Title: *Transcriptionally active nitrogen fixation and biosynthesis of diverse secondary metabolites by* Dolichospermum *and* Aphanizomenom*-like Cyanobacteria in western Lake Erie* Microcystis *blooms*

Authors: Colleen E. Yancey[1], Olivia Mathiesen[1,+], Gregory J. Dick[1,2*]

(1) Earth and Environmental Science Department, University of Michigan, Ann Arbor, USA, 48109

(2) Cooperative Institute for Great Lakes Research (CIGLR), University of Michigan, 4840 South State Road, Ann Arbor, MI 48108 USA

(+) Present Address for OM: EA Engineering, Science, and Technology Inc., 555 University Avenue, Suite 110, Sacramento, CA, 95825 USA

***Corresponding Author:** Dr. Gregory J. Dick,
Address: 1100 North University Ave, Rm. 2014, Ann Arbor, MI, USA,
email address: gdick@umich.edu

Highlights

- Through metagenomic approaches, we generated two near-complete metagenome assembled genomes from two distinct species that are dispersed across the ADA clade of cyanobacteria.
- These ADA cyanobacteria are inferred to produce a variety of known and novel secondary metabolites, some of which may encode known toxins or taste and odor compounds, at different stations and phases of bloom as distinguished by differential transcript abundance
- Metagenomic and metatranscriptomic analyses suggest that ADA are the dominant nitrogen fixers at the stations and times sampled during the 2014 bloom.
- This works highlights the diversity of cyanobacteria in western Lake Erie blooms despite the continued dominance by *Microcystis,* and that these less abundant cyanobacteria may produce unmonitored secondary metabolites and influence N availability in blooms through N-fixation.

Abstract

Cyanobacterial harmful algal blooms (cyanoHABs) in the western basin of Lake Erie are dominated by microcystin producing *Microcystis* spp., but other cyanobacterial taxa that coexist in these communities may play important roles in production of toxins and shaping bloom dynamics and community function. In this study, we used metagenomic and metatranscriptomic data from the 2014 western Lake Erie cyanoHAB to explore the genetic diversity and biosynthetic potential of cyanobacteria belonging to the *Anabaena, Dolichospermum, Aphanizomenon* (ADA) clade. We reconstructed two near-complete metagenome-assembled

genomes from two distinct ADA clade species, each containing biosynthetic gene clusters that encode novel and known secondary metabolites, including those with toxic and/or known taste and odor properties, that were transcriptionally active. However, neither ADA metagenome-assembled genome contained genes encoding guanitoxins, anatoxins, or saxitoxins, which are known to be produced by ADA. The ADA cyanobacteria accounted for most of the metagenomic and metatranscriptomic reads from nitrogen fixation genes, suggesting they were the dominant N-fixers at the times and stations sampled. Despite their relatively low abundance, our results highlight the possibility that ADA taxa could influence the water quality and ecology of *Microcystis* blooms, although the extent of these impacts remains to be quantified.

Keywords: ADA clade cyanobacteria, metagenomics, metatranscriptomics, cyanoHABs, secondary metabolites, N-fixation

1. Introduction

For the past 20 years, western Lake Erie has endured annual cyanobacterial harmful algal blooms (cyanoHABs) (Bridgeman et al., 2013; Stumpf et al., 2012) dominated by *Microcystis* spp. (Berry et al., 2017; Bridgeman et al., 2013; Rinta-Kanto et al., 2005, 2009). Excessive N-loading may have enabled the rise of *Microcystis;* higher concentrations of fixed nitrogen support the ecological success of non-diazotrophic cyanobacteria (Davis et al., 2010; Gobler et al., 2016). These contemporary blooms contrast those observed during the 1960s and 1970s, which were more diverse and dominated by diazotrophs (Davis, 1964; Makarewicz, 1993; Munawar and Munawar, 2011; Paerl et al., 2018). CyanoHABs in this system were largely reduced by the 1980s due to the Great Lakes Water Quality Agreement that implemented reductions on phosphorus (P) inputs (Matisoff and Ciborowski, 2005), which was initially identified as the major driver of cyanoHABs (Schindler, 1974), while nitrogen (N) inputs were

left unmanaged. Eventually this led to N and P co-limitation (Paerl et al., 2016; Paerl and Scott, 2010) and likely shifted species composition and contributed in part to the eventual reemergence of cyanoHABs in the western basin (Watson et al., 2016a).

While *Microcystis* dominates western Lake Erie blooms and is mainly responsible for production of the hepatotoxin microcystin, bloom communities contain other cyanobacteria including *Cyanobium* spp., *Pseudanabaena* spp., *Planktothrix* spp. (near river mouths and bays), and members of the *Anabaena, Dolichospermum, Aphanizomenon* (ADA) clade of cyanobacteria (Berry et al., 2017; Bullerjahn et al., 2016; Chaffin et al., 2013; Gobler et al., 2016). However, little is known about their role in these communities, their impact on bloom dynamics, or their potential contribution to toxin production in western Lake Erie.

The ADA clade of cyanobacteria are of particular concern as they are prolific producers of diverse and potent cyanotoxins (Dreher et al., 2021; Driscoll et al., 2017; Österholm et al., 2020). These toxins include microcystin (Fewer et al., 2008; Tonk et al., 2009; Vaitomaa et al., 2003), hepatotoxic cylindrospermopsin (Dreher et al., 2021; Stüken and Jakobsen, 2010), neurotoxic anatoxin (Carmichael et al., 1975) and guanitoxin (Lima et al., 2022), and paralytic shellfish toxin (PST) saxitoxin, which is classified as a bioweapon via the Chemical Weapons Convention (Al-Tebrineh et al., 2010; Sierra and Martínez-Álvarez, 2020). Whereas microcystin production has largely been attributed to *Microcystis* in western Lake Erie (Berry et al., 2017; Rinta-Kanto et al., 2005, 2009; Steffen et al., 2017), and anatoxin and saxitoxin have yet to be widely detected in this system (McKindles et al., 2020), the presence of ADA species in these blooms merits further investigation given that toxin production is variable and sporadic across this clade (Dreher et al., 2021; Österholm et al., 2020). Indeed, ADA have dominated Lake Erie cyanoHABs in the past (Davis, 1964), saxitoxin genes have been detected in early summer

Dolichospermum blooms in the central basin of Lake Erie (Chaffin et al., 2019), and mixing and spreading of blooms from the western to central basin have been observed (Michalak et al., 2013). A better understanding of ADA distribution, ecophysiology, and biosynthetic potential will improve our understanding of how changing nutrient regimes under various climate and nutrient management scenarios are likely to affect the abundance, distribution, and toxicity risk of these organisms.

Within western Lake Erie, ADA cyanobacteria may also play a role in cyanoHAB community dynamics via supply of nitrogen (N) and/or competition with *Microcystis*. ADA are distinct from *Microcystis* as they can fix nitrogen through differentiated heterocyst cells (Kumar et al., 2010; Wolk et al., 1994), thereby ameliorating N limitation during deplete conditions (Wood et al., 2010). However, ADA cyanobacteria also share similarities with *Microcystis* as they have large biosynthetic potential and produce diverse secondary metabolites (Dreher et al., 2021; Driscoll et al., 2017; Kehr et al., 2011; Österholm et al., 2020), may regulate their buoyancy through gas vesicles (Li et al., 2016; Walsby et al., 2007), and have genomes that are rich with mobile elements (Driscoll et al., 2018; Wang et al., 2012).

To explore the genetic diversity and toxin producing potential of ADA cyanobacteria more deeply in natural cyanoHAB communities, we analyzed metagenomic and metatranscriptomic datasets from the 2014 western Lake Erie cyanoHAB. *De novo* assembly was used to recover two high quality metagenome assembled genomes (MAGs) belonging to the ADA clade. From these MAGs we assessed the genetic diversity of ADA cyanobacteria in western Lake Erie, identified biosynthetic gene clusters which encode secondary metabolites, and tracked the relative expression of biosynthesis and nutrient uptake and metabolism genes.

2. Materials and Methods

116 *2.1 Study Site and Sample Collection*

117 Three core stations were sampled and are part of weekly cyanoHAB monitoring by the
118 NOAA Great Lakes Environmental Research Laboratory (GLERL) (Cooperative Institute for
119 Great Lakes Research, 2019) within western Lake Erie. Samples were collected weekly from
120 mid-June to late October in 2014 at core stations WE2, WE4, and WE12. WE2 is close to the
121 inlet for the Maumee River inlet (41° 45.743'N, 83° 19.874' W), WE4 is considered an offshore
122 site closer to the center of the basin (41° 49.595'N, 83° 11.698'W), and WE12 is near the Toledo
123 drinking water crib (41° 42.535'N, 83° 14.989'W).

124 20L water samples were collected via integrated depth water sampling, in which the
125 entire water column, from surface to 1 meter above the bottom, was continuously sampled by
126 pump-cast. Physiochemical measurements such as pH, water temperature, and specific
127 conductivity were measured during the research cruise. Biomass was collected by filtering 2L of
128 integrated depth water though a 100 µm polycarbonate mesh filter. The biomass on the filter was
129 then collected and filtered through a 0.22 µm filter and preserved in 1 mL of RNALater™
130 (Invitrogen™, Ambion™) and placed on ice. In addition to collecting the 100 µm fraction for
131 analysis, integrated whole water samples were also collected and filtered directly through 0.22
132 µm filters for select paired metagenomic analyses of whole water samples. These samples were
133 used to investigate free-living bacteria not associated with *Microcystis* particulates/bloom
134 biomass.

