Fig. S1. Phylogeny and metadata of 1) 46 isolates of *Microcystis aeruginosa* collected from 14 inland lakes of Michigan, USA, 2). 20 publicly available sequences collected in multiple locations across six continents, and 3.) the cyanobacterium *Synechococcus* as an outgroup comparison. Multi-locus sequence typing was used to construct a phylogeny with RAxML based on five housekeeping genes (ftsZ, glnA, gltX, gyrB and pgi). Isolates originating from oligotrophic Michigan lakes are noted in dark blue, i.e. Low Phosphorus Lake/Low Phosphorus Genotype isolates. Isolates originating from eutrophic and mesotrophic Michigan lakes that clustered with oligotrophic lakes, i.e. High Phosphorus Lake/Low Phosphorus Genotype isolates, are noted in light-blue. All other isolates originating from eutrophic and mesotrophic lakes, i.e. High Phosphorus Lake/High Phosphorus Genotype isolates, are noted in green. Isolates obtained from NCBI are noted in gray.

Table. S1. Associated metadata about each of the 46 *M. aeruginosa* isolates originating from inland lakes of Michigan, USA. Note number of observations and years sampled corresponds to all data on nutrients and chlorophyll-a. Collection date corresponds to when the *M. aeruginosa* strains were collected.

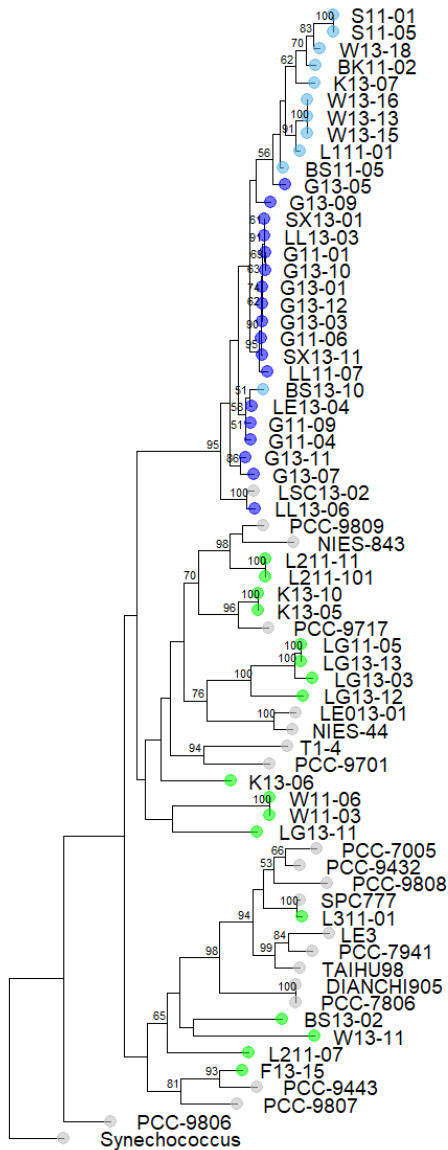| Strain Identification	Strain Collection Date	Location	Years Sampled	# Samplings	Mean TP (µg/L)	TP Range (µg/L)	Trophic State	Mean Chl-a (µg/L)	Mean SRP (µg/L)	Mean NH4 (µg/L)	Mean NO3 (µg/L)	Latitude	Longitude
BK11-02	8-Aug-11	Baker Lake, MI, USA	2009, 2011, 2013	4	28.0	11.5-54.6	meso.	33.7	4.6	34.9	52.3	42.64643	-85.50279
BS11-05;BS13-02,BS13-10	8-Aug-11	Baseline Lake, MI, USA	1998-1999, 2009, 2011, 2013	6	36.1	24.7-48.0	eutro.	47.1	3.6	19.0	31.1	42.42421	-85.5677
F13-15	15-Aug-13	Ford Lake, MI, USA	2009, 2011, 2013	3	65.0	44.8-99.6	eutro.	56.2	10.2	465.1	514.0	42.20619	-83.566
G11-01;G11-04;G11-06,G11-09	5-Jul-11	Gull Lake, MI, USA	1998-2014	201	7.6	2.3-13.1	oligo.	3.7	1.3	22.1	277.0	42.39651	-85.40936
G13-01;G13-03;G13-05;G13-07;G13-09; G13-10;G13-11;G13-12	8-Aug-13	Gull Lake, MI, USA	1998-2014	201	7.6	2.3-13.1	oligo.	3.7	1.3	22.1	277.0	42.39651	-85.40936
K13-05;K13-06;K13-07;K13-10	15-Aug-13	Kent Lake, MI, USA	2009, 2013	3	23.6	15.2-31.5	meso.	23.1	3.3	67.2	44.4	42.52346	-83.666
L111-01	19-Aug-11	MSU Lake 1, MI, USA	2009, 2011, 2013	3	163.5	71.0-209.9	eutro.	12.2	155.8	42.5	134.5	42.68059	-84.4831
L211-07;L211-01;L211-11	16-Aug-11	MSU Lake 2, MI, USA	2009-2013	79	196.8	105.3-456.2	eutro.	240.8	7.3	6.8	157.6	42.68059	-84.4871
L311-01	6-Aug-11	MSU Lake 3, MI, USA	2011, 2013	3	128.7	115.4-153.0	eutro.	53.3	4.1	9.4	543.7	42.67928	-84.4849
LE13-04	6-Aug-13	Lee Lake, MI, USA	2003, 2009, 2011, 2013	5	9.0	3.4-13.2	oligo.	4.0	1.9	5.3	40.4	42.17991	-85.11844
LG11-05	5-Aug-11	Lake Lansing, MI, USA	2011, 2013	2	17.1	16.6-17.6	meso.	5.5	5.1	11.1	89.0	42.76324	-84.405
LG13-02;LG13-03;LG13-11;LG13-12;LG13-1	14-Aug-13	Lake Lansing, MI, USA	2011, 2013	2	17.1	16.6-17.6	meso.	5.5	5.1	11.1	89.0	42.76324	-84.405
LL11-07	12-Jun-11	Little Long Lake, MI, USA	2011-2014	42	8.0	3.2-13.4	oligo.	4.1	1.0	100.0	312.5	42.41803	-85.44348
LL13-03;LL13-06	8-Aug-13	Little Long Lake, MI, USA	2011-2014	42	8.0	3.2-13.4	oligo.	4.1	1.0	100.0	312.5	42.41803	-85.44348
S11-01;S11-05	8-Aug-11	Sherman Lake, MI, USA	2009, 2011, 2013	4	13.7	4.8-24.0	meso.	9.1	2.8	0.0	64.9	42.35212	-85.38545
SX13-01;SX13-11	7-Aug-13	Sixteen Lake, MI, USA	2009, 2013	3	8.8	6.4-10.6	oligo.	5.0	1.0	127.5	1917.6	42.56518	-85.61352
W11-03;W11-06	2-Aug-11	Wintergreen Lake, MI, USA	2009, 2011, 2013	4	47.8	26.2-92.1	eutro.	21.4	2.7	136.0	129.7	42.39757	-85.38536
W13-11;W13-13;W13-15;W13-16;W13-18	8-Aug-13	Wintergreen Lake, MI, USA	2009, 2011, 2013	4	47.8	26.2-92.1	eutro.	21.4	2.7	136.0	129.7	42.39757	-85.38536
DIANCHI905		Dianchi Lake, China											
LE013-01		Lake Erie, USA											
LE3		Lake Erie, USA											
LSC13-02		Lake St. Clair, MI, USA											
NIES-44		Lake Kasumigaura Ibaraki, Japan											
NIES-843		Lake Kasumigaura Ibaraki, Japan											
PCC-7005		Lake Mendota, WI, USA											
PCC-7806		Braakman Reservoir, Netherlands											
PCC-7941		Little Rideau Lake, Canada											
PCC-9432		Little Rideau Lake, Canada											
PCC-9443		Fish Pond, Landjia, Central Afr. Rep.											
PCC-9701		Guerlesquin Dam, France											
PCC-9717		Rochereau Dam, France											
PCC-9806		Oskosh, WI, USA											
PCC-9807		Hartbeespoort Dam, South Africa											
PCC-9808		Malpas Dam, Australia											
PCC-9809		Lake Michigan, USA											
SPC777		Billings Resevoir, Brazil											
T1-4		Bangkok, Thailand											
TAIHU98		Lake Taihu, China											

