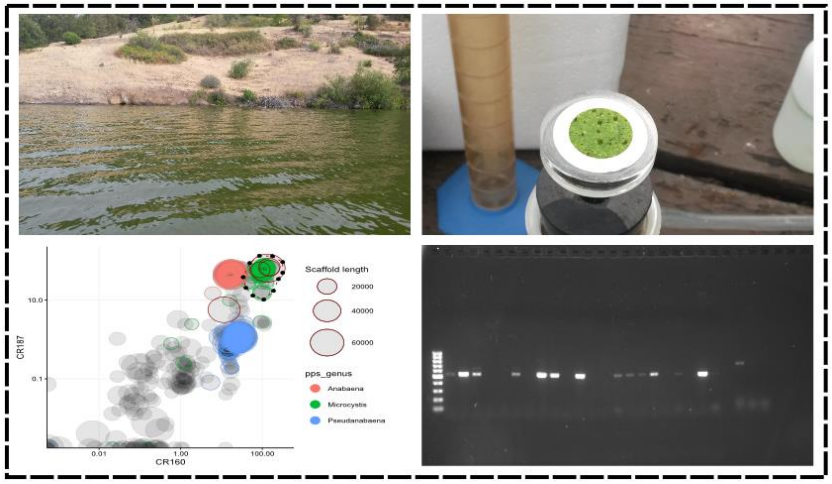
Multi-year analysis of *Microcystis* population structure and toxigenicity in Copco and Iron Gate Reservoirs



PREPARED BY: Dr. Tim Otten & Dr. Theo Dreher
Oregon State University, Department of Microbiology
226 Nash Hall, Corvallis, OR 97331

Final Report - 4/7/2017

TABLE OF CONTENTS

OVERVIEW

This report provides the results of analyses on *Microcystis* population structure in Copco and Iron Gate reservoirs over a 9-year period (2007-2015). We used a combination of analysis of variance (ANOVA) and change point statistics to determine if changes in *Microcystis* population structure (i.e., strain composition) coincided with specific environmental conditions. Changes in *Microcystis* population structure were inferred from time-series assessments of the c-phycocyanin gene (*cpcBA*) sequences and the grouping of these sequences into distinct operational taxonomic units (OTUs); OTUs can be thought of as strains or subspecies. In Copco and Iron Gate reservoirs, data from the past 8-9 years indicates that one strain of *Microcystis*–a microcystin producer–tended to be the dominant strain during most years. However, there were periods when this strain was replaced by other strains that we believe to be nontoxic on the basis of our prior work. Sequencing of the *cpcBA* genetic marker indicated that four or five different allelotypes (clustered into OTUs) comprised the Klamath River *Microcystis* population. We propose that an understanding of the environmental drivers that promote nontoxic strain dominance could lead to novel reservoir management strategies aimed at reducing microcystin toxicity within the reservoirs and the river below.

STUDY ASSUMPTIONS

1) Environmental conditions determine strain dominance through niche adaptation. Top-down controls such as zooplankton grazing, viruses, or other predators were not assessed.
2) Any changes in environmental conditions that led to alterations in *Microcystis* strain composition will be large enough to identify with statistical confidence.
3) Relationships between cells and toxins assume that all toxin occurs intracellularly (i.e., particulate fraction) since the extracellular fraction was not measured.

MATERIALS & METHODS

In this study, only samples collected near the surface [0.1 meter (m) depth for public health samples or 0.5 m depth for environmental samples] were considered in the analysis. Public health samples and environmental metadata were available from 2005 – 2015. Genetic analyses did not begin until 2007 in Copco Reservoir and 2008 in Iron Gate Reservoir. All physicochemical and toxin data was generated according to Klamath Hydroelectric Settlement Agreement (KHSA) water quality monitoring program protocols. These data are presumed to have already undergone quality assurance and quality control checks. The genetic data were generated and analyzed as detailed below.

Genetics analyses. Water samples were collected from 0.5 m depth grabs from upstream of Copco (KR19874) and Iron Gate (KR19019) dams and stored on wet ice in 1 liter (L) plastic bottles. These bottles were delivered to the Dreher laboratory at Oregon State University, where they were concentrated by vacuum filtration onto glass fiber filters (1.2 micrometer (µm) pore size, 25 millimeter (mm) diameter). As much volume as possible, without clogging, was passed through each filter. The filters were placed into microcentrifuge tubes and stored at -80°C

until further processing. Total deoxynucleic acid (DNA) was extracted from each filter using commercially available DNA extraction kits (GeneRite EZ-DNA RWOC1 extraction kits or Fisher SurePrep Soil DNA Isolation kits). Both kits incorporate a combination of physical (bead beating) and chemical lysis steps. The purified DNA was used in all subsequent analyses. The prevalence of individual *Microcystis* strains present in each sample was determined by deep sequencing of the c-phycocyanin photopigment genes (*cpcBA*) that all cyanobacteria possess. Approximately 650 base pairs (bp) of these genes were amplified in triplicate polymerase chain reaction (PCR) assays consisting of 10 microliters (µL) of 2.5X Taq Master Mix (5-Prime), 1 µL DNA, 12 µL nuclease-free water and 1 µL each (10 µM) of the following forward and reverse primers:

(PCBF; 5'- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCYGCTTGYYTACGSGACA-3')

(PCAR; 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACCCRGTWCCACCAGCAACTA-3')

Both primers were amended to contain Illumina adapter sequences (shown in blue in the sequences above) in order to allow for sample indexing (Nextera XT index primers) and sequencing on an Illumina MiSeq sequencer. Thermal cycling conditions for the first round of PCR consisted of an initial denaturation step at 94°C for 3 min, followed by 22 cycles of 94°C for 45 sec, 54.5°C for 60 sec, 72°C for 90 sec, then a final elongation step at 72°C for 10 min. After the first round of PCR, amplicons from each set of triplicate PCR reactions were pooled together and purified using an Agencourt AMPure XP magnetic bead kit. The purified amplicons were subjected to an additional eight cycles of PCR in order to ligate the Illumina index (barcodes) onto the amplicons. The PCR products were purified a second time using an Agencourt AMPure XP magnetic bead kit. The barcoded amplicons were normalized to equimolar concentrations using a SequalPrep Normalization Plate kit, pooled and sequenced using a single flow cell of an Illumina MiSeq sequencer (v2 chemistry; 2x250 bp). A total of 56 samples were bioinformatically analyzed from Copco Reservoir, 45 from Iron Gate Reservoir, 14 from Upper Klamath Lake middle of the trench (MDT; USGS site ID #422305121553800), 4 from JC Boyle Reservoir, and 2 each from Link River Dam and Keno Reservoir.

After sequencing, the adapters and primers were removed and the sequences were quality filtered with the program Cutadapt using a minimum quality score of 26, which corresponds to 99.75% or greater accuracy in base calls (A,T,G,C) across the full sequencing read. Additionally, reads shorter than 200 nucleotides were discarded, since shorter reads contain less information and are more prone to incorrect OTU assignment. The quality assurance/quality control (QA/QC) trimmed forward reads were loaded into the software package QIIME (Quantitative Insights Into Microbial Ecology) for alignment and *de novo* clustering using the algorithm UCLUST. Chimeras and singletons were removed from the dataset. Samples yielding less than 50 sequencing reads (10 of 133 samples) were excluded from the analysis in order to minimize stochasticity. The *Microcystis* OTUs produced by this analysis were aligned with reference *cpcBA* sequences previously acquired for the Klamath River and assigned to one of four categories corresponding to operational taxonomic units (OTUs) 1-4 (Bozarth et al., 2010). A fifth OTU was discovered in Copco Reservoir in September 2013; this appears to be a subtype of OTU2. Note that each OTU can be thought of as comprising a unique strain or genotype.

In addition to amplicon sequencing, real-time quantitative PCR employing TaqMan chemistry was used to estimate total (*cpcB*) and toxigenic (*mcyE*) *Microcystis* cell equivalents from each sample as previously described (Otten et al., 2015). Reactions were carried out in 25 µL volumes consisting of: 12.5 µL 2X Maxima Hot Start Probe Master Mix (Thermo Scientific), 1 µL forward primer (10 mM), 1 µL reverse primer (10 mM), 1 µL probe (5 mM), 5 µL 1:10 diluted DNA template and 4.5 µL dH_2O. All assays were run in duplicate on an Applied Biosystems ABI 7500 Fast thermal cycler; any samples with cycle threshold differences greater than one were rerun. The thermal cycling conditions were the same for both assays and consisted of an initial denaturation step at 95°C for 2 min, followed by 45 cycles of 94°C for 15 sec and 56°C for 45 sec. Gene copy numbers for environmental samples were inferred by standard curve interpolation. A single synthetic gene construct (gBlock; IDTDNA) containing both the *cpcB* and *mcyE* gene sequences was used to create serially diluted standard curves spanning from 10^8-100 gene copies. The reaction efficiency for the *cpcB* assays ranged from 95.9-101.3% and 92.0-102.8% for the *mcyE* assays. All gene copy estimates were multiplied by 2.5X in order to account for extraction efficiency as previously described (Otten et al., 2015).

Shotgun metagenomics. We used the *cpcBA* deep sequencing results to determine when *Microcystis* strain turnover events had occurred. Samples collected from both reservoirs corresponding to periods immediately before, during or after a change were selected (n=23) for shotgun metagenomic sequencing. Shotgun metagenomics is the process of fragmenting total DNA from a sample into short reads that can be accurately and massively sequenced. The short reads were bioinformatically reassembled into longer stretches of contiguous DNA sequences called contigs. Although components of the entire microbial community are represented in these sequencing data sets, here we only focus on the cyanobacteria. In the context of this study's objectives, the shotgun assemblies were used to validate the presence of 4-5 strains of *Microcystis* as predicted from the *cpcBA* sequencing; 2) to determine which strains possess microcystin biosynthesis genes; and if possible, 3) to determine the flexible genome content that is expected to govern the unique physiological capabilities of each strain, interpreted in the context of the prevailing environmental conditions in which each strain dominated. For example, if one of the strains is only dominant when nitrogen concentrations are below average, we may investigate whether or not this strain possesses any nitrogen uptake pathways (e.g., urea transporter genes) that the other strains do not.

