

Supporting information for:

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

John Gale ¹, Carey Sweeney ¹, Sara Paver ², Maureen L. Coleman ², Anne W. Thompson^{1*}

¹ Department of Biology, Portland State University, Portland, 97207, Oregon, USA

² Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

* Corresponding author

Abstract.

The Laurentian Great Lakes provide economic support to millions of people, drives biogeochemical cycling, and is 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 multi-laser 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.

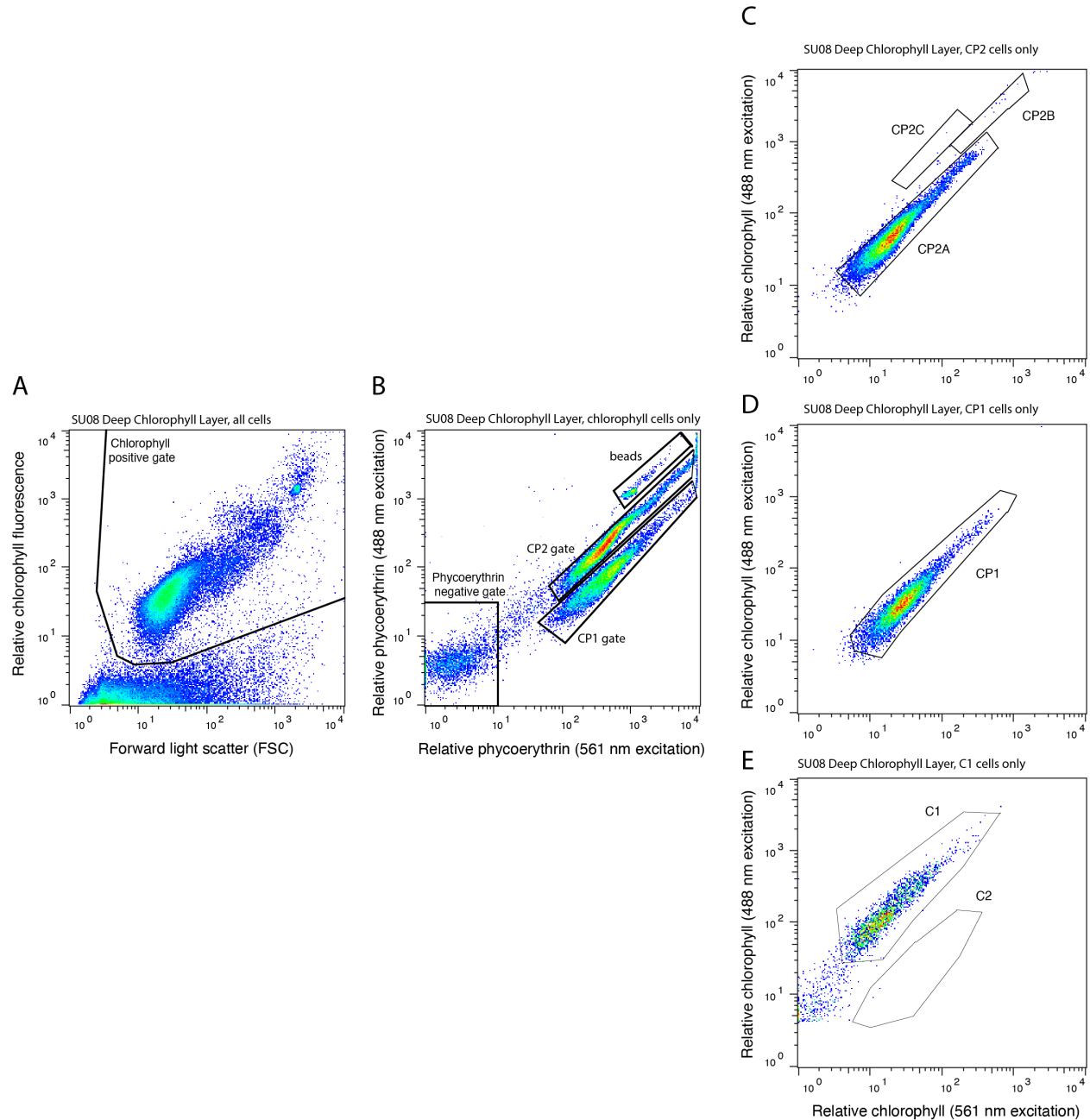
Supplemental Table 1. Details on sampling stations and depth.						
Lake	Station	Latitude	Longitude	total depth (m)	sampling depth (m)	Depth Description
Erie	ER09	42.538333	-79.616667	50.3	2	surface
Erie	ER09	42.538333	-79.616667	50.3	17.1	lower epilimnion
Erie	ER15M	42.516667	-79.893333	63.5	2	surface
Erie	ER15M	42.516667	-79.893333	63.5	11	lower epilimnion
Erie	ER15M	42.516667	-79.893333	63.5	30	upper hypolimnion
Erie	ER15M	42.516667	-79.893333	63.5	53	bottom
Erie	ER78M	42.116667	-81.25	23.8	1.9	surface
Erie	ER78M	42.116667	-81.25	23.8	14.1	lower epilimnion
Erie	ER78M	42.116667	-81.25	23.8	22.8	bottom
Erie	ER91M	41.840833	-82.916667	11.1	2	surface
Erie	ER91M	41.840833	-82.916667	11.1	5.6	lower epilimnion
Erie	ER91M	41.840833	-82.916667	11.1	10	bottom
Huron	HU15M	44	-82.35	67.8	2	surface
Huron	HU15M	44	-82.35	67.8	40	mid-hypolimnion
Huron	HU15M	44	-82.35	67.8	51.2	deep chlorophyll layer
Huron	HU15M	44	-82.35	67.8	58.5	bottom
Huron	HU45M	45.136667	-82.983333	98.9	2	surface
Huron	HU45M	45.136667	-82.983333	98.9	37	deep chlorophyll layer
Huron	HU45M	45.136667	-82.983333	98.9	50	mid-hypolimnion
Huron	HU45M	45.136667	-82.983333	98.9	89.2	bottom
Huron	HU54M	45.516667	-83.416667	125.2	1.1	surface
Huron	HU54M	45.516667	-83.416667	125.2	23.4	deep chlorophyll layer
Huron	HU54M	45.516667	-83.416667	125.2	50.6	mid-hypolimnion
Huron	HU54M	45.516667	-83.416667	125.2	114	bottom
Michigan	MI18M	42.733333	-87	150.4	2	surface
Michigan	MI18M	42.733333	-87	150.4	36	deep chlorophyll layer
Michigan	MI18M	42.733333	-87	150.4	50.2	mid-hypolimnion
Michigan	MI18M	42.733333	-87	150.4	149.7	bottom
Michigan	MI27M	43.6	-86.916667	104	1.9	surface
Michigan	MI27M	43.6	-86.916667	104	38.1	deep chlorophyll layer
Michigan	MI27M	43.6	-86.916667	104	50	mid-hypolimnion
Michigan	MI27M	43.6	-86.916667	104	94.3	bottom
Michigan	MI41M	44.736667	-86.721667	260	2	surface
Michigan	MI41M	44.736667	-86.721667	260	38.9	deep chlorophyll layer
Michigan	MI41M	44.736667	-86.721667	260	100.6	mid-hypolimnion