Fig. S2. Location map of 14 *Microcystis* source lakes in the lower peninsula of Michigan, USA.
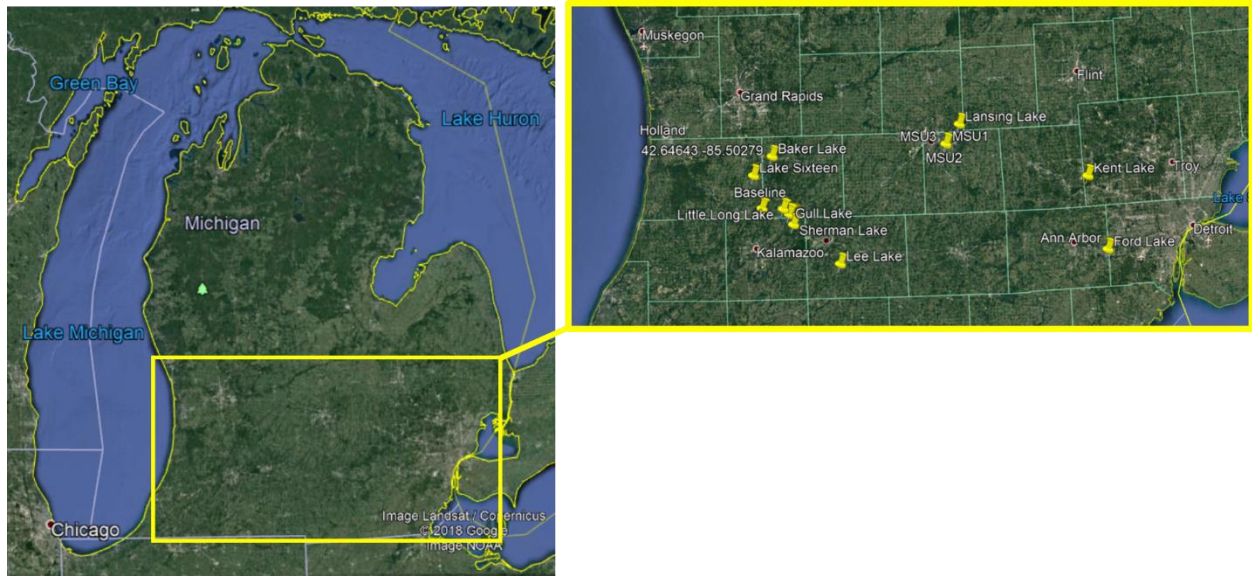
Fig. S3. Digital micrographs showing growth of an *M. aeruginosa* colony during a 6 – day growth assay. Images of F11-05 (isolated from Ford Lake, Michigan in 2011) were taken at 100x using a light microscope (Nikon Eclipse E600) interfaced with a digital camera (Diagnostic Instruments) and are shown to scale. Photo Credits: Jeffrey D. White.
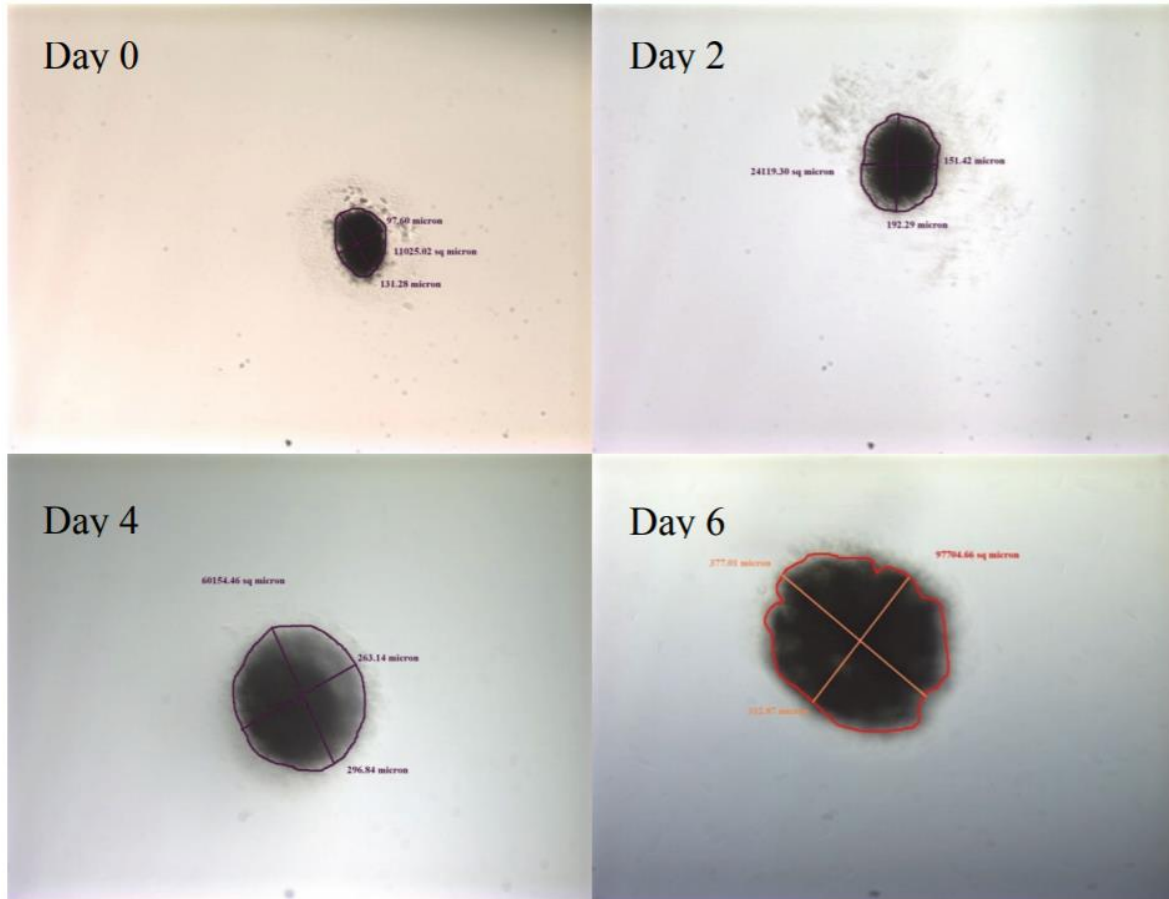
Fig. S4. A) Assembly statistics for *Microcystis aeruginosa* genomes, which includes all contigs at least 2kb in length that were binned using VizBin plus additional contigs that were shorter in length but assigned as a *Microcystis* spp. with ncbi-blast. Genomes in each of the three phylogenetic groups tended to have similar levels of coverage, but the Low Phosphorus Lake/Low Phosphorus Genotype and High Phosphorus Lake/Low Phosphorus Genotype genomes were more fragmented, as indicated by a greater number of contigs of a smaller mean contig length. B) Also shown is a comparison of assembly statistics when including versus excluding contigs under 2kb in length.

A.)
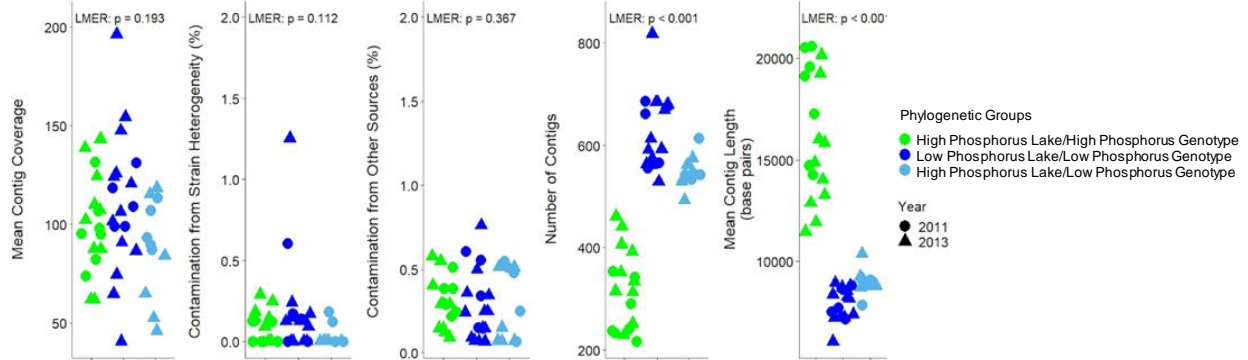
B.)
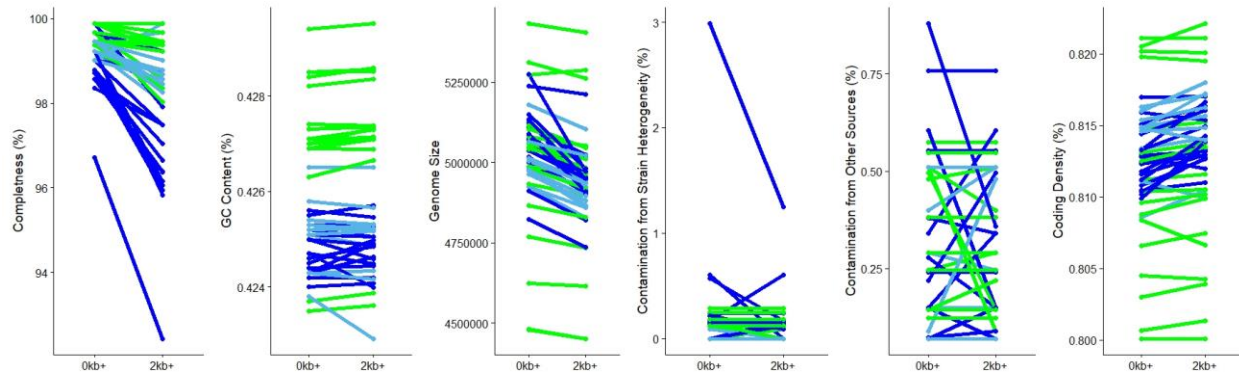
Fig S5. Abundance distributions of coverages for all contigs within the VizBin *Microcystis aeruginosa* bins are shown in black. Contigs that were taxonomically annotated as a *Microcystis* spp. according to ncbi-blast, but were below 2kb in length and therefore not included in VizBin binning, are shown in gray. Contigs were removed from our main analysis when coverage fell below the cutoffs illustrated in red.

