Cabello-Yeves et al. 2022). Most sorted metagenomes yielded multiple genomic sequence bins, demonstrating that different genotypes coexisted with the same flow phenotype. Note that we were not able to obtain sorted metagenomes from all the populations analyzed by *petB* sequencing (i.e., CP1 subpopulation from Erie; Supporting Information Table S3 shows the cell numbers sorted from each population, for each sequencing approach).

The supertree inferred from ribosomal proteins agreed with the *petB* analysis in that each flow phenotype included multiple lineages of picocyanobacteria and that the genotype of a flow phenotype differed between lakes (Fig. 8). Specifically, CP1 from Erie contained distinct sequences affiliated with both Subcluster 5.2 and 5.3, and some Superior CP1 sequences clustered along with CP2A within Subcluster 5.3 (Fig. 8). In contrast to the *petB* analysis, single copy core genes from Superior CP1 revealed a second lineage within Subcluster 5.2.

## Discussion

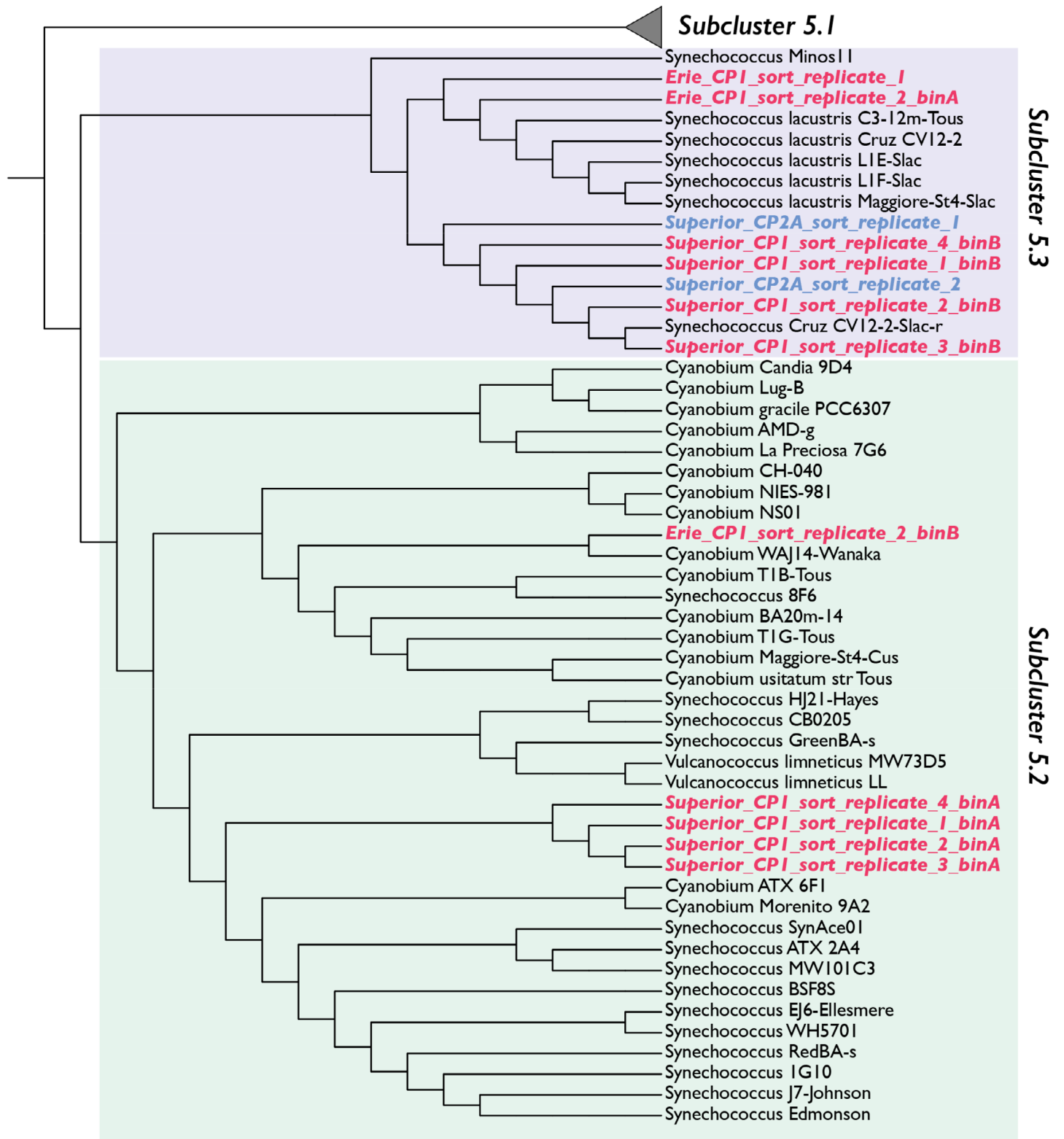
We measured the abundance and distribution of the picoplankton assemblage across a broad range of environmental condition throughout the pelagic region of the Great Lakes ecosystem. The existence of diverse picophytoplankton populations in the Great Lakes was previously known (Crosbie et al. 2003; Wilhelm et al. 2006; Ivanikova et al. 2007, 2008; Carrick et al. 2015; Bramburger and Reavie 2016; Paver et al. 2020), but this is the first time a comprehensive survey of this type was done.