Michigan	MI41M	44.736667	-86.721667	260	250.5	bottom
Ontario	ON33M	43.596667	-78.801667	138.1	2	surface
Ontario	ON33M	43.596667	-78.801667	138.1	16	deep chlorophyll layer
Ontario	ON33M	43.596667	-78.801667	138.1	50	mid-hypolimnion
Ontario	ON33M	43.596667	-78.801667	138.1	128.5	bottom
Ontario	ON55M	43.443333	-77.438333	191.5	2	surface
Ontario	ON55M	43.443333	-77.438333	191.5	23	deep chlorophyll layer
Ontario	ON55M	43.443333	-77.438333	191.5	100	mid-hypolimnion
Ontario	ON55M	43.443333	-77.438333	191.5	181.5	bottom
Superior	SU01M	46.993306	-85.16111	94.3	2	surface
Superior	SU01M	46.993306	-85.16111	94.3	30	deep chlorophyll layer
Superior	SU01M	46.993306	-85.16111	94.3	50	mid-hypolimnion
Superior	SU01M	46.993306	-85.16111	94.3	84.6	bottom
Superior	SU08M	47.60583	-86.81778	291.1	2	surface
Superior	SU08M	47.60583	-86.81778	291.1	37	deep chlorophyll layer
Superior	SU08M	47.60583	-86.81778	291.1	100	mid-hypolimnion
Superior	SU08M	47.60583	-86.81778	291.1	280.3	bottom
Superior	SU12	47.85611	-88.04194	231.9	2	surface
Superior	SU12	47.85611	-88.04194	231.9	23	deep chlorophyll layer
Superior	SU17M	47.16444	-89.66194	198.5	2	surface
Superior	SU17M	47.16444	-89.66194	198.5	29.9	deep chlorophyll layer
Superior	SU17M	47.16444	-89.66194	198.5	50.6	mid-hypolimnion
Superior	SU17M	47.16444	-89.66194	198.5	188.6	bottom

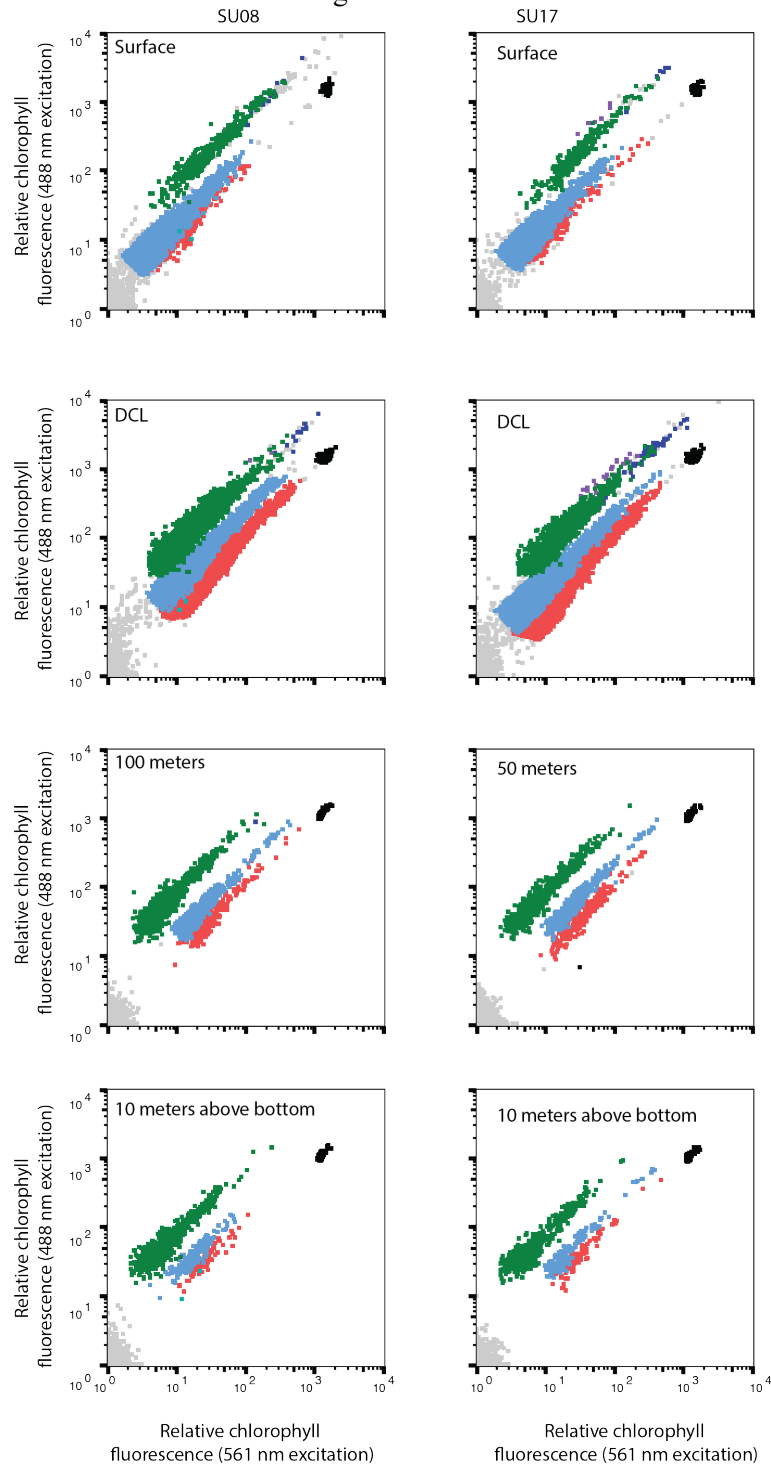
Supplemental Table 2. Optical configuration of 3-laser BD Influx to detect chlorophyll (Chl) and phycoerythrin (PE) fluorescence from unknown cells.					
Wavelength (nm)	Brand	Power (mW)	Rejection band or longpass filter (nm)	Dichroic filter (nm)	Bandpass filter (nm)
488	Coherent Sapphire	100	488	610	692/40 (Chl) 572/27 (PE)
561	Melles Griot	50	568	610	692/40 (Chl) 572/27 (PE)
642	Coherent Cube	30	648	610	692/40 (Chl)