High Phosphorus Lake/High Phosphorus Genotype Group:
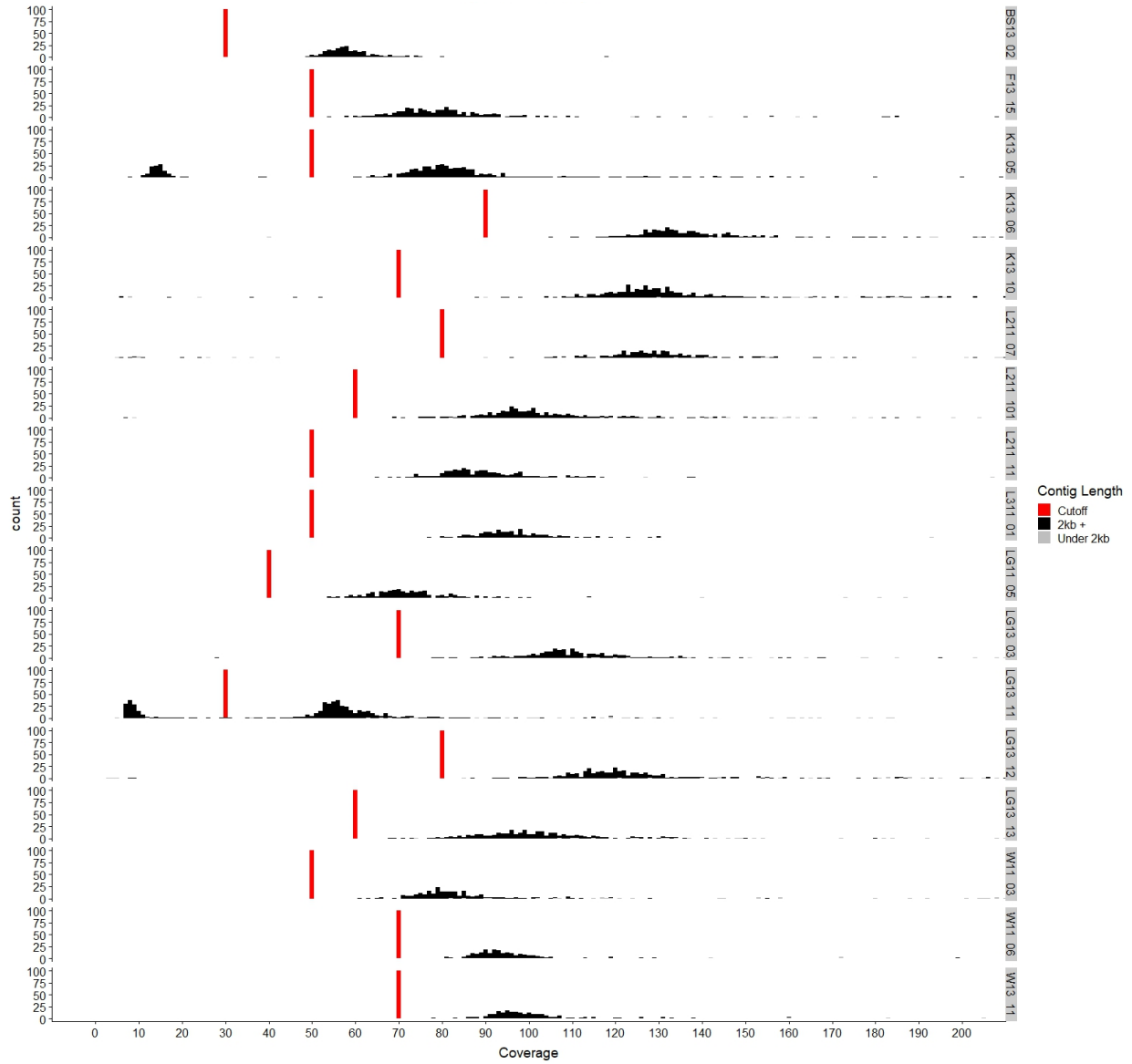
High Phosphorus Lake/Low Phosphorus Genotype Group:
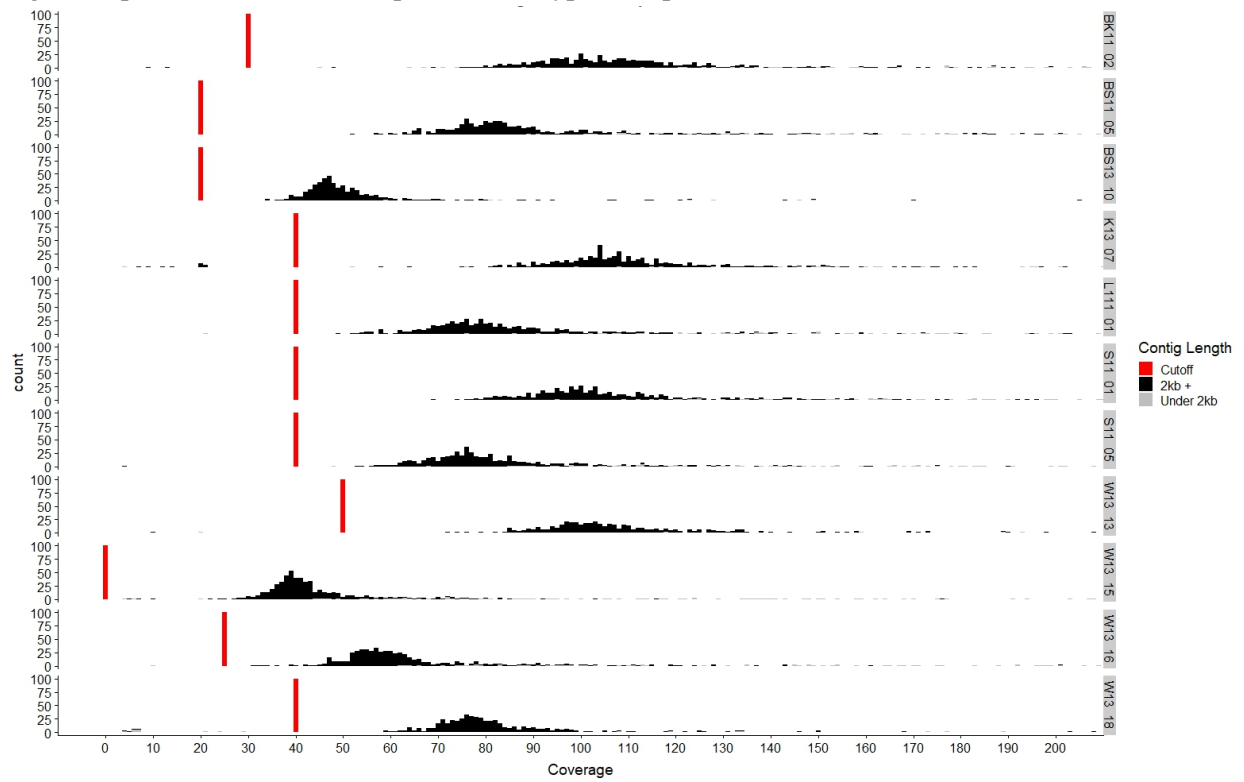
Low Phosphorus Lake/Low Phosphorus Genotype Group:
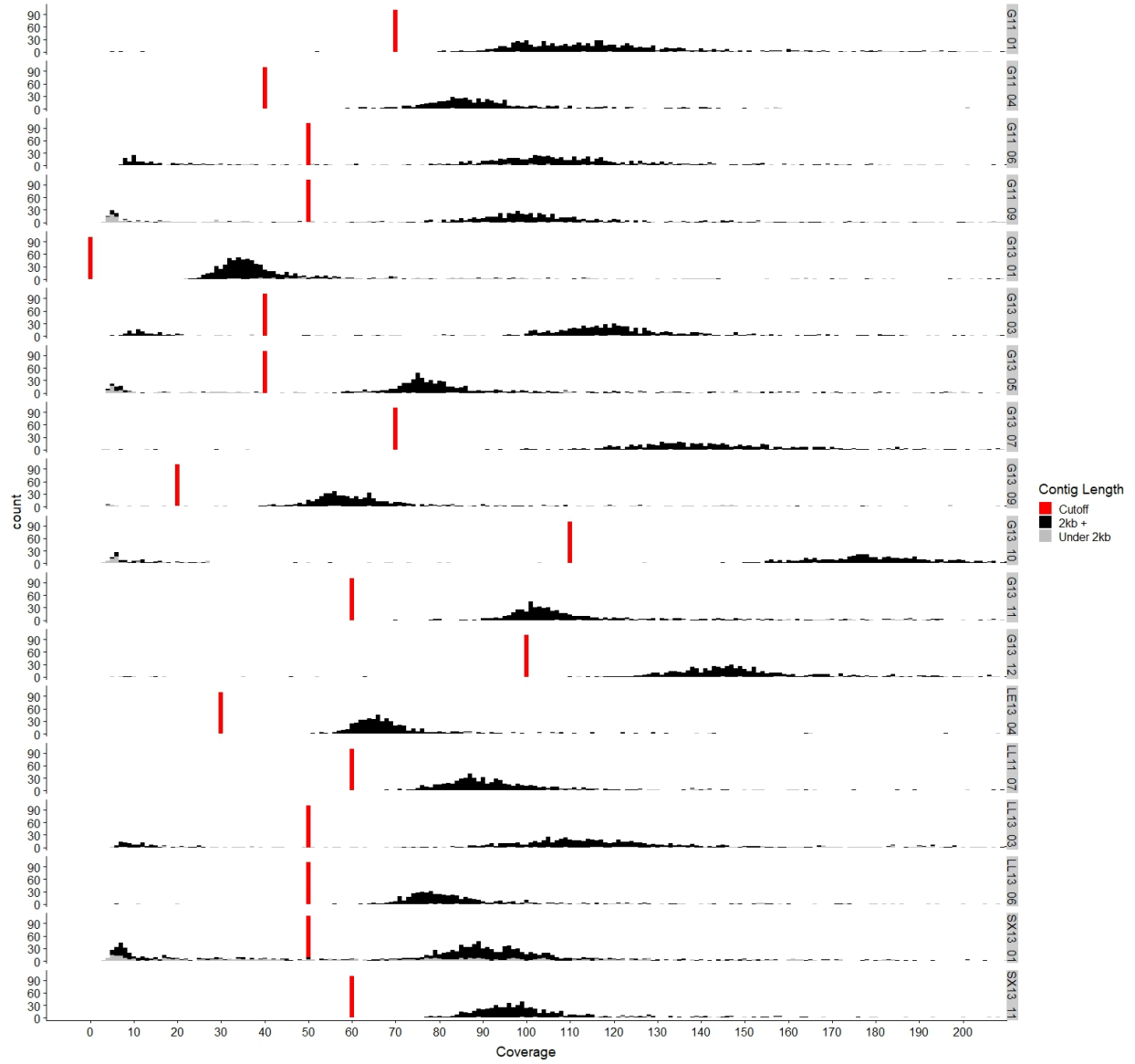
Fig. S6. The percentage of the reference versus alternate variant for each polymorphic site within a genome is illustrated with boxplots for each genome within the A) High Phosphorus Lake/High Phosphorus Genotype group, B) High Phosphorus Lake/Low Phosphorus Genotype group, or C) Low Phosphorus Lake/Low Phosphorus Genotype group. Genomes with a sizable distribution of low coverage contigs, likely caused by non-clonal cellular variation within colonies, are illustrated with a black fill color.

A)

