



Genome Streamlining, Plasticity, and Metabolic Versatility Distinguish Co-occurring Toxic and Nontoxic Cyanobacterial Strains of *Microcoleus*

 Hwee Sze Tee,^a Susanna A. Wood,^b Keith Bouma-Gregson,^c  Gavin Lear,^a  Kim M. Handley^a

^aSchool of Biological Sciences, The University of Auckland, Auckland, New Zealand

^bCawthron Institute, Nelson, New Zealand

^cU.S. Geological Survey, California Water Science Center, Sacramento, California, USA

ABSTRACT Harmful cyanobacterial bloom occurrences have increased worldwide due to climate change and eutrophication, causing nuisance and animal deaths. Species from the benthic cyanobacterial genus *Microcoleus* are ubiquitous and form thick mats in freshwater systems, such as rivers, that are sometimes toxic due to the production of potent neurotoxins (anatoxins). Anatoxin-producing (toxic) strains typically coexist with non-anatoxin-producing (nontoxic) strains in mats, although the reason for this is unclear. To determine the genetic mechanisms differentiating toxic and nontoxic *Microcoleus*, we sequenced and assembled genomes from 11 cultures and compared these to another 31 *Microcoleus* genomes. Average nucleotide identities (ANI) indicate that toxic and nontoxic strains are distinct species (ANI, <95%), and only 6% of genes are shared across all 42 genomes, suggesting a high level of genetic divergence among *Microcoleus* strains. Comparative genomics showed substantial genome streamlining in toxic strains and a potential dependency on external sources for thiamine and sucrose. Toxic and nontoxic strains are further differentiated by an additional set of putative nitrate transporter (nitrogen uptake) and cyanophycin (carbon and nitrogen storage) genes, respectively. These genes likely confer distinct competitive advantages based on nutrient availability and suggest nontoxic strains are more robust to nutrient fluctuations. Nontoxic strains also possess twice as many transposable elements, potentially facilitating greater genetic adaptation to environmental changes. Our results offer insights into the divergent evolution of *Microcoleus* strains and the potential for cooperative and competitive interactions that contribute to the co-occurrence of toxic and nontoxic species within mats.

IMPORTANCE *Microcoleus autumnalis*, and closely related *Microcoleus* species, compose a geographically widespread group of freshwater benthic cyanobacteria. Canine deaths due to anatoxin-a poisoning, following exposure to toxic proliferations, have been reported globally. While *Microcoleus* proliferations are on the rise, the mechanisms underpinning competition between, or coexistence of, toxic and nontoxic strains are unknown. This study identifies substantial genetic differences between anatoxin-producing and non-anatoxin-producing strains, pointing to reduced metabolic flexibility in toxic strains, and potential dependence on cohabiting nontoxic strains. Results provide insights into the metabolic and evolutionary differences between toxic and nontoxic *Microcoleus*, which may assist in predicting and managing aquatic proliferations.

KEYWORDS *Phormidium autumnale*, anatoxins, benthic cyanobacteria, comparative genomics, cyanobacterial proliferation, resource trade-off, toxic and nontoxic cyanobacteria

Citation Tee HS, Wood SA, Bouma-Gregson K, Lear G, Handley KM. 2021. Genome streamlining, plasticity, and metabolic versatility distinguish co-occurring toxic and nontoxic cyanobacterial strains of *Microcoleus*. *mBio* 12:e02235-21. <https://doi.org/10.1128/mBio.02235-21>.

Editor Mark J. Bailey, CEH-Oxford

Copyright © 2021 Tee et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Kim M. Handley, kim.handley@auckland.ac.nz.

Received 27 July 2021

Accepted 20 September 2021

Published 26 October 2021

Cyanobacteria first appeared on Earth over two billion years ago and are credited with the evolution of aerobic life (1). Despite this contribution, the increased occurrence of cyanobacterial planktonic blooms and benthic proliferations, due to anthropogenic activity, has severely disrupted aquatic habitats and deteriorated water quality (2). Cyanobacterial blooms and proliferations are often composed of both toxic and nontoxic strains, with various amounts of toxin detected spatially and temporally (3–7). However, the genomic differences that lead to the dominance of either toxic or nontoxic strains is unknown.

Here, we focus on the benthic cyanobacterial genus *Microcoleus*, which inhabits freshwater systems worldwide (8–10). Under favorable environmental and hydrological conditions, *Microcoleus* forms cohesive mats, which can cover large areas in lakes and rivers (11). Some *Microcoleus* species produce a neurotoxin (anatoxin-a) that has been linked to animal deaths in many countries, including the United States, the Netherlands, New Zealand, Germany, and France (12), and can accumulate in aquatic organisms (13). Both anatoxin-producing and non-anatoxin-producing *Microcoleus* strains (henceforth referred to as toxic and nontoxic strains, respectively) are often found co-occurring within a single mat (e.g., coating a riverbed cobble), where their relative abundances determine the toxin concentration in the mat (5, 11, 14, 15). Although benthic proliferations of only nontoxic strains have been documented (11), proliferations containing toxic strains are always reported as a mixture of toxic and nontoxic strains (16).

Unlike planktonic cyanobacteria, which often bloom at high nutrient concentrations (17, 18), *Microcoleus* proliferates under low-phosphate and slightly elevated nitrogen conditions by adopting diverse strategies to scavenge phosphorus and nitrogen (8, 19). Toxic *Microcoleus* strains have been reported to occur at higher nitrogen levels than nontoxic strains (8), although cellular anatoxin concentrations have been experimentally shown to be lowest under high-nitrogen and high-phosphorus conditions (14), as reported for several other cyanobacteria and cyanotoxins (20). In contrast, nontoxic *Microcoleus* strains are reported to have higher cell concentrations and higher maximum growth rates than toxic strains, regardless of nutrient concentrations (phosphorus and nitrogen), suggesting higher energy costs associated with toxin production (14).

Comparative genomics has highlighted marked genotypic and phenotypic plasticity within cyanobacterial species, such as planktonic *Raphidiopsis raciborskii* and *Microcystis aeruginosa*, and benthic and planktonic *Planktothrix* spp. (6, 21–25). Striking differences in genes found between toxic and nontoxic *R. raciborskii* strains suggest that toxin production may be associated with stress response (26). However, pangenomic comparisons differentiating other toxic and nontoxic cyanobacterial species are in general lacking, and none exist for benthic species or anatoxin producers (27, 28). As such, the genetic mechanisms that regulate the ecological successes of toxic and nontoxic benthic cyanobacteria, such as *Microcoleus*, remain elusive.

We investigated the genetic differences between toxic and nontoxic *Microcoleus* strains to determine the relationship between these often cohabiting and closely related benthic cyanobacterial species. For this, we compared 42 *Microcoleus* metagenome-assembled genomes (MAGs), including 30 nontoxic and 12 toxic strains, sourced from enrichment cultures or *Microcoleus*-dominated mats from different rivers in New Zealand and the United States. We hypothesized that there would be marked differences in the genomes of toxic and nontoxic strains, linked to their different growth rates associated with N availability. Our results help to explain why toxic strains never proliferate in the absence of nontoxic strains and the association between toxic strains and higher N availability.

RESULTS AND DISCUSSION

Genome reduction among toxic *Microcoleus* strains. To compare the genomes of toxic and nontoxic *Microcoleus*, we first sequenced and assembled the genomes (metagenome-assembled genomes, MAGs) of 11 nonaxenic isolates (3 toxic and 8 nontoxic)

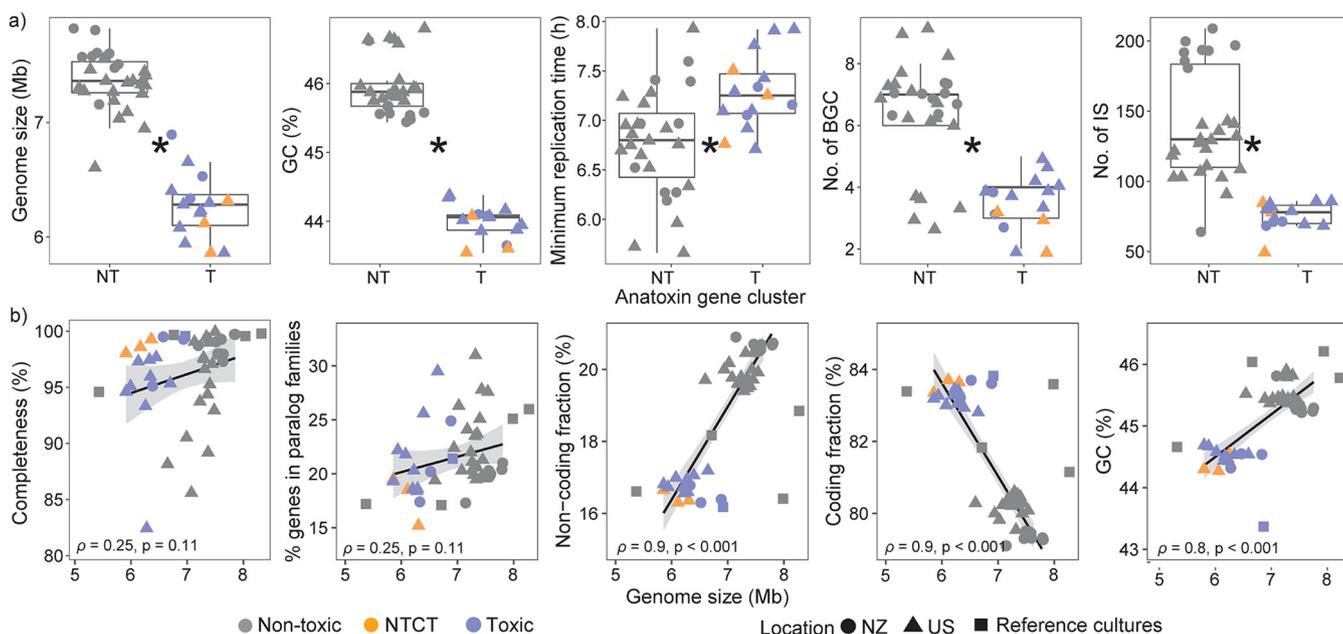


FIG 1 Genomic differences between nontoxic and toxic *Microcoleus* strains. The symbol color indicates genomes that are nontoxic (NT; gray), nontoxic clustered with toxic (NTCT; orange), and toxic (T; blue), and the shape represents the reference culture (square) or *Microcoleus* strains that originated from New Zealand (NZ; circle) and the United States (US). (a) Boxplots of nontoxic and toxic *Microcoleus* genomes showing estimated genome sizes, number of biosynthetic gene clusters (BGC), number of insertion sequences (IS), percent GC content, and predicted minimum replication time. Boxes represent the interquartile range (IQR) between the first and third quartiles, and the horizontal line inside the box represents the median. Whiskers represent the lowest and highest values within 1.5 times the IQR from the first and third quartiles, respectively. The asterisk (*) indicates a significant difference between nontoxic and toxic *Microcoleus* genomes. (b) Correlations between genome size and the coding/noncoding fraction, GC content, percentage of genes in paralog families, and genome completeness. Regression lines, Pearson correlation coefficients (ρ), and the associated P values are shown in the plots.

related to *Microcoleus autumnalis* (previously *Phormidium autumnale*) (29). Production of anatoxin-a, homoanatoxin-a, or chemical variants of these was confirmed for the 3 isolates with anatoxin gene clusters by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) (Table S1). We then compared these to a further 31 MAGs previously obtained from nonaxenic isolates (3 toxic) and mats sampled from rivers (6 toxic, 22 nontoxic) (8, 19, 30), where the 6 mat-derived MAGs with anatoxin gene clusters were spatially associated with anatoxins (8, 19). The 42 *M. autumnalis*-like genomes were classified based on the presence/absence of the anatoxin gene cluster. Each had an estimated completeness of $>80\%$ (contamination scores of $0.7 \pm 1.3\%$; Table S1).