DNA libraries were created from DNA extracts using a Nextera XT prep kit and sequenced (151 bp, paired-end reads) on an Illumina HiSeq 3000 instrument at the OSU Center for Genome Research and Biocomputing. Sequencing reads were quality screened (phred score > 25, minimum length > 100 nts) and trimmed using Trimmomatic. All reads passing QA/QC from each sample were individually assembled into unscaffolded contigs using IDBA-UD. This assembler was used because it is designed to handle highly uneven coverage depths characteristic of metagenomic datasets. The assemblies from each discrete sample were concatenated into a single, master assembly using the program minimus2 with a pairwise read cut-off setting of 98%. The sequencing reads from each sample were then mapped to the meta-assembly using Burrows Wheeler Aligner. These mapping files were then used to estimate the read coverage depth for each contig using the bioinformatic program suite samtools. PhylopythiaS+ was used to taxonomically classify contigs and single-copy universally conserved (core) genes were

identified within the contigs using a hidden markov model (HMMER3). The annotated contigs and their relative read coverage depths were visualized in R using the mmgenome package. Lastly, gene coding sequences were predicted from the assembled contigs using the program Prodigal and the putative function of each open reading frame was assigned using the bioinformatics program PROKKA.

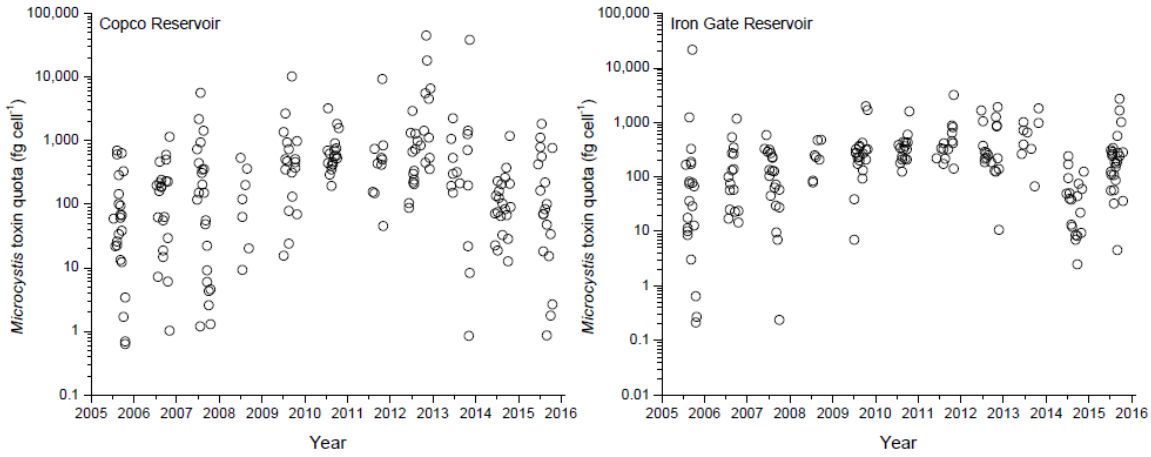
Statistics. One way analysis of variance (ANOVA) was conducted on each water quality variable collected from June 1 to September 30 for each year in order to determine if there were significant differences between each season's means. Additionally, in order to test the null hypothesis that no groups are different, the data must follow a normal distribution. Anderson-Darling and Shapiro-Wilk normality tests were used to assess the normality of each variable; any with probability values > 0.05 were log transformed. Variables with significant ANOVA results were further assessed by posterior testing (Tukey's Honest Significant Difference test). In addition to seasonal testing, those variables exhibiting significant differences were further assessed using change point analysis (CPA). CPA is used to determine more precisely when a change has occurred. In general, we anticipate that some period of environmental stability is needed in order for one strain to supplant another; this is especially true when one considers that the average doubling time for *Microcystis* is often less than one doubling every two days (Wilson et al., 2006). The R package 'changepoint' was used with the segmentation neighborhood method, normal distribution test statistics, asymptotic penalty scoring (i.e., assumed type I error of 0.05) and the maximum number of change points (Q) set equal to the number of years analyzed per variable.

RESULTS & DISCUSSION

Microcystis toxicity. One of the core objectives of this study was to determine precisely when changes in *Microcystis* population structure occur and how these changes influence microcystin concentration. In order to assess if there are indeed periods of higher and lower toxicity, we first present microcystin cell quotas from Copco and Iron Gate Reservoir from 2005-2015 (Fig. 1), derived from KHSA monitoring.

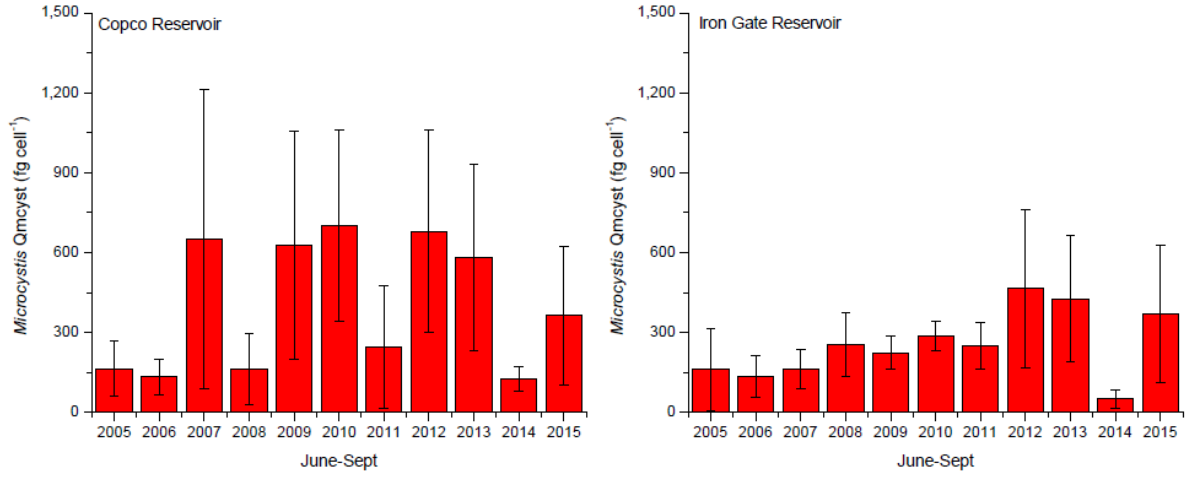
The average cellular concentrations observed over the study period were not markedly different between the reservoirs, although extremes in Qmcyst were at times 1-2 orders of magnitude greater in Copco than in Iron Gate Reservoir. These outliers tended to occur late in the fall and are most likely caused by samples being collected soon after a cell lysis event, when there would be disproportionately higher concentrations of extracellular microcystin relative to intact cells. While these outliers were typical of late-season data, there were a few data points from the peak summer months that also exhibited unusually high Qmcyst, and these too could be due to cell lysis events which would reduce cell numbers but not total toxin. An important observation to draw from this figure is that Qmcyst values showed tighter ranges and were higher during the period 2009 through the first half of summer 2013. Then from August 2013 through most of 2014, Qmcyst concentrations were attenuated relative to the preceding several years; this pattern was most pronounced in Copco Reservoir.

Figure 1. Estimates of microcystin cell quotas (Qmcyst) from public health and environmental monitoring cell counting and microcystin data.



In order to statistically compare these data, the mean and 95% confidence intervals from the summer (Jun-Sept) of each year are presented and the seasonal means were compared by one way ANOVA (Fig. 2). By restricting the analysis to only the summer months, most of the outliers were removed from the dataset. However, there were still two atypical values that were treated as outliers and removed from the microcystin dataset; these included: 10,052 fg/cell on 9/14/09 in Copco Reservoir and 21,365 fg/cell on 9/20/05 from Iron Gate Reservoir. In both cases, there were other observations (retained in the dataset) from different sites within the reservoirs that were collected on or near the same date which should help to remove observation biases that may stem from the removal of these over-leveraged data points.

Figure 2. Mean and 95% confidence intervals for *Microcystis* microcystin cell quotas (fg/cell) during summer.



Note that this approach also produced results consistent with observations of *Microcystis* Qmcyst from other studies. During the summer months, the grand mean for Qmcyst in Copco Reservoir was 414 fg/cell and for Iron Gate Reservoir it was 242 fg/cell; these values are in-line with those reported by Lyck (2004), which depending on cell growth phase, were found to range from ~110 to 400 fg/cell. The ANOVA probability (p-value) for inter-annual differences in Qmcyst for Copco Reservoir was 0.06, making it just outside of the statistical significance cutoff.

However, in Iron Gate the ANOVA indicated there was a significant difference (p=0.03), although the probability value from Tukey's HSD was 0.065 between 2012 and 2014. As such, we can conclude that there was a weakly significant difference between 2012, when Qmcyst was highest, and 2014 when Qmcyst was lowest. Overall, these results suggest that cell toxicity has varied over the time-series of inquiry, although due to the large standard deviations in the data set (likely owing to variability in extracellular microcystin content or counting error) statistical significance was generally not realized. This analysis could be refined if toxin measurements were conducted for the total and extracellular fractions, from which the particulate (i.e., cell-associated) fraction could be more accurately determined (by difference).

In addition to microcystin cell quotas, toxigenicity was assessed over the study period by comparing estimates of total *Microcystis* (*cpcBA*-possessing) with those of toxigenic *Microcystis* (*mcyE*-possessing). For this analysis, only environmental samples collected from in front of the dam log booms were included (CR01 or KR19874 and IR01 or KR19019). The results indicated that there were inter-annual differences in the proportion of toxigenic cells, with *mcyE* gene concentrations often exceeding those of the *cpcBA* genes (Fig. 3). It is not clear whether this is an analytical artifact or whether there is a subset of the population that contains more copies of the *mcyE* gene than *cpcBA*. In our previous study (Otten et al., 2015), we verified the complementarity of the QPCR primers and probes and found both sets to be perfect matches for the Klamath *Microcystis* population, and both assays showed similar reaction efficiencies (see Materials and Methods). In any case, the QPCR results corroborate key observations from the microcystin cell quota data, specifically that there are periods of higher and lower toxigenicity and that the second half of summer 2013 and all of 2014 were characterized by a strong reduction in toxigenicity relative to several years prior. Interestingly, Iron Gate Reservoir in 2015 appeared to contain a lower proportion of *mcy*-possessing cells than average, although this is generally inconsistent with the cell quota estimates for 2015 presented in Fig. 2. The difference is presumably derived from the fact that Fig. 2 contains both environmental and public health samples from various sites, whereas Fig. 3 only shows data from behind the dams. In Iron Gate Reservoir (2015), microcystin concentrations did in fact remain low at the log boom site, never exceeding 1 µg/L during the summer months. These toxin data corroborate the QPCR results, which showed low relative toxigenicity over this period.

