Metabolite diversity of Microcystis strains shows tight correspondence to genotype and may contribute to ecotype specificities

Aurore Huré, Maiwenn Le Meur, Charlotte Duval, Jean-Pierre Bouly, Lou Mary, Manon Quiquand, Michella Dawra, Muriel Gugger, Sébastien Halary, Benjamin Marie

communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-026-09599-7

Metabolite diversity of *Microcystis* strains shows tight correspondence to genotype and may contribute to ecotype specificities

Aurore Huré[1,2], Maiwenn Le Meur[1], Charlotte Duval[1], Jean-Pierre Bouly[1,3], Lou Mary[1,4], Manon Quiquand[1], Michella Dawra[1], Muriel Gugger[5], Sébastien Halary[1,6] ✉ & Benjamin Marie[1,6] ✉

Microcystis is one of the most common bloom-forming cyanobacteria colonizing freshwater ecosystems worldwide. This genus remarkably produces numerous bio-active accessory metabolites, which are believed to be potentially involved in different ecological and/or physiological processes. However, their genuine contribution to the evolutionary success of *Microcystis* blooms remains undetermined. To better depict the potential relationship between the local genetic diversity of blooming *Microcystis* populations and their respective associated chemical diversity, we conducted a joint genomic and metabolomic analysis of 65 *Microcystis* strains collected from various lakes in France and surrounding Western European countries. Interestingly, both core and pan-gene phylogenetic analysis place 57 of these strains in 11 distinct genotypes with at least 2 genomes, being widely distributed along the entire *Microcystis* phylogeny and presenting specific signatures of accessory metabolite biosynthesis. The direct chemical analysis of metabolite diversity produced by these strains, cultured under laboratory conditions, reveals the production of stable metabolite cocktails, with minimal variations over replication, growth phases and culture conditions. Remarkably, these strains belonging to 11 different genotypes correspond to 13 distinct chemotypes according to an accurate one-chemotype-for-one-genotype rule. Furthermore, these genotypes also appear distinguishable regarding their respective ecotoxicological traits and might be considered as specific toxico-ecotypes. Overall, our investigations reveal that the production of accessory metabolites constitute well conserved chemical traits across the different *Microcystis* genotypes, suggesting these molecules may be involved in key adaptive and selective processes, that still remain under-explored.

The inter- and intra-specific differentiation within micro-organisms has recently been revealed by comparative genomics, suggesting that lineage-specific genetic variation significantly contributes to adaptive diversity[1] Phylogenomic investigations based on complete genome analysis including core and pan-genes may reflect both evolutionary and adaptive traits and provide valuable insights for delimiting clades and ecotypes[2] The accessory genes shared within clades, defining the boundaries between them, constitute genetic micro-diversity of microbes. Identifying these gene sets might provide insights into ecotype adaptation to diverse local environments, supporting the diversification of microbial

[1]Molécules de Communication et Adaptation des Micro-organismes (MCAM), équipe "Cyanobactéries, Cyanotoxines et Environnement", UMR 7245, CNRS, MNHN, Muséum National d'Histoire Naturelle, RDC Bâtiment de Cryptogamie - CP 39, Paris, France. [2]UMR-I 02 SEBIO, Université de Reims-Champagne-Ardennes, Reims, France. [3]UFR 927, Sorbonne Université, Paris, France. [4]Institut Systématique Evolution Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, Sorbonne Université, CNRS, EPHE, Paris, France. [5]Collection des Cyanobactéries, Institut Pasteur, Université Paris-Cité, Paris, France. [6]These authors contributed equally: Sébastien Halary, Benjamin Marie. ✉e-mail: sebastien.halary@mnhn.fr; benjamin.marie@mnhn.fr

species[3,4]Harmful algal blooms caused by cyanobacteria are global ecological problems, but they also present an opportunity to investigate how certain proliferating cyanobacteria, such as *Microcystis*, can sustainably colonize a large range of environmental planktonic niches worldwide[5] It is likely that this ubiquity is based on a considerable diversity supporting ecological plasticity[6] *Microcystis* demonstrates remarkable ecological adaptability supporting efficient primary production, competition with other species, and resilience to various environmental constraints[5] This ecological strategy enables *Microcystis* to dominate the phytoplankton communities in freshwater lakes by forming intense blooms which outcompete neighboring phototroph diversity[7] across a wide range of physicochemical conditions in lakes globally and within lakes over time and seasonal changes[8,9] As a result, *Microcystis* is often studied as a model for genetic evolution and niche adaptation[10].

The clade *Microcystis* is characterized by coccoid unicellular cells ranging from 1 to 9 µm in diameter, exhibiting a variety of colonial morphologies used to differentiate a dozen morpho-species observed so far in natural conditions[11]. Over decades, numerous *Microcystis* isolates have been obtained from diverse aquatic environments worldwide, characterized for their microcystin production and genetically sequenced. This genomic investigation effort has revealed a wide intra-genera genetic diversity[12–15], challenging previous morpho-species distinctions[15,16]. Since then, fully sequenced genomes of *Microcystis* strains have been progressively providing unprecedented taxonomic resolution and have described a distinct phylogenetic picture compare to those previously identified through morphology-based taxonomy[17–21]. These phylogenomic analyses have contested the coherence of *Microcystis* morphospecies and now call for a clear unification of a single globally distributed *Microcystis* species complex[22].

Despite high sequence identity in their core genomes, *Microcystis* genomes exhibit a great deal of genotypic diversity, with sizes ranging from above 3.5 to 6 Mb, and a variable fraction of accessory gene repertoires[5,10,18,23,24]. Phylogenomic analysis of *Microcystis* has allowed for the distinction of various phylogenetic clades[22], but has not clearly segregated strains based on their habitats in terms of biogeography or ecological niches[6,23,25].

Overall, like many other colonial cyanobacteria, *Microcystis* exhibits complex genomes containing numerous large biosynthetic gene clusters (BGCs) that are involved in the production of diverse arsenals of accessory metabolites[18,26,27]. The content of these BGC sets, which belong to the flexible genome part, varies greatly among strains and can represent up to 7% of the total *Microcystis* genome length[22]. These biosynthetic arsenals are assumed to have a high cellular energetic cost due to gene replication and biosynthetic enzyme complex production[28]. While some of these metabolites, like microcystins, are toxic to certain predators or competitors, they may also play a role in allelopathic or nutritive interactions, as well as in physiological adaptation to support photosynthetic or metabolic capabilities[29–32]. The potential functions of cyanobacterial secondary metabolites are diverse and include nitrogen storage[33], UV protection[34], metal chelation[35], and interactions with other organisms[36]. Research has focused on characterizing the involvement of cyanobacteria metabolites in various biotic interactions, such as defense against predation, allelopathy, quorum sensing, parasitism or symbiosis[37–40]. However, the processes behind the evolution of accessory metabolite production in cyanobacteria remain a puzzling question. It is possible that the production of these specialized metabolites could be related to specific habitats and adaptative traits.

Environmental sampling has shown a wide variety of chemical structures of secondary/accessory metabolites present during *Microcystis* blooms. However, the relationship between this chemical diversity and the abundance of different *Microcystis* genotypes within a bloom is not fully understood[25,41–43]. The extent to which genomic variability within *Microcystis* clades influences metabolite composition and defines ecotype-specific features is currently underexplored.

Some studies have explored BGCs in *Microcystis* genomes[18,19,22,23,44], while others have directly characterized the metabolites produced by *Microcystis* strains, including peptides like microcystins, aeruginosins, cyanopeptolins, anabaenopeptins, microginins, aerucyclamides, and microviridins[26,27,45]. Combining metabolomic profiling with genome mining analyses has been a powerful approach for characterizing the production of accessory metabolites in different cyanobacteria[46,47]. However, there is still much to learn, as some BGCs in *Microcystis* remain uncharacterized[48,49]. Initial investigations suggest a stochastic distribution of metabolite biosynthesis among different strains[18], but our present systematic analysis now shows that metabolite production constitutes a key conserved characteristic of different *Microcystis* clades.

Results

By developing microbial reverse ecology[50] based on genomic, metabolomic and ecotoxicological investigations of an unexplored aspect of *Microcystis* diversity, our present work offers new understandings of potentially toxic bloom-forming cyanobacteria (Figs. S1–S3). This original approach clearly allows us to explore the ecological functional aspect of this diversity within the *Microcystis* tree of life, and advances the exploration of the role of chemical diversity and related ecological traits within *Microcystis* clade evolution.