We identified and enumerated eight populations of small (< 2.7  $\mu\text{m}$ ) planktonic cells (i.e., picophytoplankton) from the Laurentian Great Lakes based on flow cytometric (i.e., phenotypic) analyses of their pigment fluorescence. Through analysis of the distributions of our eight pigmented flow phenotypes, and the underlying genomes of a select few, we revealed dramatic differences in the community structure of picophytoplankton phenotypes and genotypes across the Great Lakes, which suggests strong environmental filtering for locally adapted populations across the Lakes.

## Diverse picophytoplankton partition niches of the Great Lakes

Picophytoplankton across the lakes were diverse in terms of phenotype and genotype, which suggests unique functional roles in the environment, niche partitioning, and varying sensitivity to environmental gradients and change. In all lakes, picophytoplankton were evenly composed of prokaryotic and eukaryotic members (Fig. 6), suggesting a large range of diversity in phototrophic metabolisms, life cycles, trophic interactions, and cell sizes in every lake environment. This is consistent with previous flow cytometry and microscopic analysis of picoplankton in the Lakes system that showed a range of cell size, morphologies, pigment types, and cell assemblages (Ivanikova et al. 2008; Carrick 2017). For example, we recovered sequences of diverse heterotrophic bacteria in association with eukaryotic phytoplankton sorts (C1), suggesting direct passage of primary production to bacterial production. C1 may encompass the “*Chlorella*-like red fluorescing phototrophic plankton” identified and cultivated by Fahnenstiel et al. (1991b), which displayed highest primary productivity in early spring conditions. For the ultra-oligotrophic Lake Superior especially, this even balance of prokaryotic and eukaryotic phototroph abundance contrasts with oligotrophic marine systems where prokaryotes (i.e., *Prochlorococcus*) numerically dominate the phototrophic community (e.g., Station ALOHA) and highlights the question of relative primary production rates between these coexisting cells.

Great Lakes picophytoplankton were also diverse in terms of their pigment composition, which we determined by their ratios of chlorophyll (or phycoerythrin) fluorescence excited by light of different wavelengths (Fig. 2C). Through sequencing, we identified three of these populations as *Synechococcales*, consistent with previous understanding of Great Lakes picophytoplankton identity (Fahnenstiel et al. 1991a; Ivanikova et al. 2007, 2008). This lineage is known to take on numerous arrangements of light harvesting proteins that enable partitioning of spectral niches in aquatic systems (Six et al. 2007, 2021; Grébert et al. 2018). In all the samples, multiple pigment phenotypes were present, suggesting that niche differentiation based on light quality is a common feature across the Great Lakes. There are three main pigment types in *Synechococcus* corresponding to the phycobiliprotein composition of rods in photosynthetic antenna called phycobilisomes. Type 1 have phycocyanin and no phycoerythrin, so would only exhibit chlorophyll fluorescence (i.e., “C” flow populations). Type 2 have phycocyanin and phycoerythrin I. Type 3 have phycocyanin, phycoerythrin I, and phycoerythrin II. Type 3 strains have both phycoerythrobilin and phycourobilin chromophores and have subtypes that differ in absorbance properties based on the ratio of these chromophores. Relative absorbance across the visible light spectrum differs based on pigment type (Six et al. 2007). While we are unable to assign flow populations to specific



