AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

Check for updates

| Spotlight Selection | Environmental Microbiology | Full-Length Text

Metagenomics reveals spatial variation in cyanobacterial composition, function, and biosynthetic potential in the Winam Gulf, Lake Victoria, Kenya

Lauren N. Hart,[1,2,3] Brittany N. Zepernick,[4] Kaela E. Natwora,[5] Katelyn M. Brown,[3,6] Julia Akinyi Obuya,[7] Davide Lomeo,[8] Malcolm A. Barnard,[9,10] Eric O. Okech,[11] 2022-23 NSF-IRES Lake Victoria Research Consortium, E. Anders Kiledal,[3,12] Paul A. Den Uyl,[13,14] Mark Olokotum,[13] Steven W. Wilhelm,[3,4] R. Michael McKay,[3,15] Ken G. Drouillard,[15] David H. Sherman,[1,2,3,16,17] Lewis Sitoki,[18] James Achiya,[7] Albert Getabu,[19] Kefa M. Otiso,[20] George S. Bullerjahn,[3,6] Gregory J. Dick[1,3,12,14]

AUTHOR AFFILIATIONS See affiliation list on p. 22.

ABSTRACT The Winam Gulf in the Kenyan region of Lake Victoria experiences prolific, year-round cyanobacterial harmful algal blooms (cyanoHABs) which pose threats to human, livestock, and ecosystem health. To our knowledge, there is limited molecular research on the gulf's cyanoHABs, and thus, the strategies employed for survival and proliferation by toxigenic cyanobacteria in this region remain largely unexplored. Here, we used metagenomics to analyze the Winam Gulf's cyanobacterial composition, function, and biosynthetic potential. *Dolichospermum* was the dominant bloom-forming cyanobacterium, co-occurring with *Microcystis* at most sites. *Microcystis* and *Planktothrix* were more abundant in shallow and turbid sites. Metagenome-assembled genomes (MAGs) of *Dolichospermum* harbored nitrogen fixation genes, suggesting diazotrophy as a potential mechanism supporting the proliferation of *Dolichospermum* in the nitrogen-limited gulf. Over 300 biosynthetic gene clusters (BGCs) putatively encoding the synthesis of toxins and other secondary metabolites were identified across the gulf, even at sites where there were no visible cyanoHAB events. Almost all BGCs identified had no known synthesis product, indicating a diverse and novel biosynthetic repertoire capable of synthesizing harmful or potentially therapeutic metabolites. *Microcystis* MAGs contained *mcy* genes encoding the synthesis of hepatotoxic microcystins which are a concern for drinking water safety. These findings illustrate the spatial variation of bloom-forming cyanobacteria in the Winam Gulf and their available strategies to dominate different ecological niches. This study underscores the need for further use of genomic techniques to elucidate the dynamics and mitigate the potentially harmful effects of cyanoHABs and their associated toxins on human, environmental, and economic health.

IMPORTANCE The Winam Gulf (Kenya) is a vital resource that experiences prolific cyanobacterial harmful algal blooms (cyanoHABs). Bloom-forming cyanobacteria produce cyanotoxins, threatening human and environmental health, recreation, and fishing. However, cyanotoxin production in the gulf has not been linked to a specific type of cyanobacteria. We used DNA sequencing of whole microbial communities to track the species of cyanobacteria present across the gulf and investigate the genes responsible for synthesis of known and novel toxins. Our results reveal *Dolichospermum* as the main bloom-forming cyanobacteria in the gulf, often co-occurring with high abundance of toxigenic *Microcystis.* Over 300 unique gene clusters were found, with most predicted to encode the synthesis of uncharacterized molecules. These results provide initial insights into the diverse biosynthetic potential encoded by cyanobacteria in the Winam Gulf and

Editor Jennifer F. Biddle, University of Delaware, Lewes, Delaware, USA

Address correspondence to Gregory J. Dick, gdick@umich.edu.

The authors declare no conflict of interest.

See the funding table on p. 23.

Received 8 August 2024
Accepted 11 November 2024
Published 8 January 2025

[This article was published on 8 January 2025 with the Importance section missing. The Importance was added in the current version, posted on 6 February 2025.]

Copyright © 2025 Hart et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

underscore the need to further elucidate and investigate the effects of known and novel molecules produced in cyanoHABs in this region.

KEYWORDS African Great Lakes, eutrophication, *Dolichospermum*, *Microcystis*, climate change, metagenomics

Cyanobacterial harmful algal blooms (cyanoHABs) are a severe, global threat to freshwater systems and are expected to intensify with climate change (1–3). The overgrowth of cyanobacteria and production of cyanotoxins threaten human and environmental health, freshwater supply, livestock, recreation, tourism, and fishing (4–6). Two of the most ubiquitous freshwater, bloom-forming cyanobacterial genera are *Microcystis*, a non-diazotroph notorious for its ability to produce toxic microcystins, and *Dolichospermum*, a filamentous diazotroph (7–10). These two genera can produce a variety of hepatotoxins, neurotoxins, and other secondary metabolites, including microcystins, saxitoxins, anatoxins, anabaenopeptins, aeruginosins, microginins, microviridins, and cyanopeptolins (10, 11).

Considerable attention has been directed toward understanding and monitoring *Microcystis*- and *Dolichospermum*-dominated cyanoHABs in North America and Europe (12–15). Yet, cyanoHAB characteristics, occurrences, and consequences in the global south remain less studied (16). Notably, Lake Victoria, the world's second-largest lake by surface area, is experiencing rapid eutrophication and proliferation of toxic, *Microcystis*-dominated cyanoHABs (17, 18). CyanoHABs are prevalent throughout the Lake Victoria basin, which has a coastline shared by Kenya, Uganda, and Tanzania. Moreover, most cyanoHABs are concentrated within embayments such as Winam Gulf (i.e., Nyanza Gulf) (17, 18).

The Winam Gulf is a shallow inlet (average depth 7 m) near the northeastern corner of Lake Victoria that experiences year-round cyanoHABs (17). The gulf provides water for drinking, laundry, bathing, nearshore subsistence, and commercial fishing, as well as aquaculture, to Kisumu, Kenya's third largest city, and many nearby fishing villages such as Homa Bay and Kendu Bay (directly supporting ~4.5 million people) (19, 20). A 2015 study found that 50% of households sampled near Kisumu reported sole use of raw lake water for drinking and household use, with greater than 30% of water samples from the area exceeding World Health Organization (WHO) drinking water guidelines for microcystins (21). Sixty percent of samples exceeded microcystin thresholds for children and immunocompromised individuals suggested by the United States Environmental Protection Agency, which is of particular concern due to the high prevalence of immunomodulating diseases in Kisumu like HIV and malaria (21–24). Thus, resource-dependent populations could face many routes and multifaceted effects of cyanotoxin exposure.

Since the 1980s, the Winam Gulf has suffered profound ecological changes due to invasive water hyacinth and other species, oxygen depletion, and increasing phytoplankton biomass (20). These changes are accompanied by substantial human population growth and increasing agricultural, industrial, urban, and wastewater contaminant runoff into the gulf (25). The main rivers flowing into the gulf are the Nyando and Sondu-Miriu Rivers which enter from the southeast corner, delivering loads of silt, nitrogen (N), and phosphorus (P) from agricultural, industrial, and wastewater sources. Smaller rivers, such as the Awach-Tende, Awach-Kibuon, Kibos, and Kisat rivers, deliver sewage and urban runoff from Kisumu (Fig. 1). Sugarcane processing, tea and cereal farming, and urban areas within the Winam Gulf catchment are the main sources of pollution, with Kisumu as a significant contributor (26, 27). Total nitrogen (TN) is higher in the gulf than the open lake, while total phosphorus (TP) is lower in the gulf than the open lake. Despite abundant sources of nitrogen delivery to the gulf, TN:TP ratios remain lower than 20, indicating nitrogen limitation relative to phosphorus and potentially selecting for diazotrophic organisms (26). Despite this, non-diazotrophic cyanobacteria are commonly found in the gulf (i.e., *Microcystis*). This indicates that there is an abundance of available

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

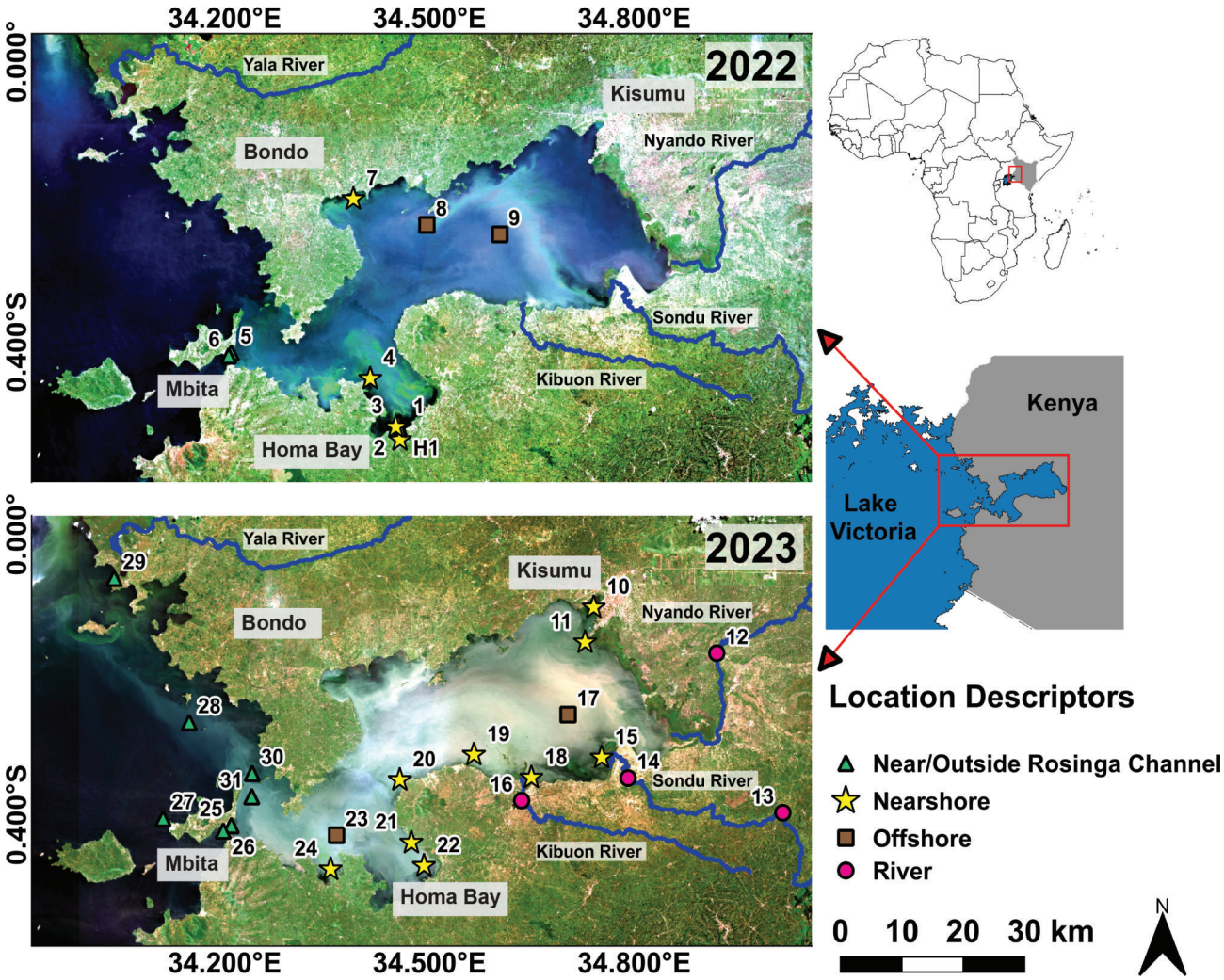

FIG 1 True color images of Winam Gulf, Lake Victoria, Kenya. The images were captured by Sentinel-2 on 30 June 2022 (top panel) and 26 May 2023 (bottom panel) and obtained as Level-2A (Surface Reflectance—atmospherically corrected by ESA using Sen2Cor) from Google Earth Engine. The images were collated together using QGIS LTR version 3.28.14. Sites sampled in 2022 are shown in the top panel, while sites sampled in 2023 are shown in the bottom panel. The location descriptor of a site is indicated by a symbol (star, circle, square, or triangle) with the site number indicated next to it on the map. Corresponding information for each site is reported in Table 1.

nutrients in the gulf, allowing non-diazotrophs to proliferate even under nitrogen-limited conditions by employing alternative strategies (28).

The climate of the Winam Gulf is characterized by two wet seasons, ranging from March to May and October to December. Two studies, from 2008 to 2009 and 2020, found maximum phytoplankton biovolumes coincide with wet (March, April, and May) and seasonal upwelling (July and August) seasons in the gulf (17, 29). Based on morphological classifications of phytoplankton, the 2008–2009 study found *Microcystis* dominates cyanoHABs in the gulf, with increased abundance of the diazotrophic cyanobacterial genus *Anabaena*, now renamed *Dolichospermum*, in September when nitrogen became more limited (30). Microcystin concentrations often exceeded WHO guidelines and increased in tandem with *Microcystis*, suggesting *Microcystis* to be the dominant microcystin producer in the gulf (17). The 2020 study identified *Microcystis* and *Dolichospermum* as the two main bloom-forming cyanobacteria at high abundance from May through July (29). Another study in 2015 and 2016 found similar phytoplankton patterns and identified *Microcystis* and *Dolichospermum* in community waters and near the drinking water intake in Kisumu (21).

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

TABLE 1 Summary of sampled stations included in the study

Station number	Site name	Sample date	Location descriptor
1	Homa Bay	24 June 2022	Nearshore
2	Homa Bay Pier	24 June 2022	Nearshore
3	Homa Bay Water Intake	24 June 2022	Nearshore
4	Soklo	24 June 2022	Nearshore
5	Mbita East[a]	24 June 2022	Near/Outside Rusinga Channel
6	Mbita West[a]	24 June 2022	Near/Outside Rusinga Channel
7	Asembo Bay	25 June 2022	Nearshore
8	Ndere Island	25 June 2022	Offshore
9	Mid Gulf	25 June 2022	Offshore
10	Kisumu Harbor	29 May 2023	Nearshore
11	Dunga	29 May 2023	Nearshore
12	Nyando River	29 May 2023	River
13	Sondu River Bridge	29 May 2023	River
14	Sondu River Shore	29 May 2023	River
15	Sondu Miriu	29 May 2023	Nearshore
16	Kibuon River	30 May 2023	River
17	Southern Mid-Gulf	29 May 2023	Offshore
18	Awach River Mouth	30 May 2023	Nearshore
19	Bala Rawi	30 May 2023	Nearshore
20	Ingra	30 May 2023	Nearshore
21	Kowuor	30 May 2023	Nearshore
22	Oluch	31 May 2023	Nearshore
23	Sikli	31 May 2023	Offshore
24	Mirunda	31 May 2023	Nearshore
25	Mbita East[a]	1 June 2023	Near/Outside Rusinga Channel
26	Mbita West[a]	1 June 2023	Near/Outside Rusinga Channel
27	Bridge Island	1 June 2023	Near/Outside Rusinga Channel
28	Bondo	2 June 2023	Near/Outside Rusinga Channel
29	Yala River Mouth	2 June 2023	Near/Outside Rusinga Channel
30	Rusinga Channel	3 June 2023	Near/Outside Rusinga Channel
31	South Rusinga	3 June 2023	Near/Outside Rusinga Channel

[a]Indicates sampling at site in both years.

