Widespread presence of cyanotoxins and taste-and-odor compounds within benthic algae of human-disturbed rivers

By

© 2022

Zane D. Rider

B.S., University of Tennessee, 2016

Submitted to the graduate degree program in Civil, Environmental, and Architectural Engineering and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

_____

Chair: Admin Husic, Ph. D.

_____

Belinda Sturm, Ph. D.

_____

Ted Harris, Ph. D.

Date Defended: 6 September 2022

The thesis committee for Zane D. Rider certifies that this is the approved version of the following thesis:

Widespread presence of cyanotoxins and taste-and-odor compounds within benthic algae of human-disturbed rivers

Chair: Admin Husic, Ph. D.

Date Approved: 15 September 2022

Abstract

Cyanotoxins and taste-and-odor (T&O) compounds are produced by cyanobacteria, including benthic varieties that grow on the bed of freshwater streams. Cyanotoxins, such as microcystin, anatoxin-a, and saxitoxin, can poison or even cause death if consumed in high quantities, while T&O compounds, such as geosmin and 2-Methylisoborneol (MIB), can cause a noxious taste and smell in drinking water. The negative effects of cyanotoxins and T&O compounds have the potential to worsen as the effects of climate change and urbanization add additional pressures to already sensitive stream ecosystems. Understanding how urbanization impacts the production of cyanotoxins and T&O compounds is crucial to lowering the impact these metabolites have on our health, safety, and environmental well-being. To understand this relationship, we surveyed three streams in the rapidly urbanizing Johnson County, KS, which range from highly rural to highly urban. A total of nine benthic algae mats were sampled repeatedly over the summer and early fall of 2021. Benthic algae mat locations were analyzed using a combination of in situ sensors and laboratory testing to assess the production of cyanotoxins and T&O compounds as well as the potential physicochemical and biological drivers of this production. Water quality parameters, including temperature, pH, specific conductivity, and nutrients, were measured during field visits as were benthic algae parameters, including cyanobacterial concentration, taxonomic make-up, stable isotopes, toxin analysis, and genetic analysis. To explain the drivers of toxin and T&O production, we develop predictive linear regression and random forest models. The results of our field sampling showed widespread prevalence of cyanotoxins and T&O compounds, irrespective of upstream land use to a benthic mat. Microcystin and geosmin were detected in every benthic algae sample we collected, whereas anatoxin-a, saxitoxin, and MIB were found in 65%, 26%, and 84% of samples, respectively.

Results of the linear regression and random forest models generally indicate that toxin gene abundance, cyanobacteria abundance, temperature, and specific conductivity are the dominant predictors of microcystin, anatoxin-a, and saxitoxin production, and that cyanobacteria abundance, temperature, and specific conductivity are the dominant predictors of geosmin and MIB production. In summary, our results show that any human-derived environmental impact, whether it be urbanization or ruralization, can lead to the occurrence of microcystin, anatoxin-a, saxitoxin, geosmin, or MIB within freshwater benthic cyanobacteria. Cyanotoxin and T&O compound forming benthic algae appear to be a common feature of human-disturbed environments and our results can assist with informing water utility and municipality managers with cyanobacterial monitoring strategies and remediation.

Ackowledgements

I firstly would like to thank my advisor, Dr. Admin Husic, for if not for his willingness to interview a quality engineer living in Kentucky, his flexibility in helping a student new to the Water Resources field succeed, and his mentorship in how to hold myself to a higher scientific standard I would not have had this amazing opportunity. Next, I would like to thank the Kansas National Science Foundation Established Program to Stimulate Competitive Research (NSF EPSCoR) FIRST Award Program and the United States Geological Survey (USGS) Kansas Water Resources Institute (KWRI) 104(b) Grant Program for their direct assistance in funding this research project. This opportunity allowed me to pursue my dream of attaining a Master of Science degree in a field I think will have a huge impact on bettering our future, Water Resources Engineering. I also want to thank my fiancée, Avery Day, whose love and support helped motivate and focus me during difficult sections, and always gave me time to listen to my presentation drafts, look at my posters, and even endure a water resource rant or two.

I want to thank the other professors in my committee, Dr. Ted Harris and Dr. Belinda Sturm, especially for their feedback from our progress meetings and enthusiasm in responding to my questions. I would also like to thank the professors in the water resources department, Dr. Admin Husic (again), Dr. Amy Hansen, and Dr. Joshua Roundy, who taught me exciting concepts in hydrology, hydraulics, water quality, modeling, and crucial water sources. Outside of water resources, I experienced fantastic professors in geology including, Dr. Sharon Billings, Dr. David Fowle, and Dr. Marcia Schulmeister, and expanded my horizons with their lectures and discussions.

In processing the many variables this research entailed, many labs were used both inside and outside of the University of Kansas. Dev Hiripitiyage taught me how to perform several

analyzes in the lab and handled the arduous task of processing the qPCR data for all our samples. Brooks Welner, Kyle Eckert, Luis Carretero and the team at WaterOne were fantastic to work with and their GC-MS analysis of our geosmin and MIB data was tremendously helpful. Joe Yelton and the team at Johnson County Water Quality Lab were always flexible and accepting of our water nutrient samples and their analysis was extremely appreciated. John Beaver at the BSA Environmental labs provided great taxonomic data on our samples, and I appreciate his quick response to my questions. The University of Arkansas Stable Isotope Laboratory were also fantastic to work with on analyzing our samples for isotopic results, and I appreciate their quick feedback and turn-around. For our Chl-a analysis we reached out to two wonderful labs, Jessica Scholz at the University of Missouri's Limnology Laboratory and Dr. Alan Wilson at the University of Auburn's WilsonLab. They were both wonderful to work with and tried to help get the data back in the best shape, but unfortunately the results were not used in this analysis. I want to thank Kent Dye and David Woody who were always good-humored even when the steel-plate mechanisms we designed for algae growth platforms were not used in the research.

Finally, I want to thank my lab-mates and friends, Amirreza Zarnaghsh, Zack McVey, Aaron Tuschhoff, Devin Gouger, Abby Percich, Sarah Hogan, Austin Gardiner, Maral Khodadadi, and Jack Respeliers who helped during wildly hot field visits, wading through algae filled streams, monotonous laboratory processes, and more. Ecohydraulics people rule!

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Freshwater benthic cyanobacteria provide many benefits to their surrounding freshwater ecosystem, including nitrogen fixation and primary production, however these same benthic cyanobacteria can have widespread and costly impacts on water quality as well as health and safety (Bauer et al., 2020; Dunlap et al., 2015; Scott and Marcarelli, 2012). Benthic cyanobacteria are a portion of the phylum of cyanobacteria that grow attached to substrates in water systems. They typically grow in colonies of many different algae and bacteria in what is referred to as a mat which describes the entire structure of the benthic community. Benthic mats include extracellular material, sometimes called the matrix, that helps bind the algal colonies together and provides other benefits for the algal community, including acting as a carbon source and sink (Stuart et al., 2016). Some cyanobacteria in benthic communities are capable of making secondary metabolites that are toxic to many creatures (Kaloudis et al., 2022). Death from the consumption of the benthic mat material has been linked to the toxic metabolites within the benthic cyanobacteria (Bauer et al., 2020; Kaloudis et al., 2022; Mez et al., 1997; Wood et al., 2010). These toxins include a hepatotoxin, microcystin, which can cause lesions in the liver, and two neurotoxins, anatoxin-a and saxitoxin, which cause respiratory tract paralysis and paralytic shellfish poisoning, respectively (Sivonen and Jones, 1999). There are more cyanotoxins, including nodularin, cylindrospermopsin, and others, but they were not analyzed as part of this study. In addition to toxins, cyanobacteria can also produce nuisance T&O compounds like geosmin and 2-Methylisoborneol (MIB) (Jüttner and Watson, 2007). These compounds are detectable at very small concentrations (near 10 ng/L) (Watson et al., 2008), and, if left untreated in drinking water, can impart an earthy or musty smell and noxious taste that nearly all humans can detect. While these effects are primarily undesirable rather than dangerous, they have the potential to cause

distrust in the local drinking water treatment and raise concerns for public services around water. Therefore, drinking water treatment plants spend considerable funds on filtration and treatment processes to remove geosmin and MIB (Dietrich, 2006; Dunlap et al., 2015). Most research has focused on the harmful cyanobacteria blooms of the floating planktonic variety in larger bodies of water like oceans, lakes, and large rivers (Wood et al., 2020). However, benthic varieties of cyanobacteria are still capable of producing these harmful metabolites and are relatively understudied.

Benthic cyanobacteria are primary producers, so like many other photosynthetic bacteria and algae, there are many physicochemical and biological drivers that impact the growth and spread of the biomass. Cyanobacteria have been known to thrive in a huge variety of climates, including polar streams or downstream of geothermal vents (Scott and Marcarelli, 2012). As photosynthesizing organisms require light, turbidity and dissolved solids in the water have been known to reduce total biomass, however, cyanobacteria are particularly adept at low-light and can dominate in these environments (Scott and Marcarelli, 2012). Also, while nutrients such as nitrogen and phosphorus have been shown to have correlations with benthic cyanobacteria biomass (Busse et al., 2006), there are nitrogen-fixing species of cyanobacteria that can dominate in low nitrogen environments as well (Howarth et al., 1988). Cyanobacteria mats can detach from the substrate once a sufficient accumulation of oxygen has been captured in the mat matrix and the buoyant force detaches the biomass from the substrate (Bouma-Gregson et al., 2017), further increasing their chance of being eaten if the buoyant mat accumulates on the shore. While connections have been made between cyanobacteria and various physicochemical and biological drivers in previous studies, there is considerable uncertainty in how these drivers promote the production of specific cyanotoxins and T&O compounds within benthic cyanobacterial mats.

Understanding how the many drivers impact benthic algae and their production of these cyanotoxins and T&O compounds in freshwater stream systems, including those with anthropogenic sources, will be important to our present and future water security.

Human-caused impacts on freshwater streams can be seen and measured in many ways, particularly in urban environments. Urban streams tend to widen due to increased bank erosion from flashy events and reduced vegetation, and a widened stream will have shallower flows resulting in greater light penetration to the substrates as well as more riffle and run sections (Brown et al., 2009). Due to impervious surfaces, runoff from a much wider area will make its way to the stream, and it will carry nutrients, sediment, and other chemicals such as insecticides (Brown et al., 2009; Zarnaghsh and Husic, 2021). Finally, water treatment and wastewater treatment facilities with discharges into the streams can release water into the stream with higher nutrient loads and higher temperatures (Carey and Migliaccio, 2009; Kinouchi et al., 2007). The combination of increased light exposure, better flow stability, higher temperatures, and increased nutrients all lead to environments that allow benthic cyanobacteria to flourish. On top of this, climate change is bringing about warmer waters as well as increased nutrient loads which both will lead to increased cyanobacteria biomass and bloom occurrence (Paerl and Paul, 2012; Paerl and Barnard, 2020).

Numerous methods and approaches for quantifying and understanding cyanotoxins and T&O compounds have been developed in recent decades. Recently, a portable in situ fluorometer device, called the BenthoTorch (bbe Moldaenke GmbH, Schwentinental, Germany), was developed to measure three types of benthic algal groups (cyanobacteria, green algae, and diatoms) within seconds in the field. The BenthoTorch has been used to varying degrees of success to measure cyanobacteria concentrations on both natural and artificial substrates in streams, and it has shown promising correlations with both laboratory chlorophyll-*a* (chl-*a*) concentrations and

taxonomic analyses (Echenique-Subiabre et al., 2016; Harris and Graham, 2015; Rosero-López et al., 2021). Taxonomic identification processes utilize microscopy to visually identify algae and bacteria species from a sample, and based on the counts of each species, sample concentrations can be determined. Enzyme linked immunosorbent assays (ELISA) are used to measure concentrations in water-based samples for the specific cyanotoxin. For benthic cyanobacteria samples, this usually involves lysing the cells to expose the intracellular cyanotoxins for easier quantification. Another breakthrough process is quantitative polymerase chain reaction (qPCR) analysis which can be used to find specific gene sequences that have been found to be genetic code for a certain step in the production of the toxin or T&O compound in question. By dividing the ELISA toxin concentration result by the qPCR genetic concentration, a toxin quota can be determined (Thomson-Laing et al., 2020). From this, one can not only calculate the quantity of cyanotoxins or T&O compounds released by a cyanobacteria sample, but also how many cyanobacteria cells have the genetic code to create them in the first place. This metric indicates whether the available cyanobacteria are producing cyanotoxins in large or small quantities given the relative abundance of a gene copy. This discrepancy can provide further depth and analysis into the relationships between the physicochemical and biological drivers being studied and the response of the cyanobacteria. Finally, predictive multivariate statistical models, like linear regression models, random forest models, cubist models, and generalized additive mixed modeling (GAMMs) can clarify the understanding of which environmental drivers are the most impactful to cyanobacteria assemblages, cyanotoxins, and T&O compounds (Harris and Graham, 2017; McAllister et al., 2018). While the aforementioned methods have seen considerable disparate use in the benthic algae mat literature, we find no studies that combine in situ sensing, taxonomic identification and enumeration, toxin analysis, stable isotope analysis, genomic analysis, land use

characteristics, and predictive modeling under a single framework for understanding the drivers of toxin and T&O compound production.

The main objective of this research was to identify the drivers of cyanotoxin and T&O compound production within the benthic algae mats of human-disturbed rivers. We hypothesized that – because urbanization results in higher nutrient concentrations, less riparian canopy, and shallower flow depths – urban rivers would be characterized by 1) greater proliferation of benthic algae mats, 2) larger concentrations of benthic cyanotoxins and taste-and-odor compounds, and 3) increased density of toxin producing genes within mat matrices. We tested these hypotheses by sampling benthic material from nine stream reaches, across a steep rural-to-urban land use gradient, on a weekly basis over the course of three months. Thereafter, we conducted physicochemical, isotopic, taxonomic, toxicological, and genomic analyses on sampled mat material. Lastly, we developed predictive linear regression and random forest models to quantify the relative importance of the environmental factors driving harmful and nuisance compound production within benthic algae mats.

Chapter 2: Study Site

This study was performed across three streams in Johnson County, Kansas, which lies directly southwest of Kansas City, Missouri, a rapidly expanding metropolitan area (Figure 1). The streams are Indian Creek (168 km^2), Mill Creek (150 km^2), and Blue River (170 km^2). Indian Creek is the most urban watershed with 93% of the basin area characterized as urban (MRLC, 2019). Further, approximately 47% of the Indian Creek watershed is classified as impervious (Rasmussen and Gatotho, 2014). Mill Creek is a mixed-use watershed and shows a combination of urban (68%) and rural (32%) features (Figure 1). Finally, Blue River is the rural watershed with 30% urban land-use and 55% pasture and cropland. Indian Creek and Mill Creek receive waste-water treatment facility discharge, with the Mill Creek facility processing 3 MGD, and the two Indian Creek facilities processing a combined 22 MGD. The climate and soil types across each watershed are similar due to their proximity to each other, so they present a unique opportunity to study differences in water quality based on anthropogenic disturbances and activities. Individual sites were of similar sizes, but some exhibited varying characteristics nearby, such as bridges, streambank stabilization, or were mostly untouched (Figure 2).

In summer 2020, a longitudinal inventory of algae mats was conducted by surveying each stream and visually identifying mat material (Figure 3). In total, we walked approximately 30 km of stream length, 10 km per river. We found a total of 83 stream reaches with benthic algae mats: Indian Creek had 47, Mill Creek had 31, and Blue River had 5. While the urban sites ostensibly have a greater density of benthic mats, the aim of this longitudinal assessment was not to enumerate every mat in the streams, but rather to get an inventory of potential mat locations where we could perform a greater in-depth analysis at a later time. The true number and density of mats is likely greater than what we report, particularly for rural Blue River. This is because turbid water, deeper

flow depths, and riparian brush in Blue River made it more difficult to perform rapid visual assessment of mats than in the more-urban streams.

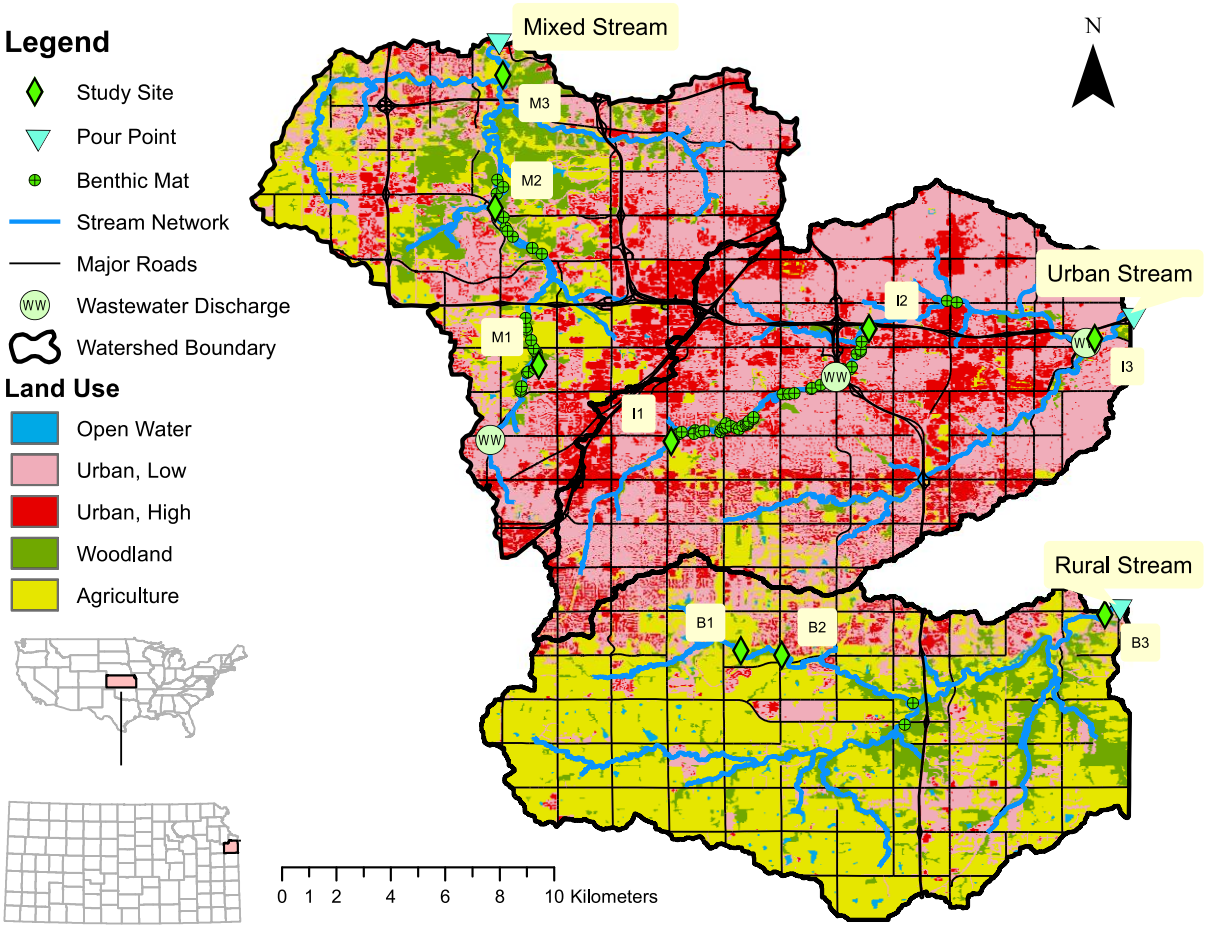

Figure 1: Locations of sampling sites. The three watersheds are displayed here, with the Mixed Stream being Mill Creek, the Urban Stream being Indian Creek, and the Rural Stream being Blue River. Land use, wastewater treatment plants, and other benthic mats identified are on the map as well. Land use data from NLCD Land Cover (Conus) 2019 (MRLC, 2019).

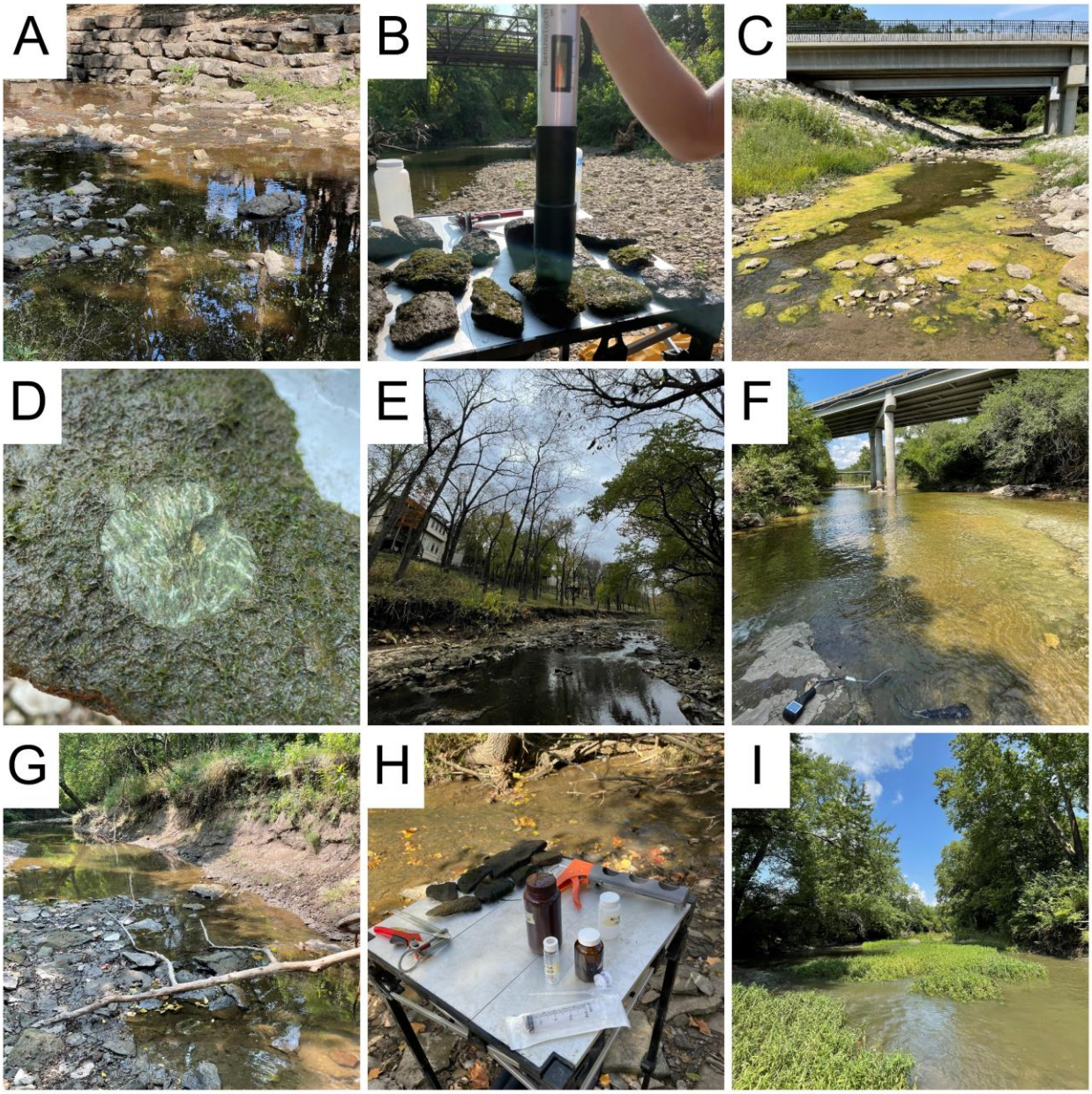

Figure 2: A grid of images taken from field visits of various study sites and field methods used. (**A**) Indian Creek site 1 (I1), latitude: 38.896, longitude: -94.760. (**B**) BenthoTorch being used on a cobble removed from Mill Creek site 3 (M3), latitude: 39.017, longitude: -94.815. (**C**) Bridge that was under construction during sampling period and algae bloom at Blue River site 2 (B2), latitude: 38.826, longitude: -94.723. (**D**) Picture taken of in process algae scrape sampling from cobble showing filamentous benthic algae mat growth. (**E**) I1 site again, but from final field visit to show loss of vegetation cover due to leaf fall. (**F**) Mill Creek site 2 (M2) with Horiba U-52 at bottom of image, latitude: 38.973, longitude: -94.818. (**G**) Blue River site 1 (B1), latitude: 38.827, longitude: -94.737. (**H**) Field table with various sampling tools used including bar-clamp sampler, various bottles, syringe, syringe filters, caulk gun, tool cleaning solution, and some cobbles. Not all tools are shown. (**I**) Blue River site 3 (B3), latitude: 38.839, longitude: -94.616.