**Fig. 8.** Consensus supertree based on 22 individual single-copy core-genes (SCGs) extracted from CP1 and CP2A sorted flow population metagenome assemblies and existing freshwater and marine *Synechococcales* genomes. Different genomic bins from the same sort replicate are designated as “binA” or “binB,” if multiple genomic bins were present.

pigment types based on fluorescence properties, our previous work with cultures shows that fluorescence properties at wavelengths corresponding to chlorophyll and phycoerythrin reflect phycobilisome composition (Thompson and van den Engh 2016).

The coexistence of distinct pigment types is a common feature of aquatic systems that harbor picocyanobacteria (Ivanikova et al. 2008; Cabello-Yeves et al. 2018, 2022), however our work shows for the first time how their relative abundance and distribution shift across freshwater gradients. By quantifying the pigment phenotypes, we found that Erie and Ontario samples harbored a set of phenotypes absent from the other lakes (i.e., C2, CP3A, and CP3B), suggesting a shift in the available spectral niches between the more oligotrophic upper lakes (bluer waters) and more eutrophic lower lakes (greener waters). This pattern is consistent with previous work cloning and culturing picocyanobacteria from Erie and Superior, where vastly different picocyanobacterial populations were found (Ivanikova et al. 2008). We hypothesize that these Erie/Ontario-specific flow phenotypes are green-light specialists, consistent with the abundance of green light in particle heavy eutrophic ecosystems.

However, we cannot untangle niche partitioning based on available wavelengths of light from niche partitioning based on nutrient availability or other physiochemical parameters (e.g., temperature) in the Lakes. We found several environmental factors to be strongly correlated to flow phenotype community composition (Fig. 5), especially nitrogen and phosphorus. Thus, the Erie-specific populations (CP3A and CP3B), and the populations that thrive in Erie (C2 and CP1) may have higher phosphorus requirements than upper lake-specific populations and/or novel mechanisms for acquiring nitrogen in addition to green-light adapted light harvesting systems.

### Genotype shifts underlie ubiquitous pigment phenotypes

This work identifies the Great Lakes system as a natural laboratory to integrate research on picocyanobacterial evolution and ecology, specifically for the globally abundant and relatively well-characterized lineage *Synechococcus*. Previous work on marine *Synechococcus* has shown the incongruence of vertically inherited core genome phylogeny with pigment type (Grébert et al. 2022) and the importance of acclimation to light quality in *Synechococcus* ecology (Six et al. 2007; Grébert et al. 2018). More recent work proposes that *Synechococcus* pigment type relies on relatively rapid, or recent, horizontal acquisition of light harvesting gene sets through a mobile DNA element in response to spectral conditions (Grébert et al. 2022). Consistent with this understanding, for two distinct *Synechococcus* core-gene lineages that are spatially separated (Superior vs. Erie), we found subclades presenting two distinct pigment types (Fig. 7). Development of this system further, specifically assembling near-complete genomes and bringing these cells into culture, offers possibilities to better

understand the timescale, environmental pressures, and molecular mechanisms of microbial evolution and ecology in an isolated system with streamlined cells.