135 *2.2 DNA Extraction and Sequencing*

136 Extraction and sequencing have been described previously in greater detail (Yancey et
137 al., 2022a). Qiagen DNeasy Blood and Tissue Kits were used to extract DNA while The Qiagen

RNEasy kit was used to complete RNA extraction and cDNA library preparation (Qiagen, Hilden, Germany). Shotgun DNA and RNA sequencing was completed at the University of Michigan Sequencing Core using the Illumina® HiSeq™ platform (2000 PE 100, Illumina, Inc., San Diego, CA, USA). Raw read datasets are publicly available under the NCBI BioProject PRJNA464361. SRA BioSample numbers include SAMN09102072 to SAMN09102087.

2.3 Bioinformatic Analyses

De novo assemblies were conducted on single samples to recover metagenome assembled genomes (MAGs). These methods are extensively detailed in a previous study (Yancey et al., 2022b). Briefly, assemblies were completed using Megahit (Li et al., 2015), with differential coverage read mapping achieved through bowtie2 (Langmead and Salzberg, 2012). Multiple binning software were implemented to recover the highest quality MAGs and included Concoct (Alneberg et al., 2013), Metabat (Kang et al., 2015), Tetra-ESOM (Ultsch and Mörchen, 2009) and VizBin (Laczny et al., 2015). MAGs generated from all four binning software were run through DASTool (Sieber et al., 2018) to assess and choose the highest qualities MAGs for further analysis. For further MAG refinement, bins were manually assessed and curated using Anvi'o v.5 (Eren et al., 2015) and taxonomic confirmation, completion, strain heterogeneity, and contamination was further assessed with CheckM (Parks et al., 2015). From this effort, several cyanobacteria MAGs were generated including 9 *Microcystis* MAGs (Yancey et al., 2022b), 2 ADA MAGs, and 6 *Pseudanabaena* MAGs (data not shown).

To understand the phylogenetic context of recovered ADA MAGs, GToTree was used to generate a phylogenomic tree with previously described ADA genomes (Lee, 2019). Briefly, a maximum likelihood tree was generated using 251 marker genes and included ADA genomes from previous studies (Dreher et al., 2021; Österholm et al., 2020; Sheik et al., 2022) with an

outgroup of Cyanobium *gracile* PCC 6307 (NCBI: PRJNA158695). Additionally, pyani v.0.2.10 (Pritchard et al., 2015) was used to calculate average nucleotide identity between the recovered ADA MAGs.

Relative abundance of ADA cyanobacteria was quantified as follows. Metagenomic reads were mapped to several cyanobacteria MAGs generated from the 2014 dataset. These MAGs included the two identified ADA MAGs, and a *Microcystis* MAG identified and examined in a previous study (C. Yancey et al., 2022), that were generated via *de novo* assembly from these samples. Reads mapped via Basic Alignment Search Tool (BLAST) (Madden, 2013) v.2.11.0 were kept for quantification if they satisfied a 95% identity and 80% alignment length cut off which was previously shown to be sensitive and sufficient for metagenomic read mapping (Yancey et al., 2022a). Reads per kilobase per million reads (RPKM) was calculated for each MAG to determine relative abundance using the equation below (Dick, 2018):

$$Relative\quad a\ i\ m\,Abu\ \ da\ \ ce = RPKM\,pe\quad e\ \ ome$$

$$= \frac{total\ \#\ of\ \ \ ead\ \ \ mapped\,pe\ \ \ MAG}{(le\quad th\ of\ the\ MAG) * (total\ \ \ umbe\ \ \ of\ \ \ ead\ \ \ pe\quad ample)} * 10^6$$

The software antiSMASH v.6.0 (Blin et al., 2021) was used to mine and annotate the ADA MAGs for biosynthetic gene clusters which encode secondary metabolites. Default parameters were run under the "relaxed" annotation settings. Identified BGCs were further deeply annotated on a by gene basis using blastP (Madden, 2013), and generated output from antiSMASH.

2.4 Relative Transcript Abundance

Relative abundance of transcripts was calculated as follows. Metatranscriptomic reads were mapped onto a database that contained all BGCs identified from the ADA MAGs as well as

183 BGCs identified from *Microcystis* MAGs from a previous study (C. Yancey et al., 2022). This
184 was done to ensure competitive mapping, especially for anabaenopeptin genes present in both the
185 ADA and *Microcystis* genomes. Reads were mapped using BLAST and were kept if they had at
186 least 95% identity and 80% alignment to the query which has previously shown to be sensitive
187 and sufficient for relative read mapping (Yancey et al., 2022a). Singular top hits that met these
188 criteria were then quantified. Similarly, a set of phosphorus and nitrogen genes were also
189 subjected to metatranscriptomic read mapping under the same parameters listed above to
190 determine nutrient uptake and metabolism throughout the bloom. The list of genes used for this
191 analysis can be found of Table S1. Metatranscriptomic reads were also mapped to entire ADA
192 MAGs using the same methods listed above. This was used to calculate the relative transcript
193 abundance for each BGC and nutrient gene using the equation below (Dick, 2018):

194 $$elative\ t\ \ a\ \ \ c\ \ ipt\ abu\ \ da\ \ ce = \frac{(\ \ ead\ \ \ mapped\ pe\ \ \ \ e\ \ e\ \div le\ \ \ th\ of\ \ \ e\ \ e\ (bp)}{(\ \ ead\ \ \ mapped\ pe\ \ \ MAG\ \div le\ \ \ th\ of\ MAG\ (bp)}$$

195 We completed an additional analysis to evaluate the sensitivity and specificity of selected
196 cut-offs for determining relative metatranscriptomic read abundance and ensure that there was no
197 ambiguous mapping of reads to genomes of different taxa. Metatranscriptomic reads mapped
198 based on the criteria above were extracted and aligned by BLAST to each ADA MAG, again
199 using a threshold of 95% identity over 80% of the alignment. Positive hits were then assessed to
200 determine if reads that mapped to BGCs from one ADA MAG mapped to contigs from the other
201 ADA MAG. No sequence reads mapped to both ADA MAGs, confirming that there was no
202 ambiguous mapping and indicating that the cut-off parameters are sensitive and sufficient for this
203 analysis.

204 *2.5 Nitrogen Fixation Gene and Transcript Abundance*

To assess the potential contribution of ADA cyanobacteria to N-fixation in western Lake Erie, we queried both 100 µm filter fractions and whole water samples for N-fixation genes, as both cyanobacteria and heterotrophic bacteria have been attributed with the ability to fix N in aquatic systems (Davis et al., 2015; Halm et al., 2011; Raymond et al., 2004). Assemblies and metagenomic QC reads from the 100 µm fraction were searched for *nif* operons while whole water QC reads were searched for *nifH, nifD,* and *nifK* using BLAST. Multiple *nif* genes were used to determine the presence of N-fixation pathways as pseudo-*nifH* genes, found in genomes that lack other *nif* genes, and therefore the ability to fix N, overpredict potential N-fixation when used as a single marker (Mise et al., 2021). Gene databases for each *nif* gene were selected from the FunGene repository (Fish et al., 2013) for each respective gene (accessed November 2022) to capture a wide variety of *nif* homologs. Contigs or reads with hits to *nif* genes were then aligned to the non-redundant nucleotide database from NCBI to determine taxonomic identity.

2.6 Figure Generation

Figures were generated using R and RStudio (Allaire, 2015) with the packages ggplot2 (Wickham, 2011), and ggpubr (Kassambara, 2020). Adobe Illustrator was used to render multiple paneled figures ("Adobe Illustrator," 2019).

3. Results

3.1 Overview of ADA cyanobacteria MAGs

From fifteen samples analyzed, d*e novo* metagenomic assembly produced two high quality ADA clade MAGs from two distinct samples (Fig 1A): station WE12 on August 4th 2014, which was near the Toledo drinking water intake during the "do not drink" advisory and first observed peak of cyanobacterial biomass and peak of microcystins concentration (Berry et

al., 2017); and WE4, an offshore station, which had low cyanobacterial biomass and concentration of microcystins on September 29th 2014. These MAGs have high completion (>97%) and low contamination (<1%) as well as large N50s (>25,000 bp) (Fig 1B).

LE14-WE12 and LE14-WE4 had low pairwise average nucleotide identity (ANI) (88.8%), and phylogenomic analysis showed that they clustered into distinct ADA subclades (Fig. 2), indicating that they are different species. LE14-WE12 clusters within a clade of nearly exclusively *Aphanizomenon* and is most closely related to *Aphanizomenon flos aquae* DEX188 (bootstrap =1, NCBI Genbank: GCA_017346855.1), which was isolated from the Dexter Reservoir in Oregon, USA (Dreher et al., 2021). LE14-WE4 belongs to a clade consisting mainly of *Dolichospermum* and is most closely related to a strain isolated from the western arm of Lake Superior, SB001 ((Sheik et al., 2022), NCBI BioSample: SAMN16655444). LE14-WE4 and SB001 fall within a branch that contains *Anabaena* sp. strains LE011-02 and AL09 (Driscoll et al., 2018) isolated from Lake Erie and Lake Ontario respectively (Fig. 2), thus representing a subclade from the Great Lakes.