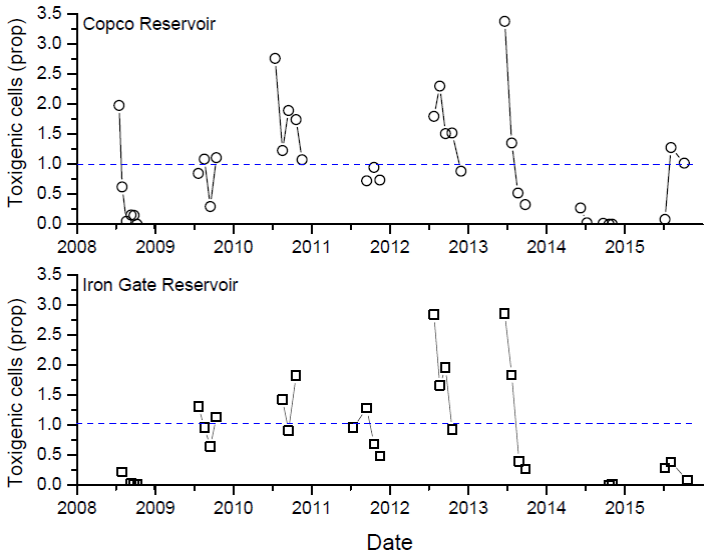
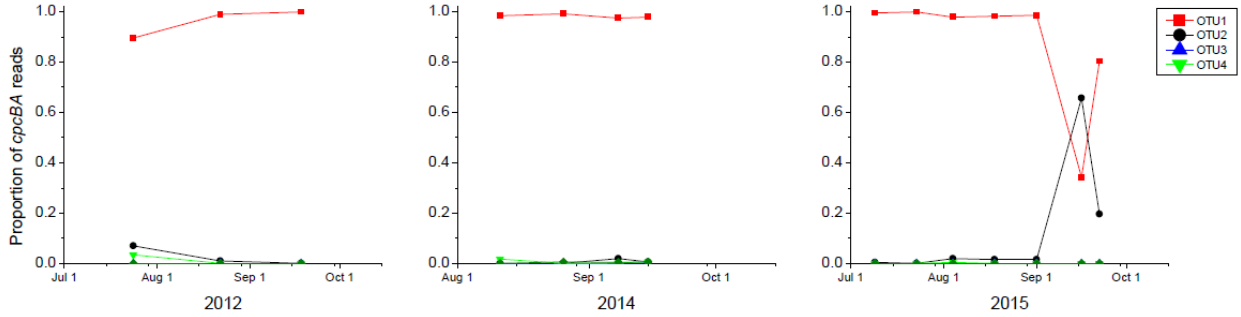


Figure 3. Changes over time in *Microcystis* toxigenicity. QPCR estimates of *Microcystis mcyE* gene number relative to *Microcystis cpcB* gene number from Copco (KR19874, top panel) and Iron Gate (KR19019, bottom panel) dam routine monitoring sites

Amplicon Sequencing results. A total of 299,096 *cpcBA* sequences corresponding to *Microcystis* passed QA/QC and were analyzed from 133 discrete Klamath samples. Only samples with 50 or more sequencing reads were further analyzed; 123 samples met this criterion. The average number of *Microcystis* reads per sample was 2,249 and the median was 1,106 reads. Beginning near the headwaters of the system, samples were analyzed from Upper Klamath Lake from 2012, 2014 and 2015 (Fig. 4). These results revealed that the dominant strain over this period was OTU1, with the exception of a brief period in September, 2015 when OTU2 was dominant. For 2012, we attempted to analyze more than the three samples shown, but due to low *Microcystis* abundance, most of these samples failed to amplify. On the contrary, all samples from 2014 and 2015 were successfully amplified. *Microcystis* blooms appear to be increasing in magnitude and persistence in recent years in UKL, possibly due to low snow-pack which leads to water column temperatures warming earlier than in previous years.

Figure 4. Upper Klamath Lake *Microcystis* population structure assessed by *cpcBA* deep sequencing



Below UKL, neither Link nor Keno Reservoir samples (all collected in 2012) generated enough sequencing reads for analysis. Similar to UKL, *Microcystis* seems to have been relatively rare in these reservoirs in 2012. Congruent with this pattern, the lone sample analyzed from JC Boyle Reservoir from 2012 failed to produce enough sequencing reads for analysis. On the contrary, all three samples collected from JC Boyle in 2015 generated ample sequences for analysis. All three of these samples were comprised almost entirely of *Microcystis* OTU1 (greater than 99% of the reads).

The genetic record for Copco Reservoir extends back nine years, five of which were shown to be dominated by OTU1 from start to finish, with three additional years exhibiting lengthy periods of OTU1 dominance (Fig. 5). A very different pattern was observed in 2014, with low OTU1 representation and dominance by OTUs 3 and 4. Our previous investigations have linked OTU1 (formerly known as CPC-A) to microcystin toxigenicity (Bozarth et al., 2010). Its low prevalence throughout 2014 corresponded with low cellular microcystin quotas (Fig. 2), consistent with OTU1 as a toxin-producer, while OTU's 3 and 4 may not be. However, there was still some microcystins observed in 2014, albeit lower than average, and these data cannot definitively determine if another microcystin producer may have been present. The possibility for additional toxin-producing genotypes was investigated using shotgun metagenomics and is described later in the report.

The departure from OTU1 dominance began in the second half of 2013, and marked the first time in over four years that a strain other than OTU1 was dominant in either reservoir. Remarkably, the first non-OTU1 strain to be dominant in 2013 was a strain not prevalent in the previous years (Fig. 5). A temporal comparison of UKL and Copco Reservoir strain composition indicates that OTU1 was dominant at both sites in 2012, but the diversity seen in 2014 in Copco Reservoir was never detected in UKL. Similarly, the period of OTU2 dominance in UKL was not reflected in Copco Reservoir in 2015. These observations are consistent with primarily internal processes determining the *Microcystis* strain compositions in both lakes.

The genetic record for Iron Gate Reservoir extends back to 2008. The *Microcystis* population patterns in Iron Gate and Copco Reservoirs were generally congruent (Figs. 5 & 6); this was most apparent in the multi-year dominance of OTU1 in both reservoirs from 2009 to mid-2013. There were also periods when the *Microcystis* population structure was quite distinct from that in Copco Reservoir. For example, in 2008, the population in Iron Gate was dominated by OTU3 throughout the summer and fall, whereas Copco Reservoir was dominated by OTU1 during two months of the summer. Other differences between the two reservoirs included the presence of four different OTUs in Iron Gate Reservoir in 2014 when only two major strains were observed in Copco Reservoir, and the presence of three strains in Iron Gate in 2015 when only one significant strain was detected in Copco Reservoir. Due to their close proximity, it is highly likely that some percentage of *Microcystis* cells from Copco Reservoir make their way into Iron Gate Reservoir. However, these rather significant differences in population structure also indicate that population structure within each reservoir is most likely driven by its own internal processes and overwintering *Microcystis* seed stocks. Iron Gate and Copco Reservoirs are fairly distinct in their physicochemical properties. Most notably, Copco Reservoir is shallower than Iron Gate, which allows thermal stratification to set-up sooner in the year, and it tends to have higher nitrogen and phosphorus concentrations due to upstream sources that are largely stripped out of the water column before water reaches Iron Gate Reservoir. Considering these significant differences between the reservoirs, it is interesting that both reservoirs experienced the same period of OTU1 dominance from 2009-2013. This suggests that there may have been a key environmental factor that remained stable in both reservoirs over this period, and that this period of stability was broken in late 2013 and throughout 2014.

Figure 5. Copco Reservoir *Microcystis* population structure assessed by *cpcBA* deep sequencing