Microcystis phylogenomic, genospecies, and genotype distinction

Given the challenge associated with obtaining reliable information on the colony morphology of *Microcystis* strains and the discrepancies between morpho-species and phylogenetic inferences, we have opted to rely solely on the classification derived from whole-genomes studies as outlined by Cai and co-wokers[22]. Initially, we reconstructed a phylogenomic tree using 791 orthologous genes from the core-genome of *Microcystis* (347 genomes, see Figs. 1 and S4). At the time, only four *Microcystis* genomes from Western European strains were accessible, while most of the samples originated from North and South America or Asia[22,51]. However, the available 16S-ITS sequence suggested a potential underestimation of European *Microcystis* diversity[5]. Our sequencing efforts have revealed the extensive diversity among the investigation of 65 strains originating from France and surrounding countries, which are largely distributed across the phylogenomic tree (Fig. 1). To further enhance their position, we determined the potential genospecies to which the constituent subclades belong. Following the commonly accepted definition of prokaryotic genospecies, wherein a monophyletic group shares >94–95% Average Nucleotide Identity (ANI)[52], *Microcystis* genomes in our dataset belong to a single genomospecies. Stricter ANI thresholds have successfully delineated clades in alignment with morphological data and recombination rates[22,23,53]. By applying an ANI threshold of >97%, we identified that these 347 *Microcystis* strains belong to 21 genospecies (each with at least 2 strains), denoted from A to U, with member counts ranging from 2 (e.g., B or H) to 69 (e.g., U). However, the delineation of these *Microcystis* genospecies (based on ANI > 97%) does not rely on ecological or physiological knowledge. They may not be entirely relevant or operationally effective[52,54], given that gene flow barriers may enable the distinction of several genetic clades, whereas the characterization of their biological specificities requires supplementary exploration. To further refine the phylogenetic delineation of the 65 Western European strains, we grouped those sharing an average identity of >99% together (Supplementary Fig. S5A, B). Out of the 65 strains, 57 were clustered into 11 distinct genetic clades, each comprising at least two strains, labeled as genotypes from d1 to t1, with correspondence with the respective genospecies they belong to. The remaining eight strains were found to be closely related to reference strains from other continents and were classified as singletons (see Fig. 1).

Microcystis diversity from France and surrounding countries and local genotype richness

Our investigation of various Western European *Microcystis* strains (mostly from France) shows that most strains belong to genospecies already observed in other geographic areas. The majority of *Microcystis* genotypes have a general ubiquitous distribution, as previously observed[55]. For

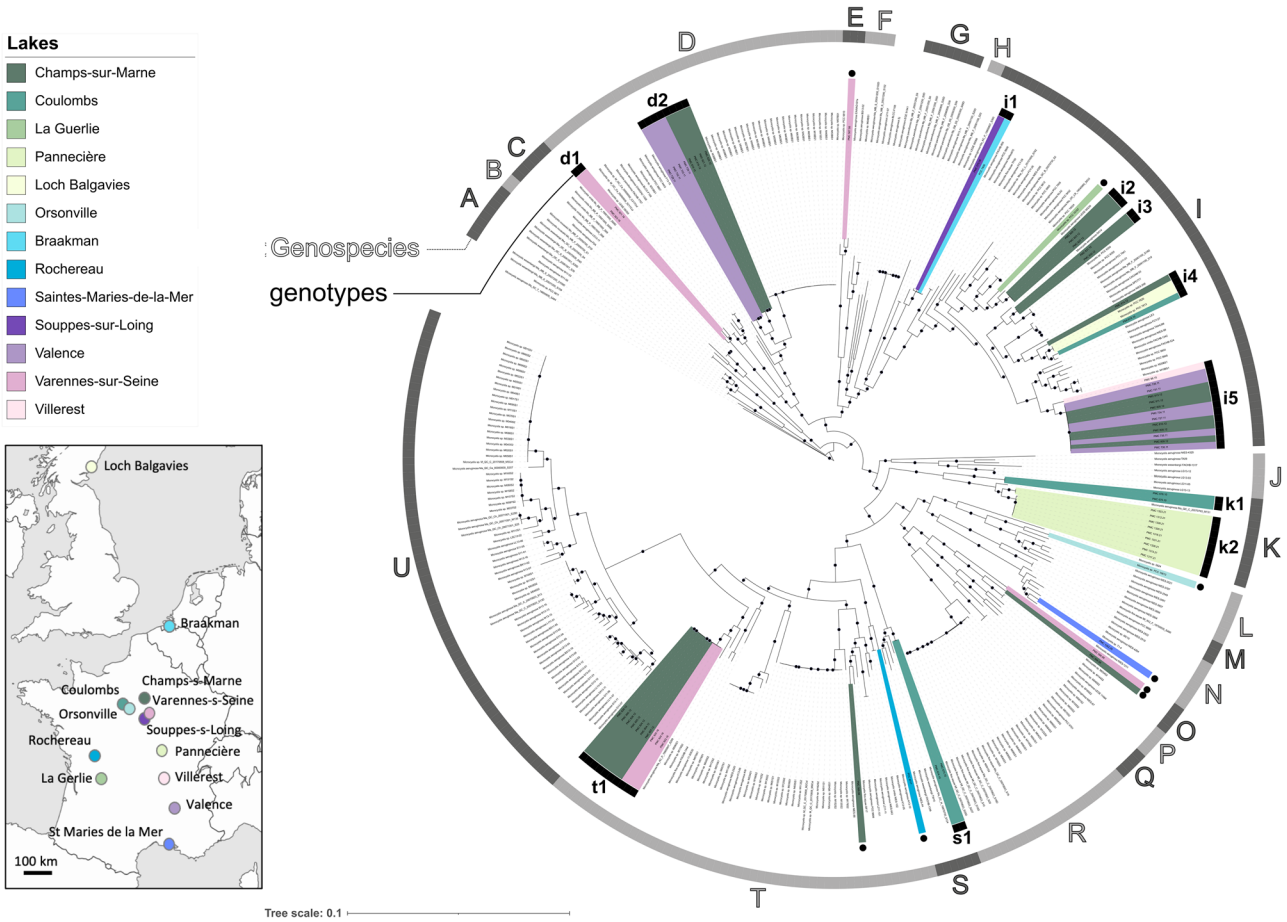

Fig. 1 | Rooted maximum likelihood phylogenetic tree inferred for 347 *Microcystis* genomes using the concatenated alignment of 791 single-copy core genes. A total of 686,669 nucleotide positions were used with *M. aeruginosa* Ma_SC_T_19800800_S464 used as a *Microcystis* roost, as identified in a previous study[127]. The colored shadow pies indicate the geographic sublocation of the genomes of all 65 Western European strains collected from PMC and PCC (see Supplementary Data S1 for details); refer to the color code on the figure. The outer circle indicates the genospecies to which each strain belongs, labeled from A to U. The inner circle indicates the genotype of each of the Western European strains studied. A green circle marks the singletons. Tree branches supported by bootstrap values over 60% are indicated by a black circle.

example, strains from genotype I isolated from Lakes Varennes-sur-Seine, Villerest and Champs-sur-Marne, have the closest *Microcystis* strains (e.g., NIES 98, PCC 9804, and M19BS1) originating from Asia, Australia and South America, respectively (Fig. 1 and Supplementary Data S1). Interestingly, the specific genotype referred to as genotype k2 was the only one observed among the nine strains collected in Lake Pannecière, resulting in the lowest diversity score among lakes (Fig. 2). This observation suggests that this particular genotype may present a remarkable ability to colonize this lake at this sampling time, possibly due to local adaptations or may alternatively result from ongoing primo-colonization. Tromas and coworkers[56] have suggested that genetic distance between *Microcystis* strains tends to increase with geographic distance at the American continent scale, but this assumption was not supported by our dataset (Supplementary Fig. S5C, $R^2 = 0.0045$), as in other global studies based on ITS or 16S rRNA sequences[55,57]. Although the sampling effort was not uniform in each lake (from 1 to 25 strains), we observed that the level of sampled strain diversity varies considerably among certain lakes. While some environments have low diversity scores, others exhibit much higher genotype diversity with up to 7 different genotypes observed for 25 strains collected in Lake Champs-sur-Marne during August or October 2011 (Fig. 2 and Supplementary Data S1), resulting in a diversity score of 2.56. Based on the analysis of the isolated strains, at least five of the studied lakes hosted populations consisting of several genotypes, confirming that a blooming Cyanobacteria population can be made up of complex consortia of different lineages[10,20,23,58].

Marine picocyanobacteria *Synechococcus* and *Prochlorococcus* have a continuous genotypic distribution worldwide[59], along a gradient of environmental conditions such as light, temperature, salinity, and nutrient availability[60]. In contrast, continental freshwater cyanobacteria, such as *Microcystis*, are distributed among patchy ecosystems that are far apart and vary in environmental conditions not necessarily correlated with the distance between them. This might explain why, as various other freshwater Cyanobacteria[61], *Microcystis* exhibits much larger genomes (ranging in this study from 3.8 to 5.9 Mb), associated with a greater number of potentially functional genes enabling each genotype to adapt to a wider range of environmental conditions.

