Paul A. Den Uyl

Genetic Characterization of Saxitoxin Producing Cyanobacteria Associated with Western Lake Erie Harmful Algal Blooms

submitted in partial fulfillment of the requirements for the degree of
Master of Science in Earth and Environmental Sciences
Department of Earth and Environmental Sciences
The University of Michigan

Accepted by:

	Gregory J. Dick	2/29/24
Signature	Name	Date
	Jenan J Kharbush	4/3/24
Signature	Name	Date
	Julia Cole	May 13, 2024
Department Chair Signature	Name	Date

I hereby grant the University of Michigan, its heirs and assigns, the non-exclusive right to reproduce and distribute single copies of my thesis, in whole or in part, in any format. I represent and warrant to the University of Michigan that the thesis is an original work, does not infringe or violate any rights of others, and that I make these grants as the sole owner of the rights to my thesis. I understand that I will not receive royalties for any reproduction of this thesis.

- [x] Permission granted.
- [ ] Permission granted to copy after: ________________
- [ ] Permission declined.

Author Signature

Earth and Environmental Sciences
UNIVERSITY OF MICHIGAN

Abstract

Saxitoxins (STXs) are a group of closely related neurotoxins and among the most potent natural toxins known. While genes encoding STX biosynthesis have been observed in Lake Erie, the organism(s) responsible for producing STXs in the region have not been confirmed. We used metagenomic tools to identify a full suite of STX biosynthesis genes in a high-quality metagenome-assembled genome (MAG) from the Anabaena-Dolichospermum-Aphanizomenon (ADA) clade of cyanobacteria. The order and sequence of *sxt* genes suggest the Lake Erie MAG is capable of producing STX as well as other STX congeners. The absence of a *sxtX* gene suggests an inability to produce neoSTX, one of the most potent congeners of STX. We also recovered highly similar ADA MAGs that did not contain STX genes, implying gene loss or horizontal gene transfer. The full suite of STX biosynthesis genes was observed in metagenomic datasets across 13 of 123 unique sampling dates between 2014 and 2022, with occurrences spanning from July 6 to September 19. Our results also reveal trends in the increasing abundance of ADA MAGs during late-season blooms. Collectively, this study provides a foundation for understanding a potential new front of threats to Lake Erie water quality.

Introduction

Saxitoxins (STXs) are primarily sourced from freshwater cyanobacteria and marine dinoflagellates and have been extensively studied as the causative agent for paralytic shellfish poisoning (PSP) in marine environments. Research on STXs in freshwater ecosystems is limited compared to other cyanotoxins like microcystin. Despite no reported human fatalities from STX exposure in freshwater, incidents of sheep mortality due to STX-contaminated cyanobacterial blooms highlight potential risks (Negri et al., 1995; Smith, 2000). Given the significance of many freshwater bodies as drinking water sources, recognizing STX presence during water treatment processes is imperative (Orr et al., 2004; Westrick, 2008). In freshwater, instances of STX exposure have been linked to recreational activities, with exposure able to occur through inhalation and skin contact (Rapala et al., 2005). STXs can also bioaccumulate in freshwater fish and shellfish, posing concerns for human consumption (J. P. Berry & Lind, 2010; da Silva et al., 2011; Galvão et al., 2009; Ibelings & Chorus, 2007; Negri & Jones, 1995). The proliferation of cyanobacteria harmful algal blooms (cyanoHABs) due to climate change further amplifies concerns about STX presence in freshwater environments (Paerl & Huisman, 2009).

More than 50 STX congeners have been identified, including non-sulfated (STX and neoSTX), mono-sulfated (gonyautoxins (GTXs 1-6)), dicarbamoyl (dcSTX, dc-neoSTX, dcGTXs 1-4) and di-sulfated (C1-4) forms (Christensen & Khan, 2020). STX congeners display variable toxicity, with STX, neoSTX, and gonyautoxins (GTX) consistently identified as the most potent (Ballot et al., 2016; Selwood et al., 2017; Suarez-Isla, 2014; United Nations Environment Programme, 1984). Due to its extreme toxicity, STX is classified as a bioweapon under the Chemical Weapons Convention (Al-Tebrineh et al., 2010; Sierra & Martínez-Álvarez, 2020).

Nine genes in cyanobacteria (*sxtA*, *sxtG*, *sxtB, sxtD*, *sxtS, sxtU*, *sxtI*, *sxtH*, and *sxtT*) have been proposed to play a direct role in STX biosynthesis (Kellmann et al., 2008). Additional genes have been hypothesized to alter STX into distinct congeners (*sxtL*, *sxtN*, *sxtX*) (Mihali et al., 2009). The *sxtA* gene takes a central role in initiating STX biosynthesis operon through catalyzing a unique Claisen condensation reaction. The crucial role played by the *sxtA* gene in

STX biosynthesis makes it an effective qPCR target for accurately quantifying STX producers in the environment (Al-Tebrineh et al., 2010; Pearson et al., 2010).

Western Lake Erie (WLE) is especially susceptible to toxic cyanoHABs, including those with the potential to produce saxitoxin (Watson et al., 2016). This vulnerability is underscored by the presence of several municipal water intakes, notably including Toledo, OH, where a nearly three day "do not drink" advisory was issued in 2014 as a result of microcystins persisting in finished water (Steffen et al., 2017; Wynne & Stumpf, 2015). The 2014 Toledo Water Crisis left over 400,000 residents without access to safe tap water. Additionally, the area relies heavily on recreation, tourism, and a significant commercial fishing industry, potentially heightening the exposure risks to cyanoHABs. While considerable efforts have been dedicated to monitoring and understanding the development, movement, and toxicity of cyanoHABs in WLE, research into cyanotoxins beyond microcystin in this region remains notably understudied. A recent investigation into the diversity of cyanobacterial biosynthetic gene clusters in WLE, which identified the presence and expression of cyanotoxins and cyanopeptides beyond microcystins, highlights the need for expanded research on this topic (Yancey et al., 2023).

Use of ELISA to identify saxitoxin and qPCR to detect *sxtA* genes in Lake Erie has confirmed the presence of both the toxin and key genetic markers responsible for STX production (Chaffin et al., 2019; Nauman et al., 2024; Ohio Environmental Protection Agency, 2022). Despite preliminary evidence from Chaffin et al. (2019) suggesting a potential association between the presence of the *sxtA* gene and the cyanobacterial genus *Dolichospermum* based on flowCAM observations, a confirmed source of STXs in WLE or the Great Lakes has not been established. STXs have been found to be produced by several cyanobacterial genus present in WLE. Potential STX producers include *Microcystis*, the dominant genus of WLE cyanoHABs, along with members of the *Anabaena*, *Dolichospermum*, and *Aphanizomenon* (ADA) cyanobacteria clade, as well as *Cylindrospermopsis*, and *Lyngbya* (Conroy et al., 2007; Naknaen et al., 2021; Wiese et al., 2010). STX genes have not been identified in any ADA-clade or *Microcystis* genomes characterized in the Great Lakes (Sheik et al., 2022; Yancey et al., 2023). The potency of STX paired with the potential for production by multiple, currently unidentified, species of cyanobacteria creates a situation of concern for cyanoHABs in WLE.

The existence of extensive monitoring efforts in WLE offers opportunities to investigate the presence of STXs. In this study, metagenomic sequence data collected from both routine and opportunistic cyanoHABs monitoring efforts (Boegehold et al., 2023) was used to assess the potential for STX production and its distribution across WLE. The DNA sequence database used in this study, the Great Lakes Atlas of Multi-omics Research (GLAMR, greatlakesomics.org), included more than 500 samples of sequenced metagenomic data collected in WLE between 2014 and 2022 (Chaffin et al., 2021; Colleen E. Yancey et al., 2022; Den Uyl et al., 2022; Koeppel et al., 2022; McKindles Katelyn M. et al., 2020). Incorporating GLAMR into this research provides a robust methodology for pinpointing neurotoxin-producing organisms and understanding their spatial and temporal distribution. This framework is essential for advancing knowledge and awareness of the associated risks of STXs in WLE.

Results

Identification of a saxitoxin (STX) source and additional ADA MAGs

Within GLAMR, *de novo* assembly of metagenomic reads from 570 WLE samples produced a total of 1,844 metagenome-assembled genomes (MAGs) with greater than or equal to

90% completeness and less than 10% contamination. Dereplication of MAGs using a 98% average nucleotide identity (ANI) threshold resulted in 850 representative MAGs generated. After dereplication, a single high-quality representative MAG belonging to the cyanobacterial ADA clade was identified possessing a full suite of biosynthetic genes for STX production (STX+). This MAG, designated LE20-WE8, was generated from a sample taken at the WLE routine monitoring station WE8 on July 6, 2020. LE20-WE8, classified as the species *Dolichospermum sp000312705*, has high completion (96.78%) and low contamination (2.56%) (Table 1) and is the only representative MAG in this study possessing the biosynthetic potential to produce STXs. Ten additional representative MAGs from the GLAMR database were identified in WLE belonging to the ADA clade (Table 1). Genes specific to STX biosynthesis were not detected (STX-) in these ten MAGs. Representative ADA MAGs were also examined for the presence of the cyanotoxin genes known to produce microcystin (*mcyE gene*), anatoxin A (*anaC*, *anaF*), cylindrospermopsin (*cyrA*), guanitoxin (*gntF*), and β-Methylamino-L-alanine or BMAA (*sbnA-B*) using blastn. None of these additional cyanotoxin genes were detected.

MAG	STX Potential	Date of collection	Latitude	Longitude	Station	GTDB-Tk species classification	Completeness	Contamination	Strain heterogeneity	N50	GC	Number of contigs	Size mb
LE20-WE8	+	6-Jul-20	41.8192	-83.3592	WE8	*Dolichospermum sp000312705*	96.78	2.56	58.82	7007	0.39	903	4.577
LE17-WE2	–	7-Aug-17	41.7633	-83.3303	WE2	*Dolichospermum sp000312705*	93.05	2.83	60	18103	0.38	348	4.338
LE17-WE12	–	11-Sep-17	41.7032	-83.254	WE12	*Dolichospermum circinale*	90.54	6.81	62.5	2791	0.38	1678	4.108
LE20-WE12	–	5-Oct-20	41.704	-83.2552	WE12	*Dolichospermum circinale*	95.2	7.85	82	3051	0.37	1664	4.352
LE21-WE12	–	21-Sep-21	41.7041	-83.2529	WE12	*Dolichospermum circinale*	90.7	1.08	85.71	11487	0.38	414	3.569
LE22-WE12	–	30-Sep-22	41.7058	-83.2556	WE12	*Dolichospermum circinale*	99.55	0.22	100	24272	0.37	342	4.773
LE21-ER30	–	7-Apr-21	42.4295	-81.2044	ER30	*Dolichospermum flosaquae*	98.22	0.22	50	33828	0.37	190	3.905
LE16-WE8	–	12-Sep-16	41.8322	-83.3589	WE8	*Dolichospermum flosaquae*	90.06	4.83	40.74	3373	0.37	1734	3.787
LE19-WE12	–	3-Sep-19	41.7034	-83.2558	WE12	*Dolichospermum flosaquae*	98.91	0.78	50	20287	0.37	293	4.238
LE20-WE2-Aug	–	10-Aug-20	41.7614	-83.3329	WE2	*Cuspidothrix issatschenkoi*	97.33	4.19	80.95	4975	0.38	1200	4.345
LE20-WE2-Sep	–	9-Sep-20	41.7628	-83.3297	WE2	*Cuspidothrix issatschenkoi*	92.37	1.85	92.31	8046	0.38	598	3.602

Table 1. Metadata and summary statistics of representative ADA MAGs obtained from WLE samples using the GLAMR database (greatlakesomics.org).

Both LE20-WE8 (STX+) and LE17-WE2 (STX-) clustered into a distinct ADA branch that contains several strains isolated from the Great Lakes, including *Dolichospermum* sp. SB001 (Lake Superior metagenome), *Anabaena* sp. AL09 (Lake Huron culture), *Anabaena* sp. LE011-02 (Lake Erie culture) and LE14-WE4 (Lake Erie metagenome) (Driscoll et al., 2018; Sheik et al., 2022; Yancey et al., 2023) (Fig. 1). GTDBTK annotations designate individuals within this branch as *Dolichospermum sp000312705*, whereas earlier research has sometimes used the name *Dolichospermum lemmermannii* (Sheik et al., 2022). Additional ADA MAGs recovered here also fell into branches with previously identified species. LE20-WE2-Aug and LE20-WE2-Sep clustered with *Cuspidothrix issatschenoki*. LE16-WE8, LE19-WE12, and LE21-ER30 were in a branch with *Dolichospermum flosaquae* and are very closely related to LE14-WE12 (97.9% - 99.6% ANI) which came from WLE in August 2014 (Yancey et al., 2023). The largest diversity of MAGs identified in this study (LE17-WE12, LE20-WE12, LE21-WE12, and LE22-WE12) clustered in a branch with *Dolichospermum circinale*. These MAGs were closely related (96.8% - 97.1% ANI) to an STX producer, *D. circinale* AWQC131C isolated from Lake Cargelligo, NSW Australia (Llewellyn et al., 2001). There does not appear to be any relationship between the genetic potential to produce STXs and strain placement within the phylogenomic tree of the ADA clade (Fig. 1).

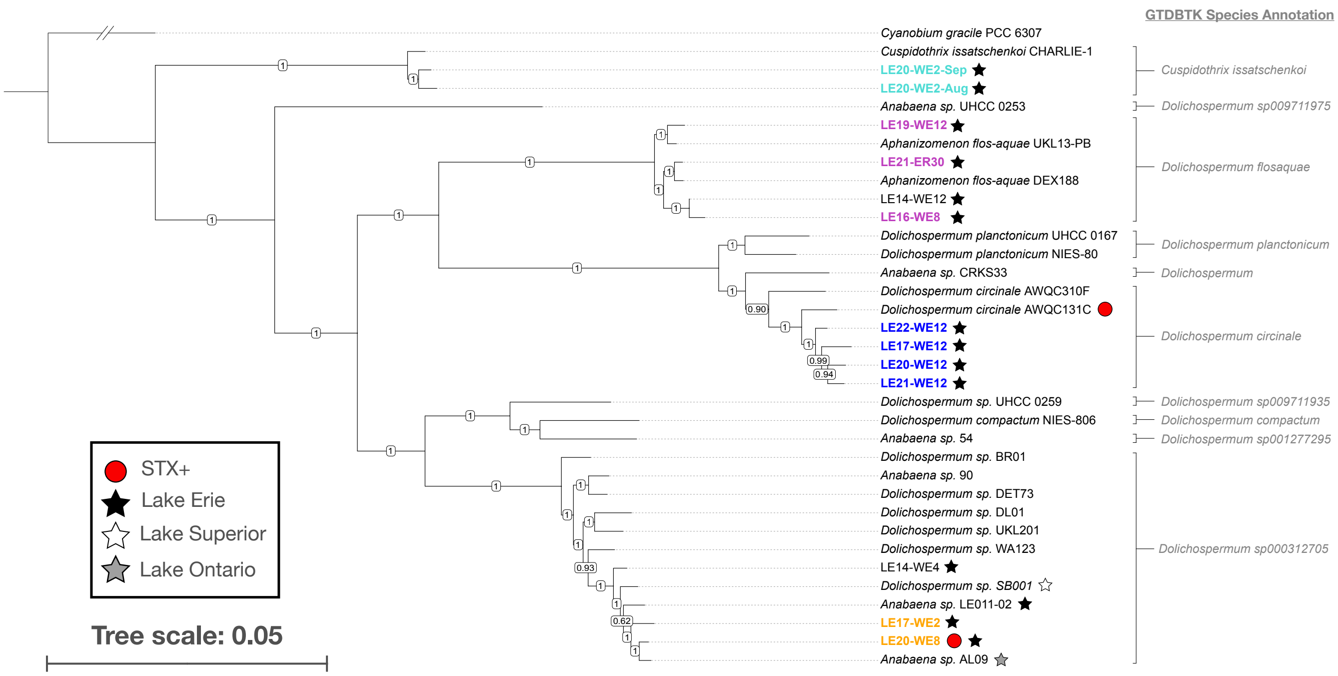

Figure 1. Phylogenomic tree of representative ADA clade members constructed using 251 housekeeping genes and a maximum likelihood approach through the GToTree workflow. The dataset incorporated ADA genomes from prior investigations (Dreher et al., 2021; Österholm et al., 2020; Sheik et al., 2022; Yancey et al., 2023). MAGs obtained from this study are highlighted in colored and bolded text. Colors reflect phylogeny and correspond to Figure 4 to aid comparison. Stars indicate ADA strains from the Great Lakes, sourced from: Lake Erie (black), Lake Superior (white), and Lake Ontario (grey). Red circles identify STX+ strains. Numbers on branches are bootstrap values, support metrics that reflect the robustness of relationships among taxa based on resampling.

24 additional MAGs from the ADA clade were identified using a stricter ANI threshold for dereplication compared to the criteria used for establishing representative MAGs (99% ANI instead of 98%) (Fig. S1). These 24 MAGs were also examined for the presence of cyanotoxin-encoding genes, but none were identified. Despite appearing directly adjacent to LE20-WE8 (STX+) on the phylogenomic tree, the MAG samp_2047_concoct_1911 lacked genes associated with STX production (STX-). We identified that the *sxt* genes were instead assigned to an *alphaproteobacteria* MAG (samp_2047_concoct_1969) during the binning process by CONCOCT. *sxt* genes were ultimately determined not to belong to the *alphaproteobacteria* bin after manual QC and genome refinement via Anvi'o (Eren et al., 2015). There is no record of *alphaproteobacteria* being able to produce STXs, further suggesting the contigs were incorrectly assigned. These contigs containing *sxt* genes were annotated to belong to *Dolichospermum* by kaiju. Binning by MaxBin also assigned these *sxt* genes to a MAG annotated as

Dolichospermum sp000312705, which matches the species annotation of LE20-WE8 (STX+). We conclude *sxt* genes contained within the MAG samp_2047_concoct_1969 were misassigned during the generation of bins using the automated binner CONCOCT. To confirm that STX biosynthesis genes weren't misassigned for other MAGs that were closely related to STX+ organisms, such as LE17-WE2 (STX-), metagenomic sample reads were examined using blastn for *sxt* genes. There were no *sxt* genes present in the sample reads from which LE17-WE2 (STX-) was generated. Further analysis of MAGs will be focused only on representatives determined using the ANI threshold of 98%, referred to hereafter as representative MAGs (n =11, Table 1).

STX biosynthesis genes

The STX biosynthesis operon identified in LE20-WE8 was divided across three contigs (Fig. 2). Each contig contained *sxt* genes in nearly the same order and orientation as the STX biosynthesis operon identified in *Aphanizomenon* sp. NH-5, a culture isolated from a small pond near Durham, New Hampshire (Mihali et al., 2009). The most notable differences between the gene clusters identified in this study and *Aphan.* sp. NH-5 are the absence of *sxtX, sxtW*, and a partial length *sxtD* gene in LE20-WE8.

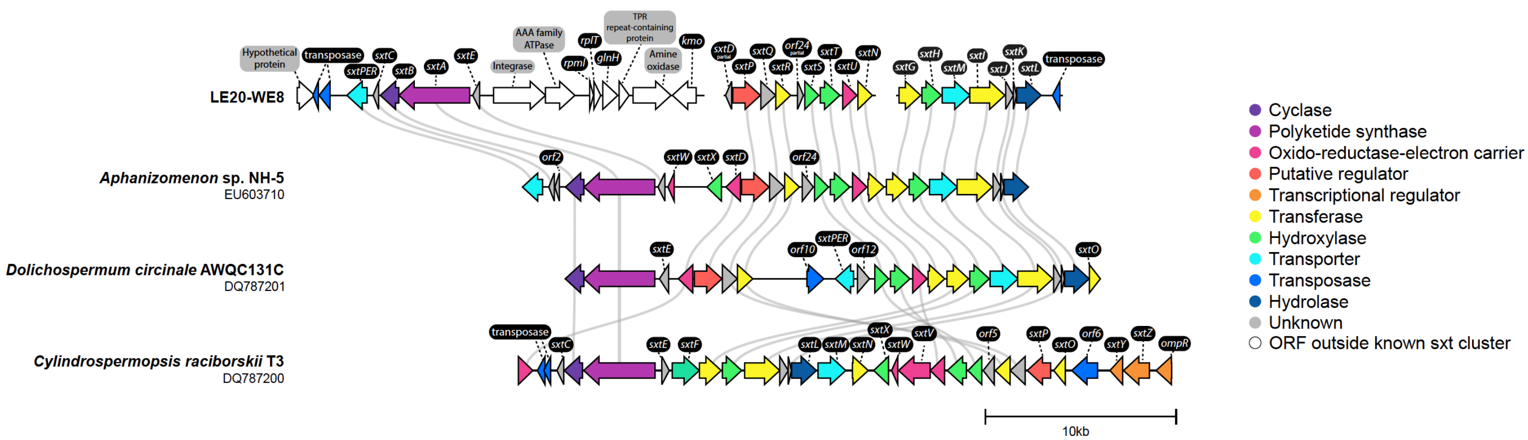

Figure 2. Comparison of STX gene cluster structure between LE20-WE8, *Aphanizomenon* sp. NH-5, *Dolichospermum circinale* AWQC131C, and *Cylinderospermopsis raciborskii* T3. GenBank accession numbers are listed below organism names. Gray lines indicate homologous genes sharing ≥80% nucleotide sequence identity. Gray labels show CoDing Sequence (CDS) annotations. Figure adapted from Mihali et al., 2009.

Only 156bp of the 759bp *sxtD* gene sequence from *Aphan.* sp. NH-5 aligned with LE20-WE8 contigs (99.4% ID; 21% coverage) via blastn. An unassigned contig (402bp) from sample LE20-WE8 showcased high sequence similarity (blastn 99.0% ID; 100% coverage) to a distinct section of the *sxtD Aphan.* sp. NH-5 reference. The length of this contig (402bp) makes it too short to reliably bin using automated methods. Considering this additional fragment and the partial *sxtD* gene identified in LE20-WE8, 74% of the *sxtD* gene is accounted for with greater than 99% blastn ID.

To examine the distribution and diversity of STX biosynthesis genes across WLE, metagenomic reads were mapped back to the assembled *sxt* genes from the three LE20-WE8 contigs (Fig. S2). Metagenomic read mapping resulted in 22 samples having >60% coverage of each of the three reference contig sequences (Table S1). All results with >60% coverage of each of the three reference sequences had at least 95% nucleotide identity between each mapped alignment and its complimentary reference sequence (Fig. S3-S5). The 22 samples with substantial read mapping alignment to *sxt* genes were collected across 13 unique sampling dates, encompassing a seasonal range from as early as July 6 to as late as September 19, spanning years 2016 to 2021.

ADA MAG metabolic pathways

KEGGDecoder was used to annotate metabolic pathways in the 11 representative ADA MAGs. Notable pathways of LE20-WE8 include ferrous iron transporter *feoB*, which was only found in the two representative *D. sp000312705* MAGs (Fig. 3). The pathway for synthesis of the carotenoid zeaxanthin diglucoside was also possessed by *D. sp000312705* MAGs, LE20-WE8 (STX+), LE17-WE2 (STX-) and one *C. issatschenoki* MAG, LE20-WE2-Aug. Nitrogen fixation pathways were present in KEGGDecoder annotations for all *Dolichospermum* MAGs, ranging between 33-66% complete, but were absent in *Cuspidothrix* MAGs. The nitrogen fixation pathway for LE20-WE8 was identified by KEGGDecoder as 66% complete, with two out of the three enzymes required for nitrogen fixation, *nifK* (K02591) and *nifD* (K02586), present. Genome annotation using Bakta revealed the remaining *nif* gene, *nifH* (K02588) in LE20-WE8. Additional *nifH*, *D*, *K* genes, beyond those annotated in KEGGDecoder, were also identified in six additional *Dolichospermum* MAGs through Bakta annotation (Table S3). No additional *nif* genes were identified in *Cuspidothrix* MAGs by Bakta.

All representative MAGs exhibited pathways facilitating the uptake of additional nutrients via urea, phosphate, and phosphonate transporters. Urease subunits were identified in Bakta annotations for all eleven representative MAGs, including the alpha, beta, and gamma subunits in LE20-WE8 (Table S3). Additionally, all representative MAGs featured pathways involved in sulfolipid biosynthesis, as well as the production of carotenoids nostoxanthin, myxoxanthophylls, and astaxanthin (Fig. 3).

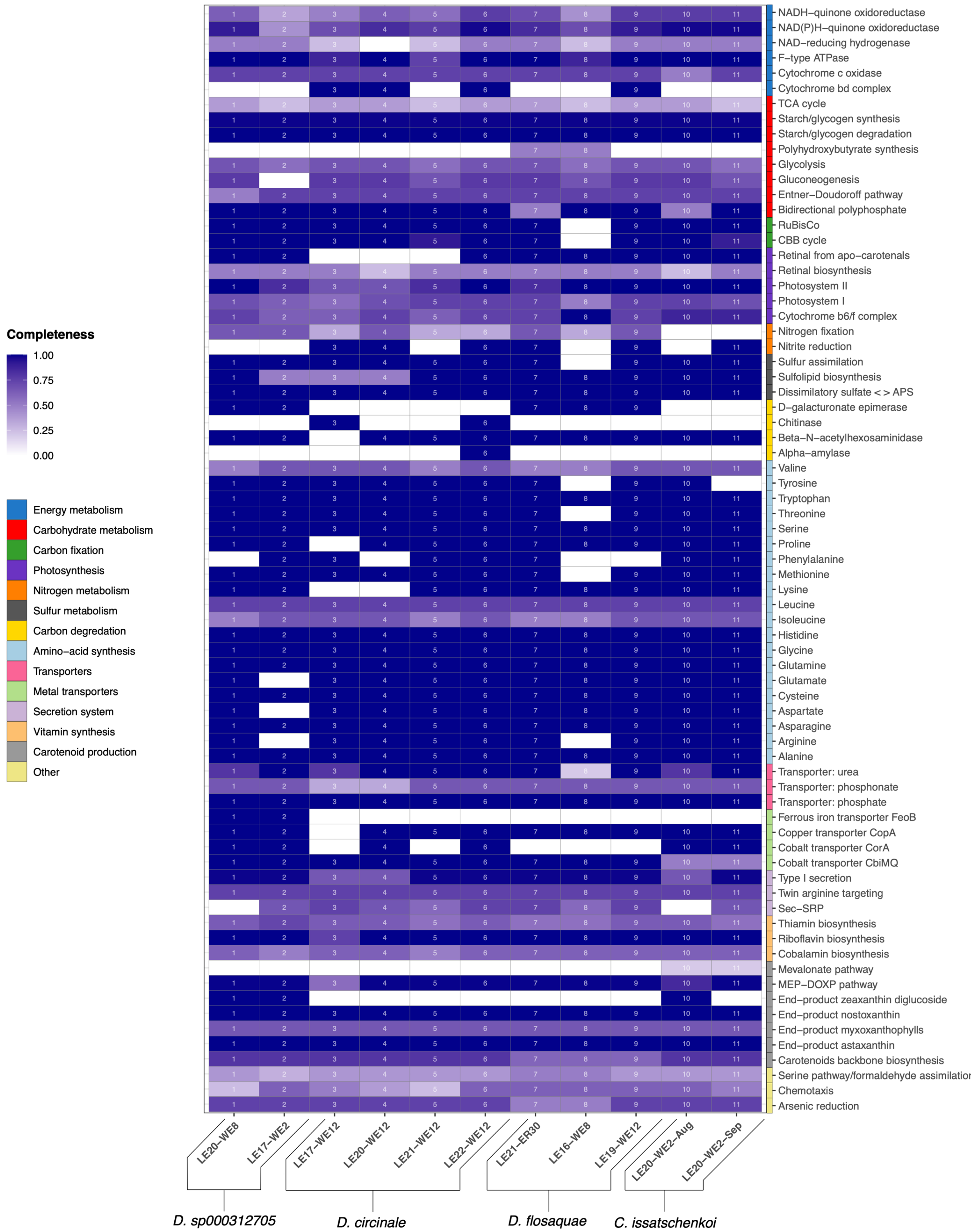

Figure 3. Heatmap visualization of KEGGDecoder functional analysis for representative MAGs

Temporal and spatial abundance of ADA MAGs

Metagenomic read mapping to representative ADA MAGs allowed for the calculation of relative organism abundance across stations and dates of sample collection (Fig. 4). Given the difficulty of short read mapping to discriminate different strains within the same species, reads per kilobase million (RPKM) values were combined for MAGs sharing the same species annotation. Across the sampling season, ADA MAGs were consistently identified in metagenomes, indicating a persistent presence. However, their abundance and distribution displayed substantial variability between years. In years 2014, 2017, and 2018, ADA MAGs were detected for only 4 weeks, highlighting the dynamic nature of their occurrence. Multiple species often co-occurred in the same sample (37% of samples). When just one species was detected, it was most often *D. circinale* (42% of samples), followed by *D. flosaquae* and *D. sp000312705* (15% and 5% of samples, respectively).

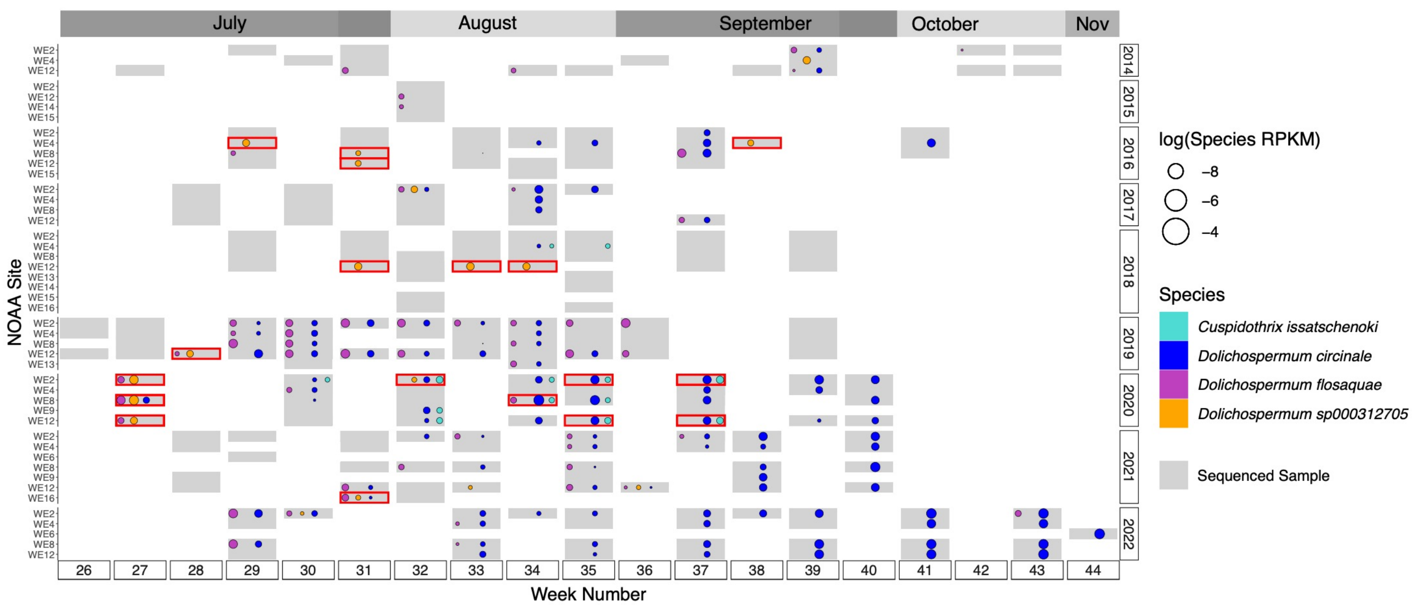

Figure 4. Reads per kilobase million (RPKM) determined from metagenomic read mapping of MAGs. Dot size determined from summing RPKM of MAGs that belong to the same species. WE20-WE8 (STX+) and LE17-WE2 (STX-) = *D. sp000312705*; LE17-WE12, LE20-WE12, LE21-WE12, and LE22-WE12 = *D. circinale*; LE21-ER30, LE16-WE8, and LE19-WE12 = *D. flosaquae*; and LE20-WE2-Aug and LE20-WE2-Sep = *C. issatschenoki*. Only the highest RPKM value is shown for each unique species and week combination. Red outlines indicate the presence of a STX biosynthesis operon determined through read mapping to assembled *sxt* genes from LE20-WE8 contigs (≥60% coverage and ≥95% nucleotide identity) (Fig. S3-S5). NOAA Site based off of routine WLE monitoring stations (Boegehold et al., 2023) and expanded to include samples collected within a 5 km radius of each station.

