Report No. 497

INVESTIGATING THE MICROBIAL CULPRITS OF TASTE/ODOR ISSUES IN CITY OF DURHAM DRINKING WATER RESERVOIR LAKE MICHIE AND ALGICIDAL MITIGATION TACTICS

By
Ryan Paerl[1], Hwa Huang[1] and Linda Ehrlich[2]

1. Department of Marine, Earth and Atmospheric Sciences
   North Carolina State University
   Raleigh, North Carolina

2. Spirogyra Diversified Environmental Services
   Burlington, North Carolina

wrri.ncsu.edu
NC WATER RESOURCES RESEARCH INSTITUTE

UNC-WRRI-497

The research on which this report is based was supported by funds provided by the North Carolina General Assembly and/or the US Geological through the North Carolina Water Resources Research Institute.

This report fulfills the requirements for a project completion report of the North Carolina Water Resources Research Institute. This report has not been peer reviewed. The authors are solely responsible for the content and completeness of the report. Completion of this grant requirement in no way impacts the authors' ability to publish final peer reviewed results.

WRRI Project No. 20-08-W
January 31, 2022

wrri.ncsu.edu
NC WATER RESOURCES RESEARCH INSTITUTE

WRRI Final Project Report

Title: Investigating the Microbial Culprits of Taste/Odor Issues in City of Durham Drinking Water Reservoir Lake Michie and Algicidal Mitigation Tactics

Investigators: Dr. Ryan W. Paerl* - Assistant Professor; Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC
Hwa Huang - PhD graduate student; Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC
Dr. Linda Ehrlich – Spirogyra Diversified Environmental Services, Burlington, NC

***PI contact information:**
Jordan Hall 4135
2800 Faucette Drive
Raleigh, NC 27695-8208
rpaerl@ncsu.edu

WRRI project number: 20-08-W

Date of report: 01/31/2022

Abstract:

Project title: Investigating the Microbial Culprits of Taste/Odor Issues in City of Durham Drinking Water Reservoir Lake Michie and Algicidal Mitigation Tactics. Taste and odor (T&O) compounds, especially geosmin and 2-methylisoborneol (2-MIB) are of concern to water managers worldwide and those serving large cities within the Piedmont region of North Carolina (NC). **Project objectives**: 1) Identify the taste/odor producing populations in Lake Michie (LM) by microscopy or flow cytometry surveying, genetic data, and/or correlation between hydrologic, biochemical conditions and T&O compound concentrations; **2)** Evaluate if flux of T&O compounds from cells can be minimized by reduced dosages of already used algicide EarthTec. **Methods:** Archived data for LM from City of Durham Water Management (CDWM) were examined to determine longer-term (6 year) trends in biochemical and physical conditions along with T&O compound concentrations, as well as contextualize higher resolution sampling in 2020. LM water was collected in 2020 from 2x the Secchi depth for multiple chemical and biological analyses - including the collection of biomass from two different size fractions (>8µm and 8-0.8µm) for DNA extraction and downstream PCR amplicon or metagenomic sequencing. Samples were also collected for microscopic analysis of plankton diversity and biovolume, as well as flow cytometry-based cell counts (phytoplankton and bacteria). Bottle incubations were used to assess the efficacy of the $CuSO_4$ containing algaecide EarthTec in reducing concentrations of geosmin and 2-MIB. **Results:** Geosmin reached higher average and maximum concentrations in LM samples in the 6 years of sampling LM and is increasing over time, although datapoints supporting the trend are limited. 2-MIB also exceeded the human detection threshold of 10 ng/L in 5 of 6 years and remains a T&O concern. Geosmin often peaked earlier than 2-MIB in late spring into early summer. *Dolichospermum* biomass and relative abundance of *geoA* sequences increased in parallel with geosmin in our intensive 2020 sampling campaign. Metagenome assembled genomes (MAGs) also point to *Dolichospermum* populations as the dominant geosmin producers, while also uncovering *Chloroflexota* populations as geosmin sources which has not been previously reported. One putative 2-MIB producing *Microcoleus* (N-fixing cyanobacterial genus) population was uncovered based on metagenomic sequencing. **Conclusions & recommendations:** Filamentous N-fixing cyanobacteria (esp. *Dolichospermum* spp., putative *Microcoleus*) are key sources of T&O to LM. Previously unrecognized *Chloroflexota* are sources of geosmin and their contribution of the compound to the greater lake, and other lakes, requires further investigation. Increases in the relative abundance of *Dolichospermum geoA* sequences paralleled increases in geosmin, thus PCR-based assays hold potential as a rapid geosmin indicator tool. Overall, geosmin concentrations are on an upward trajectory and this is likely to continue with continued increases in average air temperature in the southeastern US and low total N to P ratios in LM that favor N-fixing geosmin producers. Reductions in P may help limit the growth of T&O producing N-fixers, but likely will require substantial efforts and/or financial investment (e.g. dredging). EarthTec treatment of LM water at concentrations previously used by CDWM often caused increases in geosmin (and 2-MIB at times) in our experiments. Lower EarthTec additions (ca. 0.0125 mg/L $CuSO_4$) appear more effective to limit algal growth and T&O concentrations and are recommended.

Acknowledgments

We wish to acknowledge: Bobby Whisnant as well as Jimmy Gamble and his field collection team at the City of Durham Water Management (CDWM) Department for collection of Lake Michie water and sharing of CDWM archived data, Cara Rusmisell, Gina Kimble and MaryAnne Jones at Charlotte Water for measurements of geosmin 2-MIB via gas chromatography - mass spectrometry (GC/MS) and assistance with sample shipments, Jason Whitham (NCSU) for bioinformatic assistance, Kemoni Hughes (North Carolina Central University) for assistance with flow cytometry counts, Mark Elmore – Lake Johnson Park Manager and his staff for hosting Mobile Aquatic Microbial Laboratory (MAML) outreach events, and Joel Sanchez (NCSU) for general laboratory assistance.

1. Introduction

Taste and odor (T&O) compounds produced by bacteria, esp. geosmin and 2-methylisoborneol (2-MIB) (Fig. 1), are a significant concern globally as they indicate to consumers that water is poor quality (whether true or not). Geosmin and 2-MIB in particular and are readily detected by humans in minute quantities, e.g. ≤10 ng/L while also are considered non-toxic (Young et al. 1996; Izaguirre and Taylor 2004; Jüttner and Watson 2007; Watson et al. 2016). Beyond tainting water and air, high T&O production (esp. geosmin) can impact the taste of shellfish and finfish consumed from aquatic systems (Guttman and van Rijn 2008; Houle et al. 2011). Arguably T&O compounds have garnered greater attention from regions of the US where cyanobacterial blooms are well-documented and overt (Otten et al. 2016; Harris and Graham 2017). Diverse cyanobacteria, both filamentous and unicellular, are believed to produce geosmin and/or 2-MIB, but non-photosynthetic bacteria, esp. Actinobacteria, are also well-studied T&O compound producers (Jüttner and Watson 2007; Watson et al. 2016).

Notably few published studies describe T&O compound concentrations, dynamics, or sources within southeastern US reservoirs (van der Ploeg et al. 1992; Izaguirre and Taylor 2004). Recently surveys of two municipal reservoirs (Lake Bowen, Municipal Reservoir #1) in the Piedmont region of South Carolina focus more on geosmin due to periodic exceedances of 10 ng/L relative to 2-MIB (Journey et al. 2011, 2013). Moreover, *Oscillatoria* or other known geosmin producing cyanobacterial genera exhibited the strongest correlations with geosmin concentrations, rather than total cyanobacterial biovolume (biomass) (Journey et al. 2011, 2013).

In North Carolina (NC), water management groups are particularly concerned about T&O compounds from the perspective of drinking water quality. The City of Durham Water Management (CDWM) is one such group, as it serves >260,000 customers in Durham Co. and receives annual complaints about T&O from customers - with highest complaint numbers during spring and summer months (per. comm. J. Gamble, CDWM).

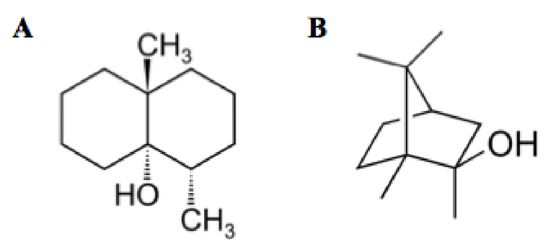


Figure 1. Chemical structure of geosmin (A) and 2-methylisoborneol (2-MIB).

Lake Michie and Little River are the primary reservoirs used by CDWM to meet customer drinking water demands. Both are similar in size (507 versus 530 acres) and are considered 'deep water' piedmont reservoirs with cycles of stratification and turnover occurring annually. Lake Michie (LM) is the older of the two, est. in 1926 and is part of the Neuse River watershed and is fed primarily by the Flat River to the northeast of the reservoir (City of Durham Water Management 2007). LM is also a notable hub for recreational activity, including fishing, boating, camping, and hiking, within the Piedmont and adjacent to the Research Triangle (Durham Parks & Recreation 2022).

Notably, microbial sources (e.g. bacteria or phytoplankton) of T&O compounds in NC reservoirs and LM are not well described. A diversity of photosynthetic bacteria (cyanobacteria - largely filamentous spp.) and non-photosynthetic bacteria (Deltaproteobacteria, Gammaproteobacteria, and Actinobacteria) are now recognized as producers of geosmin and/or 2-MIB (Watson et al. 2016; Churro et al. 2020). An advanced ability to identify T&O producers has come with increased cultivation efforts as well as genetic sequencing. In regard to the latter, the identification of individual genes key in coding for geosmin synthase and 2-MIB synthase (*geoA*, *mic*; (Ludwig et al. 2007; Giglio et al. 2008)) enable the identification of these same genes (and thus T&O production capacity) amongst complex mixtures diverse populations present in natural samples (Otten et al. 2016).

In LM, filamentous cyanobacteria are likely sources as they are common in both reservoirs (per. comm. Linda Ehrlich), and many (yet not all) produce geosmin or 2-MIB in cultures or possess genes that code for production of either compound (Watson et al. 2016; Wang et al. 2019; Churro et al. 2020). Few monitoring efforts include evaluating taxa specific data due to the labor involved in light microscopy and even fewer use flow cytometry for monitoring. Rather, chlorophyll a (conc. or fluorescence) is typically monitored and often yields no significant correlations with T&O compounds (Journey et al. 2011, 2013; Cain 2017).

Related, a collaborative pilot study in 2018 with Linda Ehrlich (Spirogyra Inc.) and CDWM was performed to examine if flow cytometry and picophytoplankton numbers could be indicative of increased T&O concentrations in Lake Michie (LM). The results showed that the highest net growth of picocyanobacteria at 2x the secchi depth occurred alongside maximum geosmin concentrations – suggesting picocyanobacteria have a role in geosmin production (directly or indirectly via stimulation of co-occurring geosmin producers) (Fig.2) . In contrast, filamentous cyanobacteria increased later in the summer where a lower geosmin peak occurred, but MIB concentrations were elevated (Fig. 2). Reports of geosmin production by picocyanobacteria are in the scientific literature, but evidence of this is much more sparse than evidence of filamentous cyanobacteria as key sources of T&O (Watson et al. 2016; Godo et al. 2017). Thus, identifying a picocyanobacteria source of T&O compounds in LM would be a notable discovery. In general, there is much more evidence that filamentous cyanobacteria, and at times actinomycetes, are important sources of T&O compounds to drinking water reservoirs and treatment pipelines (Zaitlin et al. 2003; Izaguirre and Taylor 2004; Jüttner and Watson 2007; Journey et al. 2013; Lukassen et al. 2017).

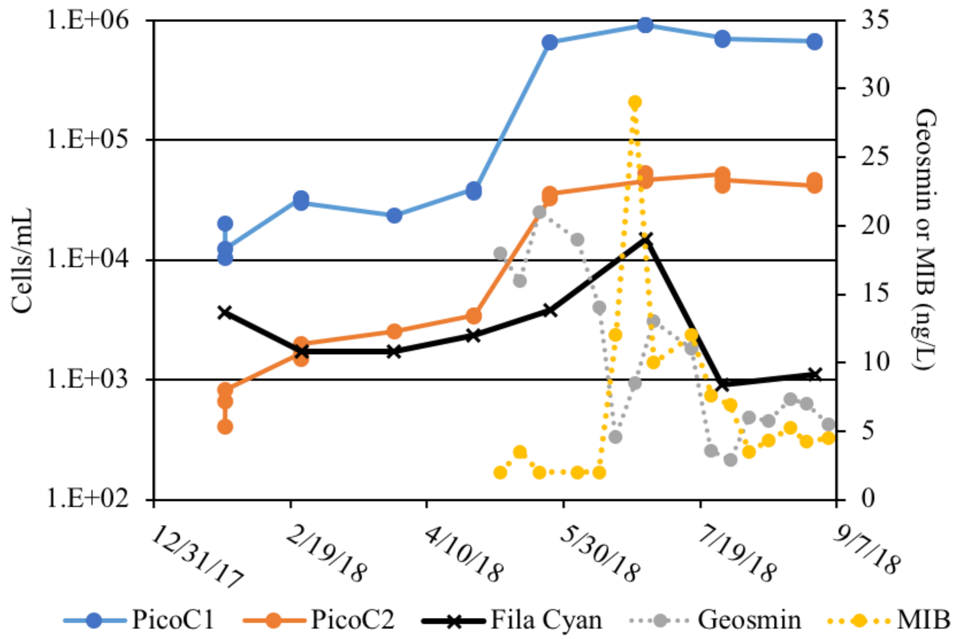


Figure 1. Abundances of two dominant picocyanobacteria (PicoC) groups, filamentous cyanobacteria, and geosmin concentrations in euphotic LM waters in 2018. Water was collected from 2x the Secchi depth adjacent to the LM dam.

Presently, NC water management groups, including CDWM, speculate on the sources of spikes in odor/taste compounds. Without knowledge of the targets and controls on taste/odor compound production, water management groups are unable to effectively anticipate future conditions and take fiscally efficient action to combat T&O issues. Instead, they are forced to implement year-round to semi-annual treatments (e.g. copper sulfate treatments) that are costly, possibly sub-optimally effective at reducing T&O concentrations, and introduce elevated heavy metals and possibly cyanobacterial lysate (including cyanotoxins) into the treated system (Kibuye et al. 2021). In the case of CDWM, a copper sulfate-based algaecide is used during the phytoplankton growing season (May – Oct) to reduce algal biomass and concentrations of T&O (per. comm. J. Gamble CDWM). Notably, copper sulfate is known to lyse cells (Gibson 1972; Fan et al. 2013) which would likely liberate T&O compounds into the dissolved phase. Reduced treatments of copper sulfate algaecide may halt cell growth, keep T&O localized to cells for capture in the downstream ferric sulfate aggregation step of treatment, and overall lower T&O concentrations in the dissolved phase.

This project focused on addressing the problematic unknowns above by identifying: **1)** taste/odor producing populations (by cell type surveying and genetics) and **2)** evaluating if flux of T&O compounds from cells can be minimized by reduced dosages of already used algicide. In addition, historical data collected by key collaborator Dr. Linda Ehrlich (Spirogyra Diversified Environmental Services) and colleagues at CDWM from LM were synthesized in order to summarize conditions, change in T&O compound concentrations, and context to our more intensive 2020 study.

2. Methods

Discrete water samples were collected via Van Dorn bottle sampler from 2x the Secchi depth from the LM dam, just west of the spillway (Fig. 3). Water samples were collected weekly (in some cases +/- a day due to weather or holidays) between the months of April and October in 2020. Collected water was drained into cleaned polycarbonate bottles or used to fill volatile organic carbon (VOC) vials by CDWM for measurements of geosmin and 2-MIB. Filled VOC vials without bubbles were shipped overnight on ice to Eurofins for chemical measurement using the 'Taste and Odor Compounds by GC/MS Method V210'. Turbidity, pH, alkalinity, and dissolved O2 were measured by CDWM and provided by J. Gamble (CDWM) for statistical analyses.

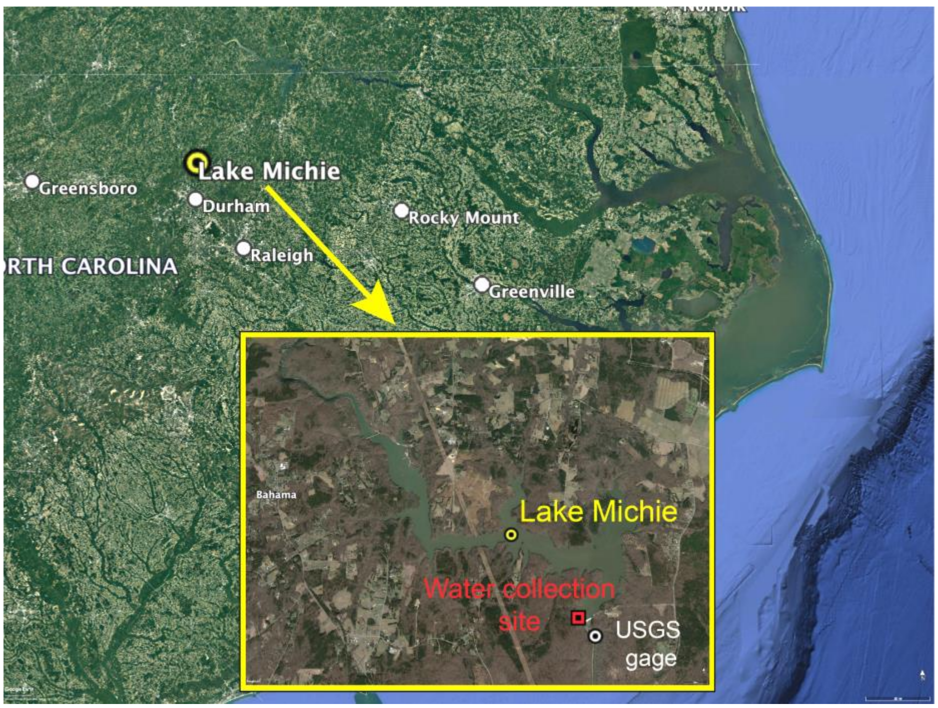


Figure 3. Location of Lake Michie and a magnified inset map showing the location of water collection for this study as well as the proximal USGS gauge 02086500. Google Earth Pro was used for map generation.

Polycarbonate bottles with LM water were kept in a cooler and transported to the Williams water treatment plant, and then transported to Jordan Hall at NC State University for further processing the same day. At Jordan Hall, samples were fixed with 0.25% final EM grade glutaraldehyde and stored at -80℃ for flow cytometric counting. Other aliquots of LM water were preserved with Lugol's Iodine solution and stored at room temperature in the dark in amber plastic bottles for further microscopic analysis. Aliquots were also used for collecting biomass upon GF/F filters for Chlorophyll a extractions or nitrocellulose filters for DNA extraction. Last, on select collection days LM water was used for incubations with or without (in the case of controls) additions of EarthTec algaecide.

Flow cytometry counts of small phytoplankton were obtained using a Guava EasyCyte HT equipped with a blue and red excitation laser (Paerl et al. 2020b). Bacterial counts were also obtained by flow cytometry using SYBR green I staining, without any heating of the stained sample (Brussaard et al. 2010).

LM samples were examined by microscopy by Dr. Linda Ehrlich (Spirogyra Diversified Environmental Services Inc.) in order to identify common populations and respective biovolumes. For microscopic analysis of plankton, Lugol's preserved samples were examined with an Accuscope EXI-310 inverted compound light microscope. Prior to quantitative analysis, aliquots of the whole water samples were concentrated by sedimentation. Each whole water sample was mixed by inversion 100 times, after which an aliquot was decanted into an appropriately sized Utermöhl sedimentation chamber (Utermöhl 1958; Lund et al. 1958). Concentrated aliquots were transferred to a Sedgewick-Rafter counting cell for microscopic examination. Identifications and enumerations were conducted at 40X magnification with an Accuscope EZI-310 inverted compound microscope equipped with a calibrated Whipple reticule. Cellular dimensions were measured with the calibrated Whipple reticule. Enumerations were performed according to the Sedgewick-Rafter strip counting method (Baird, R., & Bridgewater 2017). Data were entered directly into a Microsoft Excel spreadsheet for conversion of raw counts to numbers of units or cells per milliliter.Identifications were made with standard taxonomic references. Cellular biovolumes were estimated by applying the cellular dimensions to the most closely matching solid geometric formula (Hillebrand et al. 1999).

Biomass vacuum-filtered on GF/F filters, respective filtrate, as well as raw water were all stored at -20℃ for nutrient analysis. Frozen samples were transported to the lab of H. Paerl at UNC-Chapel Hill Institute of Marine Sciences for elemental analysis of particulate CHN (elemental analysis), as well as colorimetric measurements of dissolved NOx, NH4, PO4, and SiO2. Total P and total dissolved N (TDN) were also determined. All measurements were made using previously published protocols, which are commonly employed by the MODMON monitoring project (Wetz and Paerl 2008; Gaulke et al. 2010).

Chlorophyll a (Chl a) was extracted from biomass filtered on GF/F filters using the acetone, non-acidification method (Welschmeyer 1994). The GF/F filters were wrapped in aluminum foil with biomass inward and stored at −20 °C. Chl a was extracted from biomass on GF/F filters using acetone (90%) and sonication.

Discharge and gauge height data were retrieved from a given sampling day (daily average) the USGS National Water Information System: Web Interface, and specifically the gauge USGS 02086500 FLAT RIVER AT DAM NEAR BAHAMA, NC. Air temperature data from 2015-2020 was retrieved from Weather Underground (www.weatherunderground.com) from RDU International Airport Station (KRDU) as a representative station for LM air temperature conditions. The average air temperature for the day of LM sampling and the two days prior were collectively averaged, and that temperature was matched with other LM sample data for that day as well as in downstream analyses.

Archival CDWM and Spirogyra Inc. data from LM monitoring (2015-2019) were used for contextualizing results from 2020 and statistical analyses. Samplings in these years were fewer in number but generally spanned the months of May to October and were collected weekly to biweekly.

Geosmin and 2-MIB were measured in EarthTec addition experiments by Charlotte Water (Charlotte, NC) using the Gerstel 'Twister' stir bar sorptive extraction (SBSE) method (SM 6040D). Water from EarthTec experiments was collected in VOC bottles (as described above) and transported on ice or with frozen cold packs to Charlotte Water.

Biomass was harvested from water onto filters by serial vacuum filtration for downstream DNA extraction. LM water was filtered through a 8 µm pore-sized 47 mm nitrocellulose (Millipore) filter and then the filtrate was filtered through a 0.8 µm pore-sized 47 mm nitrocellulose filter. Vacuum dried filters were immediately folded and stored in microcentrifuge tubes at -80℃. Bacterioplankton smaller than 0.8 µm diameter are not considered sources of T&O compounds or possess genes coding for T&O compounds, for example members of the Actinobacterial clade acI (Neuenschwander et al. 2018). Thus, 0.8 µm pore-sized filters were used in order to enrich for bacterial cells that may produce T&O compounds, including picocyanobacteria, while allowing common and small non-producers to pass through the filter. DNA was extracted from filters with biomass using the Qiagen PowerSoil kit.