Gene content supports *Microcystis* genome diversity

At a genospecies level, the clustering based on core-genome analysis is in complete agreement with that derived from flexible gene content analysis (Fig. 3). However, contrasting results emerge when considering secondary metabolite biosynthetic gene clusters (BGCs). All BGCs reported at this point, as corresponding to already described BGC on Mibig 3.0 database[62], are part of the surveyed *Microcystis* accessory genome and represent an average of over 7% of their total genome. Several BGC types were consistently identified by AntiSMASH, although their distribution varied across genospecies, but was remarkably conserved within the same genotypes (Fig. 4 and Supplementary Data S2A–P). Overall, this comparison of BGC content retrieved from automatic genome search with direct metabolite

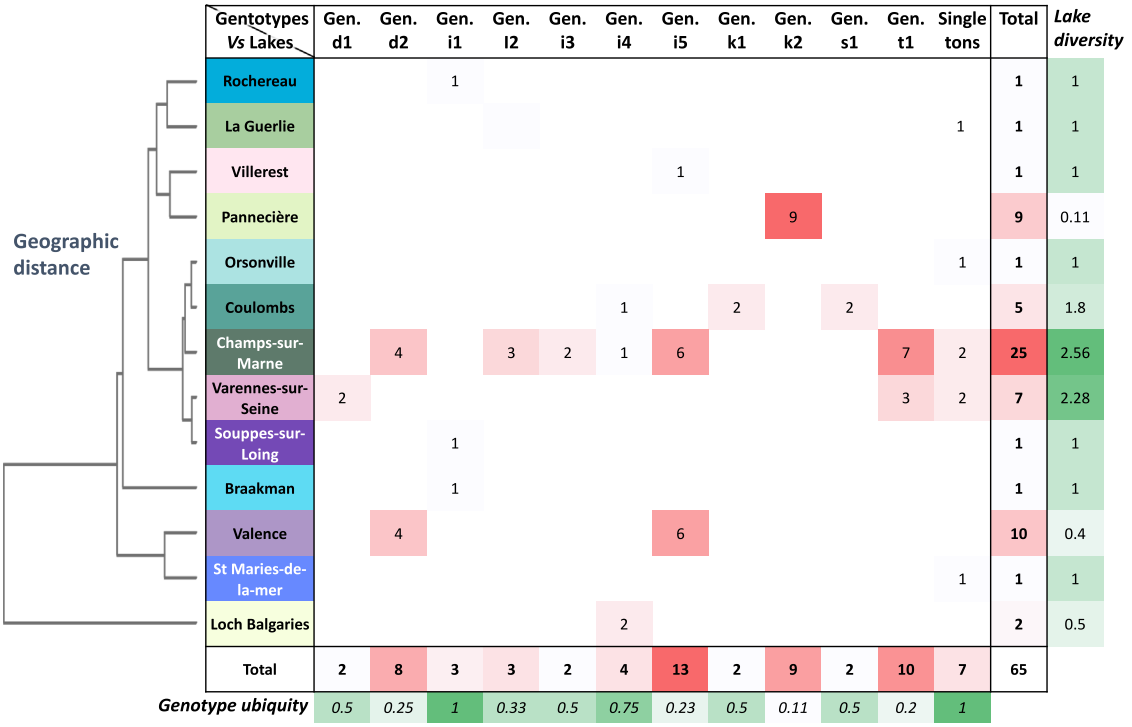

Fig. 2 | Geographical distance tree of strain lake collection and corresponding genotype location matrix. Despite the variation in strain sampling effort across the lakes (ranging from 1 to 25 strains), it is still possible to compare local and genotype diversity patterns. The custom diversity index for each lake is calculated by dividing the square of the number of genotypes per lake by the number of strains per lake. Likewise, the custom ubiquity index for each genotype is determined by dividing the number of lakes from which each genotype was collected by the number of strains within each genotype.

detection based on LC-MS/MS investigation confirms the conserved production of various metabolite families within the different genotypes, but also illustrates that the genome mining approach was not equally efficient in predicting all BGC families. For example, the shinorine/MAA BGC was often annotated as an unknown LAP (RiPP), illustrating the necessity of implementing the MiBig reference database to increase efficiency for *Microcystis* BGC annotation. These observations also highlight the global limitations of the automatic annotation of RiPP-BGC products – based on leader peptide and accessory enzyme recognition - as the high structural diversity generated by limited mutation events on pro-peptide fell to be established by LC-MSMS, while NRPS, NRPS/PKS, and PKS products offer more automatic annotations according to the GNPS-based metabolomic pipeline. Among the 17 distinct BGCs belonging to four BGC types (NPS, PKS, NRPS/PKS, and RiPP), the aeruginosin cluster is the most widespread, present in 80% of the studied genomes, except for genotypes k1, k2, and s1. In contrast, microcystin BGCs present a distinct pattern, being present in the genomes of Western European strains from various genotypes (namely d1, d2, i1, i3, i4, and t1) widely spread within the *Microcystis* cladogram (Fig. 4).

While the BGC presence for NRPS/PKS and/or their corresponding metabolite detection are highly conserved within genotypes, but not into the genomo-species, some RiPP-BGCs present intra-genotype variations (microcylamide in i5, aeruginosamide in k2 and piricyclamide and microviridin in t1, Fig. 4). This confirms the previous observation that BGCs can be acquired and lost within genospecies[23] and suggests that the genotype (supported by <99%-ANI scores) represent a better operational taxonomic unit to consider than the genospecies (supported by <97%-ANI scores) in regards to the metabolic content specificities of the different level of *Microcystis* clade distinction (Fig. 4). Overall, these results illustrate the large variation observed between the different Western European *Microcystis* strains from France and surrounding countries (ranging from 3.8 to 5.9 Mb), together with an overall homogeneity within the distinct genetic clades. The accessory metabolites of some genotypes are highly similar despite the diversity of the strain origins (e.g., genotype i5 containing isolated strains from north and south of France), while chemotypes within certain genotypes appeared much more diversified than between certain others (e.g., within I compared to between D and T; Supplementary Fig. S6). It is also worth noticing that although the two neighboring strains belonging to the genospecies P (namely PMC 566.08 and PMC 568.08) did not constitute a genuine genotype (their ANI score being above 97%, but below 99% thresholds), the metabolomes of these two strains were grouped together within a single chemotype (see Fig. 5).

Interestingly, a meta-analysis of prokaryote metagenomes reveals that the majority of BGCs (over 75%) are biome specific, suggesting that co-occurring genotypes with divergent BGC contents could constitute ecotypes adapted to distinct niches within the environment[63]. Additionally, previous studies have observed positive correlations between the genome size of prokaryotes and subsequent accessory gene contents, on one hand, and environmental nutrient concentration and subsequent biomass productivity, on the other[64]. The bloom-forming *Microcystis* with its relatively large genome and rich BGC content seems to follow a similar trend[18]. Conversely, less proliferative and small genome owning clades, such as marine picocyanobacteria, present low BGC contents, restricted to the production of carotenoid accessory pigments[61]. So far, several of the corresponding accessory compounds produced by the different BGCs have been proposed to provide various ecotoxicological benefits[29], but their genuine contribution to the ecological distinctiveness among *Microcystis* subclades remains undetermined.

Does the specific production of accessory metabolites by *Microcystis* subclades constitute adaptative traits?

Our extended LC-MS-based metabolomic analyses clearly indicate that Western European *Microcystis* strains (i.e., from France and surrounding countries) belonging to different genetic clades produce specific and distinct metabolite corteges, representing characteristic chemotypes even when cultured under standard laboratory conditions (Fig. 5). We observed that strain clusters discriminated at the genomic level considering either core-