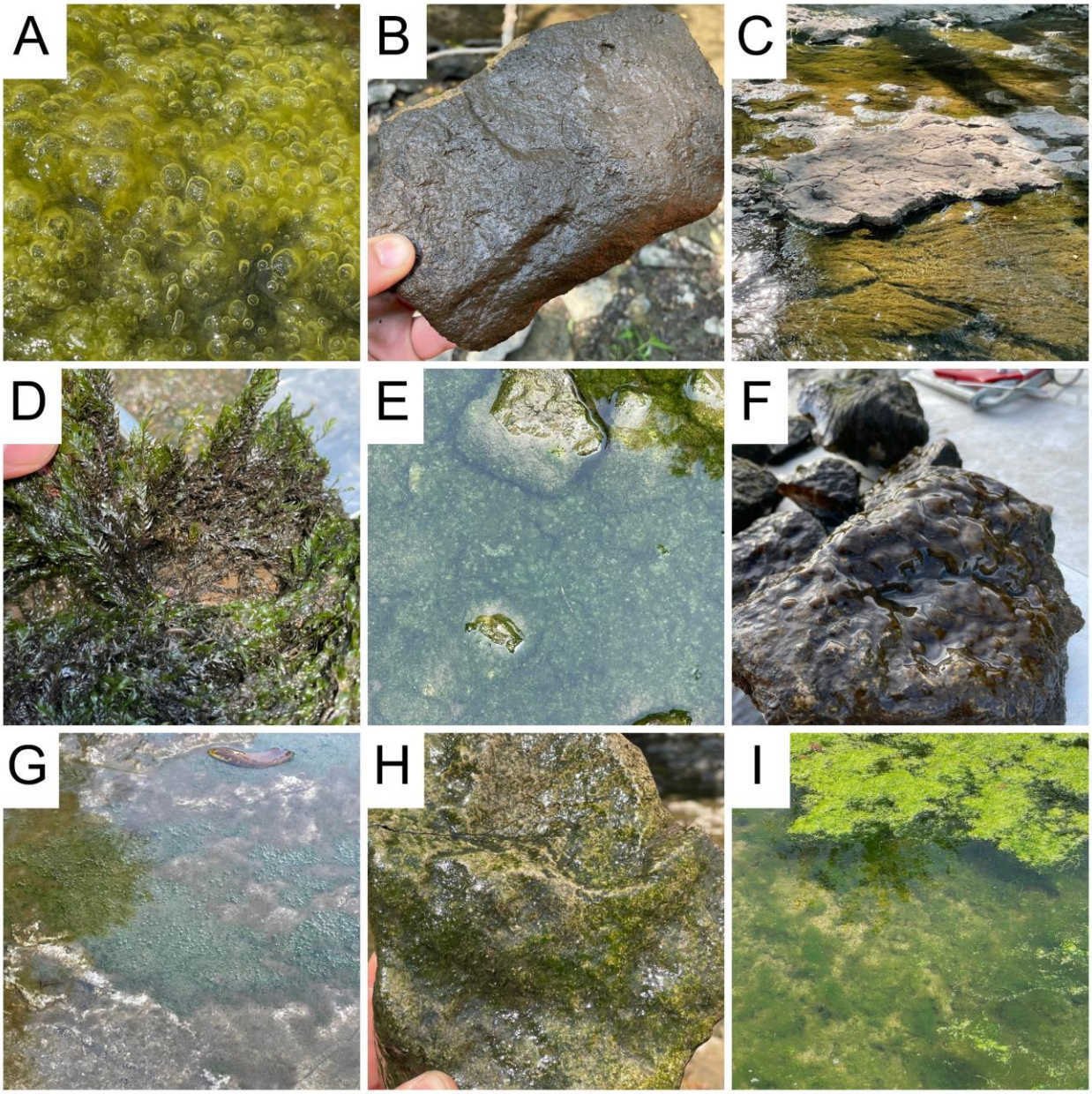

Figure 3: A grid of images taken from field visits of various cobble samples and algae mats. (**A**) Bubbles trapped in thick algae matrix (**B**) Sediment covered cobble (**C**) Long filamentous algae mat growth (**D**) Macroalgae growth on cobble (**E**) Dense, green algae growth near water surface (**F**) Non-filamentous algae growth on cobble (**G**) Thin film of algae growing with trapped bubbles (**H**) Example of a cobble with very thin algae growth after previous growth senesced (**I**) Thin algae mat growth with detached buoyant mats above

Chapter 3: Methods

3.1 Study Sampling Design

Between the 2021 late summer and early fall season (August 7th to November 5th), we conducted a field sampling campaign of 9 stream reaches across our study area (Figure 1). The nine study reaches were selected based on the extent of algae present at each site and a preliminary BenthoTorch measurement to confirm cyanobacteria presence. Among the nine sites, three were selected per stream with an effort to evenly space the sites longitudinally within each basin. Blue River site selection was most complicated due to stream access difficulties and private land surrounding a substantial portion of stream length. Nonetheless, the two Blue River sites that are close to one another have considerable local-scale differences with the middle site experiencing suburban bridge construction during sampling (Figure 2c) while the upstream site was adjacent to croplands.

In total, we made eight field visits, on a weekly-to-biweekly basis, collecting benthic material and assessing field conditions at the nine sites. Due to the length of time required to visit all nine locations and conduct all sampling tasks, field visits were typically split into two days, usually covering 5 to 6 sites on day one and 3 to 4 sites on day two. Sampling visits were selected such that limited hydrologic activity occurred in between the two-day collection periods. The order in which the sites were sampled was shuffled each time to prevent time-of-day biases on measurements. At each sampling site, several sampling methods were used including in situ sensors, water grab samples, and benthic mat scrape samples (Figure 4). For benthic assessment, fifteen cobbles were sampled along a 100 m longitudinal transect at each reach site. Most cobbles were taken from water depths ranging from <1 cm to 40 cm, with rare instances of 40-60 cm depths. Water samples were taken from the runs just upstream of the largest riffle in the study site.

Field duplicates were performed on the last 4 sampling trips, where one site in each trip had every field and laboratory sample duplicated.

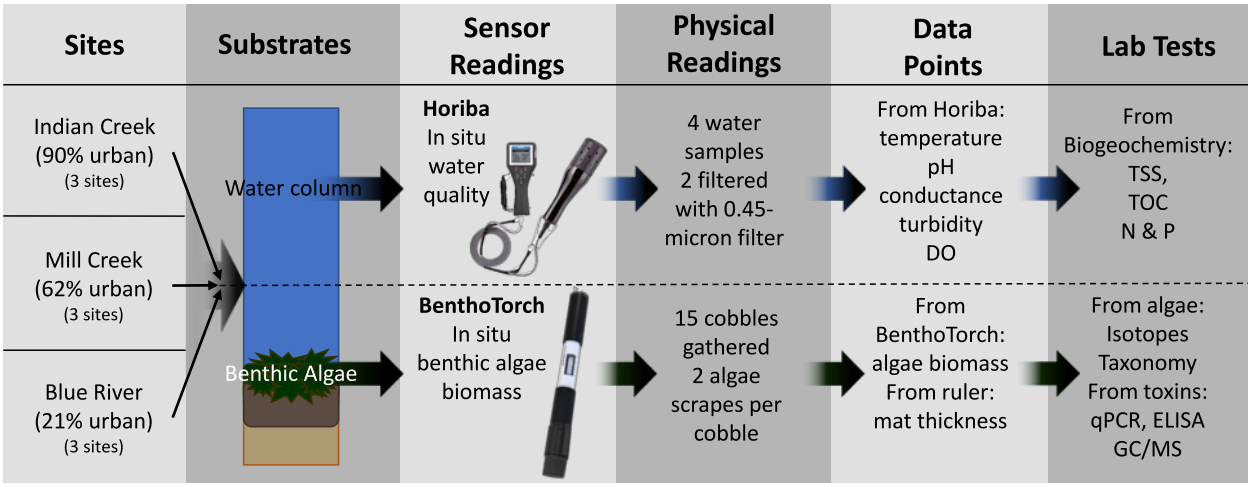

Figure 4: Flow chart of sampling done in study listing sites, substrates data taken from, *in situ* sensors used, source of some data points, and lab sourced data.

3.2 Physicochemical Drivers of Benthic CyanoHABs

3.2.1 In Situ Sensor Sampling

The physicochemical parameters temperature, pH, conductance, turbidity, dissolved oxygen (DO), and total dissolved solids (TDS) were recorded using a multi-parameter water quality probe (Horiba U-52, HORIBA Advanced Techno Co. Ltd., Kyoto, Japan). The device was calibrated before each sampling run using Horiba calibration solution. During each sampling, the sensor's probe was placed in a riffle just upstream of where cobble samples were to be taken to prevent excess sediment being kicked up and washed through the probe's sampling volume (Figure 2F). The probe was left in the water while other sampling activities were performed, providing the sensor components time to acclimate to field conditions. After 10 minutes or more, readings were recorded. The device's KCl internal reference solution, used to calibrate pH, was past its recommended usage date during the first three runs, so those pH data points were removed. Thereafter, the internal reference solution was replaced, the device was re-calibrated, and subsequent samples showed no issues.

3.2.2 Field Sampling of Physicochemical Data

Benthic mat thickness was taken by ruler measurements on several of the selected cobbles. These readings were averaged together for an aggregate mat thickness. Light penetration was gathered by visual estimation of shade cover of the stream and recorded as a percent value, where 100% describes zero shade cover and 0% represents total shade. Raw stream water samples were taken for total suspended sediment (TSS) (500 mL plastic bottle) as well as total nitrogen and total phosphorus samples (125 mL plastic bottle). For water column nutrients, a filtered stream water sample, using a 0.45 μm Whatman syringe filter, was taken for nutrient analysis (nitrate, ammonia, and orthophosphate). Another filtered water sample, using a 0.45 μm Whatman syringe filter, was taken for dissolved organic carbon (DOC) analysis. All bottles were placed on ice in a cooler within 30 minutes of collection. Nitrate, ammonia, and orthophosphate samples were delivered to the Johnson County Water Quality Lab within 48 hours of collection. TSS and DOC samples were stored in the refrigerator and TN/TP samples were frozen upon returning to the lab.

3.2.2 Laboratory Processing of Physicochemical Data

For TSS, the volume of collected sample was first recorded and then the contents were filtered out using a 1 μm Whatman filter, dried in an oven at 100°C, and then weighed on a scale. Dissolved organic carbon was preserved with phosphoric acid, and then processed using a TOC combustion analyzer (TOC Torch, Teledyne Tekmar, Mason, Ohio, USA) following the EPA 9060A methodology (USEPA, 2004). Total nitrogen and total phosphorous were processed once all samples had been collected, and each process was done in duplicate for each sample. For total phosphorus, a standard curve was created using standard dilutions of potassium phosphate (0, 15, 50, 100, 500, and 1000 μg/L standards, $R^2 = 0.998$). Phosphorus in each sample was digested by adding potassium persulfate and autoclaving using the liquid setting for 45 minutes. After adding

the mixed reagent, the TP readings were measured using a spectrometer at 885 nm (UV160, Shimadzu, Kyoto, Japan). TN was processed similarly, except using appropriate standard solutions (0, 0.5, 1.0, 1.5, 2.0, 5.0, 10.0, and 20.0 mg/L standards, R^2 = 0.799), digestion solutions, and reagents, and it was read at 220, 225, and 230 nm. A second derivative calculation was done on the 3 results to get the absorbance per sample. At Johnson County Water Quality Lab, Nitrate, Ammonia, and Orthophosphate were measured using a process derived from the EPA 353.2, EPA 350.1, and EPA 365.1, respectively (O'Dell, 1996a, 1996b, 1996c). Detection limits for nitrate were 0.03 mg/L, for ammonia were 0.04 mg/L, and for orthophosphate were 0.05 mg/L.

3.3 Cyanotoxins and Taste-and-Odor Compounds

3.3.1 In Situ Sensor Sampling

At each site, 15 cobbles were collected from the stream bed at random from 1 to 3 different riffles or runs (Rasmussen et al., 2012). A BenthoTorch was used to collect a single benthic algae concentration measurement from each of these cobbles, except for the 4 sampling sites where duplicates were recorded on the same cobble in a different, but similar looking part of the cobble. The end of the BenthoTorch with the sensor and rubber seal cover are placed on the area of benthic material and a measurement is taken, which takes approximately 20 seconds to record as the sensor proceeds through its fluorometric analysis steps (Figure 2B). The BenthoTorch reports *in situ* benthic algae concentration (in $\mu g/cm^2$) measurements of diatoms, cyanobacteria, and green algae. The BenthoTorch pulses LEDs at 470nm, 525nm, and 610nm and then read the emission wavelength emitted by the algae at 680nm. The last LED is used in tandem at 700nm to compensate for the effects of background reflection. GPS data is also recorded for each measurement by the BenthoTorch.

3.3.2 Field sampling of benthic biomass

On the same 15 cobbles, a bar clamp sampler was used to grasp each cobble. The top of the bar clamp sampler has a welded ring with a rubberized gasket to press against the cobble to prevent any scraped sample from leaving the scrape area (Figure 5). Scraping was done with a clean laboratory spatula, and after each site, the spatula was cleaned with Luminox™ cleaner solution. The area scraped from each cobble was 13.16 cm^2. Scrapings from the 15 cobbles were aggregated into a single opaque Nalgene bottle and placed into a cooler within 30 minutes of removal from the stream. The last 4 sample runs also included a smaller area (1 cm^2) that was scraped and stored in an amber glass jar with stream water and Lugol's Iodine to preserve for taxonomic analysis. All collected samples were stored on ice in a cooler until the return to the lab.

Figure 5: Bar-clamp sampler use for holding cobble and pressing a fixed-area circle into the top of the cobble where the algae would be scraped from.

3.3.3 Laboratory Processing of Chlorophyll-a, Isotopic Data, and Taxonomic Analysis

Once the benthic mat scrape sample was returned to the lab, it was frozen. Frozen samples were then freeze-dried at around -52°C and 0.150-0.180 mBar over 2-3 days to ensure complete water extraction (FreeZone 6, Labconco, Kansas City, MO, USA). Freeze dried samples were then transferred to a Whirl-Pak bag, ground to a fine powder by pulverizing within the bag using finger pressure and swirling motions, and then weighed. If extra homogenization was required, several other techniques were used including a rubber hammer, tissue grinder, and pellet pestle tool. Once homogenized, the bulk dry benthic sample could be sub-sampled for the various analyses, including chl-*a* and phycocyanin pigment analysis (250 mg), carbon and nitrogen isotopic analysis (500 mg), cyanotoxin analysis (350 mg), and genomic analysis (250 mg). The pigment analysis samples were placed in 40mL amber glass vials, the isotopic subsamples were placed in 10 mL culture tubes, cyanotoxin subsamples were placed in 50 mL sterile centrifuge bottles, the qPCR subsamples were placed in sterile 5 mL culture tubes.

The chl-*a* and phycocyanin subsample was extracted using a 1 μm Whatman filter. DIDO water was used to rinse leftover sample that stuck to the bottle's walls when emptying into the filter apparatus. Once all standing water was filtered out of the sample, the filter with sample was removed, folded in half, placed in an aluminum tray, and folded again to seal out light from the sample. The folded, aluminum-enclosed samples were placed in the freezer. The WilsonLab group at the University of Auburn analyzed the samples for chl-*a* using a 24-hour 90% ethanol extraction process in darkness at 4 °C (Kasinak et al., 2015), measuring absorbance in a spectrophotometer at 665 and 750 nm wavelengths, and repeating the spectrophotometer measurements after an acidification process to correct for the phaeophytin readings. Phycocyanin samples were measured fluorometrically, extracting each sample in a 50 mM phosphate buffer also in darkness at 4 °C, and filtering (<0.2 mm) each sample (Kasinak et al., 2015). Note: while analysis of chl-*a* and

phycocyanin was performed, the results were not produced in time for the completion of this thesis. However, we plan to integrate these data into future work.

The carbon and nitrogen isotopic analysis was performed by the University of Arkansas Stable Isotope Laboratory. The reported data were $\delta^{13}C$, $\delta^{15}N$, C%, and N%. Samples were prepped in tin capsules and then loaded in a 50 well autosampler on an isotope ratio mass spectrometer attachment device for the combustion and reduction steps (EA Isolink CN Flash IRMS, Thermo Fisher Scientific, Waltham, MA, USA). Once combusted and reduced, it then passes to the isotope ratio mass spectrometer (Delta V Plus IRMS, Thermo Fisher Scientific, Waltham, MA, USA). The results are then normalized using data from USGS 41a and USGS 8573 (Qi et al., 2016, 2003). For %C and %N data, chalky soil was used to normalize the results.

Of the forty taxonomic analysis scrape samples, 24 were selected to be analyzed by microscopy at BSA Environmental Services (Beechwood, Ohio) for taxonomic counts, density, and biovolume. These samples were selected to cover a wide range of sampling dates, target and predictor variable quantities and some that were visually similar to cyanobacteria-dominated mats described in literature (Quiblier et al., 2013) in order to see what groups of benthic algae were being measured in Johnson County, KS, and what specific strains of cyanobacteria were present. These samples were all quantified on a per milliliter basis using the Utermohl method (Lund et al., 1958) in accordance with the American Public Health Association Standard Method 10200. Analysis was performed at 800X and 1260X magnification (Leica DMi1, Leica, Wetzlar, Germany). Counts for diatoms, green algae, cyanobacteria, cryptophyta, and euglenophyta were all recorded, but in the analysis, the latter two will be ignored.

3.3.4 Laboratory Processing of Cyanotoxin and T&O Concentrations

To assess the intracellular content of cyanotoxins and T&O compounds, we had to lyse the algal mass into solution. First, the freeze-dried algae subsample was placed into a 50 mL sterile centrifuge bottle, which was filled with 40 mL of DIDO water. We decided against adding 0.1% formic acid, as others have done in the past (McAllister et al., 2018), because the formic acid caused some errors in the determination of T&O compounds using the GC/MS instrument in WaterOne's lab. The sample was then freeze-thawed three times to lyse toxins out of the benthic mass. The process included freezing at -20°C for 2-3 hours, thawing in an ultrasonic water bath at 25°C for 30 minutes (OZL-6A, ONEZILI, Guangzhou, China), and re-freezing. Once thawed the third time, we added anatoxin/saxitoxin diluent at a ratio of 10:1 to the DIDO water, as suggested by the cyanotoxin kit manufacturer (Eurofins Abraxis). After adding the diluent, the centrifuge tubes were centrifugated at 5,900 x g for 20 minutes (Sorvall Legend XFR Centrifuge, Thermo Fisher Scientific, Waltham, MA, USA). The supernatant was then poured into a sterile 60 mL syringe, and then two subsamples were pushed through a 0.45 μg glass syringe filter. One subsample of the supernatant was around 35-40 mL, which was used for T&O compound analysis. Approximately 1.5 to 1.8 mL were subsampled into a 2 mL glass vial for cyanotoxin analysis. The T&O vials were refrigerated until they were transported to the WaterOne water quality laboratory for analysis, typically within 48 hours of extraction. The cyanotoxin vials were frozen until analysis.

Cyanotoxin analysis was done using enzyme-linked immunosorbent assays for microcystin-ADDA, anatoxin-a, and saxitoxin (ELISA (96-test), Eurofins Abraxis, Warminster, PA, USA). The ELISA kits have all of the necessary components to perform the analysis following the EPA 546 method for microcystin analysis (USEPA, 2016), and manufacturer instructions were followed for all analyses (Figure 6). Individual samples were distributed in the wells of the

microplate, with duplicates for every sample. Microplates were read on a microplate reader (Eon Microplate Spectrophotometer, BioTek, Santa Clara, California, USA). within the recommended timeframe. Toxin results are fitted to the standard curve determined by the kit's standard solutions, and any toxin samples resulting in concentrations outside of the standards' range are diluted and rerun. The ELISA procedure was run in batches, with samples being frozen between sampling runs until the batched ELISA analysis was performed. The ELISA procedure was run twice with the first 4 sampling runs distributed in the first tray and the last 4 sampling runs distributed in the second tray. Resulting concentrations are calculated to μg/L. However, all cyanotoxin results were normalized in three ways for future analysis: microgram of toxin per gram dry-weight of freeze-dried subsampled algae (μg/g), microgram of toxin per gram dry-weight of freeze-dried subsampled algae multiplied by percent carbon of algae sample (μg/gC), and microgram of toxin per area of scrape sample from cobble (μg/cm^2).

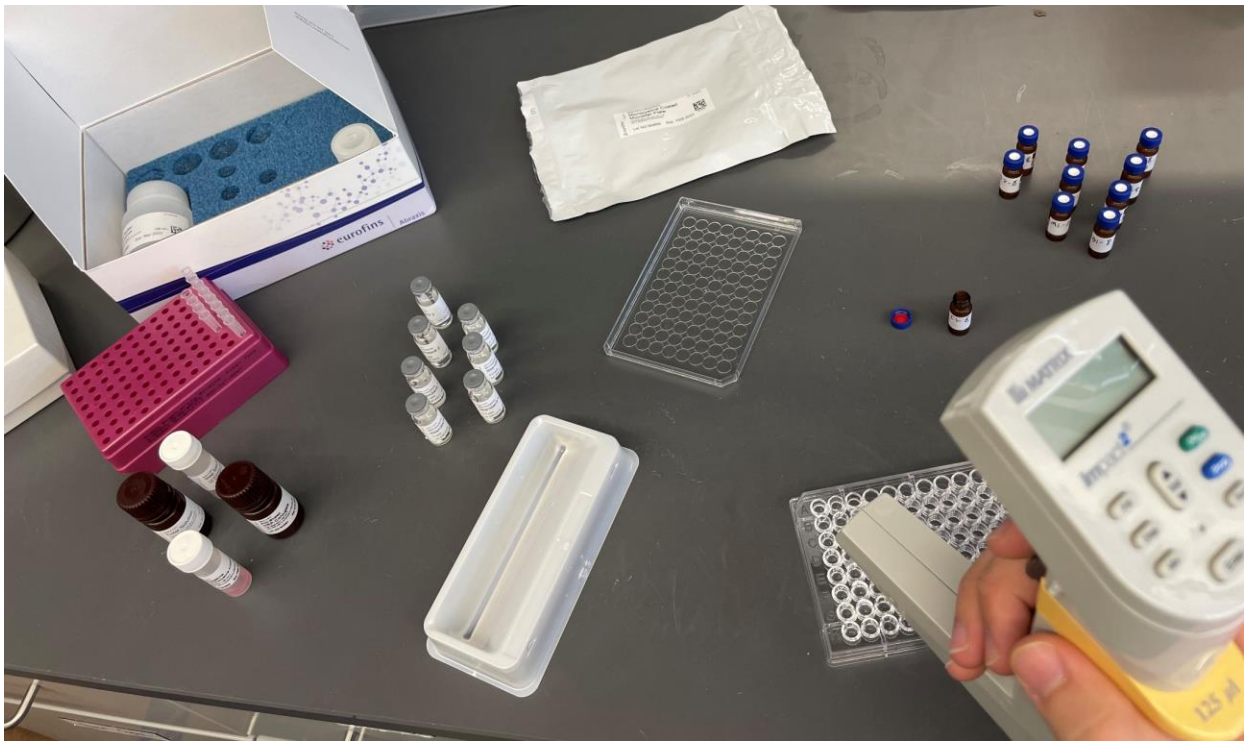

Figure 6: In process image of ELISA processing using Abraxis kits for Microcystin analysis in samples.

T&O compounds, geosmin and MIB, were analyzed using solid phase microextraction and GC-MS following a process based on the SM 6040D method at the WaterOne laboratory (Standard Methods Committee of the American Public Health Association, 2020). The detection limits for both T&O compounds are <1.0 ng/L. T&O results were normalized in similar ways as the cyanotoxins except the numerator is in nanograms (ng) rather than micrograms (μg). Sample run 6 geosmin and MIB data had to be scrapped due to machine issues. With the amount of dilution needed, sample run 6 results for geosmin and MIB had the potential to be similar to or higher than sample run 7 and 8 results. With this in mind, geosmin and MIB mean values presented may be conservative to the potential mean values had the sample run 6 processed fully.

In the results section, I will limit most result listings to just gram dry weight normalizations to reduce lengthiness of results. However, other normalizations will be discussed when their discrepancies from gram dry weight is worth mentioning, especially in the statistical modeling results.