A one-gallon bottle of Earthtec (Earth Science Laboratories Inc.) was purchased in early 2020 and used for bottle incubations where LM water was treated with EarthTec and varying concentrations. Treatments were from 0.0005 mg/L to 0.05 mg/L - the highest being the concentration used previously by CDWM (per comm. Jimmy Gamble). Untreated water samples served as negative controls and were incubated under the same condition as treated water. LM water was incubated in 10% HCl washed and thoroughly rinsed (with tap and then DI water) 150 mL clear polycarbonate bottles under white light (~100 µE m^{-2} s^{-1}) at near ambient temperature (+/- ~3℃) within a climate-controlled room. Incubations were run for 48h and bottles were slightly agitated (swirled by hand) 1-2 times daily.

Previously published primers for amplification of *geoA* and *mic* genes (Kutovaya and Watson 2014; Wang et al. 2016) were used in an attempt to amplify the respective genes from extracted DNA. CS1 and CS2 linker sequences were included on the ends of the originally published forward and reverse primers (Naqib et al. 2018). Kappa polymerase was used in reactions and 10-12 ng of sample DNA was added to each reaction tube. Published cycling conditions were used in PCR reactions for *geoA* amplification, which were 95℃ for 5 minutes, 30 cycles of 94℃ for 1 minute, annealing at 56℃ with 0.5℃ decrease per cycle for 1 min, 72℃ for 1 min, and finally, extension at 72℃ for 12 minutes (Kutovaya and Watson 2014). For the *mic* amplification, the cycling conditions were 94℃ for 3 minutes, 35 cycles of 94℃ for 30 seconds, annealing at 59℃ for 30 seconds, and 72℃ for 1 minute (Wang et al. 2016). PCR product was visualized by electrophoresis using an 1% agarose gel. Nonspecific amplification occurred in *geoA* and *mic* reactions, however *geoA* PCR also produced a PCR product of an expected size in unknowns and controls (DNA from Maynard Pond - confirmed as enriched in geosmin by Charlotte Water GC/MS measurements). Thus, the desired *geoA* band (227bp) was excised via razor blade from stained agarose gels using a gel illuminator. A ZymoClean Gel DNA Recovery kit was used to purify DNA from excised gel 'plugs'.

Sequencing of *geoA* PCR amplicons was done at the Rush University Genomics and Microbiomes Core Facility (GMCF), using the Illumina MiniSeq platform (2x150bp). Sequences were denoised and analyzed using the DADA2 pipeline (Callahan et al. 2016). Phyloseq was used to analyze amplicon sequence variant tables and generate relative abundance plots (McMurdie and Holmes 2013). Other plotting and statistical analyses were done using R (R Core Team 2014) as well as packages vegan, ggvegan and mgcv (Oksanen et al. 2013; Simpson 2015; Wood 2017). Transformation based redundancy analyses (tb-RDA) was performed using Box Cox transformed geosmin and 2-MIB data and z-score transformed environmental data.

Shotgun (metagenomic) sequencing of select LM DNA extracts from 8 µm and 0.8 µm filters was done at the Genomic Science Laboratory at NC State University using the NovaSeq S4 Illumina 2x150bp platform, targeting ~2.5 billion reads per lane. The DNA extracts selected were from periods of elevated and reduced concentrations of geosmin and 2-MIB during 2020.

Kbase and multiple software modules within it were used to perform quality control screening of sequence data as well as multiple sequence analyses (Arkin et al. 2018). This includes quality control screening of reads (JGI RQCFilter pipeline (BBTools v38.22) (Bushnell 2014)), contig assembly (MEGAHIT v1.2.9 (Li et al. 2015)), annotation (RASTtk - v1.073 (Brettin et al. 2015)), construction of metagenome assembled genomes (MAGs) or bins from individual samples (METABAT2 (Kang et al. 2019)), and taxonomic assignment of reads (Kaiju; (Menzel et al. 2016)). Geosmin synthase genes (*geoA*) and 2-MIB genes (*mic*) were recovered from contigs using BLASTn (Altschul et al. 1990) and in-house scripts. Sequences used for BLASTn queries are provided in Appendix 1. Positive hits were manually examined via antiSMASH (https://antismash.secondarymetabolites.org/; (Blin et al. 2021)) and BLASTn (including discontiguous megablast - for more distant reads) searching against the NCBI NR database (Altschul et al. 1990).

3. Results

3.1 Environmental conditions at Lake Michie 2015-2020

Multiple environmental factors could impact microbial producers of T&O compounds; thus, several datasets of environmental variables were examined during 2015-2020 in parallel with T&O concentration data. Air temperature near LM (RDU airport weather station data) during the 2015-2020 sampling period was on average 23.1℃ and showed minor deviations from this value annually (Table 1). Highest maximum and lowest minimum temperatures during sampling times occurred in 2019. Notably, a greater number of samples were collected from 2019 and 2020 (28 and 30 respectively), including samples collected in late October and November (Table 1). Air temperature at time of water collection from LM was overall higher than the average for year within Durham County (ca. 16℃) (NOAA 2022).

Table 1. Average temperatures during LM samplings as well as maximum and minimum temperatures for each year.

Year	LM Samplings	Avg temp (∘C)	Max temp (∘C)	Min temp (∘C)
2015	12	23.1	28.6	17.1
2016	23	23.8	28.6	16.4
2017	19	22.9	27.4	13.9
2018	20	24.8	28.4	19.4
2019	28	23.1	30.1	9.0
2020	30	21.2	28.6	12.6
AVG		23.1	28.6	14.7

External inputs of water to LM, reflected by annual discharge rate out of LM, varied by upwards of 4-fold between 2015-2020, with minimal (dry year) and maximal (wet year) annual discharge occurring in 2015 and 2019 respectively (Table 2). Discharge values from discrete sampling days from LM dam generally match the dry and wet conditions reflected in the annual data; however, in 2019 discharge rates were notably lower during LM samplings, than reflected in the annual value. This discrepancy is due to high discharge in early 2019 (January to April) when samples were not collected.

Table. 2. Flow conditions from USGS gauge 02086500 FLAT RIVER AT DAM NEAR BAHAMA, NC. Annual data were directly obtained from USGS web resources. Other data were manually obtained from dates where a corresponding LM sample was collected. NA = Data not available from USGS.

Year	Annual Mean Discharge (ft3/s)	Annual Mean Gauge Height (ft)	LM Samplings	Avg Discharge (ft3/s)	Stdev discharge (ft3/s)	Avg gauge height (ft)	Stdev gauge height (ft)
2015	77.5	1.139	12	13.09	37.75	0.61	0.46
2016	224.4	1.921	23	140.09	282.92	1.64	0.96
2017	161.3	1.497	19	21.61	30.73	0.76	0.46
2018	137.4	NA	20	69.81	93.84	1.26	0.69
2019	338.1	2.108	28	19.95	33.89	0.79	0.48
2020	178.1	1.523	30	131.59	311.18	1.43	1.07
AVG	186.13	1.64		66.02	57.79	1.08	0.42

Table 3. Turbidity measurements collected from 2015-2020.

Year	Samplings	Avg turbidity (NTU)	Max turbidity (NTU)	Min turbidity (NTU)
2015	12	5.42	20.4	1.75
2016	21	4.05	11	1.7
2017	23	6.97	50.8	1.21
2018	19	6.77	31.8	0.704
2019	26	3.19	8.75	0.821
2020	29	4.70	14.9	1
AVG	21.7	5.19	22.94	1.20

3.2 T&O concentrations 2015-2020

During 2015 to 2020, geosmin concentrations annually exceeded the human detection threshold (10 ng/L) and highest annual concentrations occurred in the spring but also during early summer, i.e. calendar day ~130 to ~210 (Fig. 5). The average annual geosmin concentration for the period of 2015-2020 was 11.8 ng/L, with a maximum of ≥35 ng/L reached every year except 2018 (21 ng/L). The highest geosmin concentration detected was in spring 2019 and reached 130 ng/L, ca. 10x the human detection limit. Punctuated peaks in geosmin concentration occurred during spring and early summer across 2015-2020. In several instances geosmin concentration fluctuated dramatically between weekly samples, e.g. ~130 ng/L to ~30 ng/L in early 2019. In some years dual peaks occurred in spring as well as early summer. Exceedances of 10 ng/L later in the summer and early fall, i.e. after calendar day ~230, were rate occurring two out of the six years and were not >15 ng/L (Fig. 5).

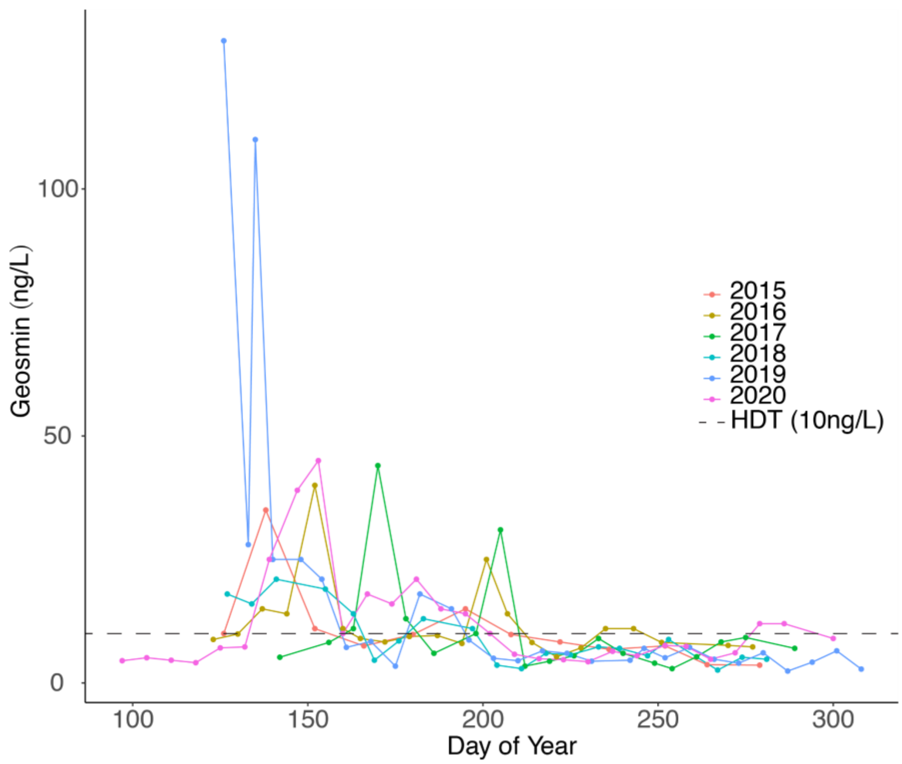


Figure 5. Geosmin concentrations from LM samplings from 2015 to 2020. The dashed line marks the 10 ng/L concentration level.

Concentrations of 2-MIB in LM also exceeded the human detection threshold (~10 ng/L) during 2015-2020, and were only slightly less frequent than geosmin exceedances, occurring in five of the six years (Fig. 6). The average 2-MIB geosmin concentration was 5.9 ng/L for 2015-2020 and a maximum ≥16 ng/L occurred every year except for 2019 (9.1 ng/L). Contrasting with geosmin, 2-MIB concentrations were often higher in summer (calendar day ~190-250), with 2018 as the exception where concentrations reached 30 ng/L on day 170 (the highest concentration in the dataset). Singular, yet broad (covering more time) peaks in 2-MIB were more common and fluctuation in the concentration of 2-MIB was generally less pronounced on a week to two-week time scale compared to geosmin (Fig. 5), e.g. 2-MIB maximally changed by ca. 3-fold before and after annual peak in 2018 (Fig. 6). Two peaks in 2-MIB uniquely occurred in 2020, with a late peak in fall (ca. day 290) not exceeding 20 ng/L (Fig. 6).

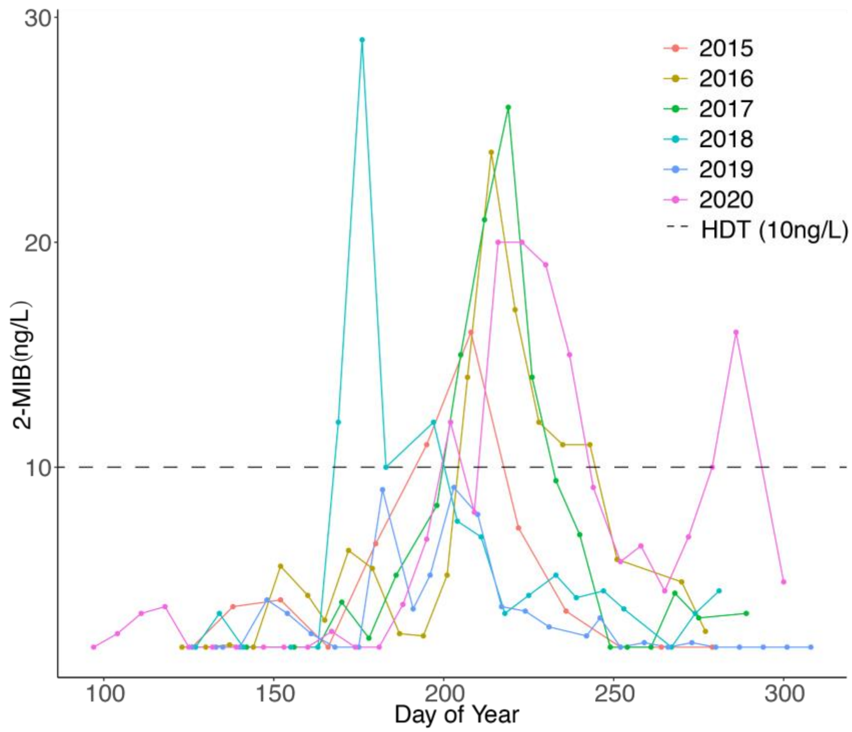


Figure 6. 2-MIB concentrations from LM samplings from 2015 to 2020. The dashed line marks the 10 ng/L concentration level.

Table 4. Geosmin and 2-MIB concentrations from 2015-2020 samples from LM.

Year	LM Samplings	Avg geosmin (ng/L)	Max / min geosmin (ng/L)	Avg 2-MIB (ng/L)	Max / mix 2-MIB (ng/L)
2015	12	10.7	35 / 3.6	5.2	16 / 2
2016	21	11.8	40 / 5.3	6.9	24 / 2
2017	19	10.2	44 / 2.9	7.1	26 / 2
2018	20	9.2	21 / 2.6	6.2	29 / 2
2019	28	17.2	130 / 2.4	3.3	9.1 / 2
2020	29	11.6	45 / 4.1	6.9	20 / 2
AVG		11.8		5.9	

The average annual geosmin concentration in LM samples shows an increasing trend from 2015-2020. Conversely, the annual average 2-MIB is slightly decreasing (Fig. 7). Of note, the linear fits of the data used to assess trends over time are not statistically supported (n=6; not normally distributed data) thus trends in the limited yet valuable dataset should be interpreted with caution.

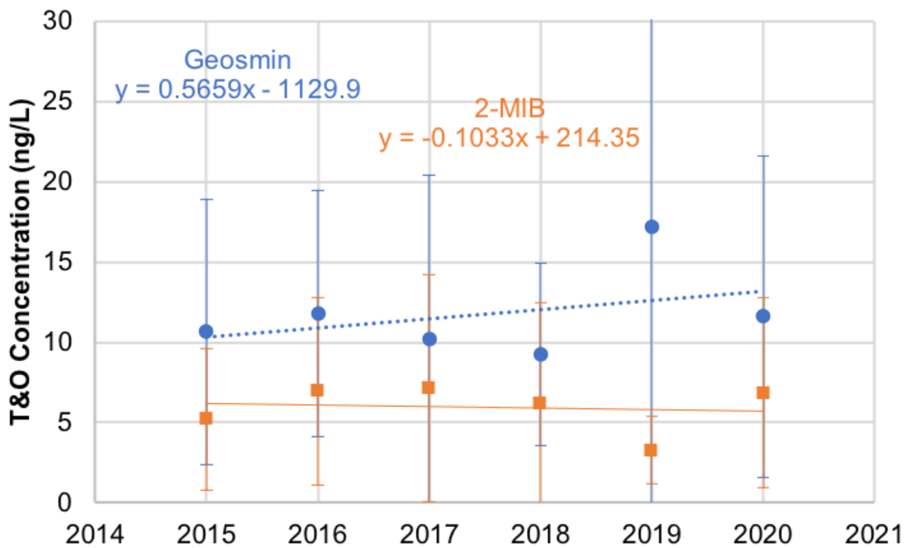


Figure 7. Average concentrations of geosmin and 2-MIB each year from 2015 to 2020. Error bars represent the standard deviation amongst concentrations recorded each year. Linear regression lines are plotted to convey trends in the annual average concentration values, but without r^2 as the data is not normally distributed (n=6).

3.3 LM sampling in 2020

Chlorophyll a (Chl a) concentrations were determined as part of 2020 sampling campaigns and ranged from ~3 to 35 µg Chl a/L in 2020 with lowest concentrations in periods of early spring and late summer as well as early fall, and highest concentrations in late spring early summer (ca. Day 160) (Fig. 8). Chl a did not exceed 40 µg Chl a/L, which is a criterion for reduced water quality within a water body in NC (NC DEQ (North Carolina Department of Environmental Quality) 2019). No parallel change in Chl a and geosmin or 2-MIB concentrations was evident across the 2020 dataset (Fig. 8).

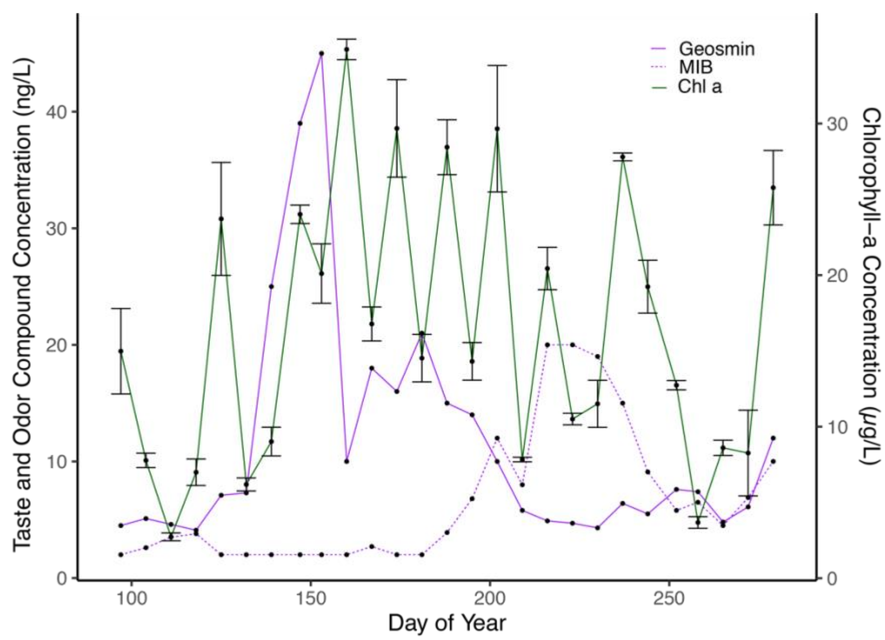


Figure 8. Change in Chl a concentration, as well as geosmin and 2-MIB concentration, in 2020 LM samples.

Notable differences in cyanobacterial biovolume are evident between years at similar periods (calendar days) of spring and summer (Fig. 9A). The weekly changes in biovolume in spring and early summer of 2020 highlight variability in cyanobacterial biovolume within LM and also the value of weekly monitoring. Likely spikes in cyanobacterial biovolume were not captured in other years due to the monthly monitoring scheme – especially in 2019 where geosmin concentrations reached >100 ng/L (Figs. 5, 9). Strong overlap between total cyanobacterial biovolume and *Dolichospermum* biovolume as well as geosmin concentration occurred in spring to early summer 2020 (Fig. 9). This highlights *Dolichospermum* spp. accounted for large portions of cyanobacterial biomass and points to the genus as an important source of geosmin. Sampling later in the year was more limited but markedly during an increase in 2-MIB the dominant cyanobacterial biovolume shifted to primarily *Planktothrix* (2875 mm^3/m^3), *Arthrospira* (2670 mm^3/m^3), and *Limnothrix* spp. (390 mm^3/m^3) (Fig. 8) – two of which are represented by 2-MIB producing strains in culture (Su et al. 2015; Kim et al. 2020) (Fig. 9). Examining linear relationships between *Dolichospermum* biovolume or Chl a concentration and T&O compound concentrations shows a strongest positive linear correlation between *Dolichospermum* biovolume and geosmin concentration (Fig. 10).

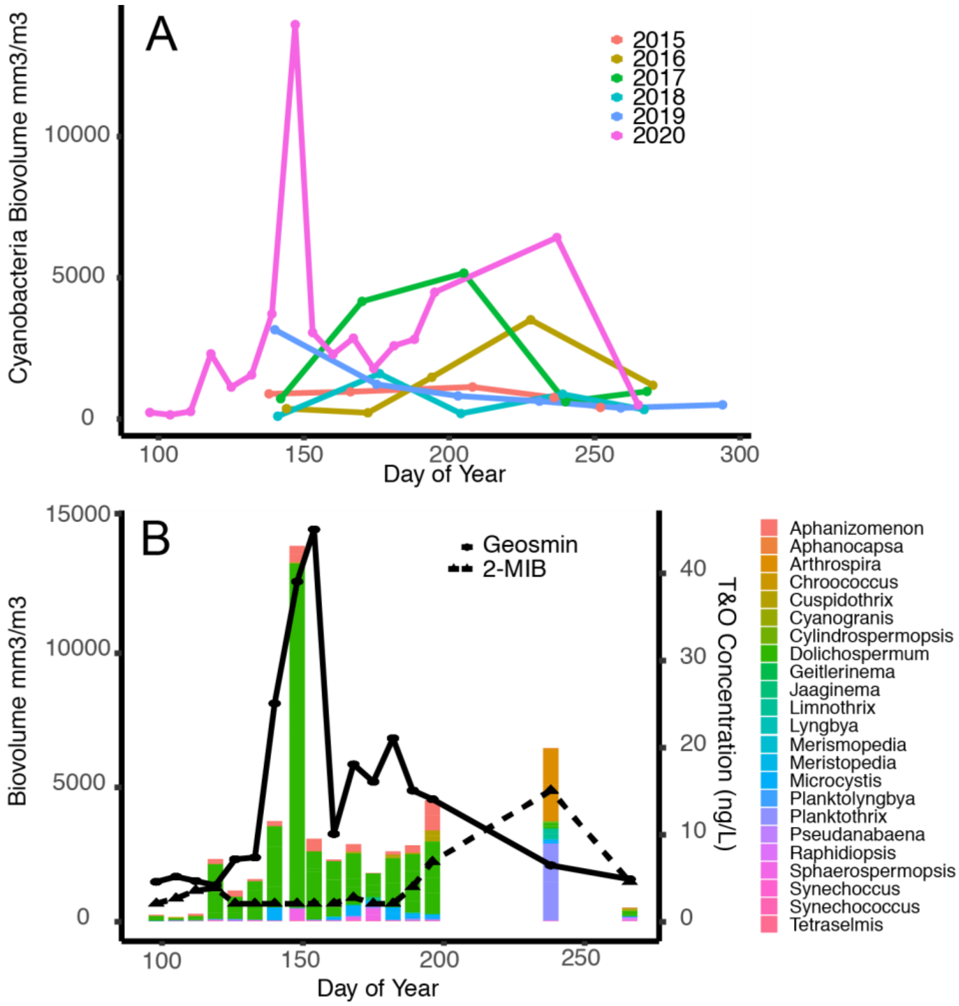


Figure 9. (A) Estimates of cyanobacterial biovolume from 2015-2020 microscopy samples and (B) genus specific biovolume estimates for 2020 as well as T&O compound concentrations.