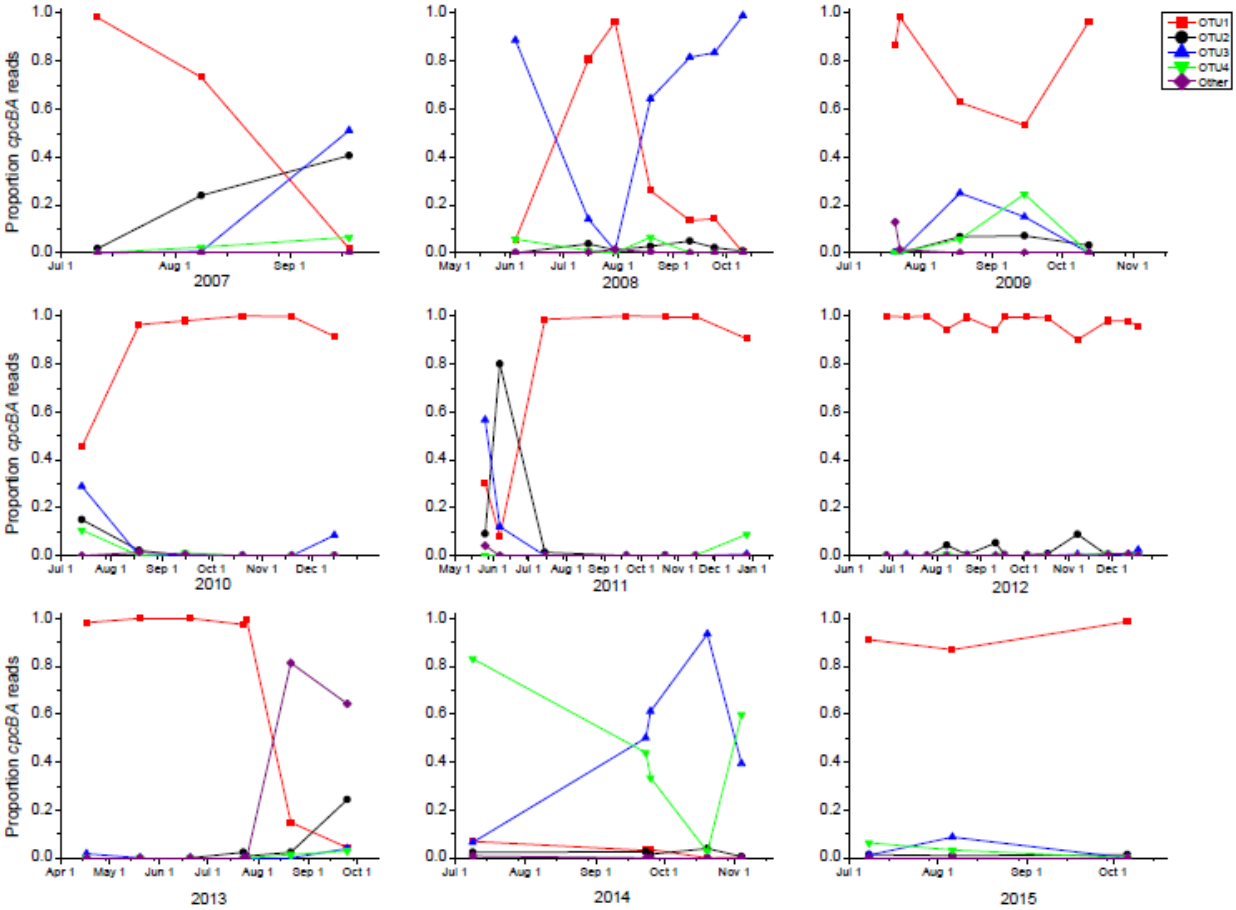
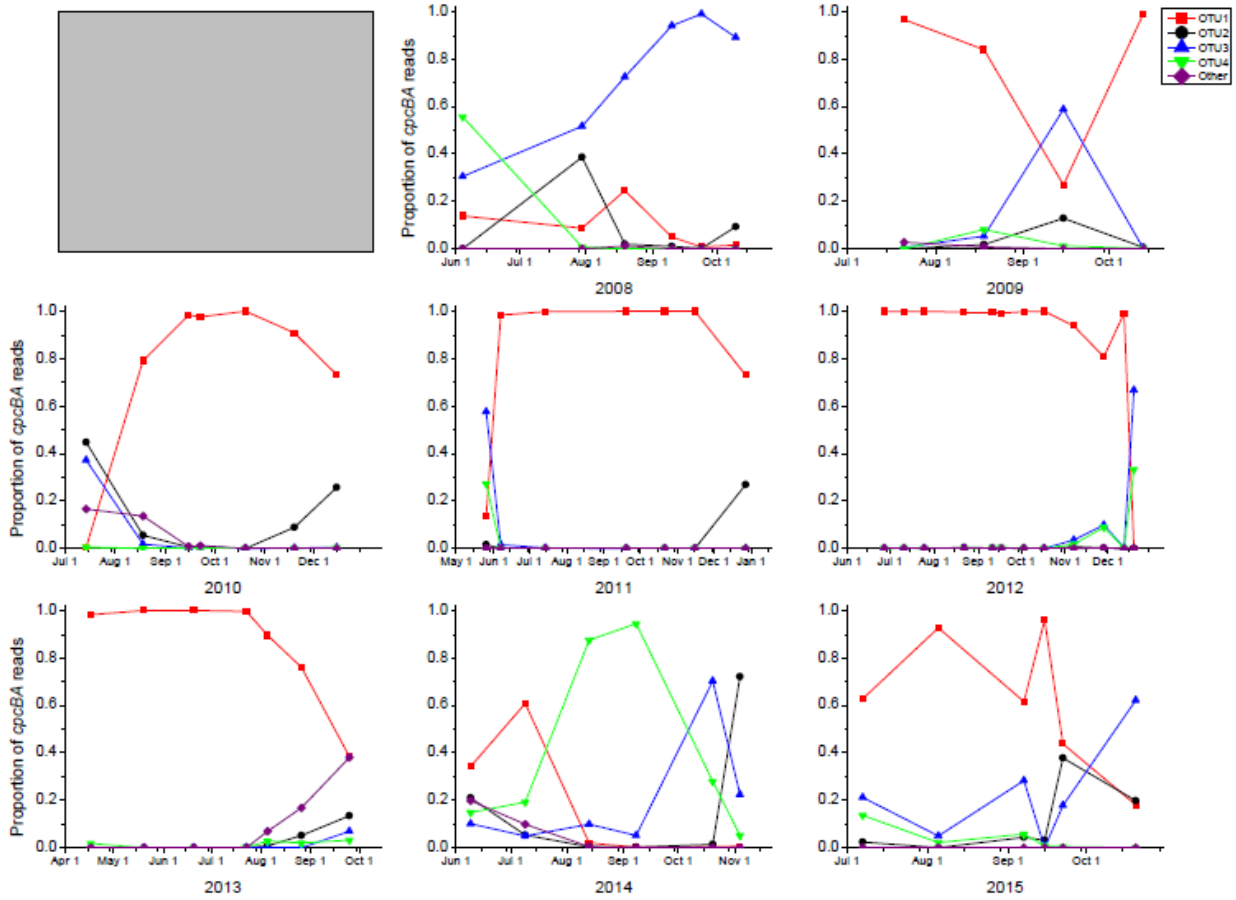


Figure 6. Iron Gate *Microcystis* population structure assessed by *cpcBA* deep sequencing



Shotgun sequencing results. A total of 23 metagenomes were sequenced, each generating over 20 million paired reads. The reads from each data set were assembled into ~4.7 million contigs; these were concatenated into a meta-assembly and like contigs (minimum 98% identical) were condensed into 1.37 million contigs spanning 1.64 billion nucleotides of genetic information. Of this total, cyanobacteria were estimated to represent ~11,000 contigs spanning over 47 million nucleotides. On the basis of universally conserved gene content, we estimate that approximately eight cyanobacterial genomes were represented and comprising the genera: *Microcystis*, *Aphanizomenon*, *Pseudanabaena* and *Synechococcus*. Of this total, *Microcystis* represented over one-third of the cyanobacterial sequences. Contigs assigned as *Microcystis* were extracted from various periods when each OTU was dominant.

Figure 7 illustrates this process on sample CR200 (Copco Reservoir - 10/20/14), the red circles are all *Microcystis* contigs and their relative abundance is indicated by the axes. In that sample, OTU3 represented the majority of the population based on *cpcBA* data, and therefore, the most abundant contigs (those around 100X coverage) should all correspond to OTU3. These contigs can be extracted and their coverage depth can be assessed across all 23 metagenome samples to generate a pattern of OTU abundance changes over time that can be compared

to the *cpcBA* deep sequencing data in Figures 5 and 6. One can also identify the toxin genes in each metagenome and estimate how their abundance changes over time; in Figure 7, the microcystin toxin genes are highlighted as darker red circles (located below 1X coverage depth). Thus, this analysis indicates that OTU3 in October 2014 was the most dominant strain as predicted from the *cpcBA* sequencing, and it also indicates that this strain lacks the *mcy* genes.

This process was completed for all 23 samples by extracting a subset of the most abundant contigs and assigning them to their respective OTU designation. The coverage depth of these extracted contigs was averaged across each extraction and compared with that of microcystin synthetase genes (*n*=3 contigs) recovered from the assembly (Fig. 8).

Figure 7. Example graph depicting how *Microcystis* contigs are extracted and how their relative abundance (contig coverage depth) is used to discern OTU strain abundance. Microcystin genes are highlighted in dark red circles. Both axes indicate coverage depth.