3.3.5 Genomic Processing of Cyanobacteria, Cyanotoxin, and T&O Sequences

For genomic analysis, the 250 mg subsamples of each freeze-dried algae sample were delivered to the University of Kansas Environmental Engineering and Genomics Laboratory. Here DNA was extracted from each sample using Qiagen DNeasy PowerSoil Pro DNA Isolation kit (Qiagen, Hilden, Germany). Extraction was performed following manufacturers protocol on a QIAcube Connect automated extraction system (Qiagen, Hilden, Germany). Extracted DNA was quantified using the Qubit HS 1X Assay on a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA). Quantitative PCR (qPCR) was performed to quantify genes of Cyanobacteria, *Phormidium*, microcystin producing bacteria, saxitoxin producing bacteria, anatoxin producing bacteria, and geosmin producing bacteria. Target genes, sequences and annealing temperatures of

primers used to quantify these groups are presented in (Table 1). At the time of analysis, MIB gene primers for qPCR analysis were less well known and studied, and coupled with budget restraints, they were not analyzed in this study. qPCR reactions were prepared using Bio-Rad SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA) following manufacturer's protocol and a total reaction volume of 20 µL per sample. A dilution factor of 50 was used for all sample runs, with a final volume in each template measuring 5 µL. Thermal cycling was carried out on a Bio-Rad CFX Connect Real-Time PCR Detection System (Bio-Rad, Hercules, CA). Standard curves for each gene of interest were generated by serial dilution of IDT gBlocks (IDT, Coralville, IA) matching the PCR product of the respective primers. Data Analysis of the qPCR data was performed using Bio-Rad CFX Maestro version 2.1 (Bio-Rad, Hercules, CA). Resulting gene concentrations were output in gene copies per µL, and the three normalization techniques utilized by the cyanotoxin concentrations were used for the qPCR gene sequence concentration results as well.

Table 1: Gene primers for the target cyanotoxins, microcystin, anatoxin, and saxitoxin as well as gene primers for the T&O compound, geosmin, are listed here. Also, the gene primers for *Phormidium* and a more widespread cyanobacteria identifying primer were analyzed as well. The annealing temperature is displayed here, too.

Target Organism	Target Gene	Primer Name	Sequence (5`-3`)	Annealing Temp (°C)	Reference
Cyanobacteria	16s rRNA	CYAN 108F	ACG GGT GAG TAA CRC GTR A	60	(Rinta-Kanto et al., 2005)
		CYAN 377R	CCA TGG CGG AAA ATT CCC C		
Microcystin synthetase gene	mycE	mycE127F	AAG CAA ACT GCT CCC GGT ATC	58	(Sipari et al., 2010)
		mycE247R	CAA TGG GAG CAT AAC GAG TCA A		
Saxitoxin	sxtA	sxtF	GGA GTG GAT TTC AAC ACC AGA A	60	(Al-Tebrineh et al., 2012)
		sxtR	GTT TCC CAG ACT CGT TTC AGG		
Anatoxin-a synthetase gene	anaC	anaC-F	TCT GGT ATT CAG TCC CCT CTA T	58	(Rantala-Ylinen et al., 2011)
		anaC-R	CCC AAT AGC CTG TCA A		
Phormidium	16s rRNA	Phor-580	GCG AAA GGG ATT AGA TAC CC	56	(Marquardt and Palinska, 2007)
		Phor-710	CCG TCA ATT CCT TTG AGT TTC		
Geosmin A synthetase gene	geoA	geo_cya543F	ATC GAA TAC ATY GAR ATG CG	55	(Auffret et al., 2011; Kutovaya and Watson, 2014)
		geo_cya728R	ACT TCT CTY TGR TAG GA		

3.4 Data and Statistical Analyses

3.4.1 Linear Regression Analysis

We conducted multiple linear regression of the potential for our explanatory variables to describe microcystin concentrations in benthic algae mat samples. We developed three suites of explanatory variables to test the data requirements for adequate characterization. These three suites of variables include (1) physicochemical-only, referred to as ***Physicochemical***, (2) biological-only, referred to as ***Biological***, and (3) both physicochemical and biological, referred to as ***All***. For each suite of variables, we decided to include predictor variables such that the number of non-detects or missing data points were limited within any given regression. Variables were abbreviated in some cases in order to fit figures and tables better (Table 2).

Table 2: Breakdown of model predictor variables used, the units they are in, and the abbreviations used in figure and table outputs from the model.

Model	Variable	Unit	Abbreviation
Physicochemical	Temperature	°C	Temp.
	pH	-	pH
	Specific conductivity	mS/cm	Sp. C.
	Turbidity	NTU	Turb.
	Dissolved oxygen	mg/L	DO
	Nitrate	mg/L	NO_3^-
	Orthophosphate	mg/L	PO_4^{-2}
	Total suspended solids	mg/L	TSS
	Dissolved organic carbon	mg/L	DOC
Biological	Concentration of cyanobacteria as measured by BenthoTorch	µg/cm²	BT_{bga}
	Percent composition of cyanobacteria to total chl-*a* as measured by BenthoTorch	%	BT_{bga}^{frac}
	Concentration of cyanobacteria 16S sequence as measured by qPCR	copies/g; copies/gC; copies/cm²	$16S_{bga}$
	Concentration of *Phormidium* 16S sequence as measured by qPCR	copies/g; copies/gC; copies/cm²	*Phor*.
	Concentration of toxin/T&O gene sequence as measured by qPCR or Concentration of toxin/T&O as measured by ELISA/GC-MS (if target variable is sequence concentration from qPCR)	copies/g; copies/gC; copies/cm² µg/g; µg/gC; µg/cm² (or ng numerator for T&O)	$qPCR_{MC\ (or\ AA,\ SX,\ GE)}$ MC (or AA, SX, GE)
All (including all rows above except for TSS)	Percent carbon	%	%-C
	Percent nitrogen	%	%-N
	Mean mat thickness	mm	Mat
	Mean light penetration	%	Light
	Percent land cover by urban land	%	%-Urban
	Percent land cover by cropland	%	%-Crop
	Percent land cover by forest	%	%-Forest
	Percent land cover by impervious surfaces	%	%-Imperv

For the ***Physicochemical*** regression, temperature, pH, conductivity, turbidity, DO, nitrate, orthophosphate, TSS, and DOC were all selected (Table 2). Because TN and TP were only run for half of the samples, half of the samples have missing values for TN and TP, thus the model produced errors and ultimately TN and TP were not selected for regression. For the ***Biological*** regression, we used four outputs from the BenthoTorch measurements, chl-*a* concentration, qPCR gene counts for Cyano 16S, *Phormidium* 16S, and the corresponding toxin's or geosmin's qPCR-based gene sequence concentration, depending on which is being modeled (Table 2). For example, if microcystin concentration was the target variable, the qPCR results for mycE gene counts were

used as a ***Biological*** regression predictor variable along with the others listed. For the ***Biological*** linear regression models for cyanobacteria concentrations, only the cyanobacteria 16S and *Phormidium* 16S concentrations results were used as predictor variables. The gene sequence values are normalized the same three ways as the cyanotoxin and T&O compounds, and so each time the target variable's normalization was changed, so too was the gene sequence normalization. The BenthoTorch data include the cyanobacteria concentration (ug/cm^2) and percent cyanobacteria concentration to total algae concentration (%). The last type of regression, ***All***, includes both ***Physicochemical*** and ***Biological*** variables as well as a few others, including isotopic results, land-use results, and two other physicochemical variables were used. These include $\delta^{13}C$, $\delta^{15}N$, %C, %N, mat average thickness, percent light penetration, percent urban land use, percent cropland, percent forested land, and percent impervious surface (Table 2).

The target variables the linear regression was run for include microcystin concentrations, mycE genomic cell count concentrations, microcystin toxin quota, and cyanobacteria concentrations and % compositions. Linear regression models for anatoxin-a, saxitoxin, geosmin, and MIB were not able to be ran due to too many missing data points in either the toxin or T&O concentration field or the gene sequence concentration field. Microcystin concentrations and gene count concentrations were all run for the 3 normalization techniques used: μg/g, μg/gC, and μg/cm^2. Cyanobacteria target variables include cyanobacteria concentration from the BenthoTorch, percent composition of cyanobacteria calculated from cyanobacteria concentration from the BenthoTorch, and percent composition of cyanobacteria calculated from cyanobacteria density from the taxonomic analysis. Finally, microcystin toxin quota was only modeled for μg per cell copy (μg/copy) as changing the normalization for both the numerator and denominator at the same time result in the same quota value. Each linear regression model that was ran also had

ANOVA (analysis of variance) run on the resulting model to learn the most impactful variables in each regression. The regression fitting and ANOVA analyses were performed using MATLAB 2021a.

3.4.2 Random Forest Analyses

Regression tree analysis was performed on the same group of predictor and target variables as in the linear regression analysis. Also, because missing data does not impact random forest modeling like it does linear regression, anatoxin-a, saxitoxin, geosmin, and MIB models were able to be developed and evaluated, including gene sequence and quota models for anatoxin-a (anaC sequence) and saxitoxin (sxtA sequence). Regression trees are particularly suited for complex ecological data, where non-linear relationships, high-order interactions, and missing variables may be present (De'Ath and Fabricius, 2000). Regression trees work by explaining the variance in a target variable by splitting the data into groups using a combination of categorical and/or numeric predictor values. Individual regression trees can be plotted graphically, which aids in an intuitive interpretation of results. A single regression tree can be overfit to the present data; thus, many trees can be constructed creating a 'boosted ensemble' or 'random forest' of predictions (Wang et al., 2021). This random forest output takes into consideration the uncertainty associated with the data and regression tree branching.

Both single tree and random forest regressions were run using built in functions in MATLAB, 'fitrtree' and 'TreeBagger' respectively. Single tree regressions have 3 branches to reduce opportunity of overfitting. For the random forest regression, 2000 trees were run, also with only 3 branches. Shapley values, which are model-agnostic (i.e., they are unaffected by underlying model structure) and describe the most influential predictor variables in each regression tree (Wang et al., 2021), were identified for both the single tree regression and random forest model for each

run. Shapely values were chosen only from the 30 best performing models within the 2000 total trees run for the random forest. Linear regressions were then run with the single tree and random forest model outputs to find the R^2 performance of each model type. The model performance of the random forest regression was then compared to the observed data and 95% prediction bounds were estimated to see where along the observed value range the model was strongest or weakest.

Chapter 4: Results

As listed in the site characteristics section, Indian Creek had the most benthic algae mats identified over the survey periods, at 47 total sites. Mill Creek had 31 sites identified, and Blue River had 5 sites identified. While individual mat proliferation area was not measured, this equates to the most urban stream having the greatest spread of benthic algae mats and the mixed stream having slightly fewer mats. However, both of those streams have substantially more identified algae mats than the most rural stream.

4.1 Physicochemical Drivers of Benthic CyanoHABs

4.1.1 In Situ Sensor Results

Water temperature showed little spatial variability between our nine sampling sites (Figure 7), but we observed considerable temporal variability from the early-summer to mid-fall. The first three sampling runs (of eight total) were during the summer period and experienced an average temperature of 27.5°C across all sites. As fall began, the mean temperature of the next three sampling runs dropped to 22.1°C. The mean temperature dropped further to 15.4°C in the 7th sampling run and 11.5°C in the final sampling run. Water temperature peaked at 34.1°C in site M2 during the first sampling run and reached a low of 9.6°C in site I1 during the last sampling run. The average pH readings across for the watersheds ranged from 7.76 to 8.2 (Figure 7), with individual readings ranging from 7.52 to 8.34 across all samples. Specific conductivity varied considerably between sites, from 0.30 mS/cm to 1.35 mS/cm (Figure 7), with the rural sites (B1, B2, and B3) having relatively low concentration of ions, and the more-urban sites (M1, M2, M3, I1, I2, and I3) having a higher specific conductivity, like as a result from a greater concentration of salts due to wastewater effluent. Mean turbidity varied from 5.28 NTU to 43.3 NTU, with the

rural sites having significantly higher average turbidity (Figure 7), except for site I3, which was directly downstream of the largest wastewater treatment plant within the study site area (Figure 1).

The average DO for all sampling runs was 9.09 mg/L (Figure 7) with region-wide mean low DO levels of 6.18 mg/L and 5.42 mg/L, observed during the sampling runs 5 and 6, respectively. Increased phytoplankton productivity in the stream water column might suggest the drop in DO during the two aforementioned sampling runs, as the water was visibly discolored during these sampling runs (brown: Figure 2A; clear: Figure 2E both are the same site, just looking different directions). Light penetration ranged from no overhead shading vegetation (100% light penetration) to only 10% of the sky being open (10% light penetration) (Figure 7). The last sampling run occurred after most leaves had fallen, and light penetration increased by 20%. Benthic algae mat thickness averaged around 2.86 mm, except for sampling run 2, which occurred not long after a significant storm event and had an averaged mat thickness of 0.84 mm (Figure 7).

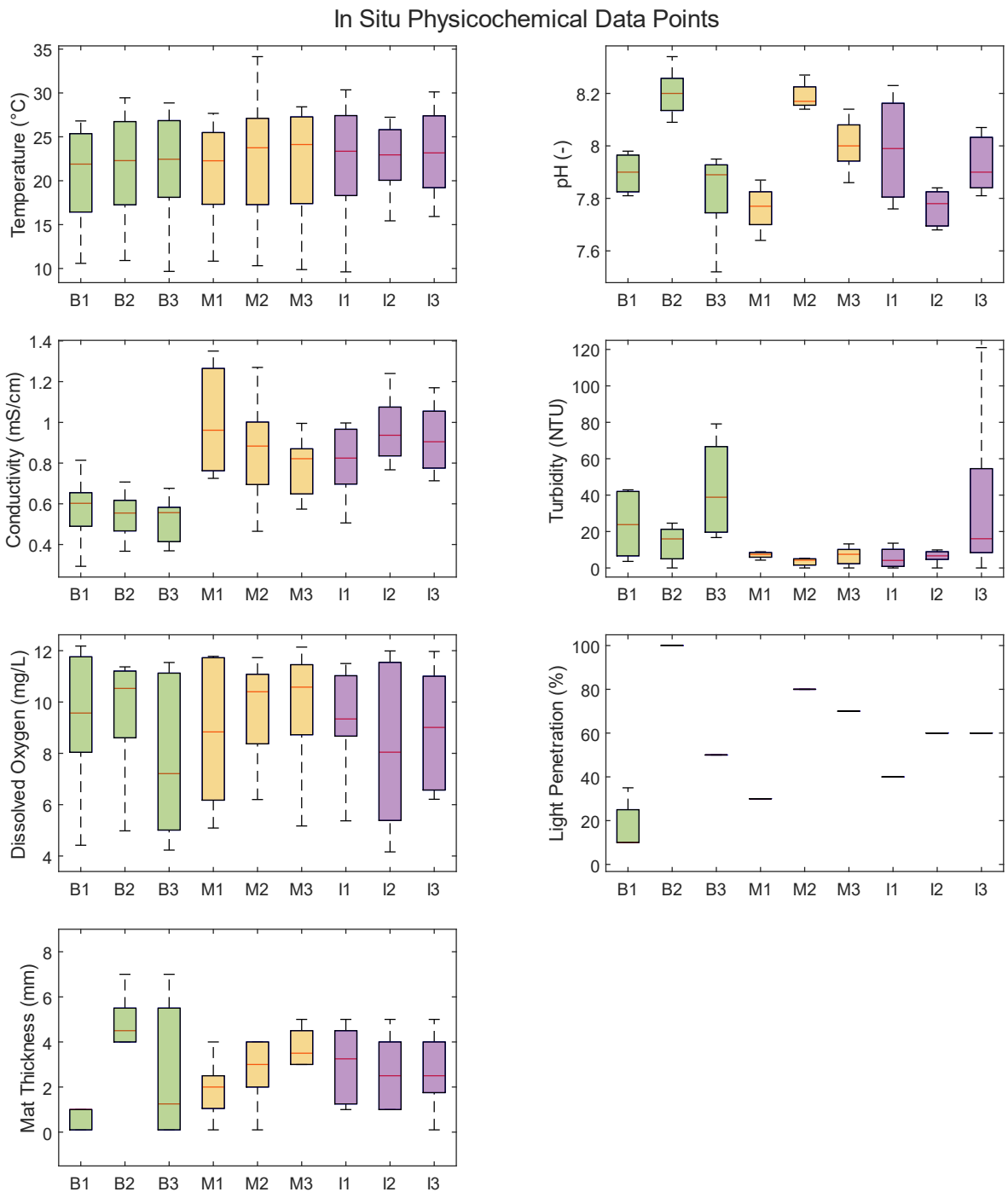

Figure 7: *In situ* physicochemical data box plots for each study site. This includes the *in situ* sensor data points from the multiparameter water quality device, the mean periphyton biomass thickness, and the amount of light penetration. Green box plots indicate Blue River sites, yellow box plots indicate Mill Creek sites, and purple box plots indicate Indian Creek. Note different y-axes.

4.1.2 Physicochemical and Nutrient data results

Dissolved organic carbon readings averaged from 5.61 ppm to 9.72 ppm across the sampling runs (Figure 8). Duplicate DOC samples (n = 4) had an error of 3.2%. The average total suspended sediment ranged from 5.05 to 13.5 mg/L (Figure 8). Duplicate TSS samples (n = 4) showed an error of 63%; however, this error is somewhat exaggerated as mean TSS magnitudes were small (< 4 mg/L) at half of the sites, such that minor changes of 1 or 2 mg/L disproportionately impact percent difference between duplicates without having much physical impact to how the data can be interpreted.

Of the 72 ammonia samples, only 12 recorded a detection, with the average of the 12 samples being 0.315 mg/L. Nitrate recorded much more detections, with a mean range of 0.39 mg/L to 10.7 mg/L (Figure 8). The rural stream regularly recorded significantly lower nitrate than the other two more urban streams ($p < 0.05$). The sample run mean range for orthophosphate ranged from 0.055 mg/L to 0.85 mg/L (Figure 8). A significant difference exists between orthophosphate in the most rural and the most urban watersheds ($p < 0.05$), but not in the intermediate-urban watershed. The mean error from duplicates (n = 4) measured at the Johnson County Water Quality Lab was 4.82%. Total nitrogen and total phosphorus results were similar to the nitrate and orthophosphate results (Figure 8), but we only collected this data on four sample runs (out of eight), thus their interpretability is more limited. However, all of the nutrient results suggest that the benthic nutrient availability is not a limiting factor, as nutrient values in the samples across all watersheds are all well above levels found for at sites of toxic benthic *Phormidium* growth (McAllister et al., 2018). In fact, the high nutrients may pose a negative effect known as the subsidy-stress concept, with possible levels indicating the start of this concept as DIN > 0.2 mg/L and DRP > 0.0014 mg/L (McAllister et al., 2018).

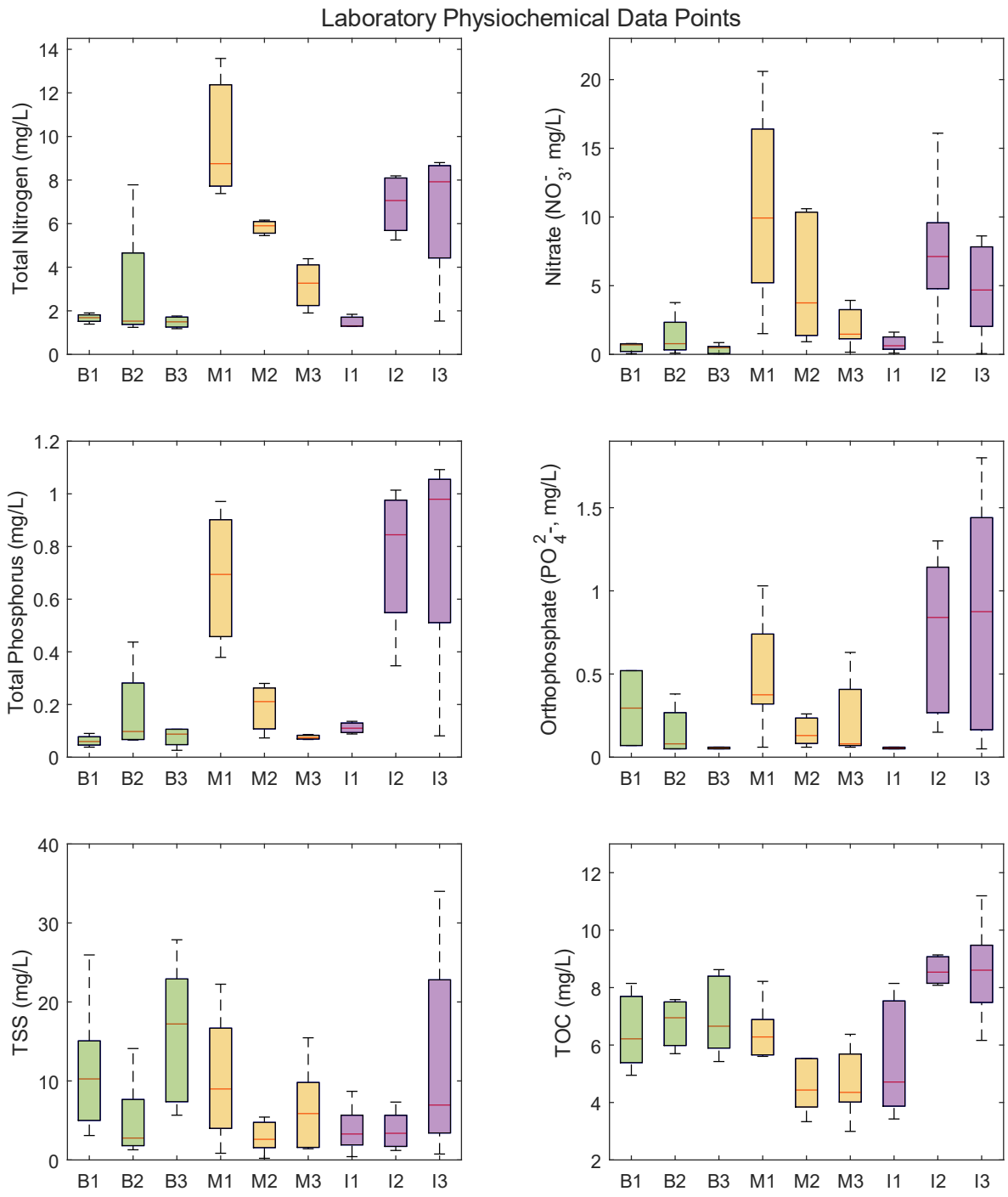

Figure 8: Nutrient and physicochemical data box plots for each study site gathered from post sampling laboratory processes. Green box plots indicate Blue River sites, yellow box plots indicate Mill Creek sites, and purple box plots indicate Indian Creek. Note different y-axes.

4.2 Cyanotoxins and Taste-and-Odor Compounds

4.2.1 BenthoTorch results

In all three streams, BenthoTorch readings suggest that diatoms were the most common periphyton measured and green algae were the least common (Figure 9) and that cyanobacteria were present at all locations. Benthic cyanobacteria concentrations measured from the BenthoTorch ranged from 0.318 μg/cm^2 to 3.20 μg/cm^2 (Figure 9). Mean cyanobacteria concentrations did not significantly change throughout the sampling period, with a mean cyanobacteria concentration of 1.44 μg/cm^2. Watershed variations in cyanobacteria concentrations were also not significantly different, except for site B1, which had a mean cyanobacteria concentration of 0.688 μg/cm^2 across all sampling runs, compared to 1.54 μg/cm^2 at the other sites. Mean cyanobacteria percent composition, as a ratio of total algae measured by the BenthoTorch, was 38.1% (Figure 9), with diatoms usually making up the majority and green algae making up less than cyanobacteria. Duplicate field BenthoTorch measurements (n = 60) were not significantly different ($p = 0.47$). The coefficient of variation for BenthoTorch readings per site ranged from 20% to 115%, with most coefficients of variation being above 30%. The considerable variation in measurements may be attributable to the 15 different cobbles that are used to determine a composite site-average BenthoTorch composition. Each of the 15 cobbles were selected without knowledge of their composition, thus some may be dominated by diatoms while other are dominated by cyanobacteria.