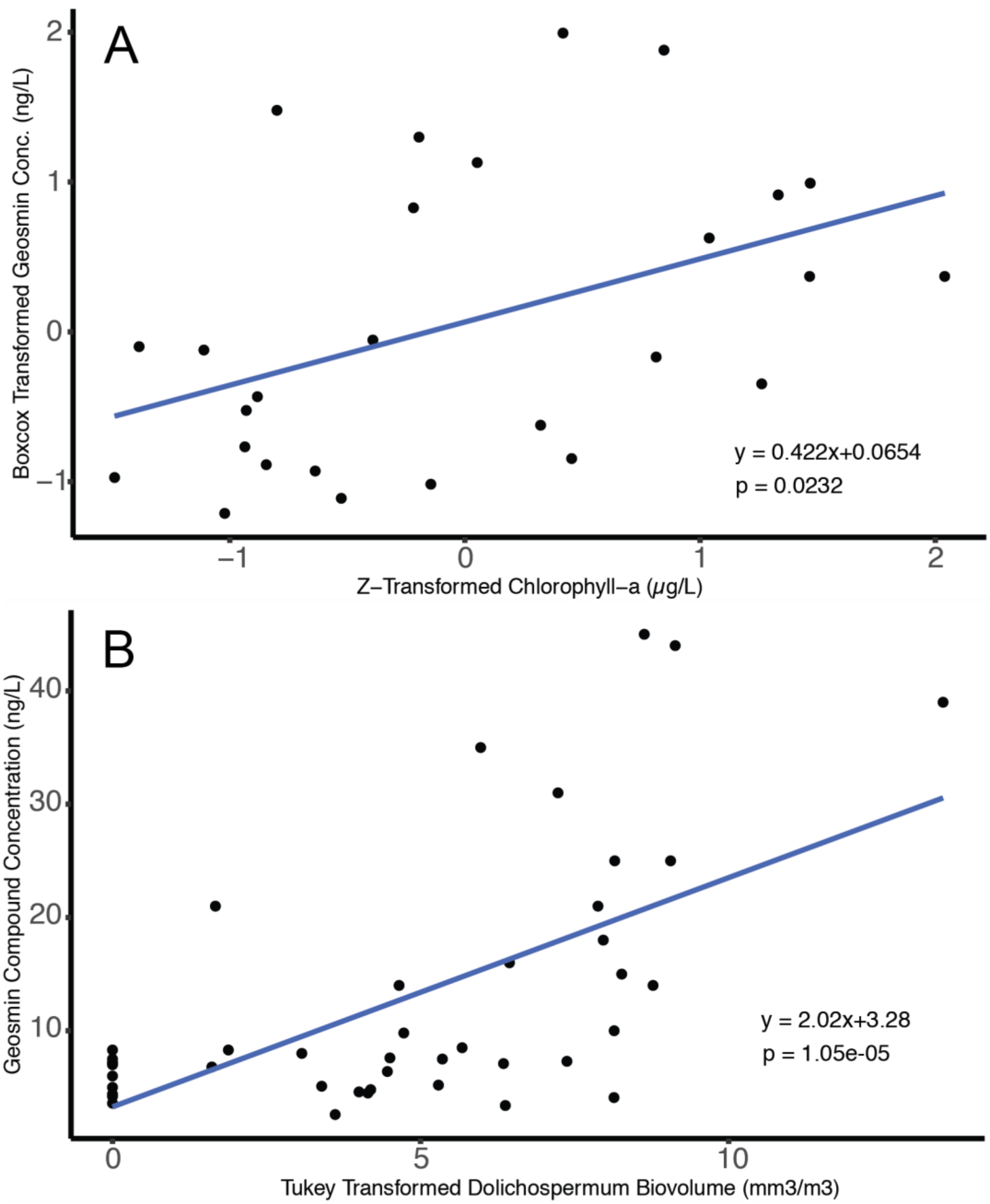


Figure 10. Non-parametric linear regressions applied to transformed geosmin, Chl a and *Dolichospermum* biovolume data from 2020. In (A) is geosmin and Chl a concentration data for 2020 and in (B) is geosmin and Dolichospermum biovolume data. Note the higher p value for the linear fit in (B).

Dissolved and particulate nutrient species were measured alongside LM collections in 2020 (Fig. 11). In brief, nitrate + nitrite (NOx) concentrations were generally elevated (>200 µg/L) until after day 147, and after day 188 (6/7/2020) NOx declined thereafter and became undetectable for the rest of the summer until day 265 (9/26/20). A slight increase in NOx was observed at the end of the study period. PO4 concentrations were below the method detection limit (6.5 µgP/L) in most samples but total P (TP) ranged from ~36 to 200 µgP/L (average = 74 µgP/L) indicating a high amount of dissolved and/or particulate P is present in LM (Fig. 11). The ratio of TN:TP was on average 8.7 based on mass comparison (µg/L); this ratio is generally considered an N-limiting condition for freshwater phytoplankton and N-fixation is favored (Guildford and Hecky 2000; Scott et al. 2019).

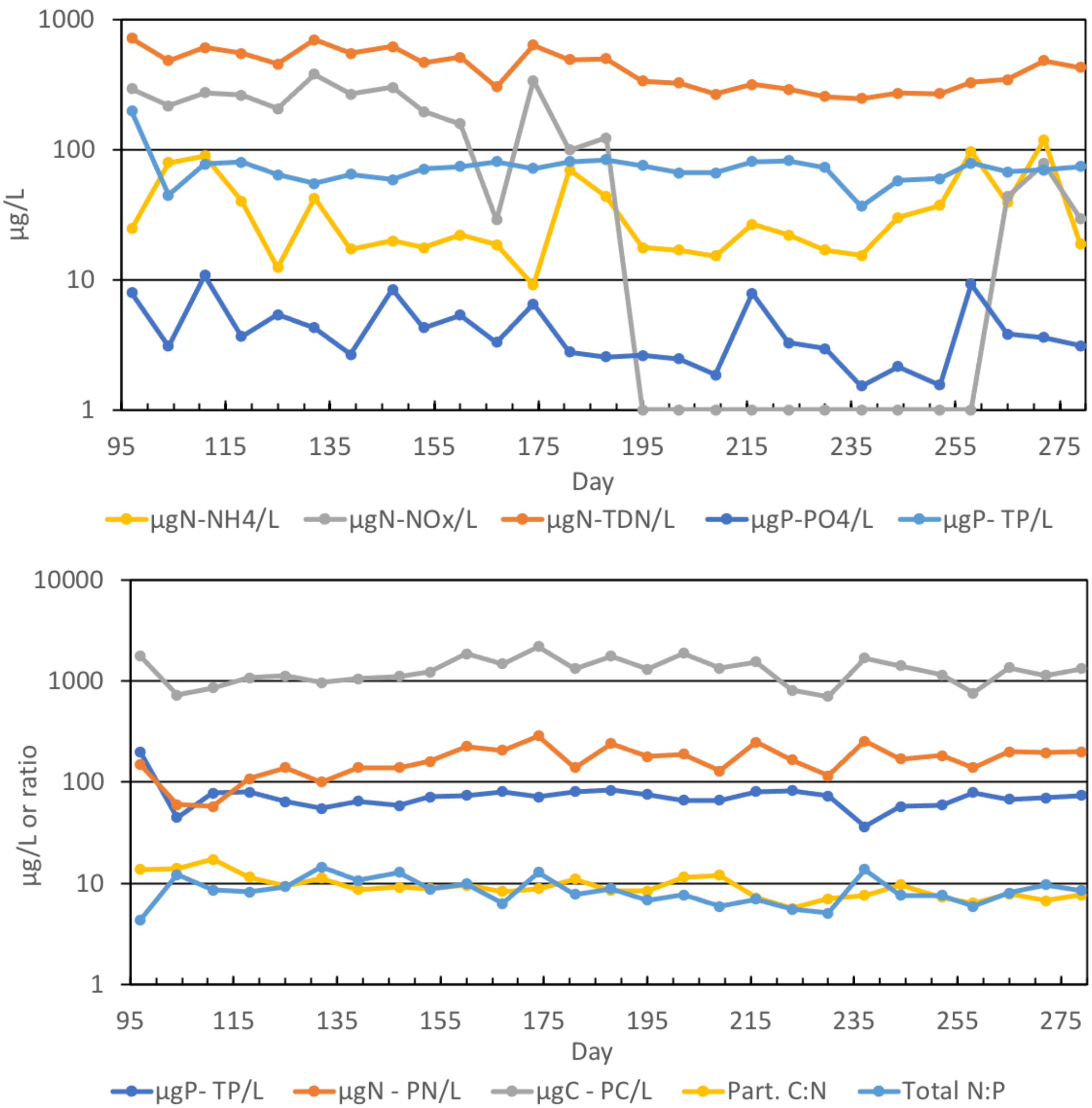


Figure 11. Nutrient concentrations in LM samples collected in 2020. The top panel shows measured dissolved nitrogen (N) species (TDN= total dissolved N), dissolved phosphate (PO4) and total phosphorus (TP). The bottom panel shows TP as well as particulate measurements, particulate C:N, and total N:P ratios.

Transformation-based (tb) redundancy analysis (RDA) of the larger 2020 dataset was performed to examine how much of the variation in geosmin and 2-MIB concentrations could be explained by change in measured hydrologic and biological (Chl a) parameters (Fig. 12). Biovolume data was not used as it was an incomplete dataset biased towards early samplings in the year. Overall, 84% of the variation in T&O concentrations was explained by change in hydrologic and biological (Chl a) parameters. Strongest relationships were with 2-MIB concentrations, where NOx, TDN, and turbidity exhibited strong negative relationships with 2-MIB concentration, and Secchi depth showed a positive relationship with 2-MIB (Fig. 12). All considered, the results support a narrative that 2-MIB is promoted by stable, low nutrient conditions that are common in mid to late summer.

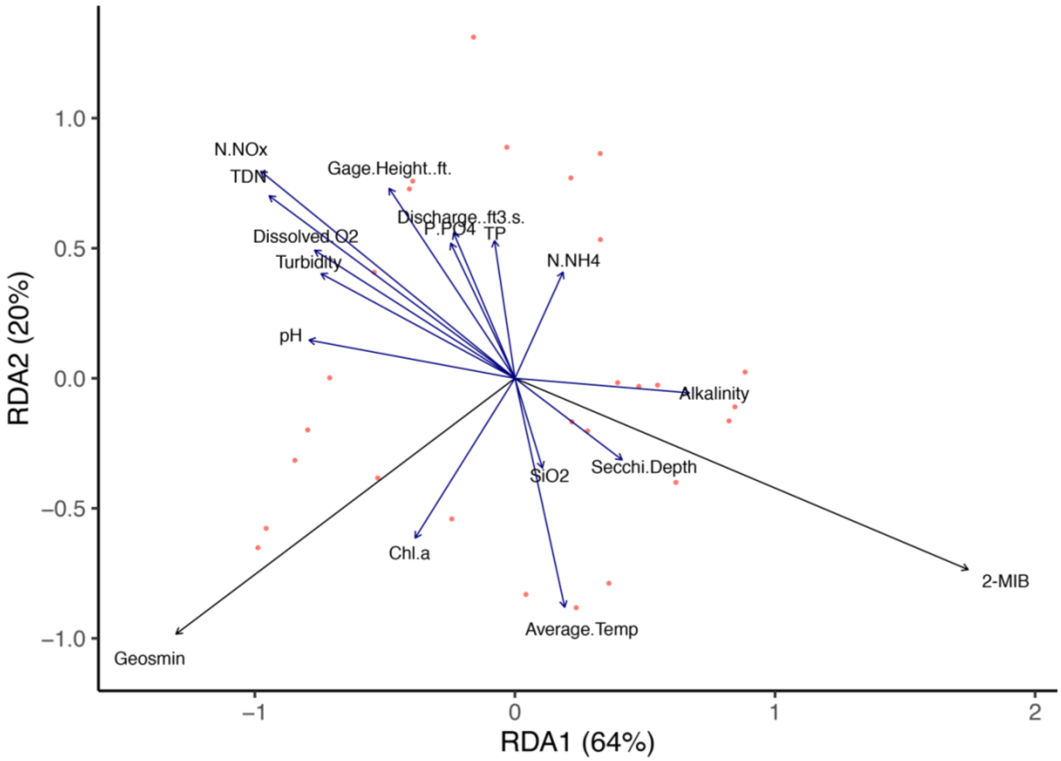


Figure 12. A tb-RDA plot showing the relationships between T&O concentrations (black lines, geosmin and 2-MIB) and hydrologic as well as biochemical measurements in 2020.

3.4 Flow cytometry (FCM)-based cell counts

Four different phytoplankton populations were identified using flow cytometry in 2020 samples (Fig. 13), including putative cyanobacteria and eukaryotic phytoplankton based on pigment and size signatures (Collier 2000; Paerl et al. 2020b). Cell abundances of these populations were generally typical for euphotic freshwater and estuarine euphotic waters – with small eukaryotic phytoplankton being similar in numbers to green phycocyanin-rich (PC-rich) picocyanobacteria early in the year (spring) and then PC-rich picocyanobacteria becoming more dominant in warmer summer to early fall months (Fig. 14). PC-rich picocyanobacteria (cells <3 µm diameter) were the most abundant phytoplankton detected (reaching ca. 5 x 10^5 cells/mL) in all but two 2020 samples from LM and peaked in warmer summer months (Fig. 14). The size and fluorescence characteristics of this population is similar to those previously reported for *Synechococcus* or *Cyanobium* cells in freshwater habitats (Callieri 2008; Paerl et al. 2020b). No coherent change was evident between the abundance of any FCM-determined phytoplankton group and T&O compound concentration (Fig. 14). Other notable patterns in the data include: 1) a generally higher concentrations of all FCM defined groups in the summer (ca. study day 170 -200), which is common for freshwater and marine systems (Stockner and Antia 1986) and 2) a decrease in abundance of multiple eukaryotic phytoplankton groups for three weeks in late spring into early summer (study day 139-153) when filamentous cyanobacterial biomass increased (Figs. 8, 11). Bacterioplankton abundance was generally stable across 2020 samplings with highest abundances (ca. 10^6 cells/mL) later in the summer and early fall (Fig. 14).

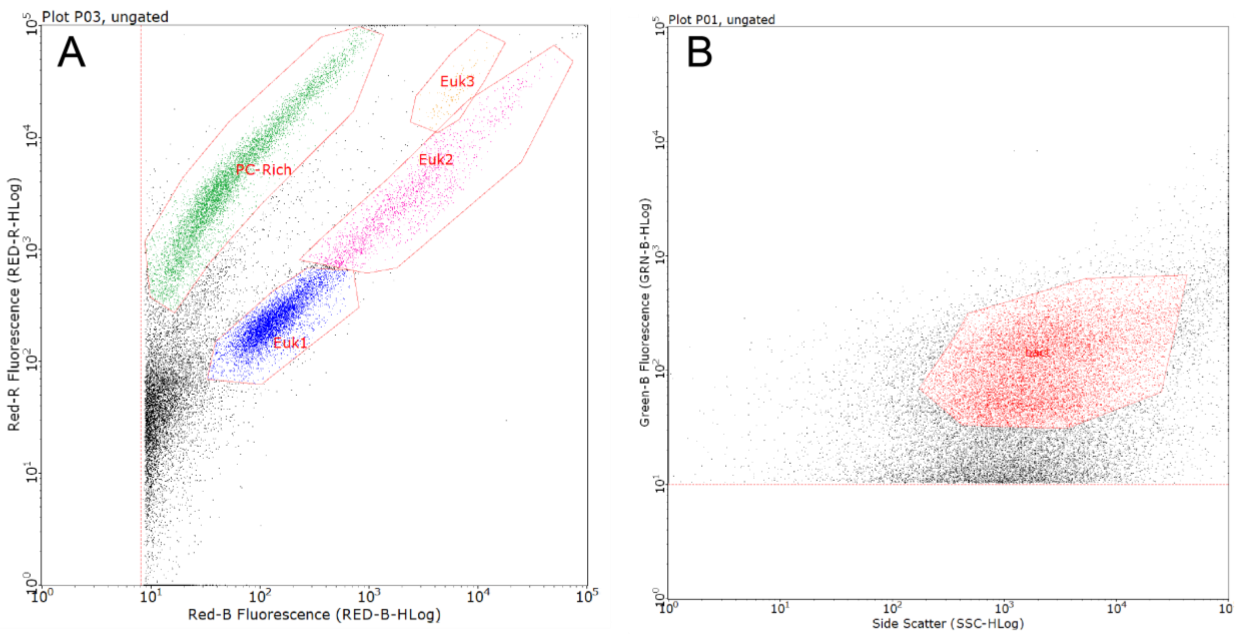


Figure 13. Example gating of phytoplankton (A) and bacterioplankton (B) populations counted by FCM.

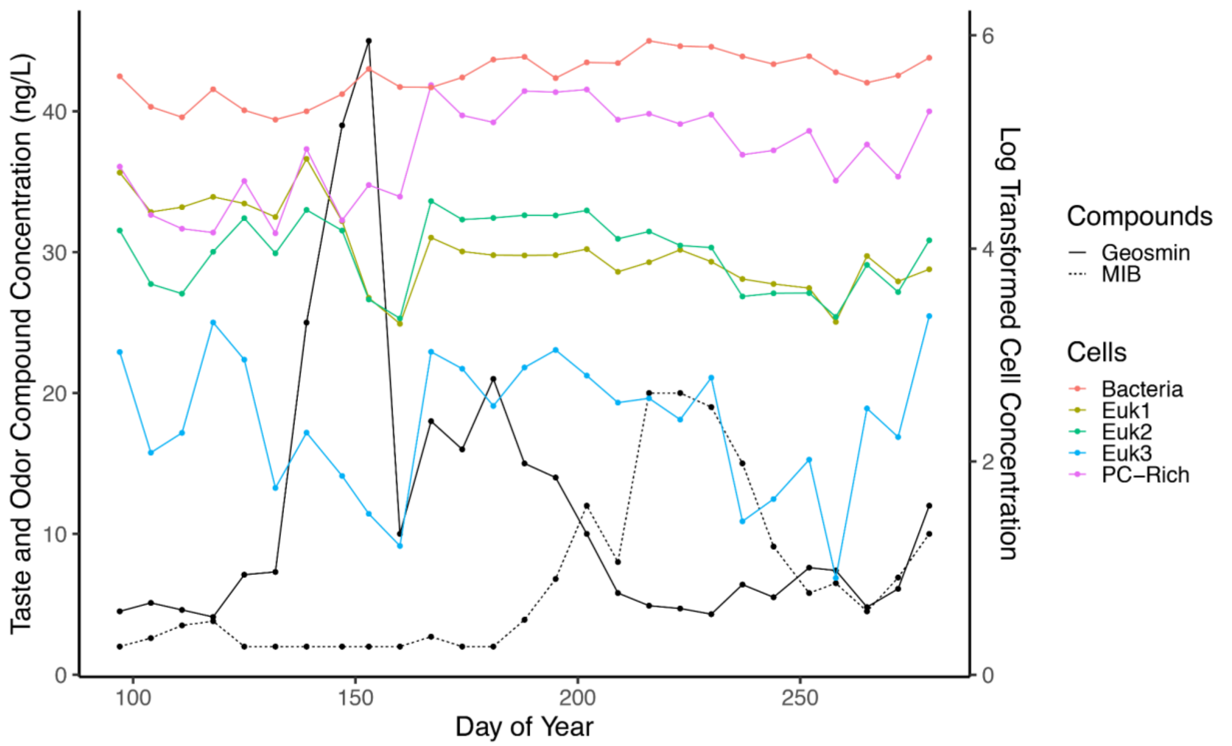


Figure 14. Abundances of FCM plankton counts from LM water in 2020.

3.5 PCR results & geoA amplicon sequences

Only *geoA* could be successfully amplified from LM samples using previously published primers (see methods). Nonspecific amplification was evident in *geoA* PCR reactions along with the expected *geoA* band (Fig. 15). Amplification of *geoA* was successful in largely PCR reactions using DNA from 8 µm pore-size filters, which capture more large planktonic cells like filamentous cyanobacteria. Exceptions to this were three samples (0.8 µm pore-size filters) obtained later in the time series (days 258- 272, mid to late August) (Table 5).

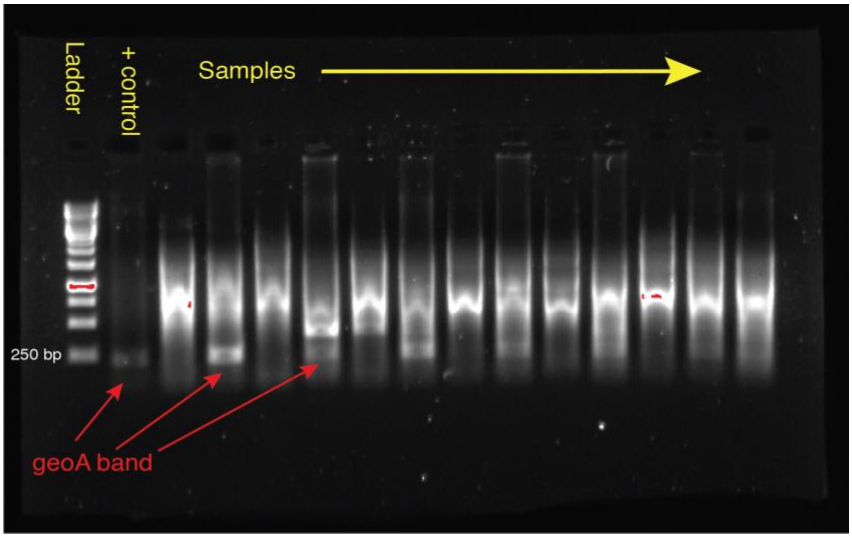
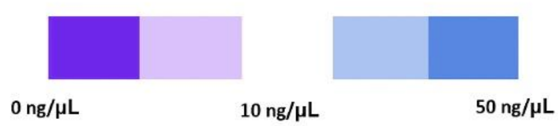


Figure 15. Example results of *geoA* PCR, including samples with successful amplification where a *geoA* band of the expected size (~220bp) could be identified. A 1kb ladder is shown for reference. ‘Smearing’ seen in samples is suspected to be the result of non-specific PCR amplification.