3.2 Temporal and Spatial Variation in Abundance of MAGs from ADA and Cyanobacteria

Mapping of metagenomic reads to MAGs provided estimates of ADA relative abundance across stations and sampling dates. LE14-WE12 reached maximum abundance when *Microcystis* was most abundant at nearshore stations August 4th and September 29th (Fig. 3A), and the relative abundances of LE14-WE12 and *Microcystis* were positively correlated (R=0.859, p=4.06e-05, Table S2). LE14-WE4 was less abundant throughout but reached greatest relative abundance at offshore station WE4 on 29-Sep, when *Microcystis* abundance was low. There were no correlations between the relative abundance of LE14-WE4 and *Microcystis* (Table S2).

Microcystis spp. consistently dominated throughout the bloom and was 2-3 orders of magnitude more abundant than the ADA taxa (Fig 3A).

3.3 Biosynthetic Gene Clusters encoding secondary metabolites

Genes encoding the biosynthesis of microcystin (*mcy*), saxitoxin (*sxt*), anatoxin (*ana*), and guanitoxin (*gnt*) were not detected in either MAG. However, both MAGs contained over 10 biosynthetic gene clusters (BGCs) with nonribosomal peptide synthases (NRPSs) and terpenes being the most common (Fig 1B). BGC mining via antiSMASH (Blin et al., 2021) also revealed the presence of 2 hgIE-KS clusters, which are believed to be important for heterocyst glycolipid formation (Fig 1B). LE14-WE4 contained more and a greater diversity of BGCs, including those encoding a cyanobactin and a linear azol(in)e-containing peptide (LAP) cluster (Fig 1B). Two clusters identified from LE14-WE4 were 100% similar (contained all genes to closest known BGC from the Minimum Information about a Biosynthetic Gene Cluster (MiBIG) database with significant BLAST hits) to previously described cyanopeptolin (Österholm et al., 2020; Rounge et al., 2007) and geosmin (Watson et al., 2016b) encoding clusters (Fig. 4A, 4B). Cyanopeptolins inhibit several enzymes including eukaryotic proteases such as chymotrypsin (Bister et al., 2004; Gademann et al., 2010), while geosmin is a taste and odor compound (Gerber and Lechevalier, 1965; Izaguirre and Taylor, 2004). The metagenome of LE14-WE12 contained NRPS clusters with low similarity hits to trypsin inhibitor psuedospumigin A/B/C/D/E/F ((Jokela et al., 2017), 66%), and protease inhibitor aeruginoside ((Ishida et al., 2007), 29%), suggesting they encode uncharacterized compounds that may be related to aeruginosin type compounds (Fig 4C, 4D). However, due to the low similarity score of these clusters to known clusters, we are not able to confirm the identify of their products with the available data. Other BGCs with percent similarity hits to sequences in the MiBIG database are described in Table S3.

Deeper gene annotation revealed the presence of genes for biosynthesis of known compounds. The NRPS cluster putatively encoding cyanopeptolin contained several *mcn* homologs, as well as a tryptophan halogenase (Tooming-Klunderud et al., 2007) suggesting the possibility of synthesized halogenated-cyanopeptolin compounds (Fig 4A). The NRPSs with low similarity hits to aeruginosin class compounds pseudospumigin and aeruginoside contained several genes with functional similarity to genes in the *aer* operon, which encode aeruginosin-like compounds (Ishida et al., 2007) (Fig 4C, 4D). Complete gene annotations for BGCs in Figure 4 are illustrated in Table S4. While these bioinformatic insights provides clues about possible synthesis products, further characterization is required to definitively identify the compounds they produce.

3.4 Relative transcript abundance of ADA cyanobacterial genes

We next assessed the transcriptional activity of genes from the two ADA strains with metatranscriptomic data, which was available only for the 100 µm fraction. BGC expression, as estimated through relative transcript abundances, was greatest at WE12 and peak and late phases of the bloom for both ADA strains (Fig 5). There was no measurable expression during July at WE2 and WE4, when both ADA strains were low in relative abundance. BGC expression generally increased during August, and late September-October (Fig 5). Some of the most highly expressed clusters include terpenes, which may encode taste and odor compounds or other novel secondary metabolites (Dittmann et al., 2015). The NRPS cluster with low similarity to an aeruginoside encoding BGC (29%) from LE14-WE12 was one of the most highly expressed clusters from this strain both at nearshore and offshore stations (Fig 5A) while terpene 5 from LE14-WE4 was consistently expressed from peak to late phases at all three stations (Fig 5B).

294 Expression of genes for nutrient metabolism and uptake also varied between strains and
295 throughout the bloom (Table 1, Fig. 5B). During August and late September, both ADA strains
296 had greatest relative transcript abundance of *pusC*, an ABC phosphate transporter, suggesting
297 these ADA strains were actively taking up exogenous phosphate from the water column. On 25
298 August, LE14-WE12 had highest relative transcript abundance of nitrogen fixation genes *nifD*
299 and *nifH* (Fig. 5A). A similar pattern was observed in LE14-WE4, although relatively high *nif*
300 transcript levels persisted throughout the rest of the bloom into early October. Higher relative
301 transcript abundance of another N-fixation gene*, nifK* was also observed in LE14-WE4, as well
302 as heterocyst glycolipid gene clusters which may aid in heterocyst formation ((Kampa et al.,
303 2013), Fig. 5B). Transcriptomic read mapping revealed the majority of *nifD* transcripts in the
304 100 µm fraction were attributed to ADA cyanobacteria (Fig. 6B).

305 *3.5 Quantification of N-fixers*

306 To assess ADA's relative abundance among N-fixers in the broader cyanoHAB
307 community, we searched metagenomic assemblies and reads from the 100 µm fraction and
308 metagenomic reads from whole water for nitrogen fixation genes belonging to other bacterial
309 taxa, which are capable of fixing N in aquatic systems (Bentzon-Tilia et al., 2015; Davis et al.,
310 2015; Shiozaki et al., 2014). In the whole water metagenomes, we detected low relative
311 abundance of alpha-proteobacteria and betaproteobacteria containing *nifHDK* genes. Their
312 relative abundance in comparison to ADA cyanobacteria was low and patchy, with greatest
313 abundance at WE2 (Fig. 6A). These findings suggest ADA are the primary N-fixers in western
314 Lake Erie during the 2014 cyanoHAB (Table S5, Fig. 6).

315 **4. Discussion**

316 Metagenomics and metatranscriptomics provide detailed insights into the diversity of
317 natural cyanoHAB communities, the ecophysiology and biosynthetic potential of the constituent
318 species, and how they respond to environmental conditions. The recovery of highly complete
319 MAGs and paired metatranscriptomic data presented an opportunity to assess the diversity,
320 function, and secondary metabolism of ADA cyanobacteria *in situ* through changing
321 environmental conditions and community composition across the season in western Lake Erie.
322 Although ADA cyanobacteria never dominated the 2014 cyanoHAB, our results raise the
323 possibility that they may make important contributions to production of secondary metabolites
324 and bioavailable N via N-fixation, as discussed below.

325 The ADA taxa were minor but pervasive members of the microbial community at all
326 stations and times sampled during the 2014 western Lake Erie cyanoHAB, consistent with
327 previous studies (Fig 3A) (Berry et al., 2017; Jankowiak et al., 2019; Steffen et al., 2017). Their
328 relative abundance was dynamic, and the two ADA MAGs showed distinct spatiotemporal
329 trends. These findings generally agree with previous studies (Berry et al., 2017; Jankowiak et al.,
330 2019), although our approach implements metagenomic mapping onto entire MAGs instead of
331 gene-targeted approaches. The *Aphanizomenon*-like MAG (LE14-WE12), so called due to its
332 close phylogenomic relationship with strains currently designated as *Aphanizomenon*, was
333 associated with *Microcystis* blooms and most abundant at nearshore stations. In contrast, the
334 *Dolichospermum*-like LE14-WE4 MAG was most abundant offshore, late in the season, when
335 *Microcystis* abundance was low. *Dolichospermum* spp. can also dominate Great Lakes
336 cyanoHABs, for example, following *Microcystis* blooms in the western basin of Lake Erie
337 (Michalak et al., 2012), preceding *Microcystis* as the dominant cyanobacterium early in blooms
338 of the central basin of Lake Erie (Chaffin et al., 2019) , or in Lake Superior (Sheik et al., 2022).

339 *Aphanizomenon* spp. dominated Lake Erie blooms in the 1960s and 1970s (Davis, 1964;
340 Munawar and Munawar, 2011). Although they appear to be rarer now in the open waters of the
341 Great Lakes, they can dominate blooms in tributaries (McKay et al., 2020). The dynamic
342 distribution of ADA organisms in time and space, taken together with evidence for potential
343 biosynthesis of saxitoxin (Chaffin et al., 2019) and guanitoxin (Lima et al., 2022) in the Great
344 Lakes, although not detected in this study, underscore the need to determine environmental and
345 ecological controls on the abundance and repertoire of toxins of ADA organisms.