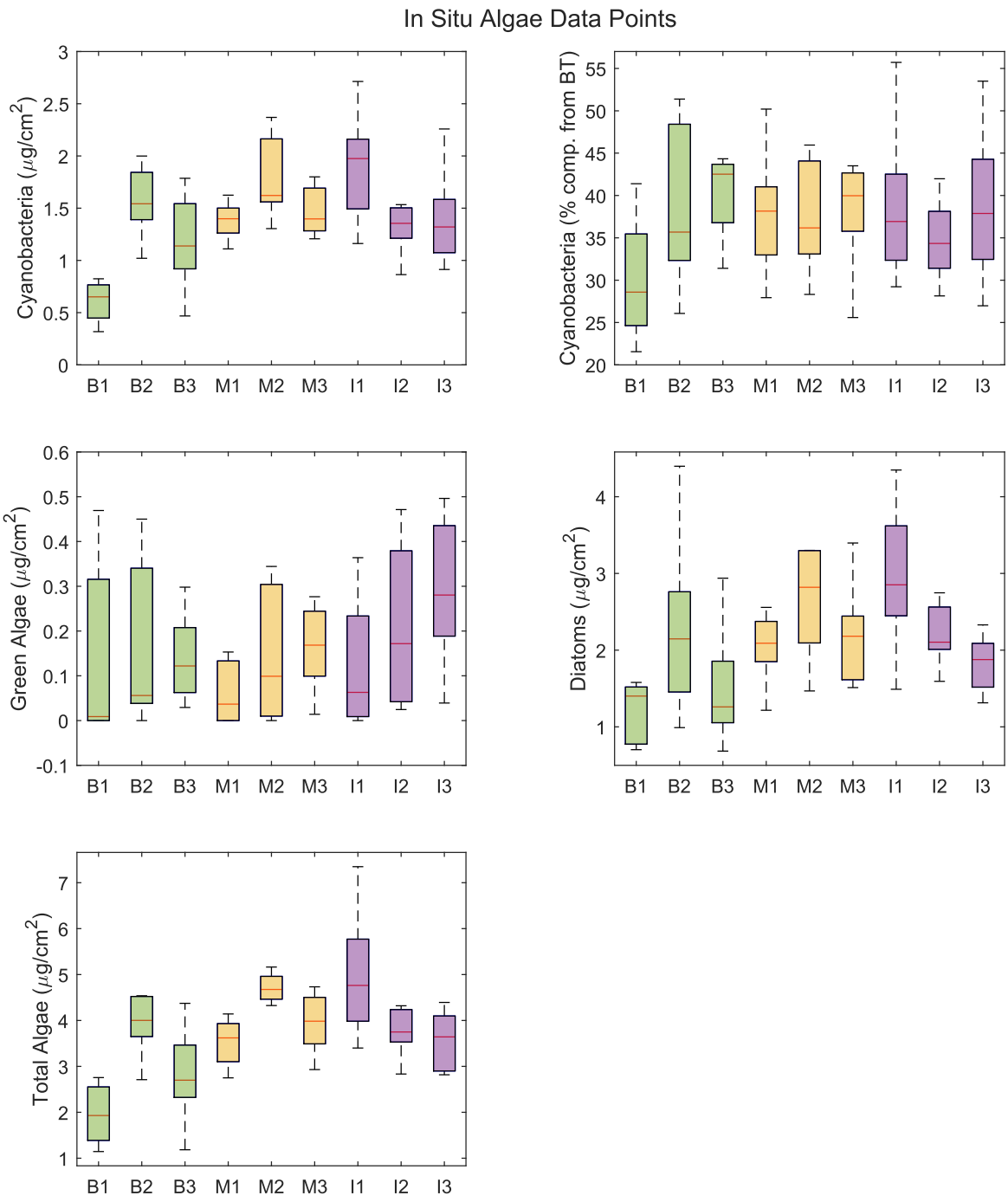

Figure 9: *In situ* algae data box plots for each study site, specifically from the BenthoTorch. Green box plots indicate Blue River sites, yellow box plots indicate Mill Creek sites, and purple box plots indicate Indian Creek. Note different y-axes. The top right plot is the calculated percent composition of cyanobacteria in the sampled periphyton based on the measured cyanobacteria and total algae concentrations

4.2.2 Benthic biomass results

Unfortunately, chl-*a* laboratory results were not fully processed in time for inclusion in this study. The mean chl-*a* measurement from the BenthoTorch was 3.82 μg/cm^2. When comparing individual site sample results for both BenthoTorch total chl-*a* and the laboratory chl-*a* measurements, a linear relationship is usually seen with R^2 values around 0.47 to 0.85 (Harris and Graham, 2015; Rosero-López et al., 2021). However, preliminary results of our chl-*a* samples were 2 orders of magnitude higher than our BenthoTorch results, which is not what other studies have found, so there seems to be an error in either the storage or processing of the chl-*a* samples in this study. Therefore, the chl-*a* data, even the preliminary data, will not be in this study.

The mean carbon and nitrogen percentages of the algae samples were 10.1% and 0.91%, respectively, while the carbon isotope ($\delta^{13}C$) and nitrogen isotope ($\delta^{15}N$) means were -22.7‰ and 8.66‰, respectively (Figure 10). The $\delta^{15}N$ values in our study are more than 1 standard deviation greater than the mean $\delta^{15}N$ value reported on a global dataset of benthic algae values (Ishikawa et al., 2018). On the other hand, the $\delta^{13}C$ values are within 1 standard deviation of the same global dataset (Ishikawa et al., 2018). These variations could be attributable to nutrient availability, as the mean nitrate concentration in the study areas from the Ishikawa study was 0.045 mg/L, much lower than our study's nitrate range.

The taxonomic identification and enumeration analysis indicated that diatoms were the dominant periphyton group, broadly agreeing with BenthoTorch results (Figure 11). The mean percent cyanobacteria composition (11.6%) (Figure 10) was relatively low in taxonomic microscopy results compared to the BenthoTorch results (38.1%) (Figure 9). *Phormidium* percent composition was only 3.4% (Figure 10) and was detected in 46% of the samples that had cyanobacteria in them. We found several notable cyanobacteria that have been shown to produce

the target cyanotoxins and T&O compounds in the taxonomic results (Wood et al., 2020), including *Phormidium animale*, *Leptolyngbya* sp., and *Chroococcus* sp.

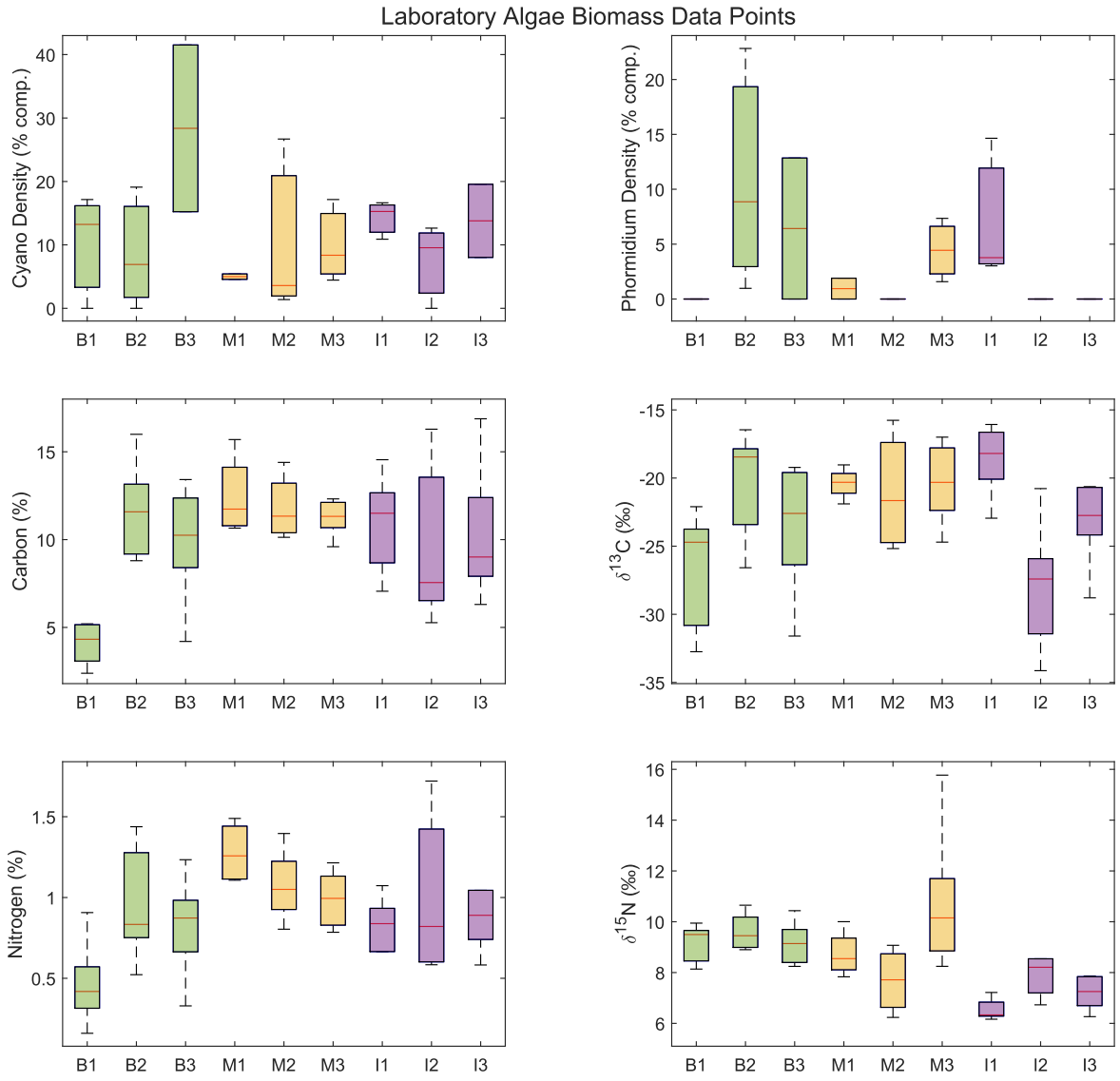

Figure 10: Laboratory analyses on benthic biomass sample box plots for each study site. The top two plots are from taxonomic analysis, and the data is converted into a percent composition amount. The bottom 4 plots are from the isotopic analysis. Green box plots indicate Blue River sites, yellow box plots indicate Mill Creek sites, and purple box plots indicate Indian Creek. Note different y-axes.

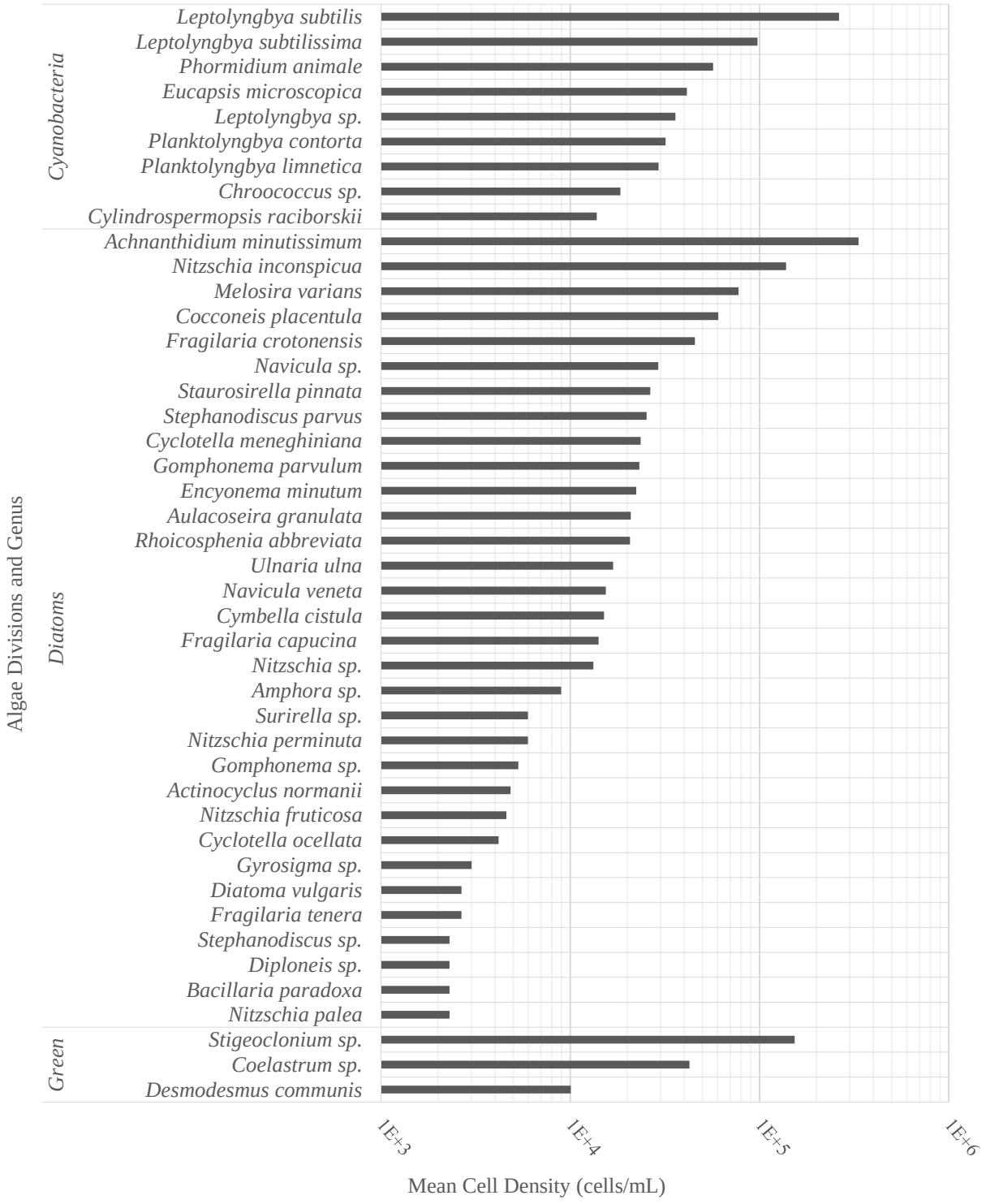

Figure 11: Taxonomic microscopy cell density results from 24 taxonomic samples spread across all sites plotted for each genus identified. In the cyanobacteria group, *Cylindrospermopsis* was the genus listed to us by the laboratory that did the taxonomic analysis, but this genus is also referred to as *Raphidiopsis*. Note the logarithmic scale on the x-axis.

4.2.3 Cyanotoxin and T&O Concentrations

Microcystin was present in every sample regardless of watershed land-use (Figure 12). Microcystin amounts increased slightly in the cooler weather with the greatest sample run mean occurring on the last sample run. Site I2 had a considerably higher mean microcystin (0.217 μg/g) than the other sites (0.063 μg/g). Thirty-five percent of samples registered non-detect for anatoxin-a with the remaining samples having a mean of 0.060 μg/g (Figure 12). Saxitoxin had a greater fraction of non-detects (74%) with the detectable samples having a mean of 0.0041 μg/g (Figure 12). Over the course of the sampling run, anatoxin-a and saxitoxin appeared to have a relative handoff in production with more anatoxin-a being produced after late September and early October and saxitoxin decreasing in relative abundance at the same time (Figure 13). Similar to microcystin, the T&O compound geosmin was present in many samples across all watersheds. Mean geosmin was 4.14 ng/g across all results (Figure 12). MIB was present in higher amounts (8.03 ng/g) than geosmin when detected but had around 16% non-detects (Figure 12). Notably, the last and coldest sampling run in November had a considerable jump in MIB concentration (mean of 41.1 ng/g).

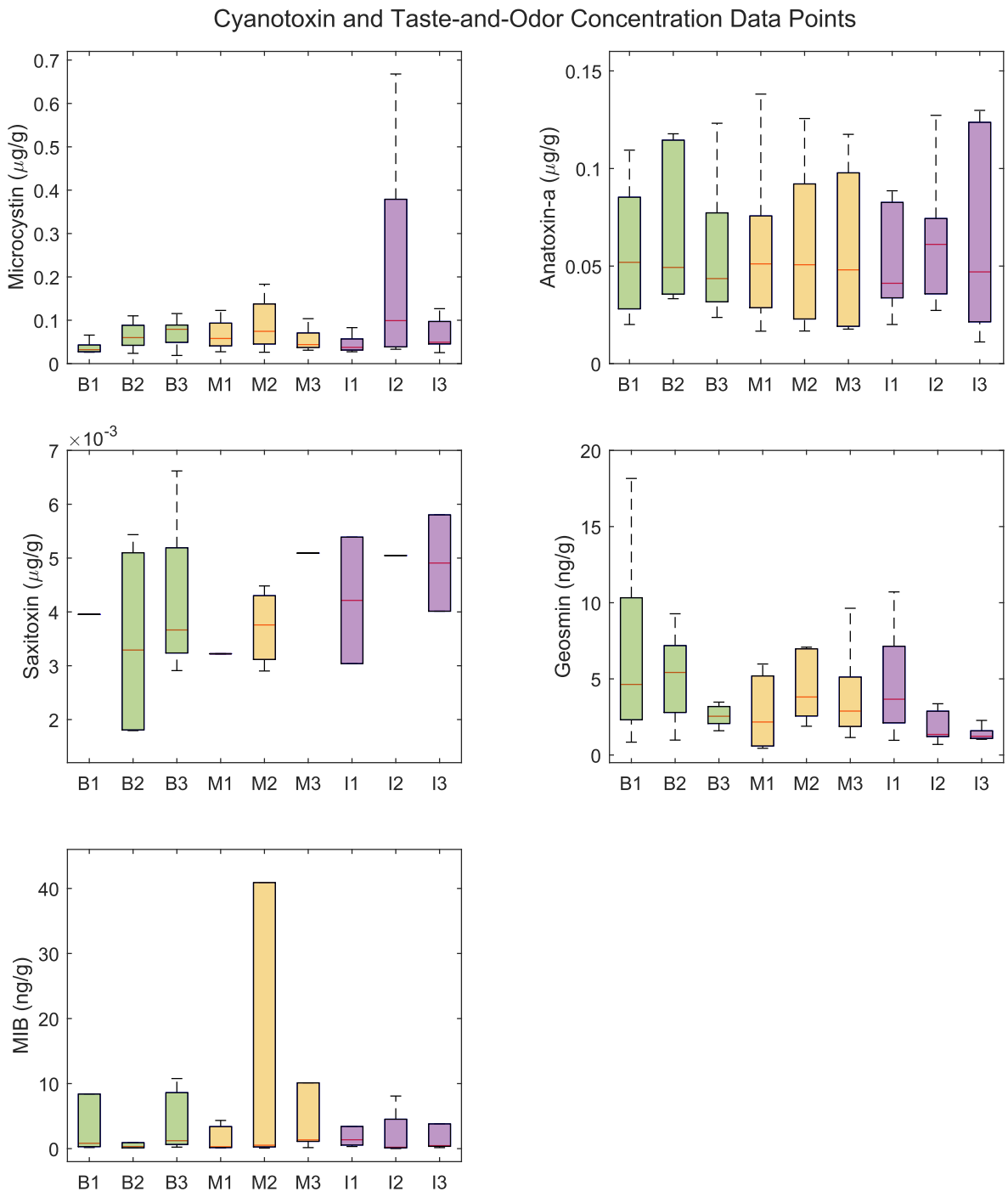

Figure 12: Cyanotoxin and T&O concentration data box plots for each study site from ELISA results. All results are normalized to dry weight of original algae scrape sample. Green box plots indicate Blue River sites, yellow box plots indicate Mill Creek sites, and purple box plots indicate Indian Creek. Note different y-axes.

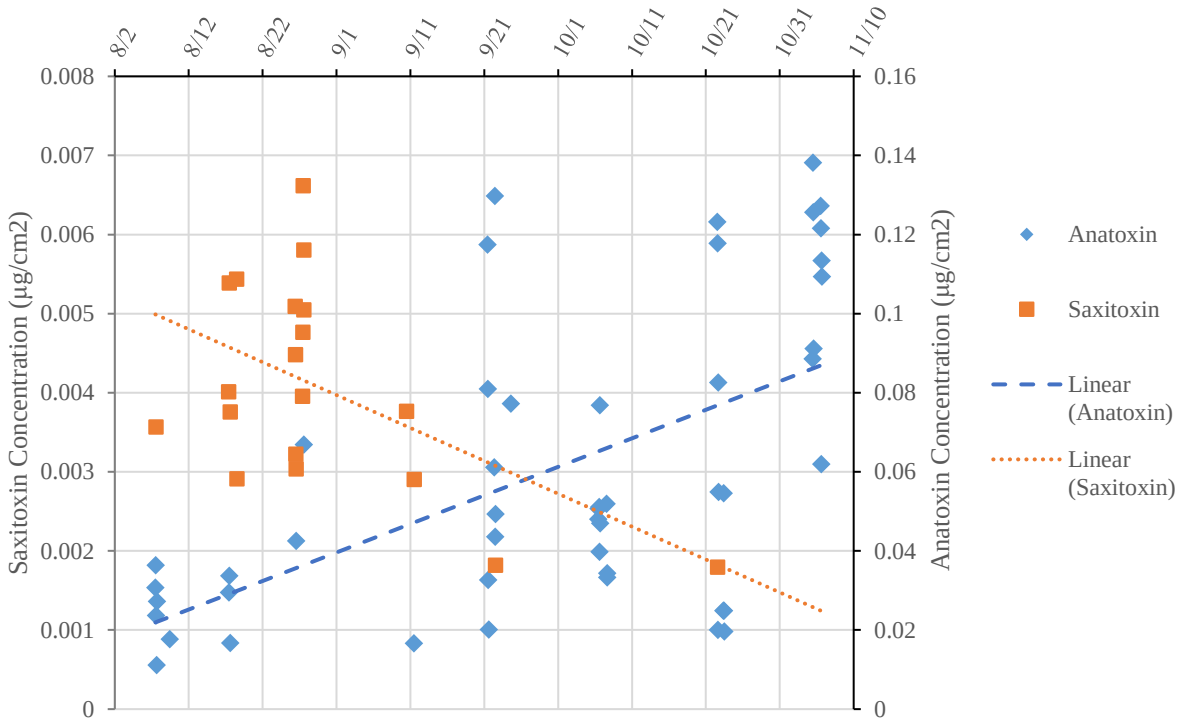

Figure 13: All detected Anatoxin-a and saxitoxin concentrations over time plotted on the same figure with separate y-axes to show relative abundance. All dates shown are in 2021 when the sampling period took place. The changeover appears to occur in mid to late September 2021, with more saxitoxin present in dates earlier than Sept 21st, and more anatoxin-a present around the same date.

4.2.4 Genomic Cyanobacteria, Cyanotoxin, and T&O results

Gene sequences for Cyano 16S and *Phormidium* 16S were detected in all samples (Figure 14). Both results were normalized to the dry weight algae sample mass. Mean Cyano 16S concentrations were 8.44×10^8 gene copies/g. Mean *Phormidium* 16S concentrations were 3.00×10^9 copies/g. Both 16S gene primers were detected in greater concentrations in the rural and mixed streams than the urban stream. As mentioned earlier, there is occasional disagreement in whether cyanobacteria and/or *Phormidium* are present in the samples, as some taxonomic results indicate no presence whereas the BenthoTorch and gene sequence data indicate the presence of cyanobacteria. This could arise due to potential sampling bias, as the sampled quantity of algal mat for taxonomic analysis was typically less than 15% of total algal mass and was scraped from

a spot next to the main sample scrape area, thus may not be entirely representative of the composite sample used in genomic analysis.

Like with the microcystin concentrations gathered from the ELISA analysis, the gene primer mycE was detected in every sample (Figure 14). The gene sequence concentration results had a wide range of detections, from 1.47×10^0 copies/g to 1.21×10^6 copies/g. The mean microcystin gene concentration was 28,700 copies/g. The mean concentration for the anatoxin-a gene, anaC, chosen in this study was 16,200 copies/g (Figure 14). Primer detection for the anatoxin-a gene had 63% non-detects. The saxitoxin gene, with a mean concentration of 7890 copies/g, had zero non-detects for its target gene (Figure 14), despite 74% non-detect in the toxin analysis. Finally, the geosmin gene, with a mean of 282,000 copies/g, had the fewest detections, with 89% of all samples being a non-detect (Figure 14), despite 100% detection in the toxin analysis.