Supplemental Table 3. Number of cells sorted from each flow phenotype population and sample for either metagenome sequencing (top set) or <i>petB</i> gene sequencing (bottom set). Number indicates the number of cells sorted in each replicate with the number of replicates in parentheses.				
For metagenomes	Population			
Sample	CP1	CP1sub	CP2A	C1
Lake Superior (SU12), deep chl. maximum, August 2019	20,000 (2)	0 (0)	20,000 (2)	5,000 (1)
Lake Erie (ER09) lower epilimnion, August 2019	20,000 (2)	0 (0)	0 (0)	2,000 (1)
For <i>petB</i> cloning	Population			
Sample	CP1	CP1sub	CP2A	C1
Lake Superior (SU12), deep chl. maximum, August 2019	20,000 (1)	0 (0)	20,000 (1)	5,000 (1)
Lake Erie (ER09) lower epilimnion, August 2019	20,000 (1)	5,000	500 (1)	2,000 (1)

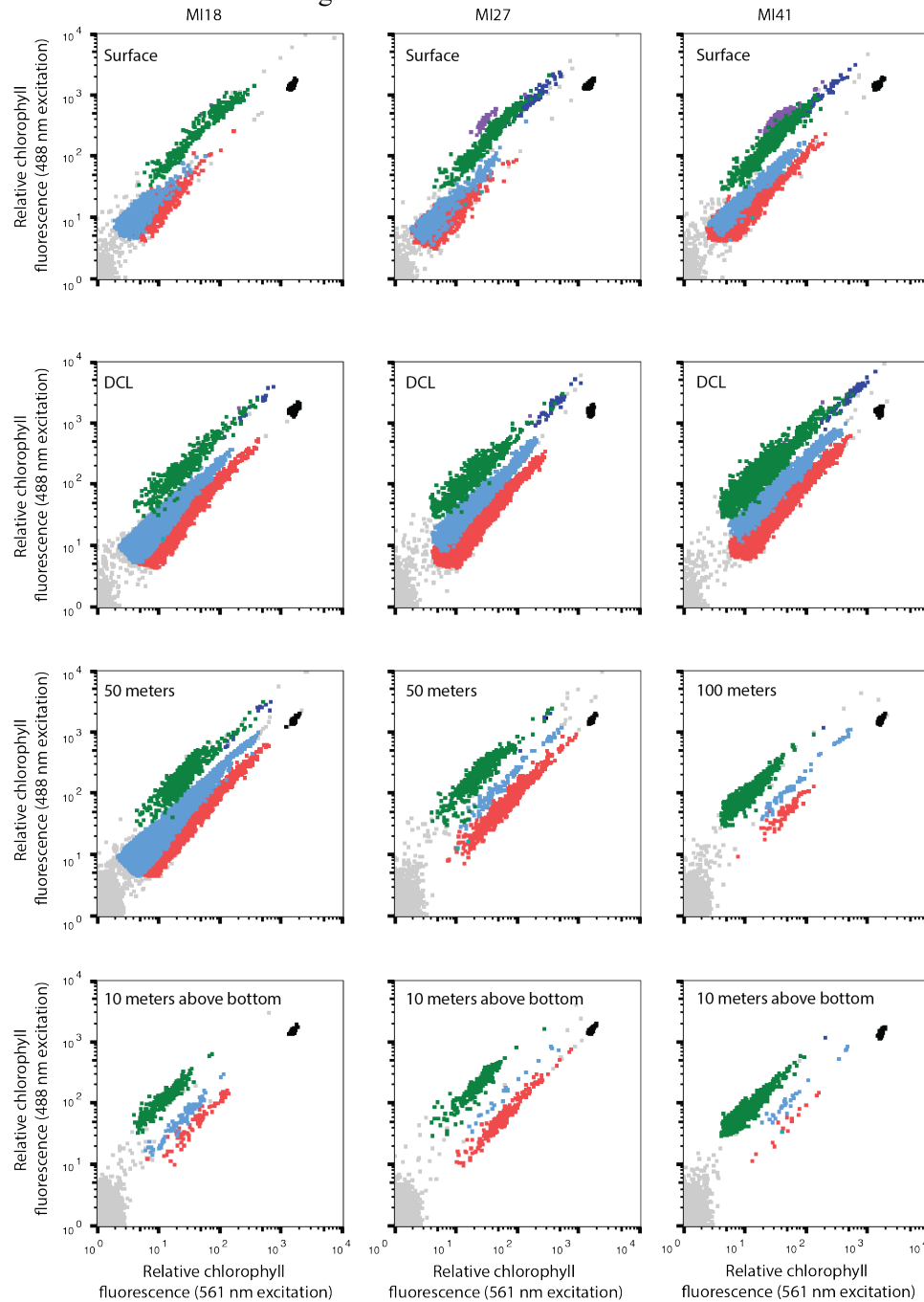
Supplemental Figure 1. Example of gating on SU08 deep chlorophyll layer (DCL) sample. A) Initial gating of all chlorophyll containing cells on bivariate plot of FSC vs chlorophyll fluorescence. B) View of chlorophyll positive cells only (gate in A) on bivariate plot of phycoerythrin fluorescence excited by green (x-axis) and blue (y-axis) lasers. Gates of major populations are labeled. C) View of CP2 cells only (gated in B) on bivariate plot of chlorophyll fluorescence excited by green (x-axis) and blue (y-axis) lasers with subpopulations labeled and gated. D) View of CP1 cells only (gated in B) on bivariate plot of chlorophyll fluorescence excited by green (x-axis) and blue (y-axis) lasers with subpopulations labeled and gated. E) View of C1 cells only (gated in B) on bivariate plot of chlorophyll fluorescence excited by green (x-axis) and blue (y-axis) lasers with subpopulations labeled and gated. Please refer to Methods for details on minimum number of cells that meet the limit of detection.