Most molecular studies of the Winam Gulf have focused on fish pathogens such as *Escherichia coli*, *Vibrio cholerae*, and *Salmonella* spp. (31–33). A study by Brown et al. in 2022 used 16S rRNA gene sequencing to characterize the microbial community in the gulf, identifying the diazotrophic genus *Dolichospermum* as the dominant cyanobacterial organism (28). This study also found high copy numbers across the gulf of *cyrA* and *mcyE*, conserved genes in the biosynthetic gene clusters (BGCs) for cylindrospermopsin and microcystin, respectively (28). Despite these investigations, the producer(s) of these cyanotoxins in the gulf remains unknown. Additionally, the functional and biosynthetic capabilities of bloom-forming cyanobacteria that contribute to their fitness across different regions of the gulf have not been studied.

The aim of this study was to genetically characterize the functional and biosynthetic potential of bloom-forming cyanobacteria in the Winam Gulf using whole-genome shotgun metagenomic sequencing. We conducted a 2-year sampling expedition (2022–2023) from coastal, offshore, and river sites within and outside of the gulf. We found that cyanoHAB events were not limited to coastal zones, or even within the gulf, and were most prolific near Homa Bay in 2022 and outside of the gulf near Bondo and the Yala River mouth in 2023. These cyanoHAB events were dominated by *Dolichospermum*, with *Microcystis* co-occurring but at lower abundance. Highly diverse and abundant BGCs were encoded by many genera of cyanobacteria, with the most biosynthetic potential

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

encoded at sites with the highest abundance of bloom-forming cyanobacteria. Most BGCs could not be linked to a known biosynthesis product, highlighting the novel chemical repertoire encoded by cyanobacteria in the gulf. Lastly, our data suggest that *Microcystis* is the main and sole producer of microcystin in the gulf during the sampling period.

RESULTS

Water physicochemistry indicates nitrogen limitation throughout the gulf

Thirty-one sites throughout the gulf and nearby rivers were sampled for metagenomic analysis in May and June of 2022 and 2023 (Fig. 1; Table 1). The sites sampled span spatial, temporal, physical, and chemical (physicochemical) gradients. Physicochemical analyses were performed at each site to infer water quality (Fig. 2; Table S1), excluding river sites. Turbidity varied across sites ranging between 10 and 48.1 nephelometric turbidity units (NTUs), with a median of 23.7 NTUs. Dissolved oxygen (DO) measured between 4.26 and 11.6 mg/L across sites, with sites within the Rusinga Channel (26, 27, and 31), on the higher end of the range measured at 9.9, 10.98, and 11.6 mg/L, respectively. TN and TP were measured in 2023, and ratios (TN:TP) ranged between 5.1 and 25 with a median of 9.06. All sites except site 19 fell below a TN:TP ratio of 20, suggestive of nitrogen limitation (34) (Fig. 2). Site 19 on the southern coast of the gulf had a TN:TP ratio of 25, falling in the ambiguous nitrogen or phosphorus limited range.

River and gulf sites vary in microbial community makeup

Overall, the microbial community composition of the gulf closely resembled other eutrophic lakes with the phyla *Actinobacteria*, *Cyanobacteria*, *Proteobacteria*, and *Planctomycetota* being highly abundant (35) (Fig. 3; Table S2). The communities in 2022 and 2023 were similar in their compositions and dominant microbial phyla. The community composition at river sites (sites 12, 13, 14, and 16) diverged from that of gulf sites primarily because of lower *Cyanobacteria* ($P = 0.1$) and significantly higher *Bacteroidota* ($P < 0.001$) abundance observed in the samples. Cyanobacteria represented less than 1% of the total microbial community in each river site. Cyanobacteria were

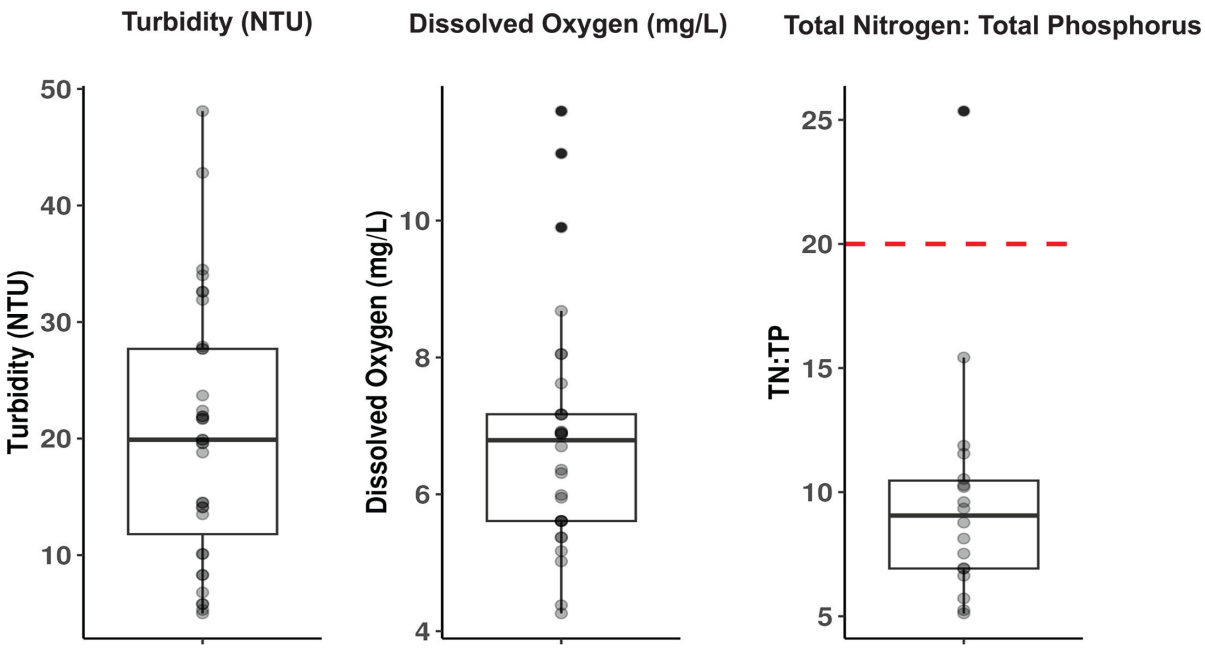

FIG 2 Turbidity (NTU), dissolved oxygen (mg/L), and TN:TP (molar) ratio measurements. Individual data points for each variable are plotted on top of each box plot. The red line in the TN:TP panel indicates the threshold for nitrogen-limitation, with all values below it indicating nitrogen-limiting scenarios according to Guildford and Hecky (2000) (34). TN:TP was only measured for sites in 2023.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

present at all sites, with relative abundance ranging between approximately 1% at site 31 and 78% at site 28.

Dolichospermum dominates large CyanoHAB events detected in distinct gulf regions

Dolichospermum, *Cyanobium*, *Vulcanococcus*, and *Microcystis* were the major cyanobacterial genera found in the gulf and rivers (Fig. 3; Table S2). Minor cyanobacterial genera present across the sites were *Snowella*, *Synechococcus*, *Elainella*, *Pseudanabaena*, *Cuspidothrix*, *Nodosilinea*, and *Caenarcaniphilales*. *Sphaerospermopsis* and *Planktothrix* had high relative abundance at sites 27 and 10, respectively, making up more than 36% and 65% of the cyanobacterial community at these two sites, respectively. *Sphaerospermopsis* and *Planktothrix* were found at four other sites at an abundance below 5% and 20% of the total cyanobacterial community, respectively. Cyanobacteria unclassified at the genus level were a substantial portion of the community (1%–16%) at most sites.

River sites had similar cyanobacterial compositions despite their lower cyanobacterial abundance. *Cyanobium* and *Vulcanococcus* represented greater than 50% of the cyanobacterial population in river sites, with *Dolichospermum* and *Microcystis* representing less than 15% individually.

FIG 3 Microbial and archaeal composition makeup in Winam Gulf, LV. Relative abundance (percentage) of bacterial and archaeal phyla (gray panel) and cyanobacterial genera (green panel) from sampled sites in 2022 and 2023 is shown. The phyla and genera representing less than 1% of the population were grouped in the "Other" category. Cyanobacteria with no taxonomic assignment for genus were labeled as "unclassified cyanobacteria." Unmapped reads were excluded from this analysis, and mapped reads were normalized to 100% for phyla and genera. Station location indicators are shown using symbols after the name of the site number and name on the y-axis.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

Across the Gulf, there were two major scenarios, with cyanobacterial community composition dominated by either bloom-forming cyanobacterial genera (*Dolichospermum* and *Microcystis*; sites 1, 2, 3, 4, 7, and 18–24) or picocyanobacterial genera (*Cyanobium* and *Vulcanococcus*; sites 5, 6, 8, 9, 25, 26, 30, and 31). In 2022, there was a large cyanoHAB event made up of *Dolichospermum* and *Microcystis* near Homa Bay, together making up approximately 98% of the cyanobacterial community. In 2023, *Dolichospermum* and *Microcystis* dominated sites along the southern and southwestern coast of the gulf, beginning near the Awach River mouth through Homa Bay and toward Mbita, together making up between 77% and 95% of the cyanobacterial community at these sites. At the sites where *Dolichospermum* and *Microcystis* co-occurred, *Dolichospermum* was more abundant than *Microcystis* by approximately 25%–60%. In a few instances, a single genus of cyanobacteria dominated the cyanobacterial assemblage: *Planktothrix* at site 10 in Kisumu harbor; *Microcystis* at site 15 near the Sondu River mouth; *Sphaerospermopsis* at site 27 off Rusinga Island in the open waters of Lake Victoria; and *Dolichospermum* at sites 28 and 29 outside the gulf near Bondo and the Yala River mouth.

Relationships between dominant cyanobacterial genera and environmental conditions

Cyanobacterial composition was assessed in relation to selected environmental parameters through non-metric MultiDimensional Scaling (NMDS; Fig. 4A). River sites and sites dominated by picocyanobacteria *Cyanobium* and *Vulcanococcus* clustered together and showed a positive relationship with DO. Sites dominated by *Dolichospermum* and *Sphaerospermopsis* clustered near each other and had a positive relationship with dissolved phosphorus, pH, and depth. Sites with mixed communities of bloom-forming cyanobacteria including *Microcystis* and *Planktothrix* formed a large cluster and had a positive relationship with ammonia, nitrate, and turbidity.

The relative abundance of dominant cyanobacterial genera (*Dolichospermum*, *Cyanobium*, *Vulcanococcus*, *Microcystis*, *Planktothrix*, and *Sphaerospermopsis*) was assessed for their relationship with environmental parameters using Spearman's correlation analysis (Fig. 4B; Table S3). *Dolichospermum*, *Microcystis*, and *Planktothrix* shared similar relationships with environmental parameters. These genera had strong positive correlations (greater than or equal to 0.5) with ammonia, temperature, chlorophyll, and turbidity ($P < 0.001$). This group was strongly negatively correlated (less than −0.5) with depth and DO ($P < 0.001$). *Sphaerospermopsis*, *Cyanobium*, and *Vulcanococcus* clustered together based on their relationships with environmental parameters. *Sphaerospermopsis* was strongly positively correlated with DO, depth, and dissolved phosphorus, while strongly negatively correlated with ammonia, dissolved N:P, nitrate, and turbidity ($P < 0.001$). *Sphaerospermopsis'* negative correlation with dissolved N:P levels suggests it to be potentially more fit in nitrogen-limiting waters, where this negative correlation most often occurs at dissolved N:P ratios less than 1 (indicative of nitrogen deficiency). *Cyanobium* and *Vulcanococcus* were strongly positively correlated with depth and strongly negatively correlated with ammonia, temperature, and cyanobacterial detection ("cyanos") via the use of an AlgaeTorch (bbe moldaenke GmnH, Schwentinental, Germany; $P < 0.001$).

Phylogenomic analysis reveals low intra-genus genetic diversity and novel clades of bloom-forming cyanobacteria from the gulf

Assembly and binning of metagenomic sequence reads produced 203 high-quality cyanobacterial metagenome-assembled genomes (MAGs; ≥90% completion and ≤10% contamination; Table S4). Dereplication at 99.5% average nucleotide identity produced a total of 17 representative MAGs of the genera *Microcystis* (five MAGs), *Dolichospermum* (five MAGs), *Planktothrix* (three MAGs), and *Sphaerospermopsis* (four MAGs; Table 2). Phylogenomic analysis showed that *Microcystis* MAGs from the gulf formed two distinct clades (Fig. 5A). Three MAGs from sites 1, 2, and 5 clustered with the Mae4 clade (36),