Table 5. Amplification of *geoA* from 2020 DNA extracts. Light green highlights samples where a *geoA* was amplified and sequenced. Purple to blue highlights note varying geosmin concentration in matching samples.

Date	Study Day	Geosmin (ng/µL)	GeoA (8.0 µm filter)	GeoA (0.8 µm filter)	Date	Study Day	Geosmin (ng/µL)	GeoA (8.0 µm filter)	GeoA (0.8 µm filter)
4/6/20	97	4.5	Y	N	7/13/20	188	14	N	N
4/13/20	104	5.1	Y	N	7/20/20	202	10	Y	N
4/20/20	111	4.6	Y	N	7/27/20	209	5.8	N	N
4/27/20	118	4.1	Y	N	8/3/20	216	4.9	N	N
5/4/20	125	7.1	Y	N	8/10/20	223	4.7	N	N
5/11/20	132	7.3	Y	N	8/17/20	230	4.3	N	N
5/18/20	139	25	Y	N	8/24/20	237	6.4	N	N
5/26/20	147	39	Y	N	8/31/20	244	5.5	N	Y
6/1/20	153	45	Y	N	9/8/20	252	7.6	N	Y
6/8/20	160	10	Y	N	9/14/20	258	7.4	N	N
6/15/20	167	18	Y	N	9/21/20	265	4.8	N	Y
6/22/20	174	16	Y	N	9/28/20	272	6.1	Y	N
6/29/20	181	21	Y	N	10/5/20	279	12	N	N
7/6/20	188	15	Y	N	10/12/20	286	12	N	N



Sequencing of *geoA* amplicons (amplified DNA) and downstream processing recovered a total of 993 different amplicon sequence variants (ASVs; a genetically defined individual population). Assigning taxonomy to these ASVs revealed that a majority (top 20 shown; Tables 6,7) belong to *Aphanizomenon* and *Dolichospermum* spp., and a lesser number associated with *Nostoc*, *Scytonema*, *Streptomyces* and unclassifiable populations. Some non-*geoA* sequences were obtained as well, which was expected given the observed non-specific amplification in our *geoA* PCR reactions (Fig. 15).

Table 6. The top 20 most abundant ASVs and their respective *geoA* sequences recovered via PCR amplicon sequencing.

ASV	Genus	*geoA* partial Sequence
ASV1	Dolichospermum	ACGGAAGGTTGGGGGCGCACCTTGGTCAGCAGATTTAGTCGAACACGCGGCCTTTGTGGAAGTTCCGGCTAAAATTGCGGCAACTAGACCCATGCGGGTTTTAAAAGACACATTTGCTGATGGTGTACATCTCCGCAATGATCTATTC
ASV2	Aphanizomenon	GCGTAAAGTTGGTGGCGCACCCTGGTCAGCCGATCTGGTGGAACACGCCGCGTTCGTGGAAGTTCCGGCTAAAATTGCGGCAACTAGACCAATGCGGGTTTTGAAAGACACATTTGCTGATGGAGTACATCTCCGCAACGACCTATTC
ASV3*	uncultured bacterium	CCGTAAGGTAGGGGGTGCACCGTGGTCAGCCGATCTGGTCGAACACGCTGTTTTTATAGAAGTTCCCGCTAAAATTGCCGCGACTCGACCGATGCGGGTACTTAAAGACACCTTTGCCGATGGCGTGCATCTTCGCAACGACCTATTT
ASV4	Dolichospermum	GCGGAAAGTTGGGGGCGCACCTTGGTCAGCAGATTTAGTCGAACACGCGGCCTTTGTGGAAGTTCCGGCTAAAATTGCGGCAACTAGACCCATGCGGGTTTTAAAAGACACATTTGCTGATGGTGTACATCTCCGCAATGATCTATTC
ASV5	Dolichospermum	GCGGAAAGTTGGGGGCGCACCTTGGTCAGCAGATTTAGTCGAACACGCGGCCTTTGTGGAAGTTCCGGCTAAAATTGCGGCAACCAGACCCATGCGGGTTTTAAAAGACACATTTGCTGATGGTGTACATCTCCGCAATGATCTATTC
ASV6	Dolichospermum	ACGGAAAGTTGGGGGCGCACCTTGGTCAGCAGATTTAGTCGAACACGCGGCCTTTGTGGAAGTTCCGGCTAAAATTGCGGCAACTAGACCCATGCGGGTTTTAAAAGACACATTTGCTGATGGTGTACATCTCCGCAATGATCTATTC
ASV7*	uncultured bacterium	TCGCAAAGTCGGCGGCGCACCCTGGTCAGCCGATTTGGTCGAACACGCCGTATTTGTCGAAGTTCCCGCTAGAATTGCTGCCACTCGACCAATGCAGGTACTAAAAGCCACATTTTCCGATGGCGTGCATCTCCGCAACGATTTATTC
ASV8*^	uncultured Caudovirales phage	AATACATTGAGATGATGGTCGCTGAGAGGCAGTTAAAGCAGTTCTACAAAGACTTGAAAGAGATGTTTATCTATCAGTTCCAAGAACCCGGTTTGTATGACGAGTTTATGGGGCGGCTAGAGAAACTTAGGGCAGACCGTAGACAAAGAG
ASV9	Aphanizomenon	GCGTAAAGTTGGTGGCGCACCCTGGTCAGCCGATCTGGTGGAACACGCCGCGTTCGTGGAAGTTCCGGCTAAAATTGCGGCAACTAGACCAATGCGGGTTTTGAAAGACACATTTGCTGATGGCGTACATCTCCGCAATGACCTATTC
ASV10	Dolichospermum	GCGGAAAGTTGGGGGCGCACCTTGGTCAGCAGATTTAGTCGAACACGCGGCCTTTGTGGAAGTTCCGGCTAAAATTGCGGCAACTAGACCGATGCGGGTTTTAAAAGACACATTCGCTGATGGTGTACATCTCCGCAATGATCTATTC
ASV11	Scytonema	CCGTAAGGTGGGGGGAGCGCCCTGGTCAGCTAACCTAGTAGAACACGCGGTTGGGGCAGAAATTCACCCTGCGATCGCCCCCACTCGACCAATGCGTGTCCTCAAAGACACCTTTTCTGACGGGGTACATCTGCGAAATGACATCTTC
ASV12*^	No match found	GAATACATTGAGATGCAGAAGAAAATAGAGAAGGAATCGGCAAGAGCCGAAGAGTTCGAGCTTTTCAAGCTGCGTCTGCAGAGTGGCGTCTACCCGTACGCCGAGTGGGAGACTGAAGAGGGAAATTATGTGGGTGCGTGCTGCGTAGAGA
ASV13	Dolichospermum	ACGGAAAGTTGGGGGCGCACCTTGGTCAGCAGATTTGGTAGAACACGCCGTATTTGTCGAAGTTCCGGCTAAAATTGCGGCAACTAGACCGATGCGGGTTTTAAAAGACACATTTGCTGATGGTGTACATCTCCGCAATGATCTATTC
ASV14	Nostoc	CCGCAAGGTTGGTGGCGCACCGTGGTCAGCCAATCTCGTAGAACACGCCGCGTTTGTTGAGATCCCGGCTGAAATCGCCACAACTCGACCGATGCGTGTACTCAAAGACACGTTTGCCGATGGAGTGCATCTTCGCAACGATTTGTTC
ASV15*	Cylindrospermum stagnale PCC 7417	CCGCAAGGTTGGTGGCGCACCGTGGTCAGCCAATCTCGTAGAACACGCCGCGTTTGTTGAGATCCCGGCTGAAATCGCCACAACTCGACCGATGCGTGTACTCAAAGACACGTTTGCCGATGGGGTGCATCTTCGCAACGATTTGTTC
ASV16	Nostoc	CCGCAAGGTTGGTGGCGCACCGTGGTCAGCCAATCTCGTAGAACACGCCGCGTTTGTTGAGATTCCGGCTGAAATCGCCGCAACTCGACCGATGCGTGTACTGAAAGACACGTTTGCCGATGGAGTGCATCTTCGCAACGATTTGTTT
ASV17*	uncultured bacterium	GCGCAAAGTCGGCGGCGCACCCTGGTCAGCCGATTTGATCGAACACGCCGTATTTGTCGAAGTTCCGGCTAGAATTGCCGCCACTCGACCAATGCAGGTACTGAAAGCCACATTTTCCGATGTTGGACATCTCTGCAACGATTTATTC
ASV18*^	uncultured Caudovirales phage	AATACATTGAGATGATGGTCGCTGAGAGGCAACTAAAGCAGTTCTACAAAGACTTGAAAGAGATGTTTATCTATCAGTTCCAAGAACCCGGTTTGTATGACGAGTTTATGGGGCGGCTAGAGAAACTTAGGGCAGACCGTAGACAAAGAG
ASV19*^	uncultured Caudovirales phage	AATACATTGAGATGATGGCCGCTGAGAGGCAGTTAAAGCAGTTCTACAAAGACTTGAAAGAGATGTTTATCTACCAGTTCCAAGAACCCGGTCTGTATGACGAGTTTATGGGGCGGTTAGAGAAACTTAGGGCAGACCGCAGACAAAGAG
ASV20*^	No match found	ACATTGAGATGCGATTACCACGATATGCGCGCATCGTTTCTTCGGGCAGCGTAAGTAATTCTGTTGTAGATCCATCATCTTCATGAAAAAGAATGGAACCACTGGCAATTCTTCCGGGAGGATTTGGGATGAGTCTCATCACGGAGA

***ASV identification was retrieved from the top hit obtained useing a BLASTn search against the NCBI NR/NT database; ^ no match to GeoA based on BLASTx search against the NR NCBI database.**

Table 7. Relative abundance (in %) of the top 20 most abundant ASVs across the 2020 dataset.

ASV	Genus	Day97	Day104	Day111	Day118	Day125	Day132	Day139	Day147	Day153	Day160	Day167	Day174	Day182	Day195	Day237	Day244	Day258	Day265
ASV1	Dolichospermum	2.9%	4.4%	4.6%	7.9%	17.6%	43.8%	42.7%	51.5%	73.1%	60.0%	64.2%	70.4%	75.8%	72.2%	0.0%	0.2%	0.1%	60.4%
ASV2	Aphanizomenon	88.0%	84.7%	89.5%	84.0%	76.9%	50.1%	50.8%	6.6%	7.9%	6.2%	3.4%	1.2%	0.5%	0.4%	0.1%	0.4%	0.0%	1.7%
ASV3*	uncultured bacterium	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	67.6%	92.9%	95.3%	0.0%
ASV4	Dolichospermum	0.4%	0.6%	0.0%	1.0%	1.0%	2.8%	1.9%	7.8%	8.1%	7.0%	7.0%	1.9%	3.8%	6.0%	0.0%	0.0%	0.0%	4.3%
ASV5	Dolichospermum	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	28.8%	5.3%	3.2%	7.2%	0.6%	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%
ASV6	Dolichospermum	3.9%	4.2%	3.2%	2.5%	1.2%	2.2%	0.8%	2.0%	1.9%	4.3%	2.2%	10.0%	3.2%	1.2%	0.0%	0.0%	0.0%	0.0%
ASV7*	uncultured bacterium	0.0%	1.2%	0.0%	1.8%	0.0%	0.0%	3.2%	0.0%	0.3%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	1.0%	24.8%
ASV8*^	uncultured Caudovirales phage	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	10.5%	5.6%	5.9%	8.1%	0.1%	0.0%	0.0%	0.0%
ASV9	Aphanizomenon	3.4%	1.3%	0.8%	0.2%	0.5%	0.3%	0.0%	1.3%	0.6%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ASV10	Dolichospermum	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	0.6%	2.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ASV11	Scytonema	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	5.4%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%
ASV12*^	No match found	0.0%	0.0%	0.1%	0.1%	0.2%	0.1%	0.0%	0.0%	0.4%	0.1%	1.4%	0.9%	0.5%	0.1%	0.0%	0.0%	0.0%	0.0%
ASV13	Dolichospermum	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	1.7%	0.8%	0.4%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%
ASV14	Nostoc	0.0%	1.7%	0.1%	0.0%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%	0.0%	0.0%	0.0%	0.0%
ASV15*	Cylindrospermum stagnale PCC 7417	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	1.9%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%
ASV16	Nostoc	0.2%	0.0%	0.0%	0.0%	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ASV17*	uncultured bacterium	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	1.5%
ASV18*^	uncultured Caudovirales phage	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.4%	0.3%	0.4%	0.0%	0.0%	0.0%	0.0%
ASV19*^	uncultured Caudovirales phage	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%
ASV20*^	No match found	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.4%	0.3%	0.4%	0.5%

Interesting patterns emerged when examining the relative abundance of ASVs over time alongside geosmin concentration (Fig. 16). In April (days 97-118), the relative abundance of *Aphanizomenon geoA* ASVs were highest until early May (day 125) when *Dolichospermum geoA* ASVs began to increase and become >85% of *geoA* sequences by mid-May (day 132). *Dolichospermum geoA* ASVs continued to dominate in relative abundance into July (day 195), but notably by the July fewer *geoA* sequences were detected by PCR - likely reflecting fewer geosmin producers. Markedly, geosmin concentration increased in concert with the increasing relative abundance in *Dolichospermum geoA* ASVs sequences. Later in the time series around September (day 230) *geoA* sequences were only recovered from 0.8 µm pore-sized filters - suggesting smaller plankton, between 8 and 0.8 µm diameter possessed the gene (Fig. 16). It was not possible to taxonomically assign these sequences using our in-house *geoA* database of sequences from the public NCBI database and Phyloseq (see methods). Independent BLASTn searches using these ASVs showed they share closest, yet distant, similarity to uncultured bacterial sequences from aquaculture systems in Denmark (Lukassen et al. 2017).

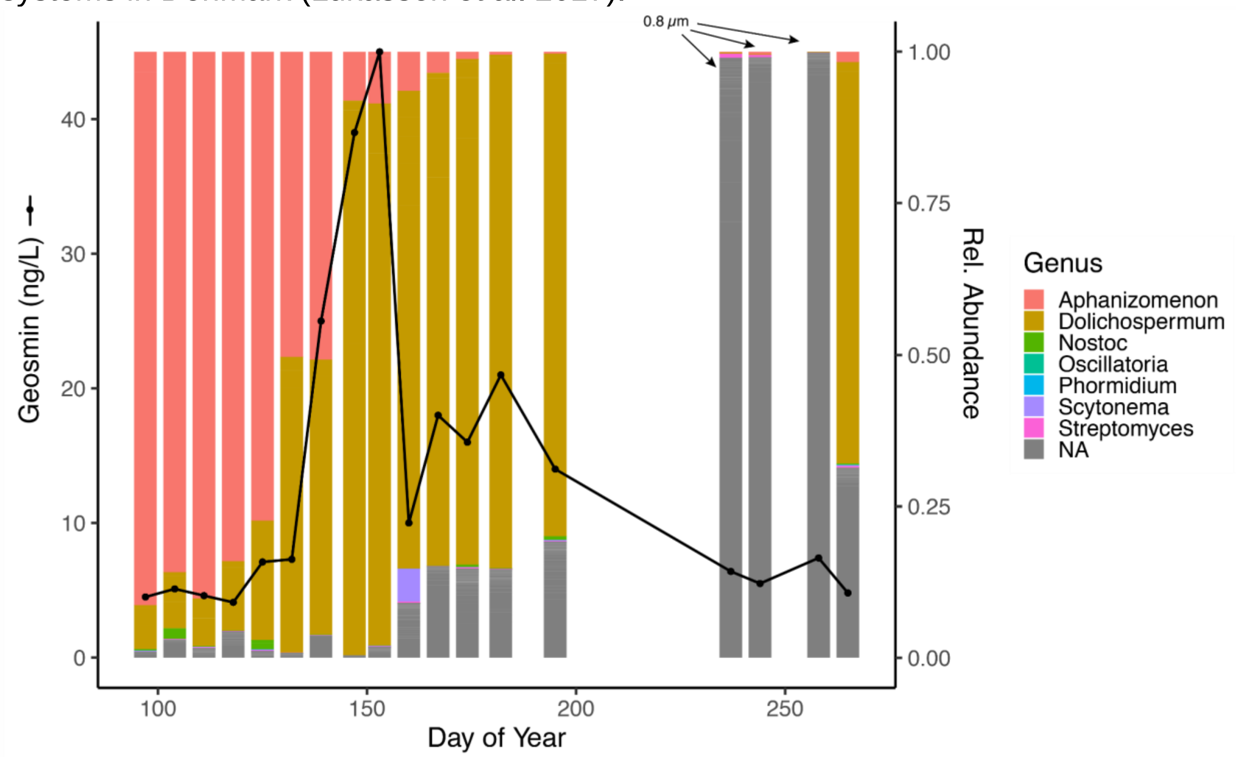


Figure 16. Relative abundance of *geoA* ASVs from 2020 LM DNA samples (columns) and geosmin concentrations (black line) measured on the same dates. All sequences were obtained exclusively from 8.0 µm pore-size filters (amplification was evident in respective 0.8 µm pore-size filters), except for three that are marked as deriving from 0.8 µm pore-size filters. Gaps in columns indicate no amplification of *geoA* from weekly samples collected within the respective time windows.

3.6 Metagenomic sequencing results

Metagenomic sequencing was used to identify T&O producing populations without relying on PCR amplification to recover genes tied to T&O production (e.g. *geoA*, *mic*).

PCR carries with it inherent biases (Acinas et al. 2005), and metagenomic sequencing potentially helps overcome these biases. Twenty different DNA samples (Table 8) were sequenced using the Illumina platform yielding a total of 4.05 x 10^{11} bases (2.7 x 10^{9} reads) and an average of 2.02 x 10^{10} bases per sample, after running quality control filtering (see methods). Samples represent peak geosmin and 2-MIB measurements as well as moderate to low concentrations (2- 45 ng/L).

At the phylum level, plankton diversity across metagenomic samples was relatively stable over time (4/6/20-10/5/20) with *Proteobacteria*, *Cyanobacteria*, *Plantomycetes*, *Verrucomicrobia* and *Bacteroidetes* being most prevalent (Fig. 17), which is typical for freshwater microbial communities (Newton et al. 2011). An increased number in eukaryotic sequences were observed on the 8 µm pore-sized filter, including typically larger sized plankton, e.g. *Ciliophora* (Ciliates), *Chlorophyta* (Chlorophyte algae) and others (Fig. 17). Distinct increases in *Chloroflexota* (Chloroflexi), Chlorobi and to a lesser extent *Acidobacteria* occurred in August samples (8/3/2020, 8/31/2020). The majority of cyanobacterial sequences belong to *Synechococcus* and *Cyanobium* spp. (Fig. 18), which are genera represented by small unicellular cyanobacteria (ca. 1-5 µm diameter) and are ubiquitous and often abundant in euphotic freshwater (Stockner and Antia 1986; Jasser and Callieri 2017). Next abundant cyanobacteria include *Aphanizomenon*, *Dolichospermum*, *Cylindrospermopsis*, *Anabaena*, *Calothrix*, *Microcystis*, *Microcoleus*, *Aphanothece* and *Planktothrix* spp. (Fig. 18).

Table 8. Samples used to generate metagenomic sequence libraries. T&O compound concentrations and retrieved hits via BLASTn searches against the NCBI NR/NT database are provided.

ID	Date	Sample Day	Filter (µm)	Geosmin (ng/L)	MIB (ng/L)	T&O Genes detected
TO_2	4/6/20	97	8	4.5	2	0
TO_3	4/6/20	97	0.8	4.5	2	0
TO_15	5/4/20	125	8	7.1	2	geoA [1]
TO_16	5/4/20	125	0.8	7.1	2	0
TO_26	5/18/20	139	8	25	2	geoA [1]
TO_27	5/18/20	139	0.8	25	2	0
TO_36	6/1/20	153	8	45	2	geoA [1]
TO_37	6/1/20	153	0.8	45	2	0
TO_45	6/15/20	167	8	18	2.7	0
TO_46	6/15/20	167	0.8	18	2.7	0
TO_56	6/30/20	181	8	21	2	geoA [4]
TO_57	6/30/20	181	0.8	21	2	0
TO_58	7/6/20	188	8	15	3.9	geoA [3]
TO_59	7/6/20	188	0.8	15	3.9	geoA [2]
TO_83	8/3/20	216	8	4.9	20	0
TO_84	8/3/20	216	0.8	4.9	20	0
TO_102	8/31/20	244	8	5.5	9.1	geoA [1]
TO_103	8/31/20	244	0.8	5.5	9.1	0
TO_125	10/5/20	279	8	12	10	geoA [2], mic [1]
TO_126	10/5/20	279	0.8	12	10	0

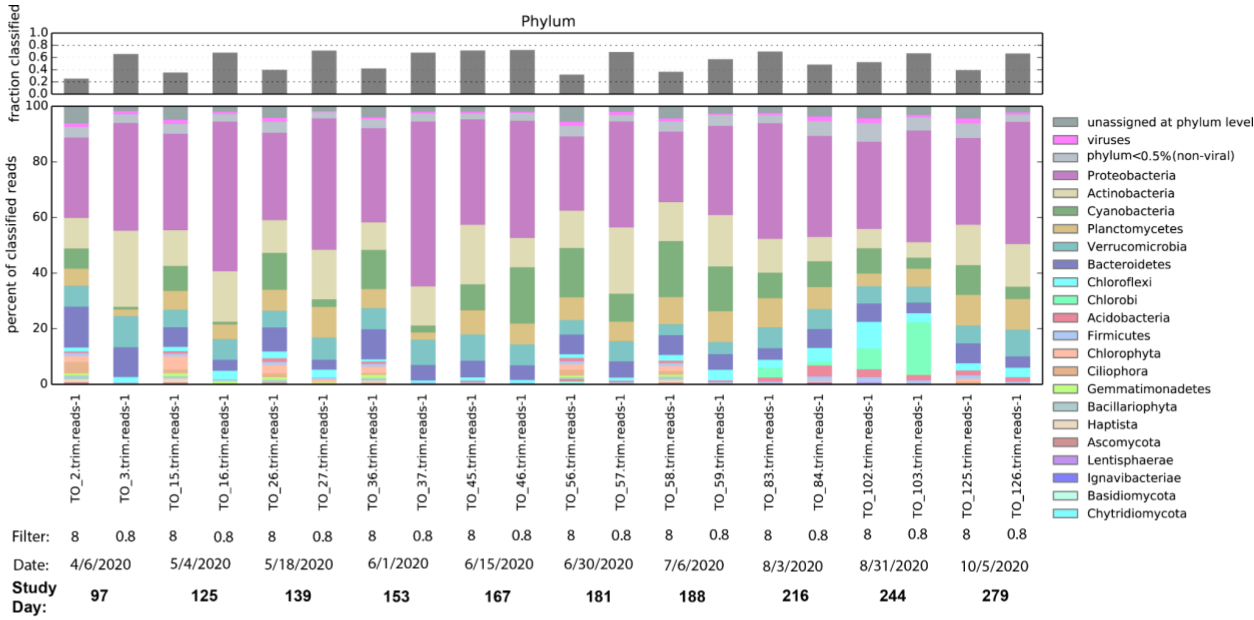


Figure 17. Relative abundance of reads assigned to specific phyla across metagenomic samples (select samples across the complete time series - see study day).