Differences in bacterial genome sizes typically reflect gain or loss of function and adaptation to more defined niches in smaller streamlined genomes (31, 32). Results show the toxic *Microcoleus* have significantly smaller estimated genome sizes ($\sim 6.3 \pm 0.3$ Mbp) and lower GC contents ($\sim 44.5 \pm 0.1\%$) than the nontoxic ones, which have estimated genome sizes of $\sim 7.4 \pm 0.3$ Mbp and $\sim 45.5 \pm 0.2\%$ GC content (Fig. 1a). This excludes three nontoxic *Microcoleus*, which have genome sizes and GC contents comparable to those of toxic *Microcoleus* and features suggestive of recent toxic gene cluster loss (discussed below). They were therefore analyzed separately (and are defined here as nontoxic clustered with toxic strains, NTCT). As most genomes of smaller toxic strains were $>90\%$ complete, their smaller size is not expected to reflect assembly or binning errors (Fig. 1b). Confirming this, we compared the genome size and assembly completeness of all toxic and nontoxic strains and observed a poor correlation ($\rho = 0.25$).

While the chemistry and biosynthesis of cyanotoxins are highly diverse, and they may function differently (20, 33), experimental and field-based evidence indicates that, like *Microcoleus*, other toxic and nontoxic cyanobacteria also differ in their requirements for nutrients (8, 34–37). We therefore sought to determine whether the observed difference in genome size in *Microcoleus* is also found in other cyanobacterial taxa. While we

observed significantly different genome sizes between toxic and nontoxic strains of *Microcystis aeruginosa* ($n = 25$, P value < 0.05), no correlation was found between genome size and toxin production of other toxin-producing cyanobacteria (i.e., *Anabaena*, *Dolichospermum*, and *Raphidiopsis raciborskii*; P value > 0.05) (Table S2). This implies that the genomic structures of *Microcoleus* spp. and *Microcystis aeruginosa* are highly distinct from other cyanobacteria. The limited number of genomes available for the other cyanobacteria ($n < 10$) may also lead to nonsignificant results, suggesting more comprehensive sequencing is needed to analyze their genome size distributions.

Significant differences in genome sizes between toxic and nontoxic *Microcoleus* strains may contribute to variations in genome plasticity and ecological adaptation (31). No association was found between genome size and the fraction of paralogous genes (Fig. 1b), a feature of genome streamlining (38). However, the non-protein-coding fractions were positively correlated with genome size and lower in toxic strains ($16.8 \pm 0.3\%$) than in nontoxic strains ($20.1 \pm 0.5\%$). This suggests that toxic strains may have undergone genome streamlining, retaining fewer genes involved in transcriptional and translational regulation (32). Some bacterial lineages streamline their genomes to select against cell complexity and reduce replication cost (31, 32). Previous research has shown that anatoxin-producing strains have lower cell concentrations and growth rates under culture conditions (14). Our results illustrate that toxic *Microcoleus* strains have significantly longer predicted minimum replication times of $\sim 7.3 \pm 0.4$ h, compared to nontoxic strains ($\sim 6.8 \pm 0.5$ h; Fig. 1a), indicating a potential trade-off between toxin production and growth. In contrast, because nontoxic strains have larger genome sizes, they may harbor enhanced metabolic capabilities that promote proliferation.

Phylogenetic and pangenome analyses demonstrate toxic and nontoxic *Microcoleus* strains comprise distinct groups of species. A phylogenetic tree of cyanobacterial 16S rRNA genes showed that all *Microcoleus* strains were placed within a highly supported cluster (98% bootstrap value) with *Tychonema* CCAP, *Microcoleus vaginatus*, *Phormidium nigroviride*, and *Phormidium autumnale* (Fig. S1), in line with previous findings (8, 29). As 16S rRNA analysis can typically only resolve to the genus level (39, 40), we generated maximum-likelihood phylogenies from alignments of 120 core single-copy marker genes and 525 core single-copy orthologs present in all genomes (Fig. 2), as well as an alignment of 12,406 single-nucleotide polymorphisms (Fig. S2). All trees revealed two distinct *Microcoleus* clades, comprising either toxic (and NTCT) or nontoxic species (clades 1 and 2, respectively) (Fig. 2). Both toxic and nontoxic strains were further divided broadly based on geography. Average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) analyses (Fig. S3) predict that the 42 *Microcoleus* genomes derived from 9 species (thresholds for ANI, 96.5%; for dDDH, 70%) (41, 42). These comprise three closely related anatoxin-producing species (clusters 4, 5, and 7), two NTCT species (clusters 6 and 8) (Fig. S3), and four non-anatoxin-producing species.

Microcoleus strains vary in their cell and filament dimensions, coloration, and morphology of their apical cells (i.e., obtusely rounded or pointed or with/without calyptra); there is no consistency with these morphological features and the phylogeny presented (3, 5). Based on our observations, we also note that these features can vary over time in culture and that strains in culture often do not have the exact same morphological features that were present when they were in the environment. Therefore, it is crucial to incorporate molecular characterization for species/strain identification.

Pangenome analysis identified a total of 17,858 gene clusters (Markov cluster algorithm [MCL] inflation = 1.5) among 42 *Microcoleus* strains and 5 cyanobacterial reference genomes. Only 6% of gene clusters were present in all genomes, and 18% were unique, suggesting huge genetic variability among *Microcoleus* strains (Fig. 3 and Table S3). Specialist cyanobacterial *Raphidiopsis* spp. share a highly conserved core genome (2,125 out of 4,715 orthologous gene clusters), while generalist *Microcystis* spp. exhibit considerable genetic divergence (413/13,884 orthologous gene clusters) (28). This suggests *Microcoleus* spp., which have relatively large genomes and high genomic diversity, may adopt a generalist approach to adapt to broad ranges of environments

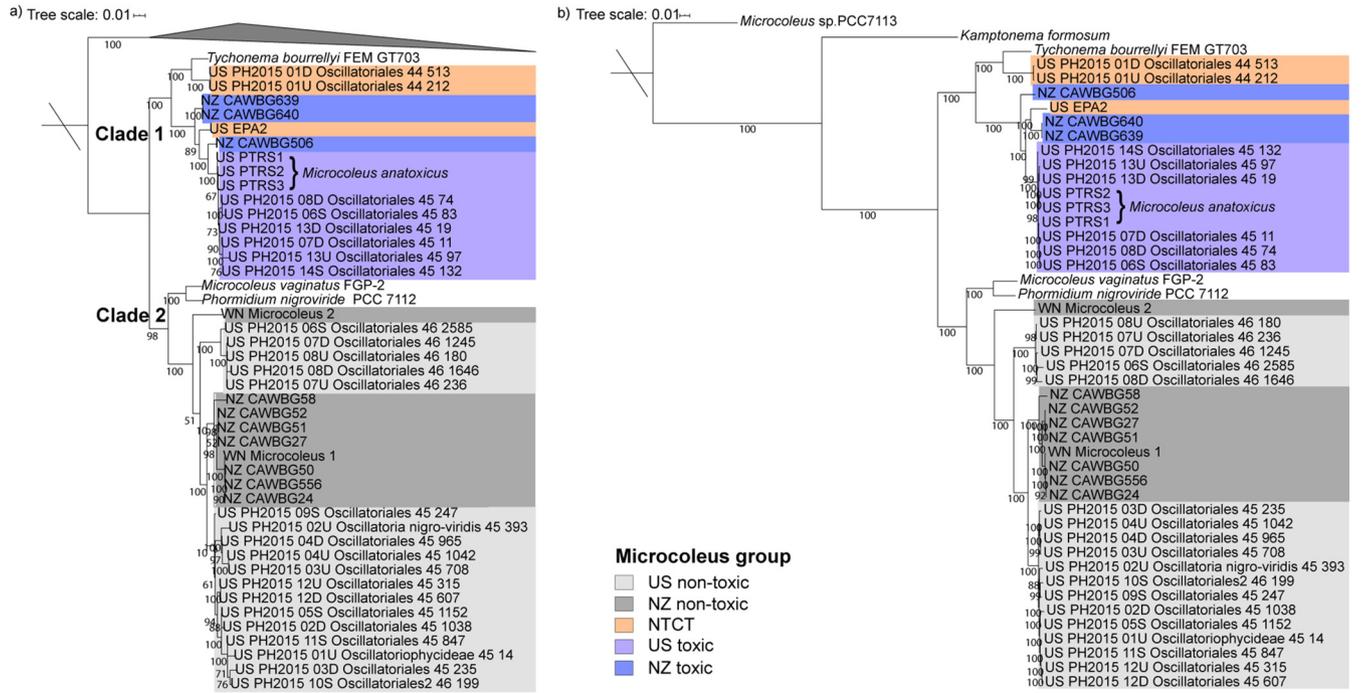


FIG 2 Maximum likelihood trees of the single-copy genes present in *Microcoleus* spp. (a) Phylogenetic tree of concatenated alignments of 120 core single-copy marker genes. The tree is rooted with 12 other cyanobacterial references as shown in Table S1. (b) Species tree based on 525 core single-copy orthologs present in all *Microcoleus* genomes. The tree is rooted with *Microcoleus* sp. strain PCC7113. For both trees, scale bars represent the number of substitutions per site. Bootstrap values over 50% are shown. Color demarks nontoxic (non-anatoxin-producing), and toxic (anatoxin-producing) *Microcoleus* groups from the United States (US) and New Zealand (NZ). US PTRS1, 2, and 3 are known as *Microcoleus anatoxicus* in Conklin et al. (30). *Oscillatoria nigroviridis* (in NCBI, UniProt) is known as *Phormidium nigroviride* in AlgaeBase (<https://www.algaebase.org/>) and *Microcoleus* sp. in Genome Taxonomy Database (<https://gtdb.ecogenomic.org/>). NTCT, nontoxic clustered with toxic.

(43). A total of 2,140 and 1,722 orthogroups/gene clusters were significantly more prevalent in nontoxic and toxic strains, respectively (Benjamini-Hochberg adjusted *P* value < 0.05; Fig. 3 and Table S4), highlighting large differences in gene content and resource allocation between toxic and nontoxic strains that may contribute to distinct physiological responses toward environmental disturbances.