346 According to a recently proposed phylogeny-based reorganization of the ADA clade into 10
347 species (Dreher et al., 2021), LE14-WE12 would be classified as "ADA-4", and LE14-WE4
348 would be "ADA-2", which both contain strains primarily from the United States, while ADA-2
349 also contains several Finnish isolates (Dreher et al., 2021). Similarly, Österholm et al., 2020
350 divided the ADA group into seven distinct clades via phylogenomics. Based on this analysis,
351 LE14-ADA 1 belongs in the ϒ clade, all of which produce aeruginosins and are from freshwater
352 sources in the United States. Genes related to those for biosynthesis of aeruginosin-like synthesis
353 were found within the LE14-WE12 MAG (Fig. 4, Table S4). LE14-WE4 clusters most similarly
354 with the α clade, which is comprised of strains from both Finland and the United States, with
355 greater BGC diversity including genes that encode for microcystins (4/9 strains), and
356 anabaenopeptins (8/9 strains) (Österholm et al., 2020). While a consensus on taxonomic
357 classification of the ADA clade has yet to be reached, these results underscore the great genetic
358 diversity of species within this group.

359 Each ADA MAG contains a distinct suite of transcriptionally active BGCs encoding
360 secondary metabolites (Fig 1B), although neither contained genes encoding saxitoxin, anatoxin,
361 or guanitoxin, which have been sporadically detected in Lake Erie cyanoHABs (Chaffin et al.,

2019; Lima et al., 2022; Miller et al., 2017). Two transcriptionally active BGCs from LE14-WE12 may encode toxic compounds related to the aeruginosin class, NRPS aeruginoside (29%) and NRPS psuedospumigin (66%) (Fig 5A). Aeruginosides belong to the peptide class aeruginosins and have been characterized in *Planktothrix* (Ishida et al., 2007). While two core NRPS biosynthetic genes required for aeruginosin biosynthesis appear to be conserved in these clusters (Fig 4D), the low percent similarity of LE14-WE12 aeruginoside cluster to the MiBIG database does not provide sufficient evidence to confirm aeruginoside synthesis. Likewise, a cluster predicted to encode pseudospumigin-like compound (Fig 4C, 66% identity), a linear peptide related to aeruginosins and spumigins, which may inhibit trypsin (Jokela et al., 2017), was also identified, but further work is needed to confirm the product synthesized from this cluster as well.

Although low in transcriptional activity, LE14-WE4 contained several BGCs with high similarity to known clusters that encode toxic compounds including related compounds cyanopeptolin (Tooming-Klunderud et al., 2007) and cyanopeptin (Rounge et al., 2007) as well as taste and odor compound geosmin (Gerber and Lechevalier, 1965; Watson et al., 2016b). Several other BGCs with measurable transcriptional activity do not have known associated compounds, including the terpene and NRPS classes observed in both LE14-WE12 and WE4. Similar trends of highly expressed but coarsely resolved BGCs, are observed in *Microcystis* genomes isolated from the same bloom (Yancey et al., 2022b). In addition, many of the ADA BGCs identified here, including several that encode the biosynthesis of compounds with known toxic and/or taste and odor properties (aeruginoside-like and pseudospumigin-like, geosmin) are not present in *Microcystis* genomes isolated from the same bloom (Yancey et al., 2022b), underscoring the potential for ADA to exacerbate water quality issues. Together, this highlights

385 our limited understanding of the breadth of cyanobacterial secondary metabolism within western
386 Lake Erie and the need the assess potential risks of these secondary metabolites to ecosystem and
387 human health.

388 The dynamic relative abundance of transcripts from nutrient uptake and metabolism genes of
389 ADA cyanobacteria suggest shifts in nutrient acquisition strategies along with changing nutrient
390 availability throughout the bloom. While our metatranscriptomic sequencing data cannot be used
391 to estimate absolute abundance of transcripts (e.g., transcripts per ml) or expression (i.e.,
392 transcripts per gene copy), our calculations of the abundance of transcripts for each gene relative
393 to all transcripts from the MAG provide estimates of transcriptional activity of each gene relative
394 to the full transcriptome of each MAG. During low phosphate conditions of the peak bloom
395 phase at WE12 (Fig. 3B), both ADA strains showed high relative abundance of transcripts for
396 phosphate ABC transporter *pusC* (Fig 5), suggesting P-limitation, stress, and/or competition with
397 *Microcystis*, which is efficient at uptake and storage of inorganic P (Jacobson and Halmann,
398 1982).

399 We also observed transcripts from various genes involved in N-fixation across time and
400 space in western Lake Erie. This finding, along with previously reported expression of *nifH*
401 (Steffen et al., 2015) and N fixation rates in western Lake Erie (Natwora and Sheik, 2021),
402 challenge the notion that current western Lake Erie cyanoHABs do not contain N-fixers (Barnard
403 et al., 2021; Newell et al., 2019), although its quantitative importance over time and space
404 remain unclear. Further, our results support the inference that N-fixing ADA associated with
405 *Microcystis* blooms underpin hot spots of N fixation in the Great Lakes (Natwora and Sheik,
406 2021). Our results also suggest that the ADA taxa are the primary N-fixers in western Lake Erie,
407 with low level abundances of non-cyanobacterial diazotrophs being most abundant at nearshore

stations in the whole water fraction (Fig 6A). This contrasts patterns observed in sites across western Lake Erie and in Sandusky Bay where N fixation has been attributed to both cyanobacteria and heterotrophic bacteria (Davis et al., 2015; Jankowiak and Gobler, 2020). Diazotrophy may be a critical strategy by which ADA cyanobacteria compete given that *Microcystis* efficiently scavenges N (Takamura et al., 1987; Wang et al., 2021); late season blooms are often limited by exogenous N (Chaffin et al., 2013), and secondary blooms of ADA cyanobacteria may be enabled by N fixation (Michalak et al., 2013). N fixation by ADA cyanobacteria may also benefit *Microcystis* by mitigating N limitation under some circumstances. N rich amino acids and ammonium may "leak" from N-fixers and provide N to cyanoHABs (Ohlendieck et al., 2000; Wetzel, 2001), thereby providing resources supporting persistence of *Microcystis* blooms. The greater relative transcript abundance of N-fixation genes during August 25th and October 6th (Fig. 5), when nitrate is depleted (Fig. 3), suggests that diazotrophy may provide an additional source of N, although its quantitative importance remains unclear based on the data currently available. Previous work has demonstrated low abundance cyanobacteria capable of nitrogen fixation, may serve as keystone species, providing bioavailable nitrogen to deplete systems via diazotrophy (Shiozaki et al., 2020). More work is needed address the potential impact of "leaky" N-fixation on Lake Erie cyanoHABs and to definitively quantify the contribution of each taxa to rates of N-fixation.

5. Conclusion

Given the prevalence and threats posed by *Microcystis,* other cyanobacteria present in western Lake Erie rarely receive attention. However, this study highlights two distinct species of ADA cyanobacteria that were pervasively present at low abundance in 2014 and have the potential to produce a variety of known and unknown secondary metabolites. Their biosynthetic

repertoire and transcriptional activity are varied and may reflect changes in bloom conditions as they compete for N and P and use alternate strategies for N acquisition and growth. N-fixation may help satisfy N demand and support the synthesis of N-rich secondary metabolites, especially during N deplete conditions, though the significance of N fixation from ADA organisms and the broader community remains to be quantified. While the monitoring of *Microcystis* and the production of its secondary metabolites is critical within western Lake Erie, this study demonstrates the need to further expand efforts toward other cyanobacteria found within these blooms as they may produce unmonitored toxins and shape observed community and ecosystem dynamics.

Acknowledgements

We thank Dack Stuart and Kent Baker for field sampling support, and the NOAA Great Lakes Environmental Research Lab and Cooperative Institute for Great Lakes Research for providing nutrient data and allowing us to sample alongside their HAB monitoring efforts.

Funding

This work was supported by NIH and NSF awards to the Great Lakes Center for Fresh Waters and Human Health (NIH: 1P01ES028939-01, NSF: OCE-1840715) and the Cooperative Institute of Great Lakes Research (NA17OAR4320152). Support was also provided by NOAA OAR 'Omics and NOAA OAR Ocean Technology Development Initiative.

Conflict of Interest

The authors state no conflict of interest.

Figure Captions and Tables

Figure 1: A) Map of Western Lake Erie and sampling stations in which ADA MAGs were recovered. LE14-WE12 was recovered from WE12 (purple) and LE14-WE4 was recovered from WE4 (orange). B) Summary statistics for ADA MAGs recovered from 2014 W. Lake Erie cyanoHAB, including sequences annotated as BGCs via antiSMASH v.6.0. Figure 1A was generated via QGIS using the Open Street Map (OSM) as a basemap (https://wiki.osmfoundation.org/wiki/Main_Page).

Figure 2: Phylogenomic tree consisting of members from the ADA clade. A maximum likelihood tree was generated using 251 housekeeping genes via the GToTree workflow and included ADA genomes from previous studies (Dreher et al., 2021; Österholm et al., 2020; Sheik et al., 2022). MAGs isolated from western Lake Erie are highlighted in bolded red text. ADA strains that were isolated from the Great Lakes are highlighted in blue text.

Figure 3: A) Relative abundance estimates of *Microcystis* and ADA as determined by mapping metagenomic reads onto MAGs for each taxon. Relative abundance was estimated via reads per kilobase million (RPKM, see methods). Note the difference in scale between *Microcystis* and ADA plots. B) Nitrate (NO_3^-), phycocyanin (PC) and soluble reactive phosphorus measurements at stations WE2, WE4, and WE12, throughout the course of the 2014 cyanoHAB. PC is used as a proxy for cyanobacterial biomass.