B)
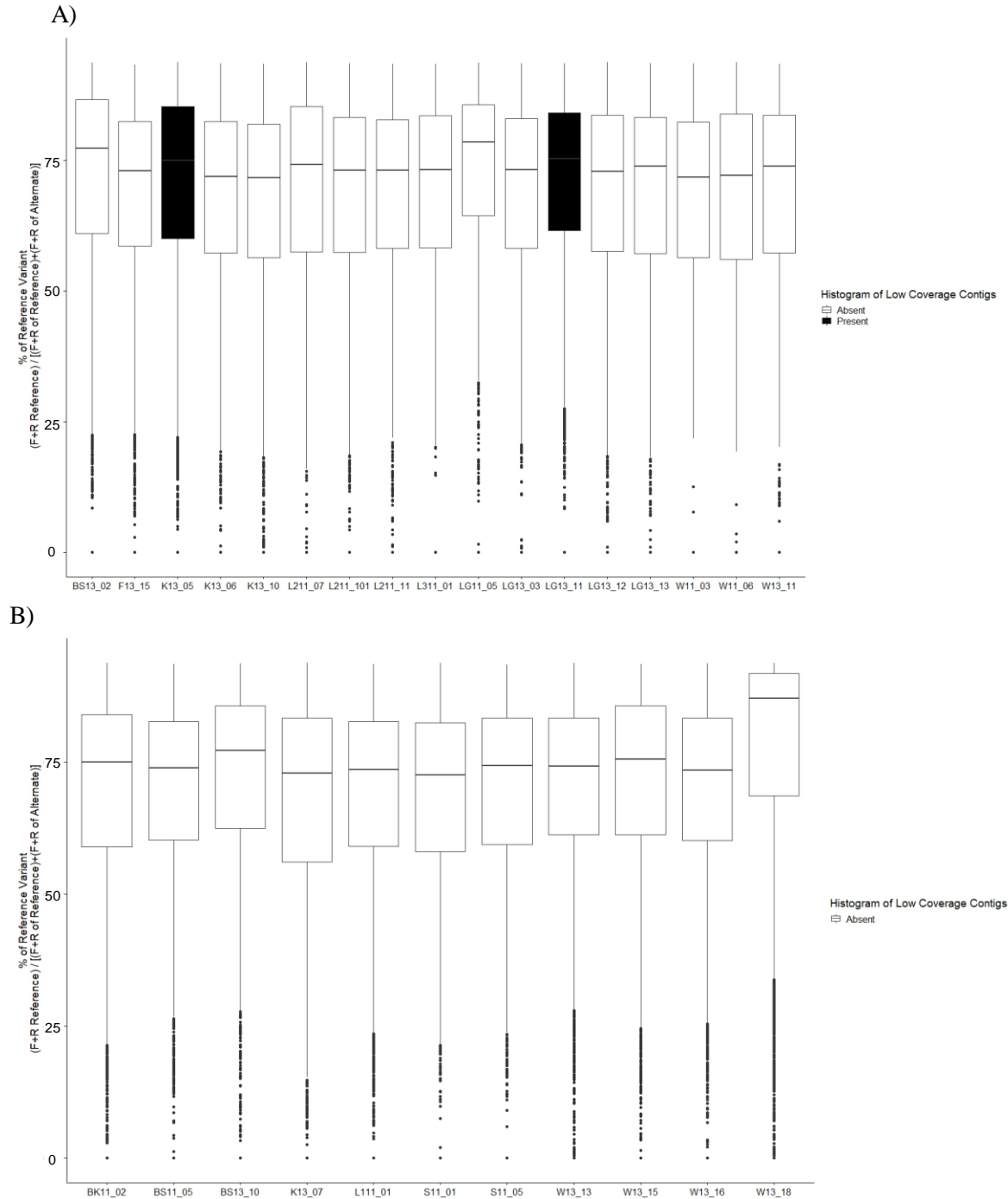
C)
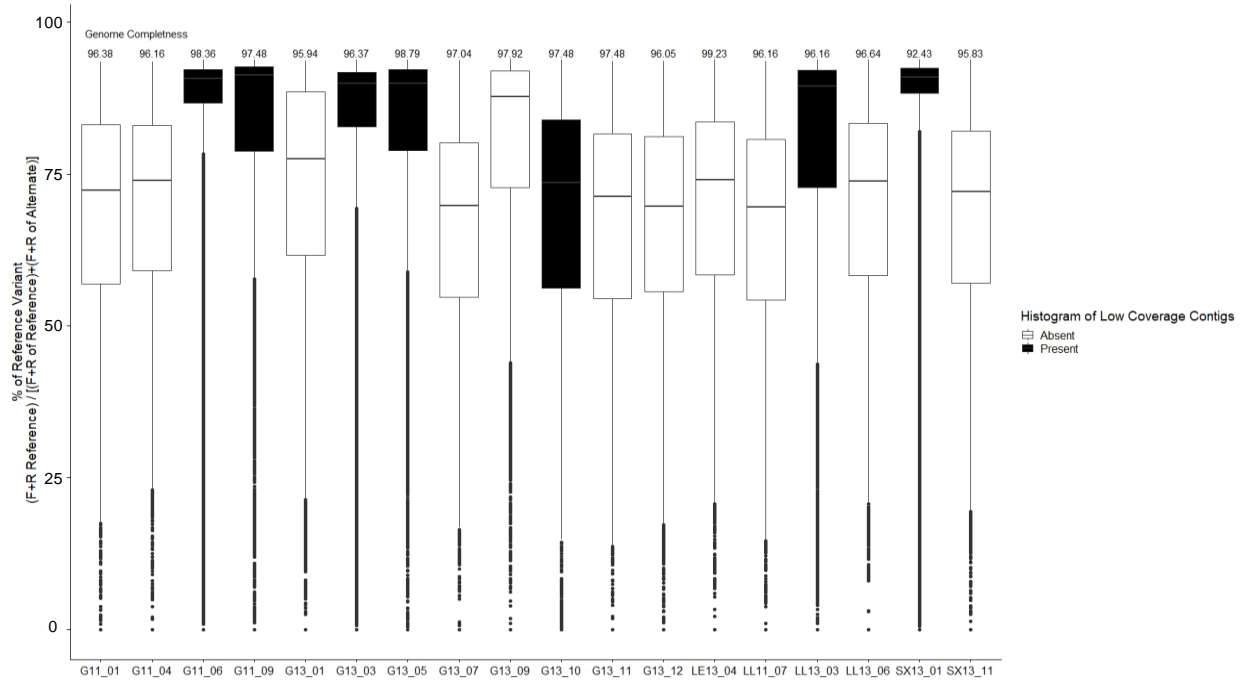
Fig. S7. As shown in Fig. S4 and Fig. 1, Low Phosphorus Lake/Low Phosphorus Genotype genomes tend to be less complete based on a survey of the occurrence of 524 core genes using checkM. The heat maps below show the percentage of genomes within each phylogenetic group that are missing each cyanobacterial core gene when considering A) all binned contigs of at least 2kb in length using VizBin, or additionally, B) all binned contigs and those under 2kb that were identified as a *Microcystis* spp. using ncbi-blast. Shown below are core genes found in fewer than 44 isolates. Black, or a value of 100, indicates all genomes within that phylogenetic group contained a particular core gene. White, or a value of 0 indicates all genomes within that phylogenetic group lacked that particular core gene.
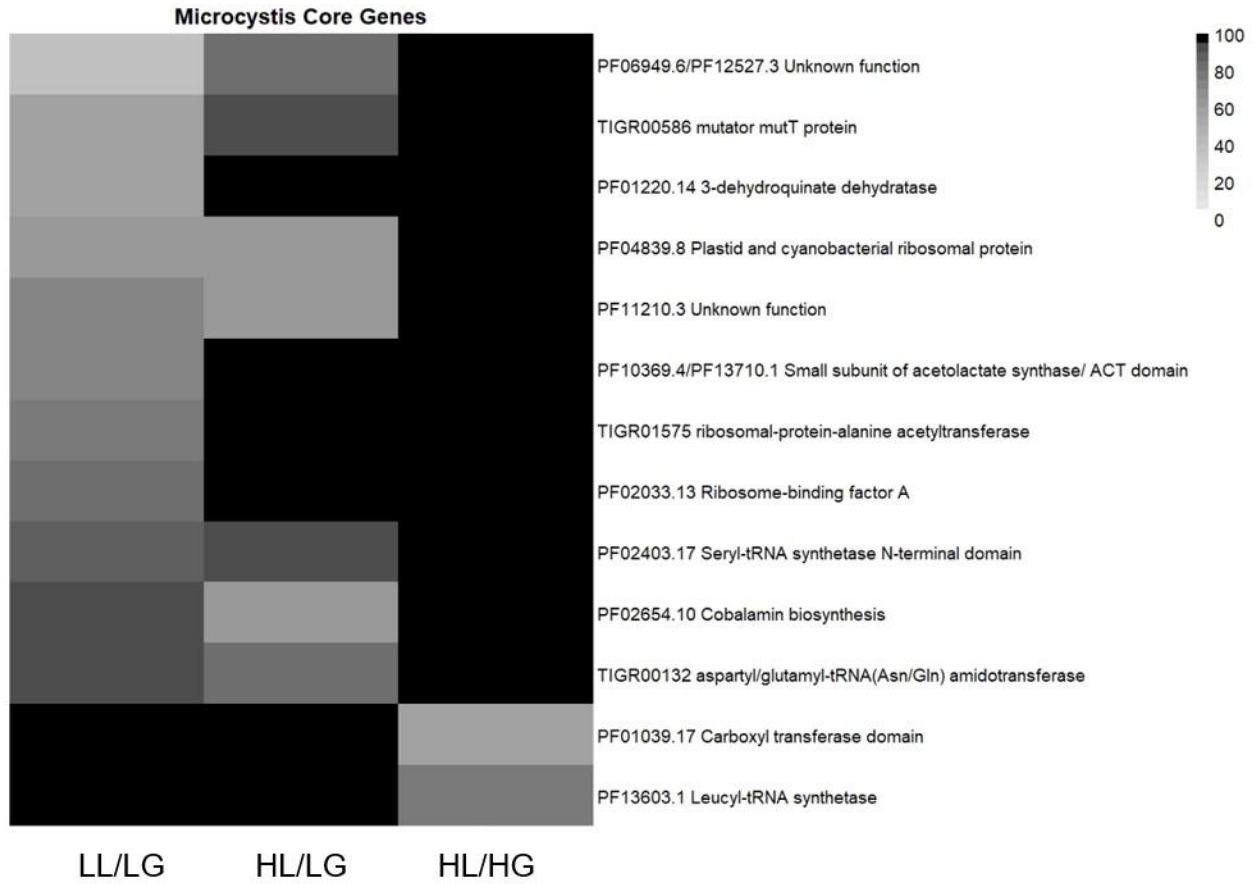
A.)

B.)

Microcystis Core Genes

PF06949.6/PF12527.3 Unknown function

TIGR00586 mutator mutT protein

PF01220.14 3-dehydroquinate dehydratase

PF04839.8 Plastid and cyanobacterial ribosomal protein

PF11210.3 Unknown function

PF10369.4/PF13710.1 Small subunit of acetolactate synthase/ ACT domain

TIGR01575 ribosomal-protein-alanine acetyltransferase

PF02033.13 Ribosome-binding factor A

PF02403.17 Seryl-tRNA synthetase N-terminal domain

PF02654.10 Cobalamin biosynthesis

TIGR00132 aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase

PF01039.17 Carboxyl transferase domain

PF13603.1 Leucyl-tRNA synthetase

LL/LG   HL/LG   HL/HG

100
80
60
40
20
0

Fig. S8 Among isolates of *Microcystis aeruginosa* collected from inland lakes of Michigan, numerous Kegg Orthology (KO) and protein families (pfam) terms were found in significantly different abundances across the three different phylogenetic groups of genomes. Note KO terms related to nitrogen and phosphorus metabolism and transport are highlighted first, followed by all other significant terms. We show results for terms that varied significantly via Analysis of Variance with a false discovery rate correction. To control for multiple isolates per lake, we input only the average gene count per isolate within each lake into the ANOVA. Heatmap color depicts average gene count per isolate for each phylogenetic group, where lighter colors indicate fewer genes per isolate occurring on average in that protein family within that phylogenetic group. Note for lakes with multiple phylogenetic groups, we include separate mean values for each group of isolates within that lake. Lettering within heatmap cells indicates which phylogenetic groups differ by Tukey's post-hoc test, where groupings sharing the same letter do not differ. A total of 671 pfams were significant at the $p < 0.05$-level. For conciseness, we show those terms with a correct p-value < 0.001, followed by those terms with a p-value < 0.01 if the pfam was entirely absent from at least one of the three phylogenetic groups.