Evidence of anatoxin gene cluster loss and potential to disrupt the DNA phosphorothioation stress response mechanism. To investigate the genomic structure of the anatoxin gene cluster in *Microcoleus* strains, we aligned the clusters, along with neighboring genes. The *anal*, *anaJ*, and *anaA* genes in all *Microcoleus* strains were rearranged and located downstream of *anaG*, compared to *Kamptonema formosum*

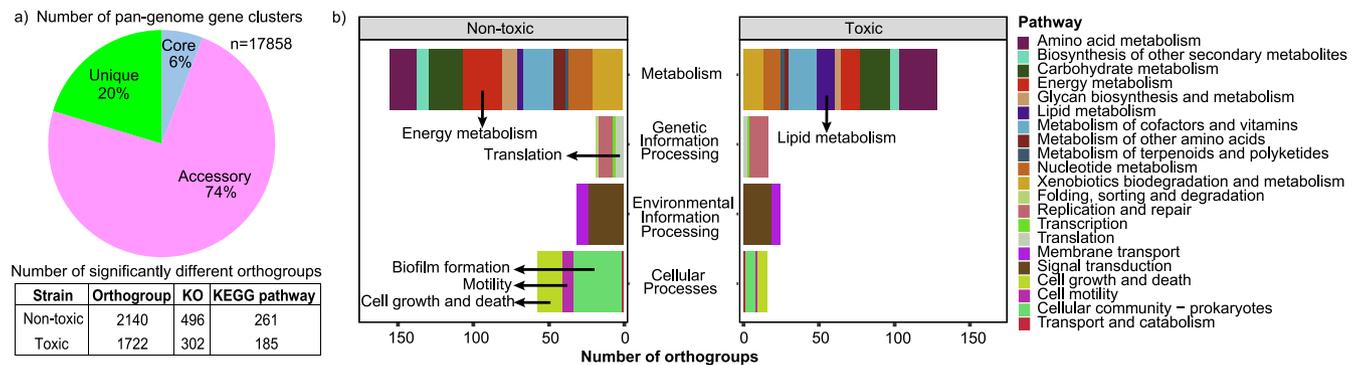


FIG 3 Pan-genome gene clusters and the orthogroups that were significantly higher in nontoxic or toxic groups. (a) The pie chart indicates the percentage of orthogroups that were classified as unique, accessory, or core according to their distribution among genomes. The orthogroups were annotated with KEGG Orthology (KO) and mapped to the KEGG pathway database. (b) Bar plots indicate the sum of the number of orthogroups that were significantly higher in either nontoxic or toxic strains. The orthogroups were categorized according to their KEGG functional pathway.

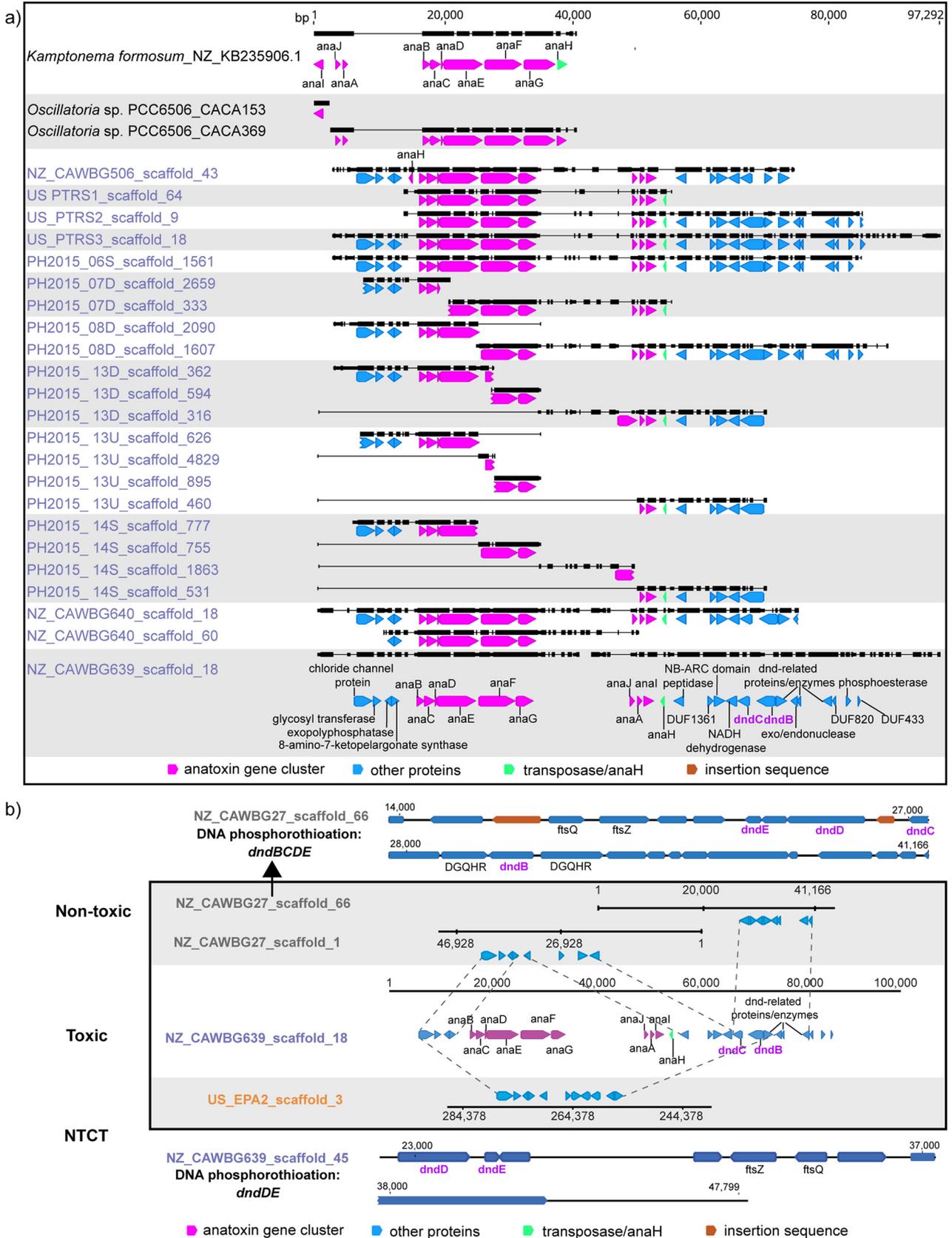


FIG 4 Multiple sequence alignments of anatoxin gene clusters and the neighboring genes. Black horizontal lines represent single contigs. Genes related to DNA phosphorothioation (*dndBCDE*) are indicated in bold purple font. Alternating white-gray shading denotes contigs from the same genome/MAG. (a) Comparison of anatoxin gene clusters present in *Microcoleus* genomes (blue font). Anatoxin gene clusters from *Kamptonema formosum* and

(Continued on next page)

and *Oscillatoria* sp. strain PCC6506 genomes (Fig. 4a). Transposase (*anaH*) genes were found next to the anatoxin gene clusters, highlighting the possibility of horizontal gene transfer and random gene rearrangement/loss. The order within *Microcoleus* anatoxin gene clusters was highly conserved, suggesting these genes were acquired through vertical gene transfer. One of the *Microcoleus* strains, NZ-CAWBG640, also harbored a second copy of the anatoxin gene cluster (Fig. 4a). These two gene cluster copies were highly similar, with 98% alignment coverage and 99.9% identity, implying either cobinching of a toxin cluster belonging to a conspecific strain (~3% predicted strain heterogeneity based on CheckM) or a gene duplication event. Potential duplication of toxin clusters would underline the importance of anatoxin production to these organisms (44, 45) and warrants further investigation. While the microcystin gene cluster is generally considered to be single copy (46, 47), a few studies have indicated that certain biosynthetic gene clusters, including nrps/pks (45), bacteriocins (48), and microginin (49), are present in multiple copies in cyanobacterial genomes.

Despite the absence of anatoxin gene clusters in the NTCT strains, genes neighboring the cluster in the toxic strains preserved a similar structural arrangement within the genomes of NTCT strains, including a collocated partial DNA modification/phosphorothioation cluster, *dndBC* (Fig. 4b). Multiple toxin gene loss events are predicted for *Microcystis* strains throughout their evolution, based on the heterogenous phylogeny of toxic and nontoxic strains (50). Evidence here suggests recent loss of the anatoxin cluster from the NTCT *Microcoleus* strains, which is further supported by the high level of ANI they share with their toxic relatives and their comparably small genome sizes (~6.1 ± 0.2 Mbp), low GC content (44.4 ± 0.1%), and low fraction of noncoding genes (Fig. 1a). Other nontoxic strains were instead equipped with a collocated, but complete, DNA modification cluster, *dndBCDE*, which functions to replace a nonbridging oxygen atom in the phosphodiester bond with sulfur to protect against nuclease activity or oxidative stress (51, 52). The presence of a DNA phosphorothioation system type-III restriction enzyme (RE) upstream of this *dndBCDE* cluster indicates that DNA phosphorothioation in *Microcoleus* likely functions as a restriction-modification (R-M) system, affording protection against phage by destroying unmodified foreign DNA (53).

Both *dndBC* and *dndDE* were present in toxic and NTCT strains but distributed across different contigs (as in toxic *M. aeruginosa* strains [54]) and separated by at least 40 kb (Fig. 4b), possibly influencing the expression and functionality of *dnd* gene products. The anatoxin gene cluster and incomplete *dnd* gene cluster (*dndBC*) are located near a type-III RE. Without a complete *dnd* gene cluster that functions as a DNA phosphorothioation system, the RE itself may induce programmed cell death (PCD) or cell lysis under stressful conditions (55) and subsequent release of the intracellular cyanotoxins into the environment (56). The functional coupling between PCD and microcystin release has been shown to promote the survival of the remaining *Microcystis* population under stress by promoting colony formation (57) and reducing grazing pressure (58). While anatoxins may behave differently than microcystin, we posit that toxic *Microcoleus* strains may undergo PCD and release intracellular anatoxin in response to stress, which likely offers protection to the wider mat community.

Nontoxic *Microcoleus* strains harbor diverse biosynthetic gene clusters. In concordance with genome size, toxic *Microcoleus* strains harbor significantly fewer biosynthetic gene clusters (BGCs) and insertion sequences (IS) than their nontoxic counterparts (Fig. 1a). Insertion sequences (IS) were widespread among all *Microcoleus* strains (1 to 3.6% of total genomic DNA; Fig. 5) although disproportionately more numerous in nontoxic strains (Fig. 1a). IS were largely associated with BGC and genomic island occurrences (Fig. 1a and 5) and likely facilitate genetic variation. BGCs are common in

FIG 4 Legend (Continued)

Oscillatoria sp. PCC6506 (black font) are used as references. (b) Comparison of genes neighboring the anatoxin gene cluster in *Microcoleus* genomes. The dotted lines represent the matching sections between toxic (unshaded box) and nontoxic strains (shaded box). The DNA phosphorothioation gene cluster, *dndBCDE*, from a representative nontoxic strain is shown at the top and is contained within a single contig. The *dndD* and *dndE* genes from a representative toxic strain are located on different contigs (bottom). NTCT, nontoxic clustered with toxic.