### Picophytoplankton community structure reflects patchwork hydrological connectivity

We investigated whether picophytoplankton community structure and characteristics reflected rapid or limited dispersal and found evidence for both at different points in the Lakes. Evidence supporting a disconnected hydrologic system comes from our observation that picocyanobacterial genotypes shifted across the extremes of Lakes gradients (Superior vs. Erie, Fig. 7), even when phenotypes are ubiquitous. This observation suggests that selection outpaces dispersal. If the system were hydrologically connected on a time scale close to picophytoplankton growth, we would have found mixing of Erie and Superior genotypes in each flow phenotype, but instead we found no overlap in the *petB* sequences present in Superior vs. Erie (Fig. 7). Further work assessing the genotypes along more gradual gradients in the Lakes system, for example comparing the genotypes found at multiple transition points between the upper Lakes, could test the idea of connectivity along shorter spatial and temporal scales.

Despite low evidence for system-wide population dispersal, we did find specific points in the Lakes system where the picocyanobacterial community composition reflects hydrological connectivity. Specifically, the transition eastward from Western Erie (ER91M) toward the other Erie stations and Ontario is a setting in which dispersal may shape regional picophytoplankton community structure. We hypothesize that niche-specific conditions favored the genomically unidentified population C2 at Erie Sta. ER91M but the population persists downstream due to dispersal and/or the ability of cells to adapt to Ontario conditions. The depth distribution pattern of C2 is also intriguing. Although appearing at all depths at shallow Sta. ER91M, C2 persists chiefly at the surface and DCL downstream, particularly in Lake Ontario, where a well-known eastward coastal jet current along its south shore may propel plankton from the warm Niagara River outflow to above-thermocline (surface) habitat at Sta. ON55M (Bai et al. 2013). Thus, ON55M could be an example of how Great Lakes currents shape microbial populations. It is important to note that current flow in the Great Lakes is complex, but with high inflow and outflow and the smaller mean depth and residence times (Dobiesz et al. 2010; US Department of Commerce), Lakes Erie and Ontario may be the most likely among the Great Lakes to manifest current-mediated dispersal.

### Western Lake Erie harbors unique picophytoplankton

Picophytoplankton population genomic diversity, abundance, and flow phenotypes in Lake Erie stood out from the other Great Lakes, particularly at its western-most Sta. ER91M (Figs. 4–6). Erie picophytoplankton total abundance was higher than the other lakes, and its phenotype diversity

was highest at ER91M where five different populations were present compared to three at all other Erie stations (Fig. 4). These data suggest that, for Lake Erie as a whole, picocyanobacteria abundance is supported by its unique environmental characteristics of warm shallow meso-eutrophic water with high total dissolved phosphorous. However, diversity richness at ER91M may be created by additional numerous and varied niches created by complex shallow bottom topography and currents with inputs from three rivers. The unique conditions of western Lake Erie are notorious for supporting harmful algal blooms (HABs) of cystic and colonial cyanobacteria in summer and early fall (Harke et al. 2016; Berry et al. 2017). While we do not know of examples of picocyanobacteria that act as harmful algal species, our findings of unique picophytoplankton community structure adds to the complexity of the western Lake Erie ecosystem with respect to harmful species. Lake Erie Sta. ER91M also represents an advantageous location to study differences between bacterial communities driven by niche partitioning and hydrologic connectivity as done in the Osterseen lakes of Bavaria (Zwirgmaier et al. 2015), especially with application to human relationships with harmful cyanobacterial blooms.

It is important to acknowledge the technical limitation we faced in distinguishing the coexisting Erie picophytoplankton populations by flow cytometry. In some cases, cell populations overlapped partially (i.e., CP1 subpopulations; Supporting Information Figs. S7, S8), making careful quantification of these populations difficult. Additionally, we did not identify the Erie-specific populations, as we did not sample the appropriate stations and/or could not sort enough cells for sequencing. Strategic sampling to capture these populations will provide additional opportunities to link picophytoplankton phenotype to genotype and ecosystem function. In addition, application of computer-assisted gating could help resolve the diversity of the phytoplankton populations in these complex environments.