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

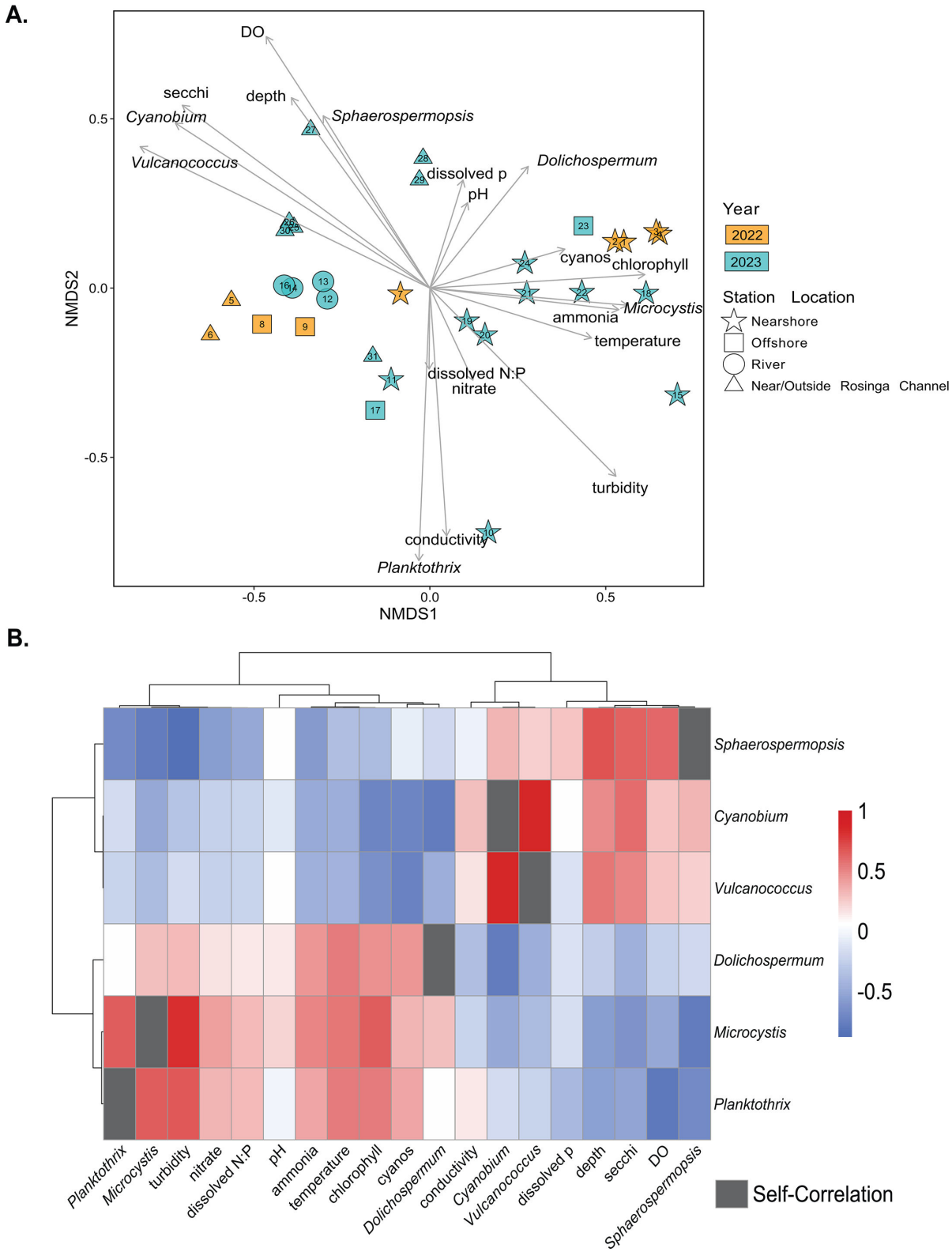

FIG 4 Cyanobacterial composition of Winam Gulf in relation to physicochemical parameters. (A) Non-metric multidimensional scaling (NMDS) analysis of the relative abundance of all cyanobacterial genera from each site, in relation to all other microbial community members (stress = 0.074). Labeled arrows show the direction and strength of the relationship between the ordination axes and associated environmental parameters and the relative abundance of the top six most abundant cyanobacterial genera across the sites. Shapes of ordination points indicate station location. Samples from 2022 are colored in amber and samples from 2023 in teal. (B) Spearman correlation coefficients between the six most abundant cyanobacterial genera across the sample sites and environmental parameters. Rows and columns are hierarchically clustered using correlation. Gray squares indicate self-correlation.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

including strain LE19-84.1 isolated from Western Lake Erie (37). Two MAGs from sites 26 and 27 fell within the Mae5 clade. *Dolichospermum* MAGs from sites 1, 25, 26, 27, and 28 formed a single, tight sister clade to a clade of *Dolichospermum circinale* and two MAGs from Lake Erie, including one MAG that has the saxitoxin BGC (Fig. 5B) (38). *Planktothrix* MAGs from sites 10, 15, and 17 formed a single clade most closely related to *Planktothrix* strains with benthic properties (Fig. 5C). *Sphaerospermopsis* MAGs from sites 25, 26, 27, and 30 formed a single, tight clade most closely related to *Sphaerospermopsis* sp. LEGE 00249, an isolate from a reservoir in Portugal (Fig. 5D) (39).

Nitrogen fixation genes harbored primarily by *Dolichospermum* and identified at sites with large bloom events

Due to nitrogen limitation in the gulf and its potential selection for diazotrophic organisms, we assessed the microbial community's ability to fix nitrogen by identifying *nif*HDK genes (42). *Dolichospermum* and *Anabaena*, two members of the *Anabaena/Dolichospermum/Aphanizomenon* (ADA) clade, had the most *nif*HDK gene matches, and very few hits were to bacteria outside of the Cyanobacteria phylum (Fig. 6) (30). *nif*HDK genes belonging to ADA clade members were found at sites where there were large cyanoHAB events in 2022 and 2023. The only instance where the three ADA clade members did not harbor most of the *nif* genes was at site 27 near Bridge Island where *Spaherospermopsis*, a closely related genera, dominated the community (Fig. 3) (10). At this site, *Sphaerospermopsis* and other *Nostacales* cyanobacteria were responsible for the *nif*HDK genes detected.

Cyanotoxin- and cyanopeptide-encoding BGCs detected in regions with and without visible cyanoHABs

Mapping metagenomic sequence reads to BGCs from the MiBIG database (44) revealed high spatiotemporal variation in genetic potential for biosynthesis of cyanotoxins and cyanopeptides (Fig. 7). No cyanobacterial BGCs were found in the four river sites, or at sites 6, 8, 9, 11, 19, 20, and 31, which had limited abundance of cyanobacteria. BGCs encoding metabolites from the classes aeruginosins, anacyclamide, microcystins, micropeptins, and microviridins were most common, detected in over half of the sites. The microcystin and *Anabaena* anacyclamide 1 BGCs were the most abundant BGCs across the data set. Although genes encoding cylindrospermopsin (28, 45) did not assemble using our metagenomic approach, read mapping revealed the presence of four highly conserved cylindrospermopsin genes (*cyr*AEBC) at a low abundance at site 27 near Bridge Island (Fig. S1). Deeper sequencing could fully elucidate this BGC and its taxonomic identity in the gulf.

Genus-specific biosynthetic potential revealed through over 300 cryptic BGCs

BGCs were mined from cyanobacterial MAGs to assess total biosynthetic potential of the Winam Gulf cyanobacterial assemblage. One thousand BGCs from cyanobacterial MAGs were identified, 300 of which were unique, from 11 different biosynthetic classes and 14 different cyanobacterial genera (Fig. 8). Notably, all but three of these BGCs did not match any known BGCs in the MiBIG database, indicating their cryptic nature (Fig. S2). BGC richness differed between the 14 cyanobacterial genera. *Synechococcus*, *Vulcanococcus*, *Cyanobium*, and the unclassified cyanobacterial genera groups all encoded an average of one BGC per MAG, with *Cyanobium* encoding the largest range of BGCs per MAG, ranging between 1 and 8. *Pseudanabaena*, *Raphidiopsis*, *Nodosilinea*, *Snowella*, and *Elainella* encoded an average of two to four BGCs per MAG. *Planktothrix*, *Cuspidothrix*, *Dolichospermum*, *Microcystis*, and *Sphaerospermopsis* encoded the most BGCs, ranging between 6 and 10 BGCs per MAG on average.

Terpenes, the largest and most structurally diverse class of natural products including compounds such as geosmin, were encoded by all 14 cyanobacterial genera analyzed (47). The next most common biosynthetic class was non-ribosomal peptide synthetases

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

TABLE 2 Summary and statistics of high-quality, dereplicated cyanobacterial MAGs

MAG	Site	Completeness (%)	Contamination (%)	Strain heterogeneity (%)	Species (based on GTDBtk annotations)	Size (Mbp)	No. of contigs	GC content (%)	N50	NCBI accession
samp_4315_VAMB_280	Mbita East	92.91	5.74	60	*Microcystis panniformis* A	4.230761	2,568	43.5	2,116	SAMN41711324
samp_4322_concoct_168	Homa Bay Pier	98.79	1.24	55.56	*Microcystis panniformis A*	5.518758	724	42.6	11,220	SAMN41711381
samp_4325_VAMB_332	H1. Homa Bay Coast	95.24	6.77	37.21	*Microcystis* sp.	5.43631	4,522	42.8	1,77	SAMN41711408
samp_4444_concoct_1899	Mbita West	90.3	7.15	69.05	*Microcystis panniformis* A	3.847615	1,713	43.2	2,584	SAMN41711474
samp_4445_concoct_1450	Bridge Island	90.56	6.79	78.05	*Microcystis aeruginosa* H	3.729636	1,804	43.1	2,269	SAMN41711483
samp_4326_maxbin_35	H1. Homa Bay Coast	97.22	0.22	0	*Dolichospermum* sp.	4.077973	370	37.5	16,022	SAMN41711415
samp_4443_metadecoder_676	Mbita East	93.74	1.01	33.33	*Dolichospermum* sp.	3.767753	443	37.2	11,273	SAMN41711470
samp_4444_maxbin_761	Mbita West	91.61	0.81	33.33	*Dolichospermum* sp.	3.834118	345	37.1	15,823	SAMN41711477
samp_4445_metadecoder_820	Bridge Island	95.11	0.44	50	*Dolichospermum* sp.	3.814968	232	37.1	24,166	SAMN41711487
samp_4446_metabat2_471	Bondo	90.07	9.23	58.33	*Dolichospermum* sp.	4.539608	629	36.9	9,935	SAMN41711489
samp_4432_maxbin_231	Kisumu Harbor	97.82	0.95	80	*Planktothrix* sp.	5.017227	441	39.3	19,782	SAMN41711420
samp_4434_metabat2_31	Sondu Miriu	98.69	1.24	100	*Planktothrix* sp.	5.003513	421	39.2	18,186	SAMN41711432
samp_4435_metabat2_87	Southern Mid-Gulf	98.47	1.69	63.64	*Planktothrix* sp.	5.119342	423	39.2	18,944	SAMN41711440
samp_4443_maxbin_260	Mbita East	91.67	0.44	33.33	*Sphaerospermopsis kisseleviana* A	4.743136	543	37.6	11,832	SAMN41711466
samp_4444_maxbin_786	Mbita West	92.89	1.12	83.33	*Sphaerospermopsis kisseleviana* A	4.915663	624	37.6	10,851	SAMN41711478
samp_4445_VAMB_547	Bridge Island	89.96	7.53	77.55	*Sphaerospermopsis kisseleviana* A	6.110041	1,452	37.5	6,049	SAMN45857663
samp_4448_concoct_1871	Rusinga Channel	93.7	1.78	77.78	*Sphaerospermopsis kisseleviana* A	5.020957	595	37.7	10,901	SAMN41711493

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

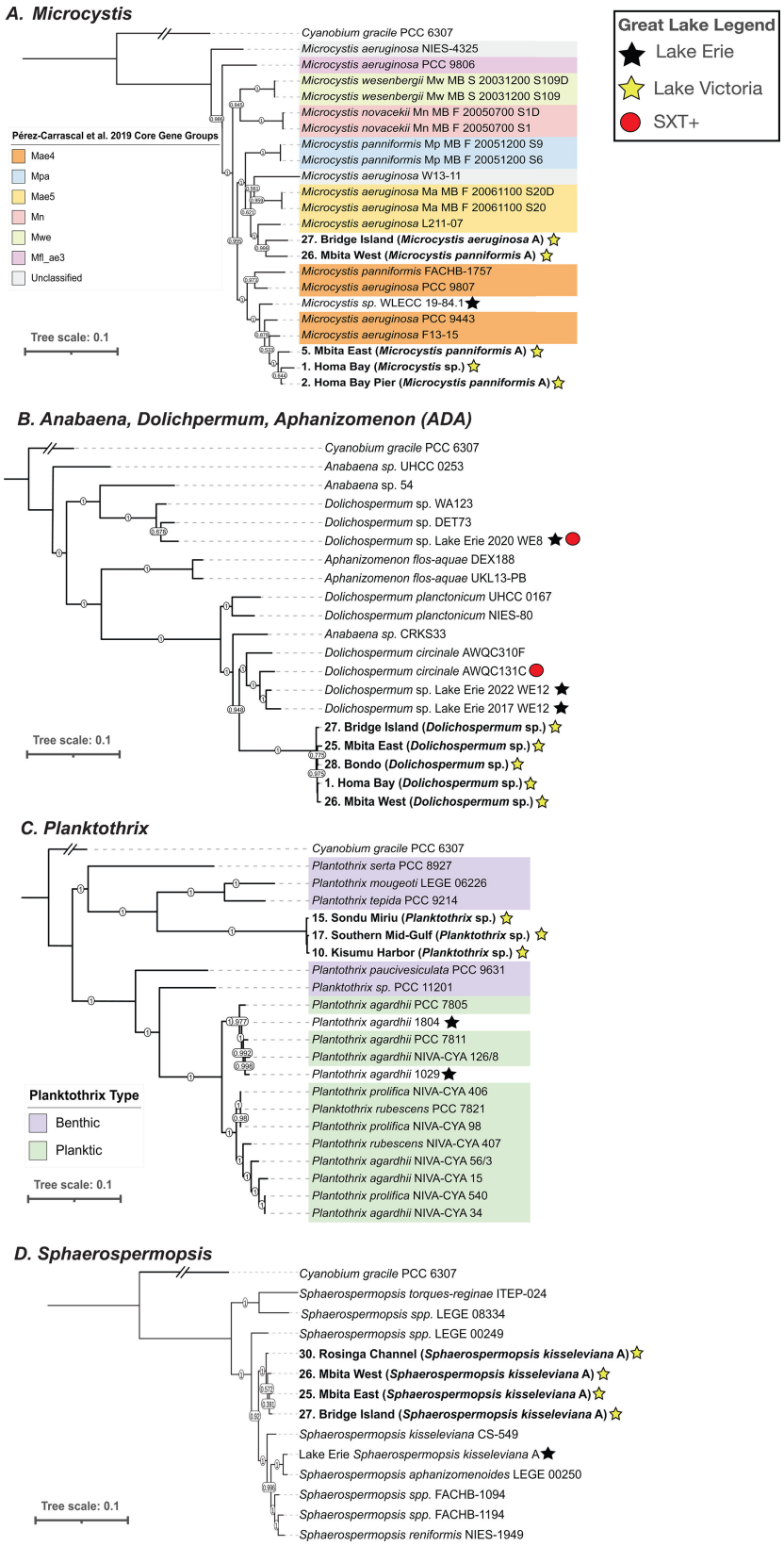

FIG 5 Phylogenomic analysis of dominant bloom-forming cyanobacteria from sampled sites with taxonomic annotations from Genome Taxonomy Database (GTDB) (40). (A) *Microcystis* phylogenomic tree built with five representative MAGs, reference genomes, and *Cyanobium gracile* PCC 6307 as the

(Continued on next page)

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

Fig 5 (Continued)

outgroup. Colored ranges indicate clades as outlined in reference (36). (B) *Dolichospermum* phylogenomic tree built with five representative MAGs, reference genomes, and *Cyanobium gracile* PCC 6307 as the outgroup. The presence of a red circle is indicative of the presence of the saxitoxin gene cluster within the genome (38). (C) *Planktothrix* phylogenomic tree built with three representative MAGs, reference genomes, and *Cyanobium gracile* PCC 6307 as the outgroup. Benthic or planktic designations for reference genomes are indicated as colored ranges (41). (D) *Sphaerospermopsis* phylogenomic tree built with four representative MAGs, reference genomes, and *Cyanobium gracile* PCC 6307 as the outgroup.