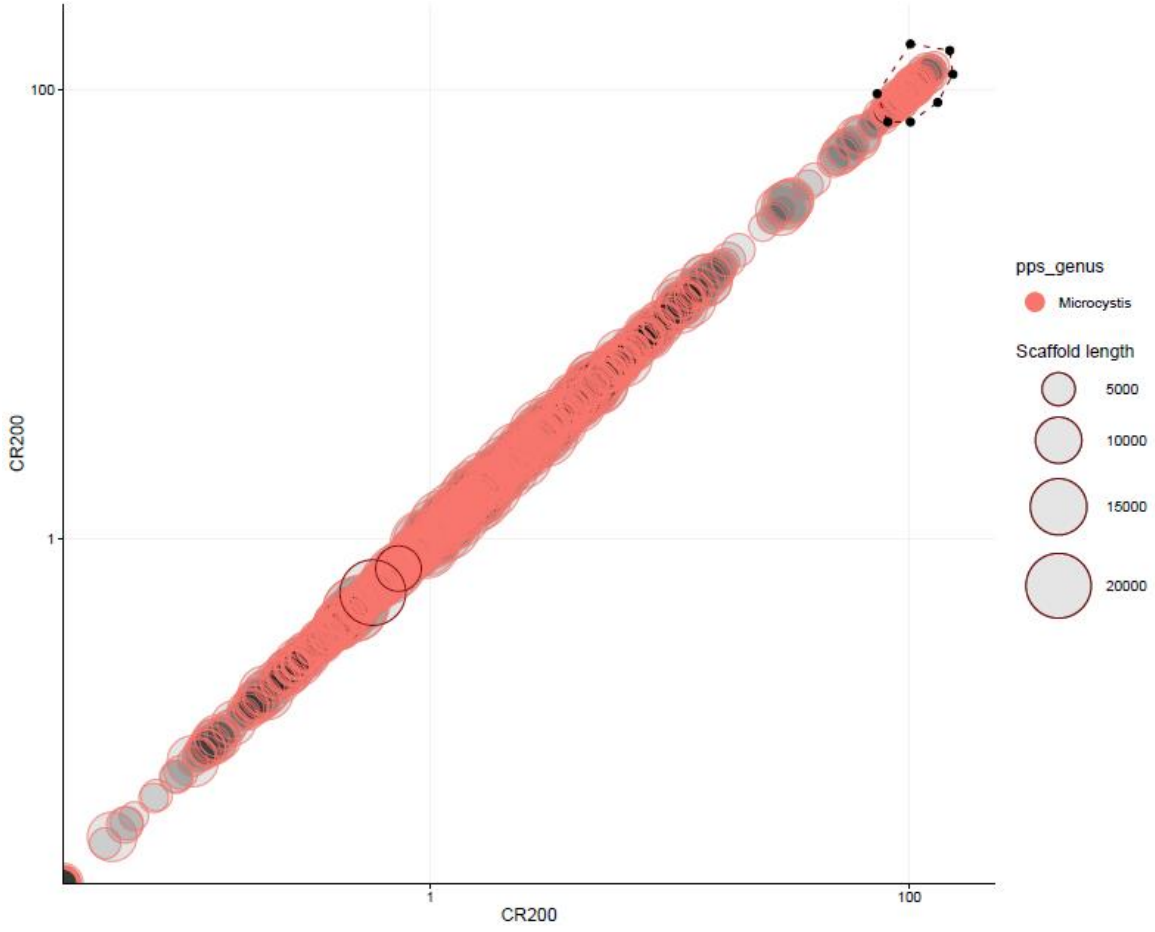
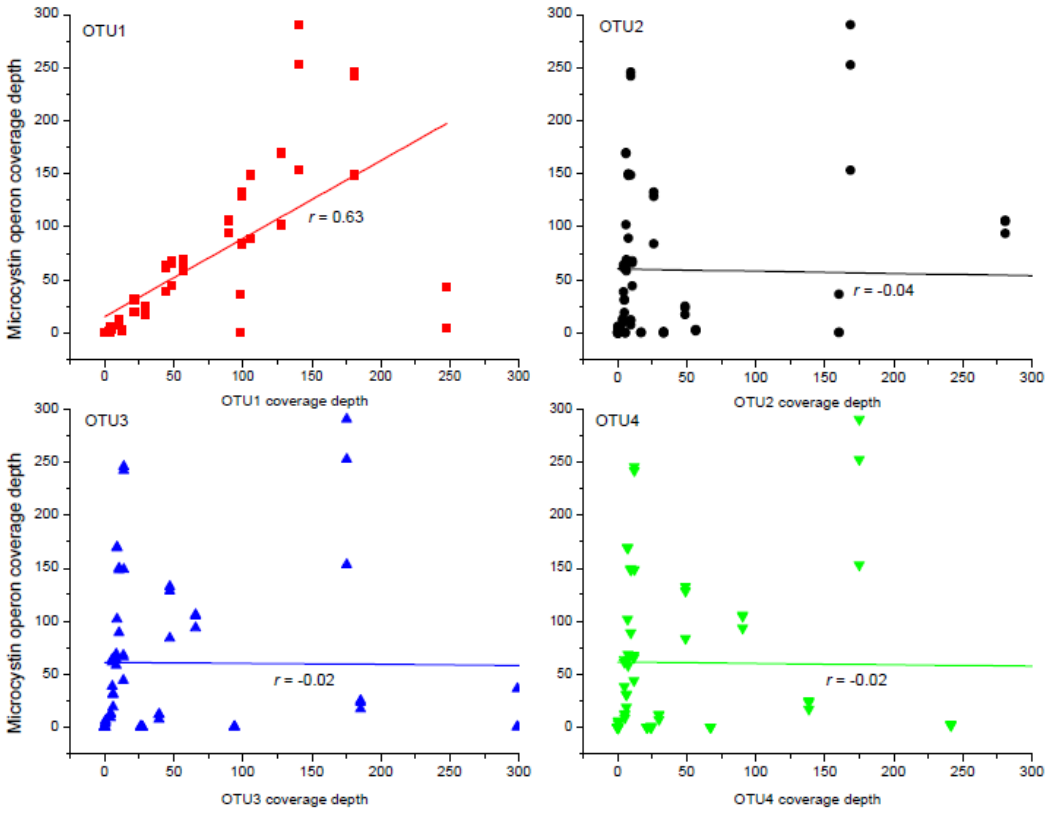


Figure 8. Comparison of shotgun sequencing read coverage depths for contigs extracted from individual *Microcystis* strains (OTUs) and contigs containing the microcystin gene operon across (*n*=23 samples).



Only OTU1 showed a positive correlation with microcystin gene abundance. This analysis also suggests that although rare, there is likely a sub-strain of OTU1 that is nontoxic based on the few data points positioned far below the trend line (i.e., high OTU1 abundance but low *mcy* gene abundance). These data points both correspond to Copco Reservoir (Aug-2008 and Sept-2014), when *Microcystis* was either rare or in decline. This conclusion is consistent with data presented in our previous studies (Bozarth et al., 2010; Dreher et. al. 2013 report).

Identification of physicochemical factors that may influence Microcystis strain composition.

In order to test the hypothesis that changes in environmental conditions precipitate changes in *Microcystis* population structure, a suite of multivariate statistics were employed. The first step was to test if any variables exhibited differences between years. One-way ANOVA was used to test for differences in seasonal means for all physicochemical variables from each individual reservoir. In Copco Reservoir, alkalinity, dissolved organic carbon (DOC), flow rate, NO_3, ORP and PO_4 were all flagged as containing at least one pair of values that significantly differed (Table 1). In Iron Gate Reservoir, alkalinity, DOC, flow rate, ORP and PO_4 were flagged as containing values that differed significantly between years (Table 1). Because ANOVA only indicates if there are differences between one or more data pairs, these variables were analyzed using Tukey's HSD test to determine which pairs were significantly different.

For variables to be further considered as potential drivers of *Microcystis* population turnover, any significant differences occurring over the time-series must also have been temporally linked with periods when strain composition was observed to vary based on the *cpcBA* sequencing results for each reservoir (Figures 5 and 6). For example, the posthoc tests indicated that oxidation-reduction potential (ORP) levels in Copco Reservoir in 2007 and 2009 were significantly different than in 2011 and 2015; however, OTU1 was dominant during each of these four years and therefore differences in ORP were unlikely to explain *Microcystis* succession patterns. Based on this framework, flow, alkalinity, DOC, NO_3, and PO_4 were advanced for further analysis from Copco Reservoir, and flow, alkalinity, DOC and PO_4 were advanced for further consideration from Iron Gate Reservoir.

In addition to the water quality analyses, we also ran ANOVA on data collected from an EPA air quality monitoring station in Siskiyou County (60932001) that collects data on particulate matter sized < 2.5 µm ($PM_{2.5}$) at a frequency of every few days. $PM_{2.5}$ is volatilized during the burning of organic carbon and it serves as a proxy for wildfires in the vicinity of the reservoirs. Wildfires are considered important because they may augment nitrogen concentrations within the basin through atmospheric deposition or runoff of charred land cover, and they also influence light intensity and photosynthetically active radiation (400-700 nm light). The ANOVA results indicated that $PM_{2.5}$ was significantly different in 2014 relative to 2005-2012 and this was further investigated by change point analysis.

Table 1. Results from one-way ANOVA and posthoc tests indicating significant differences in seasonal means between years

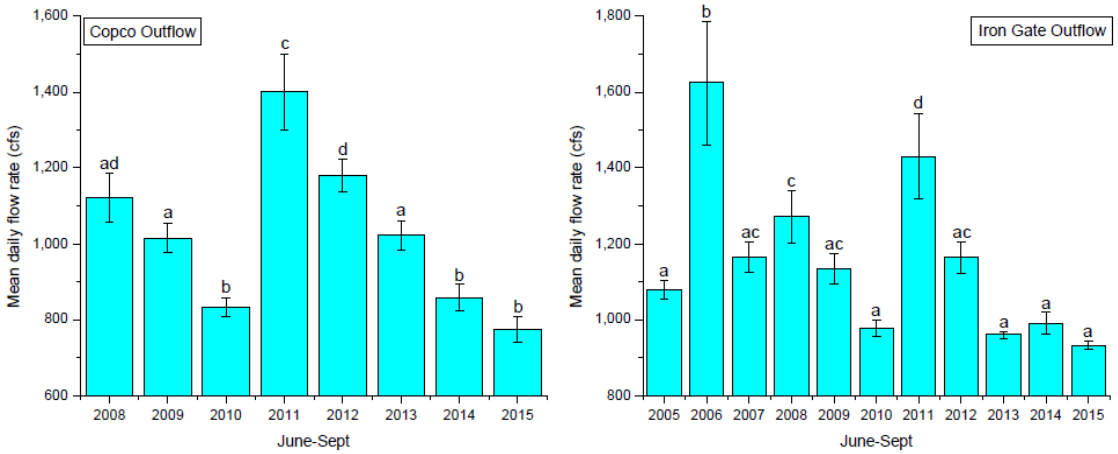
Copco Reservoir (June - September)					
Variable	**Date Range**	**Mean**	**±95% CI**	**ANOVA**	**Tukey's HSD (p < 0.05)**
ALK	2007-2015	64.2	2.1	0.031	2011 ≠ 2013
CHLA	2007-2015	58.7	70.3	0.503	
COND	2007-2015	162.1	9.3	0.223	
DOC	2007-2015	5.67	0.38	<0.001	2014 & 2015 ≠ 2008, 2009, 2011-2013
DOPER	2007-2015	137.4	12.3	0.445	
FLOW	2008-2015	1025.8	22.2	<0.001	Numerous, see Figure 7
NH_3	2007-2015	0.018	0.008	0.469	
NO_3	2007-2015	0.211	0.071	<0.001	2008 ≠ 2007, 2010-2013, 2015
ORP	2007-2015	160.3	13.8	<0.001	2007 & 2009 ≠ 2011, 2015
pH	2007-2015	8.73	0.16	0.285	
PHEO	2007-2015	10.59	12.59	0.491	
PO_4	2007-2015	0.113	0.015	<0.001	2008 ≠ 2010, 2011, 2014, 2015
Qmcyst	2005-2015	414.0	112.5	0.068	
TEMP	2007-2015	21.1	0.86	0.397	
TN	2007-2015	1.36	0.68	0.377	
TP	2007-2015	0.211	0.074	0.277	
TSS	2007-2015	11.1	8.4	0.55	
VSS	2007-2015	12.2	13.2	0.484	
Iron Gate Reservoir (June - September)					
Variable	**Date Range**	**Mean**	**±95% CI**	**ANOVA**	**Tukey's HSD (p < 0.05)**
ALK	2007-2015	66.4	1.9	0.002	2011 ≠ 2008, 2010, 2013; 2013 ≠ 2015
CHLA	2007-2015	18.0	8.0	0.124	
COND	2007-2015	167.2	7.8	0.105	
DOC	2007-2015	5.34	0.41	0.012	2008 ≠ 2015
DOPER	2007-2015	138.4	10.3	0.69	
FLOW	2005-2015	1157.5	23.2	<0.001	Numerous, see Figure 7
NH_3	2007-2015	0.011	0.007	0.321	
NO_3	2007-2015	0.068	0.048	0.029	None significant
ORP	2007-2015	152.4	16.7	<0.001	2012 ≠ 2007, 2008, 2010
pH	2007-2015	9.28	0.58	0.492	
PHEO	2007-2015	3.7	2.7	0.099	
PO_4	2007-2015	0.093	0.014	<0.001	2008 ≠ 2009-2012, 2015; 2011 ≠ 2009, 2014
Qmcyst	2005-2015	242.4	54.4	0.034	None significant
TEMP	2007-2015	22.1	0.8	0.666	
TN	2007-2015	0.82	0.18	0.461	
TP	2007-2015	0.142	0.021	0.134	
TSS	2007-2015	5.8	2.3	0.35	
VSS	2007-2015	5.1	2.0	0.702	
Siskiyou County, CA (June - September; EPA monitoring station 60932001)					
Variable	**Date Range**			**ANOVA**	**Tukey's HSD (p < 0.05)**
$PM_{2.5}$	2005-2015			<0.001	2014 ≠ 2005-2012