## Conclusions

As pigment-bearers, freshwater picocyanobacteria and small eukaryotic phytoplankton are key functional groups of primary producers that comprise the base of food webs of the Laurentian Great Lakes. Understanding how these primary producers vary over space and time is critical to understanding their role in the Great Lakes ecosystem and other large and small freshwater systems worldwide. This work demonstrates the utility of multilaser flow cytometry coupled to molecular analysis to detect, distinguish, and enumerate eukaryotic and prokaryotic pigment types in their ecological context, complementing previous understanding of the contributions of these small cells to primary production in the Lakes (Fahnenstiel et al. 1991a). However, we found dramatic differences in the genotypes underlying these ubiquitous pigment types over the Lakes' gradients. This suggests strong local

adaptation to the biogeochemical conditions of each lake, and the possibility that horizontal gene acquisitions support adaptation to spectral niches across the Lakes, as pigment genes can reside in tychepon like regions of the genome and undergo lateral gene transfer (Grébert et al. 2022; Hackl et al. 2023). Future work will characterize the genomic adaptations distinguishing these populations, link patterns of primary productivity to underlying picophytoplankton community structure (genotypic and phenotypic), and measure responses to environmental pressures and gradients including trophic structure, seasonal-to-decadal shifts, light, nutrients, and anthropogenic influences.

## Data availability statement

Environmental data are publicly available through the U.S. EPA's Great Lakes National Program Office and can be accessed through the Great Lakes Environmental Database system (GLENDa; available at <https://cdx.epa.gov>). Raw flow cytometry data files are available at FlowRepository.org (FR-FCM-Z2UB). Sequence data are available in NCBI's sequence read archive (SRA) (PRJNA917635). All other data needed to evaluate our conclusions are available in the manuscript.

## References

- Bai, X., J. Wang, D. J. Schwab, Y. Yang, L. Luo, G. A. Leshkevich, and S. Liu. 2013. Modeling 1993–2008 climatology of seasonal general circulation and thermal structure in the Great Lakes using FVCOM. *Ocean Model.* **65**: 40–63. doi:10.1016/j.ocemod.2013.02.003
- Barbiero, R. P., B. M. Lesht, E. K. Hinchey, and T. G. Nettesheim. 2018. A brief history of the U.S. EPA Great Lakes National Program Office's water quality survey. *J. Great Lakes Res.* **44**: 539–546. doi:10.1016/j.jglr.2018.05.011
- Barbiero, R. P., and M. L. Tuchman. 2001. Results from the U.S. EPA's biological open water surveillance program of the Laurentian Great Lakes: I. introduction and phytoplankton results. *J. Great Lakes Res.* **27**: 134–154. doi:10.1016/S0380-1330(01)70628-4
- Berry, M. A., and others. 2017. Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities. *Environ. Microbiol.* **19**: 1149–1162. doi:10.1111/1462-2920.13640
- Bramburger, A. J., and E. D. Reavie. 2016. A comparison of phytoplankton communities of the deep chlorophyll layers and epilimnia of the Laurentian Great Lakes. *J. Great Lakes Res.* **42**: 1016–1025. doi:10.1016/j.jglr.2016.07.004
- Brand, A., L. Allen, M. Altman, M. Hlava, and J. Scott. 2015. Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing* **28**: 151–155. doi:10.1087/20150211
- Buttigieg, P. L., and A. Ramette. 2014. A guide to statistical analysis in microbial ecology: A community-focused, living