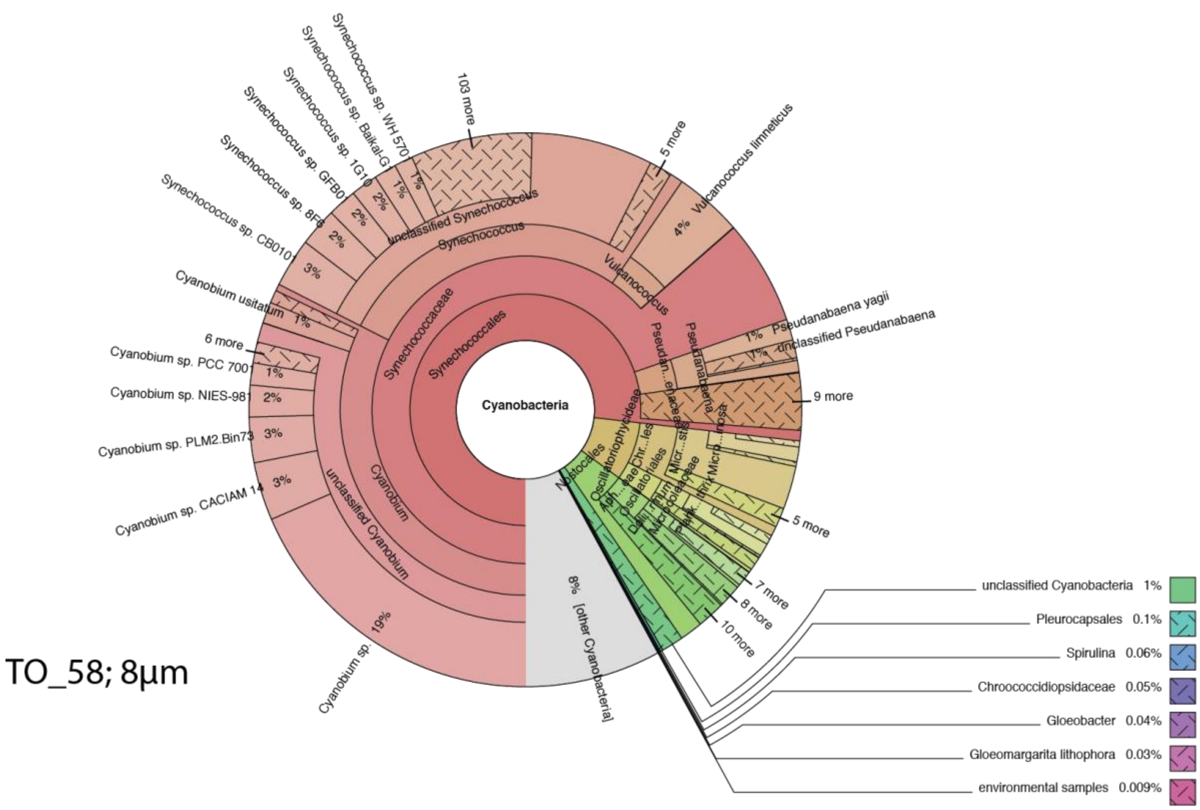


Figure 18. An example Krona plot showing the relative abundance of cyanobacterial reads in the metagenomic sample TO_58 (study day 188).

BLASTn searches of our LM metagenomes recovered 15 *geoA* sequences across contigs generated from all samples (Table 9). Seven different taxa accounted for these 15 geoA sequences based on their nearest sequence matches (using BLASTn) in the public NCBI nr/nt database, specifically: *Dolichospermum ucrainicum* CHAB1432, *Dolichospermum ucrainicum* CHAB1434, *Dolichospermum ucrainicum* CHAB2155, *Dolichospermum planctonicum* SDZ-1, *Oscillatoria limosa* CHAB7000, and 'Uncultured bacterium clone AQC_0126'. Notably, sequence similarity to *Oscillatoria limosa* CHAB7000 was quite low based on % identity and thus the LM *geoA* sequences recovered are difficult to confidently assign to that genus (Table 9). Uncultured bacterium clone AQC_0126 *geoA* sequences were from a Danish aquaculture study and highly likely non-cyanobacterial (Lukassen et al. 2017). One *mic* sequence was recovered, and its closest sequence match was to the cyanobacterium *Microcoleus pseudautumnalis* Ak1609 although the sequence identity was low (~84%) making it difficult to say with confidence the LM sequence belongs to *Microcoleus* (Table 9).

Table 9. A list of contigs in LM metagenomic libraries possessing *geoA*. Information is shown for the top hits retrieved from searching geoA sequences from contigs against the public NCBI nr/nt database.

SampleID	Date	Study day	Filter (µm)	Contig	Gene	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession ID
TO_15	5/4/20	125	8	k127_264593	*geoA*	Dolichospermum ucrainicum CHAB1432	3615	3615	100%	0	97.41	5272	HQ404996.1
TO_26	5/18/20	139	8	k127_1662810	*geoA*	Dolichospermum ucrainicum CHAB1432	3930	3930	100%	0	97.97	5272	HQ404996.1
TO_36	6/1/20	153	8	k141_1161537	*geoA*	Dolichospermum ucrainicum CHAB1432	3936	3936	100%	0	98.02	5272	HQ404996.1
TO_56	6/30/20	181	8	k127_1405190	*geoA*	Dolichospermum ucrainicum CHAB1434	2507	2507	100%	0	97.99	2256	MK213947.1
TO_56	6/30/20	181	8	k127_2257596	*geoA*	Dolichospermum planctonicum SDZ-1	1308	1308	100%	0	98.64	2256	MK213954.1
TO_56	6/30/20	181	8	k127_335534	*geoA*	Uncultured bacterium clone AQC_0126	1007	1007	62%	0	94.11	663	KX879262.1
TO_56	6/30/20	181	8	k127_493834	*geoA*	Oscillatoria limosa CHAB7000**	447	447	69%	3.00E-120	69.65	3881	MF084933.1
TO_58	7/6/20	188	8	k127_138867	*geoA*	Dolichospermum planctonicum SDZ-1	704	704	100%	0	98.26	2256	MK213954.1
TO_58	7/6/20	188	8	k127_142770	*geoA*	Uncultured bacterium clone AQC_0126	1007	1007	27%	0	94.11	663	KX879262.1
TO_58	7/6/20	188	8	k127_1575433	*geoA*	Dolichospermum ucrainicum CHAB2155	1703	1952	100%	0	97.04	5272	HQ404997.1
TO_59	7/6/20	188	0.8	k127_2545408	*geoA*	Uncultured bacterium clone AQC_0126	1007	1007	32%	0	94.11	663	KX879262.1
TO_59	7/6/20	188	0.8	k127_2606208	*geoA*	Oscillatoria limosa CHAB7000**	275	275	66%	7.00E-69	70.52	3881	MF084933.1
TO_102	8/31/20	244	8	k127_104046	*geoA*	Uncultured bacterium clone AQC_0126	743	743	65%	0	95.12	663	KX879262.1
TO_125	10/5/20	279	8	k127_1534399	*mic*	Microcoleus pseudautumnalis Ak1609***	503	503	84%	3.00E-137	71.91	4507	LC486303.1
TO_125	10/5/20	279	8	k127_1590324	*geoA*	Oscillatoria limosa CHAB7000**	287	287	71%	4.00E-72	69.24	3881	MF084933.1
TO_125	10/5/20	279	8	k127_2068863	*geoA*	Uncultured bacterium clone AQC_0126	1007	1007	27%	0	94.11	663	KX879262.1

*=Identical hits to at least one other Dolichospermum spp. **=based on discontinguous megablastn search ***=poor contig assembly

The largest number of unique *geoA* sequences were from summer (day 181; 6/30/20; 4 *different taxa*), while only one sequence recovered during the highest measured geosmin concentration on study day 153, *Dolichospermum ucrainicum* CHAB1432 (Table 8). During the highest 2-MIB concentration, no *mic* genes were detected. The one recovered *mic* sequence was obtained from later in the time series (day 279) when 2-MIB was elevated (10 ng/L; 10/5/20; Table 8).

With respect to time, *Dolichospermum geoA* sequences were most common in metagenomes from 8 µm filters from late spring to mid summer (study day 125-188) (Table 9). The *geoA* sequence variation between *Dolichospermum* spp. is subtle, yet the BLASTn search results suggest at least four different populations (sequence variations) occur in Lake Michie that possess the capability to produce geosmin. *Dolichospermum* ucrainicum CHAB1432 and CHAB1434 in particular were more prevalent in late spring and early summer, while 'uncultured bacterium' and *Oscillatoria* (as mentioned above all distantly related) *geoA* sequences in occurred mid-summer and later. The one *mic* sequence was detected in the fall (study day 279) (Table 9).

The *geoA* sequences with closest similarity to 'uncultured bacterium clone AQC_0126' occurred on contigs with neighboring protein coding genes that could be further examined to potentially identify the organism possessing the *geoA* gene. For example, on contig k127_2068863 the detected *geoA* is flanked by polyprenyl synthetase (protein model PF00348.17) as well as proteins of unknown function (Fig. 19). BLASTn searches of these neighboring genes against the NCBI NR/NT public database found that several distantly related (~70% identity) to *Chloroflexota* representatives (*Dehalococcoidis* and *Anaerolinea* spp.) - indicating that members of the phylum *Chloroflexota* in LM may be potential sources of geosmin (Fig. 19).

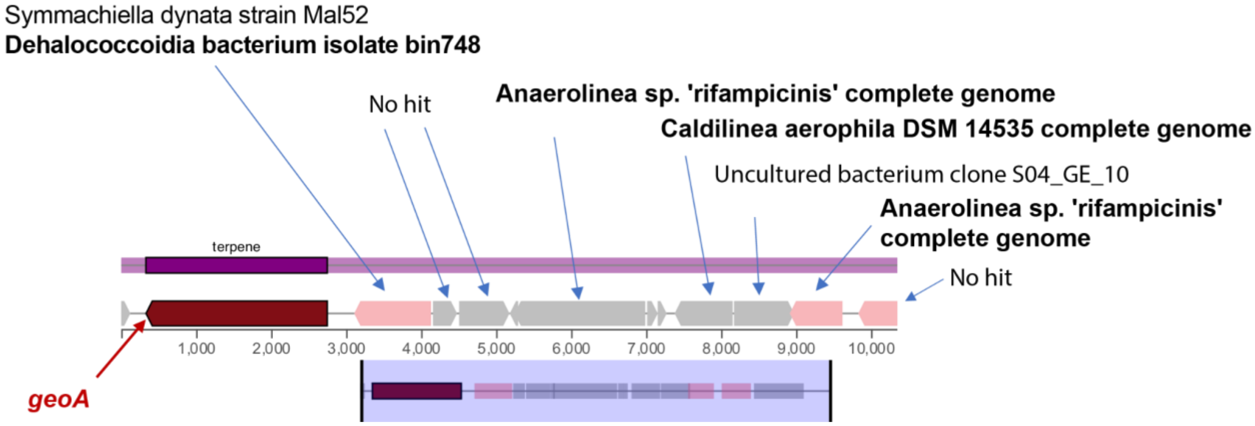


Figure 19. A map of genes on Contig 127_2068863 from sample TO_125. The contig map was generated by antiSMASH and manually annotated. Best matches based on BLASTn searches are shown where applicable and those associated with *Chloroflexota* are shown in bold.

3.7 Metagenome assembled genomes (MAGs) from metagenomic sequence libraries
A total of 271 high quality (>90% complete, <5% contamination) metagenome assembled genomes (MAGs), also known as ‘bins’ containing sequences related to a genetically similar population (Hugerth et al. 2015; Bowers et al. 2017), were obtained from Lake Michie DNA samples.

Among the high-quality cyanobacterial MAGs, two contained *geoA* genes: bin.001 from TO_15 (5/4/2020) and bin.113 from TO_26 (5/18/2020) based on searches with antiSMASH (Blin et al. 2021). Both MAGs were most similar to *Dolichospermum circinale* AWQC310F (GCA_000426905.1) based on broader genome content. Some sequence divergence (differences) is evident between the bins and *Dolichospermum circinale AWQC310F* as FASTANI scores are <99.5% (criteria for highly related genomes) (Jain et al. 2018), which suggests they are different strains. Other genes involved in secondary metabolite production were identified when searching for T&O compounds, including several pertinent to water quality and management (Figs. 20-22). First, heterocyst glycolipids were detected which highlights the ability of *Dolichospermum* to fix nitrogen and strongly compete with non-diazotrophic cyanobacteria, bacteria, and phytoplankton. Second, a number of putative toxin synthesis genes are present - most of which are understudied secondary metabolites relative to the algal toxins microcystin, anatoxin-a, and saxitoxin (Carmichael and Boyer 2016; Plaas and Paerl 2021). The type and number of these genes detected in each of these *Dolichospermum* MAGs varies. This may reflect differences in the gene repertoire of *Dolichospermum* populations, incomplete sequencing - i.e. some genes were not sequenced despite the MAGs being of high quality and near completion, or suboptimal binning (processing of sequence data). Briefly, these genes of interest include several coding for non-ribosomal peptides (NRPS) aeruginoside, anacyclamide A, cyanopeptin and anabaenopeptin (the latter only in bin.111 from TO_26) (Fig. 20-21). Related to cyanotoxin production, MAG bin.078 from sample TO_56 was closest in relation to *Dolichospermum planctonicum* (NCBI ID: GCF_005402965.1) and lacked *geoA*, but possessed genes (*mcyABC*) coding for microcystin synthesis (Fig.22). In contrast, MAGs taxonomically assigned to unicellular cyanobacterial genera (*Cyanobium*, *Synechococcus*) possessed much fewer secondary metabolite production genes and none possessed *geoA* or *mic* (Fig. 23).

Select genomic region:
Overview 122.1 173.1 312.1 456.1 508.1 561.1 615.1 635.1 644.1 686.1 690.1 709.1

Identified secondary metabolite regions using strictness 'relaxed'

Region	Type	From	To	Most similar known cluster		Similarity
Region 122.1	lanthipeptide-class-v	1	2,820			
Region 173.1	terpene	1	5,763			
Region 312.1	NRPS	1	10,805	aeruginoside 126B / aeruginoside 126A	NRP:Glycopeptide + Polyketide:Other + Saccharide:Hybrid/tailoring	41%
Region 456.1	T1PKS	1	4,388			
Region 508.1	terpene	1	3,019	geosmin	Terpene	100%
Region 561.1	terpene	1	7,837			
Region 615.1	NRPS	1	6,973	cyanopeptin	NRP	50%
Region 635.1	spliceotide , RRE-containing	1	6,938			
Region 644.1	cyanobactin	1	2,948	anacyclamide A10	RiPP:Cyanobactin	28%
Region 686.1	terpene	1	7,510			
Region 690.1	hglE-KS	1	8,029	heterocyst glycolipids	Other	57%
Region 709.1	NRPS-like	1	7,599	aeruginosin 98-A / aeruginosin 98-B / aeruginosin 98-C	NRP	21%

No secondary metabolite regions were found in these records:
k127_1002564

Fig. 20. Identified secondary metabolite coding genes (marked as regions) in MAG bin.001 from sample TO_15.

Select genomic region:
Overview 29.1 45.1 72.1 85.1 139.1 148.1 153.1 154.1 156.1 174.1 178.1 338.1 399.1

Identified secondary metabolite regions using strictness 'relaxed'

Region	Type	From	To	Most similar known cluster		Similarity
Region 29.1	lanthipeptide-class-v	1	21,377	hassallidin C	NRP + Saccharide:Hybrid/tailoring	6%
Region 45.1	NRPS-like	1	15,609			
Region 72.1	T1PKS	1	14,314			
Region 85.1	terpene	309	15,942			
Region 139.1	terpene	1	18,996	geosmin	Terpene	100%
Region 148.1	terpene	2,406	24,313			
Region 153.1	spliceotide , RRE-containing	1	3,755			
Region 154.1	terpene	1	10,711			
Region 156.1	T1PKS	1	35,934	bartoloside 2 / bartoloside 3 / bartoloside 4	Other	27%
Region 174.1	hglE-KS	1	40,209	heterocyst glycolipids	Other	57%
Region 178.1	cyanobactin	1	3,876	anacyclamide D8P	RiPP	40%
Region 338.1	hglE-KS	1	35,858	heterocyst glycolipids	Other	85%
Region 399.1	lanthipeptide-class-v	1	23,034			

No secondary metabolite regions were found in these records:
k127_1003028

Fig. 21. Identified secondary metabolite coding genes (marked as regions) in MAG bin.111 from sample TO_26.

Select genomic region:
Overview 59.1 166.1 182.1 185.1 206.1 220.1 225.1 229.1 235.1 284.1 339.1 372.1 485.1 622.1 630.1 683.1 691.1

Identified secondary metabolite regions using strictness 'relaxed'

Region	Type	From	To	Most similar known cluster		Similarity
Region 59.1	T1PKS	1	6,650			
Region 166.1	NRPS	1	5,980	microcystin LR	NRP + Polyketide:Modular type I	17%
Region 182.1	NRPS , NRPS-like	1	4,244	micropeptin K139	NRP + Polyketide	37%
Region 185.1	T1PKS	1	5,908			
Region 206.1	NRPS	1	2,580	anabaenopeptin NZ857 / nostamide A	NRP	100%
Region 220.1	hglE-KS	1	3,052			
Region 225.1	NRPS	1	3,057	aeruginosin 98-A / aeruginosin 98-B / aeruginosin 98-C	NRP	14%
Region 229.1	hglE-KS	1	9,468	ambiguine P	Alkaloid	6%
Region 235.1	T1PKS	1	10,208			
Region 284.1	T1PKS	1	3,978			
Region 339.1	T3PKS	1	4,959	merocyclophane C / merocyclophane D	Polyketide	22%
Region 372.1	lanthipeptide-class-v	1	3,302			
Region 485.1	hglE-KS	1	10,203			
Region 622.1	terpene	1	2,823			
Region 630.1	terpene	1	7,648			
Region 683.1	spliceotide , RRE-containing	1	4,328			
Region 691.1	lanthipeptide-class-v	1	10,344	hassallidin C	NRP + Saccharide:Hybrid/tailoring	6%

No secondary metabolite regions were found in these records:
k127_1008350

Figure 22. Identified secondary metabolite coding genes (marked as regions) in MAG bin.078 from sample TO_56.

Region	Type	From	To	Most similar known cluster	Similarity
Region 237.1	terpene	1	7,052		
Region 547.1	terpene	1	3,173		

Figure 23. Genes coding for putative secondary metabolites in bin.011 from sample TO_56. The MAG (bin.011) was taxonomically assigned to the family *Cyanobaiacae*, which is composed of small unicellular cyanobacteria.

Given that several contigs were identified as possessing a *geoA* sequence similar to ‘uncultured bacterium clone AQC_0126’ (Table 9), MAGs from the same metagenomic sample were searched for the respective contigs with these *geoA* sequences in an attempt to better identify the populations that truly possess these genes in LM. Contig k127_142770 was identified in bin.057 from sample TO_58 via manual search queries in Kbase. Using GTDB-TK v1.7.0 the bin was indeed affiliated with the phylum *Chloroflexota*, and further taxonomically assigned to class *Anaerolineae*, order *Thermoflexales*, and family J036 (genus unknown).

3.8 Algaecide EarthTec addition experiments

Initial bottle experiments where collected LM water was treated with EarthTec (0.05 mg/L CuSO4 final concentration) showed an increase in geosmin concentration after 48 h of incubation (Fig. 24). Of note, the water used in this experiment was from 6/8/20 (study day 160) when *Dolichospermum* and *Dolichospermum geoA* sequences were abundant (Fig. 9, 16).

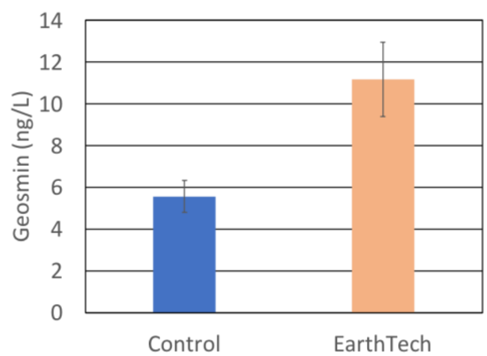


Figure 24. Concentration of geosmin in incubated LM water after 48h of incubation with (0.05 mg/L) or without EarthTech (Control). 2-MIB was also measured but was undetectable across all samples.

In bottle incubations with gradient additions of EarthTec, additions of ~0.005 mg/L CuSO4 (final conc.) or greater caused parallel increases in geosmin concentration in all bottle incubation experiments (Fig. 25). An increase in 2-MIB concentration similarly occurred in one of the three experiments (Fig. 25E). The highest addition of EarthTec (0.05 mg/L CuSO4) caused the highest increases in geosmin (and in one experiment 2-MIB) relative to the no add control (Fig. 25). In one experiment, the increase in geosmin due to 0.05 mg/L CuSO4 addition was 15-fold higher than that of the no addition control (Fig. 25A).