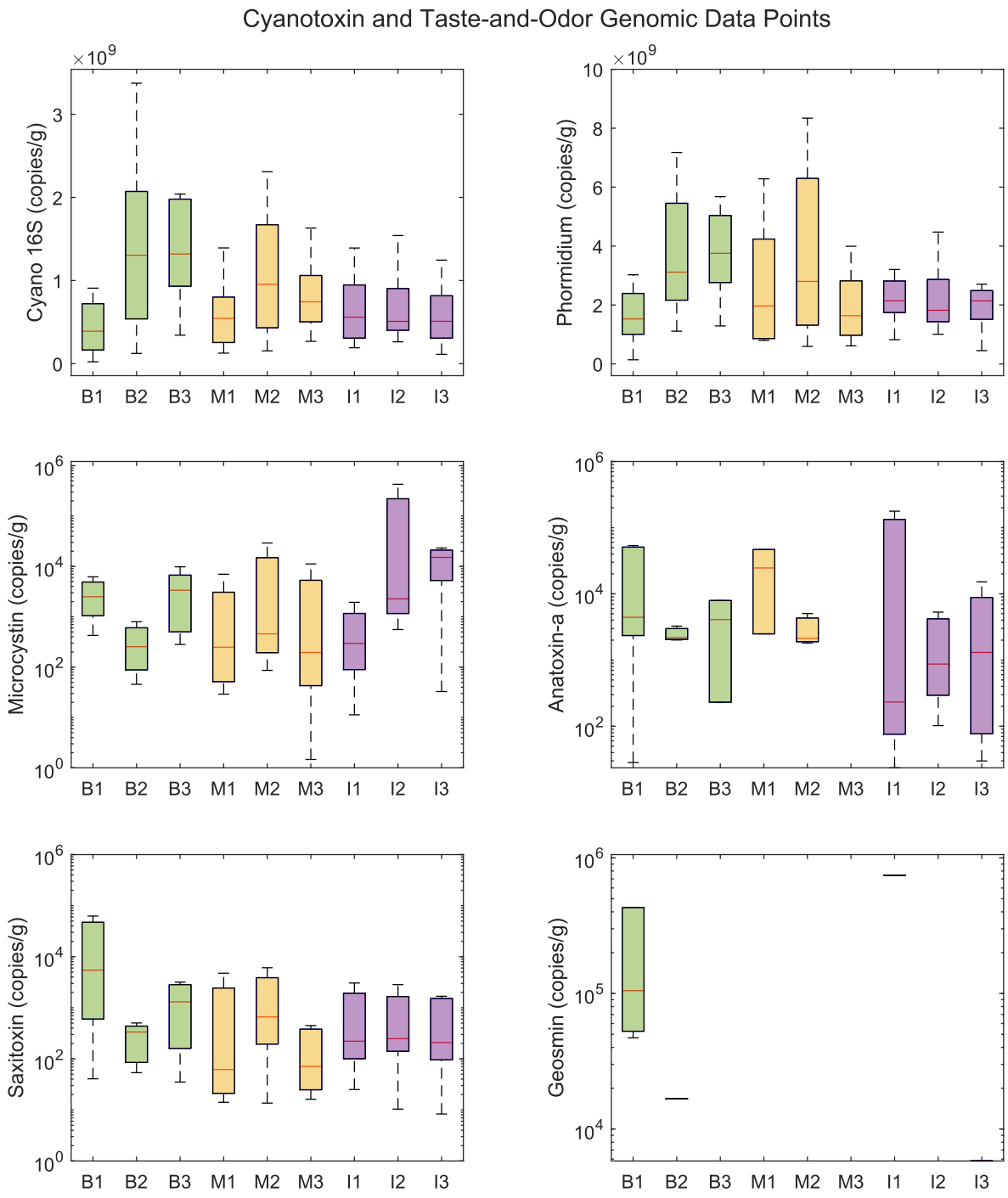

Figure 14: Cyanotoxin and T&O genomic data box plots for each study site from qPCR analysis. All results are normalized to dry weight of original algae scrape sample. Green box plots indicate Blue River sites, yellow box plots indicate Mill Creek sites, and purple box plots indicate Indian Creek. Note different y-axes, and the bottom 4 plots have a logarithmic y-axis.

The toxin quota is calculated simply by dividing the measured toxin or T&O compound concentration gathered by the ELISA/GC-MS process by the concentration of the equivalent gene sequence concentration results from to get a unit mass per gene copy quota, barring both are normalized the same way (Thomson-Laing et a;., 2020). The higher quota means a greater amount of relative toxin or T&O compound production by the cyanobacteria within a sample. The mean microcystin toxin quota was 6.46×10^{-4} µg/copy (Figure 15). Anatoxin-a's toxin quota mean was 4.13×10^{-4} (Figure 15). Saxitoxin's toxin quota mean was 5.71×10^{-5} µg/copy (Figure 15). Finally, for the T&O compounds, MIB's quota couldn't be calculated as no qPCR genetic analysis was performed, but for geosmin's quota, the mean was 7.55×10^{-5} ng/copy (Figure 15).

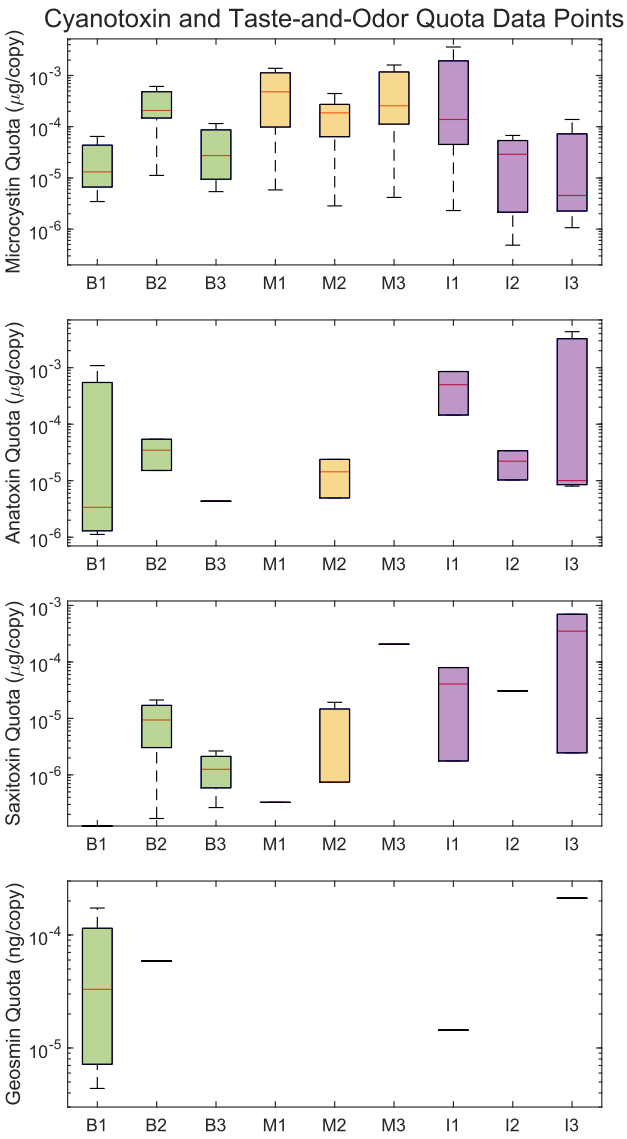

Figure 15: Cyanotoxin and T&O quota data box plots for each study site. Green box plots indicate Blue River sites, yellow box plots indicate Mill Creek sites, and purple box plots indicate Indian Creek. Note different y-axes and the logarithmic y-axes.

4.3 Data and Statistical Analyses

4.3.1 Linear Regression Analysis

Linear regression analyses were run for microcystin concentrations (Table 3), Cyanobacteria concentration and composition (Table 4), mycE gene concentrations (Table 5), and

microcystin toxin quota (Table 5). As stated above, the reason the same statistical analysis wasn't performed on anatoxin-a, saxitoxin, geosmin, and MIB was due to the amount of non-detects and their impact on the regression reliability. Non-detects reduce the size of the predictor dataset with too small of a dataset resulting in either overfitting or lack of statistical robustness. Linear regressions for microcystin only worked for the ***Physicochemical*** and ***Biological*** groups. The group with ***All*** predictor variables had too many missing rows due either to missing data or non-detects in the predictor variables. We tested ***Physicochemical*** versus ***Biological*** explanatory variables to assess which set of variables had greater explanatory strength. Further, we tested multiple normalization techniques to assess if the inferences gained from using one normalization (e.g., dry mass weight) versus another (e.g., scrape area) would change inferences regarding the drivers of cyanobacterial growth and toxin production.

The ***Physicochemical*** linear regression for microcystin (normalized by gram dry weight) had a R^2 of 0.33, and the ***Biological*** linear regression had a R^2 of 0.70. Using ANOVA, orthophosphate and TSS explained a similar amount of variance in the prediction, but not at very substantial amounts. For the ***Biological*** linear regression, the mycE gene counts had the greatest explanatory strength and resolved 64.5% of the variance in microcystin predictions (Table 6). This predictor variable and other predictors whose explanation of variance (or Shapley value percentage) are above 30% weight in their particular model can be seen in Table 6. When normalized by gram dry weight of carbon (a proxy for organic material within the mat and sediment matrix), the linear regression R^2 performance for the ***Physicochemical*** and ***Biological*** predictor variable groups improve to 0.39 and 0.77 respectively. This is driven by a higher variance explanation by both orthophosphate and TSS at over 15% for the ***Physicochemical*** group, and the ***Biological*** group saw a 74.9% explanation of the variance from the mycE gene. When normalized

by the scrape area, the ***Physicochemical*** R^2 improved again to 0.47, which appeared to have been improved due to the increase in explanatory power of specific conductivity. However, the ***Biological*** R^2 for the scrape area normalized microcystin concentration was the lowest of the three at 0.61, and mycE concentration explains 44.2% of the variance. Both the gene sequence results for cyanobacteria Cyano 16S and *Phormidium* 16S as well as the BenthoTorch results don't play a large role in the linear regression, with the greatest explanatory percent being the gene for Cyano 16S explaining only 1.89% of the variance in the gram dry weight normalization. This is an example of how other predictor variables may show up higher in the ranking of predictor variables but may only have a small explanation in the overall model.

Table 3: Microcystin concentrations as response variables from statistical models. Three normalizations are presented along with the three different variable classes the model was split into, and the results from both the linear and random forest models. These results are broken up between the coefficient of determinations for each model and the ranking of the predictor variables that either explained the most variance (ANOVA from linear regression models) or had the highest Shapley values (random forest models).

Response Variable (normalization)	Predictor Class	Model	R^2	Predictor Ranking
Microcystin (µg/g)	***Physicochemical***	linear	0.33	PO_4^{-2}, TSS, DO, Sp. C.
		random forest	0.72	Sp. C., NO_3^-, Temp., PO_4^{-2}
	Biological	linear	0.7	$qPCR_{MC}$, $16S_{bga}$, BT_{bga}, BT_{bga}^{frac}
		random forest	0.9	$qPCR_{MC}$, BT_{bga}, $16S_{bga}$, *Phor.*
	All	linear	N/A	N/A
		random forest	0.92	$qPCR_{MC}$, %-N, $16S_{bga}$, %-C
Microcystin (µg/gC)	***Physicochemical***	linear	0.39	PO_4^{-2}, TSS, pH, Sp. C.
		random forest	0.55	Turb., Sp. C., Temp., NO_3^-
	Biological	linear	0.77	$qPCR_{MC}$, $16S_{bga}$, BT_{bga}^{frac}, *Phor.*
		random forest	0.94	$qPCR_{MC}$, $16S_{bga}$, *Phor.*, BT_{bga}
	All	linear	N/A	N/A
		random forest	0.94	$qPCR_{MC}$, $16S_{bga}$, DO, Temp.
Microcystin (µg/cm^2)	***Physicochemical***	linear	0.47	TSS, PO_4^{-2}, Sp. C., Temp.
		random forest	0.44	Temp., DOC, TSS, pH
	Biological	linear	0.61	$qPCR_{MC}$, BT_{bga}^{frac}, BT_{bga}, *Phor.*
		random forest	0.75	$qPCR_{MC}$, *Phor.*, $16S_{bga}$, BT_{bga}^{frac}
	All	linear	N/A	N/A
		random forest	0.75	$qPCR_{MC}$, $16S_{bga}$, *Phor.*, BT_{bga}^{frac}

For the linear regression results of Cyanobacteria concentrations and percent composition from BenthoTorch results, the ***Physicochemical*** predictor variable group had R^2 results of 0.55 for BenthoTorch concentrations and 0.45 for BenthoTorch percent composition. For the BenthoTorch concentration linear regression model, pH resolved the greatest variation, explaining 39.3%. Due to missing data, the linear ***Physicochemical*** regression for the taxonomic percent composition isn't calculated. All of the ***Biological*** linear regressions for the cyanobacteria models show poor R^2 values (<0.2). In other words, Cyanobacteria 16S and *Phormidium* 16S gene concentration results from the qPCR process had little to no correlation with the chosen cyanobacteria target variables.

Table 4: Cyanobacteria concentrations and percent compositions from the BenthoTorch and Taxonomic analyses as response variables from statistical models. Three normalizations are presented along with the three different variable classes the model was split into, and the results from both the linear and random forest models. These results are broken up between the coefficient of determinations for each model and the ranking of the predictor variables that either explained the most variance (ANOVA from linear regression models) or had the highest Shapley values (random forest models).

Response Variable (normalization)	Predictor Class	Model	R^2	Predictor Ranking
Cyanobacteria ($\mu g/cm^2$)	***Physicochemical***	linear	0.55	pH, NO_3^-, Turb., TSS
		random forest	0.55	Turb., NO_3^-, TSS, Temp.
	Biological	linear	0.01	$16S_{bga}$, *Phor.*
		random forest	0.26	*Phor.*, $16S_{bga}$
	All	linear	N/A	N/A
		random forest	0.61	Turb, Light, %-C, %-Imperv
Cyanobacteria (% from BenthoTorch)	***Physicochemical***	linear	0.45	DOC, TSS, Sp. C., Turb.
		random forest	0.66	TSS, Temp., DO, Sp. C.
	Biological	linear	0.02	$16S_{bga}$, *Phor.*
		random forest	0.16	$16S_{bga}$, *Phor.*
	All	linear	N/A	N/A
		random forest	0.61	Mat, Temp., Sp. C., NO_3^-
Cyanobacteria (% from Taxonomic Density)	***Physicochemical***	linear	N/A	N/A
		random forest	0.78	Sp. C., Temp., DOC, Turb.
	Biological	linear	0.15	BT % Comp., $16S_{bga}$, BT_{bga}, *Phor.*
		random forest	0.78	*Phor.*, $16S_{bga}$, BT_{bga}^{frac}, BT_{bga}
	All	linear	N/A	N/A
		random forest	0.74	Sp. C., Temp., PO_4^{-2}, DOC

The next target variable is the mycE gene and the three normalizations used previously (gram dry weight, gram dry weight of carbon, and scrape area). For all three normalization

methods, the R^2 for the ***Physicochemical*** models ranged between 0.34 and 0.35. Similar to the microcystin ***Physicochemical*** linear regressions, orthophosphate and TSS played the largest roles in the explanation of variance, however no predictor had greater than 17.6% weight in the explanation of variance. As for the ***Biological*** models, the linear regression for the gram dry weight and gram dry weight times percent carbon normalizations were stronger, with R^2 values of 0.69 and 0.76 respectively and the microcystin ELISA concentration playing a large role in the explanation of the variance in these linear regression models. However, the linear regression for the scrape area normalization resulted in a very poor 0.03 linear regression. It appears the microcystin concentration data from the ELISA process had little to no linear relationship with the mycE data in this normalization. Lastly, the microcystin quota linear regression had results for both the ***Physicochemical*** and ***Biological*** predictor groups. The ***Physicochemical*** R^2 was 0.33, albeit with no predictor variables having any substantial explanation of variance. The R^2 for the ***Biological*** linear regression for the microcystin quota was even worse at only 0.05. As both the gene sequence and ELISA concentration results are used to calculate the toxin quota, both are removed from the ***Biological*** predictor variable list. All linear regressions so far for the ***Biological*** predictor group have relied on either mycE or microcystin concentration results to have a R^2 of >0.3.

Table 5: Microcystin gene concentrations and microcystin toxin quota as response variables from statistical analysis. Three normalizations are presented for the gene sequence results along with the three different variable classes the model was split into, and the results from both the linear and random forest models. These results are broken up between the coefficient of determinations for each model and the ranking of the predictor variables that either explained the most variance (ANOVA from linear regression models) or had the highest Shapley values (random forest models). The cyano 16S and *Phormidium* 16S gene sequence results are normalized to gram dry weight for the microcystin quota model.

Response Variable (normalization)	Predictor Class	Model	R^2	Predictor Ranking
mycE (copies/g)	***Physicochemical***	linear	0.34	PO_4^{-2}, TSS, pH, Sp. C.
		random forest	0.79	Turb., Temp., DOC, Sp. C.
	Biological	linear	0.69	MC, BT_{bga}, $16S_{bga}$, BT^{frac}_{bga}
		random forest	0.84	MC, BT^{frac}_{bga}, $16S_{bga}$, *Phor.*
	All	linear	N/A	N/A
		random forest	0.91	MC, BT_{bga}, %-Forest, %-Crop
mycE (copies/gC)	***Physicochemical***	linear	0.35	PO_4^{-2}, TSS, pH, Sp. C.
		random forest	0.83	Turb., Temp., Sp. C., NO_3^-
	Biological	linear	0.76	MC, $16S_{bga}$, BT_{bga}, BT % Comp.
		random forest	0.85	MC, BT^{frac}_{bga}, *Phor.*, $16S_{bga}$
	All	linear	N/A	N/A
		random forest	0.88	MC, DOC, %-Crop, BT^{frac}_{bga}
mycE (copies/cm^2)	***Physicochemical***	linear	0.34	PO_4^{-2}, TSS, pH, Sp. C.
		random forest	0.74	Temp., Turb., DOC, Sp. C.
	Biological	linear	0.03	BT^{frac}_{bga}, BT_{bga}, *Phor.*, $16S_{bga}$
		random forest	0.87	MC, *Phor.*, $16S_{bga}$, BT^{frac}_{bga}
	All	linear	N/A	N/A
		random forest	0.85	MC, %-Imperv, BT_{bga}, %-Urban
Microcystin Quota (µg/copy)	***Physicochemical***	linear	0.33	Temp., DO, PO_4^{-2}, Sp. C.
		random forest	0.77	DOC, Turb., NO_3^-, DO
	Biological	linear	0.05	BT^{frac}_{bga}, *Phor.*, $16S_{bga}$, BT_{bga}
		random forest	0.78	$16S_{bga}BT^{frac}_{bga}$, *Phor.*, BT_{bga}
	All	linear	N/A	N/A
		random forest	0.84	Turb., NO_3^-, DOC, BT^{frac}_{bga}

Table 6: Predictor variables contributing 30% or more of the total predictive power from the ANOVA/Shapley Value groups for microcystin concentration, cyanobacteria, mycE concentration, and microcystin quota models. Displayed is the target response variable, the predictor class, the type of model ran (random forest = R.F.), the normalizations relevant to the predictor, the R^2 of the whole model, the predictor variable, and the corresponding percent contribution.

Target	Predictor Class	Model	Normalization	R^2	Predictor	% Weight
Microcystin	Physico	R.F.	g	0.72	Sp. C.	38.2
Microcystin	Physico	R.F.	gC	0.55	Turb.	39.0
Microcystin	Physico	R.F.	cm^2	0.44	Temp.	49.1
Microcystin	Bio	Linear	g, gC, cm^2	0.7, 0.77, 0.61	$qPCR_{MC}$	64.5, 74.8, 44.2
Microcystin	Bio	R.F.	g, gC, cm^2	0.9, 0.94, 0.75	$qPCR_{MC}$	67.5, 74.0, 46.3
Microcystin	All	R.F.	g, gC, cm^2	0.92, 0.94, 0.75	$qPCR_{MC}$	59.6, 73.6, 41.8
Cyanobacteria	Physico	Linear	BT_{bga}	0.55	pH	39.3
Cyanobacteria	Physico	R.F.	BT_{bga}	0.55	Turb.	40.6
Cyanobacteria	Physico	R.F.	BT_{bga}	0.55	NO_3^-	31.1
Cyanobacteria	Physico	R.F.	TAX_{bga}^{frac}	0.78	Sp. C.	62.9
Cyanobacteria	Bio	R.F.	BT_{bga}, TAX_{bga}^{frac}	0.26, 0.78	*Phor.*	77.3, 57.6
Cyanobacteria	Bio	R.F.	BT_{bga}^{frac}	0.16	$16S_{bga}$	100
Cyanobacteria	All	R.F.	BT_{bga}	0.61	Turb.	37.7
Cyanobacteria	All	R.F.	BT_{bga}^{frac}	0.61	Mat	32.3
Cyanobacteria	All	R.F.	TAX_{bga}^{frac}	0.74	Sp. C.	60.5
mycE qPCR	Physico	R.F.	g, gC, cm^2	0.79, 0.83, 0.74	Turb.	32.5, 32.9, 31.9
mycE qPCR	Physico	R.F.	cm^2	0.74	Temp.	32.0
mycE qPCR	Bio	Linear	g, gC	0.69, 0.76	MC	65.8, 75.1
mycE qPCR	Bio	R.F.	g, gC, cm^2	0.84, 0.85, 0.87	MC	86.1, 88.6, 72.8
mycE qPCR	All	R.F.	g, gC, cm^2	0.91, 0.88, 0.85	MC	58.2, 75.4, 61.6
MC Toxin Quota	Physico	R.F.	N/A	0.77	DOC	45.2
MC Toxin Quota	Bio	R.F.	N/A	0.69	$16S_{bga}$	58.3

4.3.2 Random Forest Analyses

As the random forest can determine relationships outside of just linear, is not impacted by missing data, and is a composition of thousands of variations of single regression tree models (Figure 16), the random forest regression analysis can determine predictor/target variable relationships with greater reliability. Random forest models were run for all of the same variables as the linear regression models (Table 3; Table 4; Table 5), and because missing data doesn't impact random forest models in the same way linear regression models are affected, models for anatoxin-a concentration, saxitoxin concentration, geosmin concentration, and MIB concentrations were also developed and evaluated. In this study, Shapley values are converted to a percent to show how much of the model they describe. Except for the microcystin concentration model normalized to scrape area models, random forest models outperformed linear regression models of similar target and predictor variables. Mean R^2 improved from 0.39 to 0.73 for linear to random forest modeling among the models that have both linear and random forest models (microcystin concentrations, cyanobacteria, mycE concentrations, and microcystin quota).

For microcystin concentration normalized to gram dry weight, specific conductivity had a Shapley value percent weight of 38.2%, when normalized to gram dry weight carbon turbidity had a Shapley value percent weight of 39.0%, and when normalized to scrape area, temperature had a Shapley value percentage of 42.6%. The best R^2 for the ***Physicochemical*** random forest models was 0.72 for the gram dry weight normalization of microcystin concentration random forest models. The best model for the ***Biological*** random forest models for microcystin concentrations was also the gram dry weight of carbon normalization, with a R^2 of 0.94. This correlation is strong due to the high Shapley value percent weight of the mycE result, which makes up 74.0%. In the ***All*** model, R^2 values are similarly high to the ***Biological*** model, and the mycE concentration makes up a very large Shapley value percent portion as well (Figure 17).

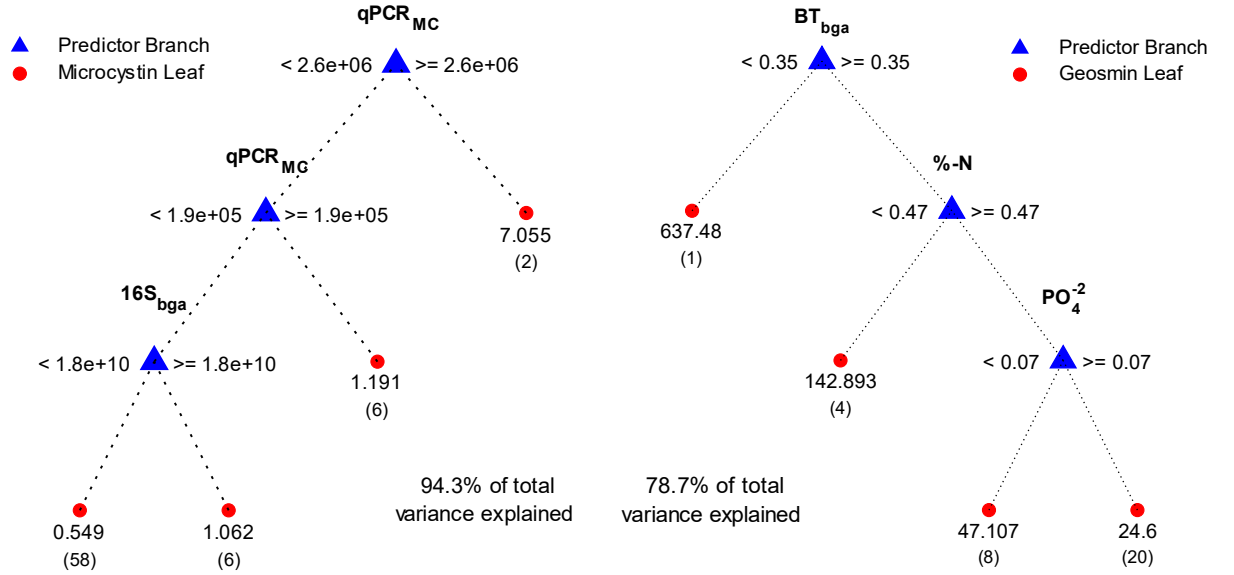

Figure 16: An example of a regression tree with good performance for predicting microcystin (left) and geosmin (right), utilizing ***All*** possible predictor values.