- review of multivariate data analyses. *FEMS Microbiol. Ecol.* **90**: 543–550. doi:10.1111/1574-6941.12437
- Cabello-Yeves, P. J., A. Picazo, A. Camacho, C. Callieri, R. Rosselli, J. J. Roda-Garcia, F. H. Coutinho, and F. Rodriguez-Valera. 2018. Ecological and genomic features of two widespread freshwater picocyanobacteria. *Environ. Microbiol.* **20**: 3757–3771. doi:10.1111/1462-2920.14377
- Cabello-Yeves, P. J., and others. 2022. Elucidating the picocyanobacteria salinity divide through ecogenomics of new freshwater isolates. *BMC Biol.* **20**: 175. doi:10.1186/s12915-022-01379-z
- Callieri, C. 2010. Single cells and microcolonies of freshwater picocyanobacteria: A common ecology. *J. Limnol.* **69**: 257. doi:10.4081/jlimnol.2010.257
- Callieri, C. 2017. *Synechococcus* plasticity under environmental changes. *FEMS Microbiol. Lett.* **364**: fnx229. doi:10.1093/femsle/fnx229
- Caron, D. A., F. R. Pick, and D. R. S. Lean. 1985. Chroococcoid cyanobacteria in Lake Ontario: Vertical and seasonal distributions during 1982. *J. Phycol.* **21**: 171–175. doi:10.1111/j.0022-3646.1985.00171.x
- Carrick, H. J. 2017. Seasonal abundance, biomass and morphological diversity of picoplankton in Lake Superior: Importance of water column mixing. *Int J Hydro* **1**: 187–197. doi:10.15406/ijh.2017.01.00034
- Carrick, H. J., E. Butts, D. Daniels, M. Fehring, C. Frazier, G. L. Fahnenstiel, S. Pothoven, and H. A. Vanderploeg. 2015. Variation in the abundance of pico, nano, and microplankton in Lake Michigan: Historic and basin-wide comparisons. *J. Great Lakes Res.* **41**: 66–74. doi:10.1016/j.jglr.2015.09.009
- Crosbie, N. D., M. Pöckl, and T. Weisse. 2003. Dispersal and phylogenetic diversity of nonmarine picocyanobacteria, inferred from 16S rRNA gene and *cpcBA*-intergenic spacer sequence analyses. *Appl. Environ. Microbiol.* **69**: 5716–5721. doi:10.1128/AEM.69.9.5716-5721.2003
- Dobiesz, N. E., and others. 2010. Metrics of ecosystem status for large aquatic systems—A global comparison. *J. Great Lakes Res.* **36**: 123–138. doi:10.1016/j.jglr.2009.11.003
- Dusenberry, J. A., R. J. Olson, and S. W. Chisholm. 2000. Field observations of oceanic mixed layer dynamics and picophytoplankton photoacclimation. *J. Mar. Syst.* **24**: 221–232. doi:10.1016/S0924-7963(99)00087-1
- Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797. doi:10.1093/nar/gkh340
- Eren, A. M., and others. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **6**: 3–6. doi:10.1038/s41564-020-00834-3
- Fahnenstiel, G. L., L. Sicko-Goad, D. Scavia, and E. F. Stoermer. 1986. Importance of picoplankton in Lake Superior. *Can. J. Fish. Aquat. Sci.* **43**: 235–240. doi:10.1139/f86-028
- Fahnenstiel, G. L., H. J. Carrick, and R. Iturriaga. 1991a. Physiological characteristics and food-web dynamics of *Synechococcus* in lakes Huron and Michigan. *Limnol. Oceanogr.* **36**: 219–234. doi:10.4319/lo.1991.36.2.0219
- Fahnenstiel, G. L., H. J. Carrick, C. E. Rogers, and L. Sicko-Goad. 1991b. Red fluorescing phototrophic picoplankton in the Laurentian Great Lakes: What are they and what are they doing? *Int Rev Ges Hydrobiol Hydrogr* **76**: 603–616. doi:10.1002/iroh.19910760411
- Fahnenstiel, G. L., and H. J. Carrick. 1992. Phototrophic picoplankton in lakes Huron and Michigan: Abundance, distribution, composition, and contribution to biomass and production. *Can. J. Fish. Aquat. Sci.* **49**: 379–388. doi:10.1139/f92-043
- Farrant, G. K., and others. 2016. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* **113**: E3365–E3374. doi:10.1073/pnas.1524865113
- Grébert, T., and others. 2018. Light color acclimation is a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc. Natl. Acad. Sci.* **115**: E2010–E2019. doi:10.1073/pnas.1717069115
- Grébert, T., and others. 2022. Diversity and evolution of pigment types in marine *Synechococcus* cyanobacteria. *Genome Biol. Evol.* **14**: evac035. doi:10.1093/gbe/evac035
- Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**: 307–321. doi:10.1093/sysbio/syq010
- Hackl, T., and others. 2023. Novel integrative elements and genomic plasticity in ocean ecosystems. *Cell* **186**: 47–62. doi:10.1016/j.cell.2022.12.006
- Harke, M. J., T. W. Davis, S. B. Watson, and C. J. Gobler. 2016. Nutrient-controlled niche differentiation of Western Lake Erie cyanobacterial populations revealed via meta-transcriptomic surveys. *Environ. Sci. Technol.* **50**: 604–615. doi:10.1021/acs.est.5b03931
- Haverkamp, T. H. A., D. Schouten, M. Doleman, U. Wollenzien, J. Huisman, and L. J. Stal. 2009. Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea. *ISME J.* **3**: 397–408. doi:10.1038/ismej.2008.118
- Hoff, K. J., and M. Stanke. 2013. WebAUGUSTUS—A web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res.* **41**: W123–W128. doi:10.1093/nar/gkt418
- Hyatt, D., G.-L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119. doi:10.1186/1471-2105-11-119
- Ivanikova, N. V., L. C. Popels, R. M. L. McKay, and G. S. Bullerjahn. 2007. Lake Superior supports novel clusters of