The presence of the species *D. sp000312705* is particularly significant due to the identification of the saxitoxin-producing capability of MAG LE20-WE08, a species member. The maximum RPKM belonging to species *D. sp000312705* reached 6.93e-03 on July 6, 2020, at WE8 (Table S2). This was the same sample that representative MAG LE20-WE8 was sourced. *D. sp000312705* prominently occurred at station WE12, being identified at that site on 8 out of 15 of the species detection dates. Dates of *D. sp000312705* observations are concentrated in the months of July (n=5), August (n=7), and September (n=3). In 2014, 2016, and 2018, whenever *D. sp000312705* appeared, it was consistently the only ADA species from this study identified in a sample (n = 10). For other years where *D. sp000312705* appeared (2017, and 2019-2022) it co-occurs with other ADA species in 10 of 11 samples.

D. sp000312705 presence corresponds well with the occurrence of STX biosynthesis genes determined by metagenomic read mapping (15 of 22 samples). Species *D. circinale* and *C. issatschenoki* were both present in the seven instances where STX biosynthesis genes were identified through read mapping and *D. sp000312705* was not. *D. sp000312705* also occurs where there are no significant STX biosynthesis genes mapped on five occasions, including WE2 in week 32 (Aug, 7th) of 2017, the sample that representative MAG LE17-WE2 (STX-) was sourced.

D. circinale stood out as the species exhibiting the highest relative organism abundance in a sample, reaching 0.207 RPKM within a late-season cyanobacterial bloom in early November of 2022 (week 44, Figure 4). Notably, the same sample also recorded the highest individual MAG abundance, corresponding to LE22-WE12 at 0.167 RPKM. The nine highest species RPKM values identified were from late season 2022 cyanobacterial bloom(s) and belonged to *D. circinale*. LE22-WE12 was the most abundant *D. circinale* MAG in each of these occurrences. *D. circinale* appears to have an increased presence in years 2020-2022, particularly after week 37 each season. *D. circinale* is present in 4 of 12 sampling dates collected after week 37 between years 2014-2019, while it is present in 16 of 16 sampling dates after week 37 between 2020-2022.

The least frequently detected species from metagenomic read mapping was *C. issatschenoki*, observed only in samples collected during 2018 and 2020 sampling years. *C. issatschenoki* did emerge as the second most frequently observed species by date in 2020 (5 of 10 sampling dates, 13 of 49 samples). *D. flosaquae* showcased a variable presence in WLE. *D. flosaquae* stood out as the most frequently observed species in 2019, by both date and sample count (15 of 22 sampling dates, 37 of 120 samples), while it was absent in all samples collected in 2018.

Discussion

To identify the organism(s) responsible for producing potential STXs, a potent group of neurotoxins (Wiese et al., 2010) with supporting biochemical (Ohio Environmental Protection Agency, 2022), environmental (Watson et al., 2016), and genetic evidence (Chaffin et al., 2019; Nauman et al., 2024) in Lake Erie, we analyzed metagenomes from 570 samples collected across WLE from 2014-2022. We identified one MAG containing a full suite of STX biosynthesis genes, attributing it to *D. sp000312705* through phylogenomic analysis. Beyond identification of organisms responsible for STX production, these results offer insights into potential biosynthetic products of the identified *sxt* genes. Additionally, the metabolic potential and environmental distribution of this STX-producing organism and related non-STX-producing organisms is discussed in more detail below.

STX biosynthesis genes – potential STX products

Comparing *sxt* genes identified in the LE20-WE8 MAG obtained from WLE to the *sxt* genes of closely related organisms with characterized STX products provides crucial insights into STX biosynthesis and potential congeners produced. The conserved order and sequence similarity observed between contigs with *sxt* genes in MAG LE20-WE8 and the STX biosynthesis operon of *Aphan.* sp. NH-5 initially suggests a capacity to produce similar products

(Fig. 2, Table S4). *Aphan.* sp. NH-5's toxin profile is primarily composed of STX and neoSTX with minor constituents also identified as 12,12dido-dcSTX, 12βdo-doSTX, dcNEO, dcSTX, 12βdo-dcSTX (LWTX-4), 12αdo-doSTX, doSTX, M4, and 12αdo-dcSTX congeners (D'Agostino et al., 2019). neoSTX was identified to compose approximately 85% of the quantified toxin profile of *Aphan.* sp. NH-5 in culture (D'Agostino et al., 2019; Mahmood & Carmichael, 1986). STX and neoSTX have been characterized to have similar toxicity and be amongst the most toxic saxitoxin congeners (2008; Selwood et al., 2017; Suarez-Isla, 2014; Wiese et al., 2010).

A closer investigation into the *sxt* genes identified in LE20-WE8 MAG suggests that it is likely not capable of neoSTX production. *sxtX*, the gene putatively responsible for N-1 hydroxylation of STX, converting STX to neoSTX (Kellmann et al., 2008; Mahmood & Carmichael, 1986; Mihali et al., 2009; Soto-Liebe et al., 2010), is absent in LE20-WE8**.** *sxtX* is also absent in previously characterized STX biosynthesis gene clusters belonging to *Raphidiopsis brookii* D9*, D. circinale* AWQC131C, and ACMB13 (formerly referred to as species *Anabaena circinalis*), all of which have been shown to not synthesize any STX variants featuring a hydroxyl group at N-1, including neoSTX (Mihali et al., 2009; Soto-Liebe et al., 2010). Subsequently, the toxin profile of the Great Lakes MAG LE20-WE8 may be suggested to consist of STX and other congeners without a hydroxyl group at N-1 observed in *Aphan.* sp. NH-5's toxin profile (12,12dido-dcSTX, 12βdo-doSTX, dcSTX, 12βdo-dcSTX (LWTX-4), 12αdo-doSTX, doSTX, M4, and 12αdo-dcSTX).

The lack of a *sxtW* gene in MAG LE20-WE8 is also shared by *R. brookii* D9 and *D. circinale* AWQC131C (Mihali et al., 2009; Stucken et al., 2010). The putative function of *sxtW* is thought to be similar to a ferredoxin and perform electron transport required for hydroxylation of products encoded by *sxtT* and *sxtH*. Mihali *et al.* (2009) suggested an endogenous ferredoxin may compensate for the lost function from a missing *sxtW* gene, but no ferredoxin-like alternative was identified in LE20-WE8. Even without *sxtW*, *R. brookii* D9 and *D. circinale* AWQC131C and have been found to have toxin profiles that include STX and other STX, GTX, and C-toxin congeners (D'Agostino et al., 2019; Llewellyn et al., 2001; Soto-Liebe et al., 2010).

The direct involvement of *sxtD* in STX synthesis raises questions about its partial presence in LE20-WE8. *sxtD* encodes a putative desaturase and is suggested to introduce double bonds into carbon-carbon bonds within the STX biosynthetic pathway. It plays a crucial role alongside *sxtS* and *sxtU*, directly contributing to STX synthesis (Kellmann et al., 2008; Mihali et al., 2009; Pearson et al., 2010). The potential absence of certain genes, including *sxtD*, could be attributed to artifacts in metagenomic assembly and binning methods. To address this, manual analyses were conducted scrutinizing the presence of *sxtD*. It is asserted that the occurrence of *sxtD* has been adequately addressed through these analyses, particularly after identifying an additional partial *sxtD* fragment in an unassigned contig post-binning. This manual examination significantly bolsters confidence in affirming the presence of a *sxtD* gene in the MAG LE20-WE8.

With over 50 documented STX congeners and an incomplete understanding of the relationship between biosynthetic genes and products, there exists a pressing need for further investigation. This is crucial to unveil the complete range of STX congeners produced by LE20-WE8. Toxicity levels among these congeners vary significantly, with STX, neoSTX, and gonyautoxins (GTX) identified as the most potent variants (Wiese et al., 2010). Metabolic transformations of mixtures of STX congeners have also been observed in vectors like shellfish and humans (Munday & Reeve, 2013; Suarez-Isla, 2014). These transformations often escalate toxicity rather than diminish it, emphasizing the critical importance of comprehensively understanding toxin profiles.

Metabolic potential and distribution

Exploring metabolic pathways of representative ADA MAGs can enhance our comprehension of cyanoHABs members and their potential interactions within WLE (Fig. 3, Fig. S6). Notably, the presence of nitrogen fixation genes in all nine *Dolichospermum* MAGs builds upon previous observations of nitrogen-fixing abilities within the WLE ADA clade (Yancey et al., 2023). This nitrogen-fixing potential in *Dolichospermum* is crucial for late-season or secondary blooms, potentially allowing *Dolichospermum* to thrive after *Microcystis* has depleted much of the available nitrogen (Chaffin, 2013; Michalak et al., 2013; Takamura et al., 1987; Wang et al., 2021). The occurrence of cyanoHABs with a large presence of ADA clade cyanobacteria, specifically *D. circinale*, are being increasingly observed later in the year (Fig. 4). Fortunately, these late-season blooms do not appear to possess the genetic potential to produce STXs. Further research is imperative to understand the evolving dynamics of late-season blooms in Lake Erie.

Mechanisms to compensate for phosphorus limitation are also present in representative ADA-clade MAGs. The presence of sulfolipid biosynthesis genes is identified in all eleven MAGs (Fig. 4). Substitution of sulfolipids for phospholipids in response to phosphorus limitation has been identified in *Microcystis* cultures, mesocosms, and *Microcystis*-dominated cyanoHAB samples from WLE (Martin et al., 2023). These non-phosphorus-containing lipids (sulfolipids) reduce the burden of phosphorus in cellular membrane construction as they can be used to substitute phosphorus containing phosphatidylglycerol.

The presence of urease genes and at least partial urea, phosphate, and phosphonate transporters across all representative MAGs highlights additional mechanisms for external nutrient acquisition in WLE ADA-clade cyanobacteria, including the use of organic nutrients. These pathways may be closely tied to agricultural practices in the Lake Erie watershed. Around 60% of land use in the Lake Erie watershed is allocated for agriculture, leading to significant contributions of nitrogen from prevalent inorganic fertilizers, commonly in the form of phosphate (Barnard et al., 2021; Mohamed et al., 2019). Increased use of organic urea in fertilizers in recent years further adds to the high nutrient setting in WLE (Belisle et al., 2016).

Transporters for phosphonate, conventionally perceived as a less bioavailable organic phosphorus form, also potentially serves as a nutrient source for ADA-clade cyanobacteria (Horsman & Zechel, 2017; McGrath et al., 2013). Glyphosate, commonly referred to as RoundUp Ready®, is a phosphonate that has seen increased use in recent decades and could promote ADA-clade cyanobacteria including strains capable of saxitoxin production (Spiese et al., 2023). The utilization of organic phosphonate by ADA-clade cyanobacteria is also relevant to the ongoing phosphorus management efforts by the United States and Canada, which rely on reduction in total phosphorus and soluble reactive phosphorus (SRP) and do not include phosphonate (Maccoux et al., 2016).

Comparing metabolic pathways in MAGs also demonstrated several potential factors governing spatial and temporal distribution of STX. The pathway for ferrous iron Fe(II) transporter *FeoB* is unique to *D. sp000312705* MAGs and may also be relevant to nitrogen fixation. The production of nitrogenase, the enzyme accountable for nitrogen fixation, relies heavily on significant quantities of iron (Larson et al., 2018). Therefore, the potential need for iron uptake to provide for nitrogenase function may support nitrogen fixation in low nitrogen environments. STXs are also nitrogen rich and demand substantial nitrogen. It has been suggested that STX production in cyanobacteria may be regulated by environmental factors such as nitrogen and phosphate (Dias et al., 2002). The exclusive presence of the *FeoB* transporter in *D. sp000312705*, the species associated with LE20-WE8 (STX+), might be attributed to nitrogen demands for STX production. This transporter aids in the uptake of exogenous Fe(II), potentially supplementing the nitrogenase requirements for nitrogen fixation, crucial in supplying the necessary nitrogen for STX production.

Other notable metabolic pathways include zeaxanthin diglucoside synthesis unique to *D. sp000312705* MAGs LE20-WE8, LE17-WE2 and *C. issatschenoki* MAG, LE20-WE2-Aug. Zeaxanthin diglucoside is a carotenoid synthesized as a protective response to environmental stressors, including high light intensity and oxidative stress commonly experienced in cyanoHAB conditions (Hellweger et al., 2022; Latifi et al., 2009; Smith et al., 2022). All MAGs also have the pathways for additional carotenoids including nostoxanthin, myxoxanthophylls, and astaxanthin. Some missing pathways may be due to incomplete MAGs especially those with lower completeness such as LE17-WE12, LE21-WE12, and LE16-WE8 (~90% complete; Table 1).

Impact of STX source identification and *sxt* gene distribution

Identification of an STX producer in WLE emphasizes the current gaps in knowledge of toxin production within the region's cyanoHABs. This limitation significantly impacts the capacity to provide comprehensive guidance for general monitoring or risk assessments in the area. The characterization of STX biosynthesis genes in LE20-WE8 signify a critical advancement in ongoing efforts aimed at understanding and tracking potential cyanotoxin producers in WLE and the broader Great Lakes.

Specific observations of concern include the prevalence of LE20-WE8 around monitoring station WE12. This station is adjacent to the city of Toledo’s water intake that supplies water to >400,000 residents. The occurrence of a full suite of STX biosynthesis genes near WE12 occurs across all summer months (July – September) and multiple years (2016, 2018 – 2020), suggesting reoccurring STX presence in the area. This occurrence of STX biosynthesis genes is also widespread across the western basin of Lake Erie, occurring at five of eight routine NOAA WLE monitoring stations. Four of these five stations (WE2, WE8, WE12, and WE16) can be described as nearshore, but are outside of more direct input from the Maumee River such as stations WE6 and WE9. The fifth NOAA station, WE4, has been characterized as offshore and receives more significant influence from the Detroit River along with station WE13, which did not have any occurrence of STX biosynthesis identified (Berry et al., 2017).

There is also a potential for other strains and species to produce STX in WLE. While the presence of LE20-WE8 corresponded well with the presence of *sxt* genes through read mapping analysis, there were several instances where complete STX operons were identified as present and the species belonging to MAG LE20-WE8, *D. sp000312705,* was not. Both species corresponding with STX genes in the instances where *D. sp000312705* was absent, *D. circinale* and *C. issatschenoki*, have strains capable of producing STXs in other environments (Mihali et al., 2009; Nogueira et al., 2004). Given the trends in late-season bloom occurrences of *D. circinale*, it is particularly important to acknowledge its ongoing potential for STX production (Fig. 4). *sxt* genes are variably present in closely related strains (Fig. 1) and do not correlate with overall organism phylogeny. This variability suggests frequent gene loss or horizontal gene transfer of *sxt* genes within the ADA-clade and incites an additional concern in determining STX sources in WLE.

Potentially positive observations also came to light through the characterization and examined distribution of LE20-WE8. The absence of the *sxtX* gene responsible for neoSTX production suggests an inability to produce one of the most toxic STX congeners (Ballot et al., 2016). The current absence of STX genes within late season cyanoHAB blooms is also a positive indication. However, the presence of STX earlier in the season alongside microcystin may pose a risk from synergistic effects between the two cyanotoxins or other secondary metabolites (Metcalf & Codd, 2020).

Conclusions

This study underscores the significance of expanding the research scope of cyanoHAB bloom monitoring to encompass various potential cyanotoxins. The identification of a *Dolichospermum* MAG (LE20-WE8) possessing a full suite of STX biosynthesis genes is a significant and necessary stride forward in determining risks to both public health and ecosystem well-being associated with these toxins. The available infrastructure currently used to investigate WLE cyanoHABs, ranging from field monitoring efforts (Boegehold et al., 2023) to data processing tools (GLAMR, greatlakesomics.org), provides an opportunity to address many subsequent questions regarding STX production in WLE. Furthermore, the importance of employing these methods in the broader Great Lakes, where many areas experience cyanoHABs, may be valuable in determining potential remaining STX producers, as well as producers of other cyanotoxins and impactful secondary metabolites.

Materials and Methods

Database

In this study, the GLAMR database was utilized to access and analyze over 500 metagenomic samples from WLE collected between 2014-2022 (greatlakesomics.org). GLAMR serves as a comprehensive repository housing metagenomic, transcriptomic, and amplicon sequencing data alongside relevant metadata, including environmental measurements and processing details. This database boasts a user-friendly interface and standardized bioinformatics pipelines, streamlining the exploration and comparison of omics datasets.

Bioinformatic Analysis

Assembly, binning, and refinement

De novo assemblies were performed on single samples to recover MAGs using GLAMR pipelines for metagenomic assembly and binning (https://github.com/Geo-omics/GLAMR_omics_pipelines). Metagenomic assemblies were completed using MEGAHIT (D. Li et al., 2015). Multiple binning software were used to maximize MAG recovery, including CONCOCT, MetaDecoder, VAMB, MetaBAT 2, MaxBin 2.0, and SemiBin (Alneberg et al., 2014; Kang et al., 2019; Liu et al., 2022; Nissen et al., 2021; Pan et al., 2022; Wu et al., 2016). MAG refinement was manually performed on representative MAGs using Anvi'o v.7 utilizing differential coverage read mapping from bowtie2 (Eren et al., 2015; Langmead & Salzberg, 2012). Kaiju 1.6.2 annotated metagenomic contigs using its nr database (accessed March 23, 2023), with results imported into Anvi'o to aid in MAG assessment and curation (Menzel et al., 2016).

MAG identification

blastn was used to check GLAMR assemblies for the presence of *sxtA*, *G*, *H*, and *I*, four genes essential to produce STXs 5/13/24 3:09:00 PM. This step and all other uses of blastn in this study set an e-value of 1e-2 to reduce the chances of missing any potential targets of interest. Bins from assemblies with positive blastn hits to *sxtA, G*, *H*, or *I* were then examined for the presence of these *sxt* genes using a second blastn analysis.

Through GLAMR pipelines, all bins were assigned taxonomy using GTDB-Tk v2.1.1 with reference data from GTDB release207 v2 (Chaumeil et al., 2020) and completion, contamination, and strain heterogeneity was assessed using CheckM v.1.2.2 (Parks, 2014). Additional ADA bins were selected for further analysis if they were annotated belonging to the family *Nostocaceae*, regardless of STX gene presence. Ultimately, only bins with greater than or equal to 90% completeness and less than 10% contamination were retained for subsequent analyses. These bins, referred to as MAGs due to their high quality, were then run through dRep to identify the highest quality representatives at two different secondary ANI cutoffs, 98 and 99 (Olm et al., 2017).

Genetic analysis and visualization

To establish the phylogenetic context of representative ADA MAGs, GToTree was employed to construct a phylogenomic tree (Lee, 2019). Using GToTree, a maximum likelihood tree encompassing 251 marker genes was generated featuring ADA genomes sourced from prior studies and an outgroup represented by Cyanobium gracile PCC 6307 (NCBI: PRJNA158695) (Dreher et al., 2021; Österholm et al., 2020; Sheik et al., 2022; Yancey et al., 2023). ANI values between representative ADA MAGs were calculated using FastANI (Jain et al., 2018). The program “clinker” was utilized to visualize contigs of LE20-WE8 containing STX genes and compare them with previously characterized STX operons (Gilchrist & Chooi, 2021; Mihali et al., 2009).

Metagenomic read mapping

Metagenomic read mapping was performed using Minimap2 with presets for short genomic paired-end reads (Li, 2018). Minimap2 output was filtered using AUGUSTUS to retain reads with minimum query cover of 50% and minimum percent ID of 80% when mapping to *sxt* genes and a minimum query cover of 80% and minimum percent ID of 90% when mapping to representative ADA MAGs (Stanke et al., 2006). The determination of the final cutoff (>60% cover) for each of the three STX reference contig sequences, indicating the presence or absence of STX genes in a sample, was validated by plotting percent coverage and mean coverage depth of read mapping results for each reference. Rsamtools was utilized to aid in calculations of percent identity, coverage, and mean depth for read mapping results (Morgan et al., 2023). RPKM values were calculated in the R statistical platform (Equation 1) (Dick, 2018). RPKM values were aggregated for MAGs of the same species to reduce unambiguous read mapping between MAGs within a species. Results were visualized using ggplot2 (Wickham, 2016).

Equation 1.

$$Relative\ Organism\ Abundance = RPKM\ per\ genomes = \frac{total\ \#\ of\ reads\ mapped\ per\ MAG}{(length\ of\ MAG)*(total\ number\ of\ reads\ per\ sample)} * 10^6$$

Functional profiling

Representative ADA MAGs were annotated to predict protein-coding genes in microbial genomes using Prodigal v2.6.3 (Hyatt et al., 2010). KofamScan v1.3.0 annotated genes based on KEGG orthology (database accessed September 16, 2022). KEGGDecoder v1.3 was used to examine KofamScan output to determine the metabolic pathway completeness (https://github.com/bjtully/BioData). Results were visualized using ggplot2. To supplement the profiling of pathways that were either not included in the original analysis or found to be incomplete by KofamScan, Bakta v1.8.1 annotations were generated using database v5.0 (Schwengers et al., 2021).

References

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146. https://doi.org/10.1038/nmeth.3103

Al-Tebrineh Jamal, Mihali Troco Kaan, Pomati Francesco, & Neilan Brett A. (2010). Detection of Saxitoxin-Producing Cyanobacteria and Anabaena circinalis in Environmental Water Blooms by Quantitative PCR. *Applied and Environmental Microbiology*, *76*(23), 7836–7842. https://doi.org/10.1128/AEM.00174-10

Ballot, A., Bernard, C., & Fastner, J. (2016). Saxitoxin and Analogues. In *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis* (pp. 148–154). https://doi.org/10.1002/9781119068761.ch14

Belisle, B. S., Steffen, M. M., Pound, H. L., Watson, S. B., DeBruyn, J. M., Bourbonniere, R. A., Boyer, G. L., & Wilhelm, S. W. (2016). Urea in Lake Erie: Organic nutrient sources as potentially important drivers of phytoplankton biomass. *Journal of Great Lakes Research*, *42*(3), 599–607. https://doi.org/10.1016/j.jglr.2016.03.002

Berry, J. P., & Lind, O. (2010). First evidence of “paralytic shellfish toxins” and cylindrospermopsin in a Mexican freshwater system, Lago Catemaco, and apparent bioaccumulation of the toxins in “tegogolo” snails (Pomacea patula catemacensis). *Harmful Algal Blooms and Natural Toxins in Fresh and Marine Waters -- Exposure, Occurrence, Detection, Toxicity, Control, Management and Policy*, *55*(5), 930–938. https://doi.org/10.1016/j.toxicon.2009.07.035

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. *Environmental Microbiology*, *19*(3), 1149–1162. https://doi.org/10.1111/1462-2920.13640

Boegehold, A. G., Burtner, A. M., Camilleri, A. C., Carter, G., DenUyl, P., Fanslow, D., Fyffe Semenyuk, D., Godwin, C. M., Gossiaux, D., Johengen, T. H., Kelchner, H., Kitchens, C., Mason, L. A., McCabe, K., Palladino, D., Stuart, D., Vanderploeg, H., & Errera, R. (2023). Routine monitoring of western Lake Erie to track water quality changes associated with cyanobacterial harmful algal blooms. *Earth Syst. Sci. Data Discuss.*, *2023*, 1–39. https://doi.org/10.5194/essd-2023-62

Chaffin, J. (2013). Nitrogen Constrains the Growth of Late Summer Cyanobacterial Blooms in Lake Erie. *Adv Microbiol.*, *3*. https://doi.org/10.4236/aim.2013.36A003

Chaffin, J. D., Bratton, J. F., Verhamme, E. M., Bair, H. B., Beecher, A. A., Binding, C. E., Birbeck, J. A., Bridgeman, T. B., Chang, X., Crossman, J., Currie, W. J. S., Davis, T. W., Dick, G. J., Drouillard, K. G., Errera, R. M., Frenken, T., MacIsaac, H. J., McClure, A., McKay, R. M., … Zhou, X. (2021). The Lake Erie HABs Grab: A binational collaboration to characterize the western basin cyanobacterial harmful algal blooms at an unprecedented high-resolution spatial scale. *Harmful Algae*, *108*, 102080. https://doi.org/10.1016/j.hal.2021.102080

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. *Journal of Great Lakes Research*, *45*(2), 277–289. https://doi.org/10.1016/j.jglr.2018.12.006

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Christensen, V. G., & Khan, E. (2020). Freshwater neurotoxins and concerns for human, animal, and ecosystem health: A review of anatoxin-a and saxitoxin. *Science of The Total Environment*, *736*, 139515. https://doi.org/10.1016/j.scitotenv.2020.139515

Colleen E. Yancey, Derek J. Smith, Paul A. Den Uyl, Osama G. Mohamed, Fengan Yu, Steven A. Ruberg, Justin D. Chaffin, Kelly D. Goodwin, Ashootosh Tripathi, David H. Sherman, & Gregory J. Dick. (2022). 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. *Applied and Environmental Microbiology*, *0*(0), e02464-21. https://doi.org/doi:10.1128/aem.02464-21

Conroy, J. D., Quinlan, E. L., Kane, D. D., & Culver, D. A. (2007). Cylindrospermopsis in Lake Erie: Testing its Association with Other Cyanobacterial Genera and Major Limnological Parameters. *Journal of Great Lakes Research*, *33*(3), 519–535. https://doi.org/10.3394/0380-1330(2007)33[519:CILETI]2.0.CO;2

da Silva, C. A., Oba, E. T., Ramsdorf, W. A., Magalhães, V. F., Cestari, M. M., Oliveira Ribeiro, C. A., & Silva de Assis, H. C. (2011). First report about saxitoxins in freshwater fish Hoplias malabaricus through trophic exposure. *Toxicon*, *57*(1), 141–147. https://doi.org/10.1016/j.toxicon.2010.10.015

D'Agostino, P. M., Boundy, M. J., Harwood, T. D., Carmichael, W. W., Neilan, B. A., & Wood, S. A. (2019). Re-evaluation of paralytic shellfish toxin profiles in cyanobacteria using hydrophilic interaction liquid chromatography-tandem mass spectrometry. *Toxicon*, *158*, 1–7. https://doi.org/10.1016/j.toxicon.2018.11.301

Den Uyl, P. A., Thompson, L. R., Errera, R. M., Birch, J. M., Preston, C. M., Ussler, W., Yancey, C. E., Chaganti, S. R., Ruberg, S. A., Doucette, G. J., Dick, G. J., Scholin, C. A., & Goodwin, K. D. (2022). Lake Erie field trials to advance autonomous monitoring of cyanobacterial harmful algal blooms. *Frontiers in Marine Science*, *9*. https://www.frontiersin.org/articles/10.3389/fmars.2022.1021952

Dias, E., Pereira, P., & Franca, S. (2002). PRODUCTION OF PARALYTIC SHELLFISH TOXINS BY APHANIZOMENON SP. LMECYA 31 (CYANOBACTERIA)1. *Journal of Phycology*, *38*(4), 705–712. https://doi.org/10.1046/j.1529-8817.2002.01146.x

Dick, G. (2018). *Genomic approaches in earth and environmental sciences*. John Wiley & Sons.

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. (2018). A closely-related clade of globally distributed bloom-forming cyanobacteria within the Nostocales. *Harmful Algae*, *77*, 93–107.

Eren, A., Esen, Ö., Quince, C., Vineis, J., Sogin, M., & Delmont, T. (2015). *Anvi'o: An advanced analysis and visualization platform for 'omics data*. https://doi.org/10.7287/PEERJ.PREPRINTS.1275

Galvão, J. A., Oetterer, M., Bittencourt-Oliveira, M. do C., Gouvêa-Barros, S., Hiller, S., Erler, K., Luckas, B., Pinto, E., & Kujbida, P. (2009). Saxitoxins accumulation by freshwater tilapia (Oreochromis niloticus) for human consumption. *Toxicon*, *54*(6), 891–894. https://doi.org/10.1016/j.toxicon.2009.06.021

Gilchrist, C. L. M., & Chooi, Y.-H. (2021). clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics*, *37*(16), 2473–2475. https://doi.org/10.1093/bioinformatics/btab007

Hellweger, F. L., Martin, R. M., Eigemann, F., Smith, D. J., Dick, G. J., & Wilhelm, S. W. (2022). Models predict planned phosphorus load reduction will make Lake Erie more toxic. *Science*, *376*(6596), 1001–1005. https://doi.org/10.1126/science.abm6791

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Ibelings, B. W., & Chorus, I. (2007). Accumulation of cyanobacterial toxins in freshwater "seafood" and its consequences for public health: A review. *Environmental Pollution*, *150*(1), 177–192. https://doi.org/10.1016/j.envpol.2007.04.012

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

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

Kellmann, R., Mihali, T. K., Jeon, Y. J., Pickford, R., Pomati, F., & Neilan, B. A. (2008). Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *Applied and Environmental Microbiology*, *74*(13), 4044–4053. https://doi.org/10.1128/AEM.00353-08

Koeppel, A. F., Goodrum, W. J., Steffen, M. M., Wurch, L. L., & Turner, S. D. (2022). Environmental DNA sequencing dataset from Lake Erie algal blooms using Oxford Nanopore MinION. *Data in Brief*, *45*, 108688. https://doi.org/10.1016/j.dib.2022.108688

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larson, C. A., Mirza, B., Rodrigues, J. L. M., & Passy, S. I. (2018). Iron limitation effects on nitrogen-fixing organisms with possible implications for cyanobacterial blooms. *FEMS Microbiology Ecology*, *94*(5), fiy046. https://doi.org/10.1093/femsec/fiy046

Latifi, A., Ruiz, M., & Zhang, C.-C. (2009). Oxidative stress in cyanobacteria. *FEMS Microbiology Reviews*, *33*(2), 258–278. https://doi.org/10.1111/j.1574-6976.2008.00134.x

Lee, M. D. (2019). GToTree: A user-friendly workflow for phylogenomics. *Bioinformatics*, *35*(20), 4162–4164. https://doi.org/10.1093/bioinformatics/btz188

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Liu, C.-C., Dong, S.-S., Chen, J.-B., Wang, C., Ning, P., Guo, Y., & Yang, T.-L. (2022). MetaDecoder: A novel method for clustering metagenomic contigs. *Microbiome*, *10*(1), 46. https://doi.org/10.1186/s40168-022-01237-8

Llewellyn, L., Negri, A., Doyle, J., Baker, P., Beltran, E., & Neilan, B. (2001). Radioreceptor assays for sensitive detection and quantitation of saxitoxin and its analogues from strains of the freshwater cyanobacterium, Anabaena circinalis. *Environmental Science & Technology*, *35*(7), 1445–1451.

Mahmood, N. A., & Carmichael, W. W. (1986). Paralytic shellfish poisons produced by the freshwater cyanobacterium Aphanizomenon flos-aquae NH-5. *Toxicon*, *24*(2), 175–186. https://doi.org/10.1016/0041-0101(86)90120-0

Martin, R. M., Denney, M. K., Pound, H. L., Chaffin, J. D., Bullerjahn, G. S., McKay, R. M. L., Zastepa, A., Jones, K. A., Castro, H. F., Campagna, S. R., & Wilhelm, S. W. (2023). Sulfolipid substitution ratios of Microcystis aeruginosa and planktonic communities as an indicator of phosphorus limitation in Lake Erie. *Limnology and Oceanography*, *68*(5), 1117–1131. https://doi.org/10.1002/lno.12333

McKindles Katelyn M., Manes Makayla A., DeMarco Jonathan R., McClure Andrew, McKay R. Michael, Davis Timothy W., & Bullerjahn George S. (2020). Dissolved Microcystin Release Coincident with Lysis of a Bloom Dominated by Microcystis spp. In Western Lake Erie Attributed to a Novel Cyanophage. *Applied and Environmental Microbiology*, *86*(22), e01397-20. https://doi.org/10.1128/AEM.01397-20

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, *7*(1), 11257. https://doi.org/10.1038/ncomms11257

Metcalf, J. S., & Codd, G. A. (2020). Co-Occurrence of Cyanobacteria and Cyanotoxins with Other Environmental Health Hazards: Impacts and Implications. *Toxins*, *12*(10). https://doi.org/10.3390/toxins12100629

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., Chaffin, J. D., Cho, K., Confesor, R., Daloglu, I., Depinto, J. V., Evans, M. A., Fahnenstiel, G. L., He, L., Ho, J. C., Jenkins, L., Johengen, T. H., Kuo, K. C., Laporte, E., … Zagorski, M. A. (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6448–6452. https://doi.org/10.1073/pnas.1216006110

Mihali, T. K., Kellmann, R., & Neilan, B. A. (2009). Characterisation of the paralytic shellfish toxin biosynthesis gene clusters in Anabaena circinalis AWQC131C and Aphanizomenon sp. NH-5. *BMC Biochemistry*, *10*(1), 8. https://doi.org/10.1186/1471-2091-10-8

Morgan, M., Pagès, H., Obenchain, V., & Hayden, N. (2023). *Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import.* (2.18.0) [Computer software]. https://doi.org/10.18129/B9.bioc.Rsamtools,

Munday, R., & Reeve, J. (2013). Risk assessment of shellfish toxins. *Toxins*, *5*(11), 2109–2137.