Figure 4: Schematics of selected BGCs as identified by antiSMASH v.6. A) NRPS with 100% similarity to cyanopeptolin encoding gene clusters. This cluster contains several genes with high similarity to *mcn* genes, as well as those involved in transport. Putative flavin-oxidoreductase genes were also identified. B) A terpene BGC with 100% similarity to geosmin encoding clusters. C) An NRPS which may encode an aeruginosin-like compound with some similarity to the gene cluster encoding pseudospumigin (66%). This cluster contained an *aerA*-like gene and a

475 putative serine/threonine protein kinase. D) An NRPS cluster which may encode an aeruginosin-
476 like compound with low similarity to aeruginoside (29%). This cluster contained 2 genes with
477 similarity to *aerG* and *aerF*.

478 **Figure 5:** Relative transcript abundance for both BGCs and nutrient uptake/metabolism genes
479 from ADA MAGs. Expression estimates were calculated by mapping metatranscriptomic reads
480 onto BGCs identified via antiSMASH. Percent similarities are listed next to BGCs with putative
481 identifications. Nutrient metabolism gene transcript abundance is depicted in the bottom row.
482 Relative transcript abundance calculations were completed for both A) LE14-WE12 and B)
483 LE14-WE4.

484 **Figure 6:** A) Relative abundance of *nifHDK* genes in the whole water fraction as estimated
485 through read mapping and B) relative transcript abundance from the 100 µm fraction
486 metatranscriptome samples. Reads were aligned to *nifHDK* genes in the FunGenes (Fish et al.,
487 2013) repository and relatively quantified via reads per kilobase per million reads (RPKM).
488 Samples are separated by station. The majority of reads mapped to ADA cyanobacteria, with low
489 level read mapping to alphaproteobacteria, betaproteobacteria, and "other cyanobacteria"
490 (*Nostoc* spp.).

491 **Table 1:** Functional roles of N and P genes. Boxes colored gray are genes involved in N fixation

Gene	Function	IMG ID	
		Dolichospermum	*Aphanizominum*
Phosphorus Metabolism Genes			
phoH	phosphate starvation inducible protein	2562231174	2914222179
phoU	phosphate transport	2562230301	2632406494
phoR	phosphate regulation histidine kinase	2562230300	2914222838
ppk	polyphosphate kinase	2562227599	2914220152
pstB 1	phosphate ABC transporter ATP-binding	2562229322	2914220293
pstB 2	phosphate transport system ATP-binding	2562228640	2914222488
pstA	phosphate ABC transporter ATP-binding	2562229321	2914222487

pstC	phosphate ABC transporter permease	2562229320	2914222486
pusC	phosphate ABC transporter permease protein	2562229319	2914223119
Nitrogen Metabolism Genes			
ntcA	nitrogen-responsive regulatory protein	2562229740	2914220114
ntcB	LysR family transcriptional regulator	2562229217	2632404561
glnB	nitrogen regulatory protein P-II	2562231361	2632405495
nrrA	two component transcriptional regulator, winged helix family	2562226995	2632404561
nifH	Mo-nitrogenase iron protein subunit	2562230999	2914221270
nifD	Mo-nitrogenase MoFe protein subunit NifD precursor	2562231091	2632405170
nifK	nitrogenase molybdenum-iron protein beta chain	2562231113	2632405420
nifB	nitrogenase cofactor biosynthesis protein	2562230995	2632405793
fdxN	4Fe-4S ferredoxin iron-sulfur binding domain-containing	2562230996	2632405441
nifS	cysteine desulfurase	2562228928	2632405440
nifU	cysteine desulfurase	2562230998	2632404346

References

Adobe Illustrator, 2019. Adobe Inc. URL https://www.adobe.com/products/illustrator.html (accessed 1.14.21).

Allaire, J.J., 2015. RStudio: Integrated Development Environment for R. Boston.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Loman, N.J., Andersson, A.F., Quince, C., 2013. CONCOCT: Clustering cONtigs on COverage and ComposiTion.

Al-Tebrineh, J., Mihali, T.K., Pomati, F., Neilan, B.A., 2010. Detection of saxitoxin-producing cyanobacteria and Anabaena circinalis in environmental water blooms by quantitative PCR. Appl Environ Microbiol 76, 7836–7842. https://doi.org/10.1128/AEM.00174-1

Barnard, M.A., Chaffin, J.D., Plaas, H.E., Boyer, G.L., Wei, B., Wilhelm, S.W., Rossignol, K.L., Braddy, J.S., Bullerjahn, G.S., Bridgeman, T.B., Davis, T.W., Wei, J., Bu, M., Paerl, H.W., 2021. Roles of Nutrient Limitation on Western Lake Erie CyanoHAB Toxin Production. Toxins (Basel) 13, 47. https://doi.org/10.3390/toxins13010047

Bentzon-Tilia, M., Severin, I., Hansen, L.H., Riemann, L., 2015. Genomics and ecophysiology of heterotrophic nitrogen-fixing bacteria isolated from estuarine surface water. mBio 6. https://doi.org/10.1128/MBIO.00929-15

Berry, M.A., Davis, T.W., Cory, R.M., Duhaime, M.B., Johengen, T.H., Kling, G.W., Marino, J.A., den Uyl, P.A., Gossiaux, D., Dick, G.J., Denef, V.J., 2017. Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities. Environ Microbiol 19, 1149–1162. https://doi.org/10.1111/1462-2920.13640

Bister, B., Keller, S., Baumann, H.I., Nicholson, G., Weist, S., Jung, G., Süssmuth, R.D., Jüttner, F., 2004. Cyanopeptolin 963A, a chymotrypsin inhibitor of Microcystis PCC 7806. J Nat Prod 67, 1755–1757. https://doi.org/10.1021/NP049828F

Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber, T., 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49, W29–W35. https://doi.org/10.1093/NAR/GKAB335

Bridgeman, T.B., Chaffin, J.D., Filbrun, J.E., 2013. A novel method for tracking western Lake Erie Microcystis blooms, 2002–2011. J Great Lakes Res 39, 83–89. https://doi.org/10.1016/J.JGLR.2012.11.004

Bullerjahn, G.S., McKay, R.M., Davis, T.W., Baker, D.B., Boyer, G.L., D'Anglada, L. v., Doucette, G.J., Ho, J.C., Irwin, E.G., Kling, C.L., Kudela, R.M., Kurmayer, R., Michalak, A.M., Ortiz, J.D., Otten, T.G., Paerl, H.W., Qin, B., Sohngen, B.L., Stumpf, R.P., Visser, P.M., Wilhelm, S.W., 2016. Global solutions to regional problems: Collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. Harmful Algae 54, 223–238. https://doi.org/10.1016/J.HAL.2016.01.003

Carmichael, W.W., Biggs, D.F., Gorham, P.R., 1975. Toxicology and Pharmacological Action of Anabaena flos-aquae Toxin. Science (1979) 187, 542–544. https://doi.org/10.1126/SCIENCE.803708

Chaffin, J.D., Bridgeman, T.B., Bade, D.L., 2013. Nitrogen Constrains the Growth of Late Summer Cyanobacterial Blooms in Lake Erie. Adv Microbiol 2013, 16–26. https://doi.org/10.4236/AIM.2013.36A003

Chaffin, J.D., Mishra, S., Kane, D.D., Bade, D.L., Stanislawczyk, K., Slodysko, K.N., Jones, K.W., Parker, E.M., Fox, E.L., 2019. Cyanobacterial blooms in the central basin of Lake Erie: Potentials for cyanotoxins and environmental drivers. J Great Lakes Res 45, 277–289. https://doi.org/10.1016/j.jglr.2018.12.006

Cooperative Institute for Great Lakes Research, U. of M.N.G.L.E.R.L., n.d. Physical, chemical, and biological water quality monitoring data to support detection of Harmful Algal Blooms (HABs) in western Lake Erie, collected by the Great Lakes Environmental Research Laboratory and the Cooperative Institute for Great Lakes Research since 2012. https://doi.org/https://doi.org/10.25921/11da-3x54

Davis, C.C., 1964. Evidence for the Eutrophication of Lake Erie from Phytoplankton Records. Limnol Oceanogr 9, 275–283. https://doi.org/10.4319/LO.1964.9.3.0275

Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., Watson, S.B., 2015. Effects of Increasing Nitrogen and Phosphorus Concentrations on Phytoplankton Community Growth and Toxicity during Planktothrix Blooms in Sandusky Bay, Lake Erie. Environ Sci Technol 49, 7197–7207. https://doi.org/10.1021/ACS.EST.5B00799

Davis, T.W., Harke, M.J., Marcoval, M.A., Goleski, J., Orano-Dawson, C., Berry, D.L., Gobler, C.J., 2010. Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of Microcystis during cyanobacterial blooms. Aquatic Microbial Ecology 61, 149–162. https://doi.org/10.3354/AME01445

Dick, G.J., 2018. Genomic Approaches in Earth and Environmental Sciences, 1st ed. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118708231

Dittmann, E., Gugger, M., Sivonen, K., Fewer, D.P., 2015. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. Trends Microbiol 23, 642–652. https://doi.org/10.1016/J.TIM.2015.07.008

Dreher, T.W., Davis, E.W., Mueller, R.S., Otten, T.G., 2021. Comparative genomics of the ADA clade within the Nostocales. Harmful Algae 104, 102037. https://doi.org/10.1016/J.HAL.2021.102037