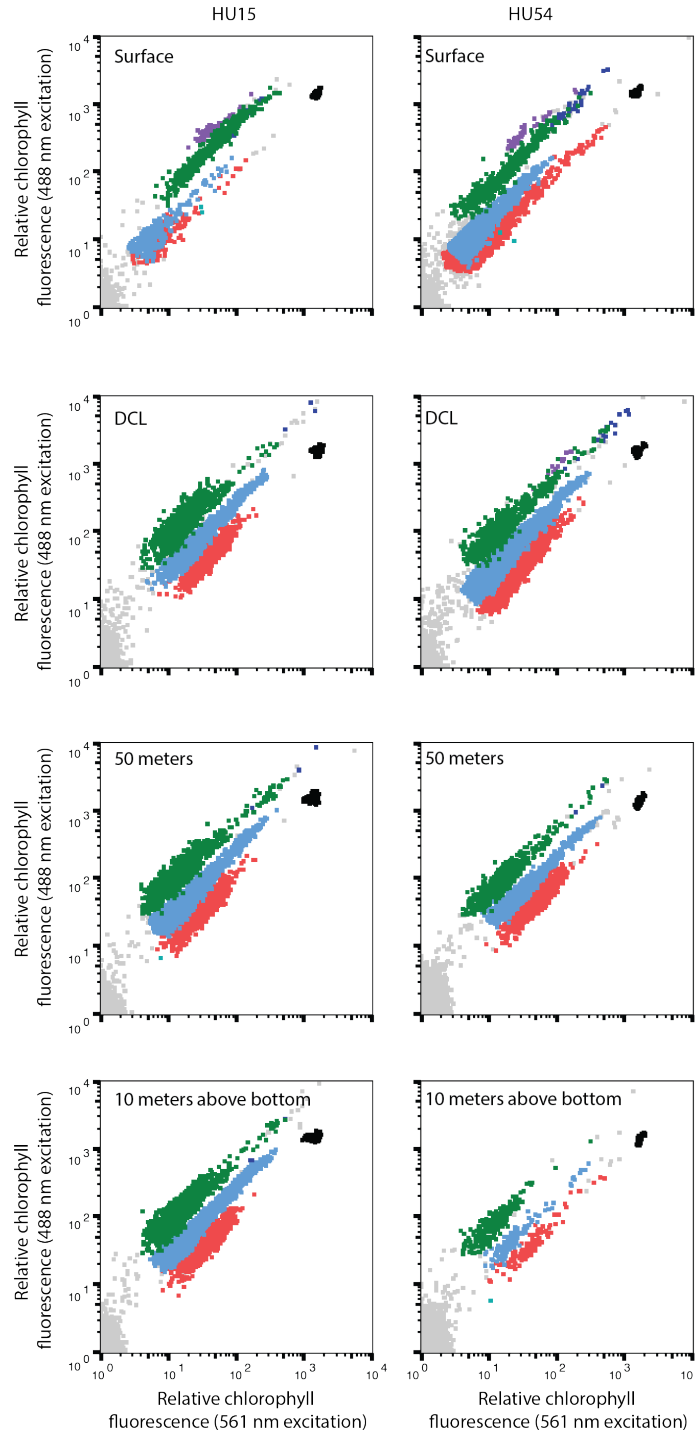
Supplemental Figure 2. Lake Superior flow populations visualized in bivariate plots of chlorophyll fluorescence excited by green (561nm, x-axis) and blue (488nm, y-axis) lasers. Colored dots are cells that are part of defined populations according to the color scheme in Figure 2. While populations overlap in these plots of chlorophyll (488 nm excited) vs chlorophyll (561 nm excited), they do not overlap in the plots of phycoerythrin (488 nm excited) vs phycoerythrin (561 nm excited), which were used to gate them. Gray dots are cells that were not gated.



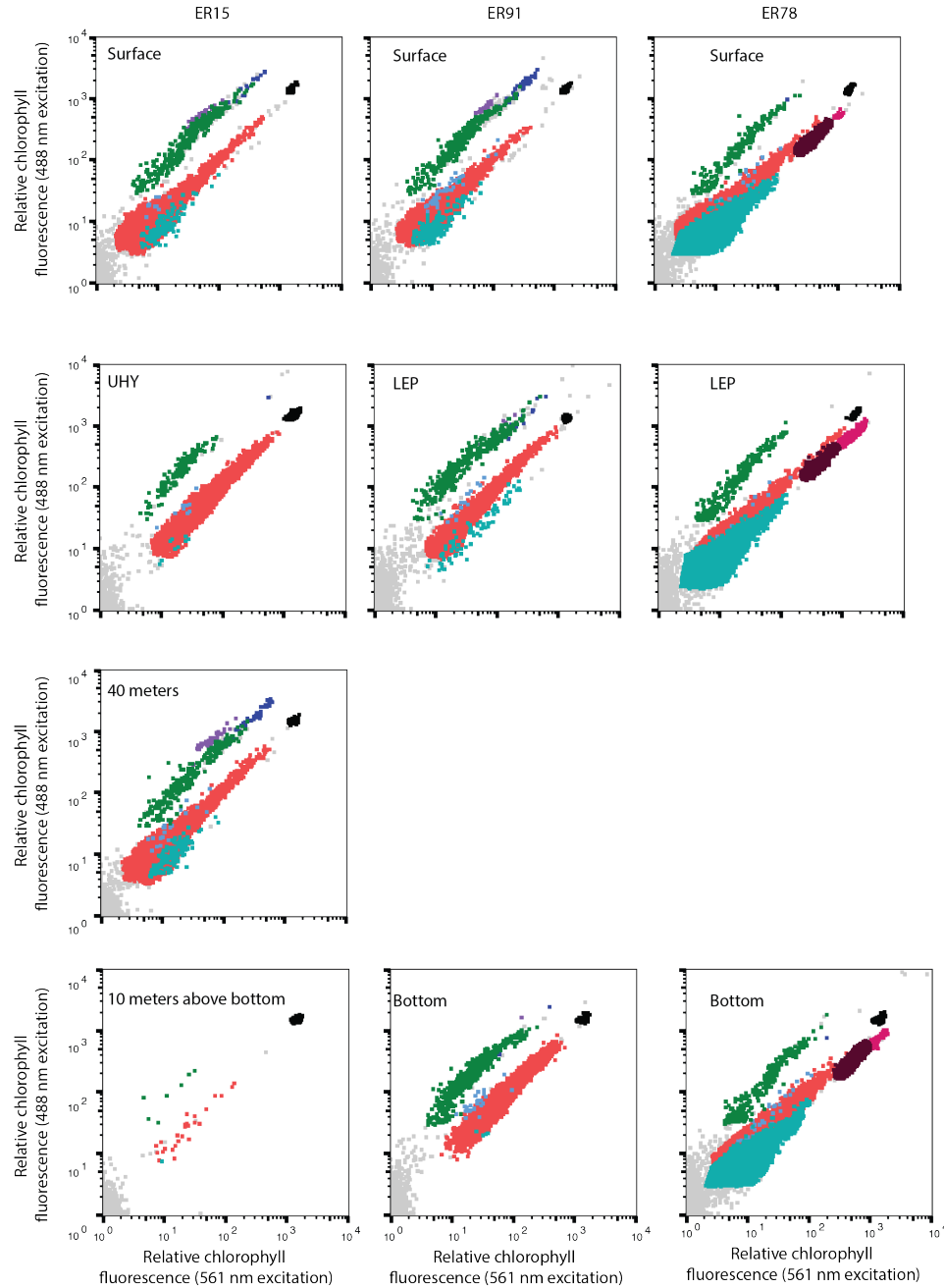
Supplemental Figure 3. Lake Michigan flow populations visualized in bivariate plots of chlorophyll fluorescence excited by green (561nm, x-axis) and blue (488nm, y-axis) lasers. Colored dots are cells that are part of defined populations according to the color scheme in Figure 2. While populations overlap in these plots of chlorophyll (488 nm excited) vs chlorophyll (561 nm excited), they do not overlap in the plots of phycoerythrin (488 nm excited) vs phycoerythrin (561 nm excited), which were used to gate them. Gray dots are cells that were not gated.