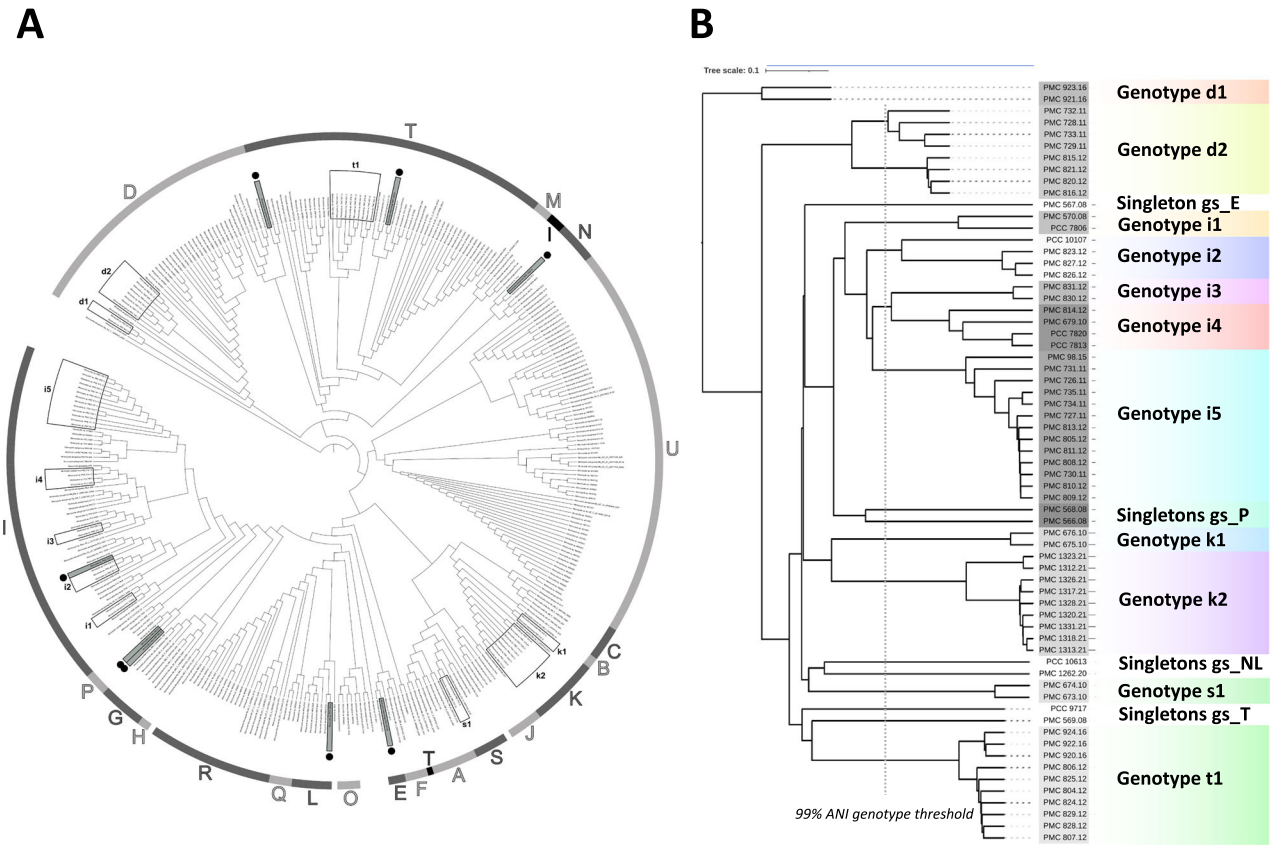

Fig. 3 | Hierarchical clustering dendrogram based on dissimilarity among gene content for each *Microcystis* strain. Cladogram based on dissimilarity matrix calculated as Jaccard distances inferred from the gene presence/absence matrix for the whole set of *Microcystis* pangenomes (**A**). The outer circle indicates the genospecies to which each strain belongs as described in Fig. 1 (97%-ANI threshold score). European strains from the same genotype are surrounded by a black frame. Singletons are grayed out and marked with a black circle. Overall correspondence between pan-genome Jaccard distances (left) and genotypes determined according to 99%-ANI threshold score (right) for the 65 Western European *Microcystis* strains only (**B**), regarding the phylogenomy performed on the core-gene set as shown in Fig. 1. Genomospecies and genotypes are named according to phylogenomy observed on Fig. 1, and genotype colors corresponding to Fig. 4.

and pan-genome or BGC contents strictly correspond to characteristic chemotypes producing well-determined metabolite categories. It is worth noticing that this chemotype distinction was remarkably identical to those retrieved from the merged data matrix obtained by combining both analyses performed on positive and negative MS mode (Supplementary Fig. S7) and that the chemotype belonging of strains was largely stable across successive cultures (Supplementary Fig. S8). Only minor variations in terms of molecular variants and/or relative metabolite quantities can be monitored regarding the biomass of strains cultured under laboratory conditions (Supplementary Fig. S9). The main metabolite families observed in the different chemotypes (in terms of count intensity signal) according to global metabolomic investigation supported by GNPS annotation are indicated on the right of Fig. 6 (Supplementary Data S3). We notice that chemotype 07 (Chem07) comprises two neighbor strains belonging to the genospecies P (namely PMC 566.08 and PMC 568.08), but that did not constitute a genuine genotype as observed in Fig. 1 (their ANI score being over 97%, but below 99% thresholds).

The biosynthesis of cyanobacterial accessory metabolites was shown to be a highly complex process that may be potentially regulated by various endogenous or exogenous factors[65–67]. Nevertheless, the occurrence of a genetic signature revealed by the genome mining strategy indicates the potential for an accessory metabolite production to be further investigated. In addition, we chose to develop a direct metabolomic approach rather than a single BGC-based one, which remains an intermediate level of determination between gene content and genuine functional molecules. Accessory metabolite annotations were thus attempted using molecular networking (Fig. 6) based on MS/MS data (see Supplementary Data S3A–C). Interestingly, several larger clusters contained annotated accessory metabolites such as aeruginosins, cyanopeptolins, microcyclamides, microginins, anabaenopeptins, microcystins, as well as anabaenolysins, biopterins and MAAs, among others. The largest clusters are also illustrated in Supplementary Fig. S10, along with metabolite annotations using GNPS, lake origins and genotypes. Such trait-based approaches offer a new perspective for investigating microbial functional capacity with a more mechanistic understanding. Notably, distinguishable *Microcystis* genetic subclades (based on either core- or pan-genome comparisons; Fig. 3) exhibit homogenous metabolomes, suggesting that this cladistic discrimination may reveal relevant ecological and evolutive chemical traits that remain remarkably conserved within the different clades (Figs. 4–6). These strains belonging to 11 genotypes correspond to 13 distinct chemotypes according to an accurate one-chemotype-for-one-genotype rule (Fig. 7A). Although, strains from chem01 and chem13 belong to the same genotype and present globally similar metabolite composition (comprising microcystins, aeruginosins, cyanopeptolins and microcyclamides) they also present noticeable molecular distinction characterized by variations in variant contents, that appears congruent with the respective genotypic position of the strains (Fig. 4 and Supplementary Data S2P). While singletons do not describe reliable clades, they also present a similar trend of correspondence between genome and metabolome relationships. For example, the different singletons belonging to geno-species I show distinct genotypes corresponding to distinct chemotypes.