Where: ALK=alkalinity, CHLA=total chlorophyll *a* , COND=conductivity, DOC=dissolved organic carbon, DOPER=dissolved oxygen (%), FLOW=flow rate, NH3=ammonia, NO3=nitrate, ORP=oxidation-reduction potential, pH= power of hydrogem, PHEO=pheophytin, PO4=orthophosphate, Qmcyst=microcystin cell quota, TEMP=temperature, TN=total nitrogen, TP=total phosphorus, TSS=total suspended solids, VSS=volatile suspended solids

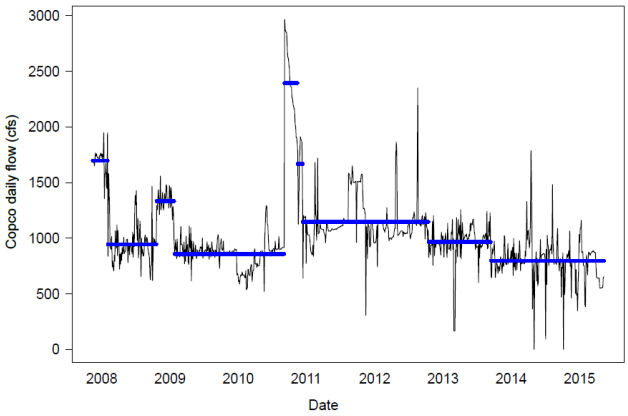
Change point analysis of retained variables

1. **Flow rates** were identified as being highly variable and statistically different between years. Figure 9 shows mean summer flows out of Copco Reservoir (2008-2015) and Iron Gate Reservoir (2005-2015). Statistically significant differences between years are indicated by the bar labels.

Figure 9. Mean seasonal flow rates and associated 95% confidence intervals. Bars from statistically similar years are labeled accordingly



The data clearly showed that 2006 and 2011 were unusually wet years, 2008 and 2012 had comparable and moderately high flows, and that 2010, 2013-2015 were relatively low flow years and correspond to drought periods. The *cpcBA* sequencing data indicate that OTU1 was completely dominant from 2009-2012, a period spanning both drought and high flow years. As such, seasonal flow rate would not seem to control *Microcystis* strain composition.



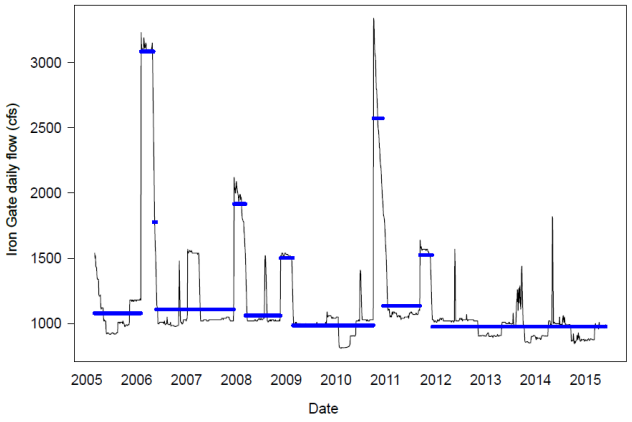
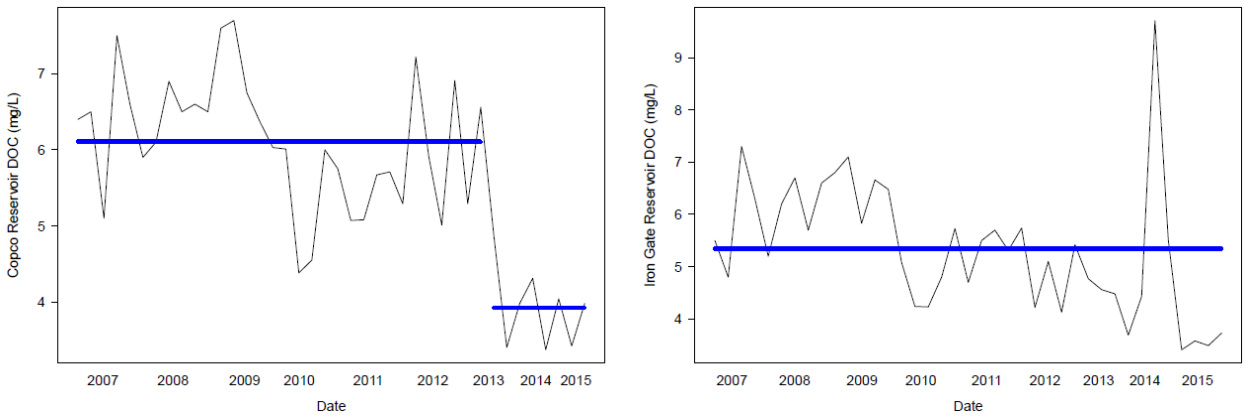


Figure 10. Change point graph denoting periods of significantly different mean daily flow rates (blue lines) for Copco powerhouse (upper panel) and Iron Gate powerhouse (lower panel)

The change point analysis (Fig. 10) confirmed the results described for Figure 9. However, by revealing higher temporal resolution, the CPA does show that during the period when OTU1 was replaced (2013-2014), there was no significant period of different flow rates in Iron Gate Reservoir relative to 2012 and 2015 when OTU1 was dominant. Similarly, 2014 and 2015 flows were comparable in Copco Reservoir, even though the *Microcystis* populations were distinct. Likewise, OTU3 dominated in Iron Gate Reservoir under both relatively high (2008) and low flow (2014) conditions. As such, we conclude that variation in flow rate is unlikely to have caused *Microcystis* strain composition changes in the reservoirs.

2. **Dissolved organic carbon** concentrations in Copco Reservoir were significantly lower in 2014 and 2015 than all prior years aside from 2007 and 2010. Because *Microcystis* strain composition in Copco Reservoir in 2014 was starkly different than all other years, this finding was further investigated. Additionally, DOC concentrations in Iron Gate were significantly different between 2008, when OTU3 dominated and 2015, when OTU1 dominated. Low DOC could mean less oxidative stress at the surface, because the interaction of UV light on dissolved organic matter leads to the production of reactive oxygen species (Cooper and Zika, 1983). Previous studies have speculated that intracellular microcystin molecules protect proteins involved in photosynthesis and carbon fixation from the protein damaging effects of free oxygen radicals (Zilliges et al., 2011) and that this may in turn provide a selective advantage for toxigenic strains to dominate during periods of high light intensity or oxidative stress (Paerl and Otten, 2013). Change point analysis indicated that significant changes in DOC concentrations occurred in Copco Reservoir around 8/21/13, although none were identified in Iron Gate Reservoir (Fig. 11). Since DOC concentrations were different between the reservoirs in 2014 when they both experienced strain dominance by OTUs 3-4, it is unlikely that DOC was a principal selection pressure.

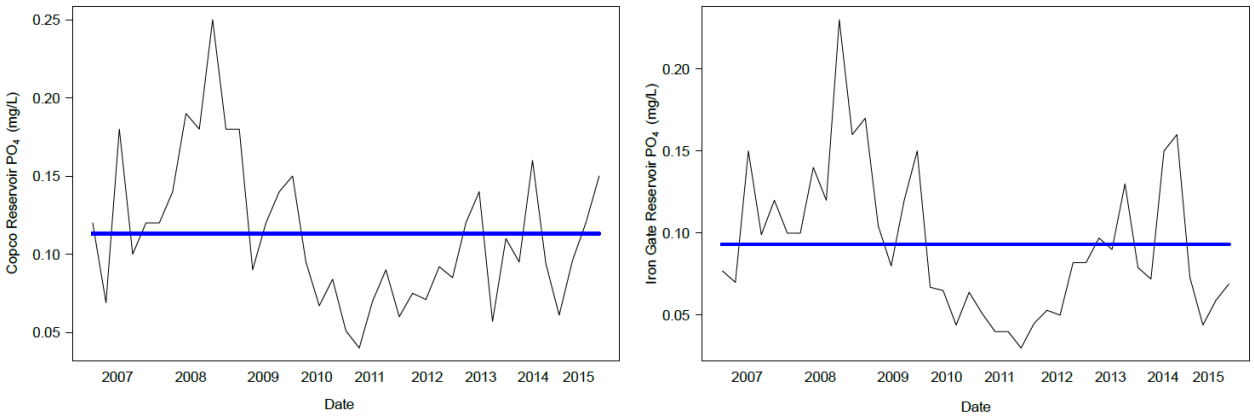
Figure 11. Change point graph denoting periods of significantly different dissolved organic carbon concentration for Copco reservoir (left panel) and Iron Gate reservoir (right panel)



3. **Orthophosphate (PO_4)** concentrations were found to be statistically different between several years in both reservoirs, and in particular, 2008 was distinct from most years. OTU3 was dominant throughout all 2008 in Iron Gate Reservoir and for a period in the early summer in Copco Reservoir. However, the CPA did not identify any

change points in either reservoir's time-series PO_4 data. As such, there is no evidence that variable orthophosphate concentrations led to the selection of one strain over another. This is likely because all strains possess alkaline phosphatases and phosphate transport systems. These results would suggest that each strain has a similar affinity for dissolved inorganic phosphate or that this nutrient is not a limiting factor for their growth. Moisander et al., (2009) conducted growth bioassay studies in Copco and Iron Gate reservoirs and found evidence that *Microcystis* growth is constrained by nitrogen-limitation during the summer months, not PO_4.