Cyanobacteria models in the random forest regression have the lower R^2 values relative to other target variables, with a R^2 range between 0.16 and 0.78, mostly impacted by the low coefficient of determinations for the ***Biological*** models for BenthoTorch concentration and percent composition. Turbidity, nitrate, specific conductivity, and TSS are strong predictors of Cyanobacteria from the ***Physicochemical*** predictor variable list. Turbidity had a Shapley value percent of 40.6% when modeling for Cyanobacteria concentration from BenthoTorch results in the ***Physicochemical*** model and was also still the top predictor in the ***All*** predictor variables model. Specific conductivity was the top predictor for the cyanobacteria percent composition from taxonomic analysis, with Shapley value percentages of 62.9% for the ***Physicochemical*** analysis and 60.5% for the ***All*** predictors analysis. From the ***Biological*** predictor variable list, *Phormidium* was the top predictor for the taxonomic model that was favorable ($R^2 = 0.78$), but the BenthoTorch-

related cyanobacteria analyses (concentration from BenthoTorch and cyanobacteria percent composition from BenthoTorch) had poor model performance.

R^2 values are strong at over 0.8 for all three normalization types. In the ***Physicochemical*** models, Turbidity was a powerful predictor, with all three Shapley value percent weight for the three normalizations being over 30%, which is being used as a significant predictor threshold for individual predictor variables. Temperature also was a powerful predictor in the scrape area normalization model. As with the microcystin concentration random forest models having high predictive strength from the qPCR results, the flip is seen when running the ***Biological*** predictor variables for the qPCR random forest models. The microcystin concentration is by far the strongest predictor, with Shapley value percentages from 72.8% to 88.6%. The predictive strength of microcystin concentrations carries over to the ***All*** predictor variable random forest model of mycE.

Next, the random forest regression model was run for the microcystin toxin quota target variable. Interestingly, Dissolved Organic Carbon (Shapley value percentage of 45.2%) was the top physicochemical predictor variable for the ***Physicochemical*** model (R^2 = 0.77). Similar to other models so far, qPCR data, in this case Cyano 16S concentration, is the top biological predictor variable in the ***Biological*** model (R^2 = 0.78). Finally, in the ***All*** predictor variable model (R^2 = 0.92), turbidity was the strongest predictor, but did not have a Shapley value percent weight above 30% (28.0 %).

Anatoxin-a did not have a linear regression model result as there were too many missing rows of data, which would have resulted in an error making the linear regression model or a potential over-fitted model due to too few predictor variables. The same happened for saxitoxin, geosmin, and MIB. Anatoxin-a had the best random forest model result from the gram dry weight normalization (Table 7). In both the ***Physicochemical*** and ***All*** predictor variable groups,

temperature had the top predictor weight (R^2 = 0.69 and R^2 = 0.67, respectively) (Table 9). The qPCR results for the Cyano and *Phormidium* 16S also had strong predictability in the ***Biological*** model with 100% of the Shapley weight coming from both the Cyano 16S and *Phormidium* 16S results combined (R^2 = 0.44) (Table 9). A random forest model was also developed and evaluated for the qPCR concentration results for anaC (Table 8), which found that it shared three of the four top Shapley values as the anatoxin-a concentration model, temperature in the ***Physicochemical*** model (R^2 = 0.71), and cyano and *Phormidium* 16S concentrations in the ***Biological*** model (R^2 = 0.57) (Table 9). In the anatoxin-a quota model, TSS and Cyano 16S concentrations made up the bulk of the Shapley value percent weights, with TSS making up 85.9% of the ***Physicochemical*** model.

For saxitoxin, the gram dry weight times percent carbon was the best normalization method (Table 7). While the ***Physicochemical*** random forest model had the best correlation (R^2 = 0.89), none of the predictors stood out as a strong weight in the Shapley values. Data from the BenthoTorch led the ***Biological*** model as well as the ***All*** model, with the cyanobacteria concentration by BenthoTorch predictor variables at 91.4% Shapley value weight in the ***Biological*** model (R^2 = 0.82 and R^2 = 0.84, respectively) (Table 9). Like, the anatoxin-a concentration model and anaC model, the sxtA model did share similar highly effective predictors to the saxitoxin concentration model, but only in the ***Biological*** model (cyanobacteria concentration from the BenthoTorch had a Shapley value percent weight of 0.6) (Table 8). In the ***Physicochemical*** model, turbidity was the top Shapley value (R^2 = 0.56), and in the ***All*** model, percent carbon was the top Shapley value (R^2 = 0.76). For the saxitoxin quota model, temperature was the top Shapley value, with percent weights of 72.7% and 59.7% in the ***Physicochemical*** (R^2 = 0.64) and ***All*** (R^2 = 0.63)

models, respectively (Table 8). Also, in the ***Biological*** model, *Phormidium* had a huge impact on the model, with a 92% Shapley value percent weight ($R^2 = 0.67$) (Table 9).

Like saxitoxin, geosmin also favored the normalization to gram dry weight times percent carbon. Specific conductivity led the ***Physicochemical*** predictor class model with 50.6% Shapley value weight ($R^2 = 0.83$) (Table 7). Also similar to the saxitoxin model, BenthoTorch data provided substantial Shapley value weight to both the ***Biological*** model and the ***All*** predictor class model ($R^2 = 0.70$ and $R^2 = 0.79$, respectively) (Table 9). Unfortunately, no random forest models could be run for geoA concentration or geosmin quota due to too many missing variables in the geoA dataset from non-detects.

Finally, MIB concentration had the best model results from the gram dry weight normalization (Table 7). Temperature was a strong predictor for MIB in both the ***Physicochemical*** and ***All*** predictor class models, with over 70% weight in both Shapley value breakdowns ($R^2 = 0.82$ and $R^2 = 0.75$, respectively) (Table 9). BenthoTorch percent composition of cyanobacteria data was the only substantial Shapley value percentage in the ***Biological*** model for MIB ($R^2 = 0.73$).

Table 7: Random Forest model for anatoxin concentration, saxitoxin concentration, geosmin concentration, and MIB Concentration. The best normalization was chosen to display for each target variable based on the average R2 Coefficient. The top 4 Shapely values for each model are displayed, along with the person make-up of any Shapley value of substantial weight.

Response Variable (normalization)	**Predictor Class**	**R^2**	**Predictor Ranking**
Anatoxin (μg/g)	***Physicochemical***	0.69	Temp., DO, DOC, Turb.
	Biological	0.44	$16S_{bga}$, *Phor.*, BT_{bga}^{frac}, $qPCR_{AA}$
	All	0.67	Temp., DO, DOC, pH
Saxitoxin (μg/gC)	***Physicochemical***	0.85	Turb., DO, DOC, TSS
	Biological	0.82	BT_{bga}, $qPCR_{SX}$, *Phor.*, BT_{bga}^{frac}
	All	0.84	BT_{bga}, %-N, DO, DOC
Geosmin (μg/gC)	***Physicochemical***	0.83	Sp. C., Turb., DOC, DO
	Biological	0.7	BT_{bga}, BT_{bga}^{frac}, *Phor.*, $16S_{bga}$
	All	0.79	BT_{bga}, Sp. C., DOC, Light
MIB (μg/g)	***Physicochemical***	0.82	Temp., Sp. C., DOC, DO
	Biological	0.73	BT_{bga}^{frac}, $16S_{bga}$, *Phor.*, BT_{bga}
	All	0.75	Temp., Sp. C., Light, DO

Table 8: Random Forest model for anatoxin's qPCR sequencing, anaC, saxitoxin concentration, geosmin concentration, and MIB Concentration. The best normalization was chosen to display for each target variable based on the average R2 Coefficient. The top 4 Shapely values for each model are displayed, along with the person make-up of any Shapley value of substantial weight.

Response Variable (normalization)	Predictor Class	R^2	Predictor Ranking
anaC (µg/g)	***Physicochemical***	0.71	Temp., Turb., TSS, NO_3^-
	Biological	0.57	$16S_{bga}$, *Phor.*, BT_{bga}, AA
	All	0.85	$16S_{bga}$, *Phor.*, Turb., Light
sxtA (µg/gC	***Physicochemical***	0.56	Turb., Temp., DO, DOC
	Biological	0.6	BT_{bga}, BT_{bga}^{frac}, $16S_{bga}$, *Phor.*
	All	0.76	%-C, %-N, BT_{bga}, Light
Anatoxin quota (µg/copy)	***Physicochemical***	0.8	TSS, Turb., Temp., Sp. C.
	Biological	0.54	$16S_{bga}$, *Phor.*, BT_{bga}^{frac}, BT_{bga}
	All	0.6	$16S_{bga}$, Turb., DO, %-Forest
Saxitoxin quota (µg/copy)	***Physicochemical***	0.64	Temp., Turb., TSS, DOC
	Biological	0.67	*Phor.*, $16S_{bga}$, BT_{bga}^{frac}, BT_{bga}
	All	0.63	Temp., *Phor.*, %-Crop, Mat

Table 9: Predictor variables contributing 30% or more of the total predictive power from the ANOVA/Shapley Value groups for the anatoxin-a concentration, saxitoxin concentration, geosmin concentration, MIB concentration, anaC concentration, sxtA concentration, anatoxin-a quota, and saxitoxin quota. Displayed is the target response variable, the predictor class, the normalizations relevant to the predictor, the R^2 of the whole model, and the predictor with corresponding percent contribution. All models were random forest models.

Target	Predictor Class	Normalization	R^2	Predictor	% Weight
Anatoxin	Physico	g	0.69	Temp.	52
Anatoxin	Bio	g	0.44	$16S_{bga}$	56.2
Anatoxin	Bio	g	0.44	*Phor.*	43.8
Anatoxin	All	g	0.67	Temp.	62.8
Saxitoxin	Bio	gC	0.82	BT_{bga}	91.4
Saxitoxin	All	gC	0.84	BT_{bga}	62.4
Geosmin	Physico	gC	0.83	Sp. C.	50.6
Geosmin	Bio	gC	0.7	BT_{bga}	59.1
Geosmin	All	gC	0.79	BT_{bga}	32.4
MIB	Physico	g	0.82	Temp.	75.5
MIB	Bio	g	0.73	BT_{bga}^{frac}	58
MIB	All	g	0.75	Temp.	73.4
anaC qPCR	Physico	g	0.71	Temp.	32.2
anaC qPCR	Bio	g	0.57	$16S_{bga}$	46.8
anaC qPCR	Bio	g	0.57	*Phor.*	31.8
sxtA qPCR	Physico	gC	0.56	Turb.	51.4
sxtA qPCR	Bio	gC	0.6	BT_{bga}	57.7
sxtA qPCR	All	gC	0.76	%-C	42.6
AA Toxin Quota	Physico	N/A	0.8	TSS	85.9
AA Toxin Quota	Bio	N/A	0.54	$16S_{bga}$	75.9
AA Toxin Quota	All	N/A	0.6	$16S_{bga}$	66.7
SX Toxin Quota	Physico	N/A	0.64	Temp.	72.7
SX Toxin Quota	Bio	N/A	0.67	*Phor.*	92
SX Toxin Quota	All	N/A	0.63	Temp.	59.7

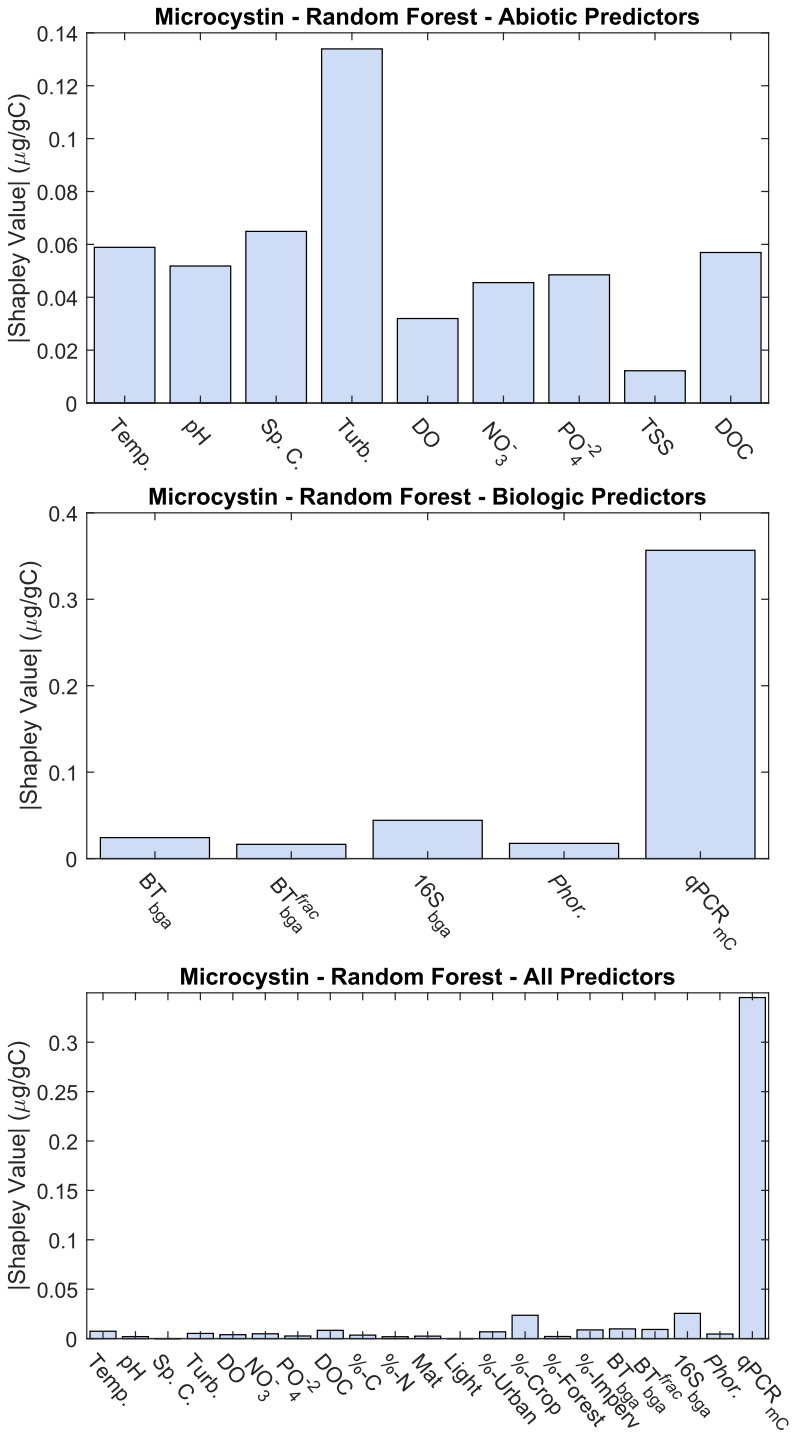

Figure 17: Shapley variable predictor percent weight for (a) ***Physicochemical***, (b) ***Biological***, and (c) ***All*** variable random forest models for microcystin concentrations normalized to gram dry weight carbon.

4.3.3 Model Selection: Linear Regression vs Random Forest

The performance comparison of linear and random forest models can be seen clearly when the random forest model is presented as a linear correlation plot rather than tree plot (Figure 18; Figure 19; Figure 20). Both the random forest model and linear regression models output interesting results for modeling of the dynamics for cyanotoxin and taste and odor compounds, cyanobacterial concentrations, genetic copy concentrations, and cyanotoxin quota. As stated in the methodology section, linear regression models are impacted by missing data. Due to non-detects, field errors, laboratory limitation, etc., the linear regression is more limited to data that is not complete or full. Since some analyses in this study, especially any analysis using the predictor class ***All*** variables, had missing data, linear regression models were impacted by this. A limit on how much missing data was acceptable was set at one third of the amount of data rows in the target variable dataset. Due to this limiter, data from ***All*** variable linear regression models were usually discarded. Linear regression models are simple to run and understand and benefit greatly from variable that share linear relationships. While ecological data usually consists of very complex and non-linear relationships, running linear regression models is simple and useful for comparisons to non-linear multi-variate models. Random forest models on the other hand are very powerful in comparing non-linear complex sets of variables such as water-quality modeling (Wang et al., 2021). Our random forest models almost always outperformed the linear regression models of similar target and predictor variables. While the random forest model worked well in our system to draw relationships between predictor and target variables, there are some model limitations that make comparing other study sites and other data sets more difficult. Random forest models cannot extrapolate data beyond the upper and lower bounds of the given dataset, as it is constructed by regression trees which use given data ranges to create branches in the tree. Sample data from another site may have a range

outside of the range our study has, and therefore wouldn't get much use out of our random forest model. At the extremes, it would only output the max or min seen in this study, whereas that other site could have microcystin concentrations well below or above our sites.

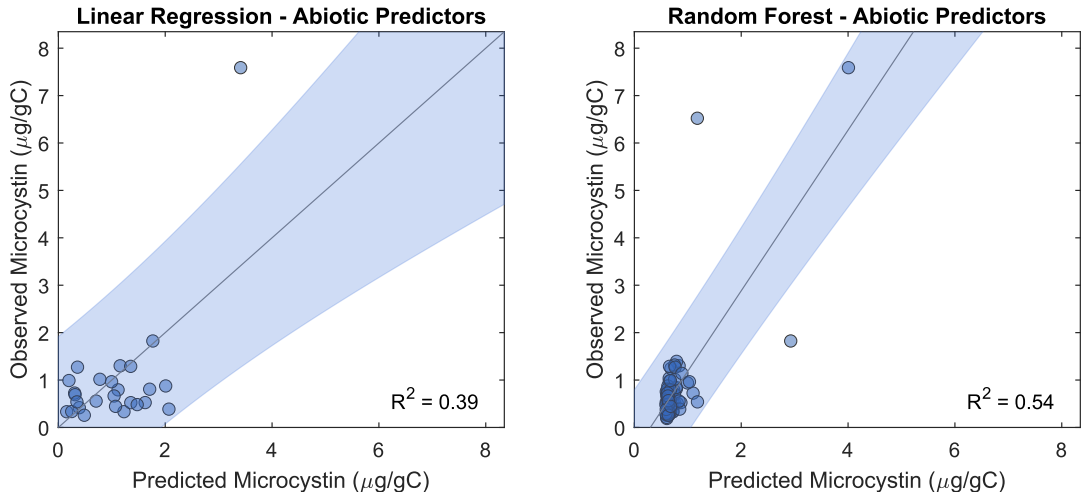

Figure 18: Comparison of best linear regression (left) and random forest (right) using only ***Physicochemical*** values. Both are normalized to gram dry weight carbon.

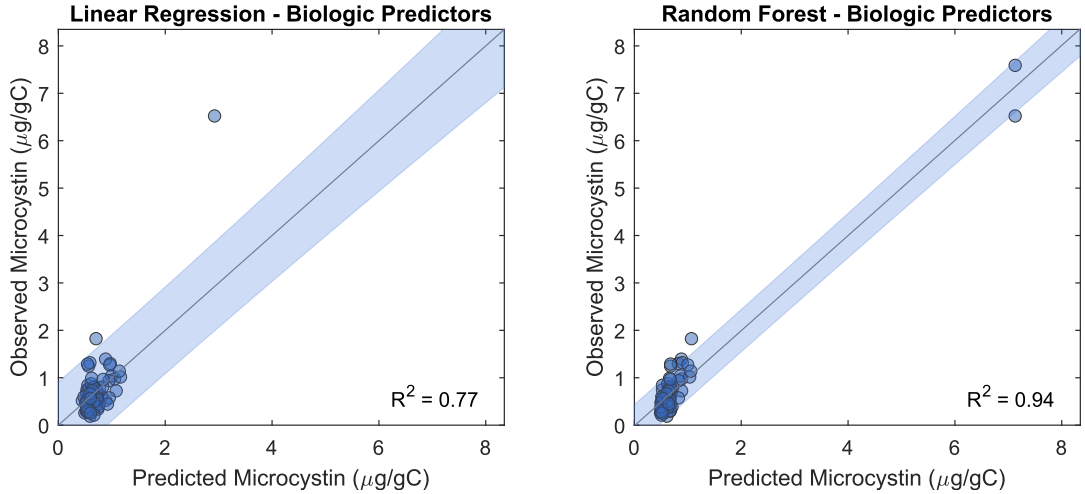

Figure 19: Comparison of best linear regression (left) and random forest (right) using only ***Biological*** values. Both are normalized to gram dry weight carbon.

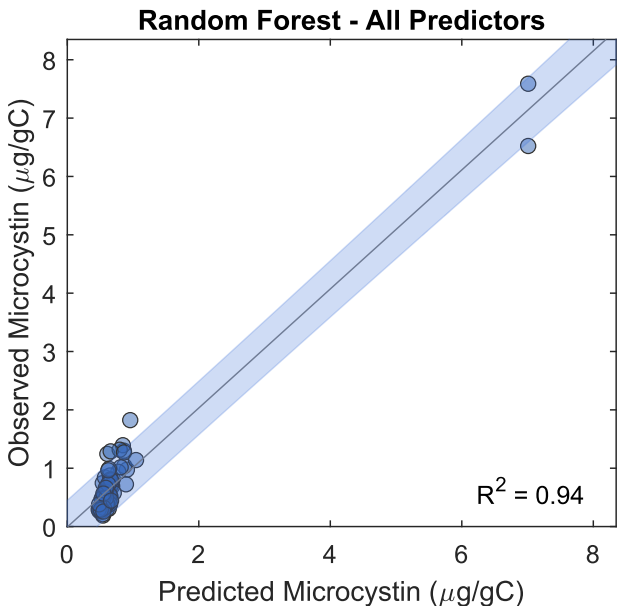

Figure 20: Plot of random forest model using ***All*** variables. The linear regression for the ***All*** predictor model had too many missing rows of data to make an accurate linear regression model, so it wasn't run. Both are normalized to gram dry weight carbon.

Chapter 5: Discussion

Using a wide combination of variables from *in situ* sensors, physicochemical field and laboratory measurements, toxin concentration analysis, genetic qPCR analysis, taxonomic identification and enumeration, isotopic analysis, and land-use characteristics, we had the goal to explain the drivers of cyanotoxin and T&O compound abundance in benthic mat samples in varying human-disturbed streams. We hypothesized that urbanized basins would experience higher cyanotoxin and T&O compound concentrations because of increasing nutrients, Shallower flow depths, and more light penetration through limited tree canopy cover. Qualitatively, there were more algae mats in the urban stream, and fewer in the rural stream. However, our results lead us to reject this original hypothesis. Instead, it appears for our basins that all anthropogenic disturbances, whether urban or rural in nature, create an environment where certain cyanotoxins and T&O compounds can be produced in detectable quantities across physicochemical and biological conditions, irrespective of the type of human disturbance.