Microcystis: Nitrogen and Phosphorus Transport/Metabolism
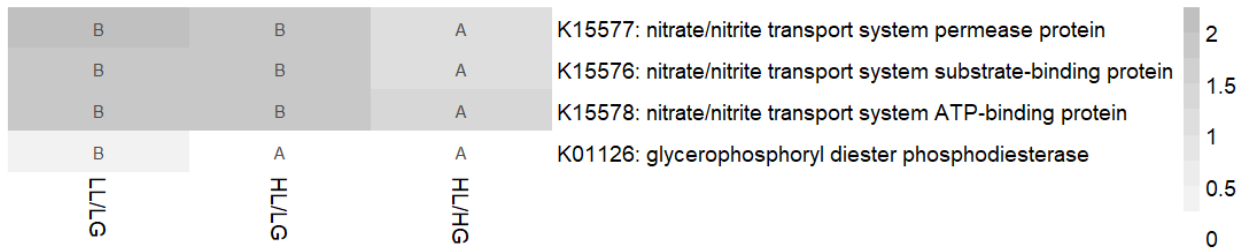
Microcystis: All other functions

LL/LG	HL/LG	HL/HG	
B	B	A	K09803: uncharacterized protein
B	B	A	K02036: phosphate transport system ATP-binding protein
B	B	A	K03320: ammonium transporter
B	B	A	K07089: uncharacterized protein
B	B	A	K03684: ribonuclease D
B	B	A	K02286: phycocyanin-associated rod linker protein
B	B	A	K06188: aquaporin Z
B	B	A	K02640: cytochrome b6-f complex subunit 5
B	A	B	K03088: RNA polymerase sigma-70 factor, ECF subfamily
B	A	B	K00555: tRNA guanine26-N2/guanine27-N2-dimethyltransferase
B	A	A	K03496: chromosome partitioning protein
B	A	A	K01091: phosphoglycolate phosphatase
B	A	A	K06871: uncharacterized protein
B	A	A	K20074: PPM family protein phosphatase
B	A	A	K00505: tyrosinase
A	B	B	K03574: 8-oxo-dGTP diphosphatase
A	B	B	K02428: XTP/dITP diphosphohydrolase
A	B	B	K01265: methionyl aminopeptidase
A	B	B	K03722: ATP-dependent DNA helicase DinG
A	B	B	K17758: ADP-dependent NAD(P)H-hydrate dehydratase
A	B	B	K17759: NAD(P)H-hydrate epimerase
A	B	B	K03786: 3-dehydroquinate dehydratase II
A	B	B	K00762: orotate phosphoribosyltransferase
A	B	B	K00031: isocitrate dehydrogenase
A	B	B	K05371: phycocyanobilin:ferredoxin oxidoreductase
A	B	B	K02705: photosystem II CP43 chlorophyll apoprotein
A	B	B	K07646: two-component system, OmpR family, sensor histidine kinase KdpD
A	B	B	K00801: farnesyl-diphosphate farnesyltransferase
A	A	B	K03671: thioredoxin 1
A	A	B	K01537: P-type Ca2+ transporter type 2C
A	A	B	K03116: sec-independent protein translocase protein TatA
A	A	B	K04035: magnesium-protoporphyrin IX monomethyl ester
A	A	B	K03707: thiaminase
A	A	B	K03789: ribosomal protein S18-alanine N-acetyltransferase
A	A	B	K02709: photosystem II PsbH protein
A	A	B	K03325: arsenite transporter
A	A	B	K01304: pyroglutamyl-peptidase
A	A	B	K19092: toxin ParE1/3/4

Scale: 0, 2, 4, 6, 8, 10, 12, 14

Microcystis

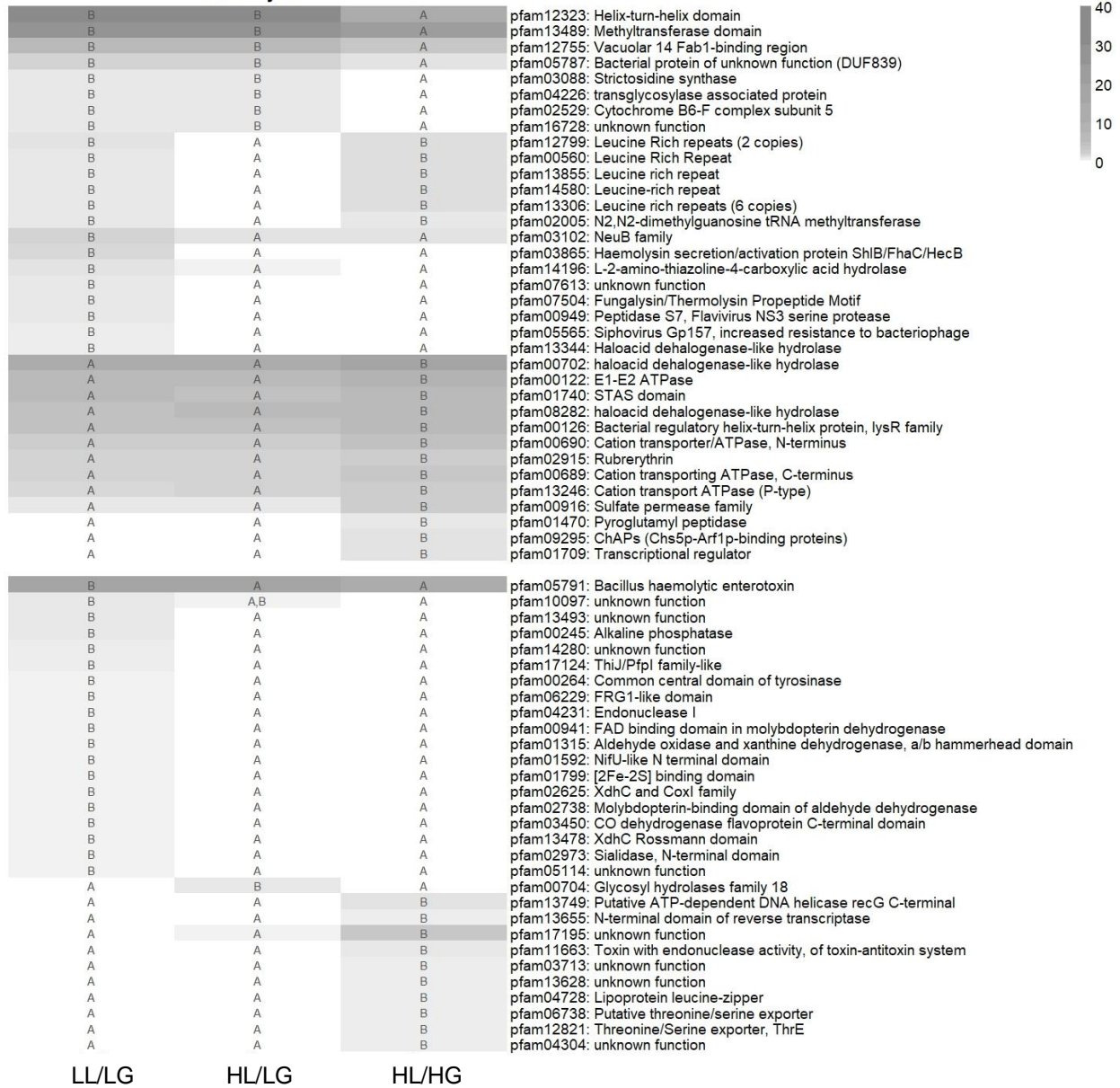| LL/LG	HL/LG	HL/HG	Pfam
B	B	A	pfam12323: Helix-turn-helix domain
B	B	A	pfam13489: Methyltransferase domain
B	B	A	pfam12755: Vacuolar 14 Fab1-binding region
B	B	A	pfam05787: Bacterial protein of unknown function (DUF839)
B	B	A	pfam03088: Strictosidine synthase
B	B	A	pfam04226: transglycosylase associated protein
B	B	A	pfam02529: Cytochrome B6-F complex subunit 5
B	B	A	pfam16728: unknown function
B	A	B	pfam12799: Leucine Rich repeats (2 copies)
B	A	B	pfam00560: Leucine Rich Repeat
B	A	B	pfam13855: Leucine rich repeat
B	A	B	pfam14580: Leucine-rich repeat
B	A	B	pfam13306: Leucine rich repeats (6 copies)
B	A	B	pfam02005: N2,N2-dimethylguanosine tRNA methyltransferase
B	A	A	pfam03102: NeuB family
B	A	A	pfam03865: Haemolysin secretion/activation protein ShlB/FhaC/HecB
B	A	A	pfam14196: L-2-amino-thiazoline-4-carboxylic acid hydrolase
B	A	A	pfam07613: unknown function
B	A	A	pfam07504: Fungalysin/Thermolysin Propeptide Motif
B	A	A	pfam00949: Peptidase S7, Flavivirus NS3 serine protease
B	A	A	pfam05565: Siphovirus Gp157, increased resistance to bacteriophage
B	A	A	pfam13344: Haloacid dehalogenase-like hydrolase
A	A	B	pfam00702: haloacid dehalogenase-like hydrolase
A	A	B	pfam00122: E1-E2 ATPase
A	A	B	pfam01740: STAS domain
A	A	B	pfam08282: haloacid dehalogenase-like hydrolase
A	A	B	pfam00126: Bacterial regulatory helix-turn-helix protein, lysR family
A	A	B	pfam00690: Cation transporter/ATPase, N-terminus
A	A	B	pfam02915: Rubrerythrin
A	A	B	pfam00689: Cation transporting ATPase, C-terminus
A	A	B	pfam13246: Cation transport ATPase (P-type)
A	A	B	pfam00916: Sulfate permease family
A	A	B	pfam01470: Pyroglutamyl peptidase
A	A	B	pfam09295: ChAPs (Chs5p-Arf1p-binding proteins)
A	A	B	pfam01709: Transcriptional regulator
B	A	A	pfam05791: Bacillus haemolytic enterotoxin
B	A,B	A	pfam10097: unknown function
B	A	A	pfam13493: unknown function
B	A	A	pfam00245: Alkaline phosphatase
B	A	A	pfam14280: unknown function
B	A	A	pfam17124: ThiJ/PfpI family-like
B	A	A	pfam00264: Common central domain of tyrosinase
B	A	A	pfam06229: FRG1-like domain
B	A	A	pfam04231: Endonuclease I
B	A	A	pfam00941: FAD binding domain in molybdopterin dehydrogenase
B	A	A	pfam01315: Aldehyde oxidase and xanthine dehydrogenase, a/b hammerhead domain
B	A	A	pfam01592: NifU-like N terminal domain
B	A	A	pfam01799: [2Fe-2S] binding domain
B	A	A	pfam02625: XdhC and CoxI family
B	A	A	pfam02738: Molybdopterin-binding domain of aldehyde dehydrogenase
B	A	A	pfam03450: CO dehydrogenase flavoprotein C-terminal domain
B	A	A	pfam13478: XdhC Rossmann domain
B	A	A	pfam02973: Sialidase, N-terminal domain
B	A	A	pfam05114: unknown function
A	B	A	pfam00704: Glycosyl hydrolases family 18
A	A	B	pfam13749: Putative ATP-dependent DNA helicase recG C-terminal
A	A	B	pfam13655: N-terminal domain of reverse transcriptase
A	A	B	pfam17195: unknown function
A	A	B	pfam11663: Toxin with endonuclease activity, of toxin-antitoxin system
A	A	B	pfam03713: unknown function
A	A	B	pfam13628: unknown function
A	A	B	pfam04728: Lipoprotein leucine-zipper
A	A	B	pfam06738: Putative threonine/serine exporter
A	A	B	pfam12821: Threonine/Serine exporter, ThrE
A	A	B	pfam04304: unknown function

Table S2. Baseline data for the core microbiome of *Microcystis aeruginosa* as determine via sequencing of the 16S rRNA gene and clustering of sequences sharing at least 97% sequence similarity into OTUs. Read depths are reported before and after normalization to the smallest sample depth. All richness and diversity metrics were calculated as defined in the phyloseq microbiome analysis package in R.