(NRPSs), a pathway known for its synthesis of cyanopeptides like anabaenopeptins and aeruginosins (48, 49). Cyanobactin BGCs encoded by *Microcystis* and *Dolichospermum* formed gene cluster families (GCFs) with known anacyclamide BGCs from the MiBIG database (Fig. S2). BGCs putatively encoding microviridin from the ribosomally synthesized and post-translationally modified peptide (RiPP) biosynthetic class were encoded by *Planktothrix*, *Microcystis*, and *Sphaerospermopsis* (Fig. S2). *Microcystis* was the only cyanobacterial genus with MAGs encoding microcystins (Fig. S2 and S3). BGCs that could not be grouped into these larger biosynthetic classes were categorized as "Other" and were putatively annotated as RiPP recognition element (RRE)-containing, ectoine, lassopeptide, spliceotide, arylprolene, crocagin, and mycosporine-like.

Biosynthetic potential converges with site-specific cyanobacterial composition

The relative abundance of cyanobacterial BGCs varied across the gulf (Fig. 9). BGC composition at each site converged with cyanobacterial composition, particularly with the dominating cyanobacterial genera at each site (Fig. S4). Specifically, sites with a high abundance of bloom-forming cyanobacteria (1–4, 23, and 27–29) had increased abundances of most BGC classes and clustered together based on the site's biosynthetic potential. BGCs putatively encoding pathways related to Anacyclamide-like, cyanobactin, heterocyst glycolipids, Lanthipeptide class V, NRPS, NRPS-like, RRE-containing, type 1 polyketide synthase, RiPP-like, spliceotide, terpene biosynthetic classes, and pathways were the most common and abundant BGC types across sites.

DISCUSSION

The Winam Gulf of Lake Victoria experiences prolific, year-round cyanoHABs, previously shown to be dominated by *Microcystis* with periodic occurrences of *Dolichospermum* based on morphological classifications (17, 50). However, due to limited molecular research on these blooms, little is known about the bloom-forming cyanobacterial function and biosynthetic potential supporting year-round growth in this region. In this study, we employed metagenomics to establish a molecular snapshot of the microbial communities found along spatial and physicochemical gradients in the Winam Gulf between 2 consecutive years. Subsequently, we investigated the functional and biosynthetic potential encoded by bloom-forming cyanobacteria in the gulf to determine potential strategies used by these organisms to grow and produce toxic metabolites.

Nitrogen limitation may provide a competitive advantage for diazotrophs during cyanoHABs

Two, large cyanoHAB events dominated mainly by *Dolichospermum* (56%–91% of the cyanobacterial community) and *Microcystis* (14%–39% of the cyanobacterial community) were detected near Homa Bay in 2022 and outside the gulf near Bondo and the Yala River mouth in 2023 (Fig. 1). Toxic cyanoHAB events, due to microcystins, in Homa Bay have presented exposure risks to this fishing community of ~60,000 residents (21, 28, 51). CyanoHABs have also been previously reported in the Yala River mouth region, which has excessive turbidity, declining water quality, and is a large source of anthropogenic

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

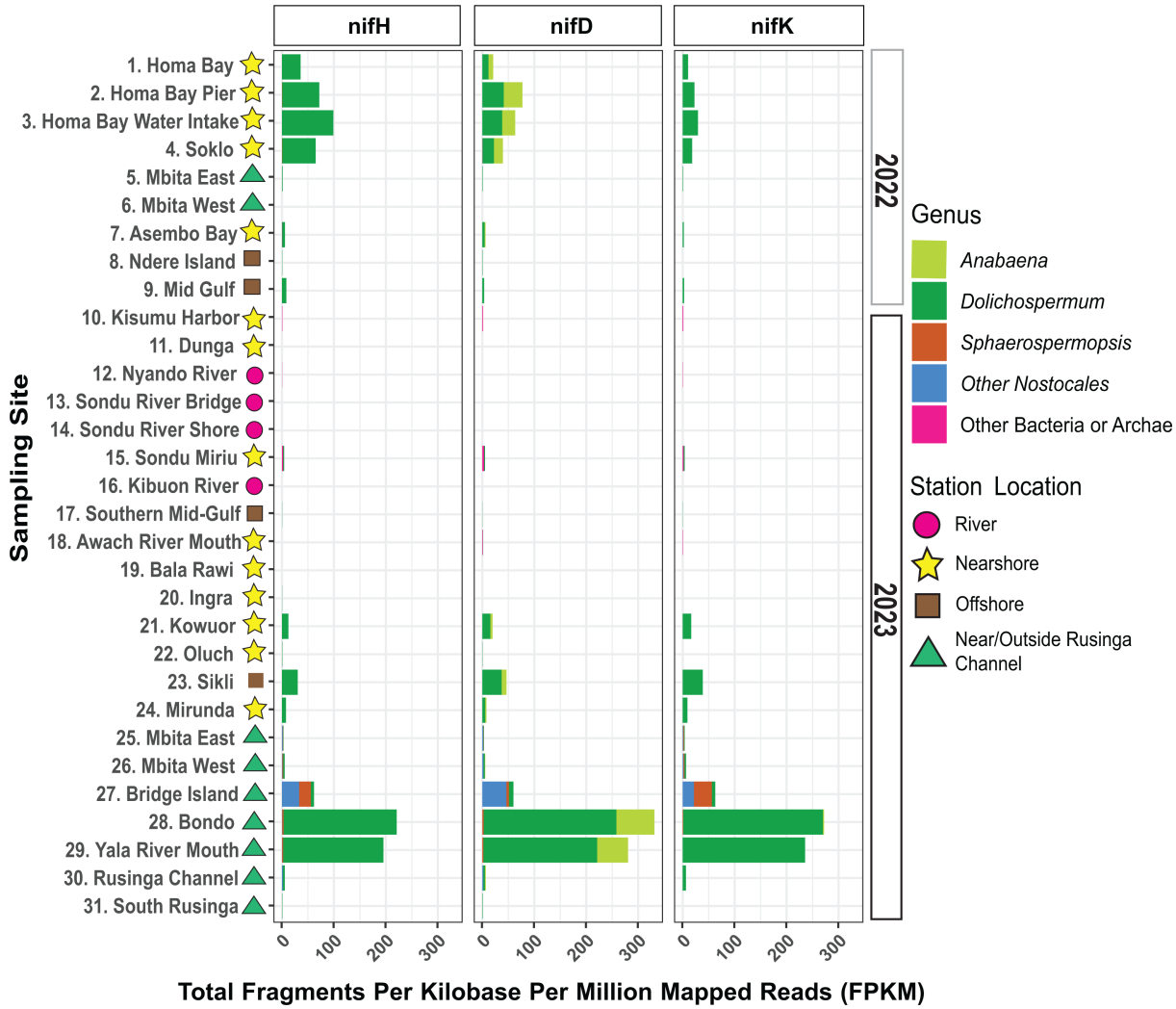

FIG 6 Relative abundance of *nifH*, *nifD*, and *nifK* genes across sites. Reads were aligned to *nifH*, *nifD*, and *nifK* genes from NFixDB and normalized using the total fragments per kilobase per million mapped reads (FPKM) metric (Equation 1) (43). The y-axis shows site names with their location descriptor symbol. The x-axis shows the FPKM of each *nif* gene in each sample. Bars are colored based on the genus the *nif* gene was extracted from.

pollution to Lake Victoria (52, 53). While this region is not as regularly monitored for cyanoHABs and cyanotoxins due to its location outside the gulf, our results provide further evidence that it should be considered in future monitoring schemes.

Organisms from the ADA clade and *Sphaerospermopsis* harbored essential genes for nitrogen fixation, such as the nitrogenase operon (*nif*HDK) and heterocyst glycolipid synthesis gene clusters, where cyanoHABs were detected in 2022 and 2023 (54). To our knowledge, this is the first report of the genus *Sphaerospermopsis*, a potentially toxigenic and diazotrophic cyanobacterium, in the Winam Gulf. These results suggest that diazotrophic cyanobacterial genera might be employing nitrogen fixation as a competitive advantage over non-diazotrophic competitors that are reliant on fixed forms of nitrogen. Low TN:TP ratios at cyanoHAB sites in 2023 (34) and 2005 (26) also indicate nitrogen limitation, with less severe nitrogen limitation closer to and outside the Rusinga channel (34). Although TN:TP was not measured in 2022, dissolved nitrate:soluble reactive phosphorus (DN:SRP) ratios were all below 1, suggestive of nitrogen deficiency (17). These results concur with Brown et al., which identified *Dolichospermum* as a dominant bloom-forming cyanobacteria in the gulf in 2022 but differ from previous reports that identified *Microcystis* as the dominating bloom-forming cyanobacteria in

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

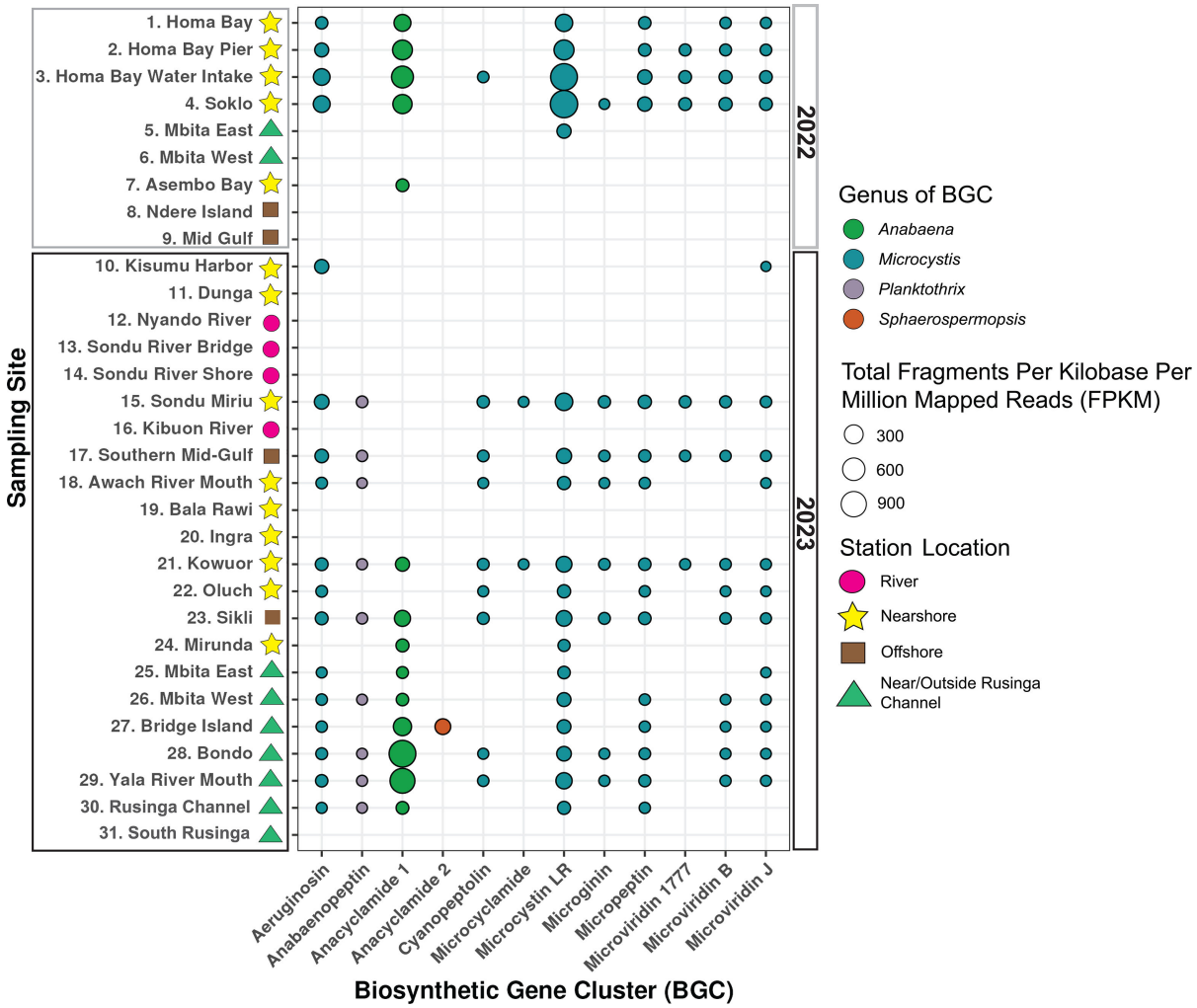

FIG 7 Known cyanopeptide BGC relative abundance across sites. Station numbers, names, and their location indicator symbols are shown on the y-axis, with BGCs encoding the biosynthesis of cyanopeptides from MIBiG (accessed December 2023) shown on the x-axis (44). The FPKM (total fragments per kilobase per million mapped reads) of each whole BGC is shown as bubble size (Equation 2). Bubbles are colored based on the cyanobacterial genus from which the reference BGC originates.

most seasons since 2008 in this region (17, 28, 55). Investigations of the phytoplankton community of the gulf in 2020 also identified *Microcystis* at slightly higher abundance than *Dolichospermum* in May and June (29). Data collection on an annual basis is vital to elucidate how changing landscapes and populations in this region are influencing cyanoHAB community structures with different toxin threats that require variant monitoring and management approaches.

According to additional studies on subtropical water bodies experiencing cyanoHABs made up of both *Microcystis* and diazotrophic cyanobacteria, high *nif*HDK (nitrogenase) gene expression was observed after large *Microcystis* blooms, with diazotrophs potentially acting as "replenishers" of bioavailable nitrogen (56). Gene expression data are needed to identify if diazotrophic cyanobacteria in the Winam Gulf serve this role in the ecosystem and could help explain the gulf's seasonal bloom succession patterns.

Non-diazotrophic cyanobacteria were identified at lower abundance in gulf regions with distinct physicochemical conditions

Microcystis and *Planktothrix*, non-diazotrophs, were at lower relative abundance at sites with large cyanoHAB events (between 1% and 39% of the cyanobacterial community).