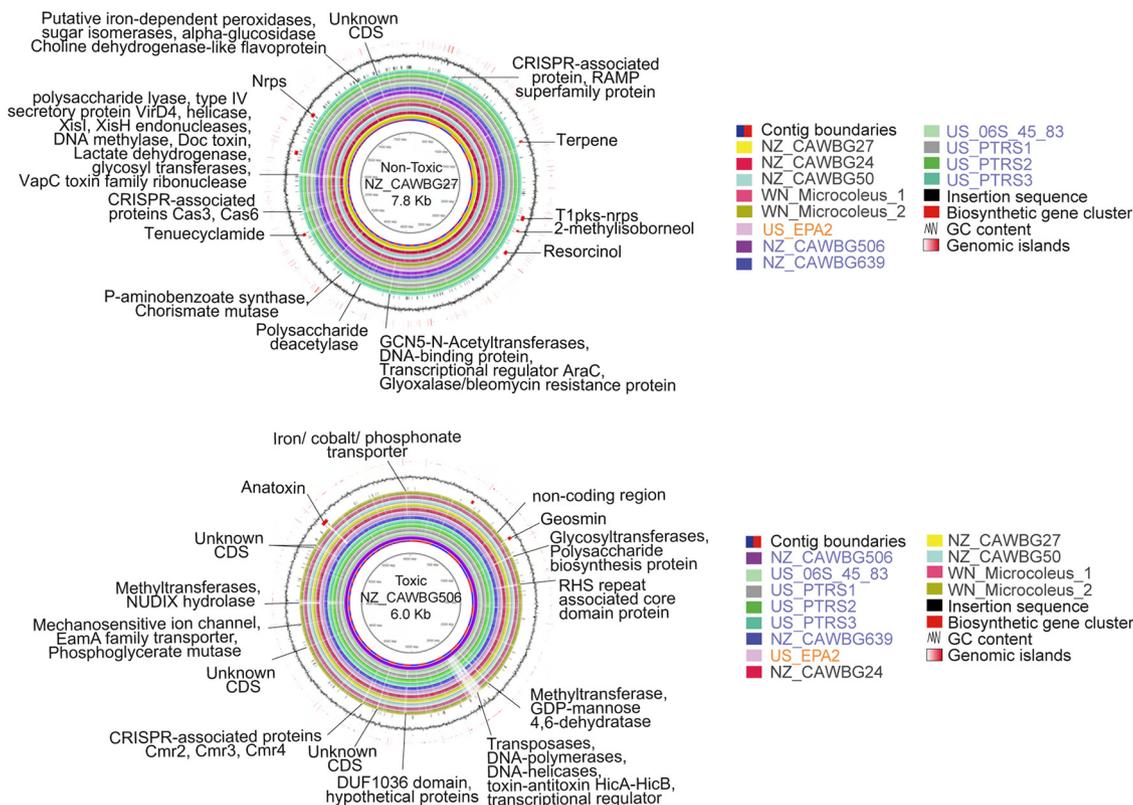


FIG 5 Whole-genome comparison plot. Circular maps of the alignment of *Microcoleus* genomes with a nontoxic (NZ-CAWBG27, top) or toxic (NZ-CAWBG506, bottom) strain as reference, and reference strain contigs ordered by size. The intensity of line color in the circular plot represents BLAST hit identity scores, while blanks indicate no match or nonsignificant matches. The first and second inner rings represent contig boundaries and reference genomes, respectively. From the 3rd to the 13th ring corresponds to a different genome. The 14th ring represents the insertion sequence, followed by biosynthetic gene clusters and GC content. The outermost ring corresponds to the genomic island score predicted with Alien Hunter v1.7.1. Notable features are labeled (e.g., BGCs and discrepancies between toxic and nontoxic genomes).

cyanobacteria and are responsible for the production of secondary metabolites, including toxins, antibiotics, and siderophores (45, 59). The production of secondary metabolites or cyanotoxins is closely linked to cyanobacterial evolution (2, 60) and may offer cyanobacteria a competitive or physiological advantage, allowing them to adapt and survive under a greater range of conditions (45, 61, 62). For example, secondary metabolites may offer protection against grazing/predation, UV radiation, and oxidative stress or promote efficient acquisition of limited nutrients (2, 60, 61).

Although toxic *Microcoleus* have few BGCs, besides the t1pks that is encoded by the ~29-kb anatoxin gene cluster (Fig. 5 and 6), geosmin and 2-methylisoborneol genes were detected in some toxic and nontoxic strains. These cause an earthy/musty odor in aquatic systems and are a nuisance to drinking water systems (63), which results indicate is a feature of both *Microcoleus* groups. However, we found nonribosomal peptide-synthetase (nrps), bacteriocin, resorcinol, and a type-I polyketide synthase (t1pks)-nrps hybrid BGCs only in nontoxic *Microcoleus* (Fig. 6). These types of BGCs are used to produce antibacterial toxin compounds, including tenuencyclamide, nostopeptolide, and nostophycin (Fig. 6), suggesting that non-anatoxin-producing *Microcoleus* can produce a suite of antibacterial compounds and may exhibit cytotoxic activity (62). Overall, results indicate that anatoxin gene loss (or lack of acquisition) may be offset by an increase in BGCs, which could drive adaptive diversification among nontoxic *Microcoleus* strains.

Large metabolic differences between toxic and nontoxic groups. KEGG pathway analysis highlighted that, aside from BGCs, toxic and nontoxic strains utilize different carbon storage/breakdown, nutrient acquisition and transport, and chemosensory and

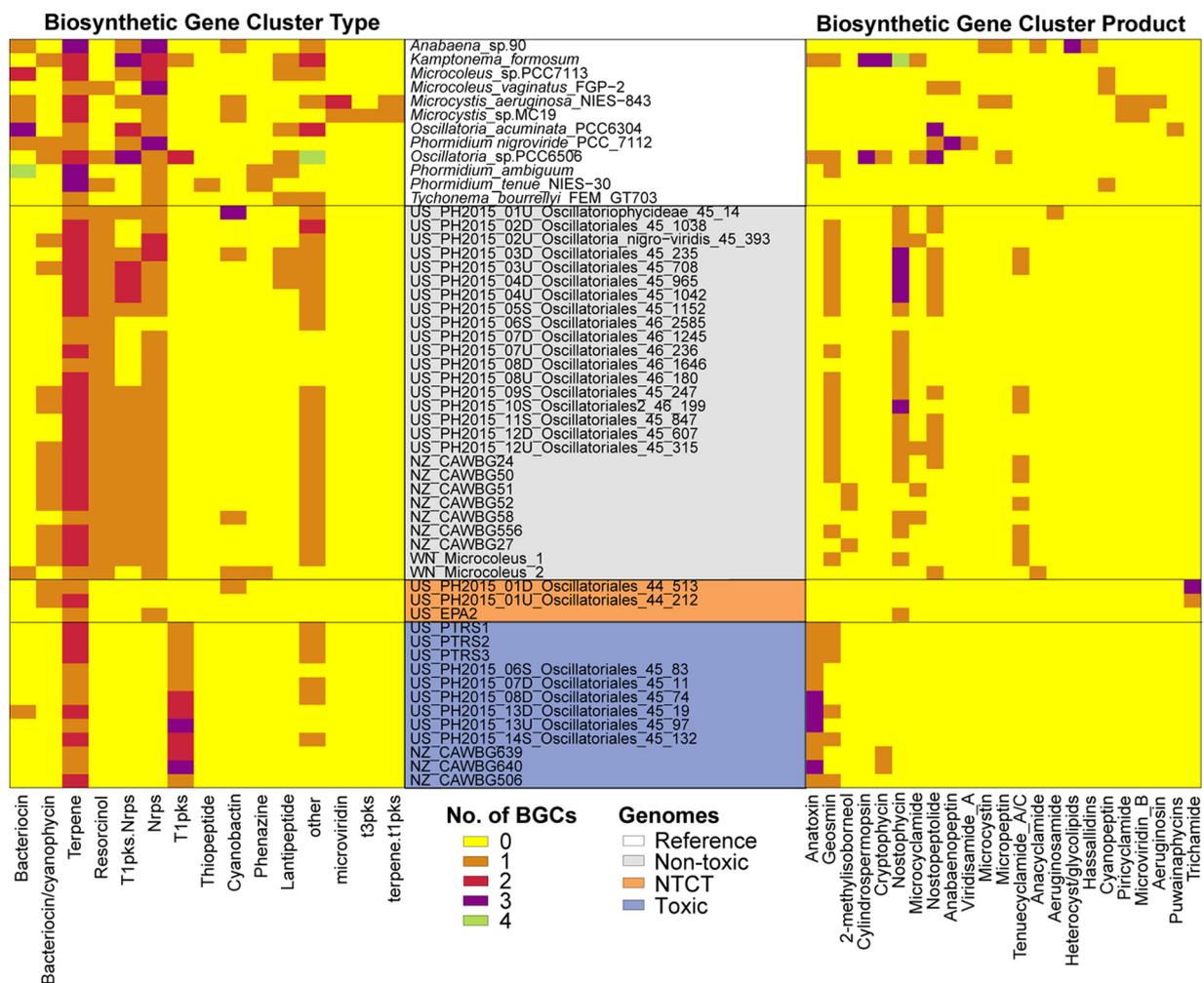


FIG 6 Heatmaps of cyanobacterial biosynthetic gene clusters (BGCs) identified among the *Microcoleus* spp. Heatmaps show the type of BGCs (left) and the known class of BGCs (right) identified among the genomes. The background color for text indicates reference genomes (black) and *Microcoleus* strains that are nontoxic (gray), nontoxic clustered with toxic (NTCT; orange), and toxic (blue). The data in the heatmaps were assigned individual colors based on the number of clusters found.

stress adaptation mechanisms (additional details in Text S1, Fig. S4, and Table S4). Significantly more orthologs involved in energy metabolism, biofilm formation, motility, and cell growth and death were found in nontoxic strains (Benjamini-Hochberg adjusted *P* value < 0.05; Fig. 3). Results further predict that nontoxic strains can synthesize sucrose and thiamine and take up alkanesulfonate as an alternative sulfur source, while toxic strains cannot (Fig. 7). Instead, there were significantly more orthologs related to lipid metabolism (FabD and MCH) in toxic strains (Fig. 3 and Fig. S4), which are implicated in anatoxin biosynthesis. The biosynthesis of anatoxin-a starts with proline adenylation by AnaC and attachment to an acyl carrier protein, AnaD (also known as FabD) (64). AnaA acts as a thioesterase (also known as MCH) and is involved in the final hydrolysis step in anatoxin production (65).

Variations in thiamine and sucrose biosynthesis. Thiamine (vitamin B₁) is an essential coenzyme that catalyzes transformations of carbon and biosynthesis of branched-chain amino acids in all living systems (66). The active form of the vitamin is thiamine diphosphate (TPP). Reductive evolution of bacterial genomes may lead to multiple auxotrophies, including vitamins and amino acids, which lead to dependent relationships with coexisting organisms (67–69). Experiments indicate that thiamine produced by phytoplankton can be used by cocultured auxotrophs to sustain their growth without