Sample	Lake	Original Read Depth	Scaled Read Depth	Observed Richness	Chao1 (± s.e.)	ACE (± s.e.)	Shannon	Simpson	Inverse Simpson	Fisher
BK11_02	Baker	48709	10169	31	40.00 (8.03)	47.09 (3.35)	1.71	0.69	3.27	3.95
BS11_05	Baseline	53613	10183	23	24.50 (2.23)	29.89 (2.84)	1.83	0.80	4.92	2.81
BS13_02	Baseline	66222	10128	54	60.11 (4.97)	65.37 (4.14)	2.52	0.86	7.22	7.49
BS13_10	Baseline	62799	10163	35	40.14 (4.65)	44.33 (3.47)	1.78	0.75	3.98	4.54
F11_05	Ford	41063	10179	25	30.00 (6.01)	33.29 (2.80)	1.22	0.49	1.95	3.09
F13_03	Ford	43554	10167	32	60.00 (21.47)	44.54 (3.27)	2.16	0.85	6.69	4.09
F13_15	Ford	38281	10163	29	31.63 (2.83)	33.29 (2.81)	0.95	0.55	2.21	3.66
G11_01	Gull	48274	10154	36	40.00 (3.69)	45.49 (3.74)	2.10	0.83	5.94	4.69
G11_04	Gull	44946	10166	36	48.00 (10.75)	43.60 (3.17)	1.91	0.80	4.99	4.69
G11_06	Gull	41686	10171	36	39.50 (3.44)	43.33 (3.05)	2.21	0.85	6.87	4.69
G11_09	Gull	55919	10164	42	49.00 (6.65)	50.30 (3.66)	1.95	0.77	4.40	5.60
G13_01	Gull	43963	10162	48	65.27 (10.23)	75.97 (6.03)	0.81	0.32	1.47	6.53
G13_03	Gull	39462	10173	24	24.50 (1.03)	25.98 (2.48)	1.46	0.62	2.63	2.95
G13_05	Gull	52079	10162	37	46.33 (8.85)	44.09 (3.20)	1.45	0.68	3.10	4.84
G13_07	Gull	65576	10177	21	24.75 (4.20)	30.35 (2.98)	1.40	0.65	2.89	2.53
G13_09	Gull	33769	10156	51	57.50 (4.88)	60.37 (3.75)	0.64	0.22	1.29	7.01
G13_10	Gull	39231	10090	93	94.57 (1.70)	97.54 (4.65)	2.83	0.91	11.06	14.15
G13_11	Gull	57219	10170	35	80.50 (34.53)	67.58 (3.91)	1.86	0.78	4.52	4.54
G13_12	Gull	48004	10162	35	42.86 (6.33)	48.36 (3.77)	1.18	0.54	2.20	4.54
K13_01	Kent	46618	10173	29	29.60 (1.18)	30.69 (2.36)	2.17	0.84	6.27	3.66
K13_05	Kent	40105	10149	48	64.50 (12.89)	74.97 (4.49)	2.74	0.91	11.55	6.53
K13_06	Kent	40594	10146	49	66.00 (10.99)	73.10 (4.58)	1.63	0.69	3.23	6.69
K13_07	Kent	66186	10169	32	33.50 (2.23)	34.40 (2.73)	2.38	0.88	8.59	4.09
K13_10	Kent	54218	10169	44	46.50 (3.16)	48.38 (3.34)	2.85	0.92	12.17	5.90
L111_01	MSU1	54821	10203	85	98.32 (7.38)	110.37 (5.42)	1.49	0.60	2.49	12.71
L211_07	MSU2	61600	10171	32	37.00 (5.53)	39.35 (2.69)	2.05	0.82	5.66	4.09
L211_101	MSU2	68003	10161	31	33.00 (2.88)	33.84 (2.79)	1.84	0.78	4.45	3.95
L211_11	MSU2	46633	10170	30	35.60 (5.34)	37.13 (2.83)	1.69	0.77	4.34	3.80
L311_01	MSU3	36485	10144	77	132.20 (32.73)	109.30 (5.28)	2.87	0.91	11.04	11.33
LE13_04	Lee	49235	10167	52	62.11 (7.19)	73.81 (5.03)	2.31	0.83	5.84	7.16
LG11_05	Lansing	41699	10166	51	62.67 (8.00)	65.80 (4.72)	2.18	0.85	6.54	7.00
LG13_02	Lansing	50838	10177	30	33.00 (4.16)	35.98 (2.58)	2.18	0.83	5.87	3.80
LG13_03	Lansing	53286	10171	33	35.00 (2.88)	36.20 (2.00)	2.24	0.84	6.26	4.24
LG13_11	Lansing	36832	10181	24	25.50 (2.23)	28.17 (2.66)	1.71	0.71	3.44	2.95
LG13_12	Lansing	49862	10169	30	36.00 (5.38)	40.33 (3.33)	1.50	0.71	3.49	3.80
LG13_13	Lansing	44345	10141	51	62.00 (8.87)	62.29 (3.80)	2.23	0.82	5.59	7.01
LL11_07	Little Long	49425	10173	24	34.50 (10.52)	34.33 (3.11)	1.95	0.83	6.03	2.95
LL13_03	Little Long	44497	10164	35	40.25 (5.37)	41.62 (3.18)	2.10	0.84	6.25	4.54
LL13_06	Little Long	56382	10158	47	50.11 (3.10)	51.15 (3.36)	2.32	0.86	7.34	6.37
S11_01	Sherman	38563	10167	51	72.11 (12.60)	83.29 (4.89)	2.01	0.82	5.68	7.00
S11_05	Sherman	55158	10159	52	56.50 (3.92)	59.01 (3.69)	2.24	0.83	5.82	7.16
SX13_01	Sixteen	47387	10152	52	54.50 (3.16)	54.35 (3.49)	2.34	0.81	5.22	7.17
SX13_11	Sixteen	46615	10168	32	41.17 (7.37)	48.13 (3.91)	1.50	0.70	3.28	4.09
W11_03	Wintergreen	48147	10140	75	95.00 (10.56)	104.84 (5.58)	1.95	0.79	4.76	10.98
W11_06	Wintergreen	62231	10139	37	44.50 (6.35)	45.82 (3.29)	1.66	0.76	4.25	4.84
W13_11	Wintergreen	53677	10142	67	91.00 (16.42)	83.38 (4.22)	2.70	0.89	8.84	9.63
W13_13	Wintergreen	47236	10156	38	46.25 (6.36)	50.52 (3.53)	1.66	0.69	3.26	4.99
W13_15	Wintergreen	41440	10150	52	71.13 (12.05)	73.12 (4.15)	1.94	0.80	4.96	7.17
W13_16	Wintergreen	34930	10154	46	57.38 (8.08)	61.88 (4.02)	2.20	0.86	6.93	6.22
W13_18	Wintergreen	40374	10159	53	87.00 (23.21)	74.71 (4.19)	2.47	0.89	9.17	7.32

Table S3. Description of the core microbiome of *Microcystis aeruginosa* as determine via sequencing of the 16S rRNA gene and clustering of sequences sharing at least 97% sequence similarity into OTUs. Fifteen bacterial OTUs were associated with > 75% of *M. aeruginosa* isolated from inland lakes of Michigan. We also note a total of 34 OTUs that were associated with at least 50% of isolates, including all OTUs with relative abundances above 2%, and OTUs disproportionately associated with isolates belonging to different phylogenetic groups. All abundances shown as mean ± standard error.