Driscoll, C.B., Meyer, K.A., Šulčius, S., Brown, N.M., Dick, G.J., Cao, H., Gasiūnas, G., Timinskas, A., Yin, Y., Landry, Z.C., Otten, T.G., Davis, T.W., Watson, S.B., Dreher, T.W., 2018. A closely-related clade of globally distributed bloom-forming cyanobacteria within the Nostocales. Harmful Algae 77, 93–107. https://doi.org/10.1016/J.HAL.2018.05.009

Driscoll, C.B., Otten, T.G., Brown, N.M., Dreher, T.W., 2017. Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. Stand Genomic Sci 12, 9. https://doi.org/10.1186/s40793-017-0224-8

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T.O., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3, e1319. https://doi.org/10.7717/peerj.1319

Fewer, D., Tooming-Klunderud, A., Jokela, J., Wahlsten, M., Rouhiainen, L., Kristensen, T., Rohrlack, T., Jakobsen, K., Sivonen, K., 2008. Natural occurrence of microcystin synthetase deletion mutants capable of producing microcystins in strains of the genus Anabaena (Cyanobacteria). Microbiology (Reading) 154, 1007–1014. https://doi.org/10.1099/MIC.0.2007/016097-0

Fish, J.A., Chai, B., Wang, Q., Sun, Y., Brown, C.T., Tiedje, J.M., Cole, J.R., 2013. FunGene: The functional gene pipeline and repository. Front Microbiol 4, 291. https://doi.org/10.3389/FMICB.2013.00291

Gademann, K., Portmann, C., Blom, J.F., Zeder, M., Jüttner, F., 2010. Multiple toxin production in the cyanobacterium Microcystis: Isolation of the toxic protease inhibitor cyanopeptolin 1020. J Nat Prod 73, 980–984. https://doi.org/10.1021/NP900818C

Gerber, N.N., Lechevalier, H.A., 1965. Geosmin, an Earthy-Smelling Substance Isolated from Actinomycetes. Appl Microbiol 13, 935–938. https://doi.org/10.1128/AM.13.6.935-938.1965

Gobler, C.J., Burkholder, J.A.M., Davis, T.W., Harke, M.J., Johengen, T., Stow, C.A., van de Waal, D.B., 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. Harmful Algae. https://doi.org/10.1016/j.hal.2016.01.010

Halm, H., Lam, P., Ferdelman, T.G., Lavik, G., Dittmar, T., Laroche, J., D'Hondt, S., Kuypers, M.M.M., 2011. Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. The ISME Journal 2011 6:6 6, 1238–1249. https://doi.org/10.1038/ismej.2011.182

Ishida, K., Christiansen, G., Yoshida, W.Y., Kurmayer, R., Welker, M., Valls, N., Bonjoch, J., Hertweck, C., Börner, T., Hemscheidt, T., Dittmann, E., 2007. Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. Chem Biol 14, 565–576. https://doi.org/10.1016/J.CHEMBIOL.2007.04.006

Izaguirre, G., Taylor, W.D., 2004. A guide to geosmin- and MIB-producing cyanobacteria in the United States. Water Science and Technology 49, 19–24. https://doi.org/10.2166/WST.2004.0524

Jacobson, L., Halmann, M., 1982. Polyphosphate metabolism in the blue-green alga Microcystis aeru-ginosa. J Plankton Res 4, 481–488. https://doi.org/10.1093/PLANKT/4.3.481

Jankowiak, J., Hattenrath-Lehmann, T., Kramer, B.J., Ladds, M., Gobler, C.J., 2019. Deciphering the effects of nitrogen, phosphorus, and temperature on cyanobacterial bloom intensification, diversity, and toxicity in western Lake Erie. Limnol Oceanogr 64, 1347–1370. https://doi.org/10.1002/lno.11120

Jankowiak, J.G., Gobler, C.J., 2020. The Composition and Function of Microbiomes Within Microcystis Colonies Are Significantly Different Than Native Bacterial Assemblages in Two North American Lakes. Front Microbiol 11, 1016. https://doi.org/10.3389/FMICB.2020.01016/BIBTEX

Jokela, J., Heinilä, L.M.P., Shishido, T.K., Wahlsten, M., Fewer, D.P., Fiore, M.F., Wang, H., Haapaniemi, E., Permi, P., Sivonen, K., 2017. Production of High Amounts of Hepatotoxin Nodularin and New Protease Inhibitors Pseudospumigins by the Brazilian Benthic Nostoc sp. CENA543. Front Microbiol 8. https://doi.org/10.3389/FMICB.2017.01963

Kampa, A., Gagunashvili, A.N., Gulder, T.A.M., Morinaka, B.I., Daolio, C., Godejohann, M., Miao, V.P.W., Piel, J., Andrésson, Ó.S., 2013. Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. Proc Natl Acad Sci U S A 110, E3129. https://doi.org/10.1073/PNAS.1305867110

599 Kang, D.D., Froula, J., Egan, R., Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single
600 genomes from complex microbial communities. PeerJ 3, e1165. https://doi.org/10.7717/peerj.1165

601 Kassambara, A., 2020. ggpubr: "ggplot2" Based Publication Ready Plots. URL https://cran.r-
602 project.org/web/packages/ggpubr/index.html (accessed 1.14.21).

603 Kehr, J.-C., Gatte Picchi, D., Dittmann, E., 2011. Natural product biosyntheses in cyanobacteria: A treasure trove of
604 unique enzymes. Beilstein journal of organic chemistry 7, 1622–35. https://doi.org/10.3762/bjoc.7.191

605 Kumar, K., Mella-Herrera, R.A., Golden, J.W., 2010. Cyanobacterial Heterocysts. Cold Spring Harb Perspect Biol
606 2. https://doi.org/10.1101/CSHPERSPECT.A000315

607 Laczny, C.C., Sternal, T., Plugaru, V., Gawron, P., Atashpendar, A., Margossian, H., Coronado, S., der Maaten, L.,
608 Vlassis, N., Wilmes, P., 2015. VizBin - an application for reference-independent visualization and human-
609 augmented binning of metagenomic data. Microbiome 3, 1. https://doi.org/10.1186/s40168-014-0066-1

610 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359.
611 https://doi.org/10.1038/nmeth.1923

612 Lee, M.D., 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35, 4162–4164.
613 https://doi.org/10.1093/bioinformatics/btz188

614 Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large
615 and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.
616 https://doi.org/10.1093/bioinformatics/btv033

617 Li, X., Dreher, T.W., Li, R., 2016. An overview of diversity, occurrence, genetics and toxin production of bloom-
618 forming Dolichospermum (Anabaena) species. Harmful Algae 54, 54–68.
619 https://doi.org/10.1016/J.HAL.2015.10.015

620 Lima, S.T., Fallon, T.R., Cordoza, J.L., Chekan, J.R., Delbaje, E., Hopiavuori, A.R., Alvarenga, D.O., Wood, S.M.,
621 Luhavaya, H., Baumgartner, J.T., Dörr, F.A., Etchegaray, A., Pinto, E., McKinnie, S.M.K., Fiore, M.F.,
622 Moore, B.S., 2022. Biosynthesis of Guanitoxin Enables Global Environmental Detection in Freshwater
623 Cyanobacteria. J Am Chem Soc 144, 9372–9379. https://doi.org/10.1021/JACS.2C01424

624 Madden, T., 2013. The BLAST Sequence Analysis Tool. The NCBI Handbook, 2nd ed.

625 Makarewicz, J.C., 1993. Phytoplankton Biomass and Species Composition In Lake Erie, 1970 to 1987. J Great
626 Lakes Res 19, 258–274. https://doi.org/10.1016/S0380-1330(93)71216-2

627 Matisoff, G., Ciborowski, J.J.H., 2005. Lake Erie Trophic Status Collaborative Study. J Great Lakes Res 31, 1–10.
628 https://doi.org/10.1016/S0380-1330(05)70300-2

629 McKay, R.M., Frenken, T., Diep, N., Cody, W.R., Crevecoeur, S., Dove, A., Drouillard, K.G., Ortiz, X.,
630 Wintermute, J., Zastepa, A., 2020. Bloom announcement: An early autumn cyanobacterial bloom co-
631 dominated by Aphanizomenon flos- aquae and Planktothrix agardhii in an agriculturally-influenced Great
632 Lakes tributary (Thames River, Ontario, Canada). Data Brief 30. https://doi.org/10.1016/J.DIB.2020.105585

633 McKindles, K., Frenken, T., McKay, R.M.L., Bullerjahn, G.S., 2020. Binational efforts addressing cyanobacterial
634 harmful algal blooms in the great lakes. Handbook of Environmental Chemistry 101, 109–133.
635 https://doi.org/10.1007/698_2020_513

636 Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K.,
637 Confesor, R., Daloglu, I., DePinto, J. v., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L.,
638 Johengen, T.H., Kuo, K.C., LaPorte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards,
639 R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., Zagorski, M.A., 2013. Record-setting algal

640 bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future
641 conditions. Proc Natl Acad Sci U S A 110, 6448–6452. https://doi.org/10.1073/PNAS.1216006110

642 Miller, T.R., Beversdorf, L.J., Weirich, C.A., Bartlett, S.L., 2017. Cyanobacterial Toxins of the Laurentian Great
643 Lakes, Their Toxicological Effects, and Numerical Limits in Drinking Water. Mar Drugs 15.
644 https://doi.org/10.3390/MD15060160