Figure 12. Change point graph denoting periods of significantly different orthophosphate concentrations in Copco reservoir (left panel) and Iron Gate reservoir (right panel)



4. **Nitrate (NO_3)** concentrations in Copco Reservoir were significantly higher in 2008, 2009 and 2014, with 2008 especially standing out as the most unique across the time series. Of particular interest, OTU3 was abundant or dominant in 2008 and 2014, respectively. However, as with the assessment of PO_4, the change point analysis did not identify any periods of significant change (with 95% confidence) across the time-series (Fig. 13).

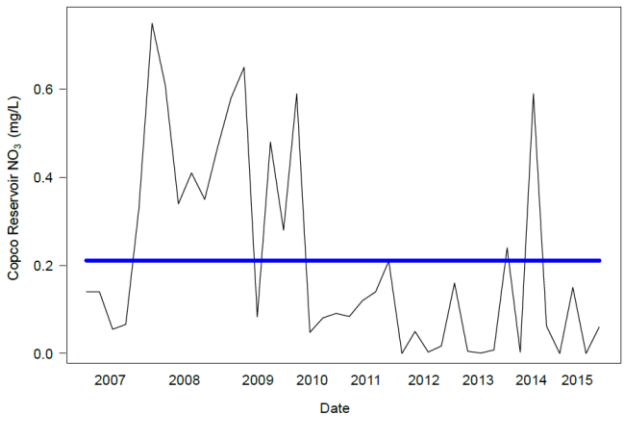
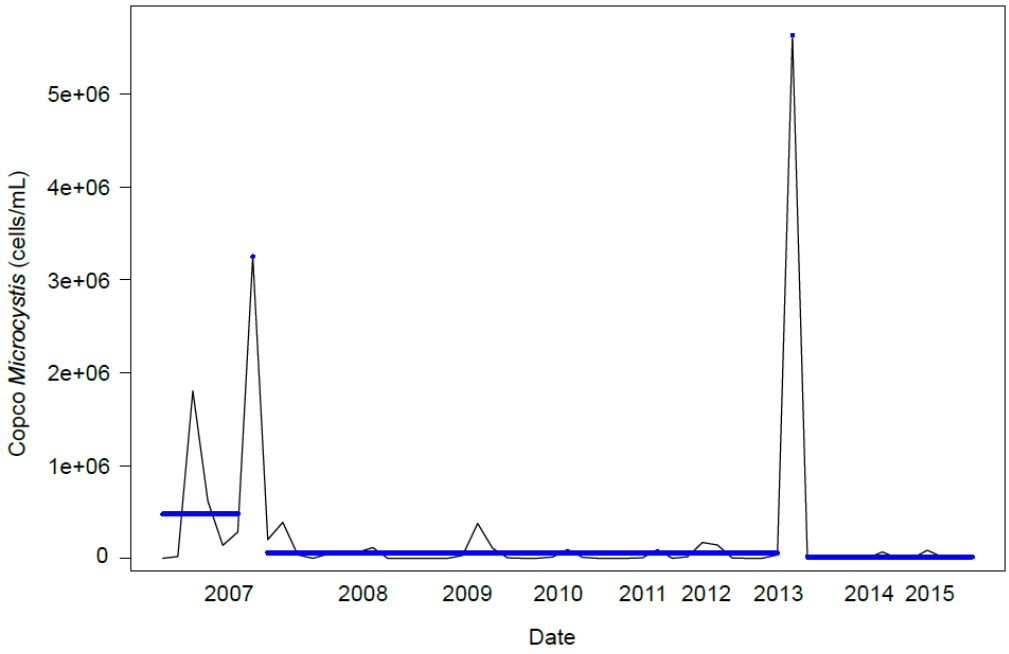


Figure 13. Change point graph denoting periods of significantly different nitrate concentrations in Copco Reservoir

Even though the CPA did not show significant changes, because 2008 and 2014 still had relatively high NO_3 concentrations we decided to investigate this further. In general, NO_3 concentrations are expected to be high in the spring due to runoff and low in the summer due to uptake and diminished terrestrial inputs. That 2008 and 2014 exhibited higher NO_3 concentrations could be due to contemporary cell lysis events, a relaxed demand for the substrate (i.e., smaller blooms) or dominance of a strain that preferentially uses other sources of nitrogen (e.g., urea

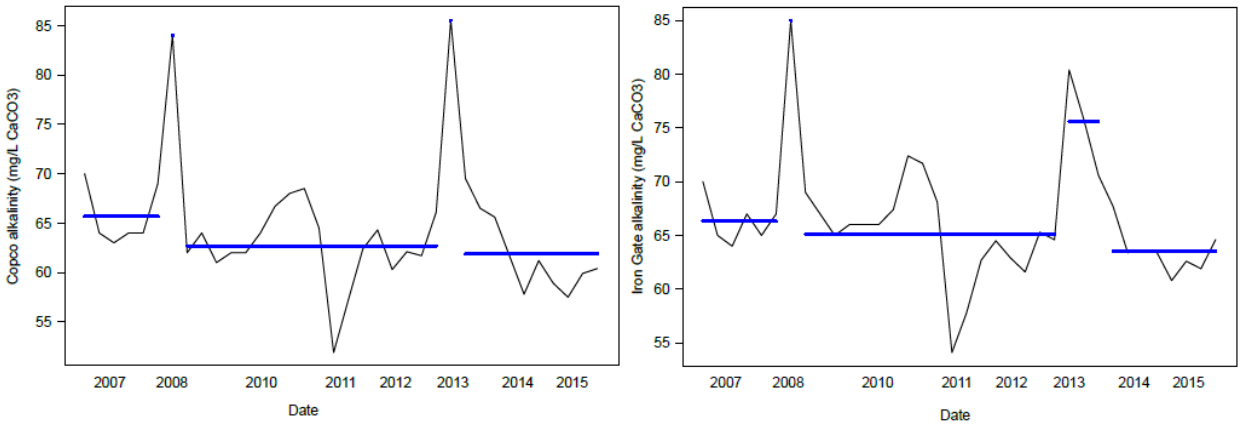
or other organic sources). The cell count data indicates that 2008 and 2014 had low to average *Microcystis* cell concentrations, whereas 2007 and 2013 were characterized as having significantly larger blooms than average (Fig. 14). Under this context, it is not surprising that NO_3 concentrations would be lower in 2007 and 2013 and that they would be higher in 2008 and 2014 since there were fewer *Microcystis* cells present in these latter years and therefore, likely less biological nitrogen demand. However, it cannot be definitively determined from these data if the higher NO_3 concentrations selected for the non-OTU1 genotypes or if it is just the by-product of fewer cells in the system. These possibilities could be further explored by sub-sampling mesocosms either diluted or amended with NO_3 in order to determine which strains are most competitive under each condition.

Figure 14. Change point graph denoting periods of significantly different *Microcystis* cell concentrations in Copco Reservoir



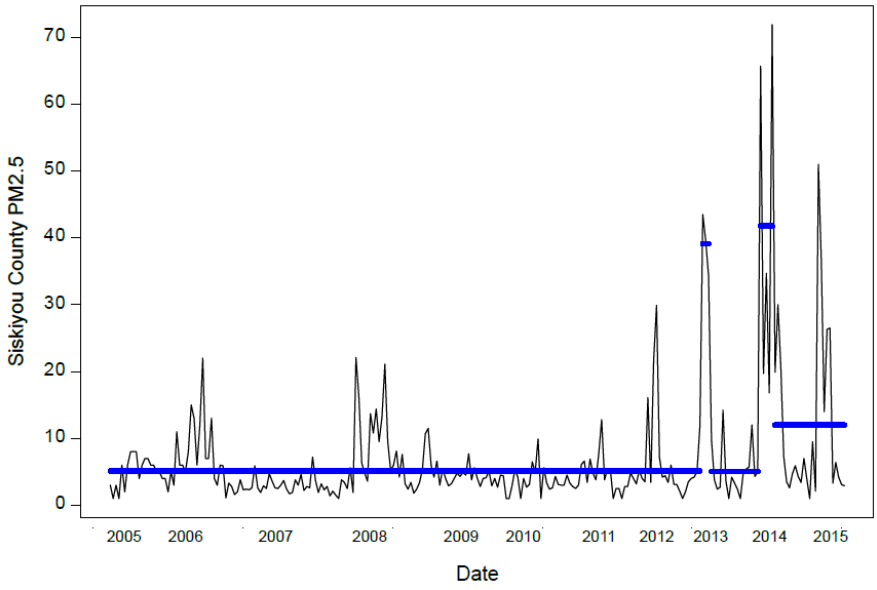
5. **Alkalinity** concentrations in both reservoirs were almost identical and exhibited similar inter-annual patterns, with 2008 and 2013 characterized by significantly higher concentrations and 2011 with lower concentrations; the latter was probably due to high flows and dilution. Note that alkalinity data were not collected in 2009. Change point analysis indicated that both reservoirs experienced significant increases in alkalinity around 6/18/2008 and 6/20/2013 (Fig. 15). During both of these dates in Copco Reservoir *Microcystis* strain OTU1 was dominant, whereas OTU3 was dominant in Iron Gate Reservoir during this period in 2008. OTU1 was dominant in Iron Gate Reservoir in 2013 throughout June and July. The lack of coherence between dates and strains suggests that alkalinity does not significantly influence strain composition.

Figure 15. Change point graph denoting periods of significantly different alkalinity concentrations for Copco reservoir (left panel) and Iron Gate reservoir (right panel).



6. Particulate matter (PM2.5) measured in Siskiyou County exhibited significantly difference concentrations throughout the study period. Change point analysis on Siskiyou $PM_{2.5}$ data indicated with greater than 95% certainty that the periods of 7/21/13 to 8/14/13 and 7/28/14 to 9/2/14 were distinctly different from all other periods (Fig. 16).