Taxonomy	Total % Abundance	LL/LG Occurrence	HL/LG Occurrence	HL/HG Occurrence	LL/LG % Abundance	HL/LG % Abundance	HL/HG % Abundance
Proteobacteria;Alphaproteobacteria;Candidatus_Phycosocius_bacilliformis	11.32 ± 2.04 (46)	18	11	17	6.34 ± 2.77	10.59 ± 3.32	19.57 ± 3.96
Proteobacteria;Alphaproteobacteria;Caulobacterales;alfII_A;Brev	7.96 ± 2.61 (46)	18	11	17	18.06 ± 6.44	3.02 ± 1.36	2.36 ± 1.23
Bacteroidetes;Cytophagia;Cytophagales;Cytophagaceae	10.47 ± 2.60 (43)	17	9	17	18.21 ± 5.60	5.69 ± 4.64	7.64 ± 2.91
Proteobacteria;unclassified Betaproteobacteria	1.67 ± 0.31 (42)	18	8	16	2.25 ± 0.59	1.57 ± 0.50	1.37 ± 0.53
Proteobacteria;Alphaproteobacteria;Caulobacterales;alfII	2.65 ± 0.69 (41)	17	10	14	5.09 ± 1.57	2.01 ± 0.90	0.48 ± 0.41
Proteobacteria;Betaproteobacteria;Burkholderiales;betI_A	1.49 ± 0.72 (39)	15	10	14	1.19 ± 0.65	0.02 ± 0.01	0.34 ± 0.26
Proteobacteria;Alphaproteobacteria;Rhodospirillales;alfVIII	2.52 ± 0.73 (38)	17	7	14	5.32 ± 1.68	0.17 ± 0.15	0.72 ± 0.49
Gemmatimonadetes;Gemmatimonadetes;Gemmatimonadales;Gemmatimonadaceae;Gemmatimonas	2.80 ± 0.82 (37)	13	9	15	0.65 ± 0.44	7.50 ± 3.06	1.68 ± 0.61
Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium	0.16 ± 0.05 (38)	16	8	14	0.09 ± 0.06	0.12 ± 0.08	0.21 ± 0.10
Proteobacteria;Alphaproteobacteria;Caulobacterales;alfII_A;Brev	1.29 ± 0.53 (39)	16	10	13	0.53 ± 0.23	1.12 ± 1.10	1.74 ± 1.10
Proteobacteria;Alphaproteobacteria;Caulobacterales;Caulobacteraceae(96);Brevundimonas	2.84 ± 1.48 (37)	16	8	13	3.59 ± 2.13	7.29 ± 5.60	0.01 ± 0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium	0.18 ± 0.12 (37)	14	10	13	0.40 ± 0.33	0.12 ± 0.11	0.03 ± 0.01
Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium	0.09 ± 0.05 (37)	12	11	14	0.02 ± 0.01	0.30 ± 0.23	0.04 ± 0.03
Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium	0.02 ± 0.00 (36)	15	9	12	0.02 ± 0.01	0.03 ± 0.01	0.01 ± 0.01
Proteobacteria;Alphaproteobacteria;Sphingomonadales;alfIII_A;Sphingo	0.89 ± 0.87 (36)	13	9	14	0.12 ± 0.08	0.02 ± 0.01	2.53 ± 2.52
Proteobacteria;Alphaproteobacteria;Sphingomonadales;alfIII_A;Sphingo	0.25 ± 0.25 (35)	14	8	13	0.68 ± 0.67	0.01 ± 0.01	0.01 ± 0.00
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae	1.20 ± 0.37 (33)	12	10	11	1.11 ± 0.43	1.46 ± 0.76	0.21 ± 0.12
Proteobacteria;Alphaproteobacteria;Rhizobiales	0.46 ± 0.16 (33)	15	7	11	0.74 ± 0.31	0.47 ± 0.36	0.27 ± 0.23
Planctomycetes;Planctomycetacia;Planctomycetales;Planctomycetaceae;Planctomyces	2.04 ± 0.80 (30)	10	8	12	0.32 ± 0.31	3.70 ± 2.47	1.13 ± 0.69
Proteobacteria;Alphaproteobacteria;Sphingomonadales;MN_122.2a	0.25 ± 0.08 (29)	12	7	10	0.13 ± 0.08	0.45 ± 0.18	0.32 ± 0.19
Cyanobacteria;Cyanobacteria;Pseudanabaena	1.98 ± 1.55 (30)	10	6	14	0.02 ± 0.01	0.04 ± 0.02	0.03 ± 0.01
Actinobacteria;Actinobacteria;Actinomycetales;acI_C2	0.01 ± 0.01 (28)	12	6	10	0.02 ± 0.02	0.01 ± 0.00	0.01 ± 0.00
Actinobacteria;Actinobacteria;Actinomycetales;acI_B1	0.01 ± 0.00 (29)	9	8	12	0.01 ± 0.01	0.01 ± 0.00	0.01 ± 0.00
Proteobacteria;Alphaproteobacteria;Caulobacterales;Caulobacteraceae	1.24 ± 0.61 (28)	9	6	13	0.04 ± 0.04	0.88 ± 0.59	2.97 ± 1.65
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingopyxis	1.24 ± 0.69 (27)	10	6	11	0.01 ± 0.01	0.62 ± 0.58	1.57 ± 1.03
Proteobacteria;unclassified Alphaproteobacteria	1.28 ± 0.43 (27)	10	5	12	3.18 ± 1.03	0.29 ± 0.28	0.23 ± 0.15
Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium	0.31 ± 0.15 (26)	9	7	10	0.22 ± 0.19	0.06 ± 0.03	0.37 ± 0.25
Unclassified proteobacteria	0.62 ± 0.37 (26)	7	6	13	0.01 ± 0.00	0.04 ± 0.03	1.79 ± 1.02
Proteobacteria;Alphaproteobacteria;Rhodospirillales;I-10	1.02 ± 0.37 (23)	9	5	9	1.11 ± 0.75	1.78 ± 0.94	0.69 ± 0.41
Proteobacteria;Alphaproteobacteria;Rhizobiales	0.60 ± 0.32 (25)	8	6	11	0.30 ± 0.21	1.62 ± 1.35	0.25 ± 0.14
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingomonas	0.78 ± 0.40 (26)	12	5	9	1.23 ± 0.71	0.03 ± 0.02	0.12 ± 0.08
Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Meganema	0.43 ± 0.17 (24)	10	7	7	0.47 ± 0.25	0.01 ± 0.00	0.21 ± 0.16
Proteobacteria;Betaproteobacteria;Burkholderiales	0.48 ± 0.31 (25)	8	6	11	0.21 ± 0.20	1.32 ± 1.30	0.36 ± 0.25
Unclassified bacteria	0.34 ± 0.13 (24)	5	8	11	0.00 ± 0.00	0.43 ± 0.26	0.67 ± 0.33
Proteobacteria;Deltaproteobacteria;Myxococcales	1.79 ± 0.88 (23)	11	5	7	3.56 ± 1.85	2.44 ± 2.42	0.01 ± 0.01
Bacteroidetes;Cytophagia;Cytophagales;Cytophagaceae;Chryseolinea	2.02 ± 1.44 (23)	10	8	5	0.09 ± 0.06	4.68 ± 4.60	2.94 ± 2.93
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Elstera	0.76 ± 0.59 (23)	8	9	6	0.49 ± 0.49	2.53 ± 2.51	0.13 ± 0.12
Proteobacteria;unclassified Alphaproteobacteria	0.48 ± 0.28 (23)	7	6	10	0.06 ± 0.04	1.94 ± 1.17	0.13 ± 0.12

Fig. S9. Illustration of the core microbiome of *Microcystis aeruginosa* as determine via sequencing of the 16S rRNA gene and clustering of sequences sharing at least 97% sequence similarity into OTUs. We illustrate (A) the taxonomic similarities among bacterial communities inhabiting the phycosphere of *M. aeruginosa* collected from the same lake. Points close in principal coordinate space are taxonomically more similar bacterial communities, where taxonomic relationships were inferred using a Bray-Curtis dissimilarity matrix on the 16S gene. (B) Protein functionality of these bacterial communities were also similar among those inhabiting host isolates originating from the same lake. Points close in principal coordinate space are functionally more similar bacterial communities, as determined using a Bray-Curtis dissimilarity metric on a gene count matrix categorized into protein families. Note for both panels, significance of separation was determined using analysis of variance on distance matrices, i.e. adonis.
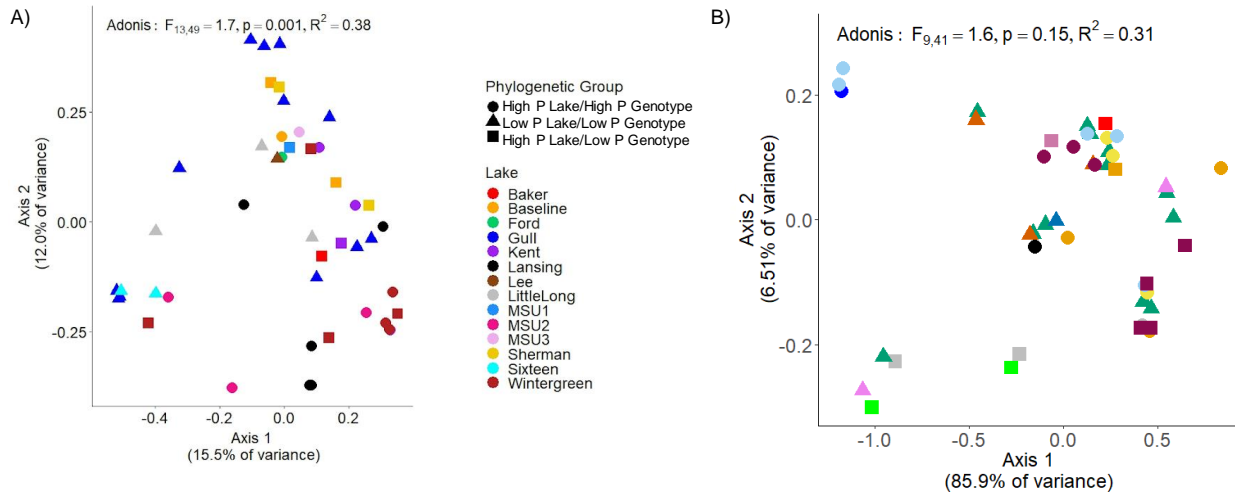
Table S4. The bacterium *Phycosocius bacilliformis* was identified via 16S rRNA sequencing in the phycospheres of each of 46 strains of *M. aeruginosa* that had been isolated from 14 lakes in Michigan, USA (see Fig. S9 for details on 16S survey data). Genomes of *P. bacilliformis* were identified from *M. aeruginosa* metagenomes using ESOM. Seven high quality genomes were identified that best represented 7 different strains of *P. bacilliformis*. Each representative genome was at least 96% complete and was divergent from all other strains by at least 0.75% average amino acid identity. We show metabolic complementarity between the *M. aeruginosa* host and associated *P. bacilliformis* using the JGI IMG Annotation Pipeline predictions for amino acid and galactose metabolism.
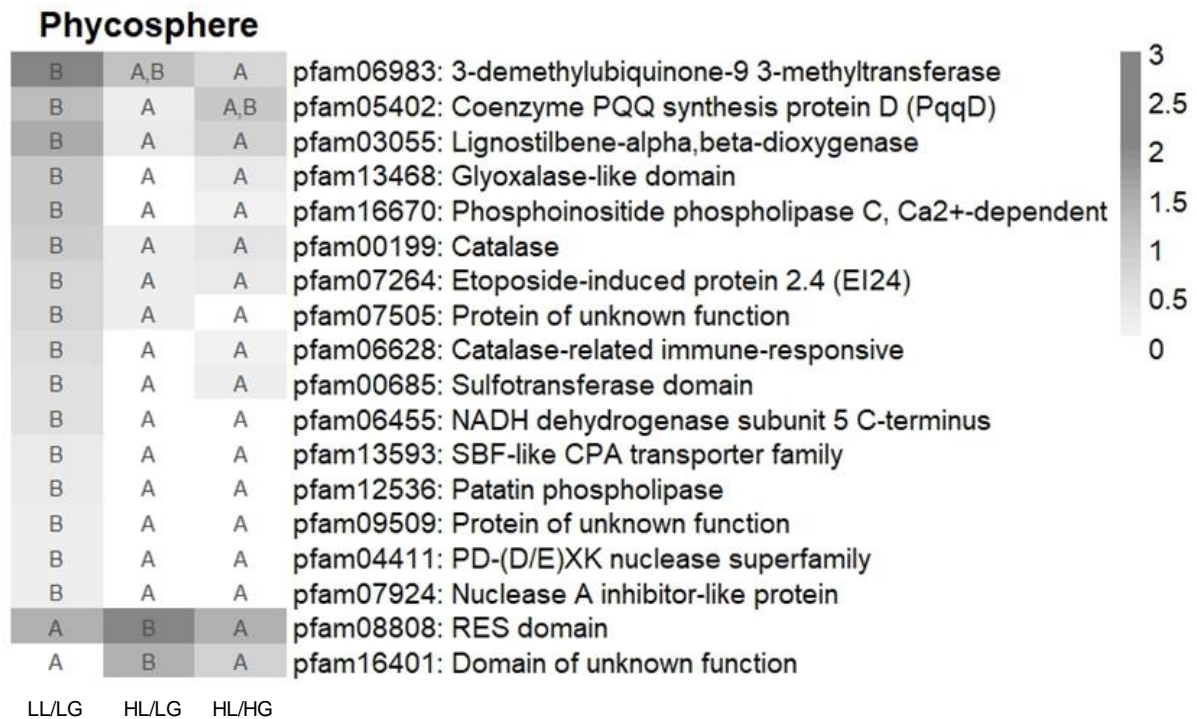
Average Amino Acid Identity between *P. bacilliformis* strains associated with *M. aeruginosa*

	LG13-03	LG13-12	W11-06	K13-06	K13-07	W13-15	G11-06
LG13-03	0	93.74	99.23	98.22	99.29	98.02	93.4
LG13-12		0	93.65	93.49	93.56	93.63	96.74
W11-06			0	98.03	99.09	98.02	93.23
K13-06				0	97.99	98.59	93.57
K13-07					0	98.01	93.31
W13-15						0	93.41
G11-06							0

		Completeness	Contamination	Threonine	Serine	Asparagine	Galactose Utilization
LG13-03	*M. aeruginosa*	NA	NA	✗	✗	✗	NA
	P. bacilliformis	97.8	0.89	✓	✓	✓	✓
LG13-12	*M. aeruginosa*	NA	NA	✗	✗	✗	NA
	P. bacilliformis	97.8	0.89	✓	✓	✓	✗
W11-06	*M. aeruginosa*	NA	NA	✗	✗	✗	NA
	P. bacilliformis	97.8	0.89	✓	✓	✓	✓
K13-06	*M. aeruginosa*	NA	NA	✗	✗	✓	NA
	P. bacilliformis	98.5	0.24	✓	✓	✓	✓
K13-07	*M. aeruginosa*	NA	NA	✗	✗	✗	NA
	P. bacilliformis	96.8	0.35	✓	✓	✓	✗
W13-15	*M. aeruginosa*	NA	NA	✗	✗	✗	NA
	P. bacilliformis	98.5	0.24	✓	✓	✓	✓
G11-06	*M. aeruginosa*	NA	NA	✗	✗	✗	NA
	P. bacilliformis	98.1	0.41	✓	✓	✓	✗

Fig. S10.