5.1 Presence and drivers of cyanotoxin production

Several studies have attempted to link certain drivers to cyanobacteria abundance, cyanotoxin and T&O compound presence, and the gene sequences that have been found to be connected to these toxins and T&O compounds. Lower flow rates, nutrient availability, higher specific conductivity, have all been linked to greater proliferations of cyanobacteria benthic cyanobacteria mats in freshwater environments (Busse et al., 2006; Graham et al., 2018; McAllister et al., 2018; Vadeboncoeur et al., 2021). Low flow rates, shallow flow depths, and nitrogen and phosphorus being present, but not in excess, leads to cyanobacteria dominating benthic assemblages (Scott and Marcarelli, 2012). While flow rates and depths were not actively measured in this study, nutrient levels were. Nitrogen and phosphorus levels in all three streams

were at levels much higher than other benthic algae studies measured, for example, Biggs et al finding soluble inorganic nitrogen and soluble reactive phosphorus levels two-to-three magnitudes lower in concentration than in the streams in this study (Biggs et al., 2000). Interestingly, chl-*a* levels in the Biggs et al study were at or above what were measured in the streams in this study, despite lower nutrient presence. Taxonomically, in studies that take place in streams with lower nutrient levels (Biggs et al., 2000; McAllister et al., 2018) there are more often mats dominated by cyanobacteria, further promoting the subsidy-stress concept. In other words, higher flow rates and/or higher nutrient concentrations promote certain types of filamentous benthic algae while non-filamentous cyanobacteria, especially well-known toxin producing varieties like *Phormidium*, get outcompeted. While our taxonomic results suggest relatively low cyanobacterial biovolume concentrations (13% on average), the BenthoTorch readings (38% on average) and detections of cyanotoxins (100% detect for microcystin) from both ELISA and qPCR methods suggest cyanobacteria are still present in large quantities across our study duration and extent.

One possible reason the taxonomic counts and cyanobacteria biovolumes are lower than the BenthoTorch measurement is the presence of pico-scale cyanobacteria (Harris and Graham, 2015). These cyanobacteria could be difficult or nearly impossible to detect with a microscope, but still contribute to toxin production. However, as the BenthoTorch measures a fluorometric response from the entire 1 cm^2 it is measuring, fluorometric responses by pico-scale cyanobacteria would still be detectable. Another point of note on the BenthoTorch measurement reliability is the spread of results gathered from each cobble at a site that are then combined to create a single composite measurement. The coefficient of variation for BenthoTorch readings per site ranged from 20% to 115%, with most coefficients of variation being above 30%. The BenthoTorch accuracy has come into question in several other studies before, often pointing out that it

underestimates chl-*a*. Harris and Graham found the BenthoTorch underestimated cyanobacteria when the concentrations measured by the BenthoTorch were greater than 4 μg/cm^2 (Harris and Graham, 2015). Also, a different study found the BenthoTorch underestimated cyanobacteria on natural substrates (Echenique-Subiabre et al., 2016). The uncertainties tied to the BenthoTorch do suggest requiring more analysis on it in future studies, but there are other studies that show high correlations with laboratory chl-*a* analysis (Rosero-López et al., 2021). Our outstanding chl-*a* samples are still waiting to be processed, but once they are finished, we can understand if the same relationship holds true for our data.

Taxonomic results were also different from the qPCR results, specifically the Cyano 16S and *Phormidium* genetic concentrations. One reason could be that the scrape site the taxonomic samples were taken from had slightly different benthic species make-up. Some cobbles collected may have only had a small area that had benthic growth on it, and while cobbles were only selected if they had enough growth for both samples, the scrape areas could sometimes be several centimeters away from each other. Another reason qPCR could have estimated high values of cyanobacteria and *Phormidium* rRNA is due to extracellular rRNA being caught in the mat matrix that came from upstream and not necessarily from within the mat itself, resulting in an overestimation of cyanobacteria and *Phormidium* qPCR counts (Ellegaard et al., 2020).

Specific conductivity is another variable that has been known to correlate with cyanobacteria concentrations (Drerup and Vadeboncoeur, 2016; Monteagudo and Moreno, 2016). In general, our random forest models performed better and had no impact from missing data like the linear regression models, and therefore are referenced more in the proceeding discussion. From the random forest models of this study, specific conductivity was found to have substantial influence in the models ran for cyanobacteria concentrations using ***Physicochemical*** variables, and

even had 61% of the Shapley value make-up for the taxonomic data-based cyanobacteria concentration using ***All*** available variables (R^2 = 0.74). In multi-city analyses on urbanization's impact on several physicochemical variables, a significant connection of conductivity to urbanization was found (Potapova et al., 2005). With urbanization causing an increase in specific conductivity in stream waters, and with specific conductivity being a substantial predictor in our models, there is good reason to more specifically look at the link between urbanization's particular impact on specific conductivity and if the particular chemical(s) being increased (and therefore driving a higher specific conductivity) is a driver of potentially harmful benthic cyanobacteria proliferations.

Microcystin was the most commonly found cyanotoxin, with a detection in every sample we collected (n = 72). Microcystin is the one of the most commonly detected cyanobacteria worldwide (Salmaso et al., 2016), and it has been detected in planktonic cyanobacteria studies near Eastern Kansas already (Graham et al., 2018; Harris and Graham, 2017). For microcystin, specific conductivity made up 38.2% of the Shapley value percent weight for the ***Physicochemical*** variable only random forest model of microcystin concentrations normalized to gram dry weight (R^2 = 0.72). The same comparison of urbanization drivers to specific conductivity and that downstream impact on cyanobacteria should be applied to microcystin release in future research as well. Temperature was another ***Physicochemical*** predictor variable with a greater than 30% Shapley value percent weight (42.6%, R^2 = 0.61). A study on the relationship between temperature and freshwater planktonic cyanobacteria found that the prime microcystin forming temperatures were from 20°C to 25°C (Walls et al., 2018). However, in our study, the direct relationship of temperature and microcystin has very poor correlation, with the higher concentration values in our dataset being found from 10°C to 23°C. With there being overlap in the ranges of peak microcystin

production, and a difference of planktonic and benthic algae, the difference could be due to species preferences, as there are many species of cyanobacteria that can produce microcystin (Ibelings et al., 2021). Another slight discrepancy in our random forest model's response to factors other studies have found to relate to microcystin are nutrients. In a study analyzing microcystin production from both lab-grown cultures and phytoplankton lake blooms, higher nitrate levels (16.1 mg/L in the culture test, and above 0.2 mg/L in the lake) were found to correlate with higher microcystin production (Wagner et al., 2021). While, in our study, nitrate does not have a positive correlation with microcystin concentrations. Nitrate and phosphate do show up as predictors with Shapely value percent weights between 8% and 19% for microcystin concentration and mycE concentration models, but these aren't significant correlators and only go to show they are involved in microcystin and mycE presence.

Based on the statistical modeling, the gene count results for mycE provided by qPCR analysis provided the strongest correlation to microcystin concentrations determined through ELISA. For example, mycE gene counts provided as much marginal contribution to the estimation of microcystin concentrations as the next 20 predictive variables combined (Figure 10c). Thus, qPCR analysis alone would be enough to make statistically backed judgements on microcystin presence. Zupančič and team found the same results through their linear regression models with a Pearson correlation coefficient of r = 0.8375 (Zupančič et al., 2021). Table 6 shows qPCR had significant contributions to both linear regression and random forest models when it was a predictive variable for microcystin concentrations, contributing between 42 and 75 percent of the predictive strength of the models, which had corresponding R^2 values between 0.61 and 0.94. However, the drawback of using a predictive model that requires gene counts as predictive

variables is that qPCR analysis is highly involved and requires specialized lab equipment and software to conduct.

Anatoxin-a was found to have high influence from temperature according to the Shapley value percent breakdown of the random forest model run (Table 9). This could be due to the temperature impact on the forms of anatoxin and which are produced (Aráoz et al., 2005). That study suggests anatoxin-a is preferentially created when temperatures are cooler, around 22 °C, while homoanatoxin-a is more common in warmer temperatures, around 25 °C. This is consistent with our study's concentrations of anatoxin-a (Figure 13). This suggests that the relationship between saxitoxin and anatoxin-a isn't as prominent as it seems, and instead the "hand-off" is potentially between homoanatoxin-a (which wasn't measured in this study) and anatoxin-a. The gene sequence measured for anatoxin-a in the study, anaC, also saw a relationship from temperature in the random forest model (R^2 = 0.71; Table 9). The anaC sequence has been linked to both anatoxin and homoanatoxin producers (Rantala-Ylinen et al., 2011), however the relationship with temperature on the possible "hand-off" between homoanatoxin and anatoxin cannot be elucidated with the anaC sequence qPCR results. The anatoxin quota ***Physicochemical*** random forest model showed TSS having a large impact on the prediction of the anatoxin quota, with a Shapley value percent weight over 80% (R^2 = 0.8). In other studies, low nutrients correlated with low anatoxin quota (Heath et al., 2016), but with the higher nutrient levels in this study, that impact was possibly avoided here. Lower TSS is favorable for cyanobacterial growth and has been found to be linked to microcystin production (Graham et al., 2018; Xue et al., 2020). It could be a similar connection with the anatoxin quota, as the ***Biological*** model having cyanobacteria 16S strain concentration as its highest Shapley value by percent weight shows cyanobacteria abundance and the drivers of that abundance both predict anatoxin quota in our model.

Saxitoxin, along with geosmin, showed the most promise for ease of analysis at reduced costs. This is because most of the predictor variables that showed the highest percent weight in their Shapley value breakdowns were from the *in-situ* sensors: DO, turbidity, conductivity, and BenthoTorch measurements. Besides qPCR, the other predictor variables, TSS and DOC, are relatively easy analyses for land managers to complete. Only qPCR is relatively specialized and requires expensive lab equipment. As far as the authors are aware, this may be one of the only analyses performed on saxitoxin with BenthoTorch data, and the high Shapley value weight of the BenthoTorch data certainly suggest further investigation might be worth it to see if the BenthoTorch is a worthy predictor of saxitoxin producing cyanobacteria. Temperature and conductivity were also found to have an impact on saxitoxin in previous studies as compiled by Neilan and colleagues (Neilan et al., 2013). One study showed that saxitoxin concentrations were higher at 28 °C, while optimal growth for the particular cyanobacteria, *Aphanizomenon*, was 22 °C (Dias et al., 2002). However, this particular cyanobacterium was not detected in our taxonomic identification and enumeration analyses. Also, the production of toxins by various cyanobacteria is potentially affected by physicochemical parameters differently. For example, a different study found the opposite affect from temperature on saxitoxin production when analyzing *Raphidiopsis (Cylindrospermopsis) raciborskii* C10 (Castro et al., 2004). As for the sxtA gene concentration, our random forest ***Biological*** model had BenthoTorch concentration as the highest predictor, like the saxitoxin concentration model as well (Table 9). Also, with turbidity being a prominent ***Physicochemical*** predictor, and with temperature being the highest predictor for the saxitoxin quota ***All*** model, the *in situ* sensors can be used effectively in developing a model to assist with monitoring of benthic saxitoxin production.

Geosmin was detected in every sample we took in the study from the GC-MS analysis, however the qPCR results only had 4 detections across all of the samples. This might suggest that the geosmin primer selected was not appropriate for quantifying the geosmin-producing genes in our samples. However, according to one study analyzing many different geosmin primers, the primer used in our study was evaluated in both benthic and planktonic cyanobacteria analyses and is listed as one that could "successfully identify the geosmin synthase for a wide range of geosmin producing cyanobacteria in both laboratory and natural samples" (Devi et al., 2021). Another reason for the discrepancy could be that the actinomycete bacteria are capable of producing geosmin as well (Jüttner and Watson, 2007), and they could have been the dominant producer in the streams. As they were not targeted by any of our analyses, this cannot be confirmed, but is simply a possibility. Therefore, it is difficult to identify the exact mechanism for why geosmin qPCR analysis resulted in so few detections while the GC-MS results showed such strong concentrations. As for the geosmin concentration modeling result, as mentioned above with saxitoxin, many of the substantial Shapley value percent weight predictors were from the BenthoTorch data (cyanobacteria concentration from mass and cell count results, as the percent composition from both as well) as well as the physicochemical data that can be measured *in situ* (specific conductivity and turbidity) or in easy laboratory procedures (TSS and DOC).

For MIB concentrations, the random forest model based on both the ***Physicochemical*** predictor class and ***All*** predictor class had temperature as the top Shapley value, both over 70% percent weight. Also, in the ***Biological*** model, percent composition from the BenthoTorch was the highest predictor, again showing the BenthoTorch is a capable device and should be incorporated into field-based monitoring programs for cyanotoxins and T&O compounds. In a 2014 study, Kakimoto and colleagues found that MIB gene expression increased at 30 °C compared to the

nominal 20 °C sample, and the cooler, 4 °C sample slightly increased and then decreased over a 10-hour period (Kakimoto et al., 2014). However, gene expression does not always correlate to toxin and T&O compound production. Also, in our study the concentration of MIB increased as the water temperature became cooler, with the MIB peaks at a water temperature of 10 °C. Unfortunately, we did not analyze MIB through the qPCR process because we lacked access to an appropriate primer at project onset. However, it does appear that a qPCR primer for potential future analysis of MIB has been tested (Kim et al., 2020).

Several datasets in our study showed results that seemed at odds with one another. While at first this may indicate contradictory results, it highlights the importance of using a multi-faceted approach whereby the benthic algae are assessed using a variety of robust methods. If a water quality stakeholder were to use just a single line of evidence, it may bias them toward an understanding that cyanobacteria are either present in high numbers or absent entirely. The reality likely falls somewhere in between these two end-members and the more data that are available, the more nuance can be attributed to cyanotoxin and T&O compound formation. In our study, we utilized multiple methods of analysis for both cyanobacteria and the secondary metabolites being studied, with hopes to both gain more insight on what impacts both the potential genes for these metabolites and the actual creation and release of them.

5.2 Proliferation of benthic cyanotoxins in human environments

In this study, three different cyanotoxins and two different T&O compounds were detected in streams across Johnson County, Kansas. One original hypothesis in the study was to compare the differences in land use in these three streams to the presence and concentrations of these secondary metabolites. However, with detections in all streams, and no clear correlation in concentration with land use or physicochemical variables associated with land use differences, we

can surmise that it isn't the level of urbanization and impervious surfaces that correlate to drastic gradients in the presence or density of cyanotoxins and T&O compounds. Instead, our results show that, for our study site, cyanotoxins and T&O compounds are present near any human disturbed land, whether residential, pasture, cropland, urban, impervious, or commercial. Even when nutrient levels are at or higher than one would expect to find cyanobacteria dominating the benthic mats (Pinckney et al., 1995), there are still enough cyanotoxin producing cyanobacteria present to pose a risk of exposure to these chemicals.

With the prevalence of cyanotoxins and T&O compounds sourced from benthic cyanobacteria mats in the waters around the world, water managers in many fields including utilities, municipalities, and recreation should invest in benthic cyanobacterial bloom monitoring. In a study on the monitoring practices and processes of utility managers around the United States, Canada, and Australia, it was found that only 20 percent of these monitor for benthic cyanobacteria (Kibuye et al., 2021). Also, the monitoring practices that are used (in either benthic or phytoplanktonic) are often not as effective as they should be. A tiered method is proposed by the Kibuye study, one which tests for biological activity in tier 1 with the cheaper and faster monitoring tools (visual assessment or chl-*a* measurements for instance). Based on our study's results, we would suggest visual assessments at locations known to have high light availability, low flow depths, and anthropomorphic land use nearby. High frequency chl-*a* sensors could also be utilized, although a connection between water column chl-*a* levels and benthic activity was not analyzed in this study. These high frequency sensors could also measure temperature and specific conductivity, which based on our analysis can be used to judge when benthic productivity is higher and the chance of cyanotoxins and T&O compounds are present. While the activity of different benthic cyanobacteria may differ from the temperature and specific conductivity levels we found

in our study, we would suggest a visual assessment of the area, if stream temperatures range from 10°C to 30°C and specific conductivity is higher than 750 mS/cm. Tier two is cyanobacteria confirmation with methods such as taxonomic microscopy confirmation. Based on our results, confirmation with multiple BenthoTorch measurements would be much faster and potentially cheaper than taxonomic analysis. However, the one benefit of the taxonomic analysis is identifying known toxin producing genus, such as *Phormidium*, which the BenthoTorch is incapable of doing. Cyanobacteria concentrations from the BenthoTorch of greater than 1.3 $\mu g/cm^2$ can be considered to have high enough concentrations of cyanobacteria that cyanotoxins and T&O compounds could be present in detectable amounts. Finally, tier 3 would confirm the presence and concentrations of metabolites, such as the cyanotoxins in this study, utilizing tools such as SPATT bags, ELISA, or qPCR. Unless the organization investigating cyanotoxins already has a GC-MS or qPCR instrument, ELISA is a much cheaper method to identify cyanotoxins. ELISA for geosmin (Bristow et al., 2019) and MIB (Chung et al., 1990) have been developed in the past, but their sensitivity is sometimes questioned. Also,they are not readily available from companies like Eurofins, where we sourced our microcystin, anatoxin, and saxitoxin ELISA kits from. Instead, GC-MS would be our suggested method for identifying geosmin and MIB. SPATT bags are another cheap way to detect the presence of cyanotoxins in the water column and have been used to detect anatoxin (Bouma-Gregson et al., 2018; Wood et al., 2018), microcystin (Kudela et al., 2011), and saxitoxin (Lane et al., 2010). SPATT Bags are not very efficient at measuring the quantitative concentration of toxin in the water (Wood et al., 2020), but detecting the presence of the toxin in the water column can add additional insight to intracellular toxin concentrations. Once a detection of cyanotoxin or T&O compound has been found to cross a certain threshold, actions should be taken to reduce public and animal exposure and to possibly eliminate the toxic

cyanobacterial bloom. The suggested tiered method, or any similarly robust monitoring method should be investigated further to determine the best way to quickly identify and act on toxic benthic cyanobacterial blooms. Models already developed (Otten et al., 2016), or ones developed in the future, that can be output as equations with a target variable of a certain cyanotoxin concentration or gene sequence abundance can be utilized as well such that the predicted cyanotoxin quantity can be output based on the input physicochemical and biological data from the utility's own analysis. Finally, these methods and thresholds determined should be regulated and utilized in large areas (such as states, regions, or countries), so that improvements made to these common practices affect a much larger population and new practices spread quicker.

However, regulatory thresholds need to be developed for benthic cyanobacteria sourced toxins and T&O compounds. For instance, it is difficult to link current regulatory suggestions for microcystin as they are usually related to concentrations in liquid water like the EPA's suggestion of 1.6 μg/L in drinking water or 8 μg/L in recreational waters (USEPA, 2019, 2015). Benthic mat microcystin is usually presented in a mass per unit dry weight and the measured toxins are mostly intracellular, meaning the consumption of the algae is usually required to experience any significant concentration of the toxins.

As stated in the introduction, the consumption of toxic benthic cyanobacteria biomass has led to animal deaths. It should be noted that concentrations of cyanotoxins found in dead animals that were found to have consumed cyanotoxin-rich benthic mats is many orders of magnitude higher than the concentrations of cyanotoxins we found in the benthic mats around Johnson County (Bauer et al., 2020). These concentrations were first converted to an equivalent per volume concentration based on the volume immediately around the mat, which the authors suggest might drive up the magnitude of the concentration. For example, anatoxin-a concentrations in a 1mL of

water simulated 68,000 μg/L. However, even the per dry weight concentrations are between 300 and 1200 μg/g dry weight, which is 6 orders of magnitude greater than concentrations found in our ELISA analysis. In the concentrations detected by this study (mean of 0.08 μg/g), the chance of poisoning in likely very low. Wood et al. established that an anatoxin-a concentration of around 35 mg/kg (mg/kg and μg/g are equivalent units) could kill a mouse (Wood et al., 2010). As this shows that toxin concentrations can reach higher concentrations in benthic cyanobacteria mats than our study area found, a continued analysis of physicochemical and biological drivers of cyanotoxins and T&O compounds should be done. Especially in these areas where very high concentrations have been found. As our random forest models would not be able to extrapolate to concentrations that high, predictive modeling should also continue to be run on the results of all toxic benthic cyanobacteria bloom environmental parameters.

5.3 Future research considerations

There are two factors that have come about through this research that were not able to be considered, but could make for interesting research in the future. The first would be the inclusion of a pristine stream, which would serve as an undisturbed end-member and provide a point of comparison to the existing agricultural urban and stream sites. In our analysis, the rural stream is still heavily impacted by human activities, including agriculture and livestock pasture. Both of these land use types have the potential to impart nutrients and increase sediment in the stream (Hoorman et al., 2008; Lenat, 1984). Adding a pristine stream in to the analysis may allow for the impact of land use and urbanization to be more noticeable in analyzing the drivers and dynamics of benthic cyanobacteria, cyanotoxins, T&O compounds, and relevant gene sequence concentrations. Pristine streams may have low enough nutrients to allow for nutrient-limitations to play a role in the analysis, and the possibly lower specific conductivity could be further analyzed

as a strong predictor in random forest models. The second factor is incorporating hydrological data such as flow rates and flow depths. These physicochemical variables have been known to impact cyanobacterial growth dynamics (McAllister et al., 2018; Scott and Marcarelli, 2012). Hydrologic variables like water residence time are also impactful on determining which taxa dominate in a particular benthic cyanobacteria mat (Paerl, 2008). High flow events have been found to correlate with densely populated areas (Brown et al., 2009), so incorporating flow data into an analysis may also help elucidate the impact of urbanization on these target variables.

Chapter 6: Conclusion

It is shown in this study that cyanotoxins and T&O compounds in human-disturbed freshwater streams are prevalent no matter the level of urbanization. Direct statistical relationships between cyanotoxins and T&O compounds and urban land use were not found in any substantial form. However, there were strong random forest model connections between specific conductivity, temperature, turbidity, and organic carbon with all cyanotoxins and T&O compounds analyzed here to varying degrees. These predictor variables have been known to be affected by increased nearby anthropogenic influences, and thus it is really that general human disturbance could drive increased cyanotoxin and T&O compound production rather than specifically urban activity. As our system did not have a pristine freshwater stream environment to use as an "undisturbed benchmark", further analysis into these conclusions is needed. Due to the complex nature in which human-disturbed environments impact freshwater ecosystems, it is suggested that more research not only be done in lab-controlled tests but also in real-world environments. The field-based research covers a larger array of ambient variables than can be simulated within lab-based analysis. On the other hand, lab-based analyses can help narrow down which specific variables impact cyanotoxin and T&O compound release with much better control over predictor variable dynamics. As climate change and urbanization are both projected to increase in the coming decades, the potential safety and environmental impact benthic-sourced cyanotoxins and T&O compounds can have will increase as well. Land managers need to create, test, and adopt benthic cyanobacterial monitoring programs and increase testing for cyanotoxins when blooms do occur to reduce instances of animal death, or to prevent human harm.

References

Al-Tebrineh, Jamal, et al. “A Multiplex QPCR Targeting Hepato- and Neurotoxigenic Cyanobacteria of Global Significance.” Harmful Algae, vol. 15, Mar. 2012, pp. 19–25. DOI.org (Crossref), https://doi.org/10.1016/j.hal.2011.11.001.

Aráoz, Rómulo, et al. “Neurotoxins in Axenic Oscillatorian Cyanobacteria: Coexistence of Anatoxin-a and Homoanatoxin-a Determined by Ligand-Binding Assay and GC/MS.” Microbiology, vol. 151, no. 4, Apr. 2005, pp. 1263–73. DOI.org (Crossref), https://doi.org/10.1099/mic.0.27660-0.

Auffret, Marc, et al. “Impact of Water Quality on the Bacterial Populations and Off-Flavours in Recirculating Aquaculture Systems.” FEMS Microbiology Ecology, vol. 84, no. 2, May 2013, pp. 235–47. DOI.org (Crossref), https://doi.org/10.1111/1574-6941.12053.

Bauer, Franziska, et al. “Mass Occurrence of Anatoxin-a- and Dihydroanatoxin-a-Producing Tychonema Sp. in Mesotrophic Reservoir Mandichosee (River Lech, Germany) as a Cause of Neurotoxicosis in Dogs.” Toxins, vol. 12, no. 11, Nov. 2020, p. 726. DOI.org (Crossref), https://doi.org/10.3390/toxins12110726.

Bouma-Gregson, Keith, et al. “Rise and Fall of Toxic Benthic Freshwater Cyanobacteria (Anabaena Spp.) in the Eel River: Buoyancy and Dispersal.” Harmful Algae, vol. 66, June 2017, pp. 79–87. DOI.org (Crossref), https://doi.org/10.1016/j.hal.2017.05.007.

Bristow, R. L., et al. “An Extensive Review of the Extraction Techniques and Detection Methods for the Taste and Odour Compound Geosmin (Trans-1, 10-Dimethyl-Trans-9-Decalol) in Water.” TrAC Trends in Analytical Chemistry, vol. 110, Jan. 2019, pp. 233–48. DOI.org (Crossref), https://doi.org/10.1016/j.trac.2018.10.032.

Brown, Larry R., et al. “Urban Streams across the USA: Lessons Learned from Studies in 9 Metropolitan Areas.” Journal of the North American Benthological Society, vol. 28, no. 4, Dec.