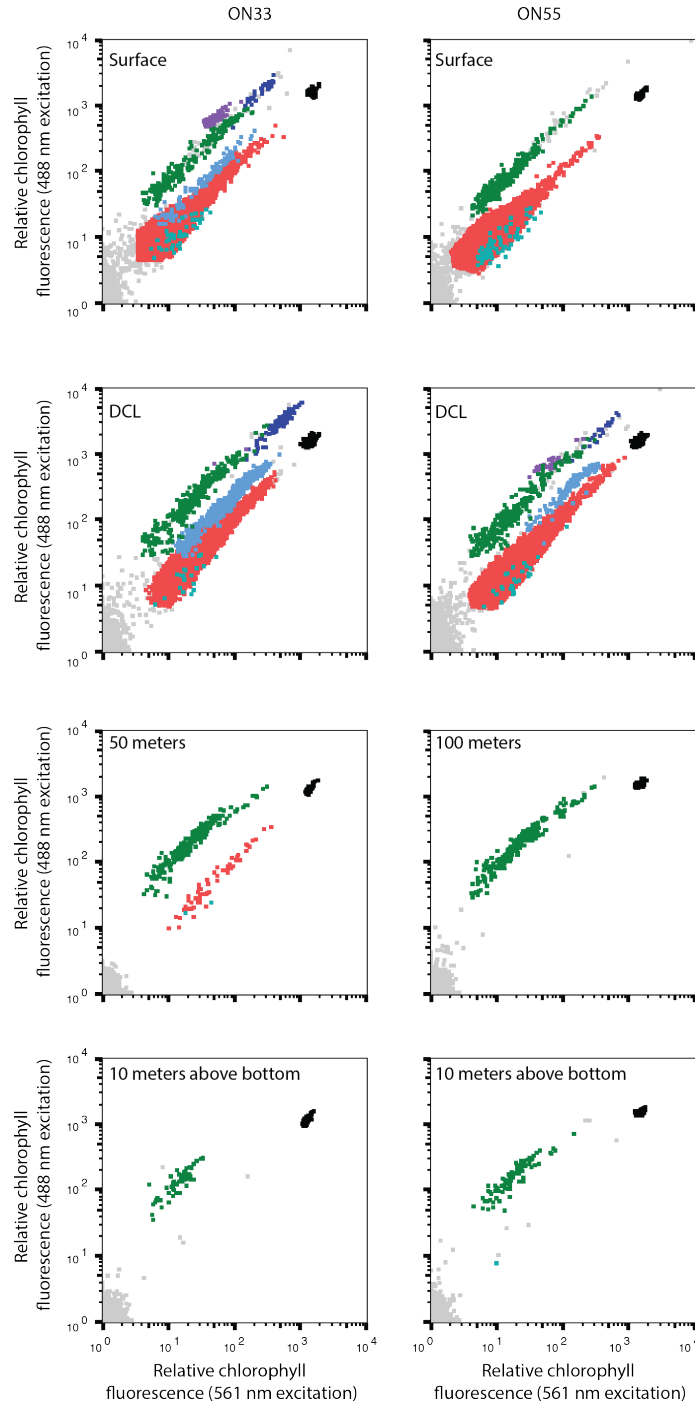
Supplemental Figure 4. Lake Huron flow populations visualized in bivariate plots of chlorophyll fluorescence excited by green (561nm, x-axis) and blue (488nm, y-axis) lasers. Colored dots are cells that are part of defined populations according to the color scheme in Figure 2. While populations overlap in these plots of chlorophyll (488 nm excited) vs chlorophyll (561 nm excited), they do not overlap in the plots of phycoerythrin (488 nm excited) vs phycoerythrin (561 nm excited), which were used to gate them. Gray dots are cells that were not gated.



Supplemental Figure 5. Lake Erie flow populations visualized in bivariate plots of chlorophyll fluorescence excited by green (561nm, x-axis) and blue (488nm, y-axis) lasers. Colored dots are cells that are part of defined populations according to the color scheme in Figure 2. While populations overlap in these plots of chlorophyll (488 nm excited) vs chlorophyll (561 nm excited), they do not overlap in the plots of phycoerythrin (488 nm excited) vs phycoerythrin (561 nm excited), which were used to gate them. Gray dots are cells that were not gated.