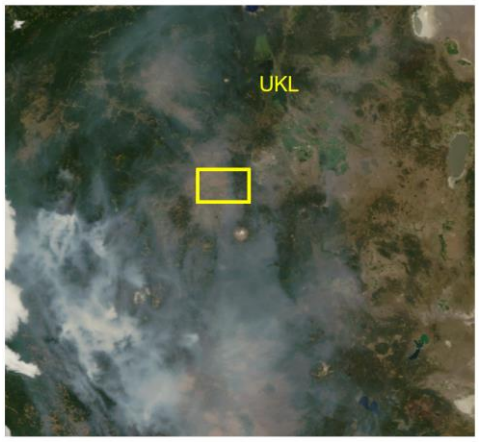
Figure 16. Change point graph denoting periods of significantly different $PM_{2.5}$ air concentrations in Siskiyou County



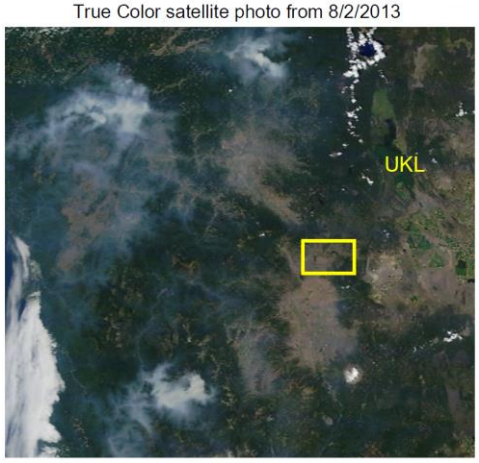
In late July 2013 wildfire smoke obscured the region and resulted in the increase in $PM_{2.5}$ (Figure 17). Concomitant with this period, the *Microcystis* population in Copco Reservoir transitioned from nearly complete dominance by OTU1 (99.1%) on 7/23/13 to 14.8% on 8/21/13 (Figure 5). The relative abundance of this strain continued to decline to 4.4% by 9/25/13, even after the air quality had improved. In Iron Gate Reservoir, this transition also occurred, but was not as dramatic; the *Microcystis* population went from complete dominance by OTU1 (99.6%) on 7/23/13 to only 38.5% relative abundance over this same period. The supplanting strains were primarily OTU2 and a previously unobserved strain labeled "Other" in Figure 6.

Figure 17. True color satellite images of Northern California wildfires during the summers of 2008, 2013 and 2014; the yellow squares indicate the study location.

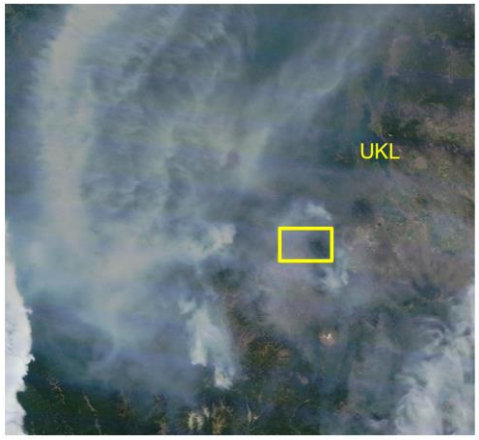
True Color satellite photo from 7/19/2008



True Color satellite photo from 8/2/2013



True Color satellite photo from 8/3/2014



On 8/3/2014, wildfire smoke was again prevalent within the region (Figure 17) and this period again coincided with a decline in OTU1 in Iron Gate Reservoir, where it was replaced by OTU3. However, unlike in Iron Gate Reservoir, the *Microcystis* population in Copco Reservoir was never comprised of more than 7% OTU1. Although less extreme than in 2013 and 2014, there was still considerable wildfire smoke throughout the region in 2008. The 2008 wildfire, named the Klamath Theater Complex, was the seventh largest wildfire in California history, consuming an estimated 192,038 acres (Cal Fire, 2017). That the 2008 wildfire appears smaller on the basis of the $PM_{2.5}$ data is probably due to local wind conditions and the location of the Siskiyou air quality monitoring station relative to the fire itself. The 2014 wildfire, named the Happy Camp Complex, was the 17th largest wildfire in state history, consuming an estimated 134,056 acres. The 2013 wildfire, called the Forks Complex, began on 7/31/13 and consumed an estimated 37,246 acres. A study on the effects of wildfire on visible light absorption during the 2008 fire documented significant light absorption as far away as Reno, NV during this event (Gyawali et al., 2009), demonstrating that smoke effects extend throughout the region and far beyond the fire epicenter itself.

It is unclear what physiological adaptation would enable strains such as OTU3 to dominate during periods of reduced light intensity or light quality. We hypothesized that this strain may contain the photopigment phycoerythrin, which would enable it to harvest additional light from the green portion of the visible

light spectrum; an adaption that enables certain cyanobacteria (e.g., *Planktothrix rubescens*) to sustain growth rates when light intensities are low. The genes for phycoerythrin biosynthesis were identified in the shotgun metagenome data set, but they all belonged to a *Pseudanabaena* strain. In addition to strain specific light harvesting pigments, we can speculate that nontoxic strains may experience a competitive advantage over toxigenic strains when the selection pressure favoring microcystin production is removed. For example, if high light and/or oxidative stress conditions enable toxigenic strains to outcompete nontoxic strains due to its protective effect within the cell, then under low light conditions microcystin metabolites provide no competitive advantage and actually come at a detrimental cost due to their metabolic expense. This concept has been demonstrated in culture studies where toxic and nontoxic strains of *Microcystis* were grown as mixed assemblages in chemostats exposed to high or low light conditions (Kardinaal et al., 2007); the nontoxic strains were found to be superior competitors under low light conditions.

CONCLUSIONS

The shotgun metagenomic sequencing data corroborated the results of other assessments (e.g., QPCR) that have implicated OTU1 as being the primary microcystin-producing strain in the system. Strain assessments by *cpcBA* sequencing and OTU clustering were well corroborated by the shotgun metagenomic sequencing. However, there are likely "blind spots" that are missed by only sequencing the *cpcBA* locus as indicated by evidence of strain diversity in shotgun metagenomes corresponding to periods when only one OTU was detected (e.g., during 2012 only one cpcBA genotype was observed but evidence of *Microcystis* diversity was present in the metagenomes). The complexity of data collected by shotgun metagenomes and short sequence lengths limits our ability to fully distinguish the genomes of co-occurring strains that share some, but not all of their genes. Producing reference genome sequences using the PacBio long-read sequencing technology would help to clarify the diversity of *Microcystis* strains present in Copco and Iron Gate reservoirs and their relationship to strain designations made on the basis of a single gene locus such as *cpcBA*. Likewise, genome sequencing would help to confirm the conclusion from this study that microcystin toxicity is only associated with OTU1, and that there are more nontoxigenic than toxigenic strains in the system.

Regarding physicochemical drivers that might influence toxicity via altered bloom composition, a number of water quality variables were found to significantly vary seasonally from year to year, although none did so in a temporally coherent manner coincident with the *Microcystis* strain succession patterns witnessed throughout the course of this study in both reservoirs investigated. Based on the available data, we believe that a reduction in light intensity and/or a shift in photosynthetically available radiation (PAR) toward the red end of the visible light spectrum due to wildfire smoke is the most plausible explanation for the observed strain turnover events. Because wildfires exert their influence regionally, differences in water chemistry between the reservoirs is largely irrelevant, which may explain why strain succession patterns between reservoirs tend to exhibit similar patterns even though the reservoirs are rather distinct with regard to their physicochemical makeup. The possibility that NO_3 may also influence strain composition was not satisfactorily resolved from the data on hand. High NO_3 concentrations did correspond with OTUs 2-3 and this could be due to lower cell biomass or due to positive selection.

AREAS OF FUTURE STUDY

We would propose conducting a mesocosm bioassay study to assess the effects that light conditions and nitrogen concentrations have on strain dominance. Directed experiments will be necessary to validate these hypotheses. These could be run within the reservoir using limnocorrals open to the air and seeded with sediments.

ACKNOWLEDGEMENTS

This study was funded by PacifiCorp for the purpose of increasing understanding of cyanobacterial dynamics in the Klamath River.

LITERATURE CITED

Bozarth CS, Schwartz AD, Shepardson JW, Colwell FS, Dreher TW. 2010. Population turnover in a *Microcystis* bloom results in predominantly nontoxigenic variants late in the season. *Appl Environ Microbiol* 76:5207-5213.

CalFire. 2017. Top 20 largest California wildfires. Accessed Feb. 8, 2017. http://www.fire.ca.gov/communications/downloads/fact_sheets/Top20_Acres.pdf

Cooper W, Zika R. 1983. Photochemical formation of hydrogen peroxide in surface and ground waters exposed to sunlight. *Science* 220:711-712.

Dreher TW, Otten TG, Mackey S. 2013. Comparison of *Microcystis* Populations in Copco/Iron Gate Reservoirs and in the Klamath System Upstream and Downstream: Are the Populations Linked or Largely Independent? Technical Report prepared for PacifiCorp, Jan-2013.

Gyawali M, Arnott WP, Lewis K, Moosmüller H. 2009. In situ aerosol optics in Reno, NV, USA during and after the summer 2008 California wildfires and the influence of absorbing and non-absorbing organic coatings on spectral light absorption. *Atmos Chem* 9:8007-8015.

Kardinaal WEA, Tonk L, Janse I, Hol S, Slot P, Huisman J, Visser PM. 2007. Competition for Light between Toxic and Nontoxic Strains of the Harmful Cyanobacterium *Microcystis*. *Appl Environ Microbiol* 73:2939-2946.

Lyck S. 2004. Simultaneous changes in cell quotas of microcystin, chlorophyll *a*, protein and carbohydrate during different growth phases of a batch culture experiment with *Microcystis aeruginosa. J Plankton Res* 26(7):727-736.

Moisander PH, Ochiai M, Lincoff A. 2009. Nutrient limitation of *Microcystis aeruginosa* in northern California Klamath River reservoirs. *Harmful Algae* 8:889-897.

Otten TG, Crosswell JR, Mackey S, Dreher TW. 2015. Application of molecular tools for microbial source tracking and public health risk assessment of a *Microcystis* bloom traversing 300 km of the Klamath River. *Harmful Algae* 46:71-81.

Paerl HW, Otten TG. 2013. Blooms Bite the Hand That Feeds Them. *Science* 342:433-434.

Wilson AE, Wilson WA, Hay ME. 2006. Intraspecific Variation in Growth and Morphology of the Bloom-Forming Cyanobacterium *Microcystis aeruginosa*. *Appl Environ Microbiol* 72:7386-7389.

Zilliges Y, Kehr JC, Meissner S, Ishida K, Mikkat S, Hagemann M, et al. 2011. The Cyanobacterial Hepatotoxin Microcystin Binds to Proteins and Increases the Fitness of *Microcystis* under Oxidative Stress Conditions. *PLOS ONE* 6(3):e17615.