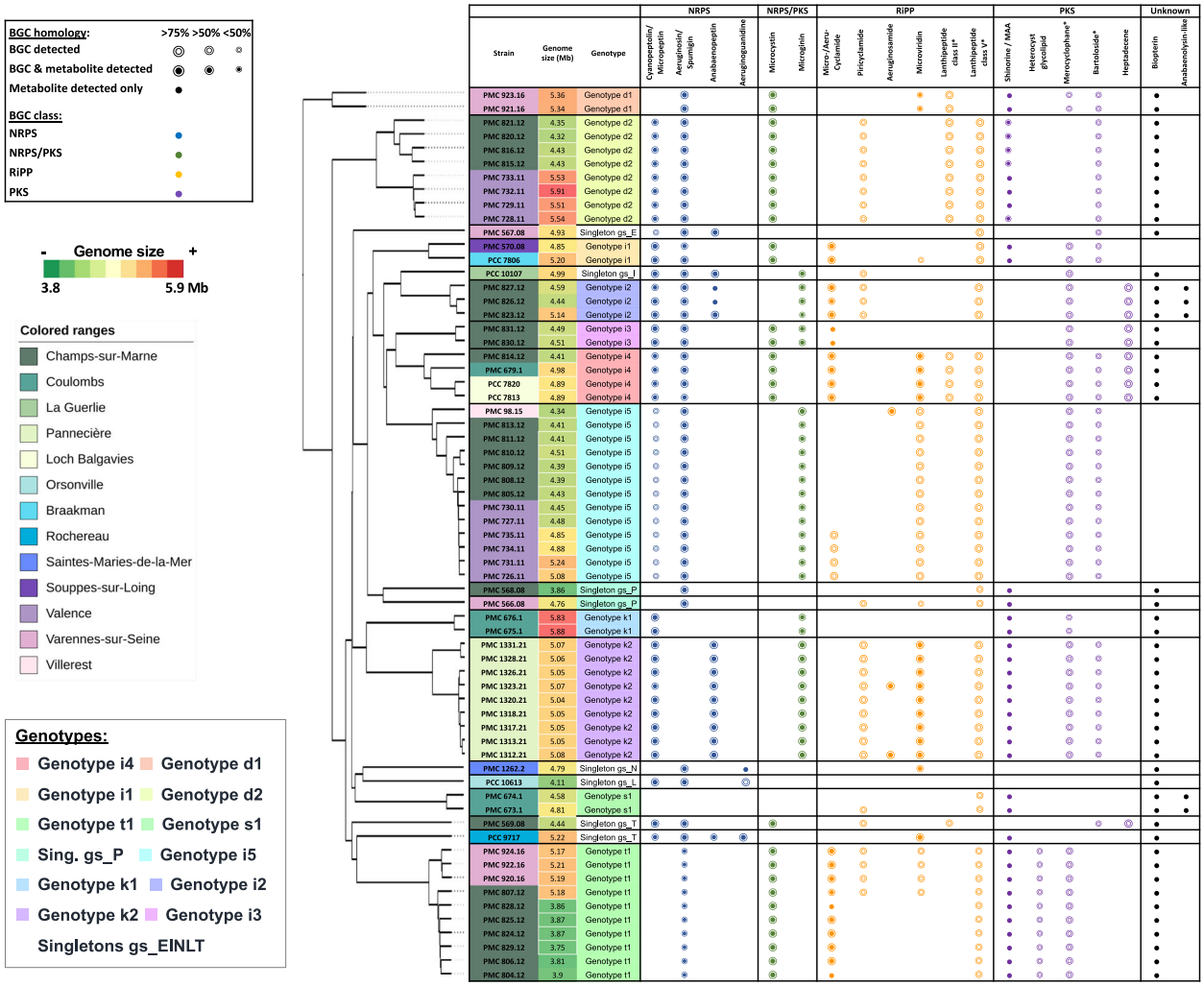

Fig. 4 | Correspondence of pan-genome distance, known BGC and metabolite contents across the 65 Western European *Microcystis* strains. From left to right: 1, dendrogram based on Jaccard's distance according to the respective presence/absence of coding genes; 2, geographical origin of the strains, genome size, and corresponding genomic clade; 3, known BGC and metabolite contents including NRPS, NRPS/PKS, PKS, and RiPPs, together with biopterins and anabaenolysins whose biosynthetic pathways remain undetermined. The terpene, LAP or thiopeptide BGCs were not considered in this analysis. The automatic BGC annotations were manually curated, as cyanobacterial genomic information was split into several contigs, leading to several false positive or low annotation scores. Refer to the color code and legend on the figure for the detection of respective BGCs, metabolites or both (empty double circle = automatic BGC detection using AntiSmash only; double filled circle = double identification by BGC detection and direct LC-MS analysis; simple filled circle = direct LC-MS analysis only). The size of circles is dependent on the percentage of identity provided by the AntiSmash automatic annotation (Supplementary Data S2A–P). Detection of metabolite family was deduced from global metabolomic investigation supported by GNPS annotation (Supplementary Fig. S10–19 and Supplementary Data S2, 3). *indicates that for *Microcystis*, the chemical structure of accessory metabolites produced by both merocyclophane-like, bartoloside-like, and lanthipeptide BGC remains unknown and thus could not be retrieved from the metabolomic analysis of the intracellular extracts of *Microcystis* strain cultures.

In this work, we have undertaken an investigation of the evolution of *Microcystis* strains and their respective ecosystems of origin, through a comparison of their genomic relation with their individual chemical traits (Fig. 7A). We hypothesized that the specific chemical traits of these various genotypes might afford them strong adaptive features likely based on genomic originalities shaped by the conditions of their originating environments[68]. Our analysis of the diversity of accessory metabolites of Western European *Microcystis* strains corroborates the possibility for environmental filtering and ecologically relevant genomic variation between strains[58]. Interestingly, Yancey and co-workers[48] recently developed a similar approach to describe the chemical diversity of 21 xenic *Microcystis* strains isolated from Great Lake Erie (Ohio, USA). They observed that several genotypes were associated with the production of distinct metabolite cocktails, suggesting and highlighting the limited understanding of the determinism of the chemical repertoire of *Microcystis*. While the specific functions of many secondary metabolites are unknown, they might mediate specific responses to other organisms and/or environmental condition variations. Therefore, *Microcystis* accessory metabolites may provide insight into certain ecotypes that may have arisen through environmental adaptation.

We further investigated 12 selected *Microcystis* strains belonging to six distinct genotypes (d2, i2, i3, i5, and k2 from Western European and a pair of strains (PCC 9804 from Australia and NIES 98 from Japan) belonging to the genospecies I for comparison with the three genotypes i) for their respective ecotoxicological impacts on Medaka fish embryos and larvae (Fig. 7B; Supplementary Fig. S20–21 and Supplementary Data S4). Overall, we observed a remarkable similarity in toxicity between the two strains belonging to the same pairs of genotypes. Both strain pairs belonging to genotypes i2 and k2 showed faint toxic effects, while the pairs from genotype d2 and i3, both being microcystin producers, induced lethality in hatched larvae but not in embryos protected by their chorion. In contrast, the strain pair belonging to genotype i5 specifically or to the genospecies I, which does

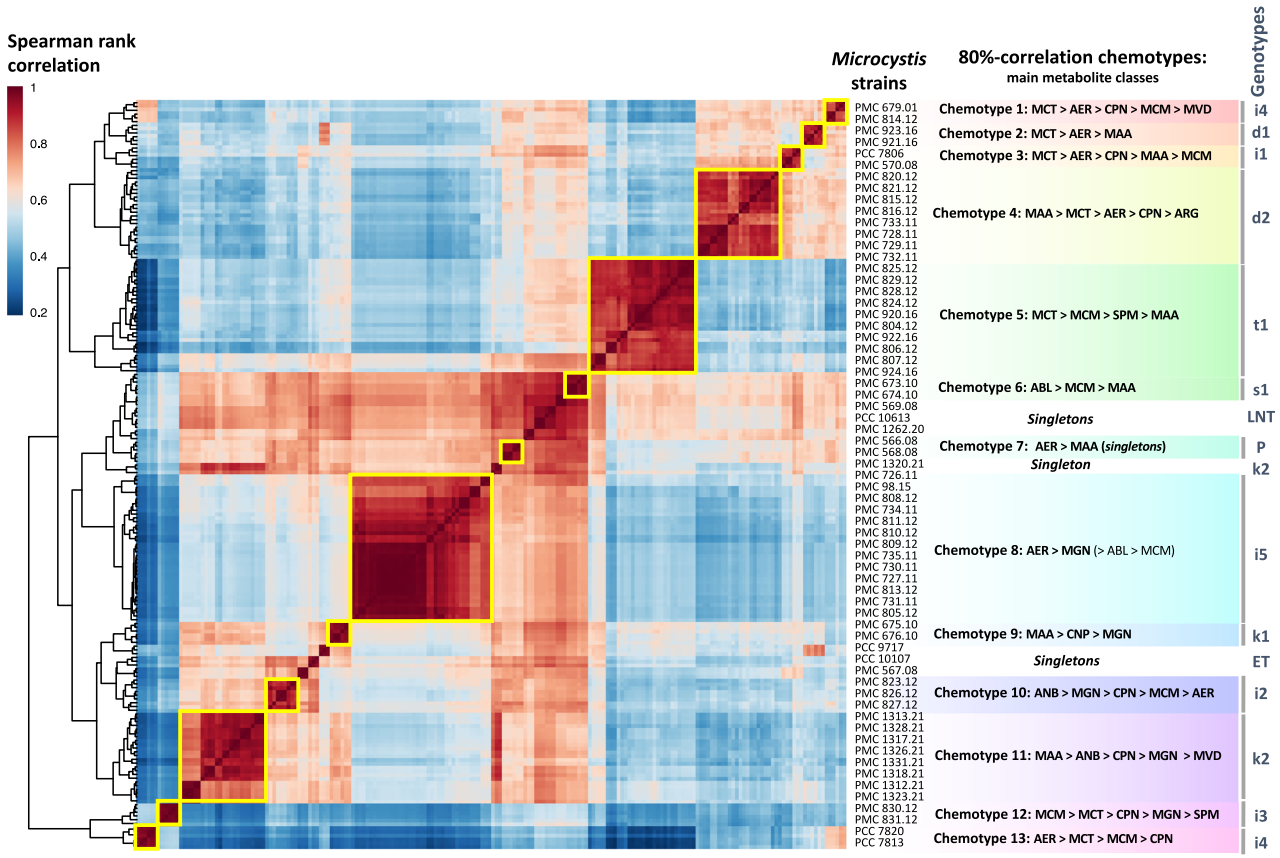

Fig. 5 | Global metabolome correlation heatmap of the 65 strains originating from France and surrounding countries according to Spearman correlation index based on LC-MSMS analyses (50-1500 *m/z*, positive modes in triplicates), combining 3873 MS2 spectra, 250 of which could have been annotated as accessory metabolites. Metabolomes presenting >80% correlation were grouped in distinct chemotypes (highlighted with yellow squares), whereas strain metabolomes that present lower correlation are considered as singletons (see Fig. 6 for color correspondence).

not produce microcystin, but synthesizes several other metabolite families, exhibited high toxicity on both embryos and larvae.

Overall, these observations highlight the conservation of ecotoxicological traits related to distinct genotypes, with respective metabolite contents potentially determining toxico-genotypes. Therefore, the ecotoxicological traits of the *Microcystis* clade may vary depending on the specific genotypes considered, since high variability is observed within *Microcystis* genospecies I[69]. Also, several key abiotic of biotic factors (e.g., nutrient availability, temperature, light intensity, or allelopathy and predation) may influence BGC expression and influence the composition of the *Microcystis* metabolite cocktails in terms of relative and absolute quantities or variant and chemical family ratio, potentially modulating the subsequent ecotoxicological threats that may be observed in both experimental and environmental contexts.

Additionally, the xenic or axenic statue of each individual strain seems presently not to blur the overall chemotype signal of the cultures that remains linked to their respective *Microcystis* genotypes. Regarding the large size difference between heterotroph bacteria of xenic cultures and the cyanobacterial cells (above 1/500-1/5000 in terms of cellular biovolume) and the high secondary production rate of cyanobacterial culture samples under the exponential growth phase (also maximizing the heterotroph/cyanobacteria ratio), the potential xenic influence on *Microcystis* strain chemotype remained presently limited.