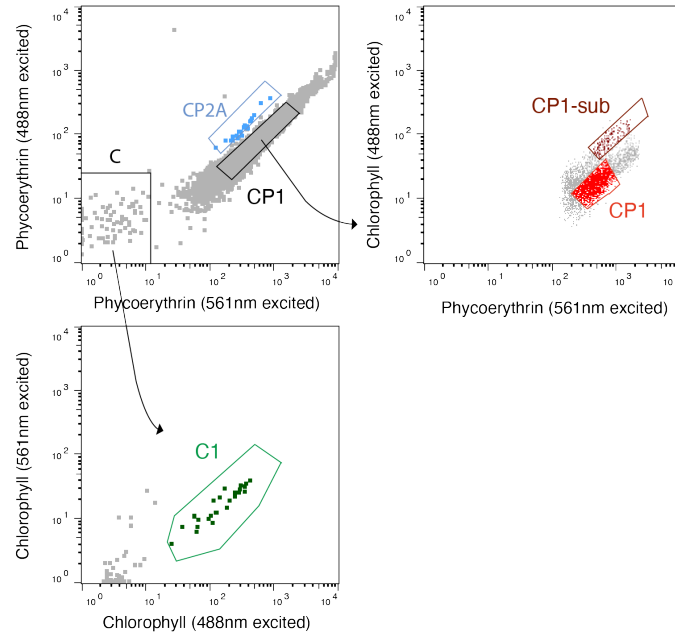


Supplemental Figure 6. Lake Ontario flow populations visualized in bivariate plots of chlorophyll fluorescence excited by green (561nm, x-axis) and blue (488nm, y-axis) lasers. Colored dots are cells that are part of defined populations according to the color scheme in Figure 2. While populations overlap in these plots of chlorophyll (488 nm excited) vs chlorophyll (561 nm excited), they do not overlap in the plots of phycoerythrin (488 nm excited) vs phycoerythrin (561 nm excited), which were used to gate them. Gray dots are cells that were not gated.

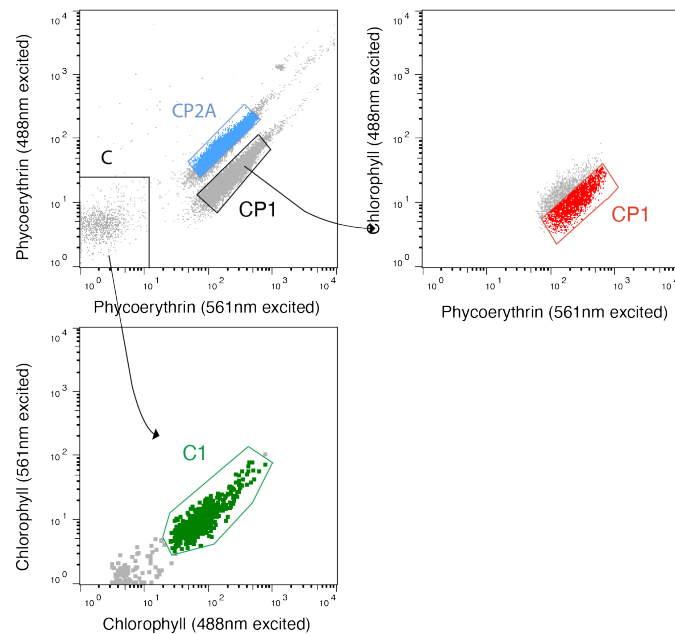


Supplemental Figure 7. Dot plots of chlorophyll-positive cells from concentrated samples used for cell sorting for Lake Erie (A) and Lake Superior (B) towards *petB* clones and/or metagenomes. All chlorophyll positive cells for each lake are shown with gray dots with parent-level gates (gray/black lines) or child-level sort gates (colored dots and lines). Cells are colored if they belong to a gate from which cells were sorted with arrows indicating the parent and child gates for each population. For Erie, CP1 cells showed a subpopulation, which was distinct enough to gate and sort (CP1-sub, Erie only) while only CP1 was apparent in the Superior sample. For some populations (e.g. Erie CP2A), more cells were sorted than are shown in this datafile of the sample. Number of sorted cells are in Supplemental Table 2.

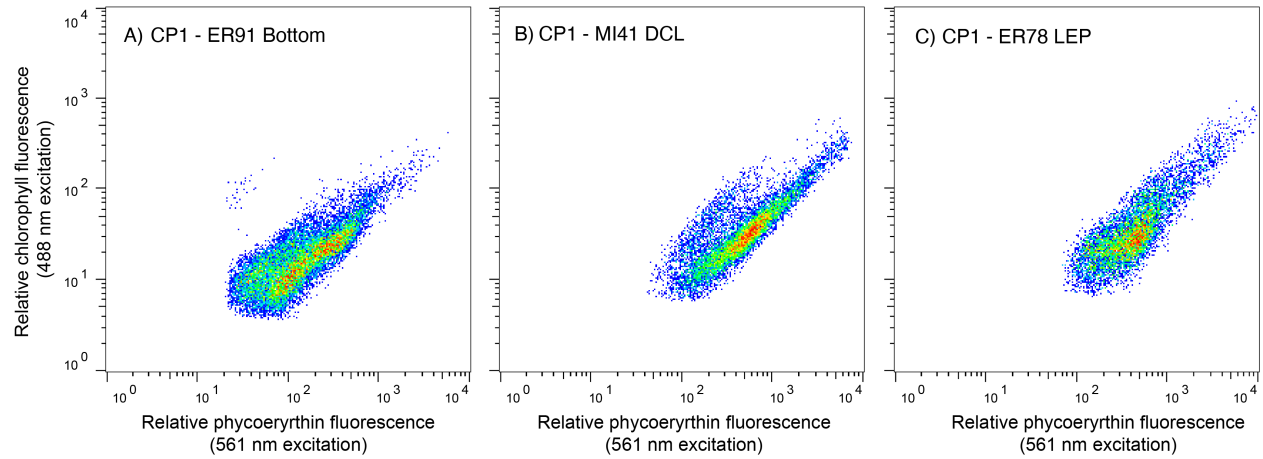
A) Superior SU12 DCL sorted sample and gates