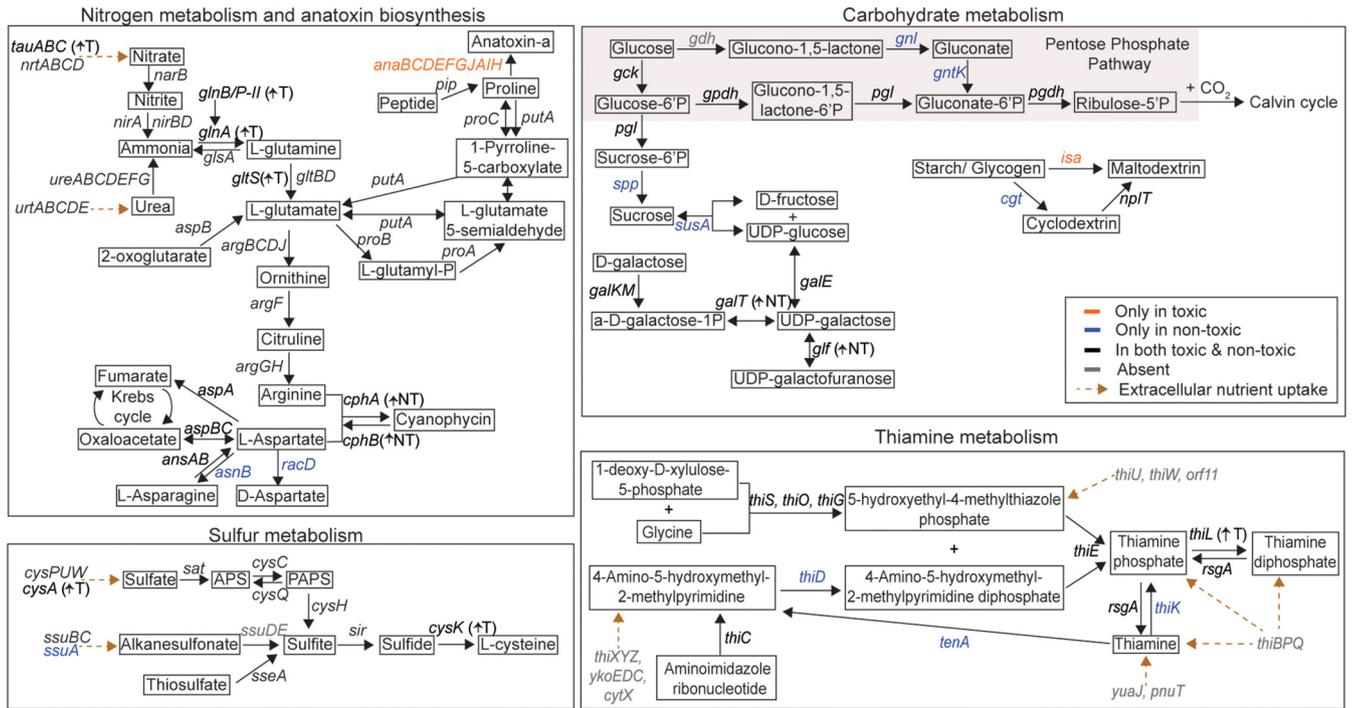


FIG 7 Schematic of metabolic pathways in *Microcoleus* spp. (gene details available in Tables S4 and S5). ↑NT represents more copies of genes in nontoxic strains, whereas ↑T depicts more copies of genes in toxic strains. Text color indicates the gene distribution across the *Microcoleus* genomes, as shown in the box on the right side of the diagram. Brown dotted arrows represent uptake of extracellular nutrients.

additional thiamine supply (70). We found a complete thiamine biosynthesis pathway (*tenA*, *rsgA*, *thiCDEGKLOS*) in the nontoxic strains, but not in toxic strains (Fig. 7 and Fig. S4). The loss of essential genes, *thiD*, *tenA*, and *thiK* for thiamine biosynthesis, salvage, and phosphorylation pathways in the reduced genomes of toxic *Microcoleus* may indicate thiamine auxotrophy and a dependency on cohabiting nontoxic strains to acquire vitamins. According to the Black Queen hypothesis, reductive evolution in free-living bacteria and associated gene loss can lead to dependencies on leaky “helper” organisms (71). Loss of function in toxic strains resulting in dependence on nontoxic strains for specific metabolites would explain why both toxic and nontoxic *Microcoleus* always co-occur in the natural environment (5, 11). We detected no thiamine uptake genes in any of the *Microcoleus* strains; however, this may be due to a lack of references. Thus far, no thiamine transporter system has been identified in cyanobacteria (72, 73). While no thiamine auxotrophy in cyanobacteria has been reported previously (72), many harmful algal bloom species, including some dinoflagellate and diatom species, are known thiamine auxotrophs (74, 75).

Orthogroups responsible for sucrose synthesis (sucrose-6-phosphatase [*spp*] and sucrose synthase [*susA*]) were only present in nontoxic strains (Fig. 7). Sucrose serves as one of the major compatible solutes among cyanobacteria in freshwater habitats, and the intracellular accumulation of sucrose helps cyanobacteria cope with salt stress (76–78). Missing genes related to sucrose synthesis suggest that toxic strains may be less tolerant to salt stress, possibly due to expendable gene loss, as they exclusively inhabit freshwater environments (12) and thus encounter minimum fluctuations in salinity.

Differences in starch and glucose utilization. Excess polysaccharide synthesized via photosynthesis during the day is stored as glycogen and starch in cyanobacteria (79). These storage molecules are then consumed to maintain levels of ATP and NADPH at night. Our data indicate that toxic and nontoxic *Microcoleus* utilize distinct enzymes for maltodextrin biosynthesis (Fig. 7 and Fig. S4)—a key intermediate product in glycogen and starch degradation. Toxic strains break down glycogen/starch into maltodextrin via isoamylase (ISA), whereas nontoxic strains first degrade starch into

cyclodextrin via cyclodextrin glucosyltransferase (CGT), which is then converted into maltodextrin via cyclomaltodextrinase (NplT) (Fig. 7). Cyclodextrin, which aids in drug/compound delivery, has been reported to improve antifungal activity in *Anabaena* spp. (80), suggesting that nontoxic *Microcoleus* may potentially produce cyclodextrin to enhance the delivery of other antibacterial or antifungal compounds.

The pentose phosphate pathway (PPP) is a crucial mechanism for the oxidation of glucose and NADPH generation in most organisms (81). The product, ribulose-5-phosphate, can be phosphorylated into ribulose-1,5-diphosphate in the dark reaction-Calvin cycle. A complete PPP that generates ribulose-5-phosphate from glucose (*gck*) and then glucose-6-phosphate (*gpdh*, *pgl*, *pgdh*) was found within *Microcoleus* genomes (Fig. 7). Results indicate that nontoxic strains are also able to use an alternative nonphosphorylated route for directing the intermediate, gluconolactone, into the PPP, which bypasses the rate-limiting enzyme glucose 6-phosphate dehydrogenase (G6PDH) (82, 83). The genes involved transform gluconic-acid/gluconolactone to gluconate-6-phosphate (*gnl* and *gntk*), which can then be converted to ribulose-5-phosphate by *pgdh*. However, the glucose dehydrogenase gene, *gdh*, which is involved in the breakdown of glucose to gluconolactone, the first step of the nonphosphorylated route, is missing from *Microcoleus* genomes. This suggests that the nontoxic strains may possess novel enzymes that substitute for glucose dehydrogenase, or they can acquire and utilize gluconolactone from the environment.

Distinct nitrogen, phosphorus, and sulfur acquisition and storage mechanisms in toxic and nontoxic groups. Analysis of pathways for nitrogen metabolism showed toxic and nontoxic *Microcoleus* differ in their capacity to take up and store nitrogen. While all *Microcoleus* strains harbor the classic high-affinity nitrate/nitrite transport (Nrt) and urea uptake (Ure) systems, toxic strains also harbor an additional nitrate/sulfonate/taurine transport system (NitT/TauT) (Fig. 7). This suggests that toxic strains have a greater capacity to acquire nitrogen and proliferate in a nitrogen-rich environment, which corroborates prior findings that river sites with higher nitrogen concentrations tended to have higher relative abundances of toxic versus nontoxic strains (8). However, results here show that nontoxic species are likely to be more resistant to variations in nitrogen supply. An alternative pathway for nitrogen assimilation among *Microcoleus* species involves cyanophycin synthesis and catabolism, by CphAB (19), which functions as temporary nitrogen/carbon storage (84). Toxic *Microcoleus* strains are equipped with one cyanophycin gene cluster (Fig. 7). In contrast, nontoxic strains are equipped with two distinct cyanophycin gene clusters. The additional gene cluster may increase maximum rates of cyanophycin metabolism, helping nontoxic strains to adapt to fluctuating nitrogen concentrations in the environment (84).

Results indicate a link between nitrogen acquisition and toxin production. Glutamate is known to be the most common precursor for proline biosynthesis (85), which plays a significant role in anatoxin production. Besides acting as a building block for secondary metabolites, glutamate is essential for nitrogen assimilation (86). Although genes responsible for proline production are present in both toxic and nontoxic strains, more copies of glutamine synthetase, *glnA*, and glutamate synthase, *gltS*, were found in toxic strains than in nontoxic strains (Fig. 7). Toxic strains, therefore, may accumulate more glutamine/glutamate, contributing to anatoxin biosynthesis and greater nitrogen assimilation. In addition, *Microcoleus* mats frequently proliferate in low-phosphate water and are able acquire/uptake nutrients under nutrient-limiting conditions (8, 12, 19). Accordingly, *Microcoleus* genomes in the present study contain diverse phosphate acquisition mechanisms (Table S5). While all *Microcoleus* strains harbor at least one copy of phosphatase and phosphate transporter orthologs, some toxic strains harbor multiple copies (Fig. S4), which potentially results in more efficient phosphate acquisition than in nontoxic counterparts.

Evidence suggests toxic strains invest greater resources into cysteine production. All *Microcoleus* strains have the capacity to acquire sulfur via a complete assimilatory sulfate reduction pathway (Fig. 7). Extracellular sulfate is transported into the cells via the high-affinity sulfate/thiosulfate uptake system, CysPUWA, reduced to sulfide via sulfite reductase (SiR), and then incorporated into cysteine by cysteine synthases (CysK or CysM). CysK can also bind and activate antibacterial toxins upon entry into target

cells (87). Multiple copies of the *cysK* cysteine synthase genes and *cysA* transporter ATP binding subunit were found in toxic strains, while on average only one copy was found in nontoxic strains, implying toxic strains utilize CysK to perform additional biological functions. Additional copies of *cysK* and *cysA* may also enhance cysteine uptake, as has been shown for other genes (e.g., additional cloned copies of threonine synthase and transporter genes were shown to increase threonine production [88]). Greater synthesis of cysteine by toxic *Microcoleus* may be needed to overcome attrition of cysteine-bearing proteins. Previous studies on hepatotoxin-microcystin have shown that microcystin can covalently bind to the cysteine residue of specific proteins, which interferes with the stability and activity of these cysteine-bearing proteins (89, 90).

Conclusion. This study highlights remarkable genomic and metabolic differences within the *Microcoleus* genus, leading to their divergent evolution. Results indicate toxic strains adopted a genome-streamlining strategy, resulting in smaller genomes, fewer BGCs, and smaller noncoding gene fractions. The estimated maximum growth rate for the toxic group is lower, suggesting that ecological trade-offs likely accompany anatoxin production in *Microcoleus*. Such trade-offs are further characterized by metabolic deficiencies, including sucrose and thiamine synthesis genes, and other stress response mechanisms. Toxic and nontoxic strains may employ both cooperative strategies (by offering protection and sharing vitamins) and competitive strategies (nutrient uptake and assimilation) to coexist in mats and utilize the same resource pool. Genomic evidence leads us to predict that nontoxic *Microcoleus* strains synthesize and share thiamine with cohabiting toxic *Microcoleus*, while anatoxin-producing strains, by triggering PCD and releasing intracellular anatoxin, provide a resource to coexisting nontoxic strains (potentially affording protection from predation, oxidative stress, or nutrient limitation). Toxic and nontoxic strains are differently equipped with additional sets of nitrate transporter and cyanophycin genes, which likely confer distinct competitive advantages under fluctuating nitrogen availability. Understanding the genomic features that differentiate nontoxic versus toxic groups, provides a basis for assessing how different environmental factors affect their selection and proliferation in freshwater systems.