Within the *M. aeruginosa* phycosphere, some protein families (pfam) and Kegg Orthology (KO) terms were found in different abundances across LL/LG, HL/HG, and HL/HG genomes. We show results for terms that varied by Analysis of Variance. As no terms were significant with a false discovery rate correction, we report terms with uncorrected p-values below 0.01 to highlight terms that may differ among groups but acknowledging the potential for more false positives. To control for multiple host isolates per lake, we input only the average gene count for each lake into the ANOVA. Heatmap color depicts average gene counts within an entire phycosphere rather than per genome. All pfams with an uncorrected p-value below 0.01 are shown. Lighter heatmap colors indicate fewer genes occurring on average in that protein family within that phylogenetic group. Note for lakes with multiple phylogenetic groups, we include separate mean values for each group of isolates within that lake. Lettering within heatmap cells indicates which phylogenetic groups differ by Tukey's post-hoc tests, where groups sharing the same letter do not differ.
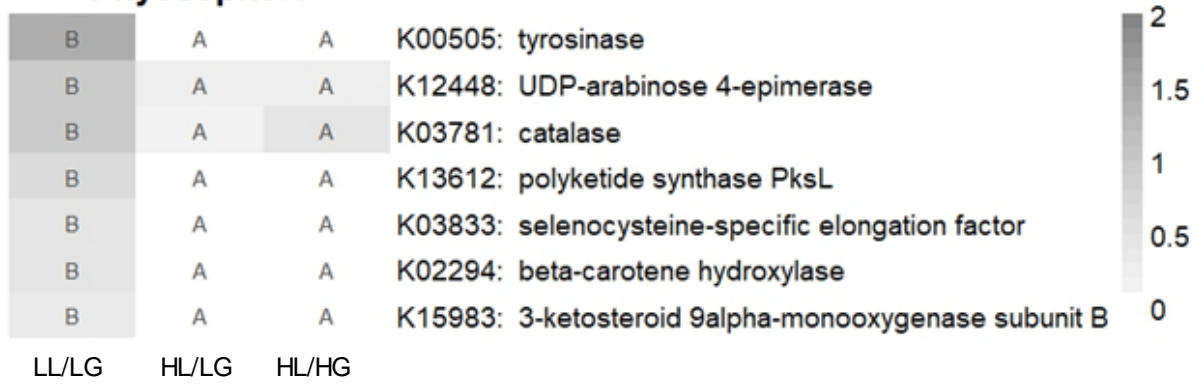
Phycosphere

LL/LG	HL/LG	HL/HG	pfam
B	A,B	A	pfam06983: 3-demethylubiquinone-9 3-methyltransferase
B	A	A,B	pfam05402: Coenzyme PQQ synthesis protein D (PqqD)
B	A	A	pfam03055: Lignostilbene-alpha,beta-dioxygenase
B	A	A	pfam13468: Glyoxalase-like domain
B	A	A	pfam16670: Phosphoinositide phospholipase C, Ca2+-dependent
B	A	A	pfam00199: Catalase
B	A	A	pfam07264: Etoposide-induced protein 2.4 (EI24)
B	A	A	pfam07505: Protein of unknown function
B	A	A	pfam06628: Catalase-related immune-responsive
B	A	A	pfam00685: Sulfotransferase domain
B	A	A	pfam06455: NADH dehydrogenase subunit 5 C-terminus
B	A	A	pfam13593: SBF-like CPA transporter family
B	A	A	pfam12536: Patatin phospholipase
B	A	A	pfam09509: Protein of unknown function
B	A	A	pfam04411: PD-(D/E)XK nuclease superfamily
B	A	A	pfam07924: Nuclease A inhibitor-like protein
A	B	A	pfam08808: RES domain
A	B	A	pfam16401: Domain of unknown function

Scale: 3, 2.5, 2, 1.5, 1, 0.5, 0

Phycosphere

LL/LG	HL/LG	HL/HG	
B	A	A	K00505: tyrosinase
B	A	A	K12448: UDP-arabinose 4-epimerase
B	A	A	K03781: catalase
B	A	A	K13612: polyketide synthase PksL
B	A	A	K03833: selenocysteine-specific elongation factor
B	A	A	K02294: beta-carotene hydroxylase
B	A	A	K15983: 3-ketosteroid 9alpha-monooxygenase subunit B

2
1.5
1
0.5
0

Fig. S11. Additional growth rate data from Wilson et al. (2006) demonstrated a consistent relationship between total phosphorus of the source lake and maximum growth rate via linear regression. These 12 additional strains originated from 12 lakes from the same geographic region (lower Michigan) as the current study. The studies included one lake in common, Gull Lake, although Wilson et al. recorded a TP level of 19.7 ug/L, which exceeds the range observed during our survey of Gull Lake TP over 16 years (see Table S1). Note growth rates from the current study were determined by repeatedly photographing single colonies over a 6-day growth assay, whereas Wilson et al. measured growth rate via cell counts in batch culture. However, we have found these approaches yield similar results (White, J. D. "Trait and environmental variation mediate the interaction between a harmful phytoplankter and an invasive grazer." PhD diss. Michigan State University, 2015.; page # 116, paired t-test, n = 8, df = 7, p = 0.71).
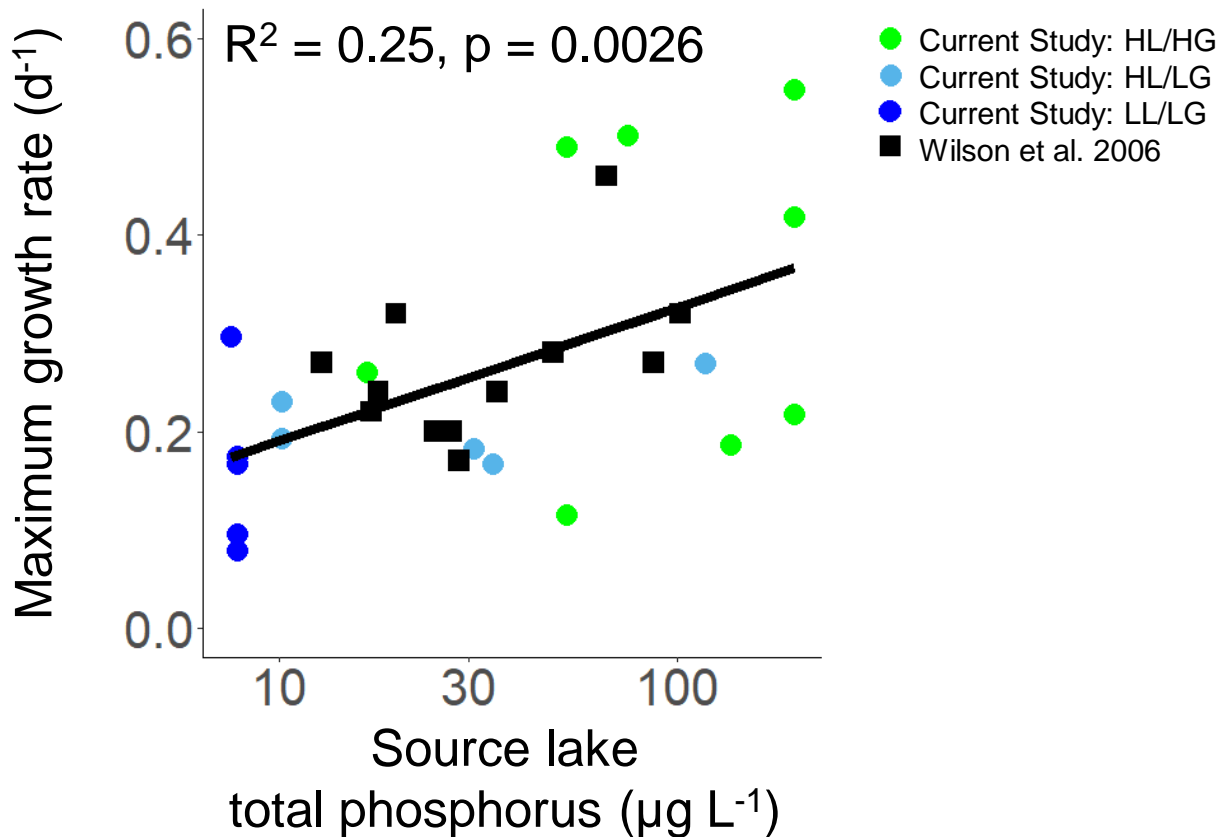
Fig. S12. Expanded figure corresponding to Fig. 2. of the main text. Phylogeny of 46 isolates of *M. aeruginosa* collected from 14 inland lakes in Michigan, USA. Multi-locus sequencing typing was used to infer evolutionary history with RAxML based on five concatenated housekeeping genes (FtsZ, glnA, gltX, gyrB and pgi). Dark blue: isolates from oligotrophic lakes ('Low Phosphorus Lake, Low Phosphorus Genotype LL/LG'); light blue: isolates from phosphorus-rich lakes, but related to oligotrophic isolates ('High Phosphorus Lake/Low Phosphorus Genotype, HL/LG'); green: isolates from phosphorus-rich lakes ('High Phosphorus Lake/High Phosphorus Genotype, HL/HG'). All significant trends, as determined using linear mixed effects models that control for collection date and lake of origin, are noted with one asterisk at the $p < 0.10$ level and two asterisks at the $p < 0.05$ level. Group means are shown with a dashed line. Except for genome size, which is shown in megabases, all metrics are percentage data. Note that genome size, completeness, and GC content consider all contigs, regardless of length, while coding DNA, paralogs, and sigma factors as a percentage of total genes considers only contigs 2kb in length and longer. Significance of post-hoc pairwise comparisons are noted with lettering above dashed lines, where groups sharing the same letter do not significantly differ from each other. Nineteen of the 20 publicly available sequences collected worldwide were most closely related to the HL/HG group (Fig. S1).