Naknaen, A., Ratsameepakai, W., Suttinun, O., Sukpondma, Y., Khan, E., & Pomwised, R. (2021). Microcystis Sp. Co-Producing Microcystin and Saxitoxin from Songkhla Lake Basin, Thailand. *Toxins*, *13*(9). https://doi.org/10.3390/toxins13090631

Nauman, C., Stanislawczyk, K., Reitz, L. A., & Chaffin, J. D. (2024). The spatiotemporal distribution of potential saxitoxin-producing cyanobacteria in western Lake Erie. *Journal of Great Lakes Research*, 102342. https://doi.org/10.1016/j.jglr.2024.102342

Negri, A. P., & Jones, G. J. (1995). Bioaccumulation of paralytic shellfish poisoning (PSP) toxins from the cyanobacterium Anabaena circinalis by the freshwater mussel Alathyria condola. *Toxicon*, *33*(5), 667–678. https://doi.org/10.1016/0041-0101(94)00180-G

Negri, A. P., Jones, G. J., & Hindmarsh, M. (1995). Sheep mortality associated with paralytic shellfish poisons from the cyanobacterium Anabaena circinalis. *Toxicon*, *33*(10), 1321–1329. https://doi.org/10.1016/0041-0101(95)00068-W

Nissen, J. N., Johansen, J., Allesøe, R. L., Sønderby, C. K., Armenteros, J. J. A., Grønbech, C. H., Jensen, L. J., Nielsen, H. B., Petersen, T. N., Winther, O., & Rasmussen, S. (2021). Improved metagenome binning and assembly using deep variational autoencoders. *Nature Biotechnology*, *39*(5), 555–560. https://doi.org/10.1038/s41587-020-00777-4

Nogueira, I. C. G., Pereira, P., Dias, E., Pflugmacher, S., Wiegand, C., Franca, S., & Vasconcelos, V. M. (2004). Accumulation of paralytic shellfish toxins (PST) from the cyanobacterium Aphanizomenon issatschenkoi by the cladoceran Daphnia magna. *Toxicon*, *44*(7), 773–780. https://doi.org/10.1016/j.toxicon.2004.08.006

Ohio Environmental Protection Agency. (2022). *Public Water System Harmful Algal Bloom Response Strategy* (pp. 1–61). https://epa.ohio.gov/monitor-pollution/maps-and-advisories/harmful-algal-bloom-monitoring

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), 2864–2868. https://doi.org/10.1038/ismej.2017.126

Orr, P. T., Jones, G. J., & Hamilton, G. R. (2004). Removal of saxitoxins from drinking water by granular activated carbon, ozone and hydrogen peroxide—Implications for compliance with the Australian drinking water guidelines. *Water Research*, *38*(20), 4455–4461. https://doi.org/10.1016/j.watres.2004.08.024

Österholm, J., Popin, R. V., Fewer, D. P., & Sivonen, K. (2020). Phylogenomic Analysis of Secondary Metabolism in the Toxic Cyanobacterial Genera Anabaena, Dolichospermum and Aphanizomenon. *Toxins*, *12*(4). https://doi.org/10.3390/toxins12040248

Paerl, H. W., & Huisman, J. (2009). Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports*, *1*(1), 27–37. https://doi.org/10.1111/j.1758-2229.2008.00004.x

Pan, S., Zhu, C., Zhao, X.-M., & Coelho, L. P. (2022). A deep siamese neural network improves metagenome-assembled genomes in microbiome datasets across different environments. *Nature Communications*, *13*(1), 2326. https://doi.org/10.1038/s41467-022-29843-y

Parks, D. (2014). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *PeerJ PrePrints*, *2*. https://doi.org/10.7287/peerj.preprints.554v1

Pearson, L., Mihali, T., Moffitt, M., Kellmann, R., & Neilan, B. (2010). On the Chemistry, Toxicology and Genetics of the Cyanobacterial Toxins, Microcystin, Nodularin, Saxitoxin and Cylindrospermopsin. *Marine Drugs*, *8*(5), 1650–1680. https://doi.org/10.3390/md8051650

Rapala, J., Robertson, A., Negri, A. P., Berg, K. A., Tuomi, P., Lyra, C., Erkomaa, K., Lahti, K., Hoppu, K., & Lepistö, L. (2005). First report of saxitoxin in Finnish lakes and possible associated effects on human health. *Environmental Toxicology*, *20*(3), 331–340. https://doi.org/10.1002/tox.20109

Schwengers, O., Jelonek, L., Dieckmann, M. A., Beyvers, S., Blom, J., & Goesmann, A. (2021). Bakta: Rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. In *Microbial Genomics* (Vol. 7, Issue 11). Microbiology Society. https://doi.org/10.1099/mgen.0.000685

Selwood, A. I., Waugh, C., Harwood, D. T., Rhodes, L. L., Reeve, J., Sim, J., & Munday, R. (2017). Acute Toxicities of the Saxitoxin Congeners Gonyautoxin 5, Gonyautoxin 6, Decarbamoyl Gonyautoxin 2&3, Decarbamoyl Neosaxitoxin, C-1&2 and C-3&4 to Mice by Various Routes of Administration. *Toxins*, *9*(2). https://doi.org/10.3390/toxins9020073

Sheik, C. S., Natwora, K. E., Alexson, E. E., Callaghan, J. D., Sailer, A., Schreiner, K. M., Steinman, B. A., Finkenbinder, M. S., Filstrup, C. T., & Bramburger, A. J. (2022). Dolichospermum blooms in Lake Superior: DNA-based approach provides insight to the past, present and future of blooms. *Journal of Great Lakes Research*, *48*(5), 1191–1205. https://doi.org/10.1016/j.jglr.2022.08.002

Sierra, M. A., & Martínez-Álvarez, R. (2020). Ricin and Saxitoxin: Two Natural Products That Became Chemical Weapons. *Journal of Chemical Education*, *97*(7), 1707–1714. https://doi.org/10.1021/acs.jchemed.9b00841

Smith Derek J., Berry Michelle A., Cory Rose M., Johengen Thomas H., Kling George W., Davis Timothy W., & Dick Gregory J. (2022). Heterotrophic Bacteria Dominate Catalase Expression during Microcystis Blooms. *Applied and Environmental Microbiology*, *88*(14), e02544-21. https://doi.org/10.1128/aem.02544-21

Smith, P. T. (2000). Freshwater neurotoxins: Mechanisms of action, pharmacology, toxicology, and impacts on aquaculture. *FOOD SCIENCE AND TECHNOLOGY-NEW YORK-MARCEL DEKKER-*, 583–602.

Soto-Liebe, K., Murillo, A. A., Krock, B., Stucken, K., Fuentes-Valdés, J. J., Trefault, N., Cembella, A., & Vásquez, M. (2010). Reassessment of the toxin profile of Cylindrospermopsis raciborskii T3 and function of putative sulfotransferases in synthesis of sulfated and sulfonated PSP toxins. *Toxicon*, *56*(8), 1350–1361. https://doi.org/10.1016/j.toxicon.2010.07.022

Spiese, C. E., Bowling, M. N., & Moeller, S. E. M. (2023). Is glyphosate an underlying cause of increased dissolved reactive phosphorus loading in the Western Lake Erie basin? *Journal of Great Lakes Research*, *49*(3), 631–639. https://doi.org/10.1016/j.jglr.2023.03.009

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(suppl_2), W435–W439. https://doi.org/10.1093/nar/gkl200

Steffen, M. M., Davis, T. W., McKay, R. M. L., Bullerjahn, G. S., Krausfeldt, L. E., Stough, J. M. A., Neitzey, M. L., Gilbert, N. E., Boyer, G. L., Johengen, T. H., Gossiaux, D. C., Burtner, A. M., Palladino, D., Rowe, M. D., Dick, G. J., Meyer, K. A., Levy, S., Boone, B. E., Stumpf, R. P., … Wilhelm, S. W. (2017). Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH. *Environmental Science & Technology*, *51*(12), 6745–6755. https://doi.org/10.1021/acs.est.7b00856

Stucken, K., John, U., Cembella, A., Murillo, A. A., Soto-Liebe, K., Fuentes-Valdés, J. J., Friedel, M., Plominsky, A. M., Vásquez, M., & Glöckner, G. (2010). The Smallest Known Genomes of Multicellular and Toxic Cyanobacteria: Comparison, Minimal Gene Sets for Linked Traits and the Evolutionary Implications. *PLOS ONE*, *5*(2), e9235. https://doi.org/10.1371/journal.pone.0009235

Suarez-Isla, B. A. (2014). Saxitoxin and Other Paralytic Toxins: Toxicological Profile. In P. Gopalakrishnakone, V. Haddad Jr., W. R. Kem, A. Tubaro, & E. Kim (Eds.), *Marine and Freshwater Toxins: Marine and Freshwater Toxins* (pp. 1–16). Springer Netherlands. https://doi.org/10.1007/978-94-007-6650-1_25-1

Takamura, N., Iwakuma, T., & Yasuno, M. (1987). Uptake of 13 C and 15 N (ammonium, nitrate and urea) by Microcystis in Lake Kasumigaura. *Journal of Plankton Research*, *9*(1), 151–165. https://doi.org/10.1093/plankt/9.1.151

United Nations Environment Programme. (1984). *Aquatic (marine and freshwater) biotoxins.* World Health Organization. https://wedocs.unep.org/20.500.11822/29334

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. *Limnology and Oceanography*, *66*(6), 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. (2016). 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

Westrick, J. A. (2008). Cyanobacterial toxin removal in drinking water treatment processes and recreational waters. In *Cyanobacterial harmful algal blooms: State of the science and research needs* (pp. 275–290). Springer.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer International Publishing.

Wiese, M., D'Agostino, P. M., Mihali, T. K., Moffitt, M. C., & Neilan, B. A. (2010). Neurotoxic alkaloids: Saxitoxin and its analogs. *Marine Drugs*, *8*(7), 2185–2211. https://doi.org/10.3390/md8072185

Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, *32*(4), 605–607. https://doi.org/10.1093/bioinformatics/btv638

Wynne, T. T., & Stumpf, R. P. (2015). Spatial and Temporal Patterns in the Seasonal Distribution of Toxic Cyanobacteria in Western Lake Erie from 2002–2014. *Toxins*, *7*(5), 1649–1663. https://doi.org/10.3390/toxins7051649

Yancey, C. E., Mathiesen, O., & Dick, G. J. (2023). Transcriptionally active nitrogen fixation and biosynthesis of diverse secondary metabolites by Dolichospermum and Aphanizomenon-like Cyanobacteria in western Lake Erie Microcystis blooms. *Harmful Algae*, *124*, 102408. https://doi.org/10.1016/j.hal.2023.102408

Supplemental Figures and Tables

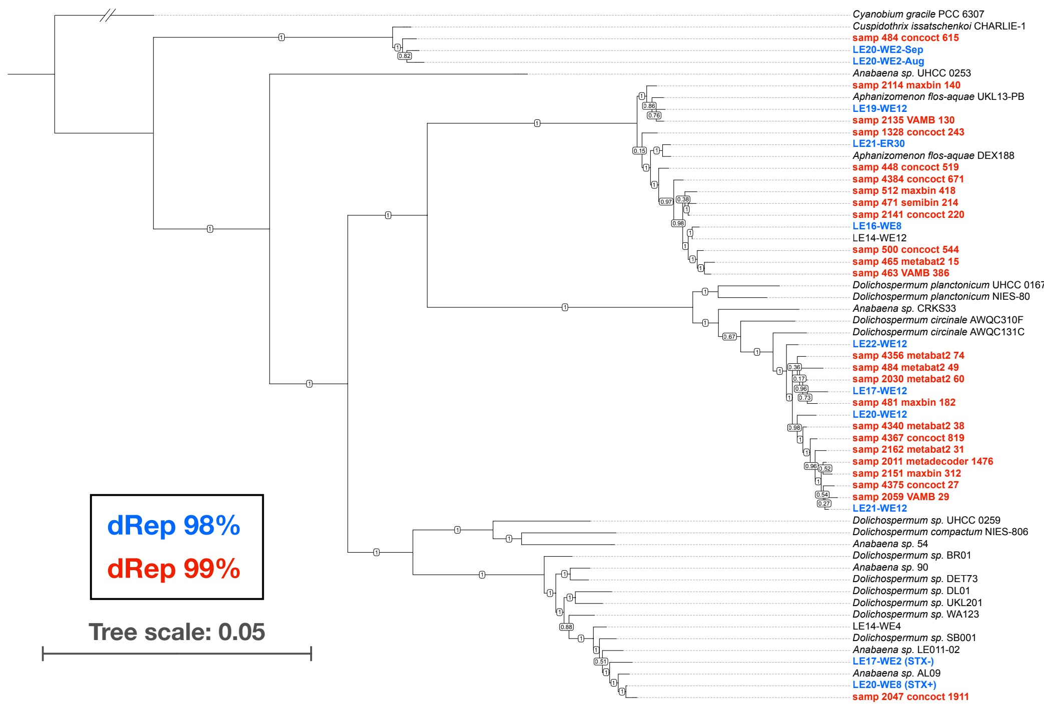

Figure S1. Phylogenomic tree comprised of ADA clade members constructed using a maximum likelihood approach, utilizing 251 housekeeping genes through the GToTree workflow. The dataset incorporated ADA genomes from prior investigations (Dreher et al., 2021; Österholm et al., 2020; Sheik et al., 2022, Yancey et al., 2023). MAGs obtained from this study are highlighted in colored and bolded text. The distinct colors represent the average nucleotide identity (ANI) thresholds (98 in blue and 99 in red) applied during MAG dereplication to select the respective strain representatives.

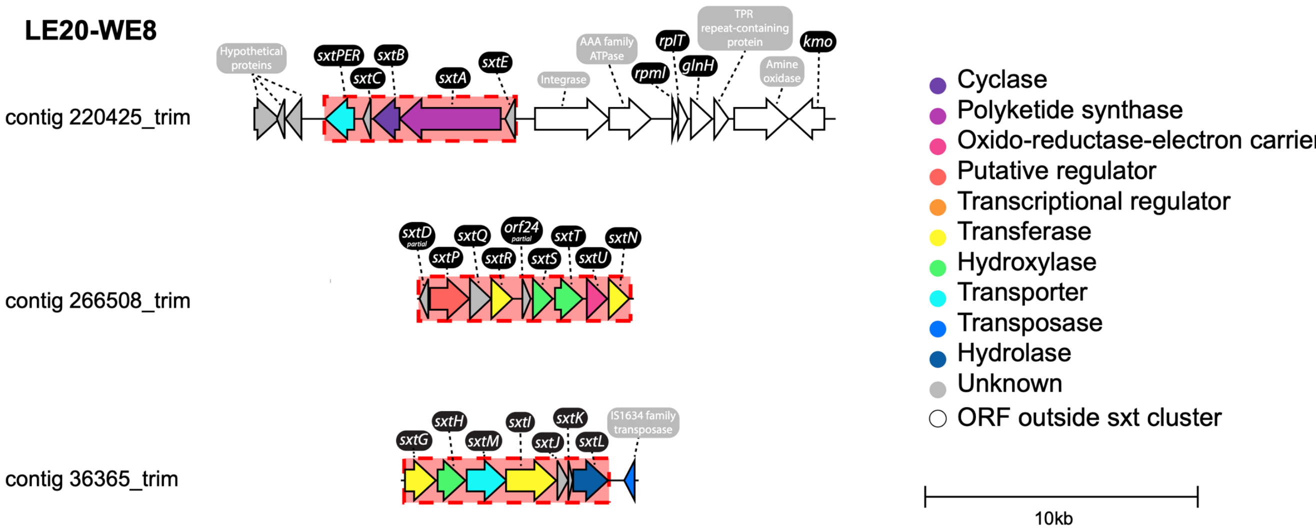

Figure S2. Sections of LE20-WE8 contigs highlighted (in red) indicate sequences utilized for genomic read mapping of *sxt* genes. This mapping serves to determine the presence or absence of STX biosynthesis potential in a sample.

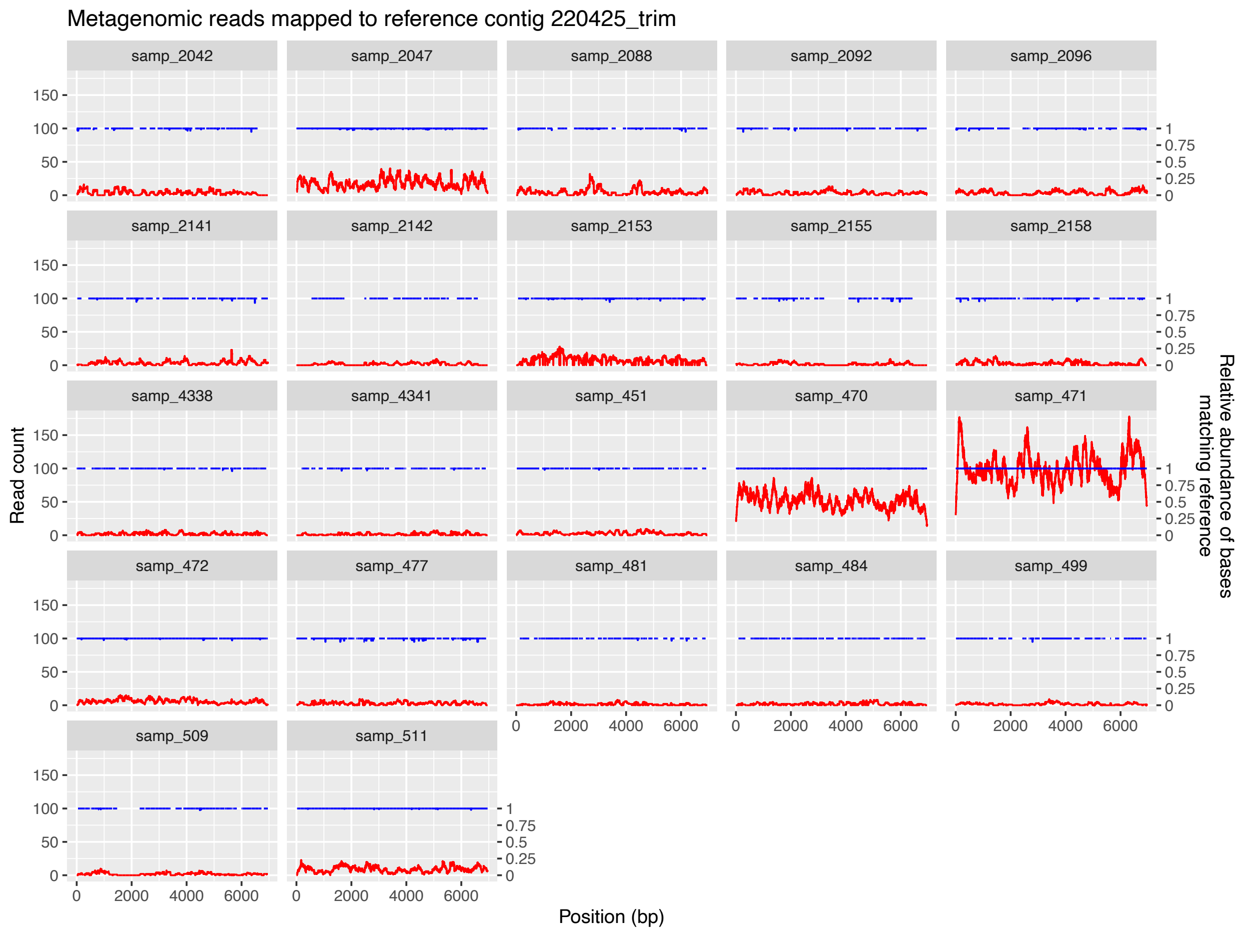

Figure S3. Read counts at each base position (red) and relative abundance of matching bases to reference averaged over 10bp intervals (blue) of for metagenomic read mapping results. Breaks in blue line and red line at 0 indicate no read mapping at that position. Reference contig 220425_trim only (Fig. S2).

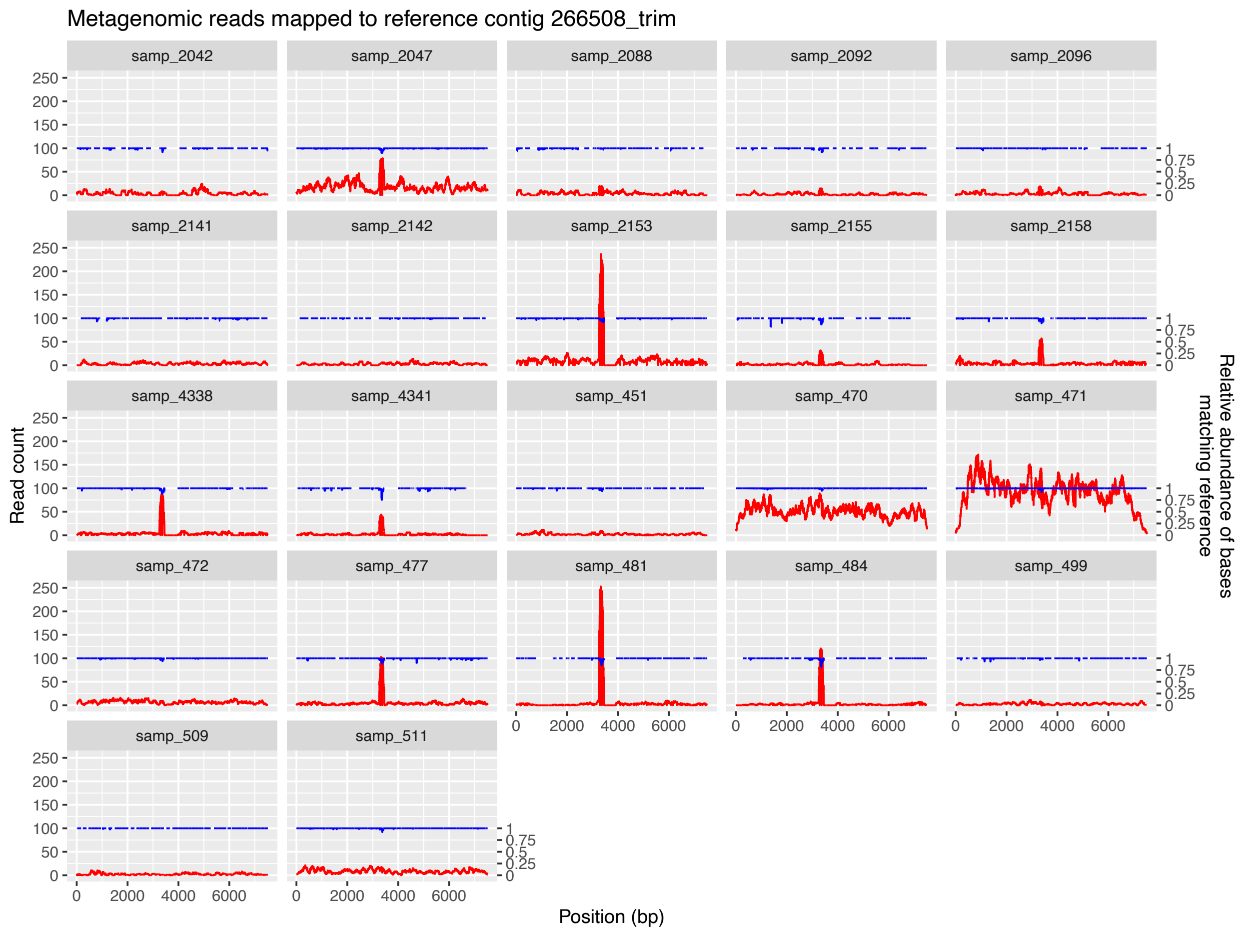

Figure S4. Read counts at each base position (red) and relative abundance of matching bases to reference averaged over 10bp intervals (blue) of for metagenomic read mapping results. Breaks in blue line and red line at 0 indicate no read mapping at that position. Reference contig 266508_trim only (Fig. S2). Investigations into the 203bp sequence causing substantial peaks in read counts does not appear to be explicitly related to saxitoxin production, appearing in genomes not capable of saxitoxin production (STX-) via web based blastn (blast.ncbi.nlm.nih.gov).

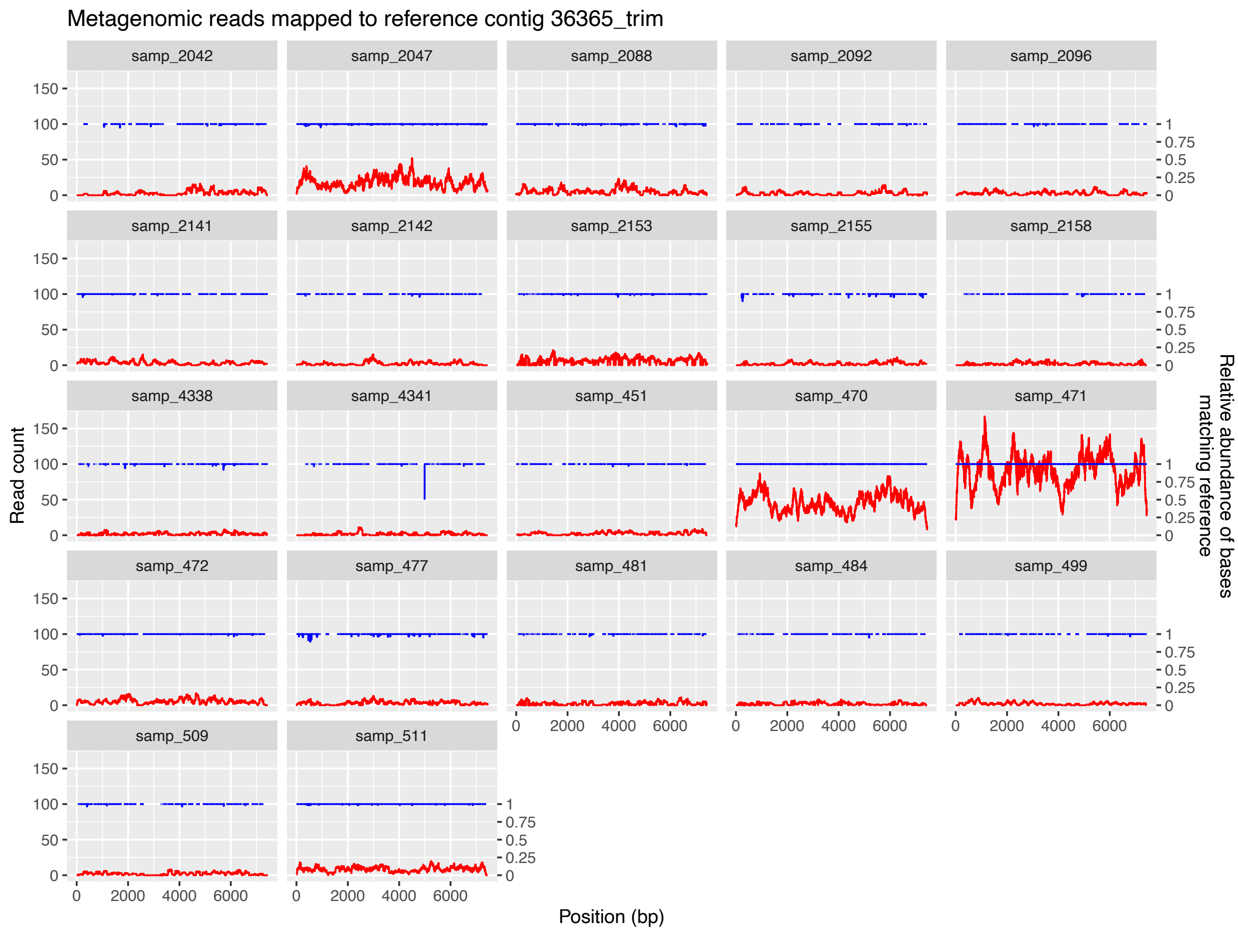

Figure S5. Read counts at each base position (red) and relative abundance of matching bases to reference averaged over 10bp intervals (blue) of for metagenomic read mapping results. Breaks in blue line and red line at 0 indicate no read mapping at that position. Reference contig 36365_trim only (Fig. S2).

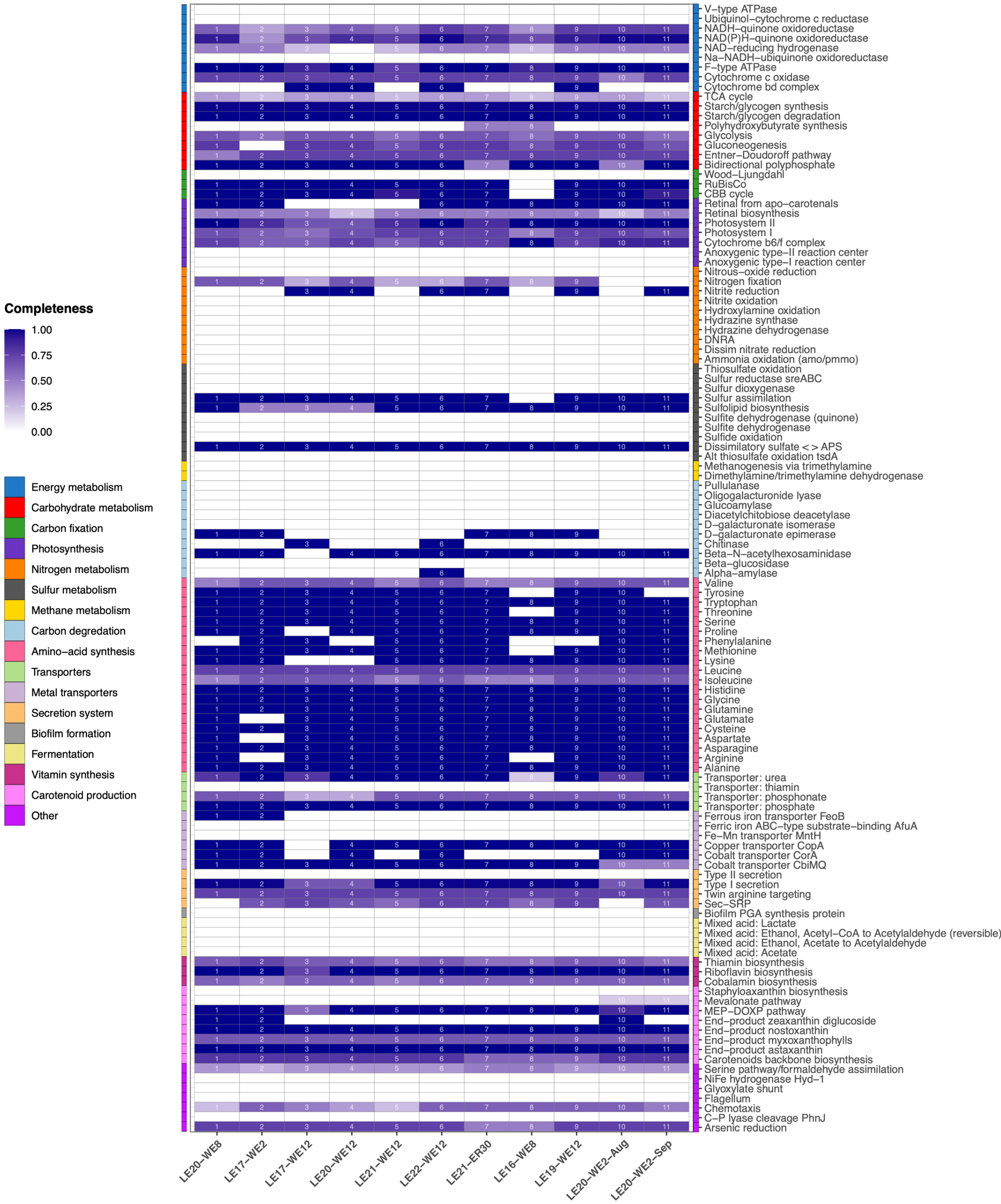

Figure S6. Heatmap visualization of KEGGDecoder functional analysis for representative MAGs, including pathways with zero completeness across examined MAGs.