- cyanobacterial picoplankton. *Appl. Environ. Microbiol.* **73**: 4055–4065. doi:10.1128/AEM.00214-07
- Ivanikova, N. V., A. R. Matteson, C. J. Gobler, L. C. Popels, M. L. McKay, S. W. Wilhelm, and G. S. Bullerjahn. 2008. Picoplanktonic cyanobacteria in lakes superior and Erie: Phylogenies of endemic populations and cultured isolates. *SIL Proc* **1922-2010**: 459–465. doi:10.1080/03680770.2008.11902166
- Kehl, J. 2018. The great lapse: Climate change, water resources and economic risks in the Great Lakes. *J Water Resour Protect* **10**: 1106–1114. doi:10.4236/jwarp.2018.1011065
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with bowtie 2. *Nat. Methods* **9**: 357–359. doi:10.1038/nmeth.1923
- Letunic, I., and P. Bork. 2019. Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**: W256–W259. doi:10.1093/nar/gkz239
- Li, D., C.-M. Liu, R. Luo, K. Sadakane, and T.-W. Lam. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676. doi:10.1093/bioinformatics/btv033
- Munawar, M., I. F. Munawar, T. Weisse, G. G. Leppard, and M. Legner. 1994. The significance and future potential of using microbes for assessing ecosystem health: The Great Lakes example. *J Aquat Ecosyst Stress Recov* **3**: 295–310. doi:10.1007/BF00115288
- Oksanen, J., and others. 2019. Vegan: Community ecology package. R package version. Vegan.
- Palenik, B. 2001. Chromatic adaptation in Marine Synechococcus strains. *Appl. Environ. Microbiol.* **67**: 991–994. doi:10.1128/AEM.67.2.991-994.2001
- Paver, S. F., R. J. Newton, and M. L. Coleman. 2020. Microbial communities of the Laurentian Great Lakes reflect connectivity and local biogeochemistry. *Environ. Microbiol.* **22**: 433–446. doi:10.1111/1462-2920.14862
- Pronk, L. J. U., and M. H. Medema. 2022. Whokaryote: distinguishing eukaryotic and prokaryotic contigs in metagenomes based on gene structure. *bioRxiv*: **8**(5): mgen000823. doi:10.1101/2021.11.15.468626
- Reavie, E. D., M. Cai, C. Meyer-Jacob, J. P. Smol, and J. P. Werne. 2021. Long-term primary production trends in the Laurentian Great Lakes: A comparison of geochemical methods. *J Paleolimnol* **65**: 299–314. doi:10.1007/s10933-020-00170-w
- Revell, L. J. 2011. Phytools: an R package for phylogenetic comparative biology (and other things), phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**: 217–223.
- Scofield, A. E., J. M. Watkins, E. Osantowski, and L. G. Rudstam. 2020. Deep chlorophyll maxima across a trophic state gradient: A case study in the Laurentian Great Lakes. *Limnol Oceanogr*, **65**: 2460–2484. doi:10.1002/lno.11464
- Six, C., J.-C. Thomas, L. Garczarek, M. Ostrowski, A. Dufresne, N. Blot, D. J. Scanlan, and F. Partensky. 2007. Diversity and evolution of phycobilisomes in marine Synechococcus spp.: A comparative genomics study. *Genome Biol.* **8**: R259. doi:10.1186/gb-2007-8-12-r259