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

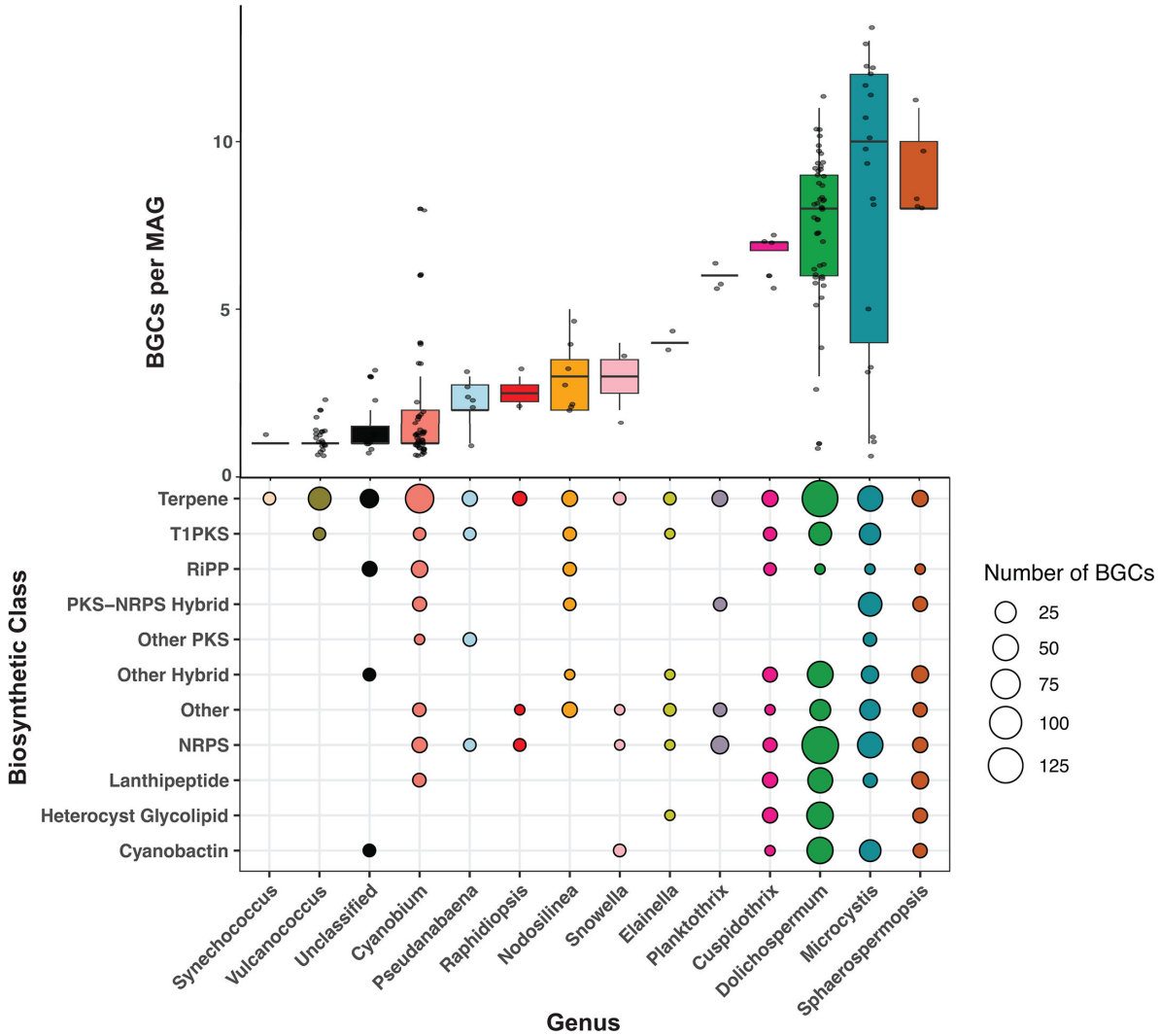

FIG 8 BGC richness and diversity in dominant cyanobacterial genera from the gulf. The top panel shows the average number of BGCs per MAG from 14 different cyanobacterial genera. The bottom panel shows the number of BGCs (bubble size) from each biosynthetic class (y-axis) identified from each cyanobacterial genus (x-axis) by antiSMASH v7.0 on the x-axis (46). Genus names are not italicized for readability.

However, *Microcystis* was observed at higher relative abundance at sites in 2023 with higher turbidity on average (between 36% and 59% of the cyanobacterial community). Although these sites were nitrogen limited, the high turbidity at these sites may give *Microcystis* an advantage in comparison to deeper and less turbid regions of the gulf. This is due to *Microcystis*' ability to form gas vesicles and thus reach the surface to circumvent light limitation, plausibly caused by the shallow depth and well-mixed water column at these sites, therefore generating shading for their non-buoyant competitors (57). *Planktothrix*, which was found to have close genomic similarity to previously characterized benthic *Planktothrix* isolates, was the dominant cyanobacterial genera in Kisumu Harbor, the shallowest site in the study (5.4 m), at the time of sampling (up to 67% of the cyanobacterial community). No *Planktothrix nif* genes were identified, indicating that *Planktothrix* was likely using strategies other than diazotrophy to overcome nitrogen limitation and proliferate in this region of the gulf (41, 58).

At sites where cyanobacteria were detected at lower abundance compared to other microbial phyla, *Cyanobium* and *Vulcanococcus* (picocyanobacteria) were the most abundant cyanobacterial genera, each making up almost half of the cyanobacterial community at these sites. There is little documentation of the picocyanobacterial

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

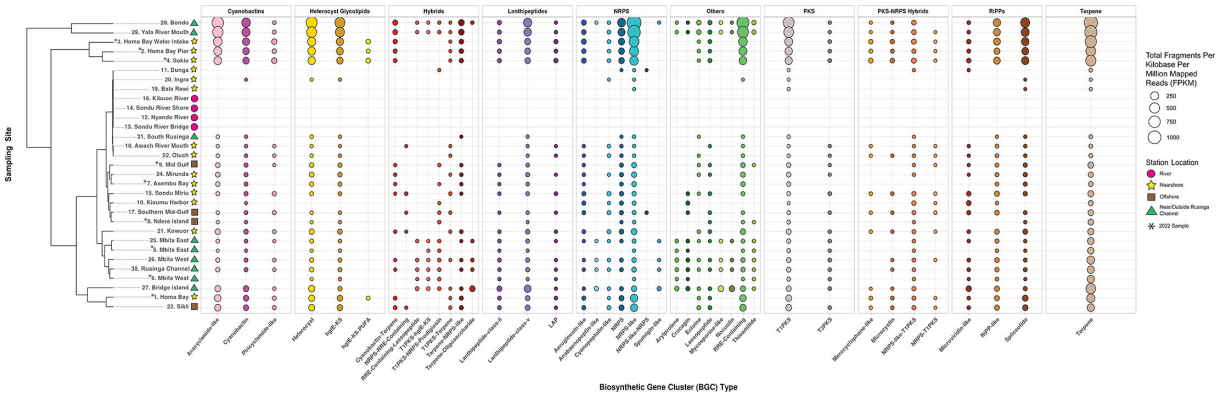

FIG 9 Total BGC relative abundance across the Winam Gulf. Station numbers, names, and their location indicator symbols are shown on the y-axis, with BGC groups shown on the x-axis. BGCs are grouped into broader biosynthetic or product classifications, labeled as the top facet strips (Table S3). The total fragments per kilobase per million mapped reads (FPKM) of each BGC was calculated (Equation 2), and the size of a point corresponds to the total FPKM of all BGCs of a particular BGC type indicated on the x-axis. Bubbles are colored to differentiate BGC groups. Samples on the y-axis are hierarchically clustered using the complete-linkage method.

community in the Winam Gulf likely due to their small cell size and thus difficult morphological classification (59). *Cyanobium* is cosmopolitan, found across the globe, and is well known for its ability to scavenge nitrogen and use light efficiently, potentially leading to its success in this ecosystem, while *Vulcanococcus* is not as well documented (60, 61). While picocyanobacteria typically do not have many BGCs, MAGs recovered in this study do encode some BGCs from terpene, PKS, and NRPS biosynthetic pathways, with *Cyanobium* encoding a far more diverse set of BGCs than *Vulcanococcus*. This finding underscores the need to include picocyanobacteria in future investigations of cyanoHABs in the Winam Gulf and freshwater systems broadly as it relates to bloom dynamics and human health (62).

Low intra-genus diversity among bloom-forming cyanobacteria in the gulf compared to temperate systems

The main bloom-forming cyanobacteria (*Dolichospermum*, *Microcystis*, *Sphaerospermopsis*, and *Planktothrix*) showed low intra-genus diversity across the sites, recovering only 3–5 MAGs from each genus. For *Microcystis* and *Planktothrix*, this result contrasts with findings from Lake Erie, where isolates from these genera share much lower average nucleotide identities (37, 63). Conversely, the results for *Microcystis* align with findings from subtropical lakes like Lake Okeechobee (Florida, USA) and Lake Taihu (China). Low intra-genus diversity was identified by Krausfeldt et al. (64) in Lake Okeechobee, where only three *Microcystis* MAGs, two of which were annotated as *Microcystis* panniformis (a commonly found species in Lake Taihu), were recovered from a large data set of metagenomes (65). Three out of five of the non-redundant *Microcystis* MAGs in our study were annotated as *M. panniformis*, while the other two MAGs were annotated as *M. aeruginosa* and *Microcystis* sp. (Fig. 5). *M. panniformis* and *M.aeruginosa* are known to share high levels of genomic similarity but differ in their colony morphology, with *M. panniformis* often forming larger colonies, potentially giving them an advantage in shading out their competitors (66). Together, these data show low intra-genus diversity and the common occurrence of *M. panniformis* across subtropical and tropical lakes, indicating a potential ecological pattern across lakes with shared climates (distinct dry and rainy seasons, warm temperatures, high humidity, and reduced seasonality). This data could not be compared directly to other studies on African Great Lakes due to limited data availability, highlighting these high-quality MAGs as a foundational resource for metagenomic research on cyanoHABs in this region.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

BGCs encoding the synthesis of cyanotoxins, cyanopeptides, and unknown metabolites found across the gulf, even in areas without visible cyanoHABs

The diverse makeup of cyanobacteria within the Winam Gulf poses threats not only due to their noxious smells and effects on fishing, tourism, recreation, and wildlife but also due to their ability to produce toxins that are harmful to humans and wildlife (21, 51). We identified the *mcy* BGC responsible for the synthesis of microcystin at over half of the sites and definitively ascribed its source for the first time, solely to *Microcystis*, resolving speculation that *Microcystis* or *Dolichospermum* could harbor these genes (28, 67). The expression of the *mcy* BGC has been shown to increase in response to temperature, presence of reactive oxygen species, light, nitrogen forms and availabilities, and growth rate, though factors that constrain the biosynthesis of microcystins remain convoluted (68, 69). The attribution of microcystin synthesis to *Microcystis* within this tropical, warm, eutrophic basin is important in accurate cyanotoxin human exposure management and prevention, especially in areas where raw water is used for drinking and domestic tasks. A previous study in the Winam Gulf identified high levels of the amidinotransferase gene (*cyrA*) in the cylindrospermopsin-encoding BGC near the Nyando River mouth in June 2022 using quantitative PCR (28, 45). In this study, these genes were not detected at high levels, and our methods were not able to assemble the cylindrospermopsin BGC or link it to an organism.

The bloom-forming cyanobacteria present in the gulf encoded the ability to produce other nitrogen-rich, hepatotoxic cyanopeptides. These BGCs were identified at over half of the sites in the study, including those that did not have a large, visible cyanoHAB event. Many of the sites in this study are heavily used for fishing, aquaculture, bathing, washing clothes, and potable water, underscoring the importance of communicating the dangers of non-visible cyanoHABs to local residents (21, 51). With changing conditions across rainy and dry seasons in the gulf, the composition and expression of these BGCs may change, producing diverse cocktails of harmful molecules beyond microcystins. Such seasonal expression changes were observed in a study in a subtropical river (Florida, USA) where the cyanopeptolin BGC, also identified in our study, was 80 times more expressed than the *mcy* operon in the winter season, while this relationship inverted in the spring season (56). BGC expression studies in tropical regions are needed to begin making hypotheses about metabolites' complex functional roles and environmental triggers. Moreover, the human health effects of these molecular combinations remain unknown, and BGC expression studies across both tropical and temperate systems indicate the expression of multiple BGCs at any given time (11, 70).

In addition to cyanopeptides and cyanotoxins, cyanobacteria in the gulf encoded over 300 unique BGCs, most of which remain cryptic. These findings illustrate the rich ability of Winam Gulf cyanobacteria to produce uncharacterized metabolites with unknown but potentially harmful or therapeutic properties. The number and diversity of BGCs per MAG in the Winam Gulf were consistent with results from cyanobacteria in Western Lake Erie (37). The diverse biosynthetic repertoire of these bloom-forming cyanobacteria is both an opportunity for discovery and has implications for protecting human health.

Conclusion

This study characterizes the cyanobacterial assemblage, function, and biosynthetic potential across different regions of the Winam gulf. The metagenomic data revealed different survival strategies available for use by major, bloom-forming cyanobacteria to harness distinct environments and conditions around the gulf, including extensive use of BGCs. Omics techniques (metagenomics, metatranscriptomics, and metabolomics) should be incorporated into current monitoring schemes in the gulf and other African Great Lakes impacted by toxic cyanobacteria to increase research capacity on cyanoHAB ecology and dynamics, inform water management plans, and protect human and ecosystem health. A multi-faceted monitoring plan that incorporates these techniques

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

will look different in various regions depending on the level of cyanoHAB exposure potential and available resources, personnel, and expertise. Moreover, filling this data gap is vital in understanding the biogeographical differences in cyanoHABs, specifically how cyanoHAB events in the global south differ from those in more temperate regions of the world. For example, our results suggest low genomic diversity of dominant cyanobacterial taxa in the gulf, differing from what is observed in temperate regions like the Laurentian Great Lakes but converging with findings from subtropical lakes like Lake Okeechobee and Lake Taihu. Insights from a diverse array of geographic regions will shed valuable insight into how cyanoHABs may respond to a rapidly warming climate across latitudinal gradients.

MATERIALS AND METHODS

Data collection

Study site and sample collection

This study was conducted on 24–25 June 2022 and 30 May to 3 June 2023 in the Winam Gulf. Thirty-one water samples were collected from a depth of 1 meter using a 2.5 L Van Dorn sampler. Microbial biomass was collected by filtering water (between 60 and 300 mL) through a 0.22 µm Polyethersulfone (PES) Sterivex cartridge filter (MilliporeSigma, Burlington, MA) and preserved in 1.5 mL of Zymo DNA/RNA shield (Zymo, Orange, CA) and placed in a −20°C freezer. Water samples for dissolved nutrient analyses were filtered through a 0.2 µm filter, while unfiltered water was collected for total nutrient analyses and frozen at −20°C. Water was filtered through GF/F filters and stored in a sterile tube for chlorophyll analysis. Nutrient, DNA, and chlorophyll samples were shipped from Kisumu, Kenya at room temperature back to University of Tennessee Knoxville in 2022 and University of Michigan in 2023 for DNA extraction, and Baylor University for nutrient and chlorophyll analyses in 2023. These shipments were accompanied by a letter of support from Kenya Marine and Fisheries Research Institute and/or a United States Department of Agriculture (USDA) permit from David H. Sherman. Physicochemical parameters (i.e., water temperature, dissolved oxygen, conductivity, total dissolved solids, turbidity, and pH) were collected using a multiparameter sonde (YSI Inc., Yellow Springs, OH) on site. Total cyanobacterial abundance was measured using an AlgaeTorch (bbe moldaenke GmnH, Schwentinental, Germany).