MATERIALS AND METHODS

Cyanobacterial strains and genome sequencing. The genomic features and metadata of 42 *Microcoleus* strains, including 28 MAGs and 14 isolates, used in this study are shown in Table S1. Representatives were selected for this study based on close similarity to *Microcoleus autumnalis* (8, 19, 30) and originated from 2 countries and 13 rivers. Of these, 11 nonaxenic *Microcoleus* cultures were obtained from the Cawthron Institute's Culture Collection of Micro-algae (CICCM, Nelson, New Zealand; <http://cultures.cawthron.org.nz>). These strains were grown in liquid MLA medium (91) and incubated under standard conditions (160 ± 20 mmol photons $m^{-2} s^{-1}$; 12:12 h light:dark cycles; $18 \pm 1^\circ C$) for at least 3 weeks to achieve adequate biomass for DNA extraction. DNA was extracted from 0.1 to 0.4 g of each sample using a DNeasy PowerSoil DNA kit (Qiagen, USA). Genomic libraries with 550-bp insert sizes were prepared with a TruSeq DNA nano library preparation kit, and 2×250 -bp sequencing was performed using the Illumina HiSeq 2500 platform with V2 chemistry at the Otago Genomics Facility (University of Otago, New Zealand). The Illumina sequence reads of three additional nonaxenic *Microcoleus* cultures from the United States (30) were downloaded from the NCBI database (92). The 28 environmental metagenome-assembled *Microcoleus* genomes and their corresponding Illumina sequence reads were obtained from prior studies (8, 19).

Read processing and genome assembly. All sequence reads associated with the 42 strains were quality checked using FastQC (93). Adapter sequences were removed, and reads were trimmed with `bbduk.sh` from BBDuk v37.93 (94); only those with a quality score of ≥ 30 and length of ≥ 80 bp were retained. Trimmed reads from each CICCM-derived culture, and the additional three nonaxenic *Microcoleus* isolates from the United States, were assembled using metaSPAdes (95) with k-mer values of 41, 61, 81, 101, and 127. Scaffolds larger than 2 kb were binned using MetaBAT (96), CONCOCT (97), and MaxBin 2.0 (98). The highest-quality nonredundant prokaryotic bins from each assembly were selected using DASTool (99) and CheckM v1.0.13 (100) and were used to estimate genome completeness and contamination. All bins (11 in total) were manually curated and validated using VizBin (101). Genome coverage was calculated by mapping reads to genomes using Bowtie v1.2.0 (102), allowing ≤ 1 mismatch per read pair, showing that >10 times read coverage was obtained for each bin (Table S1). Estimated genome size was calculated as follows: total number of bases in the genome $\times 100/\text{CheckM-estimated completeness based on lineage-specific marker genes}$.

Small subunit (SSU) rRNA gene reconstruction. Full-length 16S rRNA sequences were reconstructed from the trimmed reads of all 42 strains, over 50 iterations, using EMIRGE (103) with the SILVA SSURef NR99 132 database (104) and a clustering threshold of 99%.

Genome annotation and prediction of minimum generation time. The 42 *Microcoleus* genomes were annotated as follows. Open reading frames (ORFs) were predicted using Prodigal (105) and subjected to orthologous protein grouping using OrthoFinder v2.3.3 (106) with default parameters. The orthologous clusters were classified as core, accessory, or unique according to their distribution across the genomes. The core orthogroups comprise predicted proteins shared by all strains; the accessory clusters incorporate proteins assigned to a subset of study strains; and the unique clusters include proteins assigned to only a single strain. The orthologous groups were then annotated using KofamKOALA and the Kyoto Encyclopedia of Genes and Genomes (KEGG) release 94.0 database (Table S4) (107). Genome sequences were screened for biosynthetic gene clusters (BGCs) using antiSMASH v4.2 (108), insertion elements were predicted and classified with ISEScan v1.7.1 (109), and genomic islands were predicted using Alien Hunter v1.7.1 (110). Anatoxin gene clusters and neighboring genes were aligned using Mauve multiple alignments (111). Snippy v4.4.0 (112) was used for variant calling (SNPs or insertion/deletion) and generation of core genome SNP alignments among all *Microcoleus* genomes. Minimum generation times were estimated for all *Microcoleus* genomes based on the codon usage bias from a set of typically highly expressed genes (mainly rRNA, tRNA) using growthpred (113) with the parameters “-b -c 0 -r -T 20 -S.”

Whole-genome comparisons and phylogenetic analyses. Whole-genome comparisons were conducted via BLASTN and visualized using the BLAST Ring Image Generator (BRIG [114]). Genome sequences related to *Microcoleus* were retrieved from the NCBI genome database (92) (Table S1). The phylogenetic relationship of *Microcoleus* genomes was inferred by pairwise nucleotide-level comparisons based on digital DNA-DNA hybridization (dDDH) (41, 115) and average nucleotide identity (ANI) (116) values. Pairwise dDDH values were estimated using the Genome-To-Genome Distance Calculator (GGDC) v2.1 with a threshold of 70% for species delineation, which corresponds to at least 96.5% ANI and an alignable gene fraction of >70% (42). Reciprocal BLASTN was carried out for all genomes to calculate ANI. Maximum-likelihood trees with branch supports were constructed based on reconstructed 16S rRNA genes, 12,421 core single nucleotide polymorphisms (SNPs), and concatenated alignments of 120 single-copy core marker genes obtained from GTDB-Tk v0.2.1 (117). Trees were built using the ultrafast bootstrap approximation (118) in IQ-TREE v1.6.9 (119). The species phylogenetic tree was inferred based on 525 single-copy core orthologs (347,012 amino acid) using FastTree v2.1.10 (120).

Cyanotoxin measurement. Anatoxin-a, homoanatoxin-a, and structural variants were measured using LC-MS/MS, as described in a previous study (3). Biofilm samples were lyophilized and resuspended in 10 ml of Milli-Q water with 0.1% formic acid. Samples were then separated by liquid chromatography (Waters Acquity UPLC; Waters Corp., Massachusetts, USA) on a BEH C₁₈ column (1.7 μm, 1 × 50 mm; Waters Corp.) and quantified on a Quattro Premier XE triple quadrupole mass spectrometer (Waters-Micromass, Manchester, UK).

Statistical analysis. The following statistical analyses were carried out in R environment version 4.0.2 (121). A Wilcoxon rank-sum test was performed to determine the significant differences of genome sizes, GC content, predicted maximum growth rate, number of insertion sequences, number of biosynthetic gene clusters, and orthogroup count between toxic and nontoxic *Microcoleus* strains. Pearson correlation coefficients and regression lines between genome size and the gene content and estimated genome completeness were calculated and plotted using the `stat_cor` function in `ggpubr` (122) and `geom_smooth` function in the `ggplot2` package (123). Heatmaps with `ward.D2` hierarchical clustering were plotted using `phemap` (124).

Data availability. The data generated in this study are publicly available. All sequence data have been deposited with NCBI under BioProject PRJNA733706. Codes for read processing and genome assembly are available at https://github.com/HweeSze/Microcoleus_comparative_manuscript.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.02 MB.

FIG S1, PDF file, 1.5 MB.

FIG S2, EPS file, 0.8 MB.

FIG S3, PDF file, 2.7 MB.

FIG S4, PDF file, 0.6 MB.

TABLE S1, XLSX file, 0.02 MB.

TABLE S2, XLSX file, 0.01 MB.

TABLE S3, XLSX file, 2.6 MB.

TABLE S4, XLSX file, 0.1 MB.

TABLE S5, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

This study was supported by Genomics Aotearoa (project 1806), a New Zealand Ministry of Business, Innovation, and Employment-funded research platform (<http://www.genomics-aotearoa.org.nz/>).

We thank Laura Kelly (Cawthron Institute) for assistance in culturing the cyanobacterial isolates and Chris Nietch (U.S. EPA Experimental Stream Facility) for supplying the field material for isolation of the EPA strain. Computational resources were provided by New Zealand eScience Infrastructure.