2009, pp. 1051–69. DOI.org (Crossref), https://doi.org/10.1899/08-153.1.

Busse, Lilian B., et al. “Relationships among Nutrients, Algae, and Land Use in Urbanized Southern California Streams.” Canadian Journal of Fisheries and Aquatic Sciences, vol. 63, no. 12, Dec. 2006, pp. 2621–38. DOI.org (Crossref), https://doi.org/10.1139/f06-146.

Carey, Richard O., and Kati W. Migliaccio. “Contribution of Wastewater Treatment Plant Effluents to Nutrient Dynamics in Aquatic Systems: A Review.” Environmental Management, vol. 44, no. 2, Aug. 2009, pp. 205–17. DOI.org (Crossref), https://doi.org/10.1007/s00267-009-9309-5.

Castro, Daniela, et al. “The Effect of Temperature on Growth and Production of Paralytic Shellfish Poisoning Toxins by the Cyanobacterium Cylindrospermopsis Raciborskii C10.” Toxicon, vol. 44, no. 5, Oct. 2004, pp. 483–89. DOI.org (Crossref), https://doi.org/10.1016/j.toxicon.2004.06.005.

Chung, Si Yin, et al. “Development of an ELISA Using Polyclonal Antibodies Specific for 2-Methylisoborneol.” Journal of Agricultural and Food Chemistry, vol. 38, no. 2, Feb. 1990, pp. 410–15. DOI.org (Crossref), https://doi.org/10.1021/jf00092a015.

D’Anglada, Lesley V., and Jamie Strong. Drinking Water Health Advisory for the Cyanobacterial Microcystin Toxins. 820R15100, United States Environmental Protection Agency, 2015, p. 75.

De’ath, Glenn, and Katharina E. Fabricius. “CLASSIFICATION AND REGRESSION TREES: A POWERFUL YET SIMPLE TECHNIQUE FOR ECOLOGICAL DATA ANALYSIS.” Ecology, vol. 81, no. 11, Nov. 2000, pp. 3178–92. DOI.org (Crossref), https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.

Devi, Apramita, et al. “Quantitative PCR Based Detection System for Cyanobacterial Geosmin/2-Methylisoborneol (2-MIB) Events in Drinking Water Sources: Current Status and Challenges.” Water Research, vol. 188, Jan. 2021, p. 116478. DOI.org (Crossref), https://doi.org/10.1016/j.watres.2020.116478.

Dias, Elsa, et al. “PRODUCTION OF PARALYTIC SHELLFISH TOXINS BY APHANIZOMENON SP. LMECYA 31 (CYANOBACTERIA).” Journal of Phycology, vol. 38, no. 4, Aug. 2002, pp. 705–12. DOI.org (Crossref), https://doi.org/10.1046/j.1529-8817.2002.01146.x.

Dietrich, Andrea M. “Aesthetic Issues for Drinking Water.” Journal of Water and Health, vol. 4, no. S1, July 2006, pp. 11–16. DOI.org (Crossref), https://doi.org/10.2166/wh.2006.0038.

Drerup, Samuel A., and Yvonne Vadeboncoeur. “Elevated Specific Conductance Enhances Productivity and Biomass of Periphytic Cyanobacteria from Lake Tahoe and Lake Tanganyika.” Phycologia, vol. 55, no. 3, Mar. 2016, pp. 295–98. DOI.org (Crossref), https://doi.org/10.2216/15-100.1.

Dunlap, Catherine R., et al. “A Costly Endeavor: Addressing Algae Problems in a Water Supply.” Journal - American Water Works Association, vol. 107, no. 5, May 2015, pp. E255–62. DOI.org (Crossref), https://doi.org/10.5942/jawwa.2015.107.0055.

Echenique-Subiabre, Isidora, et al. “Application of a Spectrofluorimetric Tool (Bbe BenthoTorch) for Monitoring Potentially Toxic Benthic Cyanobacteria in Rivers.” Water Research, vol. 101, Sept. 2016, pp. 341–50. DOI.org (Crossref), https://doi.org/10.1016/j.watres.2016.05.081.

Ellegaard, Marianne, et al. “Dead or Alive: Sediment DNA Archives as Tools for Tracking Aquatic Evolution and Adaptation.” Communications Biology, vol. 3, no. 1, Dec. 2020, p. 169. DOI.org (Crossref), https://doi.org/10.1038/s42003-020-0899-z.

EPA. SW-846 Test Method 9060A: Total Organic Carbon. U.S. Environmental Protection Agency, Nov. 2004, https://www.epa.gov/hw-sw846/sw-846-test-method-9060a-total-organic-carbon.

EPA Office of Water. Method 546: Determination of Total Microcystins and Nodularins in Drinking Water and Ambient Water by Adda Enzyme-Linked Immunosorbent Assay. Aug. 2016, p. 24.

Graham, Jennifer L., et al. Water-Quality Conditions with an Emphasis on Cyanobacteria and

Associated Toxins and Tasteand- Odor Compounds in the Kansas River, Kansas, July 2012 through September 2016. Scientific Investigations Report, 2018–5089, USGS, 2018, p. 66.

Harris, Ted D., and Jennifer L. Graham. "Predicting Cyanobacterial Abundance, Microcystin, and Geosmin in a Eutrophic Drinking-Water Reservoir Using a 14-Year Dataset." Lake and Reservoir Management, vol. 33, no. 1, Jan. 2017, pp. 32–48. DOI.org (Crossref), https://doi.org/10.1080/10402381.2016.1263694.

---. "Preliminary Evaluation of an in Vivo Fluorometer to Quantify Algal Periphyton Biomass and Community Composition." Lake and Reservoir Management, vol. 31, no. 2, Apr. 2015, pp. 127–33. DOI.org (Crossref), https://doi.org/10.1080/10402381.2015.1025153.

Heath, Mark, et al. "The Role of Nitrogen and Phosphorus in Regulating Phormidium Sp. (Cyanobacteria) Growth and Anatoxin Production." FEMS Microbiology Ecology, edited by Riks Laanbroek, vol. 92, no. 3, Mar. 2016, p. fiw021. DOI.org (Crossref), https://doi.org/10.1093/femsec/fiw021.

Hoorman, J., et al. "Agricultural Impacts on Lake and Stream Water Quality in Grand Lake St. Marys, Western Ohio." Water, Air, and Soil Pollution, vol. 193, no. 1–4, Sept. 2008, pp. 309–22. DOI.org (Crossref), https://doi.org/10.1007/s11270-008-9692-1.

Howarth, Robert W., et al. "Nitrogen Fixation in Freshwater, Estuarine, and Marine Ecosystems. 1. Rates and Importance1: Nitrogen Fixation: 1." Limnology and Oceanography, vol. 33, no. 4part2, July 1988, pp. 669–87. DOI.org (Crossref), https://doi.org/10.4319/lo.1988.33.4part2.0669.

Ibelings, Bastiaan W., et al. "Understanding the Occurrence of Cyanobacteria and Cyanotoxins." Toxic Cyanobacteria in Water, edited by Ingrid Chorus and Martin Welker, 2nd ed., CRC Press, 2021, pp. 213–94. DOI.org (Crossref), https://doi.org/10.1201/9781003081449-4.

Ishikawa, Naoto F., et al. "Global Dataset for Carbon and Nitrogen Stable Isotope Ratios of Lotic

Periphyton." Ecological Research, vol. 33, no. 6, Nov. 2018, pp. 1089–1089. DOI.org (Crossref), https://doi.org/10.1007/s11284-018-1623-z.

Jüttner, Friedrich, and Susan B. Watson. "Biochemical and Ecological Control of Geosmin and 2-Methylisoborneol in Source Waters." Applied and Environmental Microbiology, vol. 73, no. 14, July 2007, pp. 4395–406. DOI.org (Crossref), https://doi.org/10.1128/AEM.02250-06.

Kakimoto, Masayuki, et al. "Culture Temperature Affects Gene Expression and Metabolic Pathways in the 2-Methylisoborneol-Producing Cyanobacterium Pseudanabaena Galeata." Journal of Plant Physiology, vol. 171, no. 3–4, Feb. 2014, pp. 292–300. DOI.org (Crossref), https://doi.org/10.1016/j.jplph.2013.09.005.

Kaloudis, Triantafyllos, et al. "Cyanotoxins in Bloom: Ever-Increasing Occurrence and Global Distribution of Freshwater Cyanotoxins from Planktic and Benthic Cyanobacteria." Toxins, vol. 14, no. 4, Apr. 2022, p. 264. DOI.org (Crossref), https://doi.org/10.3390/toxins14040264.

Kasinak, Jo-Marie E., et al. "Benchtop Fluorometry of Phycocyanin as a Rapid Approach for Estimating Cyanobacterial Biovolume." Journal of Plankton Research, vol. 37, no. 1, Jan. 2015, pp. 248–57. DOI.org (Crossref), https://doi.org/10.1093/plankt/fbu096.

Kibuye, Faith A., et al. "Utility Practices and Perspectives on Monitoring and Source Control of Cyanobacterial Blooms." AWWA Water Science, vol. 3, no. 6, Nov. 2021. DOI.org (Crossref), https://doi.org/10.1002/aws2.1264.

Kim, Keonhee, et al. "Molecular Probes to Evaluate the Synthesis and Production Potential of an Odorous Compound (2-Methylisoborneol) in Cyanobacteria." International Journal of Environmental Research and Public Health, vol. 17, no. 6, Mar. 2020, p. 1933. DOI.org (Crossref), https://doi.org/10.3390/ijerph17061933.

Kinouchi, T., et al. "Increase in Stream Temperature Related to Anthropogenic Heat Input from Urban

Wastewater." Journal of Hydrology, vol. 335, no. 1–2, Mar. 2007, pp. 78–88. DOI.org (Crossref), https://doi.org/10.1016/j.jhydrol.2006.11.002.

Koreivienė, Judita, et al. "Cyanotoxin Management and Human Health Risk Mitigation in Recreational Waters." Environmental Monitoring and Assessment, vol. 186, no. 7, July 2014, pp. 4443–59. DOI.org (Crossref), https://doi.org/10.1007/s10661-014-3710-0.

Kudela, Raphael M. "Characterization and Deployment of Solid Phase Adsorption Toxin Tracking (SPATT) Resin for Monitoring of Microcystins in Fresh and Saltwater." Harmful Algae, vol. 11, Nov. 2011, pp. 117–25. DOI.org (Crossref), https://doi.org/10.1016/j.hal.2011.08.006.

Kutovaya, Olga A., and Sue B. Watson. "Development and Application of a Molecular Assay to Detect and Monitor Geosmin-Producing Cyanobacteria and Actinomycetes in the Great Lakes." Journal of Great Lakes Research, vol. 40, no. 2, June 2014, pp. 404–14, https://doi.org/10.1016/j.jglr.2014.03.016.

Lane, Jenny Q., et al. "Application of Solid Phase Adsorption Toxin Tracking (SPATT) for Field Detection of the Hydrophilic Phycotoxins Domoic Acid and Saxitoxin in Coastal California: SPATT for Field Detection of Domoic Acid." Limnology and Oceanography: Methods, vol. 8, no. 11, Nov. 2010, pp. 645–60. DOI.org (Crossref), https://doi.org/10.4319/lom.2010.8.0645.

Lenat, David R. "Agriculture and Stream Water Quality: A Biological Evaluation of Erosion Control Practices." Environmental Management, vol. 8, no. 4, July 1984, pp. 333–43. DOI.org (Crossref), https://doi.org/10.1007/BF01868032.

Lund, J. W. G., et al. "The Inverted Microscope Method of Estimating Algal Numbers and the Statistical Basis of Estimations by Counting." Hydrobiologia, vol. 11, 1958, pp. 143–70, https://doi.org/10.1007/BF00007865.

Marquardt, Jürgen, and Katarzyna A. Palinska. "Genotypic and Phenotypic Diversity of Cyanobacteria

Assigned to the Genus Phormidium (Oscillatoriales) from Different Habitats and Geographical Sites." Archives of Microbiology, vol. 187, no. 5, May 2007, pp. 397–413. DOI.org (Crossref), https://doi.org/10.1007/s00203-006-0204-7.

McAllister, Tara G., et al. "Spatiotemporal Dynamics of Phormidium Cover and Anatoxin Concentrations in Eight New Zealand Rivers with Contrasting Nutrient and Flow Regimes." Science of The Total Environment, vol. 612, Jan. 2018, pp. 71–80. DOI.org (Crossref), https://doi.org/10.1016/j.scitotenv.2017.08.085.

Mez, Konstanze, et al. "Identification of a Microcystin in Benthic Cyanobacteria Linked to Cattle Deaths on Alpine Pastures in Switzerland." European Journal of Phycology, vol. 32, no. 2, May 1997, pp. 111–17. DOI.org (Crossref), https://doi.org/10.1080/09670269710001737029.

Monteagudo, Laura, and José Luis Moreno. "Benthic Freshwater Cyanobacteria as Indicators of Anthropogenic Pressures." Ecological Indicators, vol. 67, Aug. 2016, pp. 693–702. DOI.org (Crossref), https://doi.org/10.1016/j.ecolind.2016.03.035.

"National Land Cover Database (NLCD) 2019" MRLC, 2019, https://www.mrlc.gov/data. Accessed 1 January 2020.

Neilan, Brett A., et al. "Environmental Conditions That Influence Toxin Biosynthesis in Cyanobacteria: Regulation of Cyanobacterial Toxin Biosynthesis." Environmental Microbiology, vol. 15, no. 5, May 2013, pp. 1239–53. DOI.org (Crossref), https://doi.org/10.1111/j.1462-2920.2012.02729.x.

O'Dell, James W. "DETERMINATION OF AMMONIA NITROGEN BY SEMI-AUTOMATED COLORIMETRY." Methods for the Determination of Metals in Environmental Samples, Elsevier, 1996, pp. 434–48. DOI.org (Crossref), https://doi.org/10.1016/B978-0-8155-1398-8.50024-0.

---. "DETERMINATION OF NITRATE-NITRITE NITROGEN BY AUTOMATED COLORIMETRY." Methods for the Determination of Metals in Environmental Samples, Elsevier,

1996, pp. 464–78. DOI.org (Crossref), https://doi.org/10.1016/B978-0-8155-1398-8.50026-4.

---. “DETERMINATION OF PHOSPHORUS BY SEMI-AUTOMATED COLORIMETRY.” Methods for the Determination of Metals in Environmental Samples, Elsevier, 1996, pp. 479–95. DOI.org (Crossref), https://doi.org/10.1016/B978-0-8155-1398-8.50027-6.

Otten, Timothy G., et al. “Elucidation of Taste- and Odor-Producing Bacteria and Toxigenic Cyanobacteria in a Midwestern Drinking Water Supply Reservoir by Shotgun Metagenomic Analysis.” Applied and Environmental Microbiology, edited by G. Voordouw, vol. 82, no. 17, Sept. 2016, pp. 5410–20. DOI.org (Crossref), https://doi.org/10.1128/AEM.01334-16.

Paerl, Hans. “Nutrient and Other Environmental Controls of Harmful Cyanobacterial Blooms along the Freshwater–Marine Continuum.” Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs, edited by H. Kenneth Hudnell, vol. 619, Springer New York, 2008, pp. 217–37. DOI.org (Crossref), https://doi.org/10.1007/978-0-387-75865-7_10.

Paerl, Hans W., and Malcolm A. Barnard. “Mitigating the Global Expansion of Harmful Cyanobacterial Blooms: Moving Targets in a Human- and Climatically-Altered World.” Harmful Algae, vol. 96, June 2020, p. 101845. DOI.org (Crossref), https://doi.org/10.1016/j.hal.2020.101845.

Paerl, Hans W., and Valerie J. Paul. “Climate Change: Links to Global Expansion of Harmful Cyanobacteria.” Water Research, vol. 46, no. 5, Apr. 2012, pp. 1349–63. DOI.org (Crossref), https://doi.org/10.1016/j.watres.2011.08.002.

Pinckney, J., et al. “Impacts of Seasonality and Nutrients on Microbial Mat Community Structure and Function.” Marine Ecology Progress Series, vol. 123, 1995, pp. 207–16. DOI.org (Crossref), https://doi.org/10.3354/meps123207.

Potapova, Marina, et al. “A Comparison of the Influences of Urbanization in Contrasting Environmental Settings on Stream Benthic Algal Assemblages.” American Fisheries Society Symposium, vol. 47,

2005, pp. 333–59.

Qi, Haiping, et al. “Two New Organic Reference Materials for Δ13C and Δ15N Measurements and a New Value for the Δ13C of NBS 22 Oil.” Rapid Communications in Mass Spectrometry, vol. 17, no. 22, Nov. 2003, pp. 2483–87, https://doi.org/10.1002/rcm.1219.

Qi, Haiping, et al. “A New Organic Reference Material, L-Glutamic Acid, USGS41a, for Δ13C and Δ15N Measurements − a Replacement for USGS41.” Rapid Communications in Mass Spectrometry, vol. 30, no. 7, Apr. 2016, pp. 859–66. DOI.org (Crossref), https://doi.org/10.1002/rcm.7510.

Quiblier, Catherine, et al. “A Review of Current Knowledge on Toxic Benthic Freshwater Cyanobacteria – Ecology, Toxin Production and Risk Management.” Water Research, vol. 47, no. 15, Oct. 2013, pp. 5464–79. DOI.org (Crossref), https://doi.org/10.1016/j.watres.2013.06.042.

Rantala-Ylinen, Anne, et al. “Anatoxin-a Synthetase Gene Cluster of the Cyanobacterium Anabaena Sp. Strain 37 and Molecular Methods To Detect Potential Producers.” Applied and Environmental Microbiology, vol. 77, no. 20, Oct. 2011, pp. 7271–78. DOI.org (Crossref), https://doi.org/10.1128/AEM.06022-11.

Rasmussen, Teresa, et al. Quality of Streams in Johnson County, Kansas, 2002–10. Scientific Investigations Report, 2012–5279, USGS, 2012, p. 116.

Rasmussen, Teresa, and Jackline Gatotho. Water-Quality Variability and Constituent Transport in Streams of Johnson County, Kansas, Using Continuous Monitoring and Regression Models, 2003–11. Scientific Investigations Report, 2013–5221, USGS, 2014.

Recommended Human Health Recreational Ambient Water Quality Criteria or Swimming Advisories for Microcystins and Cylindrospermopsin. 822R19001, United States Environmental Protection Agency, May 2019, p. 249.

Rinta-Kanto, J. M., et al. “Quantification of Toxic Microcystis Spp. during the 2003 and 2004 Blooms in Western Lake Erie Using Quantitative Real-Time PCR.” Environmental Science & Technology, vol. 39, no. 11, June 2005, pp. 4198–205. DOI.org (Crossref), https://doi.org/10.1021/es048249u.

Rosero-López, D., et al. “Standardization of Instantaneous Fluoroprobe Measurements of Benthic Algal Biomass and Composition in Streams.” *Ecological Indicators*, vol. 121, Feb. 2021, pp. 1–10. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ecolind.2020.107185.Salmaso, Nico, et al. “Basic Guide to Detection and Monitoring of Potentially Toxic Cyanobacteria.” Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis, edited by Jussi Meriluoto et al., John Wiley & Sons, Ltd, 2017, pp. 46–69. DOI.org (Crossref), https://doi.org/10.1002/9781119068761.ch6.

Scott, J. Thad, and Amy M. Marcarelli. “Cyanobacteria in Freshwater Benthic Environments.” Ecology of Cyanobacteria II, edited by Brian A. Whitton, Springer Netherlands, 2012, pp. 271–89. DOI.org (Crossref), https://doi.org/10.1007/978-94-007-3855-3_9.

Sipari, Hanna, et al. “Development of a Chip Assay and Quantitative PCR for Detecting Microcystin Synthetase E Gene Expression.” Applied and Environmental Microbiology, vol. 76, no. 12, June 2010, pp. 3797–805. DOI.org (Crossref), https://doi.org/10.1128/AEM.00452-10.

Sivonen, Kaarina, and Gary Jones. Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring, and Management. Edited by Ingrid Chorus and Jamie Bartram, E & FN Spon, 1999.

“6040 TASTE AND ODOR COMPOUNDS.” Standard Methods For the Examination of Water and Wastewater, American Public Health Association, 2017, https://doi.org/10.2105/SMWW.2882.117.

Stuart, Rhona K., et al. “Cyanobacterial Reuse of Extracellular Organic Carbon in Microbial Mats.” The ISME Journal, vol. 10, no. 5, May 2016, pp. 1240–51. DOI.org (Crossref),

https://doi.org/10.1038/ismej.2015.180.

Thomson-Laing, Georgia, et al. “Broad and Fine Scale Variability in Bacterial Diversity and Cyanotoxin Quotas in Benthic Cyanobacterial Mats.” Frontiers in Microbiology, vol. 11, Feb. 2020, p. 129. DOI.org (Crossref), https://doi.org/10.3389/fmicb.2020.00129.

Vadeboncoeur, Yvonne, et al. “Blue Waters, Green Bottoms: Benthic Filamentous Algal Blooms Are an Emerging Threat to Clear Lakes Worldwide.” BioScience, vol. 71, no. 10, Oct. 2021, pp. 1011–27. DOI.org (Crossref), https://doi.org/10.1093/biosci/biab049.

Wagner, Nicole D., et al. “Nitrogen Form, Concentration, and Micronutrient Availability Affect Microcystin Production in Cyanobacterial Blooms.” Harmful Algae, vol. 103, Mar. 2021, p. 102002. DOI.org (Crossref), https://doi.org/10.1016/j.hal.2021.102002.

Walls, Jeremy T., et al. “Hot and Toxic: Temperature Regulates Microcystin Release from Cyanobacteria.” Science of The Total Environment, vol. 610–611, Jan. 2018, pp. 786–95. DOI.org (Crossref), https://doi.org/10.1016/j.scitotenv.2017.08.149.

Wang, Feier, et al. “Spatial Heterogeneity Modeling of Water Quality Based on Random Forest Regression and Model Interpretation.” Environmental Research, vol. 202, Nov. 2021, p. 111660. DOI.org (Crossref), https://doi.org/10.1016/j.envres.2021.111660.

Watson, Susan B., et al. “Taste and Odour and Cyanobacterial Toxins: Impairment, Prediction, and Management in the Great Lakes.” Canadian Journal of Fisheries and Aquatic Sciences, vol. 65, no. 8, Aug. 2008, pp. 1779–96. DOI.org (Crossref), https://doi.org/10.1139/F08-084.

Wood, Susanna A., Laura Biessy, et al. “Anatoxins Are Consistently Released into the Water of Streams with Microcoleus Autumnalis-Dominated (Cyanobacteria) Proliferations.” Harmful Algae, vol. 80, Dec. 2018, pp. 88–95. DOI.org (Crossref), https://doi.org/10.1016/j.hal.2018.10.001.

Wood, Susanna A., Mark W. Heath, et al. “Identification of a Benthic Microcystin-Producing

Filamentous Cyanobacterium (Oscillatoriales) Associated with a Dog Poisoning in New Zealand." Toxicon, vol. 55, no. 4, Apr. 2010, pp. 897–903. DOI.org (Crossref), https://doi.org/10.1016/j.toxicon.2009.12.019.

Wood, Susanna A., Laura T. Kelly, et al. "Toxic Benthic Freshwater Cyanobacterial Proliferations: Challenges and Solutions for Enhancing Knowledge and Improving Monitoring and Mitigation." Freshwater Biology, vol. 65, no. 10, Oct. 2020, pp. 1824–42. DOI.org (Crossref), https://doi.org/10.1111/fwb.13532.

Xue, Qingju, et al. "Seasonal Variation and Potential Risk Assessment of Microcystins in the Sediments of Lake Taihu, China." Environmental Pollution, vol. 259, Apr. 2020, p. 113884. DOI.org (Crossref), https://doi.org/10.1016/j.envpol.2019.113884.

Zarnaghsh, Amirreza, and Admin Husic. "Degree of Anthropogenic Land Disturbance Controls Fluvial Sediment Hysteresis." Environmental Science & Technology, vol. 55, no. 20, Oct. 2021, pp. 13737–48. DOI.org (Crossref), https://doi.org/10.1021/acs.est.1c00740.

Zupančič, Maša, et al. "Potentially Toxic Planktic and Benthic Cyanobacteria in Slovenian Freshwater Bodies: Detection by Quantitative PCR." Toxins, vol. 13, no. 2, Feb. 2021, p. 133. DOI.org (Crossref), https://doi.org/10.3390/toxins13020133.