DNA extraction and sequencing

All sequencing data for this manuscript have been summarized previously in a Microbiology Resource Announcement by Zepernick and Hart et al. (71) and are briefly described below (71). For samples from 2022, DNA was extracted from the Sterivex cartridges after flushing all DNA/RNA shields out of the filter cartridge using phosphate-buffered saline (PBS), prepared according to Cold Spring Harbor's protocol (72). Duplicate samples from each site in 2022 were collected, extracted, and sequenced. DNA extractions were performed at the University of Tennessee Knoxville using standard phenol-chloroform methods and subsequent ethanol precipitation as reported previously (73). DNA quality was assessed via the Nanodrop ND-100 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) to ensure 260/280 ratios were sufficient (260/280 > 1.80). DNA concentration was quantified using the Qubit double-stranded DNA HS assay kit (Invitrogen, Carlsbad, CA). Extracted DNA was submitted to the University of Minnesota Genomics Core for genomic library construction and sequencing. Sequencing was completed on an Illumina Novaseq system equipped with an S4 flowcell with 150 cycles (150-base pair paired-end reads; Illumina, San Diego, CA).

For samples from 2023, DNA was extracted from the Sterivex cartridges after flushing all DNA/RNA shields out of the filter cartridge using PBS, prepared according to Cold Spring Harbor's protocol (72). DNA extractions were performed at the University of

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

Michigan. Filter cartridges were opened using a pipe cutter, and the filter membrane was carefully removed using sterile forceps. DNA from the filters was extracted using the DNeasy Powerwater Sterivex Kit (Qiagen, Germantown, MD) according to the manufacturer's protocol. Quantity was assessed using the high-sensitivity DNA qubit kit (Invitrogen, Carlsbad, CA). Extracted DNA was submitted to the University of Michigan Advanced Genomics Core for genomic library construction and sequencing. Sequencing was completed on an Illumina NovaX system equipped with a 10B flowcell with 300 cycles (150-base pair paired-end reads; Illumina, San Diego, CA).

Nutrient analyses

Water samples filtered through Whatman GF/F filters pre-combusted at 550°C for 4 hours were collected in new 50 mL Falcon Tubes and frozen at −20°C until analysis. Filtered dissolved nutrient samples were analyzed for nitrate-nitrogen (NO_3-N), ammonium-nitrogen (NH_4-N), and soluble reactive phosphate (PO_4-P), using Cd reduction, phenate, and molybdate methods, respectively (74). Total dissolved nitrogen and total dissolved phosphorus were analyzed from filtered samples using the Cd reduction and molybdate methods (75). Unfiltered samples were analyzed for TN and TP using the Cd reduction and molybdate methods (75). All soluble and total nutrients were analyzed using a Lachat QuickChem 8500 Series II autoanalyzer (Lachat Instruments, Hach Company) at the Center for Reservoir and Aquatic Systems Research located at Baylor University, Waco, TX. Chlorophyll a in extracts was measured using the non-acidification method of Welschmeyer on a Turner Designs Trilogy fluorometer calibrated with pure Chlorophyll a standards (Turner Designs, Sunnyvale, CA, USA) (76).

Bioinformatic analysis

Metagenomic data processing

Metagenomes were processed with the Great Lakes Atlas for Multi-Omics Research (GLAMR) pipeline (GLAMR Portal, GLAMR Github). fastp v0.23.2 was used to deduplicate (--dup_calc_accuracy 6), remove adapters (--detect_adapter_for_pe), trim (--cut_front --cut_tail --cut_window_size = 4—cut_mean_quality --cut_mean_quality 20), and filter (--length_required 50 n_base_limit 5—low_complexity_filter --complexity_threshold 7) raw sequence reads (77). BBmap was used to remove human contaminant reads mapping to the GENCODE release 38 human genome (78). Each sample was assembled using Megahit v1.2.9 with the meta-sensitive parameter (79). Assembled contigs longer than 2,000 base pairs were used for binning. Concoct v1.1.0, MaxBin2 v2.2.7, Metabat2 v2.17, MetaDecoder v1.0.13, SemiBin v1.0.3, and VAMB v3.0.8 were used to bin contigs into MAGs (80–85). MAGs were clustered and dereplicated at 98% identity using dRep v3.2.0 (86, 87). MAG quality was assessed with CheckM v.1.2.2, and MAG taxonomy was assigned using GTDBtk and the GTDB release 214 database (40, 88).

Community composition analysis

MAGs from all 31 samples were grouped and dereplicated using dRep v3.2.0 at 98% (87). MAG quality statistics, calculated with checkM v1.2.2 and bakta v1.8.1, can be found in Table S2 (88, 89). Quality-controlled (QC) reads were mapped to the dereplicated set of MAGs using CoverM v0.6.1 (using the default mapper: MiniMap2 v2.24) to get the relative abundance of each MAG in each sample (Fig. 3) (90, 91). Observations from duplicate samples from 2022 were averaged in all applicable metagenomic methods in the study due to the replicate's similarity (Fig. S5). A Bray-Curtis dissimilarity matrix was calculated from the relative abundance of all cyanobacterial genera in each sample and assessed using NMDS using the vegan v2.6 package in the R statistical software (92). Cyanobacterial makeup was assessed in relation to environmental parameters using the envfit function in the vegan v2.6 package in the R statistical software (Fig. 4A) (92). The correlation of the six most abundant genera with each other and associated

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

environmental parameters was assessed using Spearman's correlation coefficient in the R statistical software.

Phylogenomic analysis

High-quality MAGs (checkM: ≥90% completeness and ≤10% contamination) annotated by GTDBtk as *Microcystis*, *Dolichospermum*, *Planktothrix*, and *Sphaerospermopsis* were dereplicated at 99.5% identity with dRep v3.2.0 to produce representative bins for each of these genera (40, 87, 88). A tree was built using these MAGs and reference genomes from NCBI (Table S6) using GToTree v1.7.05 (93, 94). A maximum likelihood tree for each taxa using 251 marker genes was generated. Reference genomes for the *Microcystis* (Fig. 5A), ADA (Fig. 5B), and *Planktothrix* (Fig. 5C) trees were sourced from prior studies (36, 37, 63, 95, 96). Reference genomes for the *Spaherospermopsis* tree (Fig. 5D) were chosen from the NCBI genome database if they were characterized as either a reference genome or a strain (93). An outgroup represented by *Cyanobium gracile PCC 6307* (NCBI: PRJNA158695) was used in each of the four trees. Phylogenomic trees were visualized in Interactive Tree of Life (ITOL) (97).

Nitrogen fixation genomic potential detection

QC reads from each sample were mapped to the *nif*H, *nif*D, *nif*K, *Chl*N, and *Chl*B databases in NFixDB (accessed March 2024, Database) using MiniMap2 v2.24 (43, 91). Hits to *Chl*N and *Chl*B were excluded but were used in the read mapping step to ensure a competitive database was used (98). Reads that mapped to the reference with 80% coverage and 90% identity were kept. Hits to *nif*H, *nif*D, and *nif*K were kept if 80% of the gene was covered with 90% identity (Fig. 6). *nif*HDK was used to ensure the identification of true nitrogen fixers and not pseudo-*nif*H-containing microbes (42). Total fragments per kilobase per million mapped reads (FPKM) was calculated for each gene in each sample using Equation 1 (99).

$$\text{Total Fragments per Kilobase per Million Mapped Reads (FPKM)} = \frac{\text{total no. of mapped reads}}{(\text{ length of nif gene } /1,000) \times (\text{ total no. of reads per sample })} \times 1,000,000 \quad (1)$$

Cyanopeptide and cyanotoxin BGC detection

QC reads from each sample were mapped to all BGCs in the MIBiG database from the cyanobacterial genera *Microcystis*, *Dolichospermum*, *Planktothrix*, *Sphaerospermopsis*, *Anabaena*, *Aphanizomenon*, *Cylindrospermopsis*, and *Cyanobium* using MiniMap2 v2.24 (44, 91). Reads that mapped to the reference with 80% coverage and 90% identity were kept. Hits to BGCs were kept if 70% of the BGC was covered by reads in that sample. FPKM was calculated for each BGC in each sample using Equation 2 (99).

$$\text{Total Fragments per Kilobase per Million Mapped Reads (FPKM)} = \frac{\text{total no. of mapped reads}}{(\text{ length of BGC } /1,000) \times (\text{ total no. of reads per sample })} \times 1,000,000 \quad (2)$$

Mcy and cyr operon identification and comparison to references

All assemblies in the data set were searched for the presence of genes from the *mcy* (MiBIG BGC: BGC0001016) and *cyr* (MiBIG BGC: BGC0000978) operons via BLASTn (44, 100). If genes from either of these operons were present in the assembly of a sample, the sample's MAGs were analyzed for all genes from the operons using BLASTn. Hits to genes of interest were kept if they had 80% identity and coverage in the queried MAG. MAGs with multiple genes from the reference operon present were analyzed in Anvi'o to identify if the whole reference operon was present and if it was correctly assembled and binned into the MAG (101). The *cyr* operon could not be recovered from this data set, but low levels of the core biosynthetic genes of the *cyr* operon (MiBIG

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

BGC: BGC0000978) were detected via BLASTn at Site 27. Bridge Island (Fig. S1) (100). The *mcy* operon was detected in multiple MAGs, all identified as *Microcystis* by GTDB-Tk (40). One MAG from Site 1. Homa Bay (NCBI Accession: SAMN41711332) was chosen for MAG analysis in Anvi'o to confirm the taxonomic identity of the contig with the *mcy* operon (Fig. S3). Additionally, the *mcy* operon detected in the MAG was compared to the *mcy* operon from *Microcystis aeruginosa* FACHB 1326 (GenBank: OQ291092.1) using the program "clinker" (Fig. S3) (102).

Biosynthetic gene cluster mining and abundance analysis

BGCs were mined from medium- and high-quality cyanobacterial MAGs (≥70% completion and ≤30% contamination) using antiSMASH v7.0 (46). All BGCs greater than 5,000 base pairs long were extracted and dereplicated at 99% using dRep v3.2.0 (Table S6) (87). Gene clusters were grouped into GCFs using BiG-SCAPE and visualized in cytoscape v3.9.1 ((Fig. S1) (103, 104). QC reads from each sample were mapped to the dereplicated set of BGCs using MiniMap2 v2.24 (91). Reads that mapped to the reference with 80% coverage and 90% identity were kept. Hits to BGCs were kept if 70% of the BGC was covered by reads in that sample (Fig. 9). FPKM for each BGC was calculated using Equation 2. Total FPKM for each BGC type (x-axis) was calculated by summing the FPKM of each BGC belonging to that BGC type per sample.

Statistical analysis and figure rendering

All statistical analysis and plots were created using R and R studio v4.3.3 (105). The R packages vegan, tidyverse, and RColorBrewer were used for figure creation (92, 106, 107).

ACKNOWLEDGMENTS

We thank the Kenya Marine and Fisheries Research Institute (KMFRI), Kisii University, Technical University of Kenya, and captain Frederick Okello for their collaboration, access, and assistance while on the RV Uvumbuzi. We thank Robert Hein for bioinformatic support.

This work was funded by the National Science Foundation (NSF) International Research Experiences for Students (IRES) award #1953468, National Institute of Health (NIH)-NSF awards 1P01ES0289-01 and 1840715 supporting the Great Lakes Center for Fresh Waters and Human Health at Bowling Green State University, and NIH F31 1F31ES036421-01.

The 2022–23 NSF-IRES Lake Victoria Research Consortium participants include Dorine Achieng (Kenya Marine and Fisheries Research Institute, Kisumu, Kenya), Katelyn B. Barker (Bowling Green State University, Bowling Green, Ohio, USA), George M. Basweti (Kenya Marine and Fisheries Research Institute, Kisumu, Kenya), Max Beal (University of Wisconsin—Madison, Madison, Wisconsin, USA), Aidan Byrne (King's College London, London, United Kingdom), William R. Cody (Aquatic Taxonomy Specialists, Malinta, Ohio, USA), Linet I. Kiteresi (Kenya Marine and Fisheries Research Institute, Kisumu, Kenya), Theodore Lawrence (African Center for Aquatic Research and Education), Jared B. Miruka (Kenya Marine and Fisheries Research Institute, Kisumu, Kenya), Samantha Mohney (George Mason University, Fairfax, Virginia, USA), Pamela Okutoyi (Technical University of Kenya, Nairobi, Kenya), Dennis Otieno (Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada), Omondi A. Owino (Sigalagala National Polytechnic, Kakamega, Kenya), Winnie Owoko (Kenya Marine and Fisheries Research Institute, Kisumu, Kenya), Bethwell Owuor (Kisii University, Kisii, Kenya), Anakalo Shitandi (Kisii University, Kisii, Kenya), James Njiru (Kisii University, Kisii, Kenya), Reuben Omondi (Kisii University, Kisii, Kenya), Jordyn Stoll (Michigan Trout Unlimited, Dewitt, Michigan, USA), Mariam N. Swaleh (Technical University of Mombasa, Mombasa, Kenya), Emma J. Tebbs (King's College London, London, United Kingdom), Emily Varga (Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada), Jack Abibo Adem (Jaramogi Oginga Odinga University of

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

Science and Technology, Bondo, Kenya), Anjana Adhikari (University of Wisconsin—Milwaukee, Milwaukee, Wisconsin, USA), Trinity Allan (Florida Gulf Coast University, Fort Myers, Florida, USA), Mercy Chepkirui (Kenya Marine and Fisheries Research Institute, Kisumu, Kenya, and Kisii University, Kisii, Kenya), Nusrat Nasrin Khan (Arizona State University, Tempe, Arizona, USA), Martha Moseti (Maasai Mara University, Narok, Kenya), Tonny Achieng (Kisii University, Kisii, Kenya), and Lisa Radock (Fort LeBoeuf School District, Erie, Pennsylvania, USA). Authors Lauren N. Hart, Brittany N. Zepernick, Kaela E. Natwora, Katelyn M. Brown, Julia Akinyi Obuya, Davide Lomeo, Malcolm A. Barnard, Mark Olokotum, R. Michael McKay, Ken G. Drouillard, Kefa M. Otiso, and George S. Bullerjahn are also members of the consortium.