REFERENCES

- Shih PM. 2015. Cyanobacterial evolution: fresh insight into ancient questions. *Curr Biol* 25:R192–R193. <https://doi.org/10.1016/j.cub.2014.12.046>.
- 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>.
- Heath MW, Wood SA, Ryan KG. 2010. Polyphasic assessment of fresh-water benthic mat-forming cyanobacteria isolated from New Zealand. *FEMS Microbiol Ecol* 73:95–109. <https://doi.org/10.1111/j.1574-6941.2010.00867.x>.
- Kardinaal WEA, Janse I, Agterveld MK, Meima M, Snoek J, Mur LR, Huisman J, Zwart G, Visser P. 2007. *Microcystis* genotype succession in relation to microcystin concentrations in freshwater lakes. *Aquat Microb Ecol* 48:1–12. <https://doi.org/10.3354/ame048001>.
- Wood SA, Smith FMJ, Heath MW, Palfroy T, Gaw S, Young RG, Ryan KG. 2012. Within-mat variability in anatoxin-a and homoanatoxin-a production among benthic *Phormidium* (Cyanobacteria) strains. *Toxins* (Basel) 4: 900–912. <https://doi.org/10.3390/toxins4100900>.
- Willis A, Woodhouse JN, Ongley SE, Jex AR, Burford MA, Neilan BA. 2018. Genome variation in nine co-occurring toxic *Cylindrospermopsis raciborskii* strains. *Harmful Algae* 73:157–166. <https://doi.org/10.1016/j.hal.2018.03.001>.
- Park BS, Li Z, Kang Y-H, Shin HH, Joo J-H, Han M-S. 2018. Distinct bloom dynamics of toxic and non-toxic *Microcystis* (Cyanobacteria) subpopulations in Hoedong Reservoir (Korea). *Microb Ecol* 75:163–173. <https://doi.org/10.1007/s00248-017-1030-y>.
- Bouma-Gregson K, Olm MR, Probst AJ, Anantharaman K, Power ME, Banfield JF. 2019. Impacts of microbial assemblage and environmental conditions on the distribution of anatoxin-a producing cyanobacteria within a river network. *ISME J* 13:1618–1634. <https://doi.org/10.1038/s41396-019-0374-3>.
- Heath M, Wood S, Ryan K. 2011. Spatial and temporal variability in *Phormidium* mats and associated anatoxin-a and homoanatoxin-a in two New Zealand rivers. *Aquat Microb Ecol* 64:69–79. <https://doi.org/10.3354/ame01516>.
- Echenique-Subiabre I, Zancarini A, Heath MW, Wood SA, Quiblier C, Humbert J-F. 2018. Multiple processes acting from local to large geographical scales shape bacterial communities associated with *Phormidium* (cyanobacteria) biofilms in French and New Zealand rivers. *Sci Rep* 8:14416. <https://doi.org/10.1038/s41598-018-32772-w>.
- Wood SA, Heath MW, Kuhajek J, Ryan KG. 2010. Fine-scale spatial variability in anatoxin-a and homoanatoxin-a concentrations in benthic cyanobacterial mats: implication for monitoring and management. *J Appl Microbiol* 109:2011–2018. <https://doi.org/10.1111/j.1365-2672.2010.04831.x>.
- McAllister TG, Wood SA, Hawes I. 2016. The rise of toxic benthic *Phormidium* proliferations: a review of their taxonomy, distribution, toxin content and factors regulating prevalence and increased severity. *Harmful Algae* 55:282–294. <https://doi.org/10.1016/j.hal.2016.04.002>.
- Wood SA, Kelly LT, Bouma-Gregson K, Humbert J-F, Laughinghouse HD, Lazorchak J, McAllister TG, McQueen A, Pokrzywinski K, Puddick J, Quiblier C, Reitz LA, Ryan KG, Vadeboncoeur Y, Zastepa A, Davis TW. 2020. Toxic benthic freshwater cyanobacterial proliferations: challenges and solutions for enhancing knowledge and improving monitoring and mitigation. *Freshw Biol* 65:1824–1842. <https://doi.org/10.1111/fwb.13532>.
- Heath M, Wood SA, Young RG, Ryan KG. 2016. The role of nitrogen and phosphorus in regulating *Phormidium* sp. (cyanobacteria) growth and anatoxin production. *FEMS Microbiol Ecol* 92:fiw021. <https://doi.org/10.1093/femsec/fiw021>.
- Wood SA, Puddick J. 2017. The abundance of toxic genotypes is a key contributor to anatoxin variability in *Phormidium*-dominated benthic mats. *Mar Drugs* 15:307. <https://doi.org/10.3390/md15100307>.
- Kelly LT, Wood SA, McAllister TG, Ryan KG. 2018. Development and application of a quantitative PCR assay to assess genotype dynamics and anatoxin content in *Microcoleus autumnalis*-dominated mats. *Toxins* 10:431. <https://doi.org/10.3390/toxins10110431>.
- Lu J, Zhu B, Struewing I, Xu N, Duan S. 2019. Nitrogen-phosphorus-associated metabolic activities during the development of a cyanobacterial bloom revealed by metatranscriptomics. *Sci Rep* 9:2480. <https://doi.org/10.1038/s41598-019-38481-2>.
- Schindler DW, Hecky RE, Findlay DL, Stainton MP, Parker BR, Paterson MJ, Beaty KG, Lyng M, Kasian SEM. 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci U S A* 105:11254–11258. <https://doi.org/10.1073/pnas.0805108105>.
- Tee HS, Waite D, Payne L, Middleditch M, Wood S, Handley KM. 2020. Tools for successful proliferation: diverse strategies of nutrient acquisition by a benthic cyanobacterium. *ISME J* 14:2164–2178. <https://doi.org/10.1038/s41396-020-0676-5>.
- Chorus I, Welker M. 2021. Cyanobacterial toxins, p 13–162. *In* Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management, 2nd ed. CRC Press, Boca Raton, FL.
- Abreu VAC, Popin RV, Alvarenga DO, Schaker PDC, Hoff-Riseti C, Varani AM, Fiore MF. 2018. Genomic and genotypic characterization of *Cylindrospermopsis raciborskii*: toward an intraspecific phylogenetic evaluation by comparative genomics. *Front Microbiol* 9:306. <https://doi.org/10.3389/fmicb.2018.00306>.
- Christiansen G, Molitor C, Philmus B, Kurmayer R. 2008. Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. *Mol Biol Evol* 25:1695–1704. <https://doi.org/10.1093/molbev/msn120>.
- 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>.
- Humbert J-F, Barbe V, Latifi A, Gugger M, Calteau A, Coursin T, Lajus A, Castelli V, Oztas S, Samson G, Longin C, Medigue C, de Marsac NT. 2013. A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. *PLoS One* 8:e70747. <https://doi.org/10.1371/journal.pone.0070747>.
- Pancrace C, Barny M-A, Ueoka R, Calteau A, Scalvenzi T, Pédrón 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>.
- Sinha R, Pearson LA, Davis TW, Muenchhoff J, Pratama R, Jex A, Burford MA, Neilan BA. 2014. Comparative genomics of *Cylindrospermopsis raciborskii* strains with differential toxicities. *BMC Genomics* 15:83. <https://doi.org/10.1186/1471-2164-15-83>.
- Catherine Q, Susanna W, Isidora E-S, Mark H, Aurélie V, Jean-François H. 2013. A review of current knowledge on toxic benthic freshwater cyanobacteria: ecology, toxin production and risk management. *Water Res* 47: 5464–5479. <https://doi.org/10.1016/j.watres.2013.06.042>.
- Willis A, Woodhouse JN. 2020. Defining cyanobacterial species: diversity and description through genomics. *Crit Rev Plant Sci* 39:101–124. <https://doi.org/10.1080/07352689.2020.1763541>.
- Strunecký O, Komárek J, Johansen J, Lukešová A, Elster J. 2013. Molecular and morphological criteria for revision of the genus *Microcoleus* (*Oscillatoriales*, *Cyanobacteria*). *J Phycol* 49:1167–1180. <https://doi.org/10.1111/jpy.12128>.
- Conklin KY, Stancheva R, Otten TG, Fadness R, Boyer GL, Read B, Zhang X, Sheath RG. 2020. Molecular and morphological characterization of a novel dihydroanatoxin-a producing *Microcoleus* species (cyanobacteria) from the Russian River, California, USA. *Harmful Algae* 93:101767. <https://doi.org/10.1016/j.hal.2020.101767>.
- Larsson J, Nylander JA, Bergman B. 2011. Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evol Biol* 11:187. <https://doi.org/10.1186/1471-2148-11-187>.
- Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. *ISME J* 8:1553–1565. <https://doi.org/10.1038/ismej.2014.60>.

33. Dittmann E, Fewer DP, Neilan BA. 2013. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. *FEMS Microbiol Rev* 37:23–43. <https://doi.org/10.1111/j.1574-6976.2012.12000.x>.
34. Burford MA, Davis TW, Orr PT, Sinha R, Willis A, Neilan BA. 2014. Nutrient-related changes in the toxicity of field blooms of the cyanobacterium, *Cylindrospermopsis raciborskii*. *FEMS Microbiol Ecol* 89:135–148. <https://doi.org/10.1111/1574-6941.12341>.
35. Van de Waal DB, Verspagen JM, Finke JF, Vournazou V, Immers AK, Kardinaal WEA, Tonk L, Becker S, Van Donk E, Visser PM, Huisman J. 2011. Reversal in competitive dominance of a toxic versus non-toxic cyanobacterium in response to rising CO₂. *ISME J* 5:1438–1450. <https://doi.org/10.1038/ismej.2011.28>.
36. Kardinaal WEA, Tonk L, Janse I, Hol S, Slot P, Huisman J, Visser PM. 2007. Competition for light between toxic and nontoxic strains of the harmful cyanobacterium *Microcystis*. *Appl Environ Microbiol* 73:2939–2946. <https://doi.org/10.1128/AEM.02892-06>.
37. Vézic C, Rapala J, Vaitomaa J, Seitonen J, Sivonen K. 2002. Effect of nitrogen and phosphorus on growth of toxic and nontoxic *Microcystis* strains and on intracellular microcystin concentrations. *Microb Ecol* 43:443–454. <https://doi.org/10.1007/s00248-001-0041-9>.
38. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martínez-García M, González JM, Luo H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* 110:11463–11468. <https://doi.org/10.1073/pnas.1304246110>.
39. Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 187:6258–6264. <https://doi.org/10.1128/JB.187.18.6258-6264.2005>.
40. Rodríguez-R LM, Castro JC, Kyrpides NC, Cole JR, Tiedje JM, Konstantinidis KT. 2018. How much do rRNA gene surveys underestimate extant bacterial diversity? *Appl Environ Microbiol* 84:e00014-18. <https://doi.org/10.1128/AEM.00014-18>.
41. Auch AF, von JM, Klenk H-P, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2:117–134. <https://doi.org/10.4056/signs.531120>.
42. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 43:6761–6771. <https://doi.org/10.1093/nar/gkv657>.
43. Sriswasdi S, Yang C, Iwasaki W. 2017. Generalist species drive microbial dispersion and evolution. *Nat Commun* 8:1162. <https://doi.org/10.1038/s41467-017-01265-1>.
44. Bratlie MS, Johansen J, Sherman BT, Huang DW, Lempicki RA, Drablos F. 2010. Gene duplications in prokaryotes can be associated with environmental adaptation. *BMC Genomics* 11:588. <https://doi.org/10.1186/1471-2164-11-588>.
45. Calteau A, Fewer DP, Latifi A, Coursin T, Laurent T, Jokela J, Kerfeld CA, Sivonen K, Piel J, Gugger M. 2014. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in *Cyanobacteria*. *BMC Genomics* 15:977. <https://doi.org/10.1186/1471-2164-15-977>.
46. Pacheco ABF, Guedes IA, Azevedo SMFO. 2016. Is qPCR a reliable indicator of cyanotoxin risk in freshwater? *Toxins* 8:172. <https://doi.org/10.3390/toxins8060172>.
47. Otten TG, Paerl HW, Dreher TW, Kimmerer WJ, Parker AE. 2017. The molecular ecology of *Microcystis* sp. blooms in the San Francisco Estuary. *Environ Microbiol* 19:3619–3637. <https://doi.org/10.1111/1462-2920.13860>.
48. Wang H, Fewer DP, Sivonen K. 2011. Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in *Cyanobacteria*. *PLoS One* 6:e22384. <https://doi.org/10.1371/journal.pone.0022384>.
49. Rounge TB, Rohrlack T, Nederbragt AJ, Kristensen T, Jakobsen KS. 2009. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planctothrix rubescens* strain. *BMC Genomics* 10:396. <https://doi.org/10.1186/1471-2164-10-396>.
50. Otsuka S, Suda S, Li R, Watanabe M, Oyaizu H, Matsumoto S, Watanabe MM. 1999. Phylogenetic relationships between toxic and non-toxic strains of the genus *Microcystis* based on 16S to 23S internal transcribed spacer sequence. *FEMS Microbiol Lett* 172:15–21. <https://doi.org/10.1111/j.1574-6968.1999.tb13443.x>.
51. Pu T, Liang J, Mei Z, Yang Y, Wang J, Zhang W, Liang W-J, Zhou X, Deng Z, Wang Z. 2019. Phosphorothioated DNA is shielded from oxidative damage. *Appl Environ Microbiol* 85:e00104-19. <https://doi.org/10.1128/AEM.00104-19>.
52. Vosberg HP, Eckstein F. 1982. Effect of deoxynucleoside phosphorothioates incorporated in DNA on cleavage by restriction enzymes. *J Biol Chem* 257:6595–6599. [https://doi.org/10.1016/S0021-9258\(20\)65184-5](https://doi.org/10.1016/S0021-9258(20)65184-5).
53. Wang L, Jiang S, Deng Z, Dedon PC, Chen S. 2019. DNA phosphorothioate modification: a new multi-functional epigenetic system in bacteria. *FEMS Microbiol Rev* 43:109–122. <https://doi.org/10.1093/femsre/fuy036>.
54. Frangeul L, Quillardet P, Castets A-M, Humbert J-F, Mattheijs HC, Cortez D, Tolonen A, Zhang C-C, Gribaldo S, Kehr J-C, Zilliges Y, Ziemert N, Becker S, Talla E, Latifi A, Billault A, Lepelletier A, Dittmann E, Bouchier C, Tandeau de Marsac N. 2008. Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genomics* 9:274. <https://doi.org/10.1186/1471-2164-9-274>.
55. Nagamalleswari E, Rao S, Vasu K, Nagaraja V. 2017. Restriction endonuclease triggered bacterial apoptosis as a mechanism for long time survival. *Nucleic Acids Res* 45:8423–8434. <https://doi.org/10.1093/nar/gkx576>.
56. Ibelings BW, Kurmayer R, Azevedo SMFO, Wood SA, Chorus I, Welker M. 2021. Understanding the occurrence of cyanobacteria and cyanotoxins, p 213–294. *In* Chorus I, Welker M (ed), *Toxic cyanobacteria in water*, 2nd ed. CRC Press, Boca Raton, FL.
57. Hu C, Rzymiski P. 2019. Programmed cell death-like and accompanying release of microcystin in freshwater bloom-forming cyanobacterium *Microcystis*: from identification to ecological relevance. *Toxins* 11:706. <https://doi.org/10.3390/toxins11120706>.
58. Rzymiski P, Klimaszuk P, Jurczak T, Poniedziałek B. 2020. Oxidative stress, programmed cell death and microcystin release in *Microcystis aeruginosa* in response to *Daphnia* grazers. *Front Microbiol* 11:1201. <https://doi.org/10.3389/fmicb.2020.01201>.
59. Singh RK, Tiwari SP, Rai AK, Mohapatra TM. 2011. Cyanobacteria: an emerging source for drug discovery. *J Antibiot (Tokyo)* 64:401–412. <https://doi.org/10.1038/ja.2011.21>.
60. Leão PN, Engene N, Antunes A, Gerwick WH, Vasconcelos V. 2012. The chemical ecology of cyanobacteria. *Nat Prod Rep* 29:372–391. <https://doi.org/10.1039/c2np00075j>.
61. Holland A, Kinnear S. 2013. Interpreting the possible ecological role(s) of cyanotoxins: compounds for competitive advantage and/or physiological aide? *Mar Drugs* 11:2239–2258. <https://doi.org/10.3390/md11072239>.
62. Dittmann E, Gugger M, Sivonen K, Fewer DP. 2015. Natural product biosynthetic diversity and comparative genomics of the *Cyanobacteria*. *Trends Microbiol* 23:642–652. <https://doi.org/10.1016/j.tim.2015.07.008>.
63. Izaguirre G, Hwang CJ, Krasner SW, McGuire MJ. 1982. Geosmin and 2-methylisoborneol from cyanobacteria in three water supply systems. *Appl Environ Microbiol* 43:708–714. <https://doi.org/10.1128/aem.43.3.708-714.1982>.
64. Méjean A, Paci G, Gautier V, Ploux O. 2014. Biosynthesis of anatoxin-a and analogues (anatoxins) in cyanobacteria. *Toxicol* 91:15–22. <https://doi.org/10.1016/j.toxicol.2014.07.016>.
65. Kust A, Méjean A, Ploux O. 2020. Biosynthesis of anatoxins in cyanobacteria: identification of the carboxy-anatoxins as the penultimate biosynthetic intermediates. *J Nat Prod* 83:142–151. <https://doi.org/10.1021/acs.jnatprod.9b01121>.
66. Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* 78:569–603. <https://doi.org/10.1146/annurev.biochem.78.072407.102340>.
67. van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernández JM, Jiménez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Morán F, Moya A. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci U S A* 100:581–586. <https://doi.org/10.1073/pnas.0235981100>.
68. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, Thomas BC, Banfield JF. 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *mBio* 4:e00708-13. <https://doi.org/10.1128/mBio.00708-13>.
69. Neuenschwander SM, Ghai R, Pernthaler J, Salcher MM. 2018. Microdiversification in genome-streamlined ubiquitous freshwater *Actinobacteria*. *ISME J* 12:185–198. <https://doi.org/10.1038/ismej.2017.156>.
70. Carlucci AF, Bowes PM. 1970. Vitamin production and utilization by phytoplankton in mixed culture. *J Phycol* 6:393–400. <https://doi.org/10.1111/j.1529-8817.1970.tb02413.x>.
71. Morris JJ, Lenski RE, Zinser ER. 2012. The black queen hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3:e00036-12. <https://doi.org/10.1128/mBio.00036-12>.