B) Erie ER09 LEP sorted sample and gates

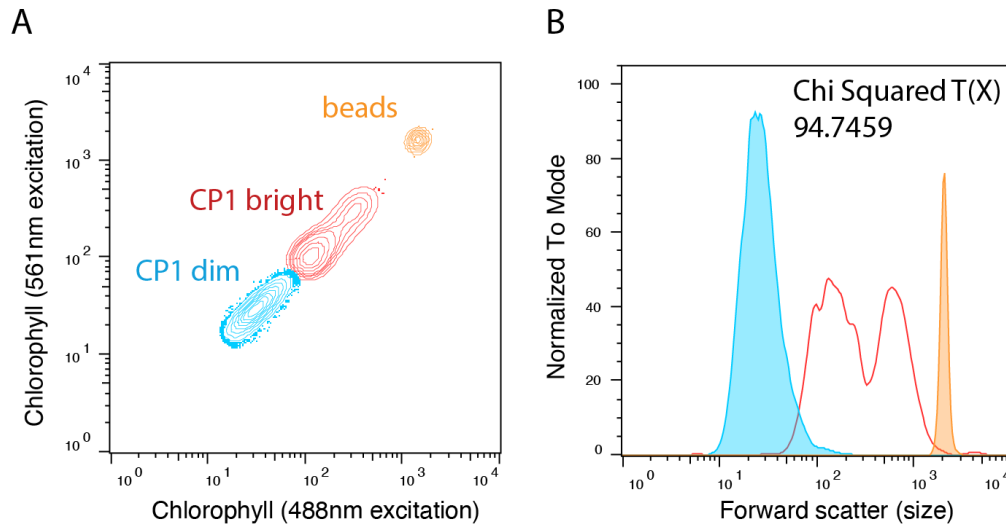


Supplemental Figure 8. Pseudo-colored density plots of CP1 cells from A) Erie Station ER91M bottom, B) Michigan Station MI41M deep chlorophyll layer (DCL), and C) Erie Station ER78M lower epilimnion (LEP) on bivariate plots of phycoerythrin excited by the 561 nm laser (x-axis) and chlorophyll excited by the 488 nm laser (y-axis). Each dot represents a cell. Dots are colored according to their density in the bivariate plot with warmer colors (red) indicating high density and cool colors (blue) representing low density. Note that that we sorted cells from a distinct subpopulation in ER09 (Supplemental Figure 7).



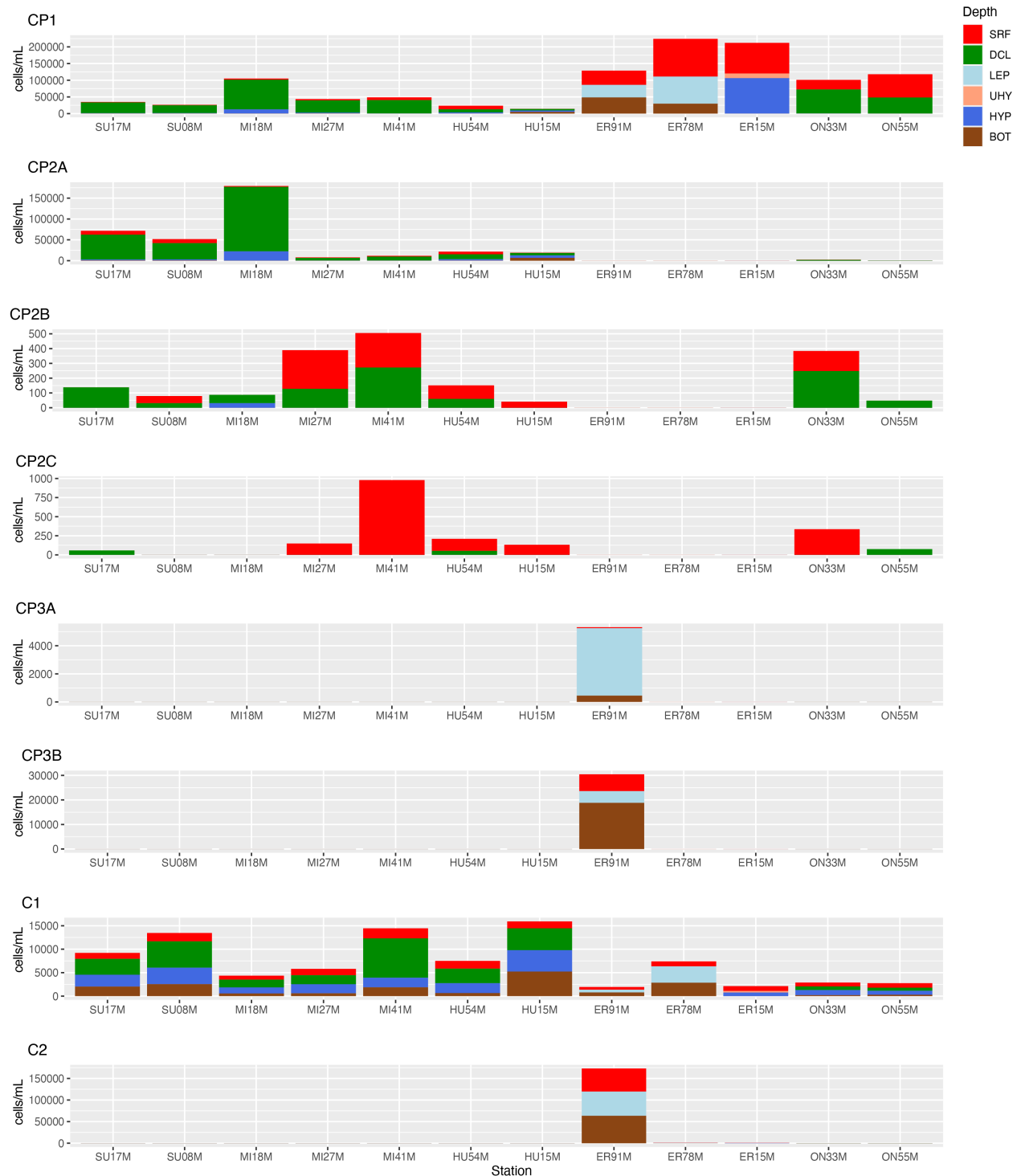
Supplemental Table 4. <i>petB</i> primers and PCR conditions.		
Primer set	Primer Sequence (5'->3')	PCR conditions
“outer” forward: petB_45F	BGTCTACGACTGGTTYVASGA	98 °C 30 seconds Repeat 10X 98 °C 10 seconds 68-58 °C 30 seconds * *decreased 1°C per cycle
“outer” reverse: petB_591R	GCTGTAGAAGCGGGTNAGNGT	72 °C 15 seconds Repeat 14X 98 °C 10 seconds 58 °C 30 seconds 72 °C 15 seconds ----- 72 °C 5 minutes 4 °C HOLD
“inner” forward: petB_59F3	TCAACGARGCNCTKGAGATYCA	98 °C 30 seconds Repeat 13X 98 °C 10 seconds 68-55 °C 30 seconds * *decreased 1°C per cycle
“inner” reverse: petB_539R2	AGYTCMACCATGAARTCDCC	72 °C 15 seconds Repeat 15X 98 °C 10 seconds 58 °C 30 seconds 72 °C 15 seconds ----- 72 °C 5 minutes 4 °C HOLD

Supplemental Figure 9. A) CP1 cells span two orders of magnitude of chlorophyll fluorescence. We gated brighter cells (CP1 bright, higher fluorescence) and dim cells (CP1 dim, lower fluorescence) to test for differences in cell size (forward scatter) relative to standard beads (3.2 μm diameter). B) CP1 bright cells have significantly higher forward scatter signals than CP1 dim cells (Chi Squared T(X)), suggesting formation of microcolonies by this population (CP1).