AUTHOR AFFILIATIONS

[1]Program in Chemical Biology, University of Michigan, Ann Arbor, Michigan, USA
[2]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, USA
[3]Great Lakes Center for Fresh Waters and Human Health, Bowling Green State University, Bowling Green, Ohio, USA
[4]Department of Microbiology, The University of Tennessee Knoxville, Knoxville, Tennessee, USA
[5]Large Lakes Observatory, University of Minnesota Duluth, Duluth, Minnesota, USA
[6]Biological Sciences, Bowling Green State University, Bowling Green, Ohio, USA
[7]Kenya Marine and Fisheries Research Institute, Kisumu, Kenya
[8]Department of Geography, King's College London, London, United Kingdom
[9]Department of Biology, Baylor University Department of Biology, Waco, Texas, USA
[10]Center for Reservoir and Aquatic Systems Research, Baylor University, Waco, Texas, USA
[11]Egerton University, Njoro, Kenya
[12]Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, USA
[13]National Fisheries Resources Research Institute (NaFIRRI), Jinja, Uganda
[14]Cooperative Institute for Great Lakes Research (CIGLR), University of Michigan, Ann Arbor, Michigan, USA
[15]Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada
[16]Natural Products Discovery Core, University of Michigan, Ann Arbor, Michigan, USA
[17]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan, USA
[18]Technical University of Kenya, Nairobi, Kenya
[19]Kisii University, Kisii, Kenya
[20]School of Earth, Environment and Society, Bowling Green State University, Bowling Green, Ohio, USA

PRESENT ADDRESS

Brittany N. Zepernick, Department of Earth, Marine, and Environmental Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
Malcolm A. Barnard, Department of Chemistry, University of Texas at Austin, Austin, Texas, USA

AUTHOR ORCIDs

Lauren N. Hart http://orcid.org/0000-0001-8553-2839
Brittany N. Zepernick http://orcid.org/0000-0003-3121-1225
Katelyn M. Brown http://orcid.org/0000-0002-8581-9490
Malcolm A. Barnard http://orcid.org/0000-0003-4192-574X
Steven W. Wilhelm http://orcid.org/0000-0001-6283-8077
R. Michael McKay http://orcid.org/0000-0003-2723-5371
George S. Bullerjahn http://orcid.org/0000-0001-5319-7896

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

Gregory J. Dick http://orcid.org/0000-0001-7666-6288

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	1953468	R. Michael McKay Kefa M. Otiso George S. Bullerjahn
HHS | NIH | National Institute of Environmental Health Sciences (NIEHS)	1P01ES0289-01	Gregory J. Dick
National Science Foundation (NSF)	1840715	Gregory J. Dick
HHS | NIH | National Institute of Environmental Health Sciences (NIEHS)	1F31ES036421-01	Lauren N. Hart

AUTHOR CONTRIBUTIONS

Lauren N. Hart, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Brittany N. Zepernick, Conceptualization, Data curation, Writing – review and editing | Kaela E. Natwora, Conceptualization, Data curation, Writing – review and editing | Katelyn M. Brown, Conceptualization, Data curation, Writing – review and editing | Julia Akinyi Obuya, Conceptualization, Data curation, Writing – review and editing | Davide Lomeo, Conceptualization, Data curation, Visualization, Writing – review and editing | Malcolm A. Barnard, Data curation, Investigation, Writing – review and editing | Eric O. Okech, Data curation, Writing – review and editing | E. Anders Kiledal, Formal analysis, Methodology, Writing – review and editing | Paul A. Den Uyl, Formal analysis, Methodology, Visualization, Writing – review and editing | Mark Olokotum, Conceptualization, Data curation, Methodology, Writing – review and editing | Steven W. Wilhelm, Data curation, Funding acquisition, Methodology, Resources, Writing – review and editing | R. Michael McKay, Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – review and editing | Ken G. Drouillard, Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review and editing | David H. Sherman, Data curation, Methodology, Supervision, Writing – review and editing | Lewis Sitoki, Conceptualization, Data curation, Methodology, Writing – review and editing | James Achiya, Data curation, Methodology, Resources, Writing – review and editing | Albert Getabu, Data curation, Methodology, Project administration, Resources, Writing – review and editing | Kefa M. Otiso, Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Writing – review and editing | George S. Bullerjahn, Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review and editing | Gregory J. Dick, Conceptualization, Data curation, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

DATA AVAILABILITY

Raw sequences are available on the NCBI SRA under BioProject PRJNA1119566 and BioSamples SAMN41659670- SAMN41659717. All MAGs have been deposited in the NCBI database under BioProject PRJNA1119566 and BioSamples SAMN41711301-SAMN41712291. Physicochemical data are available on the Biological and Chemical Oceanography Data Management Office (BCO-DMO) site (108). All code used in the analysis and visualization of the data presented in this study can be found on GitHub.

ADDITIONAL FILES

The following material is available online.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

Supplemental Material

Fig. S1 (AEM01507-24-s0001.eps). Read mapping (via blastn) from Site 27.
Fig. S2 (AEM01507-24-s0002.eps). Gene cluster families (GCFs) for identified BGCs via antiSMASH and BiG-SCAPE.
Fig. S3 (AEM01507-24-s0003.eps). Genome profile of MAG samp_4317_maxbin_26, annotated as *Microcystis* sp. (NCBI: SAMN41711332).
Fig. S4 (AEM01507-24-s0004.eps). Biosynthetic potential composition in relation to dominant cyanobacterial genera relative abundances in the Winam Gulf.
Fig. S5 (AEM01507-24-s0005.eps). Microbial and archaeal composition makeup in Winam Gulf, LV 2022 metagenome replicates.
Supplemental legends (AEM01507-24-s0006.pdf). Supplemental legends, tables, and figures.
Supplemental tables (AEM01507-24-s0007.xlsx). Tables S1 to S6.

REFERENCES

1. Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM. 2018. *Cyanobacterial blooms*. Nat Rev Microbiol 16:471–483. https://doi.org/10.1038/s41579-018-0040-1
2. Griffith AW, Gobler CJ. 2020. Harmful algal blooms: a climate change co-stressor in marine and freshwater ecosystems. Harmful Algae 91:101590. https://doi.org/10.1016/j.hal.2019.03.008
3. Ho JC, Michalak AM, Pahlevan N. 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. Nature New Biol 574:667–670. https://doi.org/10.1038/s41586-019-1648-7
4. Paerl HW. 2014. Mitigating harmful cyanobacterial blooms in a human- and climatically-impacted world. Life (Basel) 4:988–1012. https://doi.org/10.3390/life4040988
5. Chapra SC, Boehlert B, Fant C, Bierman VJ Jr, Henderson J, Mills D, Mas DML, Rennels L, Jantarasami L, Martinich J, Strzepek KM, Paerl HW. 2017. Climate change impacts on harmful algal blooms in U.S. freshwaters: a screening-level assessment. Environ Sci Technol 51:8933–8943. https://doi.org/10.1021/acs.est.7b01498
6. Igwaran A, Kayode AJ, Moloantoa KM, Khetsha ZP, Unuofin JO. 2024. Cyanobacterial harmful algae blooms: causes, impacts, and risk management. Water Air Soil Pollut 235:71. https://doi.org/10.1007/s11270-023-06782-y
7. Le Manach S, Duval C, Marie A, Djediat C, Catherine A, Edery M, Bernard C, Marie B. 2019. Global metabolomic characterizations of *Microcystis* spp. highlights clonal diversity in natural bloom-forming populations and expands metabolite structural diversity. Front Microbiol 10:791. https://doi.org/10.3389/fmicb.2019.00791
8. Harke MJ, Steffen MM, Gobler CJ, Otten TG, Wilhelm SW, Wood SA, Paerl HW. 2016. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. Harmful Algae 54:4–20. https://doi.org/10.1016/j.hal.2015.12.007
9. Capelli C, Ballot A, Cerasino L, Papini A, Salmaso N. 2017. Biogeography of bloom-forming microcystin producing and non-toxigenic populations of *Dolichospermum lemmermannii* (Cyanobacteria). Harmful Algae 67:1–12. https://doi.org/10.1016/j.hal.2017.05.004
10. Österholm J, Popin RV, Fewer DP, Sivonen K. 2020. Phylogenomic analysis of secondary metabolism in the toxic cyanobacterial genera *Anabaena*, *Dolichospermum* and *Aphanizomenon*. Toxins (Basel) 12:248. https://doi.org/10.3390/toxins12040248
11. Yancey CE, Yu F, Tripathi A, Sherman DH, Dick GJ. 2023. Expression of *Microcystis* biosynthetic gene clusters in natural populations suggests temporally dynamic synthesis of novel and known secondary metabolites in western Lake Erie. Appl Environ Microbiol 89:e0209222. https://doi.org/10.1128/aem.02092-22
12. Paerl HW, Paul VJ. 2012. Climate change: links to global expansion of harmful cyanobacteria. Water Res 46:1349–1363. https://doi.org/10.1016/j.watres.2011.08.002
13. Kaloudis T, Hiskia A, Triantis TM. 2022. Cyanotoxins in bloom: ever-increasing occurrence and global distribution of freshwater cyanotoxins from planktic and benthic cyanobacteria. Toxins (Basel) 14:264. https://doi.org/10.3390/toxins14040264
14. Burford MA, Carey CC, Hamilton DP, Huisman J, Paerl HW, Wood SA, Wulff A. 2020. Perspective: advancing the research agenda for improving understanding of *cyanobacteria* in a future of global change. Harmful Algae 91:101601. https://doi.org/10.1016/j.hal.2019.04.004
15. Svirčev Z, Lalić D, Bojadžija Savić G, Tokodi N, Drobac Backović D, Chen L, Meriluoto J, Codd GA. 2019. Global geographical and historical overview of cyanotoxin distribution and cyanobacterial poisonings. Arch Toxicol 93:2429–2481. https://doi.org/10.1007/s00204-019-02524-4
16. Plisnier P-D, Kayanda R, MacIntyre S, Obiero K, Okello W, Vodacek A, Cocquyt C, Abegaz H, Achieng A, Akonkwa B, et al. 2023. Need for harmonized long-term multi-lake monitoring of African Great Lakes. J Great Lakes Res 49:101988. https://doi.org/10.1016/j.jglr.2022.01.016
17. Sitoki L, Kurmayer R, Rott E. 2012. Spatial variation of *phytoplankton* composition, biovolume, and resulting microcystin concentrations in the Nyanza Gulf (Lake Victoria, Kenya). Hydrobiologia 691:109–122. https://doi.org/10.1007/s10750-012-1062-8
18. Frank TH, Cornelissen IJM, Vijverberg J, Nagelkerke LAJ. 2023. Spatial and seasonal variation in the phytoplankton community of Lake Victoria's Mwanza Gulf, compared to northern parts of the lake. J Great Lakes Res 49:453–462. https://doi.org/10.1016/j.jglr.2023.02.002
19. NjugunaH, RomeroJR, KhisaP. 2006. The effect of turbid inflows into Winam Gulf, Lake Victoria: a 3D modeling study with ELCOM-CAEDYM. Proceedings of the 11th World Lakes Conference. Vol. 2, p 90–94.
20. Calamari D, Akech MO, Ochumba PBO. 1995. Pollution of Winam Gulf, Lake Victoria, Kenya: a case study for preliminary risk assessment. Lakes Reservoirs 1:89–106. https://doi.org/10.1111/j.1440-1770.1995.tb00010.x
21. Roegner A, Sitoki L, Weirich C, Corman J, Owage D, Umami M, Odada E, Miruka J, Ogari Z, Smith W, Rejmankova E, Miller TR. 2020. Harmful algal blooms threaten the health of peri-urban fisher communities: a case study in Kisumu Bay, Lake Victoria, Kenya. Expo Health 12:835–848. https://doi.org/10.1007/s12403-019-00342-8
22. Codd GA, Morrison LF, Metcalf JS. 2005. *Cyanobacterial* toxins: risk management for health protection. Toxicol Appl Pharmacol 203:264–272. https://doi.org/10.1016/j.taap.2004.02.016
23. Jegede FE, Oyeyi TI, Abdulrahman SA, Mbah HA, Badru T, Agbakwuru C, Adedokun O. 2017. Effect of HIV and malaria parasites co-infection on immune-hematological profiles among patients attending anti-retroviral treatment (ART) clinic in infectious disease hospital Kano, Nigeria. PLoS One 12:e0174233. https://doi.org/10.1371/journal.pone.0174233
24. Oyieko J, Copeland NK, Otieno S, Kifude C, Ocholla S, Hutter J, Smith H, Roberds A, Luckhart S, Stewart VA. 2023. Longitudinal and cross-sectional analyses of asymptomatic HIV-1/malaria co-infection in Kisumu County, Kenya. Am J Trop Med Hyg 108:85–92. https://doi.org/10.4269/ajtmh.22-0035
25. Marriott AL, Osano OF, Coffey TJ, Humphrey OS, Ongore CO, Watts MJ, Aura CM. 2023. Considerations for environmental biogeochemistry and food security for aquaculture around Lake Victoria, Kenya. Environ Geochem Health 45:6137–6162. https://doi.org/10.1007/s10653-023-01585-w
26. Gikuma-Njuru P, Hecky RE. 2005. Nutrient concentrations in Nyanza Gulf, Lake Victoria, Kenya: light limits algal demand and abundance.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