72. Fridolfsson E, Lindehoff E, Legrand C, Hylander S. 2018. Thiamin (vitamin B1) content in phytoplankton and zooplankton in the presence of filamentous cyanobacteria. *Limnol Oceanogr* 63:2423–2435. <https://doi.org/10.1002/lno.10949>.
73. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2002. Comparative genomics of thiamin biosynthesis in prokaryotes: new genes and regulatory mechanisms. *J Biol Chem* 277:48949–48959. <https://doi.org/10.1074/jbc.M208965200>.
74. Bertrand EM, Allen AE. 2012. Influence of vitamin B auxotrophy on nitrogen metabolism in eukaryotic phytoplankton. *Front Microbiol* 3:375. <https://doi.org/10.3389/fmicb.2012.00375>.
75. Tang YZ, Koch F, Gobler CJ. 2010. Most harmful algal bloom species are vitamin B1 and B12 auxotrophs. *Proc Natl Acad Sci U S A* 107:20756–20761. <https://doi.org/10.1073/pnas.1009566107>.
76. Hagemann M. 2011. Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol Rev* 35:87–123. <https://doi.org/10.1111/j.1574-6976.2010.00234.x>.
77. Hagemann M, Marin K. 1999. Salt-induced sucrose accumulation is mediated by sucrose-phosphate-synthase in cyanobacteria. *J Plant Physiol* 155:424–430. [https://doi.org/10.1016/S0176-1617\(99\)80126-6](https://doi.org/10.1016/S0176-1617(99)80126-6).
78. Liang Y, Zhang M, Wang M, Zhang W, Qiao C, Luo Q, Lu X. 2020. Freshwater cyanobacterium *Synechococcus elongatus* PCC 7942 adapts to an environment with salt stress via ion-induced enzymatic balance of compatible solutes. *Appl Environ Microbiol* 86:e02904-19. <https://doi.org/10.1128/AEM.02904-19>.
79. Ball S, Colleoni C, Arias MC. 2015. The transition from glycogen to starch metabolism in cyanobacteria and eukaryotes, p 93–158. In Nakamura Y (ed), *Starch: metabolism and structure*. Springer, Tokyo, Japan.
80. Shishido TK, Jokela J, Kolehmainen C-T, Fewer DP, Wahlsten M, Wang H, Rouhainen L, Rizzi E, Bellis GD, Permi P, Sivonen K. 2015. Antifungal activity improved by coproduction of cyclodextrins and anabaenolysins in *Cyanobacteria*. *Proc Natl Acad Sci U S A* 112:13669–13674. <https://doi.org/10.1073/pnas.1510432112>.
81. Stincone A, Prigione A, Cramer T, Wamelink MMC, Campbell K, Cheung E, Olin-Sandoval V, Grüning N-M, Krüger A, Alam MT, Keller MA, Breitenbach M, Brindle KM, Rabinowitz JD, Ralser M. 2015. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol Rev Camb Philos Soc* 90:927–963. <https://doi.org/10.1111/brv.12140>.
82. Corkins ME, Wilson S, Cocuron J-C, Alonso AP, Bird AJ. 2017. The gluconate shunt is an alternative route for directing glucose into the pentose phosphate pathway in fission yeast. *J Biol Chem* 292:13823–13832. <https://doi.org/10.1074/jbc.M117.798488>.
83. Stanton RC. 2012. Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *IUBMB Life* 64:362–369. <https://doi.org/10.1002/iub.1017>.
84. Watzler B, Forchhammer K. 2018. Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Appl Environ Microbiol* 84:e01298-18. <https://doi.org/10.1128/AEM.01298-18>.
85. Fichman Y, Gerdes SY, Kovács H, Szabados L, Zilberstein A, Csonka LN. 2015. Evolution of proline biosynthesis: enzymology, bioinformatics, genetics, and transcriptional regulation. *Biol Rev Camb Philos Soc* 90:1065–1099. <https://doi.org/10.1111/brv.12146>.
86. Walker MC, van der Donk WA. 2016. The many roles of glutamate in metabolism. *J Ind Microbiol Biotechnol* 43:419–430. <https://doi.org/10.1007/s10295-015-1665-y>.
87. Benoni R, Beck CM, Garza-Sánchez F, Bettati S, Mozzarelli A, Hayes CS, Campanini B. 2017. Activation of an anti-bacterial toxin by the biosynthetic enzyme CysK: mechanism of binding, interaction specificity and competition with cysteine synthase. *Sci Rep* 7:8817. <https://doi.org/10.1038/s41598-017-09022-6>.
88. Adrio J-L, Demain AL. 2010. Recombinant organisms for production of industrial products. *Bioeng Bugs* 1:116–131. <https://doi.org/10.4161/bbug.1.2.10484>.
89. Kondo F, Ikai Y, Oka H, Okumura M, Ishikawa N, Harada K, Matsuura K, Murata H, Suzuki M. 1992. Formation, characterization, and toxicity of the glutathione and cysteine conjugates of toxic heptapeptide microcystins. *Chem Res Toxicol* 5:591–596. <https://doi.org/10.1021/tx00029a002>.
90. Zilliges Y, Kehr J-C, Meissner S, Ishida K, Mikkat S, Hagemann M, Kaplan A, Börner T, Dittmann E. 2011. The cyanobacterial hepatotoxin microcystin binds to proteins and increases the fitness of *Microcystis* under oxidative stress conditions. *PLoS One* 6:e17615. <https://doi.org/10.1371/journal.pone.0017615>.
91. Bolch CJS, Blackburn SI. 1996. Isolation and purification of Australian isolates of the toxic cyanobacterium *Microcystis aeruginosa* Kütz. *J Appl Phycol* 8:5–13. <https://doi.org/10.1007/BF02186215>.
92. NCBI Resource Coordinators. 2018. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 46:D8–D13. <https://doi.org/10.1093/nar/gkv1290>.
93. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
94. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. <https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner>.
95. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
96. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. <https://doi.org/10.7717/peerj.1165>.
97. 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>.
98. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>.
99. 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>.
100. 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>.
101. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, van der Maaten L, Vlassis N, Wilmes P. 2015. VizBin: an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* 3:1. <https://doi.org/10.1186/s40168-014-0066-1>.
102. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
103. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* 12:R44. <https://doi.org/10.1186/gb-2011-12-5-r44>.
104. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:590–596. <https://doi.org/10.1093/nar/gks1219>.
105. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
106. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157. <https://doi.org/10.1186/s13059-015-0721-2>.
107. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36:2251–2252. <https://doi.org/10.1093/bioinformatics/btz859>.
108. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0: improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.
109. Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33:3340–3347. <https://doi.org/10.1093/bioinformatics/btx433>.
110. Vernikos GS, Parkhill J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella

- pathogenicity islands. *Bioinformatics* 22:2196–2203. <https://doi.org/10.1093/bioinformatics/btl369>.
111. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
 112. Seemann T. 2015. Snippy. <https://github.com/tseemann/snippy>.
 113. Vieira-Silva S, Rocha EPC. 2010. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet* 6:e1000808. <https://doi.org/10.1371/journal.pgen.1000808>.
 114. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. <https://doi.org/10.1186/1471-2164-12-402>.
 115. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. <https://doi.org/10.1186/1471-2105-14-60>.
 116. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81–91. <https://doi.org/10.1099/ijs.0.64483-0>.
 117. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
 118. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>.
 119. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
 120. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
 121. R Core Team. 2020. R: a language and environment for statistical computing. <https://www.R-project.org>.
 122. Alboukadel K. 2020. ggpubr: “ggplot2” based publication ready plots. <https://cran.r-project.org/web/packages/ggpubr>.
 123. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer, New York.
 124. Raivo K. 2019. pheatmap: pretty heatmaps. <https://cran.r-project.org/web/packages/pheatmap>.