645 Mise, K., Masuda, Y., Senoo, K., Itoh, H., 2021. Undervalued Pseudo- nifH Sequences in Public Databases Distort
646 Metagenomic Insights into Biological Nitrogen Fixers . mSphere 6. https://doi.org/10.1128/MSPHERE.00785-
647 21

648 Munawar, M., Munawar, I.F., 2011. A Lakewide Study of Phytoplankton Biomass and its Species Composition in
649 Lake Erie, April–December 1970. https://doi.org/10.1139/f76-075 33, 581–600. https://doi.org/10.1139/F76-
650 075

651 Natwora, K.E., Sheik, C.S., 2021. Assessment of nitrogen fixation rates in the Laurentian Great Lakes. J Great
652 Lakes Res 47, 1288–1295. https://doi.org/10.1016/J.JGLR.2021.07.005

653 Newell, S.E., Davis, T.W., Johengen, T.H., Gossiaux, D., Burtner, A., Palladino, D., McCarthy, M.J., 2019.
654 Reduced forms of nitrogen are a driver of non-nitrogen-fixing harmful cyanobacterial blooms and toxicity in
655 Lake Erie. Harmful Algae 81, 86–93. https://doi.org/10.1016/J.HAL.2018.11.003

656 Ohlendieck, U., Stuhr, A., Siegmund, H., 2000. Nitrogen fixation by diazotrophic cyanobacteria in the Baltic Sea
657 and transfer of the newly fixed nitrogen to picoplankton organisms. Journal of Marine Systems 25, 213–219.
658 https://doi.org/10.1016/S0924-7963(00)00016-6

659 Österholm, J., Popin, R. v., Fewer, D.P., Sivonen, K., 2020. Phylogenomic Analysis of Secondary Metabolism in the
660 Toxic Cyanobacterial Genera Anabaena, Dolichospermum and Aphanizomenon. Toxins (Basel) 12, 248.
661 https://doi.org/10.3390/toxins12040248

662 Paerl, H.W., Otten, T.G., Kudela, R., 2018. Mitigating the Expansion of Harmful Algal Blooms Across the
663 Freshwater-to-Marine Continuum. Environ Sci Technol 52, 5519–5529.
664 https://doi.org/10.1021/acs.est.7b05950

665 Paerl, H.W., Scott, J.T., 2010. Throwing fuel on the fire: Synergistic effects of excessive nitrogen inputs and global
666 warming on harmful algal blooms. Environ Sci Technol. https://doi.org/10.1021/es102665e

667 Paerl, H.W., Scott, J.T., McCarthy, M.J., Newell, S.E., Gardner, W.S., Havens, K.E., Hoffman, D.K., Wilhelm,
668 S.W., Wurtsbaugh, W.A., 2016. It Takes Two to Tango: When and Where Dual Nutrient (N & P) Reductions
669 Are Needed to Protect Lakes and Downstream Ecosystems. Environ Sci Technol 50, 10805–10813.
670 https://doi.org/10.1021/ACS.EST.6B02575

671 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of
672 microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25, 1043–55.
673 https://doi.org/10.1101/gr.186072.114

674 Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G., Toth, I.K., 2015. Genomics and taxonomy in
675 diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods 8, 12–24.
676 https://doi.org/10.1039/C5AY02550H

677 Raymond, J., Siefert, J.L., Staples, C.R., Blankenship, R.E., 2004. The Natural History of Nitrogen Fixation. Mol
678 Biol Evol 21, 541–554. https://doi.org/10.1093/MOLBEV/MSH047

679 Rinta-Kanto, J., Ouellette, A., Boyer, G., Twiss, M., Bridgeman, T., Wilhelm, S., 2005. Quantification of Toxic
680 Microcystis spp. during the 2003 and 2004 Blooms in Western Lake Erie using Quantitative Real-Time PCR.
681 Environ Sci Technol 39, 4198–4205. https://doi.org/10.1021/ES048249U

682 Rinta-Kanto, J.M., Konopko, E.A., DeBruyn, J.M., Bourbonniere, R.A., Boyer, G.L., Wilhelm, S.W., 2009. Lake
683 Erie Microcystis: Relationship between microcystin production, dynamics of genotypes and environmental
684 parameters in a large lake. Harmful Algae 8, 665–673. https://doi.org/10.1016/j.hal.2008.12.004

685 Rounge, T.B., Rohrlack, T., Tooming-Klunderud, A., Kristensen, T., Jakobsen, K.S., 2007. Comparison of
686 cyanopeptolin genes in Planktothrix, Microcystis, and Anabaena strains: Evidence for independent evolution
687 within each genus. Appl Environ Microbiol 73, 7322–7330. https://doi.org/10.1128/AEM.01475-07

688 Schindler, D.W., 1974. Eutrophication and recovery in experimental lakes: Implications for lake management.
689 Science (1979) 184, 897–899. https://doi.org/10.1126/science.184.4139.897

690 Sheik, C.S., Natwora, K.E., Alexson, E.E., Callaghan, J.D., Sailer, A., Schreiner, K.M., Steinman, B.A.,
691 Finkenbinder, M.S., Filstrup, C.T., Bramburger, A.J., 2022. Dolichospermum blooms in Lake Superior: DNA-
692 based approach provides insight to the past, present and future of blooms. J Great Lakes Res.
693 https://doi.org/10.1016/J.JGLR.2022.08.002

694 Shiozaki, T., Fujiwara, A., Inomura, K., Hirose, Y., Hashihama, F., Harada, N., 2020. Biological nitrogen fixation
695 detected under Antarctic sea ice. Nature Geoscience 2020 13:11 13, 729–732. https://doi.org/10.1038/s41561-
696 020-00651-7

697 Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S., Furuya, K., 2014. Heterotrophic bacteria as major nitrogen fixers in
698 the euphotic zone of the Indian Ocean. Global Biogeochem Cycles 28, 1096–1110.
699 https://doi.org/10.1002/2014GB004886

700 Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., Banfield, J.F., 2018. Recovery of
701 genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3, 836–843.
702 https://doi.org/10.1038/s41564-018-0171-1

703 Sierra, M.A., Martínez-Álvarez, R., 2020. Ricin and Saxitoxin: Two Natural Products That Became Chemical
704 Weapons. J Chem Educ 97, 1707–1714. https://doi.org/10.1021/ACS.JCHEMED.9B00841

705 Steffen, M.M., Belisle, B.S., Watson, S.B., Boyer, G.L., Bourbonniere, R.A., Wilhelm, S.W., 2015.
706 Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial
707 communities. Appl Environ Microbiol 81, 3268–3276. https://doi.org/10.1128/AEM.04101-14

708 Steffen, M.M., Davis, T.W., McKay, R.M.L., Bullerjahn, G.S., Krausfeldt, L.E., Stough, J.M.A., Neitzey, M.L.,
709 Gilbert, N.E., Boyer, G.L., Johengen, T.H., Gossiaux, D.C., Burtner, A.M., Palladino, D., Rowe, M.D., Dick,
710 G.J., Meyer, K.A., Levy, S., Boone, B.E., Stumpf, R.P., Wynne, T.T., Zimba, P. v., Gutierrez, D., Wilhelm,
711 S.W., 2017. Ecophysiological Examination of the Lake Erie *Microcystis* Bloom in 2014: Linkages between
712 Biology and the Water Supply Shutdown of Toledo, OH. Environ Sci Technol 51, 6745–6755.
713 https://doi.org/10.1021/acs.est.7b00856

714 Stüken, A., Jakobsen, K.S., 2010. The cylindrospermopsin gene cluster of Aphanizomenon sp. strain 10E6:
715 Organization and recombination. Microbiology (N Y) 156, 2438–2451.
716 https://doi.org/10.1099/MIC.0.036988-0

717 Stumpf, R.P., Wynne, T.T., Baker, D.B., Fahnenstiel, G.L., 2012. Interannual Variability of Cyanobacterial Blooms
718 in Lake Erie. PLoS One 7, e42444. https://doi.org/10.1371/JOURNAL.PONE.0042444

719 Takamura, N., Iwakuma, T., Yasuno, M., 1987. Uptake of 13 C and 15 N (ammonium, nitrate and urea) by
720 Microcystis in Lake Kasumigaura. J Plankton Res 9, 151–165. https://doi.org/10.1093/PLANKT/9.1.151

721 Tonk, L., Welker, M., Huisman, J., Visser, P.M., 2009. Production of cyanopeptolins, anabaenopeptins, and
722 microcystins by the harmful cyanobacteria Anabaena 90 and Microcystis PCC 7806. Harmful Algae 8, 219–
723 224. https://doi.org/10.1016/j.hal.2008.05.005

Tooming-Klunderud, A., Rohrlack, T., Shalchian-Tabrizi, K., Kristensen, T., Jakobsen, K.S., 2007. Structural analysis of a non-ribosomal halogenated cyclic peptide and its putative operon from Microcystis: Implications for evolution of cyanopeptolins. Microbiology (N Y) 153, 1382–1393. https://doi.org/10.1099/mic.0.2006/001123-0

Ultsch, A., Mörchen, F., 2009. ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM Analysis of high dimensional and big data View project Genetic Foundations of Leukemia View project ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM.