Hydrobiologia 534:131–140. https://doi.org/10.1007/s10750-004-1418-9
27. Misigo AWS, Suzuki S. 2018. Spatial-temporal sediment hydrodynamics and nutrient loads in Nyanza Gulf, characterizing variation in water quality. WJET 06:98–115. https://doi.org/10.4236/wjet.2018.62B009
28. Brown KM, Barker KB, Wagner RS, Ward CS, Sitoki L, Njiru J, Omondi R, Achiya J, Getabu A, McKay RM, Bullerjahn GS, the NSF - IRES Lake Victoria Research Consortium. 2024. Bacterial community and cyanotoxin gene distribution of the Winam Gulf, Lake Victoria, Kenya. Environ Microbiol Rep 16:e13297. https://doi.org/10.1111/1758-2229.13297
29. Suba VO, Khan MS, Miruka J, Mwalati L, Njoroge B, Oluoch E, Nyongesa J, Otieno D, Li Y. 2025. Seasonal phytoplankton ecosystem dynamics in response to environmental variables in Winam Gulf, Lake Victoria, Kenya. Aquat Sci 87:4. https://doi.org/10.1007/s00027-024-01130-6
30. Driscoll CB, Meyer KA, Šulčius S, Brown NM, Dick GJ, Cao H, Gasiūnas G, Timinskas A, Yin Y, Landry ZC, Otten TG, Davis TW, Watson SB, Dreher TW. 2018. A closely-related clade of globally distributed bloom-forming cyanobacteria within the *Nostocales*. Harmful Algae 77:93–107. https://doi.org/10.1016/j.hal.2018.05.009
31. Hounmanou YMG, Baniga Z, García V, Dalsgaard A. 2022. *Salmonella salamae* and *S. waycross* isolated from Nile perch in Lake Victoria show limited human pathogenic potential. Sci Rep 12:4229. https://doi.org/10.1038/s41598-022-08200-5
32. Baniga Z, Hounmanou YMG, Kudirkiene E, Kusiluka LJM, Mdegela RH, Dalsgaard A. 2020. Genome-based analysis of extended-spectrum β-lactamase-producing *Escherichia coli* in the aquatic environment and Nile perch (*Lates niloticus*) of Lake Victoria, Tanzania. Front Microbiol 11:108. https://doi.org/10.3389/fmicb.2020.00108
33. Hounmanou YMG, Leekitcharoenphon P, Hendriksen RS, Dougnon TV, Mdegela RH, Olsen JE, Dalsgaard A. 2019. Surveillance and genomics of toxigenic *Vibrio cholerae* O1 from fish, phytoplankton, and water in Lake Victoria, Tanzania. Front Microbiol 10:901. https://doi.org/10.3389/fmicb.2019.00901
34. Guildford SJ, Hecky RE. 2000. Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: is there a common relationship? Limnology & Oceanography 45:1213–1223. https://doi.org/10.4319/lo.2000.45.6.1213
35. Yanez-Montalvo A, Aguila B, Gómez-Acata ES, Guerrero-Jacinto M, Oseguera LA, Falcón LI, Alcocer J. 2022. Shifts in water column microbial composition associated to lakes with different trophic conditions: "Lagunas de Montebello" National Park, Chiapas, México. PeerJ 10:e13999. https://doi.org/10.7717/peerj.13999
36. Pérez-Carrascal OM, Terrat Y, Giani A, Fortin N, Greer CW, Tromas N, Shapiro BJ. 2019. Coherence of *Microcystis* species revealed through population genomics. ISME J 13:2887–2900. https://doi.org/10.1038/s41396-019-0481-1
37. Yancey CE, Kiledal EA, Chaganti SR, Denef VJ, Errera RM, Evans JT, Hart LN, Isailovic D, James WS, Kharbush JJ, Kimbrel JA, Li W, Mayali X, Nitschky H, Polik CA, Powers MA, Premathilaka SH, Rappuhn NA, Reitz LA, Rivera SR, Zwiers CC, Dick GJ. 2023. The western Lake Erie culture collection: a promising resource for evaluating the physiological and genetic diversity of *Microcystis* and its associated microbiome. Harmful Algae 126:102440. https://doi.org/10.1016/j.hal.2023.102440
38. Den Uyl P. 2024. Genetic characterization of saxitoxin-producing cyanobacteria associated with western Lake Erie harmful algal blooms. University of Michigan.
39. Martins J, Leão PN, Ramos V, Vasconcelos V. 2013. N-terminal protease gene phylogeny reveals the potential for novel *cyanobactin* diversity in *cyanobacteria*. Mar Drugs 11:4902–4916. https://doi.org/10.3390/md11124902
40. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. https://doi.org/10.1093/nar/gkab776
41. Pancrace C, Barny M-A, Ueoka R, Calteau A, Scalvenzi T, Pédron J, Barbe V, Piel J, Humbert J-F, Gugger M. 2017. Insights into the *Planktothrix* genus: genomic and metabolic comparison of benthic and planktic strains. Sci Rep 7:41181. https://doi.org/10.1038/srep41181
42. Mise K, Masuda Y, Senoo K, Itoh H. 2021. Undervalued pseudo-*nifH* sequences in public databases distort metagenomic insights into biological nitrogen fixers. mSphere 6:e0078521. https://doi.org/10.1128/msphere.00785-21
43. Bellanger M, Figueroa JL, Tiemann L, Friesen ML, White RA. 2024. NFixDB (Nitrogen fixation DataBase) - a comprehensive integrated database for robust 'omics analysis of diazotrophs. bioRxiv. https://doi.org/10.1101/2024.03.04.583350
44. Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, Lee S, Meijer D, Recchia MJJ, Reitz ZL, et al. 2023. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res 51:D603–D610. https://doi.org/10.1093/nar/gkac1049
45. Mihali TK, Kellmann R, Muenchhoff J, Barrow KD, Neilan BA. 2008. Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. Appl Environ Microbiol 74:716–722. https://doi.org/10.1128/AEM.01988-07
46. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 51:W46–W50. https://doi.org/10.1093/nar/gkad344
47. Helfrich EJN, Lin G-M, Voigt CA, Clardy J. 2019. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. Beilstein J Org Chem 15:2889–2906. https://doi.org/10.3762/bjoc.15.283
48. Rouhiainen L, Jokela J, Fewer DP, Urmann M, Sivonen K. 2010. Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). Chem Biol 17:265–273. https://doi.org/10.1016/j.chembiol.2010.01.017
49. Ishida K, Christiansen G, Yoshida WY, Kurmayer R, Welker M, Valls N, Bonjoch J, Hertweck C, Börner T, Hemscheidt T, Dittmann E. 2007. Biosynthesis and structure of aeruginoside *126A* and *126B*, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. Chem Biol 14:565–576. https://doi.org/10.1016/j.chembiol.2007.04.006
50. Lung'Ayia HBO, M'Harzi A, Tackx M, Gichuki J, Symoens JJ. 2000. Phytoplankton community structure and environment in the Kenyan waters of Lake Victoria. Freshw Biol 43:529–543. https://doi.org/10.1046/j.1365-2427.2000.t01-1-00525.x
51. Obuya JA, Onyango HO, Olokotum M, Zepernick B, Natwora K, Otieno D, Owoko W, Achieng D, Otuo PW, Keyombe JL. 2024. Socioeconomic consequences of *cyanobacterial* harmful algal blooms in small-scale fishing communities of Winam Gulf, Lake Victoria. J Great Lakes Res 50:102236. https://doi.org/10.1016/j.jglr.2023.09.007
52. Aura CM, Odoli C, Nyamweya CS, Njiru JM, Musa S, Miruka JB, Owili MO, Omondi R, Raburu P, Manyala J, Mwamburi J, Ogari Z, Mbaru EK. 2020. Application of phytoplankton community structure for ranking the major riverine catchments influencing the pollution status of a lake basin. Lakes Reservoirs 25:3–17. https://doi.org/10.1111/lre.12307
53. Wanjala WSN, Olutende OM, Rose W, Joab O, Omuterema OS, Steve O. 2024. Rivers of resilience: assessing hydrological baseline data for nature-based conservation in the Yala River watershed. Open Access Libr J 11:1–22. https://doi.org/10.4236/oalib.1111206
54. Bauersachs T, Compaoré J, Hopmans EC, Stal LJ, Schouten S, Sinninghe Damsté JS. 2009. Distribution of heterocyst glycolipids in cyanobacteria. Phytochemistry 70:2034–2039. https://doi.org/10.1016/j.phytochem.2009.08.014
55. Miruka JB, Getabu A, Sitoki L, James O, Mwamburi J, George O, Chrisphine N, Odoli C. 2021. Water quality, phytoplankton composition and microcystin concentrations in Kisumu Bay (Kenya) of Lake Victoria after a prolonged water hyacinth infestation period. Lakes Reservoirs 26:e12380. https://doi.org/10.1111/lre.12380
56. Hancock TL, Dahedl EK, Kratz MA, Urakawa H. 2024. The synchronicity of bloom-forming cyanobacteria transcription patterns and hydrogen peroxide dynamics. Environ Pollut 348:123812. https://doi.org/10.1016/j.envpol.2024.123812
57. Paerl HW. 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnol Oceangr 33:823–843. https://doi.org/10.4319/lo.1988.33.4_part_2.0823
58. Hampel JJ, McCarthy MJ, Neudeck M, Bullerjahn GS, McKay RML, Newell SE. 2019. Ammonium recycling supports toxic *Planktothrix* blooms in Sandusky Bay, Lake Erie: evidence from stable isotope and metatranscriptome data. Harmful Algae 81:42–52. https://doi.org/10.1016/j.hal.2018.11.011
59. Handbook of cyanobacterial monitoring and cyanotoxin analysis. 2016. 1st ed. Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119068761.

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

60. Cabello-Yeves PJ, Picazo A, Camacho A, Callieri C, Rosselli R, Roda-Garcia JJ, Coutinho FH, Rodriguez-Valera F. 2018. Ecological and genomic features of two widespread freshwater picocyanobacteria. Environ Microbiol 20:3757–3771. https://doi.org/10.1111/1462-2920.14377
61. Burson A, Stomp M, Mekkes L, Huisman J. 2019. Stable coexistence of equivalent nutrient competitors through niche differentiation in the light spectrum. Ecology 100:e02873. https://doi.org/10.1002/ecy.2873
62. Zepernick BN, McKay RML, Martin RM, Bullerjahn GS, Paerl HW, Wilhelm SW. 2024. A tale of two blooms: do ecological paradigms for algal bloom success and succession require revisiting? J Great Lakes Res 50:102336. https://doi.org/10.1016/j.jglr.2024.102336
63. McKindles KM, McKay RM, Bullerjahn GS. 2022. Genomic comparison of *Planktothrix agardhii* isolates from a Lake Erie embayment. PLoS One 17:e0273454. https://doi.org/10.1371/journal.pone.0273454
64. Krausfeldt LE, Shmakova E, Lee HW, Mazzei V, Loftin KA, Smith RP, Karwacki E, Fortman PE, Rosen BH, Urakawa H, Dadlani M, Colwell RR, Lopez JV. 2024. Microbial diversity, genomics, and phage–host interactions of cyanobacterial harmful algal blooms. mSystems 9:e0070923. https://doi.org/10.1128/msystems.00709-23
65. Huo D, Gan N, Geng R, Cao Q, Song L, Yu G, Li R. 2021. Cyanobacterial blooms in China: diversity, distribution, and cyanotoxins. Harmful Algae 109:102106. https://doi.org/10.1016/j.hal.2021.102106
66. Zhang J-Y, Guan R, Zhang H-J, Li H, Xiao P, Yu G-L, Du L, Cao D-M, Zhu B-C, Li R-H, Lu Z-H. 2016. Complete genome sequence and genomic characterization of *Microcystis panniformis* FACHB 1757 by third-generation sequencing. Stand Genomic Sci 11:11. https://doi.org/10.1186/s40793-016-0130-5
67. Brown KM, Ward CS, Bullerjahn GS. 2024. Metagenome-assembled genome sequences of two cyanobacterial cultures from Homa Bay County, Kenya. Microbiol Resour Announc 13. https://doi.org/10.1128/mra.01205-23
68. Dziallas C, Grossart H-P. 2011. Increasing oxygen radicals and water temperature select for toxic *Microcystis* sp. PLoS One 6:e25569. https://doi.org/10.1371/journal.pone.0025569
69. Wei N, Hu C, Dittmann E, Song L, Gan N. 2024. The biological functions of microcystins. Water Res 262:122119. https://doi.org/10.1016/j.watres.2024.122119
70. Harke MJ, Davis TW, Watson SB, Gobler CJ. 2016. Nutrient-controlled niche differentiation of western Lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. Environ Sci Technol 50:604–615. https://doi.org/10.1021/acs.est.5b03931
71. Zepernick BN, Hart LN, Natwora KE, Brown KM, Obuya JA, Olokotum M, Owino E, Keating N, Lomeo D, Dick GJ, Wilhelm SW, McKay RM, Bullerjahn GS, NSF-IRES 2022-2023 Lake Victoria Research Consortium. 2024. Metagenomic sequencing of cyanobacterial dominated Lake Victoria - an African Great Lake. Microb Resour Announc 13:e00798-24. https://doi.org/10.1128/mra.00798-24
72. Phosphate-buffered saline (PBS). 2006. 2006:Pdb.Rec8247. Cold Spring Harb Protoc.
73. Zepernick BN, Truchon AR, Gann ER, Wilhelm SW. 2022. Draft genome sequence of the freshwater diatom Fragilaria crotonensis SAG 28.96. Microbiol Resour Announc 11:e0028922. https://doi.org/10.1128/mra.00289-22
74. Taylor JM, Andersen IM, Hoke AK, Kelly PT, Scott JT. 2023. In-situ N2:Ar ratios describe the balance between nitrogen fixation and denitrification in shallow eutrophic experimental lakes. Biogeochemistry 166:283–301. https://doi.org/10.1007/s10533-023-01063-6
75. Bridgewater L, Association APH, Association AWW, Federation WE. 2012. Standard methods for the examination of water and wastewater. American Public Health Association.
76. Welschmeyer NA. 1994. Fluorometric analysis of *chlorophyll a* in the presence of *chlorophyll b* and pheopigments. Limnol Oceanogr 39:1985–1992. https://doi.org/10.4319/lo.1994.39.8.1985
77. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560
78. Bushnell B. 2015. BBMap. Joint Genome Institute.
79. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033
80. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103
81. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2:26. https://doi.org/10.1186/2049-2618-2-26
82. Kang DD, 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
83. 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 10:46. https://doi.org/10.1186/s40168-022-01237-8
84. Pan S, Zhu C, Zhao X-M, Coelho LP. 2022. A deep siamese neural network improves metagenome-assembled genomes in microbiome datasets across different environments. Nat Commun 13:2326. https://doi.org/10.1038/s41467-022-29843-y
85. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, Jensen LJ, Nielsen HB, Petersen TN, Winther O, Rasmussen S. 2021. Improved metagenome binning and assembly using deep variational autoencoders. Nat Biotechnol 39:555–560. https://doi.org/10.1038/s41587-020-00777-4
86. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1
87. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126
88. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
89. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. *Bakta*: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom 7:000685. https://doi.org/10.1099/mgen.0.000685
90. Woodcroft BJ. 2024.CoverM: read coverage calculator for metagenomics.
91. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
92. Dixon P. 2003. *VEGAN*, a package of R functions for community ecology. J Vegetation Science 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x
93. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res 50:D20–D26. https://doi.org/10.1093/nar/gkab1112
94. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35:4162–4164. https://doi.org/10.1093/bioinformatics/btz188
95. Yancey CE, Mathiesen O, Dick GJ. 2023. Transcriptionally active nitrogen fixation and biosynthesis of diverse secondary metabolites by *Dolichospermum* and *Aphanizomenon*-like cyanobacteria in western Lake Erie mcrocystis blooms. Harmful Algae 124:102408. https://doi.org/10.1016/j.hal.2023.102408
96. Dreher TW, Davis EW, Mueller RS, Otten TG. 2021. Comparative genomics of the *ADA* clade within the *Nostocales*. Harmful Algae 104:102037. https://doi.org/10.1016/j.hal.2021.102037
97. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
98. Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. 2012. Distribution of Nitrogen fixation and Nitrogenase-like sequences amongst microbial genomes. 1. BMC Genomics 13:1–12. https://doi.org/10.1186/1471-2164-13-162
99. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.

switching during cell differentiation. Nat Biotechnol 28:511–515. https://doi.org/10.1038/nbt.1621
100. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421
101. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3–6. https://doi.org/10.1038/s41564-020-00834-3
102. Gilchrist CLM, Chooi Y-H. 2021. *Clinker* & *clustermap.js*: automatic generation of gene cluster comparison figures. Bioinformatics 37:2473–2475. https://doi.org/10.1093/bioinformatics/btab007
103. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68. https://doi.org/10.1038/s41589-019-0400-9
104. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303
105. Team RC. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
106. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. JOSS 4:1686. https://doi.org/10.21105/joss.01686
107. Neuwirth E. 2014. RColorBrewer: ColorBrewer palettes. https://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf.
108. Bullerjahn GS, Zepernick BN, Hart LN. 2024. Metagenomic time series of Winam Gulf, Lake Victoria from 2022-2023 (ASI Lake Victoria project). biological and chemical oceanography data management office (BCO-DMO).

Downloaded from https://journals.asm.org/journal/aem on 14 April 2026 by 41.75.185.246.