Likely the variation in T&O compound released is due to differences in the communities present at the start of the bottle experiments, for example, *Dolichospermum* was most prevalent in LM water collected (and used for experiments) on days 195 and 265 (7/13/20 and 9/21/20 respectively) (Figs. 9, 16). Declines in Chl a (total phytoplankton biomass) were evident in 7/13/20 and 9/21/20 experiments with increasing EarthTec addition, indicating the algaecide was generally effective at reducing total phytoplankton biomass. Declines in Chl a occurred with EarthTec additions of 0.0125-0.025 mg/L CuSO4 final concentration (Fig. 25). Small phytoplankton abundances (determined by FCM) minimally changed with increasing EarthTec. Some declines in FCM derived abundances were evident amongst phycocyanin rich picocyanobacteria (PC-rich) and the largest of the picoeukaryotic phytoplankton groups (Euk3), but did not decrease in parallel with Chl a in response to higher EarthTec addition (Fig. 25). This indicates larger phytoplankton, e.g. filamentous cyanobacteria and also possibly eukaryotic phytoplankton, accounted for the majority of the Chl a signal and were likely more sensitive to the EarthTec treatment than smaller phytoplankton. Typically, small cyanobacteria are highly sensitive to elevated concentrations of heavy metals, especially Cu (Brand et al. 1986; Coclet et al. 2018), but perhaps in LM these populations have a higher tolerance for Cu than other freshwater systems, or the Cu complexed primarily with the relatively higher amount of large phytoplankton biomass.

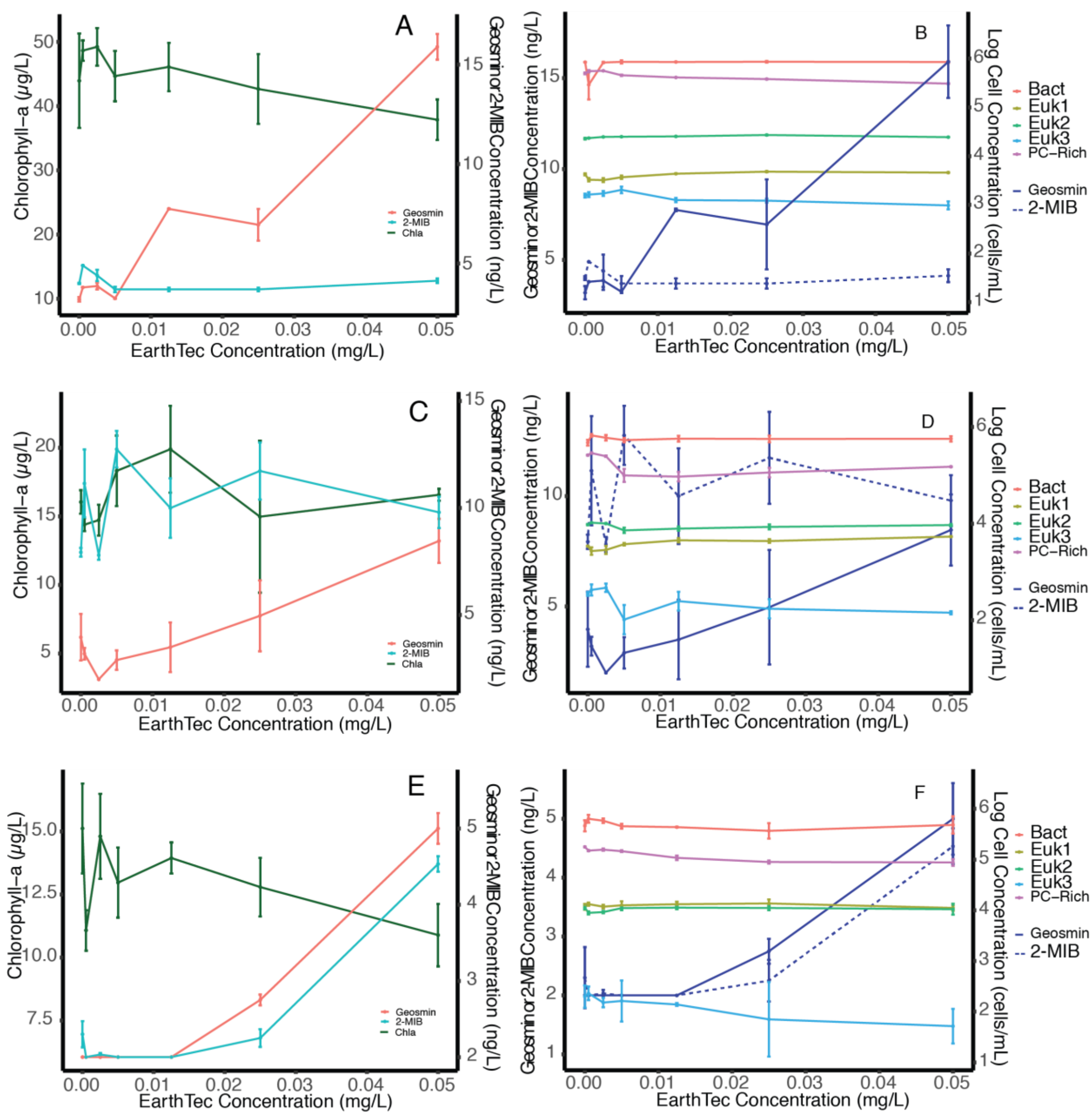


Figure 25. Chl a, FCM and T&O data from EarthTec addition experiments started on 7/13/20 (study day 195; Panels A, B), 8/24/20 (study day 237; Panels C,D), and 9/21/20 (study day 265; Panels E,F). Average values for triplicate bottles are shown, with error bars representing the standard deviation amongst the triplicate values.

4. Discussion

4.1 Contextualizing LM conditions in 2020

While average air temperatures in the southeastern US have increased since ~1960's (NOAA 2022), the air temperature during samplings in 2020 were comparable to the average across 2015-2020 data. Thus, temperature conditions in 2020 appear typical with respect to the last 6-years (Table 2). The average discharge rate for 2020 was also comparable to the 6-year average (~186 ft^3/s) (Table 2). Overall, conditions in 2020 were representative of what is typical (the average) for LM over the last 6 years.

4.2 T&O observations 2015-2020

Arguably geosmin is the T&O compound of primary concern in LM as its annual average and maximum concentration during 2015-2020 was higher than that of 2-MIB (Table 4; Figs. 5-6). Similarly, geosmin levels were higher and more prevalent in two Piedmont South Carolina reservoirs (Journey et al. 2011, 2013), which together emphasize that geosmin as a T&O compound of concern in the Piedmont of the Carolinas and reason to understand more about its sources and controls on production. The maximum geosmin concentrations detected from LM are comparable to maximum values reported for South Carolina reservoirs (Journey et al. 2011), ca. 100 ng/L, but for perspective the LM maximum is markedly lower than maxima recorded from southeastern aquaculture ponds (ca. 9 µg/L; Auburn, Alabama) (van der Ploeg et al. 1992), as well as reservoirs impacted by dense cyanobacterial blooms in Asia, Africa and Australia (all >1µg/L) (Wnorowski and Scott 1992; Jones and Korth 1995; Li et al. 2010; Lee et al. 2020).Speculatively, conditions in reservoirs of the Piedmont southeastern USA favor microbial sources of the compound, especially *Dolichospermum* spp. (more on this below).

The partitioning of higher geosmin in spring and early summer (e.g. May to June) versus 2-MIB in mid to late summer (June, July, August) is very evident in the 2015-2020 LM dataset (Figs. 5-6,9) and presumably due to shifting in plankton communities within the water column over time. Geosmin was similarly elevated during the late spring and early summer in South Carolina reservoirs (Journey et al. 2013). Gauge height and discharge rate showed no strong positive or negative relationship with geosmin concentration (Fig. 12), which suggests that terrestrial inputs of geosmin through runoff are minimal. To our knowledge though, no transect sampling of T&O compounds from shore to the central reservoir has been conducted at LM. Another interesting feature within the geosmin data is the large change in concentration week to week during spring-summer, which was more overt than in the 2-MIB data (Figs. 5-6). Without weekly monitoring spikes in geosmin could be missed. Thus, weekly monitoring – as is currently done by CDWM – will help inform decisions on drinking water treatment, whether or not to draw water from an alternative source (Little River), and/or warning the public of changes in T&O quantities. Ecologically, the relatively rapid geosmin fluctuations are interesting and likely result from change in the growth or metabolism of geosmin producers (Naes and Post 1988) and/or tight coupling of geosmin production and consumption or abiotic flux (Li et al. 2012; Guttman and van Rijn 2012).

While 2-MIB is on average less abundant in LM (ca. 5.9 ng/L), during summer into early fall it can be the dominant T&O compound, e.g. on 8/7/17 it was ca. 6-fold higher than geosmin, and >10 ng/L) (Fig. 6). Accordingly, 2-MIB should not be ignored as a T&O compound that may be off-putting to Durham County consumers. The elevated levels of 2-MIB in summer and early fall, generally later than geosmin, indicates sources of 2-MIB are likely distinct from those of geosmin in LM (more to this in the section below). This generally follows suit with the published literature as there are few cyanobacterial populations known to be producers of geosmin and 2-MIB; instead production of one or the others is more common (Watson et al. 2016). The negative correlation between discharge and 2-MIB (RDA plot) suggests the source is likely favored under more stratified conditions (Fig. 12). Also peaks in 2-MIB over time were less acute, i.e. not rapidly increasing/decreasing, compared to geosmin - with a notable exception being in late June 2018 (Fig. 6). Speculatively, the more gradual change reflects differences in production and consumption, or abiotic flux compared to geosmin (Figs. 5,6). In a prior study where production and loss of geosmin and 2-MIB was tracked in cyanobacterial cultures, 2-MIB often had the larger turnover time (half-life) and differences in biodegradation was a significant factor (Li et al. 2012). Potentially bacterial consumption of 2-MIB in LM is lower than geosmin.

4.3 Microbial sources of T&O

Predicting the microbial sources of T&O based on genus information (e.g. microscopic counts of *Dolichospermum* spp.) is potentially problematic since production varies at the species level, i.e. not all *Dolichospermum* spp. produce geosmin or possess *geoA* (Jüttner and Watson 2007; Watson et al. 2016). Identifying and studying *geoA* sequences allows for a more direct assessment of who are the putative producers of geosmin, rather than speculating on the capabilities and genes in *in-situ* populations based on that of laboratory isolates (Swan et al. 2013). On the basis of genetics, several cyanobacteria and non-cyanobacteria are capable of geosmin production in LM (Table 6,9).

Both covariation in geosmin and *Dolichospermum* biovolume (Fig. 9) as well as relative abundances of *geoA* sequences (Fig. 16) and metagenomic data (Table 9, Fig. 20-21) point to *Dolichospermum* as the genus responsible for geosmin concentrations >10 ng/L in LM. Both the ASV and metagenomic data indicates that >4 different *Dolichospermum* populations possessing *geoA* were prevalent in LM in 2020, with one *geoA* ASV being particularly dominant (Table 7; Figs. 20-21). The ecologies of these different populations could not be fully determined here but could be targeted in the future using the recovered *geoA* sequences (ASVs) and a more quantitative molecular approach like quantitative PCR (qPCR) to track the change in their abundance with depth or change in other environmental variables.

No prior studies document production of geosmin by *Chloroflexota* (previously called Chloroflexi as well as 'green non-sulfur bacteria') (Overmann 2008). Recent work highlights the presence of terpene-biosynthesis related genes (geosmin synthase falls under this umbrella) in select *Chloroflexota* strains, but not specifically *geoA* (Yamada et al. 2012). Accordingly, the discovery of a *Chloroflexota*-affiliate with *geoA* in LM water is novel and may reveal more about the function of geosmin production in microbes with further investigation into the ecology of these likely particle-associated *Chloroflexota*. From a management perspective, these bacteria could have a role in elevating geosmin concentrations in late summer following periods of stratification. *Chloroflexota*-affiliated populations in lakes and freshwater systems occupy diverse niches and with it exhibit diverse metabolic capabilities, including aerobic anoxygenic photosynthesis (Overmann 2008; Mehrshad et al. 2018). Some populations are known to dominate in the hypolimnion while others can persist in shallower epilimnion waters (Denef et al. 2015; Mehrshad et al. 2018). The *Chloroflexota*-affiliated MAG (bin.057) was obtained from an 8 µm pore-size filter (sample TO_58) which suggests the population may thrive in association with abiotic particles or larger plankton (Table 9). Other *Chloroflexota* belonging to *Anaerolineae* (as does bin.057) are anticipated to be anaerobic heterotrophs based on genomic sequencing and occur more frequently in the hypolimnion (Denef et al. 2015; Mehrshad et al. 2018). Further analysis is required to unravel the lifestyle of the bin.057 lineage. However, it is noteworthy that in late August (study day 244) *Chloroflexota* (Chloroflexi) and *Chlorobi* sequences became more abundant in 8 and 0.8 µm filter samples respectively and geosmin concentrations increased relative to the prior August sampling, although not above 10 ng/L (Fig. 17, Table 9). Perhaps finer sampling of the water column (surface to bottom) at LM and in-culture based estimates of geosmin production (following successful cultivation from natural waters) could provide greater insight of their contribution to geosmin production on the whole as well as concentrations at depth.

Debates continue about the use of geosmin as an indicator of co-occurring cyanotoxins, especially microcystin (Graham et al. 2008; Harris and Graham 2017; Shang et al. 2018). This approach does not appear worthwhile for LM as cyanobacterial populations possessing *geoA* appear to lack genes coding for microcystin production (*mcy* genes) (Fig. 22). Abundant *Dolichospermum* MAGs representing abundant populations in the Chowan River similarly possessed *geoA* and lacked *mcy* genes in a recent 2020 study (R. Paerl per. comm.). Notably though, genes linked to the production of other putative toxins were detected in LM MAGs possessing *geoA* and assigned to the genus *Dolichospermum* (Figs. 20-21). Non-ribosomal peptides (NRPS) aeruginoside, anacyclamide A, cyanopeptin and anabaenopeptin were detected in these bins and the latter of these putative toxins, anabaenopeptin, is argued to exhibit greater toxicity than microcystin (congener microcystin-RR) (Lenz et al. 2019). Accordingly, geosmin could be a useful indicator of cyanotoxins beyond microcystin. Cyanobacterial toxins (qualitatively or quantitatively) are not monitored in LM or Williams or Brown Water treatment plants to our knowledge. Based on the metagenomic sequence libraries obtained here, genes linked to the production of multiple cyanotoxins are present in abundant LM cyanobacteria. This finding may serve as motivation for future monitoring of cyanotoxins in source water as well as post-treatment processes.

The microbial sources of 2-MIB were more difficult to pinpoint. Only one *mic* gene was recovered from LM samples (Table 9), which inherently indicates 2-MIB producers are rarer – at least based on our sampling method. Potentially the *mic* gene came from a *Microcoleus* population (N-fixing filamentous cyanobacteria), although microscopy data suggests other filamentous cyanobacteria *Planktothrix*, *Arthrospira*, and *Limnothrix* were abundant during the 2-MIB peak in 2020 – so a population within one or more of the genera may produce 2-MIB. Assays (qPCR) could be developed to target the recovered *mic* gene and further clarify where the populations thrives in LM (in the water column, shallow nearshore benthos or central lake) or if it associates with a particular size fraction of plankton. Further, more metagenomic sequencing of LM plankton during periods of elevated 2-MIB could help resolve the source populations, as well as their way of life (e.g. metabolic profiles), more exhaustive testing of published *mic* PCR primers, or the development of new primers including the newly obtained *mic* sequence could be done to recover *mic* sequences from LM. Due to resource limitations weekly samples were not microscopically examined during late summer and early fall. Weekly biovolume data in parallel with 2-MIB measurements may be able to uncover covariations that are informative of the source, as in the case of *Dolichospermum* and geosmin (Fig. 9B).

4.4 No evidence of T&O production by picocyanobacteria

A small set published studies point to picocyanobacteria as possible sources of T&O compounds (Watson et al. 2016), and initially we hypothesized these small cyanobacteria were geosmin sources in LM based on FCM and T&O measurements (Fig. 2). The flow cytometry and genetic data obtained in this study strongly suggests picocyanobacteria are not sources of geosmin or 2-MIB in LM. Preliminary flow cytometry and geosmin concentration data (CDWM) indicated that *Synechococcus*-like cells may be sources of geosmin in late spring to early summer with the largest change in temperature towards warmer summer months (Fig. 2). This was not the case in 2020 (Figs. 14). Furthermore, numerous MAGs assigned to picocyanobacterial (or slightly larger unicellular cyanobacterial) genera were obtained from LM metagenomes, and none possessed *geoA* or *mic* genes (Fig. 23). It may be that rare picocyanobacterial populations possess the genes - however those likely would have little impact on total T&O production.

The lack of T&O genes aside, the presence of picocyanobacteria in LM waters is an indication of 'good' water quality– more specifically that heavy metals such as Cu and Cd are not elevated in LM. Picocyanobacteria are very sensitive to increases in bioavailable heavy metal concentrations compared to other phytoplankton (Brand et al. 1986; Coclet et al. 2018) and thus can be considered sentinels for increased metal loading or pollution. Overall, moderate picocyanobacteria abundances should be considered a component of a 'healthy' phytoplankton community in LM.

4.5 Cyanobacterial N-fixers as sources of T&O and associated future challenges

Multiple lines of evidence presented here point to filamentous N-fixing cyanobacteria as key sources of T&O in LM. Diverse N-fixing cyanobacteria are well known to impact water quality worldwide through excess biomass production, generation of toxic compounds, and T&O compound production (Chorus et al. 2000; Carmichael and Boyer 2016; Huisman et al. 2018). While they generally are capable of using diverse forms of N (inorganic and organic), N-fixing cyanobacteria are thought to outcompete other phytoplankton under low N availability relative to other key nutrients, e.g. P and trace metals (Paerl et al. 2016; Baker et al. 2018). To our knowledge N-fixation rates have not been measured in LM, so it is difficult to estimate how much biomass, especially biomass of T&O producing populations, is supported by N-fixation. Nonetheless, reductions in P and/or trace metals entering LM may help limit the growth of N-fixing and T&O producing cyanobacteria; however, accomplishing this is often challenging given that 'legacy' P accumulates in reservoirs (Sharpley et al. 2013), instead of degassing out of the system as is the case for N (through denitrification)(Seitzinger 1988). Additionally, copious amounts of Fe and other trace metals are presumably available in LM through erosion and runoff. Given the age of LM it is anticipated that substantial P has accumulated in sediment over time and is released each year to the water column - particularly during periods of deep-water anoxic conditions which have recently been observed in LM (per. comm. Dr. Linda Ehrlich). Some actions may help combat the growth of N-fixing cyanobacteria: 1) aeration of deep water to reduce P release under anoxic conditions, 2) installation of buffers or traps to reduce sediment transport and nutrients into LM, 3) Lanthanide clay treatments - however on large scales this may be ineffective and the ecosystem impacts of these additions, especially to benthic organisms, is debated (Herrmann et al. 2016), 4) sediment dredging - although the size of LM may make dredging extremely costly.

Looking to the future, N-fixing cyanobacteria are likely to continue to thrive in LM in large part due to plentiful supplies of P relative to N (Fig. 11). In addition, increased stratification and warm temperatures tends to promote N-fixing cyanobacteria (Paerl and Huisman 2008; Paerl et al. 2016), which are likely to occur if increased warming in the southeastern US and Durham County continues (NOAA 2022). Promotion of N-fixing filamentous cyanobacteria is likely to simultaneously elevate T&O concentrations. While the datapoints are limited, average geosmin concentration appears to be increasing in LM (Fig. 7) and likely this is associated with success of N-fixing cyanobacteria, such as *Dolichospermum* (Figs. 9,16). More broadly, *Dolichospermum* and its relative *Anabaena* have been the focus of increasing numbers of published 'algal bloom' research in the last decade (Paerl et al. 2020a). Considering all the data presented and trends in the literature, N-fixing cyanobacteria should receive particular focus in future monitoring efforts especially since they may increase in prominence and thereby increase T&O concentrations in the future.

4.6 Potential implementation of molecular tools for future monitoring

Rapid detection and tracking of T&O producers would enable the quickest possible responses to T&O spikes. The *geoA* and one mic sequences obtained here, especially *Dolichospermum geoA* sequences could be leveraged for rapid detection assays. Decision making based on out-sourced chemical measurements of geosmin and 2-MIB requires at least two days due to sample shipping and external analysis (e.g. Eurofins) – as is done by CDWM. Some management groups run in-house chemical analyses (gas chromatography) to determine T&O compound concentrations, however this is not implemented by numerous management groups presumably due to significant associated costs (initial investment into equipment, then operational costs - standards, operators).

Molecular (PCR-based) assays have potential as an alternative - functioning as an indicator of increased potential for T&O production (specifically geosmin in LM). Qualitative and quantitative assays, e.g. standard PCR and gel electrophoresis versus quantitative PCR, could be used depending on the budgets of management groups to identify increases in populations of concern. For instance, a relatively inexpensive PCR assay could be developed that targets the most abundant *Dolichospermum geoA* sequences obtained in this study (Table 6) and amplification can be optically determined using a standard gel imager. Based on the relative abundances of *Dolichospermum geoA* sequences from our 2020 samplings, increases in gene abundances of 4 to 8-fold (or PCR band intensity when examined via gel imager) over a week to two weeks would serve as a warning for elevated geosmin (Fig. 16).

There are caveats to consider in regard to implementing molecular assays for monitoring. They do require a DNA extraction step (and reagents, equipment) prior to performing the assays; however, there are highly simplified DNA extraction kits now available (Qiagen, Zymo, and other vendors sell easy to follow extraction kits). Additionally, an initial investment in equipment (thermal cycler, micropipettes, etc.) is needed to perform PCR or qPCR. Of these, qPCR is a larger investment, but it does not rely on gel electrophoresis and optical analysis of bands which can make data interpretation easier for less experienced lab workers. It is also worth emphasizing that PCR or qPCR assays would need to be run over time as part of 'standard' monitoring since single sample assessments of gene abundances will not be a reliable indicator (Fig. 16). A cost benefit analysis of molecular analyses versus contracted geosmin/2-MIB measurements was not performed here. Weekly microscopic analysis of cyanobacterial biovolume should also be factored into this cost benefit analysis, given the observed relationships seen here between cyanobacterial biomass and geosmin (Fig. 9); that said, microscopy is laborious and requires a skilled observer. Overall, a cost benefit analysis would be worthwhile - particularly considering the high cost per samples for geosmin/2-MIB measurements through contractors (Eurofins ~$125/sample).

4.7 Considerations regarding future EarthTec treatment of LM water

Additions of 0.05 mg/L $CuSO_4$ EarthTec, a concentration used by CDWM in the past, clearly lead to increased T&O compound concentrations (largely geosmin in our experiments) into the surrounding water. Additions at concentrations <0.05 mg/L $CuSO_4$ will be useful for reducing geosmin or 2-MIB concentrations entering the CDWM water treatment pipeline. Specifically, EarthTec additions of 0.0125 and 0.025 mg/L showed subtle reductions in Chl a concentration while not elevating T&O concentrations as high as 0.05 mg/L additions (Fig. 25). As a result, reduced EarthTec additions are recommended.