Vaitomaa, J., Rantala, A., Halinen, K., Rouhiainen, L., Tallberg, P., Mokelke, L., Sivonen, K., 2003. Quantitative real-time PCR for determination of microcystin synthetase e copy numbers for microcystis and anabaena in lakes. Appl Environ Microbiol 69, 7289–7297. https://doi.org/10.1128/AEM.69.12.7289-7297.2003

Walsby, A.E., Hayes, P.K., Boje, R., 2007. The gas vesicles, buoyancy and vertical distribution of cyanobacteria in the Baltic Sea. Eur J Phycol. 30, 87–94. https://doi.org/10.1080/09670269500650851

Wang, H., Sivonen, K., Rouhiainen, L., Fewer, D.P., Lyra, C., Rantala-Ylinen, A., Vestola, J., Jokela, J., Rantasärkkä, K., Li, Z., Liu, B., 2012. Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium Anabaena sp. strain 90. BMC Genomics 13, 1–17. https://doi.org/10.1186/1471-2164-13-613

Wang, J., Wagner, N.D., Fulton, J.M., Scott, J.T., 2021. Diazotrophs modulate phycobiliproteins and nitrogen stoichiometry differently than other cyanobacteria in response to light and nitrogen availability. Limnol Oceanogr 66, 2333–2345. https://doi.org/10.1002/LNO.11757

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Höök, T.O., Ludsin, S.A., Matisoff, G., McElmurry, S.P., Murray, M.W., Peter Richards, R., Rao, Y.R., Steffen, M.M., Wilhelm, S.W., 2016a. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. Harmful Algae 56, 44–66. https://doi.org/10.1016/J.HAL.2016.04.010

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

Wetzel, R.G., 2001. Limology Lake and River Ecosystems 3rd ed. Low Temperature Physics Academic P, 17.

Wickham, H., 2011. ggplot2. Wiley Interdiscip Rev Comput Stat 3, 180–185. https://doi.org/10.1002/WICS.147

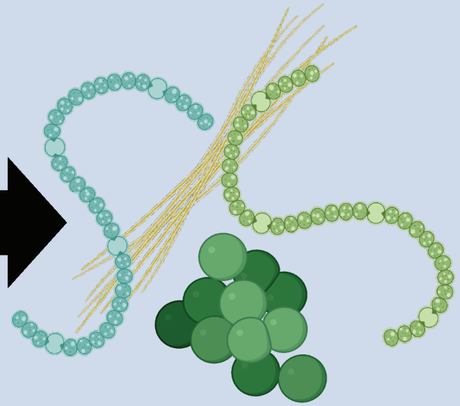
Wolk, C.P., Ernst, A., Elhai, J., 1994. Heterocyst Metabolism and Development. The Molecular Biology of Cyanobacteria 769–823. https://doi.org/10.1007/978-94-011-0227-8_27

Wood, S.A., Prentice, M.J., Smith, K., Hamilton, D.P., 2010. Low dissolved inorganic nitrogen and increased heterocyte frequency: precursors to Anabaena planktonica blooms in a temperate, eutrophic reservoir. J Plankton Res 32, 1315–1325. https://doi.org/10.1093/PLANKT/FBQ048

Yancey, C.E., Smith, D.J., Uyl, P.A. den, Mohamed, O.G., Yu, F., Ruberg, S.A., Chaffin, J.D., Goodwin, K.D., Tripathi, A., Sherman, D.H., Dick, G.J., 2022a. Metagenomic and Metatranscriptomic Insights into Population Diversity of Microcystis Blooms: Spatial and Temporal Dynamics of mcy Genotypes, Including a Partial Operon That Can Be Abundant and Expressed. Appl Environ Microbiol. https://doi.org/10.1128/AEM.02464-21

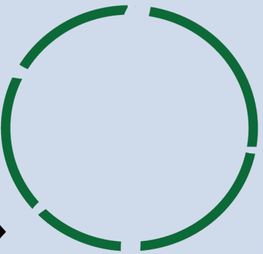
Yancey, C.E., Yu, F., Tripathi, A., Sherman, D.H., Dick, G.J., 2022b. Expression of Microcystis biosynthetic gene clusters in natural populations suggests temporally dynamic synthesis of novel and known secondary metabolites in western Lake Erie. bioRxiv 2022.10.12.511943. https://doi.org/10.1101/2022.10.12.511943

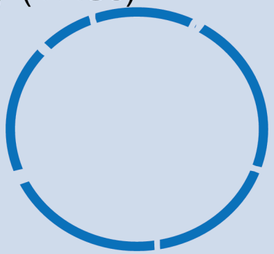
W. Lake Erie cyanoHABs



Cyanobacterial Biomass

Shotgun Sequencing



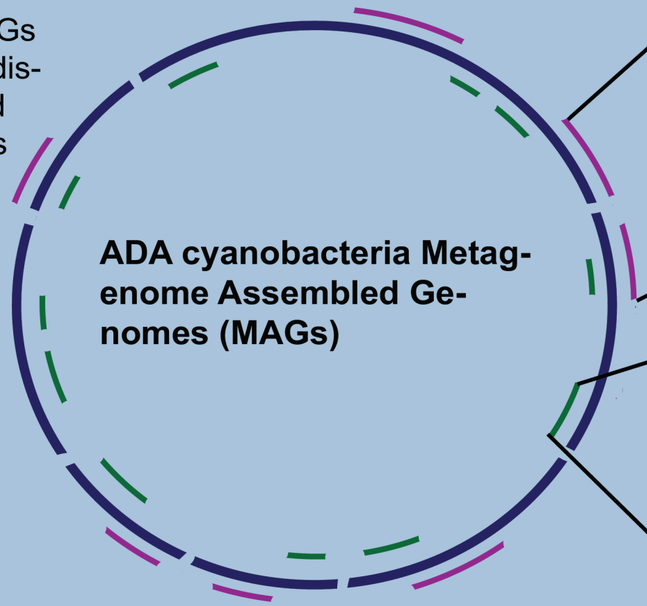


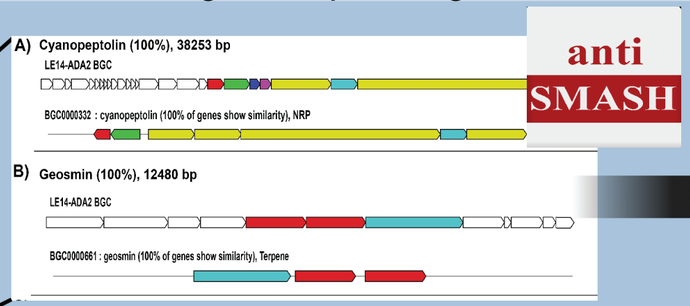
ADA cyanobacteria Metagenome Assembled Genomes (MAGs)

Genome Reconstruction

MAG	LE14-WE12	LE14-WE4
Date	4-Aug	29-Sep
Station	WE12	WE4
Size (Mb)	4.39	4.66
# of Contigs	225	259
N50	27386	25444
GC content	36.94	38.34
Completion	97.61	97.44
Redundancy	2.88	2.88
Contamination	0.78	0
Strain Het.	20	0

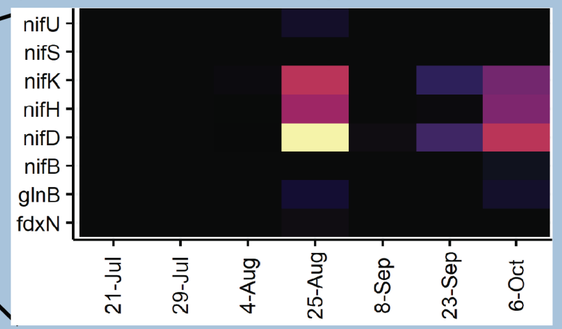
Two, unique MAGs recovered from distinct stations and sampling periods



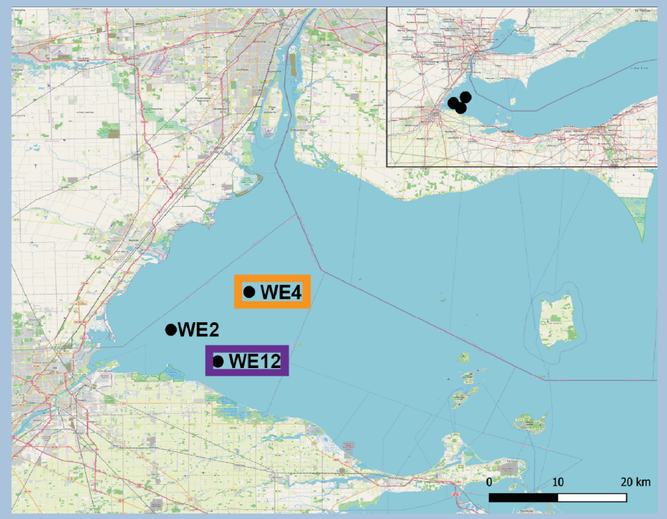


Biosynthetic Gene Cluster Mining and Identification

Potential for novel and known toxin synthesis



Differential N-fixation on spatio-temporal scales



Findings

Atleast two distinct species of ADA cyanobacteria are present in the 2014 western Lake Erie cyanoHAB.

These cyanobacteria have the potential to produce a variety of known and novel secondary metabolites

ADA cyanobacteria make up the majority of N-fixers in the 2014 cyanoHAB