Supplemental Table 5. Similarity of <i>petB</i> gene sequences cloned from sorted cells relative to other known freshwater picocyanobacteria (Cabello-Yeves et al. 2022).			
Erie			
Sort Population ID	Description Best BLASTn hit	% Identity	Accession Best Hit
CP2A	<i>Synechococcus lacustris</i> Maggiore-St4-Slac Maggiore-St4-Slac-C2	98.95	JAGQBK0100000
CP2A	<i>Synechococcus lacustris</i> Cruz CV12-2 CV12-2-C65	98.1	JAGQDO0100000
CP2A	<i>Synechococcus lacustris</i> Cruz CV12-2 CV12-2-C65	98.52	JAGQDO0100000
CP2A	<i>Synechococcus lacustris</i> Maggiore-St4-Slac Maggiore-St4-Slac-C2	99.16	JAGQBK0100000
CP2A	<i>Synechococcus lacustris</i> Cruz CV12-2 CV12-2-C65	98.52	JAGQDO0100000
CP1	<i>Synechococcus lacustris</i> C3-12m-Tous C3-12m-Tous-C2	98.73	JAGQCN0100000
CP1	<i>Synechococcus lacustris</i> Cruz CV12-2 CV12-2-C65	98.31	JAGQDO0100000
CP1	<i>Synechococcus lacustris</i> C3-12m-Tous C3-12m-Tous-C2	98.31	JAGQCN0100000
CP1	<i>Synechococcus lacustris</i> Cruz CV12-2 CV12-2-C65	98.1	JAGQDO0100000
CP1	<i>Synechococcus lacustris</i> C3-12m-Tous C3-12m-Tous-C2	98.52	JAGQCN0100000
CP1_sub	<i>Cyanobium</i> sp. T1G-Tous T1G-Tous-C47	97.88	JAGQBD0100000
CP1_sub	<i>Cyanobium</i> sp. T1G-Tous T1G-Tous-C47	95.33	JAGQBD0100000
CP1_sub	<i>Cyanobium</i> sp. N.Huapi 1H5 Nhuapi-1H5-C3	89.34	JAGQBH0100000
CP1_sub	<i>Cyanobium</i> sp. N.Huapi 1H5 Nhuapi-1H5-C3	89.24	JAGQBH0100000
CP1_sub	<i>Cyanobium</i> sp. T1G-Tous T1G-Tous-C47	97	JAGQBD0100000
Superior			
Sort Population ID	Description Best BLASTn hit	% Identity	Accession Best Hit
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.57	JAGQEK0100000
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.57	JAGQEK0100000
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.57	JAGQEK0100000
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.36	JAGQEK0100000
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.57	JAGQEK0100000
CP1	<i>Synechococcus lacustris</i> C3-12m-Tous C3-12m-Tous-C2	80.54	JAGQCN0100000
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.99	JAGQEK0100000
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.78	JAGQEK0100000
CP1	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	96.2	JAGQEK0100000
CP2A	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	94.51	JAGQEK0100000
CP2A	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.36	JAGQEK0100000
CP2A	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	94.51	JAGQEK0100000
CP2A	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.57	JAGQEK0100000
CP2A	<i>Synechococcus</i> sp. Cruz CV12-2-Slac-r CV12-2-Slac-r-C46	95.78	JAGQEK0100000

Supplemental Figure 10. Stacked barplots of the number of cells per mL of each population (panels) across all stations and depths (colors). Abbreviations: Surface (SRF), deep chlorophyll layer (DCL), lower epilimnion (LEP), upper hypolimnion (UHY), water column between bottom and DCL (HYP), bottom or near bottom (BOT). Station locations are shown on the map on Figure 1. A heatmap of the same data presented here is in Figure 6.



Supplemental Table 6. Results of Chi Squared tests T(x), performed in FlowJo with *Compare Populations* comparing chlorophyll and phycoerythrin fluorescence of populations at different depths at Station SU08M. T(x) value greater than 4 corresponds to *p-value* < 0.01 (*).

Population	Comparison	T(x) Chi Squared	
		Chlorophyll (488 nm excited)	Phycoerythrin (488 nm excited)
CP1	Surface vs DCL*	42.4794	31.4151
	DCL vs 100 meters	0	7.0262
	Surface vs 100 meters*	37.9865	14.3087
CP2A	Surface vs DCL*	363.1506	364.1685
	DCL vs 100 meters	0.3253	9.8217
	Surface vs 100 meters*	93.6371	87.5736

Supplemental Table 7. Best hits for predicted photosynthetic eukaryote proteins in sorted population C1 from Superior and Erie.		
Superior C1		
Group	Species	n
Chlorophyta (Trebouxiophyceae)	<i>Choricystis parasitica</i>	25
Chlorophyta (Mamiellophyceae)	<i>Monomastix</i> sp. OKE-1	8
Chlorophyta (Trebouxiophyceae)	multiple	15
Chlorophyta (Chlorophyceae)	multiple	3
Ochrophyta (multiple)	multiple	4
Euglenozoa (Euglenophyceae)	multiple	2
Erie C1		
Group	Species	n
Chlorophyta (Trebouxiophyceae)	<i>Choricystis parasitica</i>	12
Chlorophyta (Mamiellophyceae)	<i>Monomastix</i> sp. OKE-1	7
Chlorophyta (Trebouxiophyceae)	multiple	11
Chlorophyta (Chlorophyceae)	multiple	1
Chlorophyta (Ulvophyceae)	multiple	2
Chlorophyta (Pyramimonadophyceae)	multiple	2
Cryptophyta (Cryptophyceae)	multiple	1
Haptophyta (Prymnesiophyceae)	multiple	2
Picozoa (Picomonadea)	multiple	1
Streptophyta	multiple	1