Explanations for 0.05 mg/L EarthTec additions (highest, yet comparable to additions used by CDWM) causing increased T&O concentrations in our experiments include: 1) T&O producing cyanobacteria were lysed (damaged to induce leakage of metabolites), but the majority of phytoplankton biomass was unaffected - possibly due to increased tolerance to $CuSO_4$, which is well-documented in eukaryotic phytoplankton versus cyanobacteria (Brand et al. 1986); 2) lysis or damage to cells caused release of nutrients that stimulated growth and production of T&O producing populations. The latter seems less likely, as bacterial abundances were not elevated in the 0.05 mg/L additions (Fig. 25) and that would be expected to some degree if organics were extensively released by dying phytoplankton. EarthTec was effective in reducing Chl a concentration in two of the three experiments and in these experiments there was notable release of geosmin (Fig. 25), including the first experiment during which *Dolichospermum* was abundant (day 195; Figs. 9, 16). Considering the now recognized importance of *Dolichospermum* spp. as a geosmin source (Figs. 9, 16), additional testing of the efficacy of EarthTec at varying concentrations on *Dolichospermum* growth and subsequent T&O release would be especially useful.

5. Summary

In near-surface waters of Lake Michie (LM), a drinking water resource for Durham County, NC, geosmin reaches higher concentrations than 2-MIB in LM euphotic waters and generally peaks in spring and early summer based on 6 years of time series data. This is similar to recent reports of T&O in reservoirs in the Piedmont of South Carolina (SC). In contrast to SC reservoirs, 2-MIB frequently exceed the human detection limit in LM and thus remains a concern. Microscopic and genetic evidence point to growth of *Dolichospermum* spp. as a key source of geosmin in LM. Multiple *geoA* sequences were obtained that can be used in the future to monitor problematic geosmin producing populations in LM, but also probe for their prevalence in other systems in the regions as well as globally. The first evidence of *Chloroflexota* affiliates as geosmin producers was obtained from LM metagenomes and potentially contribute the most to geosmin during summer and late fall, and possibly at depth. 2-MIB producers were notably rarer and difficult to pinpoint in LM, but one cyanobacterial 2-MIB synthase gene (*mic*) – likely from a filamentous cyanobacterium - was recovered and could be targeted in future studies. No DNA sequence data or flow cytometry data implicated picocyanobacteria as sources of T&O compounds. Nonetheless they are prevalent in LM which generally is an indicator of low heavy metal concentrations and good water quality. A number of factors suggest T&O production (especially geosmin) in LM is likely to persist and potentially worsen in the future: (1) total N:P was often < 9, (2) removing legacy P from older reservoirs in a physical and fiscal challenge, (3) N-fixing cyanobacteria are key primary sources of geosmin, (4) the average geosmin concentration is increasing, and (5) annual average temperature in Durham County continues to increase. Last, additions of 0.05 mg/L $CuSO_4$ in the form of EarthTec in bottle experiments caused increases in geosmin or 2-MIB concentration, rather than decreasing their concentrations as intended. Reduced additions of 0.0125 mg/L to 0.025 mg/L $CuSO_4$ (final concentration) are recommended instead based on Chl a responses and reduced flux of geosmin (and 2-MIB at times) into the surrounding water. Follow on exposure experiments that specifically examine the effects of EarthTec treatments specifically on growth of and geosmin release by *Dolichospermum* spp. is also recommended.

6. Conclusions and recommendations

The results of this study highlight geosmin and 2-MIB as a problematic T&O compounds in the Lake Michie reservoir. Filamentous N-fixing cyanobacteria, esp. *Dolichospermum* sp., appear to be key producers of geosmin which is in line with recent results from South Carolina Piedmont reservoirs and likely indicates these cyanobacteria are of concern in the broader Piedmont and southeastern US. It is ecologically and microbiologically interesting to discover that *Chloroflexota* affiliates produce geosmin based on our metagenomic data. This motivates further studies to test if their production is significant compared to cyanobacteria and how it relates to their success in freshwater lakes – where in some cases they are significant primary producers.

Key microbial sources of 2-MIB were more difficult to pinpoint in LM; thus, this remains largely unaddressed and should be explored in future investigations. Nonetheless, one population with the capacity to produce 2-MIB was identified and can now be targeted in future studies to better resolve their identity and ecology (e.g. determine what conditions promote 2-MIB production by the organism).

Several lines of evidence indicate that geosmin production will continue to be an issue and may worsen in LM, which increases pressure to consider mechanisms to reduce P, develop or invest in tactics to remove T&O, and address water quality concerns of people. Molecular assays and/or microscopy-based monitoring could aid 'fast-response action' by water managers to deal with rises in geosmin (and 2-MIB). Looking beyond LM, likely geosmin is, or will be, an issue within other reservoirs as well as smaller ponds within the southeastern US that exhibit similar water chemistry (e.g. total N: total P <9) and physical conditions. Thus, it is anticipated that concerns about T&O compounds will rise in the future.

While the application of EarthTec to treat water containing T&O producers is simple, the response is more complex and the opposite of what users might expect (e.g. decrease in T&O with increased EarthTec addition). The use of EarthTec should be preliminarily assessed with gradient additions to identify the concentrations that yield a most desirable outcome before use in large scale treatment. Here we show the benefit of reducing treatment concentrations to obtain the desired reduction in T&O concentration while also reducing or stabilizing phytoplankton biomass. For LM, it is recommended that treatments of 0.0125 to 0.025 mg/L CuSO4 final concentration, in the form of EarthTec, be used when necessary.

Last, the metagenomic sequencing employed in this study not only directly identified populations capable of producing geosmin and 2-MIB but also obtained genetic information on the lifestyles (e.g. N-fixation) and broader metabolic potential of the T&O producers. For example, LM cyanobacteria were found to potentially produce well-studied and/or understudied cyanotoxins (microcystin, aeruginoside, anacyclamide A, cyanopeptin and anabaenopeptin). Accordingly, it is recommended that microcystin concentrations (in the least) be investigated in LM source water but also at other points downstream the treatment pipeline. Lower cost chemical methods are being developed for the detection of less-studied cyanotoxins (aeruginoside, anacyclamide A, cyanopeptin and anabaenopeptin) (Roy-Lachapelle et al. 2021); however, the most prominent methodology requires advanced equipment (mass spectrometers, liquid chromatographs) not readily available to most water managers (Beversdorf et al. 2018; Roy-Lachapelle et al. 2019). Given that genes coding for the production of these compounds are more prevalent than those coding for relatively well-studied toxins (e.g. microcystin), it is recommended that this class of compounds (esp. the four listed above) be considered for monitoring in the future.

7. References cited

Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl. Environ. Microbiol. **71**:

8966–8969. doi:10.1128/AEM.71.12.8966-8969.2005
Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2
Arkin, A. P., R. W. Cottingham, C. S. Henry, and others. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat. Biotechnol. **36**: 566–569. doi:10.1038/nbt.4163
Baird, R., & Bridgewater, L., ed. 2017. Standard methods for the examination of water and wastewater, 23rd ed. American Public Health Association (APHA), American Water Works Association (AWWA), and Water Environment Federation (WEF).
Baker, B. C., A. E. Wilson, and J. T. Scott. 2018. Phytoplankton N2-fixation efficiency and its effect on harmful algal blooms. Freshw. Sci. **37**: 264–275. doi:10.1086/697530
Beversdorf, L. J., K. Rude, C. A. Weirich, and others. 2018. Analysis of cyanobacterial metabolites in surface and raw drinking waters reveals more than microcystin. Water Res. **140**: 280–290. doi:https://doi.org/10.1016/j.watres.2018.04.032
Blin, K., S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema, and T. Weber. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. **49**: W29–W35. doi:10.1093/nar/gkab335
Bowers, R. M., N. C. Kyrpides, R. Stepanauskas, and others. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. **35**: 725–731. doi:10.1038/nbt.3893
Brand, L. E., W. G. Sunda, and R. R. L. Guillard. 1986. Reduction of marine phytoplankton reproduction rates by copper and cadmium. J. Exp. Mar. Bio. Ecol. **96**: 225–250.
Brettin, T., J. J. Davis, T. Disz, and others. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. **5**: 8365. doi:10.1038/srep08365
Brussaard, C. P. D., J. P. Payet, C. Winter, and M. G. Weinbauer. 2010. Quantification of aquatic viruses by flow cytometry, p. 102–109. *In* Manual of aquatic viral ecology ASLO. American Society of Limnology and Oceanography.
Bushnell, B. 2014. BBTools software package. URL http//sourceforge. net/projects/bbmap **578**: 579.
Cain, J. L. 2017. Water Quality and Stratification in Jordan Lake: Assessment of Spatial, Temporal and Inter-Annual Variability. North Carolina State University.
Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods **13**: 581–583. doi:10.1038/nmeth.3869
Callieri, C. 2008. Picophytoplankton in Freshwater Ecosystems: The Importance of Small-Sized Phototrophs. Freshw. Rev. **1**: 1–28.
Carmichael, W. W., and G. L. Boyer. 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. Harmful Algae **54**. doi:10.1016/j.hal.2016.02.002
Chorus, I., I. R. Falconer, H. J. Salas, and J. Bartram. 2000. Health risks caused by

freshwater cyanobacteria in recreational waters. J. Toxicol. Environ. Heal. - Part B Crit. Rev. **3**: 323–347. doi:10.1080/109374000436364

Churro, C., A. P. Semedo-Aguiar, A. D. Silva, J. B. Pereira-Leal, and R. B. Leite. 2020. A novel cyanobacterial geosmin producer, revising GeoA distribution and dispersion patterns in Bacteria. Sci. Rep. **10**: 8679. doi:10.1038/s41598-020-64774-y

City of Durham Water Management. 2007. Water Management - Archived Webpage. 2007.

Coclet, C., C. Garnier, F. Delpy, D. Jamet, G. Durrieu, C. Le Poupon, M. Mayer, and B. Misson. 2018. Trace metal contamination as a toxic and structuring factor impacting ultraphytoplankton communities in a multicontaminated Mediterranean coastal area. Prog. Oceanogr. **163**: 196–213.

Collier, J. L. 2000. Flow Cytometry and The Single Cell In Phycology. J. Phycol. **36**: 628–644. doi:10.1046/j.1529-8817.2000.99215.x

Denef, V. J., R. S. Mueller, E. Chiang, J. R. Liebig, and H. A. Vanderploeg. 2015. Chloroflexi CL500-11 Populations That Predominate Deep-Lake Hypolimnion Bacterioplankton Rely on Nitrogen-Rich Dissolved Organic Matter Metabolism and C1 Compound Oxidation. Appl. Environ. Microbiol. **82**: 1423–1432. doi:10.1128/AEM.03014-15

Durham Parks & Recreation. 2022. Lake Michie. 2022.

Fan, J., L. Ho, P. Hobson, and J. Brookes. 2013. Evaluating the effectiveness of copper sulphate, chlorine, potassium permanganate, hydrogen peroxide and ozone on cyanobacterial cell integrity. Water Res. doi:10.1016/j.watres.2013.05.057

Gaulke, A. K., M. S. Wetz, and H. W. Paerl. 2010. Picophytoplankton: a major contributor to planktonic biomass and primary production in a eutrophic, river-dominated estuary. Estuar. Coast. Shelf Sci. **90**: 45–54.

Gibson, C. E. 1972. The Algicidal Effect of Copper on a Green and a Blue-Green Alga and some Ecological Implications. J. Appl. Ecol. **9**: 513. doi:10.2307/2402449

Giglio, S., J. Jiang, C. P. Saint, D. E. Cane, and P. T. Monis. 2008. Isolation and characterization of the gene associated with geosmin production in cyanobacteria. Environ. Sci. Technol. **42**: 8027–8032. doi:10.1021/es801465w

Godo, T., Y. Saki, Y. Nojiri, and others. 2017. Geosmin-producing Species of Coelosphaerium (Synechococcales, Cyanobacteria) in Lake Shinji, Japan. Sci. Rep. **7**. doi:10.1038/srep41928

Graham, J. L., K. A. Loftin, A. C. Ziegler, and M. T. Meyer. 2008. Guidelines for design and sampling for cyanobacterial toxin and taste-and-odor studies in lakes and reservoirs.

Guildford, S. J., and R. E. Hecky. 2000. Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: Is there a common relationship? Limnol. Oceanogr. **45**: 1213–1223. doi:https://doi.org/10.4319/lo.2000.45.6.1213

Guttman, L., and J. van Rijn. 2008. Identification of conditions underlying production of geosmin and 2-methylisoborneol in a recirculating system. Aquaculture **279**: 85–91. doi:https://doi.org/10.1016/j.aquaculture.2008.03.047

Guttman, L., and J. van Rijn. 2012. Isolation of bacteria capable of growth with 2-methylisoborneol and geosmin as the sole carbon and energy sources. Appl. Environ. Microbiol. **78**: 363–370. doi:10.1128/AEM.06333-11

Harris, T. D., and J. L. Graham. 2017. Predicting cyanobacterial abundance, microcystin, and geosmin in a eutrophic drinking-water reservoir using a 14-year dataset. Lake Reserv. Manag. **33**: 32–48. doi:10.1080/10402381.2016.1263694

Herrmann, H., J. Nolde, S. Berger, and S. Heise. 2016. Aquatic ecotoxicity of lanthanum – A review and an attempt to derive water and sediment quality criteria. Ecotoxicol. Environ. Saf. **124**: 213–238. doi:https://doi.org/10.1016/j.ecoenv.2015.09.033

Hillebrand, H., C.-D. Dürselen, D. Kirschtel, U. Pollingher, and T. Zohary. 1999. BIOVOLUME CALCULATION FOR PELAGIC AND BENTHIC MICROALGAE. J. Phycol. **35**: 403–424. doi:https://doi.org/10.1046/j.1529-8817.1999.3520403.x

Houle, S., K. K. Schrader, N. R. Le François, Y. Comeau, M. Kharoune, S. T. Summerfelt, A. Savoie, and G. W. Vandenberg. 2011. Geosmin causes off-flavour in arctic charr in recirculating aquaculture systems. Aquac. Res. **42**: 360–365. doi:https://doi.org/10.1111/j.1365-2109.2010.02630.x

Hugerth, L. W., J. Larsson, J. Alneberg, M. V. Lindh, C. Legrand, J. Pinhassi, and A. F. Andersson. 2015. Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biol. **16**: 279. doi:10.1186/s13059-015-0834-7

Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser. 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. **16**: 471–483. doi:10.1038/s41579-018-0040-1

Izaguirre, G., and W. D. Taylor. 2004. A guide to geosmin-and MIB-producing cyanobacteria in the United States. Water Sci. Technol. **49**: 19–24.

Jain, C., L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, and S. Aluru. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. **9**: 5114. doi:10.1038/s41467-018-07641-9

Jasser, I., and C. Callieri. 2017. Picocyanobacteria – the smallest cell-size cyanobacteria. Handb. Cyanobacterial Monit. Cyanotoxin Anal.

Jones, G. J., and W. Korth. 1995. In situ production of volatile odour compounds by river and reservoir phytoplankton populations in Australia. Water Sci. Technol. **31**: 145–151. doi:https://doi.org/10.1016/0273-1223(95)00469-4

Journey, C. A., J. M. Arrington, K. M. Beaulieu, J. L. Graham, and P. M. Bradley. 2011. Limnological Conditions and Occurrence of Taste-and-Odor Compounds in Lake William C. Bowen and Municipal Reservoir# 1, Spartanburg County, South Carolina, 2006-2009. IntechOpen.

Journey, C. A., K. M. Beaulieu, and P. M. Bradley. 2013. Environmental factors that influence cyanobacteria and geosmin occurrence in reservoirs, *In* Current Perspectives in contaminant hydrology and water resources sustainability. IntechOpen.

Jüttner, F., and S. B. Watson. 2007. Biochemical and Ecological Control of Geosmin and 2-Methylisoborneol in Source Waters. Appl. Environ. Microbiol. **73**: 4395 LP – 4406. doi:10.1128/AEM.02250-06

Kang, D. D., F. Li, E. Kirton, A. Thomas, R. Egan, H. An, and Z. Wang. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ **7**: e7359–e7359. doi:10.7717/peerj.7359

Kibuye, F. A., A. Zamyadi, and E. C. Wert. 2021. A critical review on operation and performance of source water control strategies for cyanobacterial blooms: Part I-chemical control methods. Harmful Algae **109**: 102099.

doi:https://doi.org/10.1016/j.hal.2021.102099
Kim, K., Y. Yoon, H. Cho, and S.-J. Hwang. 2020. Molecular Probes to Evaluate the Synthesis and Production Potential of an Odorous Compound (2-methylisoborneol) in Cyanobacteria. Int. J. Environ. Res. Public Health **17**: 1933. doi:10.3390/ijerph17061933
Kutovaya, O. A., and S. B. Watson. 2014. Development and application of a molecular assay to detect and monitor geosmin-producing cyanobacteria and actinomycetes in the Great Lakes. J. Great Lakes Res. **40**: 404–414. doi:https://doi.org/10.1016/j.jglr.2014.03.016
Lee, J. E., S.-J. Youn, M. Byeon, and S.-J. Yu. 2020. Occurrence of cyanobacteria, actinomycetes, and geosmin in drinking water reservoir in Korea: a case study from an algal bloom in 2012. Water Supply **20**: 1862–1870. doi:10.2166/ws.2020.102
Lenz, K. A., T. R. Miller, and H. Ma. 2019. Anabaenopeptins and cyanopeptolins induce systemic toxicity effects in a model organism the nematode Caenorhabditis elegans. Chemosphere **214**: 60–69. doi:https://doi.org/10.1016/j.chemosphere.2018.09.076
Li, D., C.-M. Liu, R. Luo, K. Sadakane, and T.-W. Lam. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics **31**: 1674–1676. doi:10.1093/bioinformatics/btv033
Li, Z., P. Hobson, W. An, M. D. Burch, J. House, and M. Yang. 2012. Earthy odor compounds production and loss in three cyanobacterial cultures. Water Res. **46**: 5165–5173. doi:https://doi.org/10.1016/j.watres.2012.06.008
Li, Z., J. Yu, M. Yang, J. Zhang, M. D. Burch, and W. Han. 2010. Cyanobacterial population and harmful metabolites dynamics during a bloom in Yanghe Reservoir, North China. Harmful Algae **9**: 481–488. doi:https://doi.org/10.1016/j.hal.2010.03.003
Ludwig, F., A. Medger, H. Börnick, M. Opitz, K. Lang, M. Göttfert, and I. Röske. 2007. Identification and expression analyses of putative sesquiterpene synthase genes in Phormidium sp. and prevalence of geoA-like genes in a drinking water reservoir. Appl. Environ. Microbiol. **73**: 6988–6993. doi:10.1128/AEM.01197-07
Lukassen, M. B., A. M. Saunders, P.-D. Sindilariu, and J. L. Nielsen. 2017. Quantification of novel geosmin-producing bacteria in aquaculture systems. Aquaculture **479**: 304–310. doi:https://doi.org/10.1016/j.aquaculture.2017.06.004
Lund, J. W. G., C. Kipling, and E. D. Le Cren. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia **11**: 143–170. doi:10.1007/bf00007865
McMurdie, P. J., and S. Holmes. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One **8**: e61217.
Mehrshad, M., M. M. Salcher, Y. Okazaki, S. Nakano, K. Šimek, A.-S. Andrei, and R. Ghai. 2018. Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. Microbiome **6**: 176. doi:10.1186/s40168-018-0563-8
Menzel, P., K. L. Ng, and A. Krogh. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. **7**: 11257. doi:10.1038/ncomms11257
Naes, H., and A. F. Post. 1988. Transient states of geosmin, pigments, carbohydrates and proteins in continuous cultures of Oscillatoria brevis induced by changes in nitrogen supply. Arch. Microbiol. **150**: 333–337. doi:10.1007/BF00408303

Naqib, A., S. Poggi, W. Wang, M. Hyde, K. Kunstman, and S. J. Green. 2018. Making and Sequencing Heavily Multiplexed, High-Throughput 16S Ribosomal RNA Gene Amplicon Libraries Using a Flexible, Two-Stage PCR Protocol BT - Gene Expression Analysis: Methods and Protocols, p. 149–169. *In* N. Raghavachari and N. Garcia-Reyero [eds.]. Springer New York.
NC DEQ (North Carolina Department of Environmental Quality). 2019. Classifications and Water Quality Standards Applicable to Surface Waters of North Carolina.
Neuenschwander, S. M., R. Ghai, J. Pernthaler, and M. M. Salcher. 2018. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. ISME J. **12**: 185–198. doi:10.1038/ismej.2017.156
Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. Microbiol. Mol. Biol. Rev. **75**: 14–49. doi:10.1128/MMBR.00028-10
NOAA. 2022. NOAA National Centers for Environmental information, Climate at a Glance: County Time Series.
Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, and R. B. O'Hara. 2013. Package vegan. R Packag ver **254**.
Otten, T. G., J. L. Graham, T. D. Harris, and T. W. Dreher. 2016. Elucidation of Taste- and Odor-Producing Bacteria and Toxigenic Cyanobacteria in a Midwestern Drinking Water Supply Reservoir by Shotgun Metagenomic Analysis. Appl. Environ. Microbiol. **82**: 5410–5420. doi:10.1128/AEM.01334-16
Overmann, J. 2008. Green Nonsulfur Bacteria. eLS. doi:https://doi.org/10.1002/9780470015902.a0000457
Paerl, H. W., W. S. Gardner, K. E. Havens, A. R. Joyner, M. J. McCarthy, S. E. Newell, B. Qin, and J. T. Scott. 2016. Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. Harmful Algae **54**: 213–222.
Paerl, H. W., K. E. Havens, H. Xu, and others. 2020a. Mitigating eutrophication and toxic cyanobacterial blooms in large lakes: The evolution of a dual nutrient (N and P) reduction paradigm. Hydrobiologia **847**: 4359–4375. doi:10.1007/s10750-019-04087-y
Paerl, H. W., and J. Huisman. 2008. Blooms like it hot. Science (80-.). **320**: 57–58.
Paerl, R. W., R. E. Venezia, J. J. Sanchez, and H. W. Paerl. 2020b. Picophytoplankton dynamics in a large temperate estuary and impacts of extreme storm events. Sci. Rep. **10**: 22026. doi:10.1038/s41598-020-79157-6
Plaas, H. E., and H. W. Paerl. 2021. Toxic Cyanobacteria: A Growing Threat to Water and Air Quality. Environ. Sci. Technol. **55**: 44–64. doi:10.1021/acs.est.0c06653
van der Ploeg, M., C. S. Tucker, and C. E. Boyd. 1992. Geosmin and 2-Methylisoborneol Production by Cyanobacteria in Fish Ponds in the Southeastern United States. Water Sci. Technol. **25**: 283–290.
R Core Team. 2014. R Core Team (2014). R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria. URL http//www.R-project.org/. R Foundation for Statistical Computing.
Roy-Lachapelle, A., S. V. Duy, G. Munoz, Q. T. Dinh, E. Bahl, D. F. Simon, and S. Sauvé. 2019. Analysis of multiclass cyanotoxins (microcystins, anabaenopeptins, cylindrospermopsin and anatoxins) in lake waters using on-line SPE liquid

chromatography high-resolution Orbitrap mass spectrometry. Anal. Methods **11**: 5289–5300.