SampleName	percent_cover all_ref	percent_id all_ref	lat	lon	collection date	NOAA_Site	contig_220425_trim percent_cover	contig_266508_trim percent_cover	contig_36365_trim percent_cover	contig_220425_trim percent_id	contig_266508_trim percent_id	contig_36365_trim percent_id	sample_id GLAMR
7-18-16WLE4	74.14	99.92	41.826	-83.1947	7/18/16	WE4	79.01	70.65	73.18	99.96	99.82	99.94	samp_2042
8-1-16WLE8	99.99	99.8	41.8346	-83.3565	8/1/16	WE8	100	100	100	99.91	99.52	99.91	samp_2047
7-30-18WLE12	84.07	99.91	41.7036	-83.2547	7/30/18	WE12	82.85	77.23	92.33	99.94	99.5	99.93	samp_2088
8-13-18WLE12	75.99	99.93	41.7032	-83.2536	8/13/18	WE12	89.05	69.78	70.17	99.96	99.69	99.98	samp_2092
8-20-18WLE12	83.12	99.95	41.7031	-83.2543	8/20/18	WE12	81.86	86.57	80.92	99.94	99.68	99.98	samp_2096
7-6-20WLE8	85.75	99.95	NA	NA	7/6/20	WE8	83.79	86.79	86.6	99.95	99.92	99.97	samp_2141
7-6-20WLE12	74.7	99.98	41.7035	-83.2552	7/6/20	WE12	63.09	80.57	79.65	99.98	99.97	99.98	samp_2142
8-24-20WLE8	95.49	99.4	41.8342	-83.3604	8/24/20	WE8	97.31	93.71	98.84	99.79	98.26	99.85	samp_2153
8-31-20WLE2	70.46	99.41	NA	NA	8/31/20	WE2	65.7	67.27	78.99	99.8	98.41	99.65	samp_2155
8-31-20WLE12	85.32	99.52	NA	NA	8/31/20	WE12	84.92	92.03	81.58	99.93	98.77	99.99	samp_2158
8761-SRC-7	85.59	98.99	41.7628	-83.3297	9/9/20	WE2	85.42	87.68	86.42	99.97	97.78	99.89	samp_4338
8761-SRC-10	74.59	99.03	41.7055	-83.2542	9/9/20	WE12	76.61	72.92	76.82	99.93	97.34	99.93	samp_4341
E20160030	85.08	99.95	41.7048	-83.2544	8/1/16	WE12	85.5	84.83	84.98	99.99	99.87	99.98	samp_451
E20200011	100	99.96	41.7634167	-83.3307	7/6/20	WE2	100	100	100	99.98	99.92	99.99	samp_470
E20200013	100	99.95	41.81916	-83.35921	7/6/20	WE8	100	100	100	99.98	99.9	99.98	samp_471
E20200015	98.12	99.96	41.7034833	-83.255167	7/6/20	WE12	99.24	99.27	95.93	99.97	99.93	99.98	samp_472
E20200047	92.32	99.05	41.76135	-83.33285	8/10/20	WE2	89.82	96.45	91.15	99.72	97.85	99.61	samp_477
E20200089	77.49	97.51	41.7627833	-83.32965	9/9/20	WE2	75.15	77.5	82.37	99.94	95.28	99.97	samp_481
E20200094	80.9	98.33	41.7054833	-83.254233	9/9/20	WE12	83.43	78.79	83.37	99.99	95.9	99.99	samp_484
E20212010	80.65	99.96	41.653433	-83.177667	8/3/21	WE16-WSW	75.02	86.13	80.55	99.99	99.73	99.98	samp_499
LE17	80.87	99.98	41.8265	-83.1893	9/19/16	WE4	78.14	86.39	77.89	99.97	99.99	99.96	samp_509
LE20	99.64	99.96	41.7035	-83.2537	7/15/19	WE12	99.64	99.97	99.3	99.99	99.92	99.97	samp_511

Table S1. Metagenomic read mapping results for LE20-WE8 *sxt* reference contig sequences (Figure S2). Column headers labeled with 'all_ref' represent the aggregated results of all three contig sections.

SampleName	mag	group	group_rpkm	mag_length_mb	id_ref_mag percent	mag_cover percent	rpkm	lat	lon	noaa_site	collection date	clustered_site	study_id	year	month	day	yday	week	sample_id GLAMR
E2014-0110-100LTR	LE20-WE8	D_lem	0.00126443	4.57749	99.05	77.48	0.000873822	41.8268	-83.1947	WE4	9/29/14	WE4	set_17	2014	Sep	29	272	39	samp_40
E20160030	LE20-WE8	D_lem	0.000372243	4.57749	99.6	77.3	0.000261752	41.7048	-83.2544	WE12	8/1/16	WE12	set_35	2016	Aug	1	214	31	samp_451
E20200011	LE20-WE8	D_lem	0.003766677	4.57749	99.91	99.14	0.003120784	41.7634167	-83.3307	WE2	7/6/20	WE2	set_35	2020	Jul	6	188	27	samp_470
E20200013	LE20-WE8	D_lem	0.006929648	4.57749	99.87	99.54	0.005769564	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200015	LE20-WE8	D_lem	0.000941889	4.57749	99.74	93.7	0.000561013	41.7034833	-83.255167	WE12	7/6/20	WE12	set_35	2020	Jul	6	188	27	samp_472
E20200047	LE20-WE8	D_lem	0.000217949	4.57749	99.57	60.86	0.000167038	41.76135	-83.33285	WE2	8/10/20	WE2	set_35	2020	Aug	10	223	32	samp_477
E20210086	LE20-WE8	D_lem	0.000129569	4.57749	99.64	52.31	8.37E-05	41.7026667	-83.25035	WE12	8/16/21	WE12	set_35	2021	Aug	16	228	33	samp_490
E20212010	LE20-WE8	D_lem	0.00022155	4.57749	99.27	78.56	0.000124514	41.653433	-83.177667	WE16-WSW	8/3/21	WE16	set_35	2021	Aug	3	215	31	samp_499
LE17	LE20-WE8	D_lem	0.000458414	4.57749	99.7	63.23	0.000337667	41.8265	-83.1893	WE4	9/19/16	WE4	set_35	2016	Sep	19	263	38	samp_509
LE20	LE20-WE8	D_lem	0.000702374	4.57749	99.84	94.24	0.000560901	41.7035	-83.2537	WE12	7/15/19	WE12	set_35	2019	Jul	15	196	28	samp_511
SV_663	LE20-WE8	D_lem	0.000114083	4.57749	98.39	54.9	5.33E-05	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016
7-18-16 WLE 4	LE20-WE8	D_lem	0.000932758	4.57749	99.89	56.26	0.000750993	41.826	-83.1947	WE4	7/18/16	WE4	set_41	2016	Jul	18	200	29	samp_2042
8-1-16 WLE 8	LE20-WE8	D_lem	0.000215746	4.57749	99.76	96.27	0.000164075	41.8346	-83.3565	WE8	8/1/16	WE8	set_41	2016	Aug	1	214	31	samp_2047
8-7-17 WLE 2	LE20-WE8	D_lem	0.000612251	4.57749	99.25	55.66	0.000130652	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
7-30-18 WLE 12	LE20-WE8	D_lem	0.001231645	4.57749	99.88	58.34	0.00099205	41.7036	-83.2547	WE12	7/30/18	WE12	set_41	2018	Jul	30	211	31	samp_2088
8-13-18 WLE 12	LE20-WE8	D_lem	0.001495235	4.57749	99.9	55.5	0.001199965	41.7032	-83.2536	WE12	8/13/18	WE12	set_41	2018	Aug	13	225	33	samp_2092
8-20-18 WLE 12	LE20-WE8	D_lem	0.001018504	4.57749	99.81	67.67	0.000797739	41.7031	-83.2543	WE12	8/20/18	WE12	set_41	2018	Aug	20	232	34	samp_2096
7-6-20 WLE 8	LE20-WE8	D_lem	0.000773275	4.57749	99.5	89.52	0.000583259	NA	NA	WE8	7/6/20	WE8	set_41	2020	Jul	6	188	27	samp_2141
7-6-20 WLE 12	LE20-WE8	D_lem	0.000441884	4.57749	99.7	75.73	0.000322172	41.7035	-83.2552	WE12	7/6/20	WE12	set_41	2020	Jul	6	188	27	samp_2142
LE14	LE20-WE8	D_lem	0.00016342	4.57749	98.88	58.75	9.70E-05	41.7026	-83.2547	WE12	9/7/21	WE12	set_35	2021	Sep	7	250	36	samp_3935
E2014-0110-100LTR	LE17-WE2	D_lem	0.00126443	4.337595	98.97	73.35	0.000390608	41.8268	-83.1947	WE4	9/29/14	WE4	set_17	2014	Sep	29	272	39	samp_40
E20200011	LE17-WE2	D_lem	0.003766677	4.337595	99.49	68.63	0.000645893	41.7634167	-83.3307	WE2	7/6/20	WE2	set_35	2020	Jul	6	188	27	samp_470
E20200013	LE17-WE2	D_lem	0.006929648	4.337595	99.45	74.44	0.001160084	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200015	LE17-WE2	D_lem	0.000941889	4.337595	99.69	87.33	0.000380876	41.7034833	-83.255167	WE12	7/6/20	WE12	set_35	2020	Jul	6	188	27	samp_472
8-1-16 WLE 8	LE17-WE2	D_lem	0.000215746	4.337595	99.07	50.89	5.17E-05	41.8346	-83.3565	WE8	8/1/16	WE8	set_41	2016	Aug	1	214	31	samp_2047
8-7-17 WLE 2	LE17-WE2	D_lem	0.000612251	4.337595	99.86	99.59	0.000481599	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
E2014-0106-CNA	LE17-WE12	D_circ	0.000289378	4.108029	99.13	60.88	9.98E-05	41.76472222	-83.335	WE2	9/29/14	WE2	set_17	2014	Sep	29	272	39	samp_14
E2014-0106-100LTR	LE17-WE12	D_circ	0.00053057	4.108029	99.13	71.06	0.000186769	41.76472222	-83.335	WE2	9/29/14	WE2	set_17	2014	Sep	29	272	39	samp_15
E2014-0108-100LTR	LE17-WE12	D_circ	0.000371661	4.108029	99.33	67.66	0.000131147	41.7031	-83.2591	WE12	9/29/14	WE12	set_17	2014	Sep	29	272	39	samp_38
E2014-0108-CNA	LE17-WE12	D_circ	0.000726368	4.108029	99.5	84.88	0.00025968	41.7031	-83.2591	WE12	9/29/14	WE12	set_17	2014	Sep	29	272	39	samp_39
E20170045	LE17-WE12	D_circ	0.00205555	4.108029	99.7	73.84	0.000805334	41.764	-83.3296	WE2	8/28/17	WE2	set_35	2017	Aug	28	240	35	samp_456
E20190025	LE17-WE12	D_circ	0.003766682	4.108029	97.75	76.95	0.000828435	41.7624	-83.3306	WE2	8/5/19	WE2	set_35	2019	Aug	5	217	31	samp_462
E20190026	LE17-WE12	D_circ	0.002921565	4.108029	98.32	82.95	0.000704012	41.703	-83.2544	WE12	8/5/19	WE12	set_35	2019	Aug	5	217	31	samp_463
E20190028	LE17-WE12	D_circ	0.001933314	4.108029	97.91	72.12	0.000426497	41.7615	-83.3309	WE2	8/12/19	WE2	set_35	2019	Aug	12	224	32	samp_464
E20190029	LE17-WE12	D_circ	0.001004902	4.108029	97.84	61.63	0.000221754	41.7021	-83.2551	WE12	8/12/19	WE12	set_35	2019	Aug	12	224	32	samp_465
E20200013	LE17-WE12	D_circ	0.002026368	4.108029	97.69	59.15	0.00053967	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200047	LE17-WE12	D_circ	0.001383328	4.108029	99.15	75.11	0.000326585	41.76135	-83.33285	WE2	8/10/20	WE2	set_35	2020	Aug	10	223	32	samp_477
E20200051	LE17-WE12	D_circ	0.003088196	4.108029	99.43	88.02	0.000739122	41.7174667	-83.424367	WE9	8/10/20	WE9	set_35	2020	Aug	10	223	32	samp_479
E20200052	LE17-WE12	D_circ	0.000820946	4.108029	98.29	55.15	0.000157786	41.702	-83.254333	WE12	8/10/20	WE12	set_35	2020	Aug	10	223	32	samp_480
E20200089	LE17-WE12	D_circ	0.001813956	4.108029	99.51	83.59	0.000481451	41.7627833	-83.32965	WE2	9/9/20	WE2	set_35	2020	Sep	9	253	37	samp_481
E20200090	LE17-WE12	D_circ	0.000404428	4.108029	99.69	50.29	0.000148627	41.8275333	-83.196683	WE4	9/9/20	WE4	set_35	2020	Sep	9	253	37	samp_482
E20200094	LE17-WE12	D_circ	0.003140385	4.108029	99.61	86.56	0.000827507	41.7054833	-83.254233	WE12	9/9/20	WE12	set_35	2020	Sep	9	253	37	samp_484
E20200118	LE17-WE12	D_circ	0.012232969	4.108029	99.78	90.57	0.004389272	41.7622833	-83.330583	WE2	9/28/20	WE2	set_35	2020	Sep	28	272	39	samp_485
E20200119	LE17-WE12	D_circ	0.00417101	4.108029	99.59	58.49	0.000818092	41.826	-83.193733	WE4	9/28/20	WE4	set_35	2020	Sep	28	272	39	samp_486
E20200123	LE17-WE12	D_circ	0.000425267	4.108029	99.48	55.17	0.000105752	41.7035167	-83.254533	WE12	9/28/20	WE12	set_35	2020	Sep	28	272	39	samp_487
E20212012	LE17-WE12	D_circ	0.00041014	4.108029	99.18	55.6	0.00012014	41.76265	-83.329	WE2	8/31/21	WE2	set_35	2021	Aug	31	243	35	samp_501
E20212016	LE17-WE12	D_circ	0.002457347	4.108029	99.72	85.63	0.000819222	41.76265	-83.329	WE2	9/21/21	WE2	set_35	2021	Sep	21	264	38	samp_505
E20212018	LE17-WE12	D_circ	0.00482268	4.108029	99.78	89.4	0.001820755	41.7269	-83.42329	WE9	9/21/21	WE9	set_35	2021	Sep	21	264	38	samp_506
E20212019	LE17-WE12	D_circ	0.003604476	4.108029	99.54	86.54	0.00073032	41.70406	-83.25291	WE12	9/21/21	WE12	set_35	2021	Sep	21	264	38	samp_507
6538-PDU-21	LE17-WE12	D_circ	0.000615818	4.108029	99.57	50.37	0.000188881	41.7427	-83.3048	NF	8/9/21	WE2	set_51	2021	Aug	9	221	32	samp_1443
6538-PDU-41	LE17-WE12	D_circ	0.000480277	4.108029	99.62	50.11	0.00017283	41.8326	-83.3512	NF	8/16/21	WE8	set_51	2021	Aug	16	228	33	samp_1463
SV_658	LE17-WE12	D_circ	0.001112076	4.108029	99.33	79.7	0.000150266	41.7628	-83.3283	T10	8/24/22	WE2	set_42	2022	Aug	24	236	34	samp_2011
SV_661	LE17-WE12	D_circ	0.009220734	4.108029	99.77	90.43	0.000958323	41.7626	-83.3291	T10	9/20/22	WE2	set_42	2022	Sep	20	263	38	samp_2014
SV_663	LE17-WE12	D_circ	0.001609055	4.108029	99.53	94.64	0.000337447	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016
sv_639	LE17-WE12	D_circ	0.000241635	4.108029	99.33	58.61	7.76E-05	41.7635	-83.3288	T10	8/17/21	WE2	set_42	2021	Aug	17	229	33	samp_2025
sv_641	LE17-WE12	D_circ	0.000501963	4.108029	99.38	58.39	0.000104708	41.7654	-83.3273	T6	9/15/21	WE2	set_42	2021	Sep	15	258	37	samp_2027
sv_645	LE17-WE12	D_circ	0.011127123	4.108029	99.78	91.03	0.003517096	41.7664	-83.3268	T9	10/5/21	WE2	set_42	2021	Oct	5	278	40	samp_2031
8-15-16 WLE 8	LE17-WE12	D_circ	0.000197562	4.108029	99.38	50.97	5.98E-05	41.834	-83.364	WE8	8/15/16	WE8	set_41	2016	Aug	15	228	33	samp_2051
8-29-16 WLE 4	LE17-WE12	D_circ	0.00158174	4.108029	99.37	51.69	0.00039454	41.8302	-83.1904	WE4	8/29/16	WE4	set_41	2016	Aug	29	242	35	samp_2054
9-12-16 WLE 2	LE17-WE12	D_circ	0.001628439	4.108029	99.16	50.92	0.000477782	41.7632	-83.3315	WE2	9/12/16	WE2	set_41	2016	Sep	12	256	37	samp_2057
9-12-16 WLE 4	LE17-WE12	D_circ	0.005925267	4.108029	99.46	77.42	0.001574753	41.826	-83.1925	WE4	9/12/16	WE4	set_41	2016	Sep	12	256	37	samp_2058
9-12-16 WLE 8	LE17-WE12	D_circ	0.011649241	4.108029	99.45	87.78	0.002386868	41.8322	-83.3589	WE8	9/12/16	WE8	set_41	2016	Sep	12	256	37	samp_2059
10-11-16 WLE 4	LE17-WE12	D_circ	0.008960357	4.108029	99.67	85.87	0.002482582	41.8256	-83.1933	WE4	10/11/16	WE4	set_41	2016	Oct	11	285	41	samp_2062
8-7-17 WLE 2	LE17-WE12	D_circ	0.000412191	4.108029	99.59	81.7	0.000155293	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
8-21-17 WLE 2	LE17-WE12	D_circ	0.00677139	4.108029	99.77	92.95	0.002466219	41.7624	-83.3295	WE2	8/21/17	WE2	set_41	2017	Aug	21	233	34	samp_2077
8-21-17 WLE 4	LE17-WE12	D_circ	0.003743202	4.108029	99.74	79.85	0.001393596	41.8265	-83.1978	WE4	8/21/17	WE4	set_41	2017	Aug	21	233	34	samp_2078
8-21-17 WLE 8	LE17-WE12	D_circ	0.001866488	4.108029	99.66	89.68	0.000616727	41.8337	-83.3583	WE8	8/21/17	WE8	set_41	2017	Aug	21	233	34	samp_2079
7-22-19 WLE 2	LE17-WE12	D_circ	0.000491693	4.108029	98.67	55.39	0.000107213	41.763	-83.3289	WE2	7/22/19	WE2	set_41	2019	Jul	22	203	29	samp_2109
7-29-19 WLE 4	LE17-WE12	D_circ	0.002200539	4.108029	97.42	63.82	0.000466947	41.8261	-83.1946	WE4	7/29/19	WE4	set_41	2019	Jul	29	210	30	samp_2114
7-29-19 WLE 8	LE17-WE12	D_circ	0.000939711	4.108029	98.01	80.15	0.000217266	41.8328	-83.3625	WE8	7/29/19	WE8	set_41	2019	Jul	29	210	30	samp_2115
7-29-19 WLE 12	LE17-WE12	D_circ	0.002278847	4.108029	97.71	53.15	0.000484403	41.7028	-83.2563	WE12	7/29/19	WE12	set_41	2019	Jul	29	210	30	samp_2116
8-19-19 WLE 2	LE17-WE12	D_circ	0.000707756	4.108029	98.42	72.19	0.000120455	41.7621	-83.3303	WE2	8/19/19	WE2	set_41	2019	Aug	19	231	33	samp_2117
7-27-20 WLE 2	LE17-WE12	D_circ	0.000619176	4.108029	98.68	65.24	0.000143566	NA	NA	WE2	7/27/20	WE2	set_41	2020	Jul	27	209	30	samp_2143
8-10-20 WLE 2	LE17-WE12	D_circ	0.002007348	4.108029	99.11	81.83	0.000431152	41.7614	-83.3329	WE2	8/10/20	WE2	set_41	2020	Aug	10	223	32	samp_2147
8-24-20 WLE 2	LE17-WE12	D_circ	0.003888262	4.108029	99.32	83.58	0.000779995	41.7618	-83.331	WE2	8/24/20	WE2	set_41	2020	Aug	24	237	34	samp_2151
8-24-20 WLE 8	LE17-WE12	D_circ	0.068110198	4.108029	99.69	96.56	0.020754642	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
8-24-20 WLE 12	LE17-WE12	D_circ	0.002665071	4.108029	99.48	64.59	0.000842329	41.7031	-83.2557	WE12	8/24/20	WE12	set_41	2020	Aug	24	237	34	samp_2154
8-31-20 WLE 2	LE17-WE12	D_circ	0.01550335	4.108029	99.56	90.7	0.003399507	NA	NA	WE2	8/31/20	WE2	set_41	2020	Aug	31	244	35	samp_2155
8-31-20 WLE 8	LE17-WE12	D_circ	0.032520372	4.108029	99.64	94.12	0.007640776	NA	NA	WE8	8/31/20	WE8	set_41	2020	Aug	31	244	35	samp_2157
8-31-20 WLE 12	LE17-WE12	D_circ	0.012362757	4.108029	99.18	92.74	0.002473435	NA	NA	WE12	8/31/20	WE12	set_41	2020	Aug	31	244	35	samp_2158
9-14-20 WLE 2	LE17-WE12	D_circ	0.010189501	4.108029	99.72	83.24	0.002977486	NA	NA	WE2	9/14/20	WE2	set_41	2020	Sep	14	258	37	samp_2159
9-14-20 WLE 8	LE17-WE12	D_circ	0.00965797	4.108029	99.55	84.09	0.001819016	NA	NA	WE8	9/14/20	WE8	set_41	2020	Sep	14	258	37	samp_2161
9-14-20 WLE 12	LE17-WE12	D_circ	0.004791122	4.108029	99.57	76.56	0.001052071	NA	NA	WE12	9/14/20	WE12	set_41	2020	Sep	14	258	37	samp_2162
TWTPSurge_080219	LE17-WE12	D_circ	0.000507701	4.108029	97.64	57.77	9.18E-05	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2214
LE14	LE17-WE12	D_circ	0.000390983	4.108029	98.68	51.46	7.13E-05	41.7026	-83.2547	WE12	9/7/21	WE12	set_35	2021	Sep	7	250	36	samp_3935
LE2	LE17-WE12	D_circ	0.000844169	4.108029	99.67	87.37	0.00034238	41.7032	-83.254	WE12	9/11/17	WE12	set_35	2017	Sep	11	254	37	samp_3937
doli_UT_Erie_late_2022	LE17-WE12	D_circ	0.206590332	4.108029	99.83	76.95	0.015824039	NA	NA	NA	NA	WE6	set_35	2022	NA	NA	NA	44	samp_4333
8761-SRC-7	LE17-WE12	D_circ	0.00932581	4.108029	99.5	92.5	0.001832705	NA	NA	WE2	9/9/20	WE2	set_41	2020	Sep	9	253	37	samp_4338
8761-SRC-8	LE17-WE12	D_circ	0.002132266	4.108029	99.76	59.74	0.000770955	NA	NA	WE4	9/9/20	WE4	set_41	2020	Sep	9	253	37	samp_4339
8761-SRC-9	LE17-WE12	D_circ	0.007746412	4.108029	99.64	91.16	0.001936794	NA	NA	WE8	9/9/20	WE8	set_41	2020	Sep	9	253	37	samp_4340
8761-SRC-10	LE17-WE12	D_circ	0.0087103	4.108029	99.54	92.41	0.00181684	NA	NA	WE12	9/9/20	WE12	set_41	2020	Sep	9	253	37	samp_4341
8761-SRC-11	LE17-WE12	D_circ	0.008030088	4.108029	99.71	88.69	0.00233067	NA	NA	WE2	10/5/20	WE2	set_41	2020	Oct	5	279	40	samp_4342
8761-SRC-13	LE17-WE12	D_circ	0.005412309	4.108029	99.77	90.62	0.001909352	NA	NA	WE8	10/5/20	WE8	set_41	2020	Oct	5	279	40	samp_4343
8761-SRC-14	LE17-WE12	D_circ	0.003237957	4.108029	99.23	72.42	0.000665492	NA	NA	WE12	10/5/20	WE12	set_41	2020	Oct	5	279	40	samp_4344
8761-SRC-24	LE17-WE12	D_circ	0.009401801	4.108029	99.74	91.84	0.003166475	NA	NA	WE2	9/20/21	WE2	set_41	2021	Sep	20	263	38	samp_4353
8761-SRC-26	LE17-WE12	D_circ	0.001780052	4.108029	99.62	86.46	0.000457272	NA	NA	WE8	9/20/21	WE8	set_41	2021	Sep	20	263	38	samp_4355
8761-SRC-27	LE17-WE12	D_circ	0.005748146	4.108029	99.65	92.06	0.001512621	NA	NA	WE12	9/20/21	WE12	set_41	2021	Sep	20	263	38	samp_4356
8761-SRC-28	LE17-WE12	D_circ	0.0042785	4.108029	99.72	91.77	0.00139226	NA	NA	WE2	10/4/21	WE2	set_41	2021	Oct	4	277	40	samp_4357
8761-SRC-29	LE17-WE12	D_circ	0.004382092	4.108029	99.75	75.38	0.001493858	NA	NA	WE4	10/4/21	WE4	set_41	2021	Oct	4	277	40	samp_4358
8761-SRC-30	LE17-WE12	D_circ	0.027606805	4.108029	99.77	93.07	0.009596407	NA	NA	WE8	10/4/21	WE8	set_41	2021	Oct	4	277	40	samp_4359
8761-SRC-31	LE17-WE12	D_circ	0.008022129	4.108029	99.69	91.62	0.002250121	NA	NA	WE12	10/4/21	WE12	set_41	2021	Oct	4	277	40	samp_4360
8761-SRC-33	LE17-WE12	D_circ	0.005829398	4.108029	99.55	97.55	0.001801059	NA	NA	WE2	7/18/22	WE2	set_41	2022	Jul	18	199	29	samp_4362
8761-SRC-35	LE17-WE12	D_circ	0.001918135	4.108029	99.37	96	0.000569899	NA	NA	WE8	7/18/22	WE8	set_41	2022	Jul	18	199	29	samp_4364
8761-SRC-37	LE17-WE12	D_circ	0.002836836	4.108029	99.58	79.27	0.00037789	NA	NA	WE2	8/17/22	WE2	set_41	2022	Aug	17	229	33	samp_4366
8761-SRC-38	LE17-WE12	D_circ	0.001836272	4.108029	99.45	79.53	0.00029444	NA	NA	WE4	8/17/22	WE4	set_41	2022	Aug	17	229	33	samp_4367
8761-SRC-39	LE17-WE12	D_circ	0.001358964	4.108029	99.34	83.42	0.000206178	NA	NA	WE8	8/17/22	WE8	set_41	2022	Aug	17	229	33	samp_4368
8761-SRC-41	LE17-WE12	D_circ	0.001953536	4.108029	99.48	86.88	0.000229952	NA	NA	WE2	8/29/22	WE2	set_41	2022	Aug	29	241	35	samp_4370
8761-SRC-43	LE17-WE12	D_circ	0.002178549	4.108029	99.41	80.78	0.000258692	NA	NA	WE8	8/29/22	WE8	set_41	2022	Aug	29	241	35	samp_4372
8761-SRC-44	LE17-WE12	D_circ	0.001073596	4.108029	99.27	70.27	0.000114919	NA	NA	WE12	8/29/22	WE12	set_41	2022	Aug	29	241	35	samp_4373
8761-SRC-45	LE17-WE12	D_circ	0.010036444	4.108029	99.68	79.21	0.000984559	NA	NA	WE2	9/12/22	WE2	set_41	2022	Sep	12	255	37	samp_4374
8761-SRC-46	LE17-WE12	D_circ	0.005257714	4.108029	99.36	52.47	0.000687994	NA	NA	WE4	9/12/22	WE4	set_41	2022	Sep	12	255	37	samp_4375
8761-SRC-47	LE17-WE12	D_circ	0.010050091	4.108029	99.63	85.67	0.001000352	NA	NA	WE8	9/12/22	WE8	set_41	2022	Sep	12	255	37	samp_4376
8761-SRC-48	LE17-WE12	D_circ	0.00329929	4.108029	99.3	74.61	0.000404328	NA	NA	WE12	9/12/22	WE12	set_41	2022	Sep	12	255	37	samp_4377
8761-SRC-49	LE17-WE12	D_circ	0.026085088	4.108029	99.75	63.28	0.002381663	NA	NA	WE2	9/30/22	WE2	set_41	2022	Sep	30	273	39	samp_4378
8761-SRC-51	LE17-WE12	D_circ	0.059879046	4.108029	99.79	92.53	0.005802136	NA	NA	WE8	9/30/22	WE8	set_41	2022	Sep	30	273	39	samp_4379
8761-SRC-53	LE17-WE12	D_circ	0.148223532	4.108029	99.78	92	0.012614427	NA	NA	WE2	10/11/22	WE2	set_41	2022	Oct	11	284	41	samp_4381
8761-SRC-55	LE17-WE12	D_circ	0.097418149	4.108029	99.8	91.81	0.008584309	NA	NA	WE8	10/11/22	WE8	set_41	2022	Oct	11	284	41	samp_4382
8761-SRC-56	LE17-WE12	D_circ	0.083122387	4.108029	99.8	92.43	0.007431621	NA	NA	WE12	10/11/22	WE12	set_41	2022	Oct	11	284	41	samp_4383
8761-SRC-57	LE17-WE12	D_circ	0.142999268	4.108029	99.8	91.64	0.011607296	NA	NA	WE2	10/24/22	WE2	set_41	2022	Oct	24	297	43	samp_4384
8761-SRC-58	LE17-WE12	D_circ	0.04847821	4.108029	99.83	83.01	0.004048743	NA	NA	WE4	10/24/22	WE4	set_41	2022	Oct	24	297	43	samp_4385
8761-SRC-59	LE17-WE12	D_circ	0.11365768	4.108029	99.81	89.92	0.009198198	NA	NA	WE8	10/24/22	WE8	set_41	2022	Oct	24	297	43	samp_4386
8761-SRC-60	LE17-WE12	D_circ	0.169975815	4.108029	99.79	89.53	0.013573239	NA	NA	WE12	10/24/22	WE12	set_41	2022	Oct	24	297	43	samp_4387
E2014-0108-CNA	LE20-WE12	D_circ	0.000726368	4.351921	99.35	73.25	0.000136954	41.7031	-83.2591	WE12	9/29/14	WE12	set_17	2014	Sep	29	272	39	samp_39
E20170045	LE20-WE12	D_circ	0.00205555	4.351921	99.49	52.98	0.000380021	41.764	-83.3296	WE2	8/28/17	WE2	set_35	2017	Aug	28	240	35	samp_456
E20190025	LE20-WE12	D_circ	0.003766682	4.351921	97.51	80	0.000833351	41.7624	-83.3306	WE2	8/5/19	WE2	set_35	2019	Aug	5	217	31	samp_462
E20190026	LE20-WE12	D_circ	0.002921565	4.351921	97.95	85.12	0.000594075	41.703	-83.2544	WE12	8/5/19	WE12	set_35	2019	Aug	5	217	31	samp_463
E20190028	LE20-WE12	D_circ	0.001933314	4.351921	97.66	75.88	0.000412481	41.7615	-83.3309	WE2	8/12/19	WE2	set_35	2019	Aug	12	224	32	samp_464
E20190029	LE20-WE12	D_circ	0.001004902	4.351921	97.58	64.93	0.0002188	41.7021	-83.2551	WE12	8/12/19	WE12	set_35	2019	Aug	12	224	32	samp_465
E20200013	LE20-WE12	D_circ	0.002026368	4.351921	97.1	61.77	0.000458443	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200027	LE20-WE12	D_circ	0.00055396	4.351921	97.09	50.42	0.00013168	41.8274167	-83.195617	WE4	7/28/20	WE4	set_35	2020	Jul	28	210	30	samp_474
E20200047	LE20-WE12	D_circ	0.001383328	4.351921	98.85	73.05	0.00028726	41.76135	-83.33285	WE2	8/10/20	WE2	set_35	2020	Aug	10	223	32	samp_477
E20200051	LE20-WE12	D_circ	0.003088196	4.351921	99.19	89.46	0.00062437	41.7174667	-83.424367	WE9	8/10/20	WE9	set_35	2020	Aug	10	223	32	samp_479
E20200052	LE20-WE12	D_circ	0.000820946	4.351921	98.21	60.41	0.000186213	41.702	-83.254333	WE12	8/10/20	WE12	set_35	2020	Aug	10	223	32	samp_480
E20200089	LE20-WE12	D_circ	0.001813956	4.351921	99.31	80.41	0.000347302	41.7627833	-83.32965	WE2	9/9/20	WE2	set_35	2020	Sep	9	253	37	samp_481
E20200094	LE20-WE12	D_circ	0.003140385	4.351921	99.46	83.47	0.000598965	41.7054833	-83.254233	WE12	9/9/20	WE12	set_35	2020	Sep	9	253	37	samp_484
E20200118	LE20-WE12	D_circ	0.012232969	4.351921	99.64	90.19	0.00223258	41.7622833	-83.330583	WE2	9/28/20	WE2	set_35	2020	Sep	28	272	39	samp_485
E20200119	LE20-WE12	D_circ	0.00417101	4.351921	99.51	52.8	0.000662418	41.826	-83.193733	WE4	9/28/20	WE4	set_35	2020	Sep	28	272	39	samp_486
E20212016	LE20-WE12	D_circ	0.002457347	4.351921	99.56	79.38	0.00045433	41.76265	-83.329	WE2	9/21/21	WE2	set_35	2021	Sep	21	264	38	samp_505
E20212018	LE20-WE12	D_circ	0.00482268	4.351921	99.66	85.8	0.000876396	41.7269	-83.42329	WE9	9/21/21	WE9	set_35	2021	Sep	21	264	38	samp_506
E20212019	LE20-WE12	D_circ	0.003604476	4.351921	99.45	83.76	0.000672103	41.70406	-83.25291	WE12	9/21/21	WE12	set_35	2021	Sep	21	264	38	samp_507
SC11_50_Metagenome_2019	LE20-WE12	D_circ	0.000881982	4.351921	98.75	50.62	0.000191759	41.74394092	-83.13597092	NF	8/26/19	WE13	set_36	2019	Aug	26	238	34	samp_1296
SV_658	LE20-WE12	D_circ	0.001112076	4.351921	99.43	82.19	0.000193503	41.7628	-83.3283	T10	8/24/22	WE2	set_42	2022	Aug	24	236	34	samp_2011
SV_661	LE20-WE12	D_circ	0.009220734	4.351921	99.73	91.49	0.000825113	41.7626	-83.3291	T10	9/20/22	WE2	set_42	2022	Sep	20	263	38	samp_2014
SV_663	LE20-WE12	D_circ	0.001609055	4.351921	99.44	90.77	0.000274216	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016
sv_641	LE20-WE12	D_circ	0.000501963	4.351921	99.37	55.19	9.90E-05	41.7654	-83.3273	T6	9/15/21	WE2	set_42	2021	Sep	15	258	37	samp_2027
sv_645	LE20-WE12	D_circ	0.011127123	4.351921	99.67	90.74	0.002075738	41.7664	-83.3268	T9	10/5/21	WE2	set_42	2021	Oct	5	278	40	samp_2031
9-12-16 WLE 4	LE20-WE12	D_circ	0.005925267	4.351921	99.27	73.32	0.001208033	41.826	-83.1925	WE4	9/12/16	WE4	set_41	2016	Sep	12	256	37	samp_2058
9-12-16 WLE 8	LE20-WE12	D_circ	0.011649241	4.351921	99.38	83.49	0.002345979	41.8322	-83.3589	WE8	9/12/16	WE8	set_41	2016	Sep	12	256	37	samp_2059
10-11-16 WLE 4	LE20-WE12	D_circ	0.008960357	4.351921	99.56	80.6	0.001778492	41.8256	-83.1933	WE4	10/11/16	WE4	set_41	2016	Oct	11	285	41	samp_2062

Table S2. (Page 1 of 5) Reads per kilobase million (RPKM) calculations and sample metadata used in Figure 4.