Genome evolution within *Microcystis* clades and potential environmental drivers

Investigations of *Microcystis* by global genomics depicted a clade-specific genome relation characterized by highly plastic accessory genomes. Shared core genes consist of less than half (35%) of each genome, with up to 26% of pan genes being unique to a single strain[18,22]. However, little is known about the specific functional differences associated with the *Microcystis* micro-diversity. Functional differences between genotypes can confer significant competitive advantages and niche specialization, potentially supporting *Microcystis* bloom persistence[70]. Although the overall ecological differences between these clades remain speculative, it is assumed that these different monophyletic clades presenting both distinct core- and pan-genomes (including various BGCs among other accessory gene content) may reflect potential ecological adaptations[5,53,63]. Intraspecific genetic diversity may enable a population to adapt to dynamic environmental conditions and biotic interactions[24].

According to the reverse ecology approach, characterizing the environmental conditions of water bodies from which *Microcystis* clades originated may reveal potential ecological drivers favoring the emergence of specific genotypes[50]. In the absence of extensive environmental data to characterize ecotypes, ecological traits could also be inferred from genomic information[23]. Interestingly, Jackrel and co-wokers[10] and latter Kuijpers and co-wokers[71] have correlated global genome reductions and expansions in gene supporting specific metabolic capabilities (i.e., N- and P-related genes) in relation with specific ecological features of their originating lakes (low- or high-nutrient lakes). However, to date, numerous attempts have failed to link *Microcystis* secondary metabolite production, including microcystins, with various abiotic environmental parameters due to their potential health repercussions[72,73].

Potential biological functions of *Microcystis* BGC

In pelagic freshwater environments, several cyanobacteria such as *Microcystis*, *Planktothrix*, *Aphanizomenon*, and *Raphidiopsis* often develop intense blooms living in diversified communities comprising various

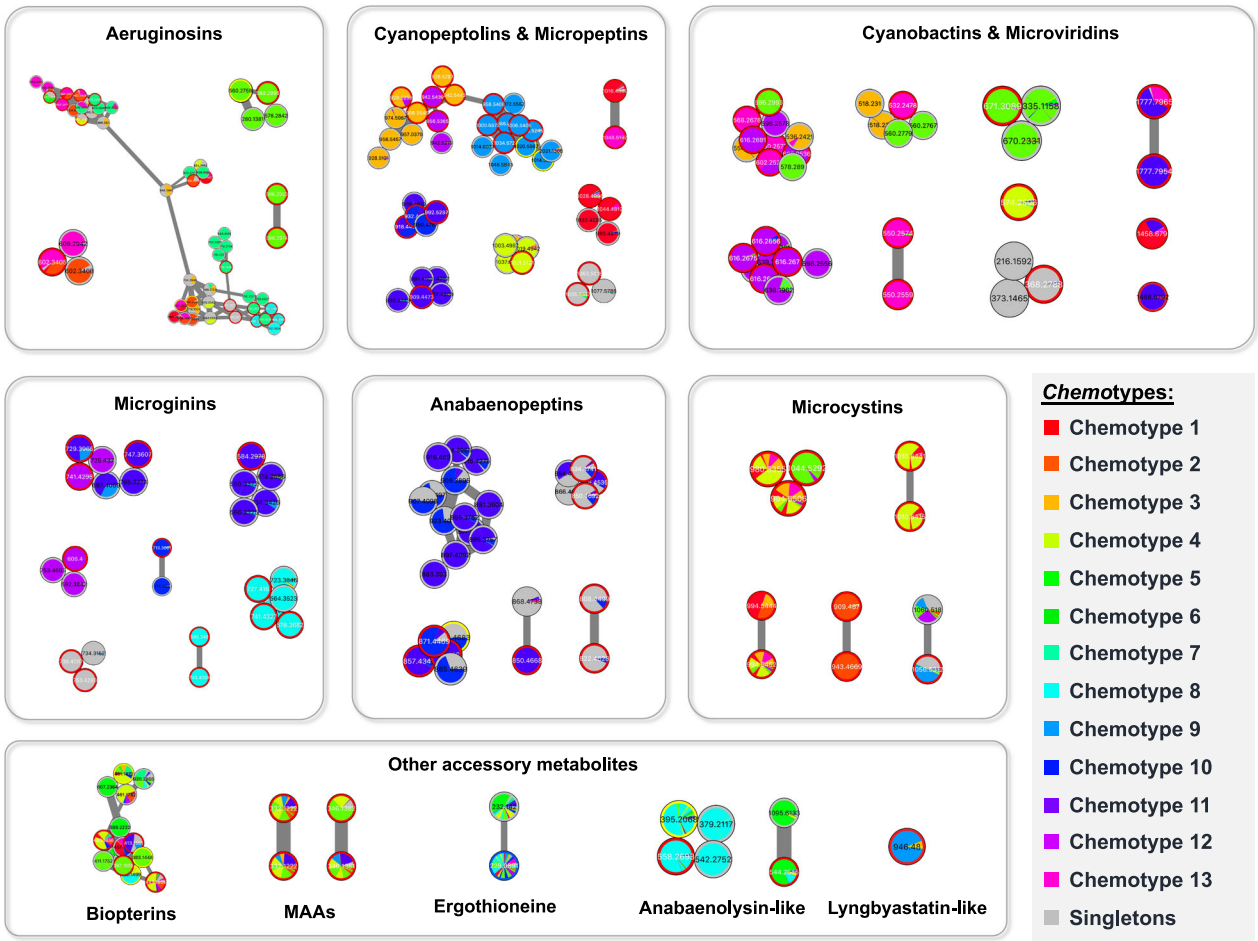

Fig. 6 | Molecular network of primary known metabolite families across the 65 Western European *Microcystis* strains based on their respective chemotypes (see color code of pie chart occurrence within strains in the figure legend) from LC-MS/MS (positive mode). It is worth noticing that a similar pattern was observed for genotype correspondence (Supplementary Fig. S10). The GNPS algorithm compares all MS/MS spectra by aligning them one by one, grouping identical molecules (those with the mass and fragmentation pattern) and assigning a cosine score ranking from 0 to 1 to each alignment. This allows for the reconstruction of the network linking each molecule based on the cosine score calculated between all molecules (with a cosine score significance threshold set at 0.7). Nodes represent molecules, when number inside the nodes and size of the nodes indicate molecular weight and depend on retention time of the molecules, respectively. Metabolite annotations were made using Metaboscape and GNPS automatic annotations based on MS and MS/MS data, respectively (see Supplementary Data S3A–C). Larger clusters were annotated based on the similarity of their fragmentation patterns. Primary metabolite clusters (e.g., amino acids, nucleic acids, lipids, saccharides, etc.) were not represented here. Several clusters contained annotated accessory metabolites such as aeruginosins, cyanopeptolins, microcyclamides, microginins, anabaenopeptins, microcystins, as well as anabaenolysins, biopterins and MAAs, among others. The largest clusters are also shown in Supplementary Fig. S10, along with metabolite annotations using GNPS, lake origins and genotypes. The selection of the principal unknown metabolite cluster is shown in Supplementary Fig. S11. The main clusters corresponding to cyanobacteria-specific accessory metabolites are detailed in Supplementary Figs. S12–19 for aeruginosins, cyanopeptolins/micropeptins, microginins, anabaenopeptins, microcystins, microcyclamides, anabaenolysins, aeruginosamides, microviridins, biopterins, and MAAs, respectively.

species[5,74–76]. They generally contain many biosynthetic gene clusters (BGCs) involved in the production of various bioactive accessory metabolites[58,61,77–80] (including cyclic peptides, lipopeptides, alkaloids, macrolides, and polyketides), that may inhibit a wide variety of prokaryotic and eukaryotic competitors or predators[81], they can also play a role in competition between *Microcystis* strains[82], quorum sensing in nutrient storage[83,84], potentially shaping the genotypic structures of blooming population.

In terms of biosynthetic capacity, *Microcystis* is one of the most prolific cyanobacterial genera[85], producing many cyclic peptides such as aeruginosins, anabaenopeptins, cyanobactins, cyanopeptolins, microginins, microviridins, and microcystins[23,86]. However, the BGC content varies greatly across *Microcystis* clades, leading to substantial diversity in the products observed through metabolomics. While numerous BGCs encode known compounds, many have not been linked to a known product[61], suggesting the potential for the discovery of novel and potentially bioactive metabolite families (Supplementary Fig. S11). Our dataset offers novel opportunities for discovering new BGCs associated within metabolites families. Future effort would now be specifically dedicated for charactering such novel molecules and potentially corresponding BGC. For example, the 2 stains from Chem06 (PMC 673 and 674) both present several uncharacterized RiPP BGCs (presenting faint similarities with PcpA, hglE and lanthipeptide class-v, respectively), together with a singular molecular cluster of 7 unknown metabolites ranging from around 1052–1092 Da, that aim at being further de novo characterized by direct structural analysis (Supplementary Fig. S11 and Supplementary Data S2a, 2G).

Despite over 50 years of intense study, the ecological and physiological functionality of cyanobacteria-specific accessory metabolites, like microcystins, remains unclear[29,85,87]. However, the occurrence pattern of these molecules and their respective BGCs align with core *Microcystis* phylogeny on the scale of sub-clades, indicating a cohesive trait among closely related strains. This suggests that the gain and loss of these genes are relatively rare, contributing to the genotype-conserved signature of each respective clade.