Roy-Lachapelle, A., M. Solliec, S. Sauvé, and C. Gagnon. 2021. Evaluation of ELISA-based method for total anabaenopeptins determination and comparative analysis with on-line SPE-UHPLC-HRMS in freshwater cyanobacterial blooms. Talanta **223**: 121802. doi:https://doi.org/10.1016/j.talanta.2020.121802

Scott, J. T., M. J. McCarthy, and H. W. Paerl. 2019. Nitrogen transformations differentially affect nutrient-limited primary production in lakes of varying trophic state. Limnol. Oceanogr. Lett. **4**: 96–104. doi:https://doi.org/10.1002/lol2.10109

Seitzinger, S. P. 1988. Denitrification in freshwater and coastal marine ecosystems: Ecological and geochemical significance. Limnol. Oceanogr. **33**: 702–724. doi:https://doi.org/10.4319/lo.1988.33.4part2.0702

Shang, L., M. Feng, X. Xu, F. Liu, F. Ke, and W. Li. 2018. Co-Occurrence of Microcystins and Taste-and-Odor Compounds in Drinking Water Source and Their Removal in a Full-Scale Drinking Water Treatment Plant. Toxins (Basel). **10**: 26. doi:10.3390/toxins10010026

Sharpley, A., H. P. Jarvie, A. Buda, L. May, B. Spears, and P. Kleinman. 2013. Phosphorus Legacy: Overcoming the Effects of Past Management Practices to Mitigate Future Water Quality Impairment. J. Environ. Qual. **42**: 1308–1326. doi:https://doi.org/10.2134/jeq2013.03.0098

Simpson, G. L. 2015. ggvegan: “ggplot2” Plots for the “vegan” Package.

Stockner, J. G., and N. J. Antia. 1986. Algal picoplankton from marine and freshwater ecosystems: a multidisciplinary perspective. Can. J. Fish. Aquat. Sci. **43**: 2472–2503.

Su, M., J. Yu, J. Zhang, and others. 2015. MIB-producing cyanobacteria (Planktothrix sp.) in a drinking water reservoir: Distribution and odor producing potential. Water Res. **68**: 444–453. doi:https://doi.org/10.1016/j.watres.2014.09.038

Swan, B. K., B. Tupper, A. Sczyrba, and others. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc. Natl. Acad. Sci. doi:10.1073/pnas.1304246110

Utermöhl, H. 1958. Zur Vervollkommung der quantitative Phytoplankton Methodik. Mitteilungen der Int. Vereinigung für Theor. und Angew. Limnol. **9**: 1–38.

Wang, Z., G. Song, Y. Li, G. Yu, X. Hou, Z. Gan, and R. Li. 2019. The diversity, origin, and evolutionary analysis of geosmin synthase gene in cyanobacteria. Sci. Total Environ. **689**: 789–796. doi:https://doi.org/10.1016/j.scitotenv.2019.06.468

Wang, Z., G. Song, J. Shao, W. Tan, Y. Li, and R. Li. 2016. Establishment and field applications of real-time PCR methods for the quantification of potential MIB-producing cyanobacteria in aquatic systems. J. Appl. Phycol. **28**: 325–333. doi:10.1007/s10811-015-0529-1

Watson, S. B., P. Monis, P. Baker, and S. Giglio. 2016. Biochemistry and genetics of taste- and odor-producing cyanobacteria. Harmful Algae **54**: 112–127. doi:https://doi.org/10.1016/j.hal.2015.11.008

Welschmeyer, N. A. 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. Limnol. Oceanogr. **39**: 1985–1992. doi:10.4319/lo.1994.39.8.1985

Wetz, M. S., and H. W. Paerl. 2008. Estuarine phytoplankton responses to hurricanes

and tropical storms with different characteristics (trajectory, rainfall, winds). Estuaries and Coasts **31**: 419–429. doi:10.1007/s12237-008-9034-y
Wnorowski, A. U., and W. E. Scott. 1992. Incidence of Off-Flavors in South African Surface Waters. Water Sci. Technol. **25**: 225–232. doi:10.2166/wst.1992.0056
Wood, S. N. 2017. Generalized additive models: an introduction with R, CRC press.
Yamada, Y., D. E. Cane, and H. Ikeda. 2012. Chapter Seven - Diversity and Analysis of Bacterial Terpene Synthases, p. 123–162. *In* D.A.B.T.-M. in E. Hopwood [ed.], Natural Product Biosynthesis by Microorganisms and Plants, Part A. Academic Press.
Young, W. F., H. Horth, R. Crane, T. Ogden, and M. Arnott. 1996. Taste and odour threshold concentrations of potential potable water contaminants. Water Res. **30**: 331–340. doi:https://doi.org/10.1016/0043-1354(95)00173-5
Zaitlin, B., S. B. Watson, J. Ridal, T. Satchwill, and D. Parkinson. 2003. Actinomycetes in Lake Ontario: habitats and geosmin and MIB production. J. Am. Water Works Assoc. **95**: 113–118.

Appendix 1: Defined abbreviations and symbols

CDWM – City of Durham Water Management
T&O – taste and odor
2-MIB - 2-methylisoborneol
Cu - copper
Cd - cadmium
Fe - iron
N – nitrogen
P - phosphorous
TN – total nitrogen
TP – total phosphorous
PO4 - phosphate
NOx – oxidized inorganic nitrogen; nitrate + nitrite
NH4 - ammonium
SiO2 - silica
PCR – polymerase chain reaction
qPCR – quantitative polymerase chain reaction
GF/F – glass fiber filter

Gene sequences used for BLASTn searches of LM metagenome-derived contigs are below (see methods):

>DolichospermumcircinaleAWQC310F_NZ_KE384690.1:c40238-37979_GeoAcya_tocompounds_1
ACCATTTAAACTGCCAGCATTTTATATGCCTTGGCCAGCGCGGCTGAATCCAAACC
TGGAAGCGGCACGAGTGCATTCCAAAGCCTGGGCTTATGAAATGGGAATACTTGG
GTCAAAAGAGGAGTCCCAAGGTGAGCCTATATGGGATGAACGTAAATTTGATGCTC

ACGACTATGCTTTACTTTGTTCTTATACTCATCCAGACACGCCCTCAACAGAGTTGA
ACCTGGTAACGGATTGGTATGTATGGGTATTCTTTTTTGATGATCACTTTCTGGAAA
TCTATAAGCGCAGTCAGGATCTAATTGGGGCAAAGGAGTATCTTGACCGACTACCG
GCATTTATGCCAATTTATCCCCAAGATAACCTTCCCTTTCCCACGAACCCAGTAGA
GCGCGGTTTAGCTGATTTGTGGTCTCGTACTGCCTTTACTAAATCCGTAGAATGGC
GGCAAAGATTCTTTGAAAGTACCAAAAATCTTTTAGATGAGTCAATGTGGGAACTAG
CAAATATTAATCAAAATCGCATTGCTAACCCCATTGAATACATTGAAATGCGGCGGA
AAGTTGGGGGCGCACCTTGGTCAGCAGATTTAGTCGAACACGCGGCCTTTGTGGA
AGTTCCGGCTAAAATTGCGGCAACTAGACCCATGCGGGTTTTAAAAGACACATTTG
CTGATGGTGTACATCTCCGCAATGATCTATTCTCCTACCAAAGAGAGGTGGAAGAG
GAAGGTGAAAATTCTAACTGTGTGCTTGTAGTTGAGCGTTTCTTGAATGTGAGTAC
CCAAGAGGCCGCTAACCTCACTAACGAACTACTCAACTCCCGTTTATACCAATTTG
ACAACACTGCTGTCACGGAATTACCCTCTCTTTTTGAGGAGTACGGAGTAGATCCA
GTAGAGCGTGTGAATGTTCTCCTTTACATTAAAGGACTTCAGGATTGGCAATCTGG
TGGTCACGAGTGGCACATGAGGTCAAGCCGCTATATGAACAAGGAAGAGCCGGAT
AATTCAGGTACATCTGTGACTGTTGGCGGTCCGACTGGGCTAGGAACATCGGCCG
CGCGGCTGGAATCCTTATCCACTACTTTGGGTGTGGGAAGGTTTAAAAGTTTTACT
CATGTTCCCTATCAAACTGTAGGACCAGTAAAACTGCCTAAGTTTTATATGCCCTTC
TCTACCACCTTAAATCCTAATTTGGATGCTGCACGAAATCATTCCAAAGAATGGGC
GCGTCGGATGGGGATGTTAGAATCATTACCGGGGATTCCTGATGCTTTTATCTGGA
ATGACCATAAGTTTGATGTTGCTGACGTGGCTTTATGCGGTGCGTGGATACATCCA
AATGGTTCTGAGCATGAGCTAAATTTAACAGCTTGTTGGCTGGTTTGGGGAACTTA
TGCGGATGATTATTTCCCAGCAATCTATGGTAATAACCGCGACTTGGCAGGTGCAA
AAGTCTTTAATGCTCGACTGTCAGCATTTATGCCTCTAGATAATTCTACACCCCCTG
TAGCAACTAATCCAGTAGAAAAGGGGTTGGCGGATATTTGGTCTCGCACCGCAGG
ACCAATGTCCTCCACAGCACGTACCGAATTTCGCCGTGCTATTCAGGATATGACTG
ATAGTTGGGTGTGGGAACTGGCTAACCAGACGCAAAATCGGATTCCAGATCCGAT
TGACTATATTGAAATGCGTCGTAAGACATTTGGATCTGATTTAACCATGAGTCTGTC
TAGACTATCCCATGGCGGCGAAATACCAATGCAGATTTATTATAGCCGACCCATGC
GATCGCTGGAAAATGCTGCCGCCGATTTTGCCTGTTTCACTAATGACATTTTTTCTT
ATCAGAAAGAAATCGAATTTGAGGGCGAAATTCATAACTGCGTGTTAGTGGTGCAG
AATTTCCTGAACTGTGATCTTCTCAAAGCTGTTGAGATTGTTAATAATCTGATGACA
GCGCGGGCGCAACAGTTTCAACACATTGTTGAGACTGAACTTCCAGCCCTTTTCGA
CGATTTCAACCTGGATAAAAACACCCGCGAGAAACTACTCAAATACATTGAGAAATT
AGAACAGTGGATGTGTGGTGTACTTAAGTGGCATACAAAGGTAGACCGATACAAAG
AATTTGAATTGCGTAATAGTGCTTCCCCCATAGTGCGACTACTCAACGGTCCGACA
GGGTTTGGCACTTCGGCTGCACATATCAGATCCTTGGTTGGTGCAACTAATTCCGT
GGGTAAAAATCCGTAA

>Streptomycesbottropensis_ATCC_25435_KB405089.1:392131-394338_GeoAbac_tocompounds_2
ATGACGCAGCCGTTCGTACTTCCGCACTTCTACATGCCGTATCCCGCGCGGCTGA
ACCCCCACCTCGAAGAGGCCCGTGCTCATTCCTCCCGGTGGGCGCGGGAGATGG
GCATGCTGGAGGGCTCCGGTGTGTGGGACCAGGCGGACCTGGACGCGCACGACT
ACGGCCTGCTGTGCGCGTACACCCATCCCGACTGCGACGCCCCGGCGCTCTCGC
TCATCACCGACTGGTACGTGTGGGTCTTCTTCTTCGACGACCACTTCCTGGAACTC
TTCAAGCGTTCCCAGGACCGTGCGGGCGGCAAGGCCCACCTCGACCGGCTCCCC
CTCTTCATGCCGCTGGACCTCTCCACCCCCGTGCCCGAGCCGCGGAACCCGGTG
GAGGCGGGCCTCGCCGACCTCTGGGCCCGTACGGTGCCGGCGATGTCGATGGAC
TGGCGCCGCCGCTTCGCCGTGGCGACCGAACACCTGCTCAACGAGTCGATGTGG
GAGCTGTCCAACATCAACGAGGGCCGGATCGCCAACCCGGTCGAGTACATCGAG
ATGCGCCGCAAGGTCGGCGGCGCGCCCTGGTCGGCCGGGCTGGTCGAGTACGC
GACCTGCGAAGTCCCGGCCTGCGTCGCGGAGTCCAGGCCCCTGAGGGTCCTGAT
GGAGACCTTCTCCGACGCCGTCCATCTGCGCAACGACCTGTTCTCGTACCAGCGT
GAGGTCGAGGAGGAGGGCGAGAACAGCAACGGCGTGCTCGTCCTGGAGACCTTC
TTCGGCTGCGGCACCCAGCAGGCCGCCGAGACGGTCAACGACATCCTCACCTCC
CGGCTGCACCAGTTCGAGAACACCGCGCTCACCGAAGTCCCCGGAATCGCCCTG
GAGAAGGGCCTCACCCCGGCCGAGACCGCCGCCGTCGCGGCGTACACGAAGGG
CCTTCAGGACTGGCAGTCCGGTGGTCACGAGTGGCACATGCGCTCCAGCCGCTA
CATGAACGAGGGGGCCGCCGCCTCGCGCGGCCCGCTCGACCTCGGCGGCACGG
TCCTGTCCGGCCCCGCCCTGGTCGCCCGGGCCGGGTTCGGCACGTCCGCCGCC
GACGTGGGGGCACTGCTCGCGACCGCCGCCGCGGAACGGCTGCGGGCGCACAC
GCATCAGCCGTACCAGAAGGTCGGACCGTCGCGACTGCCGGACTTCTTCATGCCC
TTCCAGGTGGAGCTCTCCCCGCACCTGGACGGCGCCCGCCCCCGCCTCACCGCC
TGGGCGCACGCGACGGGCATCCTCGGCGAGGGCGTCTGGGACGAGGAGCGGCT
CGCCGCCGCCGACCTGCCCCTCTGCTCGGCCGGCCTCGACCCGGACGCCACCCC
CGAGCAACTCGACCTCAGCTCCCAGTGGCTCGCCTGGGGCACCTACGGCGACGA
CTACTACCCCCTGGTCTTCGGCCACCGCCGCGACCTCGCCGCCGCCCGGCTGAC
CACGGCCCGGCTCTCCGACTGCATGCCCGTCGAGGGCGAGCCGGTCCCCCTCCC
GGTGAACGGCATGGAACGGGGCCTGATCGACCTCTGGGCGCGGACGACGGCGG
GGATGACCCCGGACCAGCGGCGCGGTCTGAAGGCATCCGTCGACAAGATGACGG
AGAGCTGGGTGTGGGAGGTGTTCAACCAGATCCAGCACCGCGTCCCCGACCCGG
TCGACTATCTGGAGATGCGCCGCGCCACCTTCGGCTCCGACCTCACCCTCAGCAT
GTGCCGGATGGGCCACGGCCCACAGATCCCGGAGGCGGTCTATCGCAGCGGCC
CCGTCCGCTCACTGGAGAACGCGGCGATCGACTACGCCTGCCTCGTCAACGACG
TCTTCTCCTACCAGAAGGAGATCGAGTACGAGGGCGAGGTCCACAACGCGATCCT
CGTCGTGCAGAACTTCTTCGGCTGCGACTACCCGACCGCGCTCGGTGTGATCCAC
GACCTGATGACCCAGCGCATGCAGCAGTTCCAACACGTCGTGGCCCATGAACTCC
CCGTCGTCTACGACGACTTCGGGCTGACAGAAGAGGCGCGCGACATCATGGGCG

GCTATGTGACCGACCTGCAGAACTGGCTGTCCGGCATCCTCAACTGGCACCGCTC
GGTGGACCGTTACAAGCCGGAGTTCCTGGCCCGCCGGGCCCACGGCTTCGTCCC
GGACCGTCCCCCGGTCCTGTCCTTCGCGCCGGCGCTCACCGGCTGA

>Pseudanabaenacinerea_NIES-4062_LC507460.1_MicCya_tocompounds_3
ATGAAAGATACCAACTTGGATAATACGAGTACCACATTTTTATTTGGTCCGACTGGT
CTTGGTACATCCGCCGCTCGCTTTGTGAGTGAGATAGCCAAGGCTGCTCGTACCG
ATAACCTAGATTCAGGACTACTCCAGACAGATCTCGCCTCGGTGGGGGATTGTGAT
CCTGCCTATGCAGAGAGAGATTGGGGCGATGGAACTGCCTCACCACTCTACTGCC
CCATTACCGAGCGATTCAACGAGCCACTAGCGGACGAGGTCGATGATCGGCTGGC
GGTCTGGGCATTGGAATGCGGGTTTGATGAGGACGAGGCGCAGAAAATTCGCAAG
GTGCGGTTTGGACGGCTGGTGATGCTAGCTCACCCAGACTGCGATGACCCATCTC
GGCTGTTGATTGGGGCCAAGTTGAACATGGCTTGGTGGGCTGCTGATGACTACTA
TGCTGACGACAGCGAGCTTGGGGCAGATCCAAAGCTGCTCCCCCCTCGTCTTCTG
TTGGCGATGACCGCTATGGACCCGCCCCCGCCAGCAGGGGAATTTACGCCCCCG
CTAGAAGCCGCGATCGCCGAAGAGCGGGTATTAGTAGCCCTCGGTAGGGGCATT
GACTATTTGAGCCAGTATGCCACACCAGAGCAAGTCCAGCGTACCTGTTACGCCA
CCTTCTCCATGTTCGTCTCTTGGGGGGCTTACGCGGCGTGGCGCTATACCGGCGA
TTACCCGCCAGCTTGGAAGTATCTGGCTGCGCGACAGCACGACAGCTTCTACACC
TCCATGACACTAATCGACCCCATCGGAGGCTACGTCCTCCCACCAGATCTTTTCTT
CGATCCACGCGTCCGTCACGCAGCGTTCCTAGCTGGAACGGCAGTCGTTCTGGTC
AACGATCTCCTTTCG

>StreptomycescoelicolorA3-2_NC_003888.3:8537030-
8538352_MicBac_tocompounds_4
ATGCCCGACTCCGGGACCCTCGGAACACCACCCCCCGAGCAGGGGCCGACGCC
GCCCACAACCCTGCCGGACGTGCCCGCCCCGGTGATCCCGAGCGCTTCGGTCAC
GTCGGCCGCGTCCGACTTCCTGGCCGCGCTGCATCCGCCGGTCACGGTTCCCGA
CCCGGCACCACCCCCGCCGCCCGCCCCCGCCGCAGGGAACCCGCCCGACACGG
TCACGGGCGACTCCGTCCTCCAGCGGATCCTGCGTGGCCGACCGGGCCGGGCA
CGACCTCGCTCGCCCCGGCCGTCAGATACGGGCGGCAACCGGGCCCGGAGGCT
CCGGCATCCGCTCCGCCCGCCGCCGGACGAGCCGTCCCCGGCCTCTACCACCAC
CCCGTCCCGGAACCCGACCCCGTGCGCGTCGAGGAGGTGAGCCGCCGTATCAAG
CGCTGGGCCGAGGACGAGGTCCAGCTGTATCCCGAGGAGTGGGAGGGGCAGTTC
GACGGCTTCTCGGTGGGCCGCTACATGGTCGGCTGCCATCCCGACGCACCGACC
GTCGACCACCTGATGCTCGCGACCCGGCTGATGGTCGCGGAGAACGCGGTGGAC
GACTGCTACTGCGAGGACCACGGCGGCTCACCCGTCGGTCTGGGCGGGCGCCTC
CTGCTCGCGCACACCGCGATCGACCACTTCCACTCCACGGCCGAATACACGCCGA
CGTGGCAAGCGTCCCTCGCAGCGGACGCCCCGCGCCGGGCGTACGACAGCGCC
ATGGGCTACTTCGTGCGGGCGGCGACTCCGTCCCAGTCCGACCGCTACCGGCAC

GACATGGCCCGGCTGCACCTGGGCTACCTCGCCGAGGGCGCCTGGGCACAGACC
GGACACGTCCCGGAGGTGTGGGAGTACCTGGCGATGCGCCAGTTCAACAACTTC
CGCCCCTGCCCCACGATCACCGACACCGTCGGCGGCTACGAACTGCCCGCGGAC
CTGCACGCCCGCCCGGACATGCAACGGGTCATCGCCCTCGCCGGCAACGCGACG
ACCATCGTCAACGACCTCTACTCCTACACCAAGGAACTCAACAGCCCCGGTCGCC
ACCTGAACCTTCCGGTGGTGATCGCCGAGCGCGAGCAGCTCTGCGAACGCGACG
CCTACCTCAAGGCCGTCGAGGTCCACAACGAGCTCCAGCACTCCTTCGAGGCCGC
CGCGGCCGACCTCGCGGAGGCCTGCCCCCTGCCCCCCGTGCTGCGCTTCCTCAG
GGGCGTCGCCGCCTGGGTCGACGGCAACCACGACTGGCACCGCACCAACACCTA
CCGCTACAGCCTGCCCGACTTCTGGTAG

Appendix 2: Items resulting from the project:

- Two prospective chapters in the thesis of PhD student Hwa Huang are anticipated to result from this project.
- An electronic copy of the final report was provided to Jimmy Gable (CDWM) for archiving and future reference.
- Information concerning taste and odor compounds in local waters was transferred to the public through three Mobile Aquatic Microbial Laboratory (MAML) events at Lake Johnson Park in Raleigh, NC. At least 10 people were directly engaged through the outreach effort and participated in teaching modules. 7 have completed pre/post assessments. Prior to the demonstration, the average participant assessment score (correct answers) was 1. The average participant score in post assessment increased to 3.14 after engagement in the teaching module. Other interactions occurred with the public but could not be documented through only pre/post assessments – as they were not willing to complete the assessment; this includes observing live plankton from Lake Johnson water as well as smelling a geosmin standard purchased from a perfume provider.
- North Carolina Central University undergraduate Kemoni Hughes gained experience with flow cytometry, data analysis, and general lab interactivity as part of this project (summer internship 2021).
- Two research presentations covering results of the project will occur at the WRRI annual meeting, March 23, 2022.
- Results from this project were presented at the 2021 NC American Society for Microbiology's virtual meeting by PhD student Hwa Huang, titled "Identifying microbial sources of taste and odor compounds in the NC piedmont drinking-water reservoir Lake Michie".
- Two extensive DNA sequence datasets were generated as part of this project: 1) geoA amplicon sequences, 2) shotgun metagenomic sequence libraries from plankton >8µm diameter and 8-0.8µm in diameter. Both datasets will be deposited to the national public database, NCBI Genbank, following acceptance of a publication using these data or by Dec. 31st, 2024.