SampleName	mag	group	group_rpkm	mag_length_mb	id_ref_mag percent	mag_cover percent	rpkm	lat	lon	noaa_site	collection date	clustered_site	study_id	year	month	day	yday	week	sample_id GLAMR
8-7-17 WLE 2	LE20-WE12	D_circ	0.000412191	4.351921	99.42	58.47	7.54E-05	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
8-21-17 WLE 2	LE20-WE12	D_circ	0.00677139	4.351921	99.66	90.9	0.001308785	41.7624	-83.3295	WE2	8/21/17	WE2	set_41	2017	Aug	21	233	34	samp_2077
8-21-17 WLE 4	LE20-WE12	D_circ	0.003743202	4.351921	99.63	63.83	0.000720065	41.8265	-83.1978	WE4	8/21/17	WE4	set_41	2017	Aug	21	233	34	samp_2078
8-21-17 WLE 8	LE20-WE12	D_circ	0.001866488	4.351921	99.48	84.9	0.000364749	41.8337	-83.3583	WE8	8/21/17	WE8	set_41	2017	Aug	21	233	34	samp_2079
7-22-19 WLE 2	LE20-WE12	D_circ	0.000491693	4.351921	98.53	60.75	0.000101262	41.763	-83.3289	WE2	7/22/19	WE2	set_41	2019	Jul	22	203	29	samp_2109
7-22-19 WLE 4	LE20-WE12	D_circ	0.000531143	4.351921	97.28	54.78	0.000129564	41.8311	-83.195	WE4	7/22/19	WE4	set_41	2019	Jul	22	203	29	samp_2110
7-29-19 WLE 4	LE20-WE12	D_circ	0.002200539	4.351921	97.26	67.85	0.000519774	41.8261	-83.1946	WE4	7/29/19	WE4	set_41	2019	Jul	29	210	30	samp_2114
7-29-19 WLE 8	LE20-WE12	D_circ	0.000939711	4.351921	97.65	84.53	0.000206517	41.8328	-83.3625	WE8	7/29/19	WE8	set_41	2019	Jul	29	210	30	samp_2115
7-29-19 WLE 12	LE20-WE12	D_circ	0.002278847	4.351921	97.53	57.6	0.000522803	41.7028	-83.2563	WE12	7/29/19	WE12	set_41	2019	Jul	29	210	30	samp_2116
8-19-19 WLE 2	LE20-WE12	D_circ	0.000707756	4.351921	98.57	76.66	0.000148936	41.7621	-83.3303	WE2	8/19/19	WE2	set_41	2019	Aug	19	231	33	samp_2117
8-19-19 WLE 8	LE20-WE12	D_circ	0.000266805	4.351921	97.45	50.13	6.01E-05	41.8336	-83.3598	WE8	8/19/19	WE8	set_41	2019	Aug	19	231	33	samp_2119
7-27-20 WLE 2	LE20-WE12	D_circ	0.000619176	4.351921	98.43	64.78	0.000139861	NA	NA	WE2	7/27/20	WE2	set_41	2020	Jul	27	209	30	samp_2143
8-10-20 WLE 2	LE20-WE12	D_circ	0.002007348	4.351921	98.95	82.55	0.000434908	41.7614	-83.3329	WE2	8/10/20	WE2	set_41	2020	Aug	10	223	32	samp_2147
8-24-20 WLE 2	LE20-WE12	D_circ	0.003888262	4.351921	99.22	85.89	0.000832171	41.7618	-83.331	WE2	8/24/20	WE2	set_41	2020	Aug	24	237	34	samp_2151
8-24-20 WLE 8	LE20-WE12	D_circ	0.068110198	4.351921	99.57	99.2	0.013609278	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
8-24-20 WLE 12	LE20-WE12	D_circ	0.002665071	4.351921	99.14	56.29	0.000565948	41.7031	-83.2557	WE12	8/24/20	WE12	set_41	2020	Aug	24	237	34	samp_2154
8-31-20 WLE 2	LE20-WE12	D_circ	0.01550335	4.351921	99.47	93.87	0.00322325	NA	NA	WE2	8/31/20	WE2	set_41	2020	Aug	31	244	35	samp_2155
8-31-20 WLE 8	LE20-WE12	D_circ	0.032520372	4.351921	99.55	97.68	0.006664626	NA	NA	WE8	8/31/20	WE8	set_41	2020	Aug	31	244	35	samp_2157
8-31-20 WLE 12	LE20-WE12	D_circ	0.012362757	4.351921	99.08	96.76	0.002724943	NA	NA	WE12	8/31/20	WE12	set_41	2020	Aug	31	244	35	samp_2158
9-14-20 WLE 2	LE20-WE12	D_circ	0.010189501	4.351921	99.6	77.03	0.002008196	NA	NA	WE2	9/14/20	WE2	set_41	2020	Sep	14	258	37	samp_2159
9-14-20 WLE 8	LE20-WE12	D_circ	0.00965797	4.351921	99.52	82.86	0.00193478	NA	NA	WE8	9/14/20	WE8	set_41	2020	Sep	14	258	37	samp_2161
9-14-20 WLE 12	LE20-WE12	D_circ	0.004791122	4.351921	99.46	73.49	0.000946782	NA	NA	WE12	9/14/20	WE12	set_41	2020	Sep	14	258	37	samp_2162
TWTPSurge_080219	LE20-WE12	D_circ	0.000507701	4.351921	97.82	65.22	0.000106491	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2214
LE14	LE20-WE12	D_circ	0.000390983	4.351921	98.92	54.88	7.86E-05	41.7026	-83.2547	WE12	9/7/21	WE12	set_35	2021	Sep	7	250	36	samp_3935
LE2	LE20-WE12	D_circ	0.000844169	4.351921	99.45	56.84	0.000138642	41.7032	-83.254	WE12	9/11/17	WE12	set_35	2017	Sep	11	254	37	samp_3937
doli_UT_Erie_late_2022	LE20-WE12	D_circ	0.206590332	4.351921	99.83	70.57	0.011840857	NA	NA	NA	NA	WE6	set_35	2022	NA	NA	NA	44	samp_4333
8761-SRC-7	LE20-WE12	D_circ	0.00932581	4.351921	99.45	96.21	0.001893882	NA	NA	WE2	9/9/20	WE2	set_41	2020	Sep	9	253	37	samp_4338
8761-SRC-9	LE20-WE12	D_circ	0.007746412	4.351921	99.53	92.59	0.001539619	NA	NA	WE8	9/9/20	WE8	set_41	2020	Sep	9	253	37	samp_4340
8761-SRC-10	LE20-WE12	D_circ	0.0087103	4.351921	99.49	94.78	0.001767362	NA	NA	WE12	9/9/20	WE12	set_41	2020	Sep	9	253	37	samp_4341
8761-SRC-11	LE20-WE12	D_circ	0.008030088	4.351921	99.56	85.68	0.00138138	NA	NA	WE2	10/5/20	WE2	set_41	2020	Oct	5	279	40	samp_4342
8761-SRC-13	LE20-WE12	D_circ	0.005412309	4.351921	99.66	88.29	0.000963899	NA	NA	WE8	10/5/20	WE8	set_41	2020	Oct	5	279	40	samp_4343
8761-SRC-14	LE20-WE12	D_circ	0.003237957	4.351921	99.23	88.71	0.000913982	NA	NA	WE12	10/5/20	WE12	set_41	2020	Oct	5	279	40	samp_4344
8761-SRC-24	LE20-WE12	D_circ	0.009401801	4.351921	99.62	90.36	0.00177516	NA	NA	WE2	9/20/21	WE2	set_41	2021	Sep	20	263	38	samp_4353
8761-SRC-26	LE20-WE12	D_circ	0.001780052	4.351921	99.49	83.98	0.000344715	NA	NA	WE8	9/20/21	WE8	set_41	2021	Sep	20	263	38	samp_4355
8761-SRC-27	LE20-WE12	D_circ	0.005748146	4.351921	99.52	92.37	0.001119933	NA	NA	WE12	9/20/21	WE12	set_41	2021	Sep	20	263	38	samp_4356
8761-SRC-28	LE20-WE12	D_circ	0.0042785	4.351921	99.58	92.1	0.000806469	NA	NA	WE2	10/4/21	WE2	set_41	2021	Oct	4	277	40	samp_4357
8761-SRC-29	LE20-WE12	D_circ	0.004382092	4.351921	99.62	61.36	0.000797711	NA	NA	WE4	10/4/21	WE4	set_41	2021	Oct	4	277	40	samp_4358
8761-SRC-30	LE20-WE12	D_circ	0.027606805	4.351921	99.66	95.13	0.004932269	NA	NA	WE8	10/4/21	WE8	set_41	2021	Oct	4	277	40	samp_4359
8761-SRC-31	LE20-WE12	D_circ	0.008022129	4.351921	99.57	92.22	0.001529009	NA	NA	WE12	10/4/21	WE12	set_41	2021	Oct	4	277	40	samp_4360
8761-SRC-33	LE20-WE12	D_circ	0.005829398	4.351921	99.28	95.88	0.000993633	NA	NA	WE2	7/18/22	WE2	set_41	2022	Jul	18	199	29	samp_4362
8761-SRC-35	LE20-WE12	D_circ	0.001918135	4.351921	99.19	92.38	0.000324652	NA	NA	WE8	7/18/22	WE8	set_41	2022	Jul	18	199	29	samp_4364
8761-SRC-37	LE20-WE12	D_circ	0.002836836	4.351921	99.54	76.71	0.000334359	NA	NA	WE2	8/17/22	WE2	set_41	2022	Aug	17	229	33	samp_4366
8761-SRC-38	LE20-WE12	D_circ	0.001836272	4.351921	99.42	77.43	0.00029617	NA	NA	WE4	8/17/22	WE4	set_41	2022	Aug	17	229	33	samp_4367
8761-SRC-39	LE20-WE12	D_circ	0.001358964	4.351921	99.32	80.7	0.00021122	NA	NA	WE8	8/17/22	WE8	set_41	2022	Aug	17	229	33	samp_4368
8761-SRC-41	LE20-WE12	D_circ	0.001953536	4.351921	99.48	88.59	0.000252693	NA	NA	WE2	8/29/22	WE2	set_41	2022	Aug	29	241	35	samp_4370
8761-SRC-43	LE20-WE12	D_circ	0.002178549	4.351921	99.44	82.76	0.000308799	NA	NA	WE8	8/29/22	WE8	set_41	2022	Aug	29	241	35	samp_4372
8761-SRC-44	LE20-WE12	D_circ	0.001073596	4.351921	99.3	75.95	0.000135294	NA	NA	WE12	8/29/22	WE12	set_41	2022	Aug	29	241	35	samp_4373
8761-SRC-45	LE20-WE12	D_circ	0.010036444	4.351921	99.65	79.13	0.000842091	NA	NA	WE2	9/12/22	WE2	set_41	2022	Sep	12	255	37	samp_4374
8761-SRC-46	LE20-WE12	D_circ	0.005257714	4.351921	99.43	58	0.000846208	NA	NA	WE4	9/12/22	WE4	set_41	2022	Sep	12	255	37	samp_4375
8761-SRC-47	LE20-WE12	D_circ	0.010050091	4.351921	99.61	87.08	0.000945632	NA	NA	WE8	9/12/22	WE8	set_41	2022	Sep	12	255	37	samp_4376
8761-SRC-48	LE20-WE12	D_circ	0.00329929	4.351921	99.4	80.77	0.000572722	NA	NA	WE12	9/12/22	WE12	set_41	2022	Sep	12	255	37	samp_4377
8761-SRC-49	LE20-WE12	D_circ	0.026085088	4.351921	99.73	56.63	0.00182166	NA	NA	WE2	9/30/22	WE2	set_41	2022	Sep	30	273	39	samp_4378
8761-SRC-51	LE20-WE12	D_circ	0.059879046	4.351921	99.77	94.14	0.004360997	NA	NA	WE8	9/30/22	WE8	set_41	2022	Sep	30	273	39	samp_4379
8761-SRC-53	LE20-WE12	D_circ	0.148223532	4.351921	99.77	92.88	0.009749807	NA	NA	WE2	10/11/22	WE2	set_41	2022	Oct	11	284	41	samp_4381
8761-SRC-55	LE20-WE12	D_circ	0.097418149	4.351921	99.78	92.36	0.00645302	NA	NA	WE8	10/11/22	WE8	set_41	2022	Oct	11	284	41	samp_4382
8761-SRC-56	LE20-WE12	D_circ	0.083122387	4.351921	99.78	93.63	0.005631753	NA	NA	WE12	10/11/22	WE12	set_41	2022	Oct	11	284	41	samp_4383
8761-SRC-57	LE20-WE12	D_circ	0.142999268	4.351921	99.79	82.48	0.008953285	NA	NA	WE2	10/24/22	WE2	set_41	2022	Oct	24	297	43	samp_4384
8761-SRC-58	LE20-WE12	D_circ	0.04847821	4.351921	99.82	73.81	0.003107366	NA	NA	WE4	10/24/22	WE4	set_41	2022	Oct	24	297	43	samp_4385
8761-SRC-59	LE20-WE12	D_circ	0.11365768	4.351921	99.8	87.68	0.007091485	NA	NA	WE8	10/24/22	WE8	set_41	2022	Oct	24	297	43	samp_4386
8761-SRC-60	LE20-WE12	D_circ	0.169975815	4.351921	99.78	85.34	0.010448414	NA	NA	WE12	10/24/22	WE12	set_41	2022	Oct	24	297	43	samp_4387
E2014-0108-CNA	LE21-WE12	D_circ	0.000726368	3.568765	99.32	81.42	0.000124657	41.7031	-83.2591	WE12	9/29/14	WE12	set_17	2014	Sep	29	272	39	samp_39
E20160044	LE21-WE12	D_circ	0.000337484	3.568765	99.52	78.35	0.00016399	41.8268	-83.192	WE4	8/22/16	WE4	set_35	2016	Aug	22	235	34	samp_455
E20170045	LE21-WE12	D_circ	0.00205555	3.568765	99.37	53.77	0.000318842	41.764	-83.3296	WE2	8/28/17	WE2	set_35	2017	Aug	28	240	35	samp_456
E20190025	LE21-WE12	D_circ	0.003766682	3.568765	98.06	95.41	0.001228248	41.7624	-83.3306	WE2	8/5/19	WE2	set_35	2019	Aug	5	217	31	samp_462
E20190026	LE21-WE12	D_circ	0.002921565	3.568765	98.7	98.17	0.001049597	41.703	-83.2544	WE12	8/5/19	WE12	set_35	2019	Aug	5	217	31	samp_463
E20190028	LE21-WE12	D_circ	0.001933314	3.568765	98.33	94.58	0.000660171	41.7615	-83.3309	WE2	8/12/19	WE2	set_35	2019	Aug	12	224	32	samp_464
E20190029	LE21-WE12	D_circ	0.001004902	3.568765	98.17	84.79	0.000332953	41.7021	-83.2551	WE12	8/12/19	WE12	set_35	2019	Aug	12	224	32	samp_465
E20200013	LE21-WE12	D_circ	0.002026368	3.568765	97.08	68.59	0.000531243	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200027	LE21-WE12	D_circ	0.00055396	3.568765	97.31	62.27	0.000167432	41.8274167	-83.195617	WE4	7/28/20	WE4	set_35	2020	Jul	28	210	30	samp_474
E20200047	LE21-WE12	D_circ	0.001383328	3.568765	99.34	92.86	0.000504761	41.76135	-83.33285	WE2	8/10/20	WE2	set_35	2020	Aug	10	223	32	samp_477
E20200051	LE21-WE12	D_circ	0.003088196	3.568765	99.55	98.94	0.001161193	41.7174667	-83.424367	WE9	8/10/20	WE9	set_35	2020	Aug	10	223	32	samp_479
E20200052	LE21-WE12	D_circ	0.000820946	3.568765	98.92	83.56	0.000323831	41.702	-83.254333	WE12	8/10/20	WE12	set_35	2020	Aug	10	223	32	samp_480
E20200089	LE21-WE12	D_circ	0.001813956	3.568765	99.6	97.23	0.000615984	41.7627833	-83.32965	WE2	9/9/20	WE2	set_35	2020	Sep	9	253	37	samp_481
E20200094	LE21-WE12	D_circ	0.003140385	3.568765	99.71	98.8	0.00110291	41.7054833	-83.254233	WE12	9/9/20	WE12	set_35	2020	Sep	9	253	37	samp_484
E20200118	LE21-WE12	D_circ	0.012232969	3.568765	99.69	99.45	0.002281874	41.7622833	-83.330583	WE2	9/28/20	WE2	set_35	2020	Sep	28	272	39	samp_485
E20200119	LE21-WE12	D_circ	0.00417101	3.568765	99.75	80.59	0.001381752	41.826	-83.193733	WE4	9/28/20	WE4	set_35	2020	Sep	28	272	39	samp_486
E20200123	LE21-WE12	D_circ	0.000425267	3.568765	99.6	72.74	0.000153112	41.7035167	-83.254533	WE12	9/28/20	WE12	set_35	2020	Sep	28	272	39	samp_487
E20212010	LE21-WE12	D_circ	0.000313041	3.568765	99.2	65.71	8.79E-05	41.653433	-83.177667	WE16-WSW	8/3/21	WE16	set_35	2021	Aug	3	215	31	samp_499
E20212011	LE21-WE12	D_circ	0.000422618	3.568765	99.26	66.21	0.000148444	41.653433	-83.177667	WE12-S	8/3/21	WE12	set_35	2021	Aug	3	215	31	samp_500
E20212012	LE21-WE12	D_circ	0.00041014	3.568765	99.36	60.99	0.000118032	41.76265	-83.329	WE2	8/31/21	WE2	set_35	2021	Aug	31	243	35	samp_501
E20212013	LE21-WE12	D_circ	0.000363166	3.568765	99.53	53.57	0.000159627	41.82726	-83.19041	WE4	8/31/21	WE4	set_35	2021	Aug	31	243	35	samp_502
E20212015	LE21-WE12	D_circ	0.00036329	3.568765	99.36	62.96	0.000164278	41.70406	-83.25291	WE12	8/31/21	WE12	set_35	2021	Aug	31	243	35	samp_504
E20212016	LE21-WE12	D_circ	0.002457347	3.568765	99.69	96.67	0.000588686	41.76265	-83.329	WE2	9/21/21	WE2	set_35	2021	Sep	21	264	38	samp_505
E20212018	LE21-WE12	D_circ	0.00482268	3.568765	99.68	98.16	0.000784526	41.7269	-83.42329	WE9	9/21/21	WE9	set_35	2021	Sep	21	264	38	samp_506
E20212019	LE21-WE12	D_circ	0.003604476	3.568765	99.77	99.39	0.001664618	41.70406	-83.25291	WE12	9/21/21	WE12	set_35	2021	Sep	21	264	38	samp_507
LE8	LE21-WE12	D_circ	0.000460002	3.568765	97.92	52.75	0.000132604	41.7028	-83.2563	WE12	7/29/19	WE12	set_35	2019	Jul	29	210	30	samp_508
LE25	LE21-WE12	D_circ	0.000206499	3.568765	99.62	50.16	0.000107405	41.82726	-83.19041	WE4	9/13/21	WE4	set_35	2021	Sep	13	256	37	samp_516
SC50_50_Metagenome	LE21-WE12	D_circ	0.000314661	3.568765	98.81	56.22	0.000121834	41.82489782	-83.21184393	NF	8/24/18	WE4	set_36	2018	Aug	24	236	34	samp_924
SC05_50_Metagenome_2019	LE21-WE12	D_circ	0.000398297	3.568765	98.45	52.63	0.000133988	41.70248614	-83.25536512	NF	8/27/19	WE12	set_36	2019	Aug	27	239	35	samp_994
SC39_50_Metagenome_2019	LE21-WE12	D_circ	0.000588186	3.568765	98.47	56.7	0.000201731	41.73074728	-83.29734036	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1288
SC31_50_Metagenome_2019	LE21-WE12	D_circ	0.000668947	3.568765	98.61	59.01	0.000236578	41.85004718	-83.19432782	NF	8/23/19	WE4	set_36	2019	Aug	23	235	34	samp_1290
SC11_50_Metagenome_2019	LE21-WE12	D_circ	0.000881982	3.568765	99.23	76.36	0.000340386	41.74394092	-83.13597092	NF	8/26/19	WE13	set_36	2019	Aug	26	238	34	samp_1296
SC27_50_Metagenome_2019	LE21-WE12	D_circ	0.000527004	3.568765	98.73	55.69	0.000180068	41.83387126	-83.35776843	NF	8/23/19	WE8	set_36	2019	Aug	23	235	34	samp_1313
SC41_50_Metagenome_2019	LE21-WE12	D_circ	0.00079012	3.568765	98.8	69.04	0.000288466	41.79870352	-83.34393421	NF	8/20/19	WE2	set_36	2019	Aug	20	232	34	samp_1317
SC39_50_Metagenome_2019	LE21-WE12	D_circ	0.000589066	3.568765	98.47	56.66	0.000201772	41.73074728	-83.29734036	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1326
6538-PDU-21	LE21-WE12	D_circ	0.000615818	3.568765	99.62	50.83	0.000164479	41.7427	-83.3048	NF	8/9/21	WE2	set_51	2021	Aug	9	221	32	samp_1443
SV_658	LE21-WE12	D_circ	0.001112076	3.568765	99.76	98.83	0.000466807	41.7628	-83.3283	T10	8/24/22	WE2	set_42	2022	Aug	24	236	34	samp_2011
SV_661	LE21-WE12	D_circ	0.009220734	3.568765	99.82	99.67	0.001183902	41.7626	-83.3291	T10	9/20/22	WE2	set_42	2022	Sep	20	263	38	samp_2014
SV_663	LE21-WE12	D_circ	0.001609055	3.568765	99.7	99.55	0.000500915	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016
sv_639	LE21-WE12	D_circ	0.000241635	3.568765	99.37	58.45	7.06E-05	41.7635	-83.3288	T10	8/17/21	WE2	set_42	2021	Aug	17	229	33	samp_2025
sv_641	LE21-WE12	D_circ	0.000501963	3.568765	99.7	86.82	0.000221953	41.7654	-83.3273	T6	9/15/21	WE2	set_42	2021	Sep	15	258	37	samp_2027
sv_642	LE21-WE12	D_circ	0.00057196	3.568765	99.63	74.29	0.000232931	41.7603	-83.329	T10	9/15/21	WE2	set_42	2021	Sep	15	258	37	samp_2028
sv_645	LE21-WE12	D_circ	0.011127123	3.568765	99.74	99.55	0.00235927	41.7664	-83.3268	T9	10/5/21	WE2	set_42	2021	Oct	5	278	40	samp_2031
8-15-16 WLE 8	LE21-WE12	D_circ	0.000197562	3.568765	99.45	52.01	5.16E-05	41.834	-83.364	WE8	8/15/16	WE8	set_41	2016	Aug	15	228	33	samp_2051
8-29-16 WLE 4	LE21-WE12	D_circ	0.00158174	3.568765	99.51	66.84	0.00055023	41.8302	-83.1904	WE4	8/29/16	WE4	set_41	2016	Aug	29	242	35	samp_2054
9-12-16 WLE 4	LE21-WE12	D_circ	0.005925267	3.568765	99.51	91.69	0.001958217	41.826	-83.1925	WE4	9/12/16	WE4	set_41	2016	Sep	12	256	37	samp_2058
9-12-16 WLE 8	LE21-WE12	D_circ	0.011649241	3.568765	99.68	98.83	0.005124286	41.8322	-83.3589	WE8	9/12/16	WE8	set_41	2016	Sep	12	256	37	samp_2059
10-11-16 WLE 4	LE21-WE12	D_circ	0.008960357	3.568765	99.72	96.98	0.002882887	41.8256	-83.1933	WE4	10/11/16	WE4	set_41	2016	Oct	11	285	41	samp_2062
8-7-17 WLE 2	LE21-WE12	D_circ	0.000412191	3.568765	99.37	59.08	7.02E-05	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
8-21-17 WLE 2	LE21-WE12	D_circ	0.00677139	3.568765	99.7	99.45	0.001301243	41.7624	-83.3295	WE2	8/21/17	WE2	set_41	2017	Aug	21	233	34	samp_2077
8-21-17 WLE 4	LE21-WE12	D_circ	0.003743202	3.568765	99.63	70.68	0.000679946	41.8265	-83.1978	WE4	8/21/17	WE4	set_41	2017	Aug	21	233	34	samp_2078
8-21-17 WLE 8	LE21-WE12	D_circ	0.001866488	3.568765	99.59	97.15	0.000448788	41.8337	-83.3583	WE8	8/21/17	WE8	set_41	2017	Aug	21	233	34	samp_2079
7-22-19 WLE 2	LE21-WE12	D_circ	0.000491693	3.568765	99.13	86.12	0.000187102	41.763	-83.3289	WE2	7/22/19	WE2	set_41	2019	Jul	22	203	29	samp_2109
7-22-19 WLE 4	LE21-WE12	D_circ	0.000531143	3.568765	97.46	65.54	0.000154694	41.8311	-83.195	WE4	7/22/19	WE4	set_41	2019	Jul	22	203	29	samp_2110
7-22-19 WLE 12	LE21-WE12	D_circ	0.008482869	3.568765	98.22	52.05	0.003022905	41.706	-83.2518	WE12	7/22/19	WE12	set_41	2019	Jul	22	203	29	samp_2112
7-29-19 WLE 2	LE21-WE12	D_circ	0.000872867	3.568765	99.2	69.32	0.000311358	41.7602	-83.3336	WE2	7/29/19	WE2	set_41	2019	Jul	29	210	30	samp_2113
7-29-19 WLE 4	LE21-WE12	D_circ	0.002200539	3.568765	97.46	80.82	0.000637369	41.8261	-83.1946	WE4	7/29/19	WE4	set_41	2019	Jul	29	210	30	samp_2114
7-29-19 WLE 8	LE21-WE12	D_circ	0.000939711	3.568765	98.11	96.85	0.000290425	41.8328	-83.3625	WE8	7/29/19	WE8	set_41	2019	Jul	29	210	30	samp_2115
7-29-19 WLE 12	LE21-WE12	D_circ	0.002278847	3.568765	98.01	74	0.000731236	41.7028	-83.2563	WE12	7/29/19	WE12	set_41	2019	Jul	29	210	30	samp_2116
8-19-19 WLE 2	LE21-WE12	D_circ	0.000707756	3.568765	99.22	96.72	0.000305128	41.7621	-83.3303	WE2	8/19/19	WE2	set_41	2019	Aug	19	231	33	samp_2117
8-19-19 WLE 8	LE21-WE12	D_circ	0.000266805	3.568765	97.88	65.07	8.37E-05	41.8336	-83.3598	WE8	8/19/19	WE8	set_41	2019	Aug	19	231	33	samp_2119
8-19-19 WLE 12	LE21-WE12	D_circ	0.000969894	3.568765	98.94	75.64	0.000407771	41.7029	-83.2539	WE12	8/19/19	WE12	set_41	2019	Aug	19	231	33	samp_2120
8-28-19 WLE 12	LE21-WE12	D_circ	0.001159543	3.568765	99.08	53.36	0.000449273	41.7424	-83.138	WE12	8/28/19	WE12	set_41	2019	Aug	28	240	35	samp_2134
7-6-20 WLE 8	LE21-WE12	D_circ	0.00066562	3.568765	97.37	54.34	0.000132209	NA	NA	WE8	7/6/20	WE8	set_41	2020	Jul	6	188	27	samp_2141
7-27-20 WLE 2	LE21-WE12	D_circ	0.000619176	3.568765	98.92	82.76	0.000203608	NA	NA	WE2	7/27/20	WE2	set_41	2020	Jul	27	209	30	samp_2143
7-27-20 WLE 4	LE21-WE12	D_circ	0.000812347	3.568765	97.26	55.64	0.000230254	NA	NA	WE4	7/27/20	WE4	set_41	2020	Jul	27	209	30	samp_2144
7-27-20 WLE 8	LE21-WE12	D_circ	0.000261384	3.568765	98.09	51.34	7.90E-05	NA	NA	WE8	7/27/20	WE8	set_41	2020	Jul	27	209	30	samp_2145
8-10-20 WLE 2	LE21-WE12	D_circ	0.002007348	3.568765	99.4	97.35	0.000787824	41.7614	-83.3329	WE2	8/10/20	WE2	set_41	2020	Aug	10	223	32	samp_2147
8-24-20 WLE 2	LE21-WE12	D_circ	0.003888262	3.568765	99.59	98.62	0.001661053	41.7618	-83.331	WE2	8/24/20	WE2	set_41	2020	Aug	24	237	34	samp_2151
8-24-20 WLE 8	LE21-WE12	D_circ	0.068110198	3.568765	99.74	99.96	0.019448503	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
8-24-20 WLE 12	LE21-WE12	D_circ	0.002665071	3.568765	99.22	66.03	0.000656179	41.7031	-83.2557	WE12	8/24/20	WE12	set_41	2020	Aug	24	237	34	samp_2154
8-31-20 WLE 2	LE21-WE12	D_circ	0.01550335	3.568765	99.73	99.69	0.006425574	NA	NA	WE2	8/31/20	WE2	set_41	2020	Aug	31	244	35	samp_2155
8-31-20 WLE 8	LE21-WE12	D_circ	0.032520372	3.568765	99.78	99.91	0.012949433	NA	NA	WE8	8/31/20	WE8	set_41	2020	Aug	31	244	35	samp_2157
8-31-20 WLE 12	LE21-WE12	D_circ	0.012362757	3.568765	99.48	99.76	0.005126116	NA	NA	WE12	8/31/20	WE12	set_41	2020	Aug	31	244	35	samp_2158
9-14-20 WLE 2	LE21-WE12	D_circ	0.010189501	3.568765	99.74	95.16	0.002921193	NA	NA	WE2	9/14/20	WE2	set_41	2020	Sep	14	258	37	samp_2159
9-14-20 WLE 8	LE21-WE12	D_circ	0.00965797	3.568765	99.78	98.96	0.004398392	NA	NA	WE8	9/14/20	WE8	set_41	2020	Sep	14	258	37	samp_2161
9-14-20 WLE 12	LE21-WE12	D_circ	0.004791122	3.568765	99.71	95.4	0.001832099	NA	NA	WE12	9/14/20	WE12	set_41	2020	Sep	14	258	37	samp_2162
TWTPSurge_080219	LE21-WE12	D_circ	0.000507701	3.568765	98.56	89.15	0.000186683	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2214
TWTPIntake_080219	LE21-WE12	D_circ	0.000278628	3.568765	98.27	59.98	8.25E-05	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2215
LE14	LE21-WE12	D_circ	0.000390983	3.568765	99.56	89.93	0.000180158	41.7026	-83.2547	WE12	9/7/21	WE12	set_35	2021	Sep	7	250	36	samp_3935
LE2	LE21-WE12	D_circ	0.000844169	3.568765	99.62	76.96	0.000207816	41.7032	-83.254	WE12	9/11/17	WE12	set_35	2017	Sep	11	254	37	samp_3937
doli_UT_Erie_late_2022	LE21-WE12	D_circ	0.206590332	3.568765	99.83	72.78	0.012356452	NA	NA	NA	NA	WE6	set_35	2022	NA	NA	NA	44	samp_4333
8761-SRC-7	LE21-WE12	D_circ	0.00932581	3.568765	99.74	99.85	0.004191493	NA	NA	WE2	9/9/20	WE2	set_41	2020	Sep	9	253	37	samp_4338
8761-SRC-9	LE21-WE12	D_circ	0.007746412	3.568765	99.73	99.72	0.002765736	NA	NA	WE8	9/9/20	WE8	set_41	2020	Sep	9	253	37	samp_4340
8761-SRC-10	LE21-WE12	D_circ	0.0087103	3.568765	99.75	99.87	0.003779341	NA	NA	WE12	9/9/20	WE12	set_41	2020	Sep	9	253	37	samp_4341
8761-SRC-11	LE21-WE12	D_circ	0.008030088	3.568765	99.61	96.81	0.001538395	NA	NA	WE2	10/5/20	WE2	set_41	2020	Oct	5	279	40	samp_4342
8761-SRC-13	LE21-WE12	D_circ	0.005412309	3.568765	99.65	98.06	0.000834322	NA	NA	WE8	10/5/20	WE8	set_41	2020	Oct	5	279	40	samp_4343
8761-SRC-14	LE21-WE12	D_circ	0.003237957	3.568765	99.3	90.8	0.000985116	NA	NA	WE12	10/5/20	WE12	set_41	2020	Oct	5	279	40	samp_4344
8761-SRC-20	LE21-WE12	D_circ	0.000114358	3.568765	99.17	51.19	3.58E-05	NA	NA	WE2	8/30/21	WE2	set_41	2021	Aug	30	242	35	samp_4349