- Six, C., M. Ratin, D. Marie, and E. Corre. 2021. Marine Synechococcus picocyanobacteria: Light utilization across latitudes. *Proc. Natl. Acad. Sci.* **118**: e2111300118. doi:10.1073/pnas.2111300118
- Stecher, G., K. Tamura, and S. Kumar. 2020. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol. Biol. Evol.* **37**: 1237–1239. doi:10.1093/molbev/msz312
- Sterner, R. W. 2010. In situ-measured primary production in Lake Superior. *J. Great Lakes Res.* **36**: 139–149. doi:10.1016/j.jglr.2009.12.007
- Sterner, R. W. 2021. The Laurentian Great Lakes: A biogeochemical test bed. *Annu. Rev. Earth Planet. Sci.* **49**: 201–229. doi:10.1146/annurev-earth-071420-051746
- Thompson, A. W., and G. van den Engh. 2016. A multi-laser flow cytometry method to measure single cell and population-level fluorescence action spectra for the targeted study and isolation of phytoplankton in complex assemblages. *Limnol Oceanogr Methods* **14**: 39–49.
- Thyssen, M., and others. 2022. Interoperable vocabulary for marine microbial flow cytometry. *Front Mar Sci* **9**. doi:10.3389/fmars.2022.975877
- Twiss, M. R., and others. 2012. Diatoms abound in ice-covered Lake Erie: An investigation of offshore winter limnology in Lake Erie over the period 2007 to 2010. *J. Great Lakes Res.* **38**: 18–30. doi:10.1016/j.jglr.2011.12.008
- Tyner, E. H., and T. A. Boyer. 2020. Applying best-worst scaling to rank ecosystem and economic benefits of restoration and conservation in the Great Lakes. *J. Environ. Manage.* **255**: 109888. doi:10.1016/j.jenvman.2019.109888
- U.S. EPA. 2019. Great Lakes biology monitoring program technical report: status and trends through 2014 for chlorophyll, phytoplankton, zooplankton and benthos; and through 2016 for mysis. EPA.
- Wilhelm, S. W., G. S. Bullerjahn, M. L. Eldridge, J. M. Rinta-Kanto, L. Poorvin, and R. A. Bourbonniere. 2006. Seasonal hypoxia and the genetic diversity of prokaryote populations in the Central Basin Hypolimnion of Lake Erie: Evidence for abundant cyanobacteria and photosynthesis. *J. Great Lakes Res.* **32**: 657–671.
- Zwirgmaier, K., K. Keiz, M. Engel, J. Geist, and U. Raeder. 2015. Seasonal and spatial patterns of microbial diversity along a trophic gradient in the interconnected lakes of the Osterseen Lake District, Bavaria. *Front. Microbiol.* **6**: 1168.

#### Acknowledgments

Funding was provided through NSF-OCE 1830002 to AWT and 1830011 to MLC. We thank Justin Podowski and María Hernández Limón for sample collection. We thank the science staff of the US EPA Great

*Gale et al.*

Lakes National Program Office and the captain, crew, and science party of the R/V *Lake Guardian* for facilitating sample collection, especially Elizabeth Hinchey Malloy, Eric Osantowski, Maxwell Morgan, and Kitty Kennedy.

**Conflict of Interest**

None declared.

*Great Lakes picophytoplankton*

*Submitted 12 May 2023*

*Revised 05 August 2023*

*Accepted 09 August 2023*

*Associate editor: Jana Himmers*