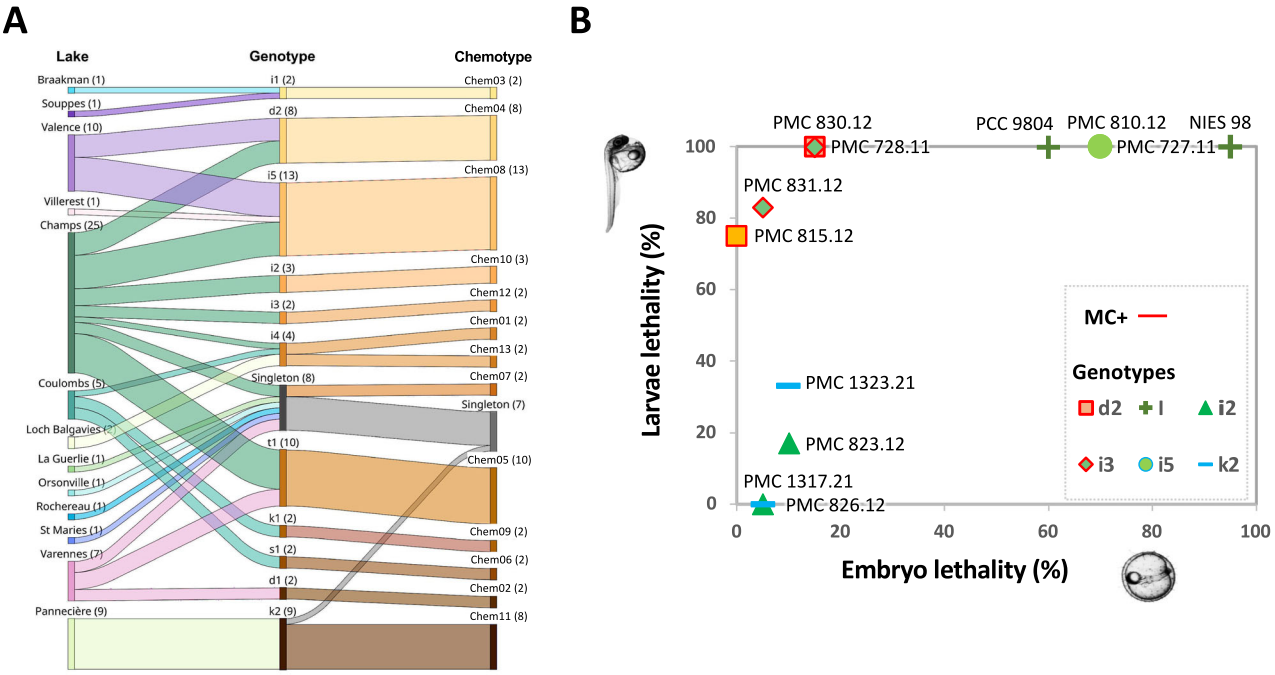

Fig. 7 | Relation between strain origins, genotypes and chemotypes and toxicity associated to 12 strains belonging to 6 distinct genotypes. Relation between strain origins, genotypes and chemotypes (**A**) and toxicity associated to 12 strains belonging to 6 distinct genotypes (**B**). **A** Sankey diagram showing the relationships between isolation sites (as described in Supplementary Fig. S4), genotypes (as shown in Fig. 1), and chemotypes determined based on LC-MSMS analyses (as described in Fig. 5) for the 65 Western European *Microcystis* strains. The overall correspondence between genotypes and chemotypes suggests that accessory metabolite production might be a key chemical trait that is conserved within clades, regardless of their respective biogeographic records. Champs indicates Champs-sur-Marne, whereas Varennes and St Maries designate Varennes-sur-Seine and St Maries de la Mer, respectively. **B** Synthetic results of toxicity tests conducted with extracts from 12 *Microcystis* strains (see Supplementary Fig. S20). The toxicity was measured in terms of lethality percentage on the embryos and larvae of Medaka fish (see Supplementary Fig. S21). Overall, the results reveal a striking similarity in toxicity between the paired strains within the same genotypes. However, pairs from the different genotypes within the same genospecies exhibit remarkably distinct toxicological effects. It is worth noticing that strains producing microcystins (MCs) show clear toxicity on hatched larvae, while embryos with chorion display milder toxic responses (PMC 830.12, PMC 831.12, PMC 728.11, and PMC 815.12). Interestingly, certain non-microcystin-producing strains (PCC 9804, NIES 08, PMC 810.12, and PMC 727.11) exhibit high toxicity levels on both embryos and larvae, while other non-MC-producing strains belonging to genotype i2 and k2 show low toxicity on both targets.

The patterns of diversity that we observe here suggest that each distinct *Microcystis* clade generates and maintains biochemical diversity in particular ways. In contrast to certain other micro-organisms, where the chemical diversity increases linearly with taxonomic repertoire and genome size[88], freshwater cyanobacteria demonstrate noteworthy examples of convergence in their accessory metabolite structural diversity, that do not appear systematically related to the genome size of the different taxa. As already documented, some *Planktothrix* and *Microcystis* strains exhibit almost identical metabolite repertoires, including microcystins, cyanopeptolins, microginins, aeruginosins, microviridins and anabaenopeptins[27,47]. Remarkably, both genera occupy similar ecological niches, but rarely co-occur, suggesting that these similar metabolite repertoires may be involved in competition or chemical mediation processes between these two taxa[89]. Indeed, several cyanobacterial metabolites have been proposed to play key ecological functions (e.g., allelopathy, anti-grazing or antimicrobial) within the microbial communities and may contribute to the blooming capabilities of these micro-organisms[32,58,87]. According to the screening hypothesis posited by Firn and Jones[90], mechanisms to generate and retain biochemical diversity could be selected, despite the global energetic costs, because of the potential adaptive benefits increased by their capability of generating molecular novelties. Thus, the phenomenon of diversification of accessory metabolite production could be considered analogous to the intrinsic capacity of the mammalian immune system to generate structural novelties of the recognition domains of antibodies, resulting in a capacity to respond to diverse antagonists of different kinds[28].

Microcystis diversification mechanisms?

The intra-lake diversity, which should be maintained by gene flow barriers between genotypes[23,91], could suggest the existence of different *Microcystis* ecotypes, each adapted to distinct spatial or temporal niches within these discontinuous ecosystems. For comparison, picocyanobacteria genome investigation shows that freshwater clades have a greater diversity and gene pool compared to their salt-adapted counterparts in relation with their respective ecological niche-richness[92–96]. However, for freshwater planktonic cyanobacteria, the potential ecological drivers that may globally shape genomes and BGC contents remain insufficiently characterized at the scale of diverse freshwater ecosystems[91,97,98].

According to the micro-diversity theory implemented for prokaryotes[99], the co-occurrence of different genotypes of the same clade within the same cyanobacterial bloom may constitute a remarkable ecotype reservoir, thereby supporting the perennity of the population proliferation beyond intra-seasonal or inter-annual environmental variations. For instance, recent genomic investigation of co-proliferating *Aphanizomenon gracile* genotypes reveals considerable genomic variations, with core genomes representing only 68% of the whole genomes[58]. So far, only whole genome sequencing appears to be appropriate to fully capture such important genomic diversity, as single gene markers such as 16S–23S rRNA sequence appear almost identical between genotypes (ANI > 99% identity). Regarding *Microcystis* micro-diversity, these classical phylogenetic markers fail to operationally distinguish genotypes[22,23,100,101]. This may be a general case for various colonial freshwater cyanobacteria, whose relatively large genomes (ranging from 4 to 8 Mb) show numerous traces of horizontal gene transfer (HGT)

distorting the global phylogenetic signal[19,51]. However, the diversification of distinct lineages within a population and the evolutionary events leading to the appearance of micro-diversity in natural populations remain mostly unknown. Interestingly, the investigation of diversity patterns of planktonic heterotrophic bacteria, such as SAR11, shows that multiple genotypes coexist within the same populations and remain stable over time, potentially constituting a general evolutionary trait for diverse prokaryotes[102].

Conclusions

Taken together, our results enhance our understanding of the adaptations of planktonic freshwater Cyanobacteria, such as the harmful bloom-forming *Microcystis*, to different environments. This highlights the importance of ecological constraints imposed by the environment in shaping the evolution of Cyanobacteria accessory metabolite repertoire. Additionally, our results indicate that a significant portion of the *Microcystis* chemical diversity remains uncharacterized. Further analytical chemical efforts guided by dereplicative approaches combining genome mining and metabolomic approaches are now needed to reveal these novel compounds[27,48].

Understanding the physiological, toxicological or biological significance of *Microcystis* chemical traits will be crucial for models aiming to predict bloom ecology and ecotoxicology. Further research, including competition experiments and examination of environmental drivers beyond the production of various secondary metabolite families, is still necessary to understand the links between genotype, phenotype and competitive outcomes.