Table S2. (Page 2 of 5) Reads per kilobase million (RPKM) calculations and sample metadata used in Figure 4.

SampleName	mag	group	group_rpkm	mag_length_mb	id_ref_mag percent	mag_cover percent	rpkm	lat	lon	noaa_site	collection date	clustered_site	study_id	year	month	day	yday	week	sample_id GLAMR
8761-SRC-22	LE21-WE12	D_circ	0.000173264	3.568765	99.5	78.77	6.85E-05	NA	NA	WE8	8/30/21	WE8	set_41	2021	Aug	30	242	35	samp_4351
8761-SRC-23	LE21-WE12	D_circ	0.000188204	3.568765	99.23	60.58	6.78E-05	NA	NA	WE12	8/30/21	WE12	set_41	2021	Aug	30	242	35	samp_4352
8761-SRC-24	LE21-WE12	D_circ	0.009401801	3.568765	99.69	99.42	0.002085465	NA	NA	WE2	9/20/21	WE2	set_41	2021	Sep	20	263	38	samp_4353
8761-SRC-25	LE21-WE12	D_circ	0.00050037	3.568765	99.63	76.34	0.000236695	NA	NA	WE4	9/20/21	WE4	set_41	2021	Sep	20	263	38	samp_4354
8761-SRC-26	LE21-WE12	D_circ	0.001780052	3.568765	99.71	98.75	0.00064797	NA	NA	WE8	9/20/21	WE8	set_41	2021	Sep	20	263	38	samp_4355
8761-SRC-27	LE21-WE12	D_circ	0.005748146	3.568765	99.72	99.74	0.00200959	NA	NA	WE12	9/20/21	WE12	set_41	2021	Sep	20	263	38	samp_4356
8761-SRC-28	LE21-WE12	D_circ	0.0042785	3.568765	99.68	99.67	0.001060734	NA	NA	WE2	10/4/21	WE2	set_41	2021	Oct	4	277	40	samp_4357
8761-SRC-29	LE21-WE12	D_circ	0.004382092	3.568765	99.66	75.7	0.000942687	NA	NA	WE4	10/4/21	WE4	set_41	2021	Oct	4	277	40	samp_4358
8761-SRC-30	LE21-WE12	D_circ	0.027606805	3.568765	99.7	99.9	0.004933266	NA	NA	WE8	10/4/21	WE8	set_41	2021	Oct	4	277	40	samp_4359
8761-SRC-31	LE21-WE12	D_circ	0.008022129	3.568765	99.72	99.71	0.002405091	NA	NA	WE12	10/4/21	WE12	set_41	2021	Oct	4	277	40	samp_4360
8761-SRC-33	LE21-WE12	D_circ	0.005829398	3.568765	99.4	99.79	0.001242405	NA	NA	WE2	7/18/22	WE2	set_41	2022	Jul	18	199	29	samp_4362
8761-SRC-35	LE21-WE12	D_circ	0.001918135	3.568765	99.56	99.67	0.00061266	NA	NA	WE8	7/18/22	WE8	set_41	2022	Jul	18	199	29	samp_4364
8761-SRC-37	LE21-WE12	D_circ	0.002836836	3.568765	99.74	96.95	0.000626485	NA	NA	WE2	8/17/22	WE2	set_41	2022	Aug	17	229	33	samp_4366
8761-SRC-38	LE21-WE12	D_circ	0.001836272	3.568765	99.73	97.92	0.000655981	NA	NA	WE4	8/17/22	WE4	set_41	2022	Aug	17	229	33	samp_4367
8761-SRC-39	LE21-WE12	D_circ	0.001358964	3.568765	99.64	98.54	0.000477768	NA	NA	WE8	8/17/22	WE8	set_41	2022	Aug	17	229	33	samp_4368
8761-SRC-41	LE21-WE12	D_circ	0.001953536	3.568765	99.76	99.46	0.000544383	NA	NA	WE2	8/29/22	WE2	set_41	2022	Aug	29	241	35	samp_4370
8761-SRC-43	LE21-WE12	D_circ	0.002178549	3.568765	99.76	99	0.000725817	NA	NA	WE8	8/29/22	WE8	set_41	2022	Aug	29	241	35	samp_4372
8761-SRC-44	LE21-WE12	D_circ	0.001073596	3.568765	99.71	97.83	0.000299951	NA	NA	WE12	8/29/22	WE12	set_41	2022	Aug	29	241	35	samp_4373
8761-SRC-45	LE21-WE12	D_circ	0.010036444	3.568765	99.76	96.98	0.001222558	NA	NA	WE2	9/12/22	WE2	set_41	2022	Sep	12	255	37	samp_4374
8761-SRC-46	LE21-WE12	D_circ	0.005257714	3.568765	99.75	91.53	0.002047036	NA	NA	WE4	9/12/22	WE4	set_41	2022	Sep	12	255	37	samp_4375
8761-SRC-47	LE21-WE12	D_circ	0.010050091	3.568765	99.77	99.26	0.001620955	NA	NA	WE8	9/12/22	WE8	set_41	2022	Sep	12	255	37	samp_4376
8761-SRC-48	LE21-WE12	D_circ	0.00329929	3.568765	99.75	98.65	0.00144835	NA	NA	WE12	9/12/22	WE12	set_41	2022	Sep	12	255	37	samp_4377
8761-SRC-49	LE21-WE12	D_circ	0.026085088	3.568765	99.75	65.15	0.002026596	NA	NA	WE2	9/30/22	WE2	set_41	2022	Sep	30	273	39	samp_4378
8761-SRC-51	LE21-WE12	D_circ	0.059879046	3.568765	99.78	99.79	0.004633158	NA	NA	WE8	9/30/22	WE8	set_41	2022	Sep	30	273	39	samp_4379
8761-SRC-53	LE21-WE12	D_circ	0.148223532	3.568765	99.76	99.19	0.010191757	NA	NA	WE2	10/11/22	WE2	set_41	2022	Oct	11	284	41	samp_4381
8761-SRC-55	LE21-WE12	D_circ	0.097418149	3.568765	99.78	99.17	0.006677353	NA	NA	WE8	10/11/22	WE8	set_41	2022	Oct	11	284	41	samp_4382
8761-SRC-56	LE21-WE12	D_circ	0.083122387	3.568765	99.79	99.78	0.005997857	NA	NA	WE12	10/11/22	WE12	set_41	2022	Oct	11	284	41	samp_4383
8761-SRC-57	LE21-WE12	D_circ	0.142999268	3.568765	99.79	84.34	0.009371473	NA	NA	WE2	10/24/22	WE2	set_41	2022	Oct	24	297	43	samp_4384
8761-SRC-58	LE21-WE12	D_circ	0.04847821	3.568765	99.82	76.31	0.003226213	NA	NA	WE4	10/24/22	WE4	set_41	2022	Oct	24	297	43	samp_4385
8761-SRC-59	LE21-WE12	D_circ	0.11365768	3.568765	99.8	93.82	0.007414863	NA	NA	WE8	10/24/22	WE8	set_41	2022	Oct	24	297	43	samp_4386
8761-SRC-60	LE21-WE12	D_circ	0.169975815	3.568765	99.77	88.03	0.010907274	NA	NA	WE12	10/24/22	WE12	set_41	2022	Oct	24	297	43	samp_4387
E2014-0106-100LTR	LE22-WE12	D_circ	0.00053057	4.773271	98.39	55.12	0.000181114	41.76472222	-83.335	WE2	9/29/14	WE2	set_17	2014	Sep	29	272	39	samp_15
E2014-0108-100LTR	LE22-WE12	D_circ	0.000371661	4.773271	98.55	53.66	0.000121579	41.7031	-83.2591	WE12	9/29/14	WE12	set_17	2014	Sep	29	272	39	samp_38
E2014-0108-CNA	LE22-WE12	D_circ	0.000726368	4.773271	99.12	76.07	0.000205077	41.7031	-83.2591	WE12	9/29/14	WE12	set_17	2014	Sep	29	272	39	samp_39
E20170045	LE22-WE12	D_circ	0.00205555	4.773271	99.3	61.01	0.000551354	41.764	-83.3296	WE2	8/28/17	WE2	set_35	2017	Aug	28	240	35	samp_456
E20190025	LE22-WE12	D_circ	0.003766682	4.773271	97.04	76.93	0.000876648	41.7624	-83.3306	WE2	8/5/19	WE2	set_35	2019	Aug	5	217	31	samp_462
E20190026	LE22-WE12	D_circ	0.002921565	4.773271	97.34	81.14	0.000573882	41.703	-83.2544	WE12	8/5/19	WE12	set_35	2019	Aug	5	217	31	samp_463
E20190028	LE22-WE12	D_circ	0.001933314	4.773271	97.06	72.57	0.000434165	41.7615	-83.3309	WE2	8/12/19	WE2	set_35	2019	Aug	12	224	32	samp_464
E20190029	LE22-WE12	D_circ	0.001004902	4.773271	97	63.18	0.000231395	41.7021	-83.2551	WE12	8/12/19	WE12	set_35	2019	Aug	12	224	32	samp_465
E20200013	LE22-WE12	D_circ	0.002026368	4.773271	96.88	63	0.000497013	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200027	LE22-WE12	D_circ	0.00055396	4.773271	96.83	51.36	0.000142651	41.8274167	-83.195617	WE4	7/28/20	WE4	set_35	2020	Jul	28	210	30	samp_474
E20200047	LE22-WE12	D_circ	0.001383328	4.773271	98.61	69.23	0.000264723	41.76135	-83.33285	WE2	8/10/20	WE2	set_35	2020	Aug	10	223	32	samp_477
E20200051	LE22-WE12	D_circ	0.003088196	4.773271	98.99	84.03	0.000563511	41.7174667	-83.424367	WE9	8/10/20	WE9	set_35	2020	Aug	10	223	32	samp_479
E20200052	LE22-WE12	D_circ	0.000820946	4.773271	97.64	54.5	0.000153116	41.702	-83.254333	WE12	8/10/20	WE12	set_35	2020	Aug	10	223	32	samp_480
E20200089	LE22-WE12	D_circ	0.001813956	4.773271	99.16	80.48	0.000369219	41.7627833	-83.32965	WE2	9/9/20	WE2	set_35	2020	Sep	9	253	37	samp_481
E20200094	LE22-WE12	D_circ	0.003140385	4.773271	99.31	82.16	0.000611002	41.7054833	-83.254233	WE12	9/9/20	WE12	set_35	2020	Sep	9	253	37	samp_484
E20200118	LE22-WE12	D_circ	0.012232969	4.773271	99.57	99.34	0.003329243	41.7622833	-83.330583	WE2	9/28/20	WE2	set_35	2020	Sep	28	272	39	samp_485
E20200119	LE22-WE12	D_circ	0.00417101	4.773271	99.67	76.27	0.001308747	41.826	-83.193733	WE4	9/28/20	WE4	set_35	2020	Sep	28	272	39	samp_486
E20212016	LE22-WE12	D_circ	0.002457347	4.773271	99.48	83.75	0.000595109	41.76265	-83.329	WE2	9/21/21	WE2	set_35	2021	Sep	21	264	38	samp_505
E20212018	LE22-WE12	D_circ	0.00482268	4.773271	99.56	97.93	0.001341002	41.7269	-83.42329	WE9	9/21/21	WE9	set_35	2021	Sep	21	264	38	samp_506
E20212019	LE22-WE12	D_circ	0.003604476	4.773271	99.26	80.89	0.000537435	41.70406	-83.25291	WE12	9/21/21	WE12	set_35	2021	Sep	21	264	38	samp_507
SV_658	LE22-WE12	D_circ	0.001112076	4.773271	99.54	98.81	0.000301499	41.7628	-83.3283	T10	8/24/22	WE2	set_42	2022	Aug	24	236	34	samp_2011
SV_661	LE22-WE12	D_circ	0.009220734	4.773271	99.92	99.96	0.006253396	41.7626	-83.3291	T10	9/20/22	WE2	set_42	2022	Sep	20	263	38	samp_2014
SV_663	LE22-WE12	D_circ	0.001609055	4.773271	99.54	99.74	0.000496477	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016
sv_645	LE22-WE12	D_circ	0.011127123	4.773271	99.63	99.76	0.003175019	41.7664	-83.3268	T9	10/5/21	WE2	set_42	2021	Oct	5	278	40	samp_2031
9-12-16 WLE 4	LE22-WE12	D_circ	0.005925267	4.773271	99.06	68.37	0.001184263	41.826	-83.1925	WE4	9/12/16	WE4	set_41	2016	Sep	12	256	37	samp_2058
9-12-16 WLE 8	LE22-WE12	D_circ	0.011649241	4.773271	99.06	76.19	0.001792107	41.8322	-83.3589	WE8	9/12/16	WE8	set_41	2016	Sep	12	256	37	samp_2059
10-11-16 WLE 4	LE22-WE12	D_circ	0.008960357	4.773271	99.37	75.05	0.001816396	41.8256	-83.1933	WE4	10/11/16	WE4	set_41	2016	Oct	11	285	41	samp_2062
8-7-17 WLE 2	LE22-WE12	D_circ	0.000412191	4.773271	99.14	64.09	0.000111335	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
8-21-17 WLE 2	LE22-WE12	D_circ	0.00677139	4.773271	99.51	88.96	0.001695142	41.7624	-83.3295	WE2	8/21/17	WE2	set_41	2017	Aug	21	233	34	samp_2077
8-21-17 WLE 4	LE22-WE12	D_circ	0.003743202	4.773271	99.46	66.75	0.000949595	41.8265	-83.1978	WE4	8/21/17	WE4	set_41	2017	Aug	21	233	34	samp_2078
8-21-17 WLE 8	LE22-WE12	D_circ	0.001866488	4.773271	99.3	81.15	0.000436224	41.8337	-83.3583	WE8	8/21/17	WE8	set_41	2017	Aug	21	233	34	samp_2079
7-22-19 WLE 2	LE22-WE12	D_circ	0.000491693	4.773271	97.75	54.93	9.61E-05	41.763	-83.3289	WE2	7/22/19	WE2	set_41	2019	Jul	22	203	29	samp_2109
7-22-19 WLE 4	LE22-WE12	D_circ	0.000531143	4.773271	97.08	57.54	0.000147654	41.8311	-83.195	WE4	7/22/19	WE4	set_41	2019	Jul	22	203	29	samp_2110
7-29-19 WLE 4	LE22-WE12	D_circ	0.002200539	4.773271	96.95	67.92	0.000576449	41.8261	-83.1946	WE4	7/29/19	WE4	set_41	2019	Jul	29	210	30	samp_2114
7-29-19 WLE 8	LE22-WE12	D_circ	0.000939711	4.773271	97.24	83.06	0.000225502	41.8328	-83.3625	WE8	7/29/19	WE8	set_41	2019	Jul	29	210	30	samp_2115
7-29-19 WLE 12	LE22-WE12	D_circ	0.002278847	4.773271	97.11	56.49	0.000540404	41.7028	-83.2563	WE12	7/29/19	WE12	set_41	2019	Jul	29	210	30	samp_2116
8-19-19 WLE 2	LE22-WE12	D_circ	0.000707756	4.773271	97.8	73.23	0.000133238	41.7621	-83.3303	WE2	8/19/19	WE2	set_41	2019	Aug	19	231	33	samp_2117
8-19-19 WLE 8	LE22-WE12	D_circ	0.000266805	4.773271	97.11	53.95	7.40E-05	41.8336	-83.3598	WE8	8/19/19	WE8	set_41	2019	Aug	19	231	33	samp_2119
7-6-20 WLE 8	LE22-WE12	D_circ	0.00066562	4.773271	97.37	57.13	0.000174791	NA	NA	WE8	7/6/20	WE8	set_41	2020	Jul	6	188	27	samp_2141
7-27-20 WLE 2	LE22-WE12	D_circ	0.000619176	4.773271	97.93	59.44	0.000132141	NA	NA	WE2	7/27/20	WE2	set_41	2020	Jul	27	209	30	samp_2143
8-10-20 WLE 2	LE22-WE12	D_circ	0.002007348	4.773271	98.54	75.64	0.000353464	41.7614	-83.3329	WE2	8/10/20	WE2	set_41	2020	Aug	10	223	32	samp_2147
8-24-20 WLE 2	LE22-WE12	D_circ	0.003888262	4.773271	98.87	77.74	0.000615043	41.7618	-83.331	WE2	8/24/20	WE2	set_41	2020	Aug	24	237	34	samp_2151
8-24-20 WLE 8	LE22-WE12	D_circ	0.068110198	4.773271	99.38	99.26	0.014297775	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
8-24-20 WLE 12	LE22-WE12	D_circ	0.002665071	4.773271	98.98	55.63	0.000600616	41.7031	-83.2557	WE12	8/24/20	WE12	set_41	2020	Aug	24	237	34	samp_2154
8-31-20 WLE 2	LE22-WE12	D_circ	0.01550335	4.773271	99.21	87.49	0.002455019	NA	NA	WE2	8/31/20	WE2	set_41	2020	Aug	31	244	35	samp_2155
8-31-20 WLE 8	LE22-WE12	D_circ	0.032520372	4.773271	99.33	96.48	0.005265536	NA	NA	WE8	8/31/20	WE8	set_41	2020	Aug	31	244	35	samp_2157
8-31-20 WLE 12	LE22-WE12	D_circ	0.012362757	4.773271	98.66	92.77	0.002038263	NA	NA	WE12	8/31/20	WE12	set_41	2020	Aug	31	244	35	samp_2158
9-14-20 WLE 2	LE22-WE12	D_circ	0.010189501	4.773271	99.52	80.28	0.002282627	NA	NA	WE2	9/14/20	WE2	set_41	2020	Sep	14	258	37	samp_2159
9-14-20 WLE 8	LE22-WE12	D_circ	0.00965797	4.773271	99.34	84.36	0.001505782	NA	NA	WE8	9/14/20	WE8	set_41	2020	Sep	14	258	37	samp_2161
9-14-20 WLE 12	LE22-WE12	D_circ	0.004791122	4.773271	99.38	78.69	0.000960169	NA	NA	WE12	9/14/20	WE12	set_41	2020	Sep	14	258	37	samp_2162
TWTPSurge_080219	LE22-WE12	D_circ	0.000507701	4.773271	96.81	60.46	0.000122733	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2214
LE2	LE22-WE12	D_circ	0.000844169	4.773271	99.09	56.1	0.000155332	41.7032	-83.254	WE12	9/11/17	WE12	set_35	2017	Sep	11	254	37	samp_3937
doli_UT_Erie_late_2022	LE22-WE12	D_circ	0.206590332	4.773271	99.92	99.97	0.166568985	NA	NA	NA	NA	WE6	set_35	2022	NA	NA	NA	44	samp_4333
8761-SRC-7	LE22-WE12	D_circ	0.00932581	4.773271	99.14	94.49	0.00140773	NA	NA	WE2	9/9/20	WE2	set_41	2020	Sep	9	253	37	samp_4338
8761-SRC-8	LE22-WE12	D_circ	0.002132266	4.773271	99.58	55.98	0.000641697	NA	NA	WE4	9/9/20	WE4	set_41	2020	Sep	9	253	37	samp_4339
8761-SRC-9	LE22-WE12	D_circ	0.007746412	4.773271	99.38	95.14	0.001504263	NA	NA	WE8	9/9/20	WE8	set_41	2020	Sep	9	253	37	samp_4340
8761-SRC-10	LE22-WE12	D_circ	0.0087103	4.773271	99.19	91.09	0.001346757	NA	NA	WE12	9/9/20	WE12	set_41	2020	Sep	9	253	37	samp_4341
8761-SRC-11	LE22-WE12	D_circ	0.008030088	4.773271	99.64	99.44	0.002779643	NA	NA	WE2	10/5/20	WE2	set_41	2020	Oct	5	279	40	samp_4342
8761-SRC-13	LE22-WE12	D_circ	0.005412309	4.773271	99.6	99.82	0.001704736	NA	NA	WE8	10/5/20	WE8	set_41	2020	Oct	5	279	40	samp_4343
8761-SRC-14	LE22-WE12	D_circ	0.003237957	4.773271	99.04	77.02	0.000673366	NA	NA	WE12	10/5/20	WE12	set_41	2020	Oct	5	279	40	samp_4344
8761-SRC-24	LE22-WE12	D_circ	0.009401801	4.773271	99.51	96.18	0.002374701	NA	NA	WE2	9/20/21	WE2	set_41	2021	Sep	20	263	38	samp_4353
8761-SRC-26	LE22-WE12	D_circ	0.001780052	4.773271	99.31	80.93	0.000330095	NA	NA	WE8	9/20/21	WE8	set_41	2021	Sep	20	263	38	samp_4355
8761-SRC-27	LE22-WE12	D_circ	0.005748146	4.773271	99.34	89.6	0.001106003	NA	NA	WE12	9/20/21	WE12	set_41	2021	Sep	20	263	38	samp_4356
8761-SRC-28	LE22-WE12	D_circ	0.0042785	4.773271	99.46	96.41	0.001019037	NA	NA	WE2	10/4/21	WE2	set_41	2021	Oct	4	277	40	samp_4357
8761-SRC-29	LE22-WE12	D_circ	0.004382092	4.773271	99.54	73.84	0.001147836	NA	NA	WE4	10/4/21	WE4	set_41	2021	Oct	4	277	40	samp_4358
8761-SRC-30	LE22-WE12	D_circ	0.027606805	4.773271	99.62	99.95	0.008144864	NA	NA	WE8	10/4/21	WE8	set_41	2021	Oct	4	277	40	samp_4359
8761-SRC-31	LE22-WE12	D_circ	0.008022129	4.773271	99.47	98.47	0.001837908	NA	NA	WE12	10/4/21	WE12	set_41	2021	Oct	4	277	40	samp_4360
8761-SRC-33	LE22-WE12	D_circ	0.005829398	4.773271	99.39	99.9	0.001792302	NA	NA	WE2	7/18/22	WE2	set_41	2022	Jul	18	199	29	samp_4362
8761-SRC-35	LE22-WE12	D_circ	0.001918135	4.773271	99.04	99.17	0.000410924	NA	NA	WE8	7/18/22	WE8	set_41	2022	Jul	18	199	29	samp_4364
8761-SRC-37	LE22-WE12	D_circ	0.002836836	4.773271	99.83	99.73	0.001498102	NA	NA	WE2	8/17/22	WE2	set_41	2022	Aug	17	229	33	samp_4366
8761-SRC-38	LE22-WE12	D_circ	0.001836272	4.773271	99.65	98.27	0.000589681	NA	NA	WE4	8/17/22	WE4	set_41	2022	Aug	17	229	33	samp_4367
8761-SRC-39	LE22-WE12	D_circ	0.001358964	4.773271	99.54	99.29	0.000463798	NA	NA	WE8	8/17/22	WE8	set_41	2022	Aug	17	229	33	samp_4368
8761-SRC-40	LE22-WE12	D_circ	0.000840262	4.773271	99.78	81.56	0.000490058	NA	NA	WE12	8/17/22	WE12	set_41	2022	Aug	17	229	33	samp_4369
8761-SRC-41	LE22-WE12	D_circ	0.001953536	4.773271	99.8	99.89	0.000926508	NA	NA	WE2	8/29/22	WE2	set_41	2022	Aug	29	241	35	samp_4370
8761-SRC-43	LE22-WE12	D_circ	0.002178549	4.773271	99.75	99.78	0.000885241	NA	NA	WE8	8/29/22	WE8	set_41	2022	Aug	29	241	35	samp_4372
8761-SRC-44	LE22-WE12	D_circ	0.001073596	4.773271	99.77	99.73	0.000523433	NA	NA	WE12	8/29/22	WE12	set_41	2022	Aug	29	241	35	samp_4373
8761-SRC-45	LE22-WE12	D_circ	0.010036444	4.773271	99.91	99.92	0.006987236	NA	NA	WE2	9/12/22	WE2	set_41	2022	Sep	12	255	37	samp_4374
8761-SRC-46	LE22-WE12	D_circ	0.005257714	4.773271	99.64	89.52	0.001676476	NA	NA	WE4	9/12/22	WE4	set_41	2022	Sep	12	255	37	samp_4375
8761-SRC-47	LE22-WE12	D_circ	0.010050091	4.773271	99.89	99.93	0.006483152	NA	NA	WE8	9/12/22	WE8	set_41	2022	Sep	12	255	37	samp_4376
8761-SRC-48	LE22-WE12	D_circ	0.00329929	4.773271	99.58	99.29	0.000873891	NA	NA	WE12	9/12/22	WE12	set_41	2022	Sep	12	255	37	samp_4377
8761-SRC-49	LE22-WE12	D_circ	0.026085088	4.773271	99.92	99.89	0.019855169	NA	NA	WE2	9/30/22	WE2	set_41	2022	Sep	30	273	39	samp_4378
8761-SRC-51	LE22-WE12	D_circ	0.059879046	4.773271	99.91	99.98	0.045082756	NA	NA	WE8	9/30/22	WE8	set_41	2022	Sep	30	273	39	samp_4379
8761-SRC-52	LE22-WE12	D_circ	0.006131792	4.773271	99.94	99.45	0.004859538	NA	NA	WE12	9/30/22	WE12	set_41	2022	Sep	30	273	39	samp_4380
8761-SRC-53	LE22-WE12	D_circ	0.148223532	4.773271	99.91	99.98	0.115667541	NA	NA	WE2	10/11/22	WE2	set_41	2022	Oct	11	284	41	samp_4381
8761-SRC-55	LE22-WE12	D_circ	0.097418149	4.773271	99.92	99.98	0.075703468	NA	NA	WE8	10/11/22	WE8	set_41	2022	Oct	11	284	41	samp_4382
8761-SRC-56	LE22-WE12	D_circ	0.083122387	4.773271	99.92	99.98	0.064061156	NA	NA	WE12	10/11/22	WE12	set_41	2022	Oct	11	284	41	samp_4383
8761-SRC-57	LE22-WE12	D_circ	0.142999268	4.773271	99.92	99.98	0.113067213	NA	NA	WE2	10/24/22	WE2	set_41	2022	Oct	24	297	43	samp_4384
8761-SRC-58	LE22-WE12	D_circ	0.04847821	4.773271	99.94	99.97	0.038095888	NA	NA	WE4	10/24/22	WE4	set_41	2022	Oct	24	297	43	samp_4385
8761-SRC-59	LE22-WE12	D_circ	0.11365768	4.773271	99.93	99.98	0.089953133	NA	NA	WE8	10/24/22	WE8	set_41	2022	Oct	24	297	43	samp_4386
8761-SRC-60	LE22-WE12	D_circ	0.169975815	4.773271	99.92	99.98	0.135046888	NA	NA	WE12	10/24/22	WE12	set_41	2022	Oct	24	297	43	samp_4387
8761-SRC-102	LE22-WE12	D_circ	0.005571287	4.773271	99.9	96.67	0.004250924	NA	NA	WE4	10/11/22	WE4	set_41	2022	Oct	11	284	41	samp_4425
E2014-0106-100LTR	LE21-ER30	D_flos	0.000797091	3.90547	99.34	73.67	0.000264536	41.76472222	-83.335	WE2	9/29/14	WE2	set_17	2014	Sep	29	272	39	samp_15
E2014-0048-100LTR	LE21-ER30	D_flos	0.000718535	3.90547	99.42	74.28	0.000190341	41.7025	-83.26305556	WE12	8/4/14	WE12	set_17	2014	Aug	4	216	31	samp_22
E20150030	LE21-ER30	D_flos	0.000665611	3.90547	99.59	82.46	0.000280483	41.7037	-83.2553	WE12	8/10/15	WE12	set_35	2015	Aug	10	222	32	samp_448
E20150032	LE21-ER30	D_flos	0.000392729	3.90547	99.56	71.82	0.000167756	41.7172	-83.016	WE14	8/10/15	WE14	set_35	2015	Aug	10	222	32	samp_450
E20190025	LE21-ER30	D_flos	0.010823969	3.90547	99.54	97.53	0.002875239	41.7624	-83.3306	WE2	8/5/19	WE2	set_35	2019	Aug	5	217	31	samp_462
E20190026	LE21-ER30	D_flos	0.013620331	3.90547	99.6	98.53	0.003806036	41.703	-83.2544	WE12	8/5/19	WE12	set_35	2019	Aug	5	217	31	samp_463
E20190028	LE21-ER30	D_flos	0.007465809	3.90547	99.62	97.51	0.002127136	41.7615	-83.3309	WE2	8/12/19	WE2	set_35	2019	Aug	12	224	32	samp_464
E20190029	LE21-ER30	D_flos	0.003531	3.90547	99.6	96.55	0.001036169	41.7021	-83.2551	WE12	8/12/19	WE12	set_35	2019	Aug	12	224	32	samp_465
E20190037	LE21-ER30	D_flos	0.001052131	3.90547	99.58	70.53	0.000324226	41.7606	-83.3326	WE2	9/3/19	WE2	set_35	2019	Sep	3	246	36	samp_467
E20190038	LE21-ER30	D_flos	0.002286958	3.90547	99.64	70.88	0.000343275	41.7034	-83.2558	WE12	9/3/19	WE12	set_35	2019	Sep	3	246	36	samp_468
E20200011	LE21-ER30	D_flos	0.001596018	3.90547	99.5	92.41	0.000608047	41.7634167	-83.3307	WE2	7/6/20	WE2	set_35	2020	Jul	6	188	27	samp_470
E20200013	LE21-ER30	D_flos	0.00736591	3.90547	99.54	92.2	0.002554295	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200027	LE21-ER30	D_flos	0.000428774	3.90547	99.27	58.54	0.000136311	41.8274167	-83.195617	WE4	7/28/20	WE4	set_35	2020	Jul	28	210	30	samp_474
E20212010	LE21-ER30	D_flos	0.001565883	3.90547	99.65	80.01	0.000541748	41.653433	-83.177667	WE16-WSW	8/3/21	WE16	set_35	2021	Aug	3	215	31	samp_499
E20212011	LE21-ER30	D_flos	0.001484297	3.90547	99.63	74.02	0.000492693	41.653433	-83.177667	WE12-S	8/3/21	WE12	set_35	2021	Aug	3	215	31	samp_500
E20212015	LE21-ER30	D_flos	0.000690298	3.90547	99.61	60.44	0.000211536	41.70406	-83.25291	WE12	8/31/21	WE12	set_35	2021	Aug	31	243	35	samp_504
LE8	LE21-ER30	D_flos	0.002582696	3.90547	99.55	94.01	0.000788547	41.7028	-83.2563	WE12	7/29/19	WE12	set_35	2019	Jul	29	210	30	samp_508
LE20	LE21-ER30	D_flos	0.000454234	3.90547	99.49	71.56	0.000197351	41.7035	-83.2537	WE12	7/15/19	WE12	set_35	2019	Jul	15	196	28	samp_511
LE21	LE21-ER30	D_flos	0.000934808	3.90547	99.51	80.85	0.000275495	41.8261	-83.1946	WE4	7/29/19	WE4	set_35	2019	Jul	29	210	30	samp_512
SC53_50_Metagenome_2019	LE21-ER30	D_flos	0.000840428	3.90547	99.57	59.24	0.000282607	41.74046627	-83.29939226	NF	8/17/19	WE2	set_36	2019	Aug	17	229	33	samp_980
SC05_50_Metagenome_2019	LE21-ER30	D_flos	0.00059693	3.90547	99.56	70.88	0.000233364	41.70248614	-83.25536512	NF	8/27/19	WE12	set_36	2019	Aug	27	239	35	samp_994
SC57_50_Metagenome_2019	LE21-ER30	D_flos	0.000556493	3.90547	99.54	54.58	0.000176676	41.74057002	-83.30542931	NF	8/16/19	WE2	set_36	2019	Aug	16	228	33	samp_1315
SC41_50_Metagenome_2019	LE21-ER30	D_flos	0.00063083	3.90547	99.44	51.7	0.000180312	41.79870352	-83.34393421	NF	8/20/19	WE2	set_36	2019	Aug	20	232	34	samp_1317
SC43_50_Metagenome_2019	LE21-ER30	D_flos	0.000793039	3.90547	99.6	73.45	0.000318715	41.80373272	-83.36106082	NF	8/20/19	WE8	set_36	2019	Aug	20	232	34	samp_1319
SC37_50_Metagenome_2019	LE21-ER30	D_flos	0.002067918	3.90547	99.55	54.15	0.000315549	41.73539193	-83.29849188	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1324
SC49_50_Metagenome_2019	LE21-ER30	D_flos	0.001574878	3.90547	99.57	80.86	0.000473519	41.73442755	-83.29631719	NF	8/17/19	WE2	set_36	2019	Aug	17	229	33	samp_1328
SV_663	LE21-ER30	D_flos	0.000467376	3.90547	99.49	81.96	0.000138954	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016

Table S2. (Page 3 of 5) Reads per kilobase million (RPKM) calculations and sample metadata used in Figure 4.

SampleName	mag	group	group_rpkm	mag_length_mb	id_ref_mag percent	mag_cover percent	rpkm	lat	lon	noaa_site	collection date	clustered_site	study_id	year	month	day	yday	week	sample_id GLAMR
sv_639	LE21-ER30	D_flos	0.000467941	3.90547	99.66	69.09	0.00012834	41.7635	-83.3288	T10	8/17/21	WE2	set_42	2021	Aug	17	229	33	samp_2025
9-12-16 WLE 8	LE21-ER30	D_flos	0.004472647	3.90547	99.52	58.58	0.000936293	41.8322	-83.3589	WE8	9/12/16	WE8	set_41	2016	Sep	12	256	37	samp_2059
8-7-17 WLE 2	LE21-ER30	D_flos	0.000463084	3.90547	99.54	66.54	0.000101219	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
7-22-19 WLE 2	LE21-ER30	D_flos	0.001406048	3.90547	99.57	93.02	0.000453391	41.763	-83.3289	WE2	7/22/19	WE2	set_41	2019	Jul	22	203	29	samp_2109
7-22-19 WLE 8	LE21-ER30	D_flos	0.014755786	3.90547	99.58	55.73	0.002935736	41.8354	-83.3584	WE8	7/22/19	WE8	set_41	2019	Jul	22	203	29	samp_2111
7-29-19 WLE 2	LE21-ER30	D_flos	0.003999353	3.90547	99.58	80.28	0.000727904	41.7602	-83.3336	WE2	7/29/19	WE2	set_41	2019	Jul	29	210	30	samp_2113
7-29-19 WLE 4	LE21-ER30	D_flos	0.006031619	3.90547	99.52	95.67	0.001336861	41.8261	-83.1946	WE4	7/29/19	WE4	set_41	2019	Jul	29	210	30	samp_2114
7-29-19 WLE 8	LE21-ER30	D_flos	0.003109758	3.90547	99.55	97.52	0.000818504	41.8328	-83.3625	WE8	7/29/19	WE8	set_41	2019	Jul	29	210	30	samp_2115
7-29-19 WLE 12	LE21-ER30	D_flos	0.007113439	3.90547	99.58	92.76	0.001725159	41.7028	-83.2563	WE12	7/29/19	WE12	set_41	2019	Jul	29	210	30	samp_2116
8-19-19 WLE 2	LE21-ER30	D_flos	0.000366966	3.90547	99.37	76.96	0.000104146	41.7621	-83.3303	WE2	8/19/19	WE2	set_41	2019	Aug	19	231	33	samp_2117
9-3-19 WLE 12	LE21-ER30	D_flos	0.003446266	3.90547	99.67	67.12	0.000681879	41.7034	-83.2558	WE12	9/3/19	WE12	set_41	2019	Sep	3	246	36	samp_2138
7-6-20 WLE 2	LE21-ER30	D_flos	0.001248611	3.90547	99.53	79.13	0.000403482	41.7634	-83.3307	WE2	7/6/20	WE2	set_41	2020	Jul	6	188	27	samp_2139
7-6-20 WLE 8	LE21-ER30	D_flos	0.005559975	3.90547	99.58	97.88	0.001863517	NA	NA	WE8	7/6/20	WE8	set_41	2020	Jul	6	188	27	samp_2141
7-6-20 WLE 12	LE21-ER30	D_flos	0.000835268	3.90547	99.5	65.98	0.000255829	41.7035	-83.2552	WE12	7/6/20	WE12	set_41	2020	Jul	6	188	27	samp_2142
8-24-20 WLE 8	LE21-ER30	D_flos	0.001169967	3.90547	97.85	69.8	0.000350499	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
TWTPSurge_080219	LE21-ER30	D_flos	0.000383046	3.90547	99.33	77.42	0.000129331	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2214
TWTPIntake_080219	LE21-ER30	D_flos	0.001809966	3.90547	99.6	95.55	0.000482774	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2215
LE2	LE21-ER30	D_flos	0.000495575	3.90547	99.54	53.23	0.000131546	41.7032	-83.254	WE12	9/11/17	WE12	set_35	2017	Sep	11	254	37	samp_3937
8761-SRC-20	LE21-ER30	D_flos	0.000380035	3.90547	99.62	74.82	0.000117691	NA	NA	WE2	8/30/21	WE2	set_41	2021	Aug	30	242	35	samp_4349
8761-SRC-22	LE21-ER30	D_flos	0.000522983	3.90547	99.62	78.33	0.000154685	NA	NA	WE8	8/30/21	WE8	set_41	2021	Aug	30	242	35	samp_4351
8761-SRC-23	LE21-ER30	D_flos	0.000891071	3.90547	99.62	85.03	0.000277323	NA	NA	WE12	8/30/21	WE12	set_41	2021	Aug	30	242	35	samp_4352
8761-SRC-33	LE21-ER30	D_flos	0.009035442	3.90547	99.65	89.35	0.002574168	NA	NA	WE2	7/18/22	WE2	set_41	2022	Jul	18	199	29	samp_4362
8761-SRC-35	LE21-ER30	D_flos	0.007307129	3.90547	99.61	91.77	0.00243914	NA	NA	WE8	7/18/22	WE8	set_41	2022	Jul	18	199	29	samp_4364
8761-SRC-57	LE21-ER30	D_flos	0.00100291	3.90547	99.44	78.48	0.000317647	NA	NA	WE2	10/24/22	WE2	set_41	2022	Oct	24	297	43	samp_4384
E2014-0106-100LTR	LE16-WE8	D_flos	0.000797091	3.786874	99.52	83.71	0.00035512	41.76472222	-83.335	WE2	9/29/14	WE2	set_17	2014	Sep	29	272	39	samp_15
E2014-0048-100LTR	LE16-WE8	D_flos	0.000718535	3.786874	99.59	95.78	0.000421657	41.7025	-83.26305556	WE12	8/4/14	WE12	set_17	2014	Aug	4	216	31	samp_22
E2014-0124-100LTR	LE16-WE8	D_flos	0.000143862	3.786874	99.46	50.57	7.14E-05	41.76388889	-83.33083333	WE2	10/20/14	WE2	set_17	2014	Oct	20	293	42	samp_35
E2014-0108-100LTR	LE16-WE8	D_flos	0.000173721	3.786874	99.45	52.18	7.68E-05	41.7031	-83.2591	WE12	9/29/14	WE12	set_17	2014	Sep	29	272	39	samp_38
E2014-0066-100LTR	LE16-WE8	D_flos	0.000417839	3.786874	99.38	59.92	0.000198859	41.705	-83.25638889	WE12	8/25/14	WE12	set_17	2014	Aug	25	237	34	samp_42
E20150030	LE16-WE8	D_flos	0.000665611	3.786874	99.65	75.2	0.000251933	41.7037	-83.2553	WE12	8/10/15	WE12	set_35	2015	Aug	10	222	32	samp_448
E20150032	LE16-WE8	D_flos	0.000392729	3.786874	99.62	63.14	0.000145826	41.7172	-83.016	WE14	8/10/15	WE14	set_35	2015	Aug	10	222	32	samp_450
E20190025	LE16-WE8	D_flos	0.010823969	3.786874	99.66	96.53	0.003378208	41.7624	-83.3306	WE2	8/5/19	WE2	set_35	2019	Aug	5	217	31	samp_462
E20190026	LE16-WE8	D_flos	0.013620331	3.786874	99.72	98.01	0.004497454	41.703	-83.2544	WE12	8/5/19	WE12	set_35	2019	Aug	5	217	31	samp_463
E20190028	LE16-WE8	D_flos	0.007465809	3.786874	99.67	96.1	0.001979967	41.7615	-83.3309	WE2	8/12/19	WE2	set_35	2019	Aug	12	224	32	samp_464
E20190029	LE16-WE8	D_flos	0.003531	3.786874	99.65	93.81	0.000955152	41.7021	-83.2551	WE12	8/12/19	WE12	set_35	2019	Aug	12	224	32	samp_465
E20190038	LE16-WE8	D_flos	0.002286958	3.786874	99.53	52.7	0.000252182	41.7034	-83.2558	WE12	9/3/19	WE12	set_35	2019	Sep	3	246	36	samp_468
E20200011	LE16-WE8	D_flos	0.001596018	3.786874	99.6	88.81	0.000631001	41.7634167	-83.3307	WE2	7/6/20	WE2	set_35	2020	Jul	6	188	27	samp_470
E20200013	LE16-WE8	D_flos	0.00736591	3.786874	99.68	92.38	0.003210115	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20200015	LE16-WE8	D_flos	0.000315886	3.786874	99.54	52.59	0.000147839	41.7034833	-83.255167	WE12	7/6/20	WE12	set_35	2020	Jul	6	188	27	samp_472
E20200027	LE16-WE8	D_flos	0.000428774	3.786874	99.53	68.13	0.000193646	41.8274167	-83.195617	WE4	7/28/20	WE4	set_35	2020	Jul	28	210	30	samp_474
E20212010	LE16-WE8	D_flos	0.001565883	3.786874	99.8	93.17	0.000792921	41.653433	-83.177667	WE16-WSW	8/3/21	WE16	set_35	2021	Aug	3	215	31	samp_499
E20212011	LE16-WE8	D_flos	0.001484297	3.786874	99.79	88.03	0.000774734	41.653433	-83.177667	WE12-S	8/3/21	WE12	set_35	2021	Aug	3	215	31	samp_500
E20212012	LE16-WE8	D_flos	0.000255854	3.786874	99.77	63.04	0.000136495	41.76265	-83.329	WE2	8/31/21	WE2	set_35	2021	Aug	31	243	35	samp_501
E20212015	LE16-WE8	D_flos	0.000690298	3.786874	99.8	80.7	0.000376048	41.70406	-83.25291	WE12	8/31/21	WE12	set_35	2021	Aug	31	243	35	samp_504
LE8	LE16-WE8	D_flos	0.002582696	3.786874	99.66	91.65	0.000869049	41.7028	-83.2563	WE12	7/29/19	WE12	set_35	2019	Jul	29	210	30	samp_508
LE20	LE16-WE8	D_flos	0.000454234	3.786874	99.59	62.74	0.000163127	41.7035	-83.2537	WE12	7/15/19	WE12	set_35	2019	Jul	15	196	28	samp_511
LE21	LE16-WE8	D_flos	0.000934808	3.786874	99.66	80.94	0.000318507	41.8261	-83.1946	WE4	7/29/19	WE4	set_35	2019	Jul	29	210	30	samp_512
SC53_50_Metagenome_2019	LE16-WE8	D_flos	0.000840428	3.786874	99.66	55.57	0.000277599	41.74046627	-83.29939226	NF	8/17/19	WE2	set_36	2019	Aug	17	229	33	samp_980
SC05_50_Metagenome_2019	LE16-WE8	D_flos	0.00059693	3.786874	99.62	62.75	0.000228839	41.70248614	-83.25536512	NF	8/27/19	WE12	set_36	2019	Aug	27	239	35	samp_994
SC57_50_Metagenome_2019	LE16-WE8	D_flos	0.000556493	3.786874	99.64	56.4	0.000209811	41.74057002	-83.30542931	NF	8/16/19	WE2	set_36	2019	Aug	16	228	33	samp_1315
SC43_50_Metagenome_2019	LE16-WE8	D_flos	0.000793039	3.786874	99.59	58.19	0.000236966	41.80373272	-83.36106082	NF	8/20/19	WE8	set_36	2019	Aug	20	232	34	samp_1319
SC37_50_Metagenome_2019	LE16-WE8	D_flos	0.002067918	3.786874	99.6	57.07	0.000330484	41.73539193	-83.29849188	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1324
SC49_50_Metagenome_2019	LE16-WE8	D_flos	0.001574878	3.786874	99.62	75.96	0.000453434	41.73442755	-83.29631719	NF	8/17/19	WE2	set_36	2019	Aug	17	229	33	samp_1328
6538-PDU-29	LE16-WE8	D_flos	0.000573355	3.786874	99.8	65.35	0.000320227	41.8381	-83.3504	NF	8/11/21	WE8	set_51	2021	Aug	11	223	32	samp_1451
SV_663	LE16-WE8	D_flos	0.000467376	3.786874	99.74	95.69	0.000246287	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016
sv_639	LE16-WE8	D_flos	0.000467941	3.786874	99.8	88.65	0.000275375	41.7635	-83.3288	T10	8/17/21	WE2	set_42	2021	Aug	17	229	33	samp_2025
sv_641	LE16-WE8	D_flos	0.000204093	3.786874	99.79	55.98	0.000142866	41.7654	-83.3273	T6	9/15/21	WE2	set_42	2021	Sep	15	258	37	samp_2027
7-18-16 WLE 8	LE16-WE8	D_flos	0.000255379	3.786874	99.84	60.69	0.000178913	41.8338	-83.3511	WE8	7/18/16	WE8	set_41	2016	Jul	18	200	29	samp_2043
9-12-16 WLE 8	LE16-WE8	D_flos	0.004472647	3.786874	99.85	96.4	0.003005798	41.8322	-83.3589	WE8	9/12/16	WE8	set_41	2016	Sep	12	256	37	samp_2059
8-7-17 WLE 2	LE16-WE8	D_flos	0.000463084	3.786874	99.82	92.2	0.00030262	41.7633	-83.3303	WE2	8/7/17	WE2	set_41	2017	Aug	7	219	32	samp_2073
8-21-17 WLE 2	LE16-WE8	D_flos	0.000216554	3.786874	99.4	50.94	0.000104793	41.7624	-83.3295	WE2	8/21/17	WE2	set_41	2017	Aug	21	233	34	samp_2077
7-22-19 WLE 2	LE16-WE8	D_flos	0.001406048	3.786874	99.7	95.05	0.000603613	41.763	-83.3289	WE2	7/22/19	WE2	set_41	2019	Jul	22	203	29	samp_2109
7-22-19 WLE 4	LE16-WE8	D_flos	0.000394011	3.786874	99.34	53.61	0.000197739	41.8311	-83.195	WE4	7/22/19	WE4	set_41	2019	Jul	22	203	29	samp_2110
7-22-19 WLE 8	LE16-WE8	D_flos	0.014755786	3.786874	99.7	66.16	0.003938318	41.8354	-83.3584	WE8	7/22/19	WE8	set_41	2019	Jul	22	203	29	samp_2111
7-29-19 WLE 2	LE16-WE8	D_flos	0.003999353	3.786874	99.7	89.97	0.000986593	41.7602	-83.3336	WE2	7/29/19	WE2	set_41	2019	Jul	29	210	30	samp_2113
7-29-19 WLE 4	LE16-WE8	D_flos	0.006031619	3.786874	99.59	93.69	0.001343201	41.8261	-83.1946	WE4	7/29/19	WE4	set_41	2019	Jul	29	210	30	samp_2114
7-29-19 WLE 8	LE16-WE8	D_flos	0.003109758	3.786874	99.69	98.21	0.001157858	41.8328	-83.3625	WE8	7/29/19	WE8	set_41	2019	Jul	29	210	30	samp_2115
7-29-19 WLE 12	LE16-WE8	D_flos	0.007113439	3.786874	99.63	90.53	0.001655178	41.7028	-83.2563	WE12	7/29/19	WE12	set_41	2019	Jul	29	210	30	samp_2116
8-19-19 WLE 2	LE16-WE8	D_flos	0.000366966	3.786874	99.52	78.57	0.000145455	41.7621	-83.3303	WE2	8/19/19	WE2	set_41	2019	Aug	19	231	33	samp_2117
9-3-19 WLE 12	LE16-WE8	D_flos	0.003446266	3.786874	99.62	55.05	0.000596764	41.7034	-83.2558	WE12	9/3/19	WE12	set_41	2019	Sep	3	246	36	samp_2138
7-6-20 WLE 2	LE16-WE8	D_flos	0.001248611	3.786874	99.66	82.61	0.000600822	41.7634	-83.3307	WE2	7/6/20	WE2	set_41	2020	Jul	6	188	27	samp_2139
7-6-20 WLE 8	LE16-WE8	D_flos	0.005559975	3.786874	99.7	97.36	0.002465395	NA	NA	WE8	7/6/20	WE8	set_41	2020	Jul	6	188	27	samp_2141
7-6-20 WLE 12	LE16-WE8	D_flos	0.000835268	3.786874	99.68	77.03	0.000423753	41.7035	-83.2552	WE12	7/6/20	WE12	set_41	2020	Jul	6	188	27	samp_2142
7-27-20 WLE 4	LE16-WE8	D_flos	0.000498846	3.786874	99.39	50.78	0.000256603	NA	NA	WE4	7/27/20	WE4	set_41	2020	Jul	27	209	30	samp_2144
8-24-20 WLE 8	LE16-WE8	D_flos	0.001169967	3.786874	98.51	79.81	0.00040569	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
TWTPSurge_080219	LE16-WE8	D_flos	0.000383046	3.786874	99.59	80.26	0.000185226	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2214
TWTPIntake_080219	LE16-WE8	D_flos	0.001809966	3.786874	99.7	94.55	0.000564481	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2215
LE14	LE16-WE8	D_flos	0.000156125	3.786874	99.58	55.63	9.52E-05	41.7026	-83.2547	WE12	9/7/21	WE12	set_35	2021	Sep	7	250	36	samp_3935
LE2	LE16-WE8	D_flos	0.000495575	3.786874	99.82	79.71	0.000292437	41.7032	-83.254	WE12	9/11/17	WE12	set_35	2017	Sep	11	254	37	samp_3937
8761-SRC-20	LE16-WE8	D_flos	0.000380035	3.786874	99.79	91.02	0.000205862	NA	NA	WE2	8/30/21	WE2	set_41	2021	Aug	30	242	35	samp_4349
8761-SRC-21	LE16-WE8	D_flos	0.000291294	3.786874	99.65	54.6	0.000156313	NA	NA	WE4	8/30/21	WE4	set_41	2021	Aug	30	242	35	samp_4350
8761-SRC-22	LE16-WE8	D_flos	0.000522983	3.786874	99.8	95.14	0.000292206	NA	NA	WE8	8/30/21	WE8	set_41	2021	Aug	30	242	35	samp_4351
8761-SRC-23	LE16-WE8	D_flos	0.000891071	3.786874	99.79	94.57	0.000473751	NA	NA	WE12	8/30/21	WE12	set_41	2021	Aug	30	242	35	samp_4352
8761-SRC-33	LE16-WE8	D_flos	0.009035442	3.786874	99.87	99.17	0.005163363	NA	NA	WE2	7/18/22	WE2	set_41	2022	Jul	18	199	29	samp_4362
8761-SRC-35	LE16-WE8	D_flos	0.007307129	3.786874	99.77	99.06	0.003682404	NA	NA	WE8	7/18/22	WE8	set_41	2022	Jul	18	199	29	samp_4364
8761-SRC-38	LE16-WE8	D_flos	0.000204981	3.786874	99.69	53.24	0.000118476	NA	NA	WE4	8/17/22	WE4	set_41	2022	Aug	17	229	33	samp_4367
8761-SRC-39	LE16-WE8	D_flos	0.000150626	3.786874	99.65	65.54	9.03E-05	NA	NA	WE8	8/17/22	WE8	set_41	2022	Aug	17	229	33	samp_4368
8761-SRC-57	LE16-WE8	D_flos	0.00100291	3.786874	99.57	86.11	0.000433205	NA	NA	WE2	10/24/22	WE2	set_41	2022	Oct	24	297	43	samp_4384
E2014-0106-100LTR	LE19-WE12	D_flos	0.000797091	4.238051	99.05	56.34	0.000177435	41.76472222	-83.335	WE2	9/29/14	WE2	set_17	2014	Sep	29	272	39	samp_15
E2014-0048-100LTR	LE19-WE12	D_flos	0.000718535	4.238051	99.22	56.77	0.000106537	41.7025	-83.26305556	WE12	8/4/14	WE12	set_17	2014	Aug	4	216	31	samp_22
E20150030	LE19-WE12	D_flos	0.000665611	4.238051	99.3	50.28	0.000133195	41.7037	-83.2553	WE12	8/10/15	WE12	set_35	2015	Aug	10	222	32	samp_448
E20190025	LE19-WE12	D_flos	0.010823969	4.238051	99.71	99.78	0.004570522	41.7624	-83.3306	WE2	8/5/19	WE2	set_35	2019	Aug	5	217	31	samp_462
E20190026	LE19-WE12	D_flos	0.013620331	4.238051	99.73	99.87	0.005316841	41.703	-83.2544	WE12	8/5/19	WE12	set_35	2019	Aug	5	217	31	samp_463
E20190028	LE19-WE12	D_flos	0.007465809	4.238051	99.77	99.78	0.003358706	41.7615	-83.3309	WE2	8/12/19	WE2	set_35	2019	Aug	12	224	32	samp_464
E20190029	LE19-WE12	D_flos	0.003531	4.238051	99.74	99.65	0.001539679	41.7021	-83.2551	WE12	8/12/19	WE12	set_35	2019	Aug	12	224	32	samp_465
E20190037	LE19-WE12	D_flos	0.001052131	4.238051	99.72	87.49	0.000544189	41.7606	-83.3326	WE2	9/3/19	WE2	set_35	2019	Sep	3	246	36	samp_467
E20190038	LE19-WE12	D_flos	0.002286958	4.238051	99.9	99.43	0.001691501	41.7034	-83.2558	WE12	9/3/19	WE12	set_35	2019	Sep	3	246	36	samp_468
E20200011	LE19-WE12	D_flos	0.001596018	4.238051	99.2	73.6	0.00035697	41.7634167	-83.3307	WE2	7/6/20	WE2	set_35	2020	Jul	6	188	27	samp_470
E20200013	LE19-WE12	D_flos	0.00736591	4.238051	99.34	88.36	0.0016015	41.81916	-83.35921	WE8	7/6/20	WE8	set_35	2020	Jul	6	188	27	samp_471
E20212010	LE19-WE12	D_flos	0.001565883	4.238051	99.36	58.24	0.000231215	41.653433	-83.177667	WE16-WSW	8/3/21	WE16	set_35	2021	Aug	3	215	31	samp_499
LE8	LE19-WE12	D_flos	0.002582696	4.238051	99.65	98.63	0.0009251	41.7028	-83.2563	WE12	7/29/19	WE12	set_35	2019	Jul	29	210	30	samp_508
LE21	LE19-WE12	D_flos	0.000934808	4.238051	99.65	90.65	0.000340806	41.8261	-83.1946	WE4	7/29/19	WE4	set_35	2019	Jul	29	210	30	samp_512
BWT_2G_3_50_Metagenome_2019	LE19-WE12	D_flos	0.000713845	4.238051	99.84	87.3	0.000482373	41.7344	-83.3005	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_974
SC53_50_Metagenome_2019	LE19-WE12	D_flos	0.000840428	4.238051	99.61	61.93	0.000280222	41.74046627	-83.29939226	NF	8/17/19	WE2	set_36	2019	Aug	17	229	33	samp_980
Mirror_ST12_50_Metagenome_2019	LE19-WE12	D_flos	0.00176153	4.238051	99.91	98.78	0.001316306	41.7025	-83.2547	NF	8/27/19	WE12	set_36	2019	Aug	27	239	35	samp_1001
Mirror_Nessie_50_Metagenome_2019	LE19-WE12	D_flos	0.000687852	4.238051	99.86	86.39	0.000445563	41.734	-83.302	NF	8/27/19	WE2	set_36	2019	Aug	27	239	35	samp_1003
SC39_50_Metagenome_2019	LE19-WE12	D_flos	0.000343659	4.238051	99.61	55.8	0.000180281	41.73074728	-83.29734036	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1288
SC35_50_Metagenome_2019	LE19-WE12	D_flos	0.00044425	4.238051	99.72	68.55	0.000282242	41.73542874	-83.30516857	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1294
SC11_50_Metagenome_2019	LE19-WE12	D_flos	0.000640505	4.238051	99.81	84.23	0.000412851	41.74394092	-83.13597092	NF	8/26/19	WE13	set_36	2019	Aug	26	238	34	samp_1296
SC57_50_Metagenome_2019	LE19-WE12	D_flos	0.000556493	4.238051	99.53	54.23	0.000170006	41.74057002	-83.30542931	NF	8/16/19	WE2	set_36	2019	Aug	16	228	33	samp_1315
SC41_50_Metagenome_2019	LE19-WE12	D_flos	0.00063083	4.238051	99.65	72.95	0.000301664	41.79870352	-83.34393421	NF	8/20/19	WE2	set_36	2019	Aug	20	232	34	samp_1317
SC43_50_Metagenome_2019	LE19-WE12	D_flos	0.000793039	4.238051	99.53	64.24	0.000237357	41.80373272	-83.36106082	NF	8/20/19	WE8	set_36	2019	Aug	20	232	34	samp_1319
SC37_50_Metagenome_2019	LE19-WE12	D_flos	0.002067918	4.238051	99.87	98.55	0.001421885	41.73539193	-83.29849188	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1324
SC39_50_Metagenome_2019	LE19-WE12	D_flos	0.000342897	4.238051	99.62	55.54	0.000179265	41.73074728	-83.29734036	NF	8/21/19	WE2	set_36	2019	Aug	21	233	34	samp_1326
SC49_50_Metagenome_2019	LE19-WE12	D_flos	0.001574878	4.238051	99.69	92.53	0.000647925	41.73442755	-83.29631719	NF	8/17/19	WE2	set_36	2019	Aug	17	229	33	samp_1328
SV_663	LE19-WE12	D_flos	0.000467376	4.238051	99.14	65.59	8.21E-05	41.7574	-83.3325	T10	7/26/22	WE2	set_42	2022	Jul	26	207	30	samp_2016
7-22-19 WLE 2	LE19-WE12	D_flos	0.001406048	4.238051	99.49	92.3	0.000349044	41.763	-83.3289	WE2	7/22/19	WE2	set_41	2019	Jul	22	203	29	samp_2109
7-22-19 WLE 4	LE19-WE12	D_flos	0.000394011	4.238051	99.31	63.27	0.000115563	41.8311	-83.195	WE4	7/22/19	WE4	set_41	2019	Jul	22	203	29	samp_2110
7-22-19 WLE 8	LE19-WE12	D_flos	0.014755786	4.238051	99.81	91.11	0.007881731	41.8354	-83.3584	WE8	7/22/19	WE8	set_41	2019	Jul	22	203	29	samp_2111
7-29-19 WLE 2	LE19-WE12	D_flos	0.003999353	4.238051	99.83	99.47	0.002284855	41.7602	-83.3336	WE2	7/29/19	WE2	set_41	2019	Jul	29	210	30	samp_2113
7-29-19 WLE 4	LE19-WE12	D_flos	0.006031619	4.238051	99.78	99.75	0.003351556	41.8261	-83.1946	WE4	7/29/19	WE4	set_41	2019	Jul	29	210	30	samp_2114
7-29-19 WLE 8	LE19-WE12	D_flos	0.003109758	4.238051	99.66	99.84	0.001133397	41.8328	-83.3625	WE8	7/29/19	WE8	set_41	2019	Jul	29	210	30	samp_2115
7-29-19 WLE 12	LE19-WE12	D_flos	0.007113439	4.238051	99.78	99.54	0.003733102	41.7028	-83.2563	WE12	7/29/19	WE12	set_41	2019	Jul	29	210	30	samp_2116
8-19-19 WLE 2	LE19-WE12	D_flos	0.000366966	4.238051	99.45	82.81	0.000117366	41.7621	-83.3303	WE2	8/19/19	WE2	set_41	2019	Aug	19	231	33	samp_2117
8-28-19 WLE 2	LE19-WE12	D_flos	0.001072999	4.238051	99.86	91.8	0.000698956	NA	NA	WE2	8/28/19	WE2	set_41	2019	Aug	28	240	35	samp_2131
8-28-19 WLE 12	LE19-WE12	D_flos	0.002972985	4.238051	99.86	95.82	0.00204478	41.7424	-83.138	WE12	8/28/19	WE12	set_41	2019	Aug	28	240	35	samp_2134
9-3-19 WLE 2	LE19-WE12	D_flos	0.006538934	4.238051	99.88	95.48	0.004772785	41.7606	-83.3326	WE2	9/3/19	WE2	set_41	2019	Sep	3	246	36	samp_2135
9-3-19 WLE 12	LE19-WE12	D_flos	0.003446266	4.238051	99.86	97.15	0.002167623	41.7034	-83.2558	WE12	9/3/19	WE12	set_41	2019	Sep	3	246	36	samp_2138
7-6-20 WLE 2	LE19-WE12	D_flos	0.001248611	4.238051	99.3	55.94	0.000244306	41.7634	-83.3307	WE2	7/6/20	WE2	set_41	2020	Jul	6	188	27	samp_2139
7-6-20 WLE 8	LE19-WE12	D_flos	0.005559975	4.238051	99.4	94.88	0.001231063	NA	NA	WE8	7/6/20	WE8	set_41	2020	Jul	6	188	27	samp_2141
8-24-20 WLE 8	LE19-WE12	D_flos	0.001169967	4.238051	97.91	55.18	0.000413778	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
TWTPIntake_080219	LE19-WE12	D_flos	0.001809966	4.238051	99.75	99.61	0.000762711	41.700044	-83.250217	NA	8/2/19	WE12	set_18	2019	Aug	2	214	31	samp_2215
8761-SRC-20	LE19-WE12	D_flos	0.000380035	4.238051	99.37	50.73	5.65E-05	NA	NA	WE2	8/30/21	WE2	set_41	2021	Aug	30	242	35	samp_4349
8761-SRC-22	LE19-WE12	D_flos	0.000522983	4.238051	99.39	56.58	7.61E-05	NA	NA	WE8	8/30/21	WE8	set_41	2021	Aug	30	242	35	samp_4351
8761-SRC-23	LE19-WE12	D_flos	0.000891071	4.238051	99.39	62.44	0.000139997	NA	NA	WE12	8/30/21	WE12	set_41	2021	Aug	30	242	35	samp_4352
8761-SRC-33	LE19-WE12	D_flos	0.009035442	4.238051	99.41	79.1	0.001297911	NA	NA	WE2	7/18/22	WE2	set_41	2022	Jul	18	199	29	samp_4362
8761-SRC-35	LE19-WE12	D_flos	0.007307129	4.238051	99.37	83.86	0.001185585	NA	NA	WE8	7/18/22	WE8	set_41	2022	Jul	18	199	29	samp_4364
8761-SRC-57	LE19-WE12	D_flos	0.00100291	4.238051	99.28	62.38	0.000252058	NA	NA	WE2	10/24/22	WE2	set_41	2022	Oct	24	297	43	samp_4384
E20200026	LE20-WE2-Aug	C_issa	0.000367634	4.344744	98.85	50.38	0.000192213	41.7622	-83.331167	WE2	7/28/20	WE2	set_35	2020	Jul	28	210	30	samp_473
E20200047	LE20-WE2-Aug	C_issa	0.001336614	4.344744	99.25	97.33	0.000784902	41.76135	-83.33285	WE2	8/10/20	WE2	set_35	2020	Aug	10	223	32	samp_477
E20200051	LE20-WE2-Aug	C_issa	0.000728399	4.344744	98.94	86.3	0.000351232	41.7174667	-83.424367	WE9	8/10/20	WE9	set_35	2020	Aug	10	223	32	samp_479
E20200052	LE20-WE2-Aug	C_issa	0.000808706	4.344744	99.01	85.19	0.000429133	41.702	-83.254333	WE12	8/10/20	WE12	set_35	2020	Aug	10	223	32	samp_480
E20200089	LE20-WE2-Aug	C_issa	0.00234427	4.344744	99.06	97.17	0.001129077	41.7627833	-83.32965	WE2	9/9/20	WE2	set_35	2020	Sep	9	253	37	samp_481
E20200094	LE20-WE2-Aug	C_issa	0.001424058	4.344744	99.04	94.43	0.000729083	41.7054833	-83.254233	WE12	9/9/20	WE12	set_35	2020	Sep	9	253	37	samp_484
SC17_50_Metagenome	LE20-WE2-Aug	C_issa	0.000249467	4.344744	98.51	53.48	0.000130075	41.85046938	-83.19494729	NF	8/31/18	WE4	set_36	2018	Aug	31	243	35	samp_791
SC33_50_Metagenome	LE20-WE2-Aug	C_issa	0.000354827	4.344744	98.49	60.43	0.00018577	41.84968278	-83.19036052	NF	8/29/18	WE4	set_36	2018	Aug	29	241	35	samp_861
SC50_50_Metagenome	LE20-WE2-Aug	C_issa	0.000247093	4.344744	98.41	51.61	0.000133252	41.82489782	-83.21184393	NF	8/24/18	WE4	set_36	2018	Aug	24	236	34	samp_924
8-24-20 WLE 2	LE20-WE2-Aug	C_issa	0.000447967	4.344744	98.26	55.51	0.000233226	41.7618	-83.331	WE2	8/24/20	WE2	set_41	2020	Aug	24	237	34	samp_2151
8-24-20 WLE 8	LE20-WE2-Aug	C_issa	0.000582846	4.344744	97.86	90.33	0.000282389	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153

Table S2. (Page 4 of 5) Reads per kilobase million (RPKM) calculations and sample metadata used in Figure 4.

SampleName	mag	group	group_rpkm	mag_length_mb	id_ref_mag percent	mag_cover percent	rpkm	lat	lon	noaa_site	collection date	clustered_site	study_id	year	month	day	yday	week	sample_id GLAMR
8-31-20 WLE 2	LE20-WE2-Aug	C_issa	0.000560606	4.344744	98.38	51.95	0.000289555	NA	NA	WE2	8/31/20	WE2	set_41	2020	Aug	31	244	35	samp_2155
8-31-20 WLE 8	LE20-WE2-Aug	C_issa	0.000459211	4.344744	98.14	65.56	0.000234755	NA	NA	WE8	8/31/20	WE8	set_41	2020	Aug	31	244	35	samp_2157
8-31-20 WLE 12	LE20-WE2-Aug	C_issa	0.001040414	4.344744	98.11	83.04	0.000532351	NA	NA	WE12	8/31/20	WE12	set_41	2020	Aug	31	244	35	samp_2158
8761-SRC-7	LE20-WE2-Aug	C_issa	0.000801379	4.344744	98.65	93.09	0.000407065	NA	NA	WE2	9/9/20	WE2	set_41	2020	Sep	9	253	37	samp_4338
8761-SRC-10	LE20-WE2-Aug	C_issa	0.000324033	4.344744	98.25	73.61	0.000170138	NA	NA	WE12	9/9/20	WE12	set_41	2020	Sep	9	253	37	samp_4341
E20200026	LE20-WE2-Sep	C_issa	0.000367634	3.60207	99.28	52.42	0.00017542	41.7622	-83.331167	WE2	7/28/20	WE2	set_35	2020	Jul	28	210	30	samp_473
E20200047	LE20-WE2-Sep	C_issa	0.001336614	3.60207	99.28	93.43	0.000551712	41.76135	-83.33285	WE2	8/10/20	WE2	set_35	2020	Aug	10	223	32	samp_477
E20200051	LE20-WE2-Sep	C_issa	0.000728399	3.60207	99.16	88.95	0.000377167	41.7174667	-83.424367	WE9	8/10/20	WE9	set_35	2020	Aug	10	223	32	samp_479
E20200052	LE20-WE2-Sep	C_issa	0.000808706	3.60207	99.25	87.57	0.000379572	41.702	-83.254333	WE12	8/10/20	WE12	set_35	2020	Aug	10	223	32	samp_480
E20200089	LE20-WE2-Sep	C_issa	0.00234427	3.60207	99.26	99.37	0.001215193	41.7627833	-83.32965	WE2	9/9/20	WE2	set_35	2020	Sep	9	253	37	samp_481
E20200094	LE20-WE2-Sep	C_issa	0.001424058	3.60207	99.18	96.57	0.000694975	41.7054833	-83.254233	WE12	9/9/20	WE12	set_35	2020	Sep	9	253	37	samp_484
SC17_50_Metagenome	LE20-WE2-Sep	C_issa	0.000249467	3.60207	98.89	54.52	0.000119392	41.85046938	-83.19494729	NF	8/31/18	WE4	set_36	2018	Aug	31	243	35	samp_791
SC33_50_Metagenome	LE20-WE2-Sep	C_issa	0.000354827	3.60207	98.78	61.85	0.000169056	41.84968278	-83.19036052	NF	8/29/18	WE4	set_36	2018	Aug	29	241	35	samp_861
SC50_50_Metagenome	LE20-WE2-Sep	C_issa	0.000247093	3.60207	98.82	52.16	0.000113841	41.82489782	-83.21184393	NF	8/24/18	WE4	set_36	2018	Aug	24	236	34	samp_924
8-24-20 WLE 2	LE20-WE2-Sep	C_issa	0.000447967	3.60207	98.49	56.92	0.000214742	41.7618	-83.331	WE2	8/24/20	WE2	set_41	2020	Aug	24	237	34	samp_2151
8-24-20 WLE 8	LE20-WE2-Sep	C_issa	0.000582846	3.60207	98.16	92.47	0.000300457	41.8342	-83.3604	WE8	8/24/20	WE8	set_41	2020	Aug	24	237	34	samp_2153
8-31-20 WLE 2	LE20-WE2-Sep	C_issa	0.000560606	3.60207	98.67	53.45	0.000271052	NA	NA	WE2	8/31/20	WE2	set_41	2020	Aug	31	244	35	samp_2155
8-31-20 WLE 8	LE20-WE2-Sep	C_issa	0.000459211	3.60207	98.47	67.26	0.000224456	NA	NA	WE8	8/31/20	WE8	set_41	2020	Aug	31	244	35	samp_2157
8-31-20 WLE 12	LE20-WE2-Sep	C_issa	0.001040414	3.60207	98.26	85.64	0.000508062	NA	NA	WE12	8/31/20	WE12	set_41	2020	Aug	31	244	35	samp_2158
8761-SRC-7	LE20-WE2-Sep	C_issa	0.000801379	3.60207	98.86	95.43	0.000394314	NA	NA	WE2	9/9/20	WE2	set_41	2020	Sep	9	253	37	samp_4338
8761-SRC-10	LE20-WE2-Sep	C_issa	0.000324033	3.60207	98.57	76.06	0.000153895	NA	NA	WE12	9/9/20	WE12	set_41	2020	Sep	9	253	37	samp_4341

Table S2. (Page 5 of 5) Reads per kilobase million (RPKM) calculations and sample metadata used in Figure 4.

mag	gtdbtk_species_classification	gene	contig	type	start	stop	strand	locus_tag	product	db_refs
LE16-WE8	Dolichospermum flosaquae	nifH	samp_2059_2985	cds	1432	1833	-	BHEOGE_16725	Nitrogenase (Molybdenum-iron) reductase and maturation protein NifH	SO:0001217, UniRef:UniRef50_A0A6B9R323, UniRef:UniRef90_W6FTP0
LE16-WE8	Dolichospermum flosaquae	ureA	samp_2059_47461	cds	27	329	+	BHEOGE_08860	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, SO:0001217, UniParc:UPI0007FBE4BB, UniRef:UniRef100_A0A1B7X098, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE16-WE8	Dolichospermum flosaquae	ureB	samp_2059_140083	cds	1291	1596	-	BHEOGE_15265	urease subunit beta	COG:COG0832, COG:E, RefSeq:WP_190386701.1, SO:0001217, UniParc:UPI00080196BD, UniRef:UniRef100_A0A1B7VZ77, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE17-WE12	Dolichospermum circinale	nifH	samp_3937_1114353	cds	682	1122	+	CELJHL_14400	Nitrogenase ATPase subunit NifH/coenzyme F430 biosynthesis subunit CfbC	COG:COG1348, COG:HP, SO:0001217, UniRef:UniRef50_A0A1Z4UNY5, UniRef:UniRef90_A0A1Z4UNY5
LE17-WE12	Dolichospermum circinale	ureA	samp_3937_676628	cds	112	414	-	CELJHL_09425	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, SO:0001217, UniParc:UPI0007FBE4BB, UniRef:UniRef100_A0A1B7X098, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE17-WE12	Dolichospermum circinale	ureB	samp_3937_256809	cds	650	955	-	CELJHL_03775	urease subunit beta	COG:COG0832, COG:E, RefSeq:WP_028090419.1, SO:0001217, UniParc:UPI00042296D1, UniRef:UniRef100_UPI00042296D1, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE17-WE2	Dolichospermum sp000312705	nifH	samp_2073_20761	cds	2087	2809	+	PHGANH_00510	Nitrogenase ATPase subunit NifH/coenzyme F430 biosynthesis subunit CfbC	COG:COG1348, COG:HP, KEGG:K02588, SO:0001217, UniRef:UniRef50_A0A076N8J7, UniRef:UniRef90_A0A5C0DSJ6
LE17-WE2	Dolichospermum sp000312705	nifK	samp_2073_767671	cds	15733	17268	-	PHGANH_07990	nitrogenase molybdenum-iron protein subunit beta	COG:COG2710, COG:HP, SO:0001217, UniRef:UniRef50_P07329, UniRef:UniRef90_D2CN98
LE17-WE2	Dolichospermum sp000312705	ureA	samp_2073_915305	cds	564	866	-	PHGANH_09445	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, SO:0001217, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE17-WE2	Dolichospermum sp000312705	ureB	samp_2073_58478	cds	233	538	+	PHGANH_01015	urease subunit beta	COG:COG0832, COG:E, SO:0001217, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE17-WE2	Dolichospermum sp000312705	ureC	samp_2073_58478	cds	582	2288	+	PHGANH_01020	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A161Y2P7
LE19-WE12	Dolichospermum flosaquae	nifD	samp_468_229199	cds	6587	7981	+	FCOGFM_07390	nitrogenase molybdenum-iron protein alpha chain	COG:COG2710, COG:HP, SO:0001217, UniRef:UniRef50_P07328, UniRef:UniRef90_UPI001E49E6DE
LE19-WE12	Dolichospermum flosaquae	nifH	samp_468_229199	cds	6023	6424	+	FCOGFM_07385	Nitrogenase (Molybdenum-iron) reductase and maturation protein NifH	SO:0001217, UniRef:UniRef50_A0A6B9R323, UniRef:UniRef90_W6FTP0
LE19-WE12	Dolichospermum flosaquae	nifK	samp_468_793005	cds	8582	10117	-	FCOGFM_19175	nitrogenase molybdenum-iron protein subunit beta	COG:COG2710, COG:HP, RefSeq:WP_190387296.1, SO:0001217, UniParc:UPI0008021008, UniRef:UniRef100_A0A1B7VZP7, UniRef:UniRef50_P07329, UniRef:UniRef90_D2CN98
LE19-WE12	Dolichospermum flosaquae	ureA	samp_468_294592	cds	20966	21268	-	FCOGFM_10075	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, SO:0001217, UniParc:UPI0007FBE4BB, UniRef:UniRef100_A0A1B7X098, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE19-WE12	Dolichospermum flosaquae	ureB	samp_468_294592	cds	20638	20943	-	FCOGFM_10070	urease subunit beta	COG:COG0832, COG:E, RefSeq:WP_190386701.1, SO:0001217, UniParc:UPI00080196BD, UniRef:UniRef100_A0A1B7VZ77, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE19-WE12	Dolichospermum flosaquae	ureC	samp_468_294592	cds	18880	20586	-	FCOGFM_10060	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A161Y2P7
LE20-WE12	Dolichospermum circinale	nifD	samp_4344_89087	cds	5588	6514	+	AIOLHN_04875	Nitrogenase Mo-Fe protein NifD/coenzyme F430 biosynthesis subunit CfbD	COG:COG2710, COG:HP, SO:0001217, UniParc:UPI000AFF0457, UniRef:UniRef100_A0A480AIX5, UniRef:UniRef50_A0A090A141, UniRef:UniRef90_K9ZEH4
LE20-WE12	Dolichospermum circinale	nifH	samp_4344_89087	cds	4566	5456	+	AIOLHN_04870	nitrogenase iron protein	COG:COG1348, COG:HP, EC:1.18.6.1, GO:0005524, GO:0009399, GO:0016163, GO:0018697, GO:0046872, GO:0051539, KEGG:K02588, RefSeq:WP_028089570.1, SO:0001217, UniParc:UPI0004280ACB, UniRef:UniRef100_A0A480ACX4, UniRef:UniRef50_P22921, UniRef:UniRef90_P0A3S1
LE20-WE12	Dolichospermum circinale	nifK	samp_4344_179825	cds	305	1840	+	AIOLHN_08975	nitrogenase molybdenum-iron protein subunit beta	COG:COG2710, COG:HP, SO:0001217, UniRef:UniRef50_P07329, UniRef:UniRef90_D2CN98
LE20-WE12	Dolichospermum circinale	ureA	samp_4344_207048	cds	3244	3546	+	AIOLHN_10135	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, RefSeq:WP_028090420.1, SO:0001217, UniParc:UPI0004144060, UniRef:UniRef100_UPI0004144060, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE20-WE12	Dolichospermum circinale	ureB	samp_4344_117061	cds	1204	1509	+	AIOLHN_06185	urease subunit beta	COG:COG0832, COG:E, RefSeq:WP_190391792.1, SO:0001217, UniParc:UPI0016801939, UniRef:UniRef100_UPI0016801939, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE20-WE12	Dolichospermum circinale	ureC	samp_4344_61552	cds	5296	7002	-	AIOLHN_03775	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A161Y2P7
LE20-WE2-Aug	Cuspidothrix issatschenkoi	ureA	samp_477_328402	cds	744	1046	+	FCGBEC_06260	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, RefSeq:WP_104386384.1, SO:0001217, UniParc:UPI000CEA416C, UniRef:UniRef100_A0A2S6CYG2, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE20-WE2-Aug	Cuspidothrix issatschenkoi	ureB	samp_477_432475	cds	351	656	+	FCGBEC_07690	urease subunit beta	COG:COG0832, COG:E, SO:0001217, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE20-WE2-Aug	Cuspidothrix issatschenkoi	ureC	samp_477_432475	cds	699	2405	+	FCGBEC_07695	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A161Y2P7
LE20-WE2-Sep	Cuspidothrix issatschenkoi	ureA	samp_481_103302	cds	4655	4957	-	JAKCJJ_05045	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, RefSeq:WP_104386384.1, SO:0001217, UniParc:UPI000CEA416C, UniRef:UniRef100_A0A2S6CYG2, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE20-WE2-Sep	Cuspidothrix issatschenkoi	ureB	samp_481_103302	cds	2450	2755	-	JAKCJJ_05030	urease subunit beta	COG:COG0832, COG:E, RefSeq:WP_104386381.1, SO:0001217, UniParc:UPI000CEA6BBA, UniRef:UniRef100_A0A2S6CYY5, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE20-WE2-Sep	Cuspidothrix issatschenkoi	ureC	samp_481_103302	cds	701	2407	-	JAKCJJ_05020	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A161Y2P7
LE20-WE8	Dolichospermum sp000312705	nifD	samp_471_4997	cds	151	1545	-	INKDHH_04075	nitrogenase molybdenum-iron protein alpha chain	COG:COG2710, COG:HP, SO:0001217, UniParc:UPI000801E20D, UniRef:UniRef100_A0A1B7UYU8, UniRef:UniRef50_P07328, UniRef:UniRef90_A0A6H2BUG4
LE20-WE8	Dolichospermum sp000312705	nifH	samp_471_4997	cds	1708	2109	-	INKDHH_04080	Nitrogenase (Molybdenum-iron) reductase and maturation protein NifH	SO:0001217, UniRef:UniRef50_A0A6B9R323, UniRef:UniRef90_W6FTP0
LE20-WE8	Dolichospermum sp000312705	nifH	samp_471_716301	cds	3226	3663	-	INKDHH_18800	Nitrogenase ATPase subunit NifH/coenzyme F430 biosynthesis subunit CfbC	COG:COG1348, COG:HP, SO:0001217, UniRef:UniRef50_A0A1Z4UNY5, UniRef:UniRef90_A0A6P1ZSQ3
LE20-WE8	Dolichospermum sp000312705	nifK	samp_471_100553	cds	5475	7010	+	INKDHH_00435	nitrogenase molybdenum-iron protein subunit beta	COG:COG2710, COG:HP, SO:0001217, UniParc:UPI0007FE4F2B, UniRef:UniRef100_A0A1B7UYS0, UniRef:UniRef50_P07329, UniRef:UniRef90_D2CN98
LE20-WE8	Dolichospermum sp000312705	ureA	samp_471_293068	cds	7701	8003	+	INKDHH_10965	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, SO:0001217, UniParc:UPI0007FCEAF5, UniRef:UniRef100_A0A1B7USE5, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE20-WE8	Dolichospermum sp000312705	ureB	samp_471_293068	cds	8658	8963	+	INKDHH_10975	urease subunit beta	COG:COG0832, COG:E, SO:0001217, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE20-WE8	Dolichospermum sp000312705	ureC	samp_471_293068	cds	9007	10713	+	INKDHH_10980	urease subunit alpha	SO:0001217, UniParc:UPI0008014104, UniRef:UniRef100_A0A1B7UST5, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_K9ZQQ4
LE21-ER30	Dolichospermum flosaquae	nifD	samp_4305_756901	cds	3621	5015	+	FKMMKH_13010	nitrogenase molybdenum-iron protein alpha chain	COG:COG2710, COG:HP, SO:0001217, UniRef:UniRef50_P07328, UniRef:UniRef90_UPI001E49E6DE
LE21-ER30	Dolichospermum flosaquae	nifH	samp_4305_756901	cds	3057	3458	+	FKMMKH_13005	Nitrogenase (Molybdenum-iron) reductase and maturation protein NifH	SO:0001217, UniRef:UniRef50_A0A6B9R323, UniRef:UniRef90_W6FTP0
LE21-ER30	Dolichospermum flosaquae	nifK	samp_4305_143453	cds	56647	58182	-	FKMMKH_01615	nitrogenase molybdenum-iron protein subunit beta	COG:COG2710, COG:HP, SO:0001217, UniParc:UPI0008009EEB, UniRef:UniRef100_A0A1B7X3U9, UniRef:UniRef50_P07329, UniRef:UniRef90_D2CN98
LE21-ER30	Dolichospermum flosaquae	ureA	samp_4305_486248	cds	63200	63502	-	FKMMKH_00855	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, SO:0001217, UniParc:UPI0007FBE4BB, UniRef:UniRef100_A0A1B7X098, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE21-ER30	Dolichospermum flosaquae	ureB	samp_4305_486248	cds	62872	63177	-	FKMMKH_00850	urease subunit beta	COG:COG0832, COG:E, SO:0001217, UniRef:UniRef50_B7K907, UniRef:UniRef90_A0A1Z4UIF0
LE21-ER30	Dolichospermum flosaquae	ureC	samp_4305_486248	cds	61114	62820	-	FKMMKH_00840	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniParc:UPI0007FE5E31, UniRef:UniRef100_A0A1B7VKE7, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A0S3TTX2
LE21-WE12	Dolichospermum circinale	nifK	samp_507_981320	cds	1503	3038	-	FLNODI_11290	nitrogenase molybdenum-iron protein subunit beta	COG:COG2710, COG:HP, SO:0001217, UniRef:UniRef50_P07329, UniRef:UniRef90_D2CN98
LE21-WE12	Dolichospermum circinale	ureA	samp_507_468850	cds	14369	14671	+	FLNODI_03475	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, RefSeq:WP_028090420.1, SO:0001217, UniParc:UPI0004144060, UniRef:UniRef100_UPI0004144060, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE21-WE12	Dolichospermum circinale	ureB	samp_507_468850	cds	16572	16877	+	FLNODI_03490	urease subunit beta	COG:COG0832, COG:E, RefSeq:WP_190391792.1, SO:0001217, UniParc:UPI0016801939, UniRef:UniRef100_UPI0016801939, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE21-WE12	Dolichospermum circinale	ureC	samp_507_468850	cds	16843	18627	+	FLNODI_03495	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A161Y2P7
LE22-WE12	Dolichospermum circinale	nifD	samp_4380_175307	cds	9160	10587	-	DPCBHP_09890	NifD element site-specific recombinase	SO:0001217, UniRef:UniRef50_K7W5E0, UniRef:UniRef90_K7W5E0
LE22-WE12	Dolichospermum circinale	nifD	samp_4380_175307	cds	11291	12217	+	DPCBHP_09900	Nitrogenase Mo-Fe protein NifD/coenzyme F430 biosynthesis subunit CfbD	COG:COG2710, COG:HP, SO:0001217, UniParc:UPI000AFF0457, UniRef:UniRef100_A0A480AIX5, UniRef:UniRef50_A0A090A141, UniRef:UniRef90_K9ZEH4
LE22-WE12	Dolichospermum circinale	nifH	samp_4380_175307	cds	6230	6670	+	DPCBHP_09865	Nitrogenase ATPase subunit NifH/coenzyme F430 biosynthesis subunit CfbC	COG:COG1348, COG:HP, SO:0001217, UniRef:UniRef50_A0A1Z4UNY5, UniRef:UniRef90_A0A1Z4UNY5
LE22-WE12	Dolichospermum circinale	nifH	samp_4380_175307	cds	10743	11159	+	DPCBHP_09895	Nitrogenase ATPase subunit NifH/coenzyme F430 biosynthesis subunit CfbC	COG:COG1348, COG:HP, SO:0001217, UniRef:UniRef50_A0A6B9R323, UniRef:UniRef90_A0A1Z4SB71
LE22-WE12	Dolichospermum circinale	nifK	samp_4380_27589	cds	1316	2851	+	DPCBHP_21475	nitrogenase molybdenum-iron protein subunit beta	COG:COG2710, COG:HP, SO:0001217, UniRef:UniRef50_P07329, UniRef:UniRef90_D2CN98
LE22-WE12	Dolichospermum circinale	ureA	samp_4380_38824	cds	8413	8715	+	DPCBHP_14045	urease subunit gamma	COG:COG0831, COG:E, EC:3.5.1.5, GO:0005737, GO:0009039, GO:0016151, GO:0043419, RefSeq:WP_028090420.1, SO:0001217, UniParc:UPI0004144060, UniRef:UniRef100_UPI0004144060, UniRef:UniRef50_Q4KJ06, UniRef:UniRef90_B2IT64
LE22-WE12	Dolichospermum circinale	ureB	samp_4380_38824	cds	10301	10606	+	DPCBHP_14060	urease subunit beta	COG:COG0832, COG:E, RefSeq:WP_028090419.1, SO:0001217, UniParc:UPI00042296D1, UniRef:UniRef100_UPI00042296D1, UniRef:UniRef50_A1TSZ5, UniRef:UniRef90_K9ZNU6
LE22-WE12	Dolichospermum circinale	ureC	samp_4380_38824	cds	10649	12355	+	DPCBHP_14065	urease subunit alpha	COG:COG0804, COG:E, SO:0001217, UniRef:UniRef50_Q7V3V2, UniRef:UniRef90_A0A161Y2P7

Table S3. Urease and nif gene annotations from Bakta for all eleven representative MAGs.

query	gene	subject	percent_identity	length	mismatch	gapopen	query_start	query_end	subject_start	subject_end	evalue	bitscore	qcovs
samp_471_220425_trim	sxtPER	sxtPER_Aphanizomenon sp. NH-5_EU603710.1	99.43	1059	6	0	5899	6957	1	1059	0.00E+00	1923	15
samp_471_220425_trim	sxtC	sxtC_Aphanizomenon sp. NH-5_EU603710.1	99.65	285	1	0	5294	5578	1	285	1.33E-149	521	4
samp_471_220425_trim	sxtB	sxtB_Aphanizomenon sp. NH-5_EU603710.1	99.17	969	8	0	4239	5207	1	969	0.00E+00	1746	14
samp_471_220425_trim	sxtA	sxtA_Aphanizomenon sp. NH-5_EU603710.1	99.19	3705	30	0	529	4233	1	3705	0.00E+00	6676	53
samp_471_220425_trim	sxtE	sxtE_Aphanizomenon sp. NH-5_EU603710.1	99.45	363	2	0	1	363	1	363	0.00E+00	660	5
samp_471_266508_trim	sxtD	sxtD_Aphanizomenon sp. NH-5_EU603710.1	99.36	156	1	0	1	156	156	1	7.36E-78	283	2
samp_471_266508_trim	sxtP	sxtP_Aphanizomenon sp. NH-5_EU603710.1	99.52	1443	7	0	219	1661	1	1443	0.00E+00	2627	19
samp_471_266508_trim	sxtQ	sxtQ_Aphanizomenon sp. NH-5_EU603710.1	98.84	777	9	0	1686	2462	1	777	0.00E+00	1386	10
samp_471_266508_trim	sxtR	sxtR_Aphanizomenon sp. NH-5_EU603710.1	99.61	777	3	0	2473	3249	1	777	0.00E+00	1419	10
samp_471_266508_trim	sxtS	sxtS_Aphanizomenon sp. NH-5_EU603710.1	98.63	729	10	0	3993	4721	1	729	0.00E+00	1291	10
samp_471_266508_trim	sxtT	sxtT_Aphanizomenon sp. NH-5_EU603710.1	99.61	1020	4	0	4805	5824	1	1020	0.00E+00	1862	14
samp_471_266508_trim	sxtU	sxtU_Aphanizomenon sp. NH-5_EU603710.1	98.8	750	9	0	5973	6722	1	750	0.00E+00	1336	10
samp_471_266508_trim	sxtN	sxtN_Aphanizomenon sp. NH-5_EU603710.1	97.32	747	14	1	6775	7515	1	747	0.00E+00	1264	10
samp_471_36365_trim	sxtG	sxtG_Aphanizomenon sp. NH-5_EU603710.1	99.12	1134	10	0	6310	7443	1134	1	0.00E+00	2039	15
samp_471_36365_trim	sxtH	sxtH_Aphanizomenon sp. NH-5_EU603710.1	99.31	1020	7	0	5229	6248	1020	1	0.00E+00	1845	14
samp_471_36365_trim	sxtM	sxtM_Aphanizomenon sp. NH-5_EU603710.1	98.97	1458	15	0	3725	5182	1458	1	0.00E+00	2610	20
samp_471_36365_trim	sxtI	sxtI_Aphanizomenon sp. NH-5_EU603710.1	98.48	1839	28	0	1899	3737	1839	1	0.00E+00	3241	25
samp_471_36365_trim	sxtJ	sxtJ_Aphanizomenon sp. NH-5_EU603710.1	99.75	405	1	0	1453	1857	405	1	0.00E+00	743	5
samp_471_36365_trim	sxtK	sxtK_Aphanizomenon sp. NH-5_EU603710.1	99.39	165	1	0	1289	1453	165	1	7.24E-83	300	2
samp_471_36365_trim	sxtL	sxtL_Aphanizomenon sp. NH-5_EU603710.1	95.54	1278	57	0	1	1278	1278	1	0.00E+00	2045	17

Table S4. Blastn results from comparison between LE20-WE8 *sxt* genes and *Aphanizomenon* sp. NH-5 reference.