Methods

Strain collection and study design

The 344 publicly available genome assemblies belonging to the *Microcystis* genus were retrieved from the NCBI FTP server (GenBank and/or RefSeq databases) using the genus name identifiers of 1125 [*Microcystis*] on 14 January 2022. Low-quality genomes, cross-references, and assemblies belonging to the same strain were then removed from the dataset, reducing its size to 276 assemblies.

Additionally, 58 and 13 strains of *Microcystis* spp. from several Western European areas were obtained from the Paris Muséum Collection and Pasteur Culture of Cyanobacteria (PMC and PCC), respectively, for biomass cultures, DNA extraction and de novo genome sequencing, for those yet unavailable (Supplementary Data S1), as described in the following text. Then, this genomic dataset was filtered by removing genomes displaying completeness <98%, and/or contamination >5% determined with CheckM (version 1.1.3) using cyanobacteria-specific marker genes and default parameters[103]. After this step, 347 unique *Microcystis* genomes (276 publicly available after filtering and 58 newly sequenced PMC and 13 PCC strains) constitute the basis for further genomic analysis (Supplementary Data S1).

We then identified monophyletic coherent clades supported by both core gene and pan-gene based phylogenies and biosynthetic gene cluster detection as described in Supplementary Figs. S1 and 2. In parallel, we performed a whole metabolomic investigation (comprising LC-MS/MS-based analyses and molecular networking, as detailed in Supplementary Fig. S3) of metabolite diversity produced by the 65 Western European *Microcystis* strains selected from both PMC and PCC collections (Supplementary Fig. S4).

DNA extraction, genome sequencing, and assembly

To sequence the new *Microcystis* genomes from PMC, cultures were grown in 150 mL of Z8 medium[104] at 25 °C in 250-mL Erlenmeyer flasks. The cultures were exposed to a photon flux density of 20 µmol m^{-2} s^{-1} and maintained on a 16:8 h light:dark cycle. Fresh cultures were then centrifuged (3200 × *g*, 15 min at 15 °C), cell biomass rinsed twice with pure water and then freeze-dried. DNA was extracted from 10 mg of freeze-dried cells using a DNeasy® PowerLyser® PowerSoil® Kit (Cat N° 12855-100, Qiagen, Germany) following the manufacturer's protocol. Mechanical lysis was carried out with a FastPrep-24™ 5 G BeadBeater (MP Biomedical, USA) for five cycles of 45 s ON, 45 s OFF at a frequency of six beats per second. For the new *Microcystis* genomes from PCC, the strains were cultured in medium BG11 + 2 mM $NaNO_3$ + 10 mM $NaHCO_3$[105], and DNA was extracted as described in Shih et al.[61]. DNA quantity was measured using a Qubit dsDNA BR assay kit (according to the manufacturer's recommendation) and a Qubit 4 fluorometer. Total DNA was sequenced as 150-bp paired-end reads using an Illumina NovaSeq platform (Eurofins, Konstanz, Germany). Reads were first trimmed and quality checked using Trimmomatic[106], then corrected and assembled using MetaSPAdes v3.12[107]. The resulting scaffolds were taxonomically assigned with CAT[108], and sequences belonging to *Microcystis* were extracted and pooled in a *Microcystis* pangenome dataset for each strain. Corrected reads were mapped against it to extract all reads originating from *Microcystis* genomes. These reads were assembled using SPAdes with default parameters. Genome completeness was then assessed using CheckM[103]. All information concerning the newly sequenced genomes is provided in Supplementary Data S1.

Phylogenomic analyses

From our dataset of 347 *Microcystis* genomes, we identified a core genome of 791 single-copy genes shared by all genomes. This was done using Roary with a sequence similarity threshold of 90% and multiple alignment performed by MAFFT[109,110] based on coding gene prediction by Prokka[111]. The concatenated core gene sequences multiple alignment, totaling 686,669 nucleotides, was used to construct a phylogenetic tree using RaxML v8.2.12[112] with the GTR-Gamma model, and 100 bootstraps. The phylogenetic tree was visualized and annotated using iTOL[111], with *M. aeruginosa* Ma_SC_T_19800800_S464 chosen as the root according to Cai and co-wokers[22]. Pairwise distances from the phylogenomic tree were analyzed using R package (version 1.4.1103) ape[113] and the correlation between genetic and geographic[114] distances was tested using linear regression. Average Nucleotide Identity between genome pairs was calculated using pyani[115]. Genospecies and genotypes were defined as clusters of strains obtained by MCL with a low inflation of 1.2 on the ANI criterion after filtering at 97% and 99%, respectively. Following the protocol outlined in Puillandre et al.[116], the pairwise ANI values obtained between each pair of genomes are ranked from smallest to largest. Additionally, the pan-genome of the 65 Western European *Microcystis* strains included in our study was analyzed based on the presence/absence matrix of coding genes. Estimated distances were calculated using Jaccard's method, and strain linkages were depicted using R software (version 1.4.1103)[117].

Biosynthetic gene clusters prediction

Putative biosynthetic gene clusters (BGCs) encoded within the 65 Western European *Microcystis* genomes currently under investigation were analyzed using AntiSMASH 6.0[118] with the detection strictness set to relaxed and all additional features activated. BGC matches were further examined if they were at least 5.5 kb in length[61] and contained Non-Ribosomal Peptide Synthase (NRPS) (cyanopeptolin, aeruginosin, anabaenopetin and aeruginoguanidine), Polyketide Synthase (PKS) (MAA, HG, merocyclophane-like, bartoloside-like and heptadecene) NRPS/PKS (microcystin and microginin), or Ribosomally synthetized and Post-translationally modified Peptides (RiPPs) (microcylamide, pyricyclamide, microviridin, or aeruginosamide). These clusters were initially annotated automatically using the MiBig reference database[62] and then validated manually to address potential inaccuracies due to incomplete sequences. Of particular interest was the microginin BGC as this cluster contains an internally repeated region that is often misassembled and split onto two distinct contigs, resulting in incorrect automatic annotations that required a manual curation, as previously described[119]. BGCs corresponding to terpenes (such as carotenoids) were present in all strains and were not further reported for strain comparison. Additionally, the aeruginosamide BGC was queried using BLAST against the PCC 9432 reference sequence[77,120].

Biomass production and metabolite extraction

The 65 Western European *Microcystis* strains (63 from France, 1 from Scotland, 1 from Netherland), 59 clonal but non-axenic from PMC and 6 monoclonal and axenic from PCC, were cultured in 150 mL BG-11 medium[105] at 22 °C in 250-mL Erlenmeyer flasks, with a photon flux density of 12 µmol m^{-2} s^{-1}, a 13:11 h light:dark cycle and a daily manual gentle agitation (Supplementary Fig. S3).

A volume equivalent to 10% of each culture (above 15–20 mL) was used to inoculate successive new cultures and grown for 28 days. The remaining 90% of cells (collected from 120 to 130 mL culture volume) were washed three times at room temperature with pure water by centrifugation for 10 min at 3200 × *g*. The supernatants were discarded, and the pellets were stored at −80 °C for at least 3 h, followed by lyophilization (−47 °C, 0.03 mBar for 16 h) (Labconco, Kansas City, MO, USA). Three successive cycles of culture-biomass sampling were performed and analyzed separately.

Simultaneously, a specific experiment was conducted over a single 28-day period to monitor variations in intracellular metabolite content. Experimental conditions (250 mL of BG-11 medium in 500-mL Erlenmeyers) were carried out in triplicates, with a 16 h:8 h photoperiod (12 µmol m^{-2} s^{-1}; 25 °C) and constant homogenization maintained by magnetic agitators. Two weekly samples of 12 mL were taken throughout the 28 days (final volume above 150 mL). Each sample was divided into (i) the extraction of extra-cellular metabolites (10 mL) and (ii) the biomass monitoring using optical density (2 mL) at 750 nm performed by in vivo spectrophotometry (Cary 500; Varian, U.S.A.).

For all intracellular metabolite extracts, five mg of freeze-dried biomass were extracted with 500 µL of methanol:water (75:25) with 0.1% formic acid solution. Cell lysis was achieved by sonication: 3 cycles of 30 s, with a 10 s break between each cycle, at 80% of the maximum intensity (SONICS VibraCell, Newton, CT, USA; 130 Watts, 20 kHz). Samples were then centrifuged (10 min, 13,400 × *g*, 4 °C), and the supernatants were collected and stored in darkness at −20 °C prior to mass spectrometry analysis.

LC-MS/MS metabolome analyses

Microcystis strain extracts were separated using ultra-high performance liquid chromatography (UHPLC). For each sample, 2 µL was injected, and molecule separation was performed by a Polar Advance II 2.5 pore C_{18} chromatographic column (Thermo Fisher Scientific, Waltham, MA, USA) at a flow rate of 300 µL min^{-1} under a linear gradient of acetonitrile acidified with 0.1% formic acid (from 5 to 90% in 15 min). Metabolite contents were analyzed using an electrospray ionization hybrid quadrupole time-of-flight (ESI-QqTOF) high-resolution mass spectrometer (Compact, Bruker, Bremen, Germany). The electrospray ionization (ESI) system was calibrated with a capillary temperature set to 200 °C, the source voltage at 3.5 kV, and gas flow rates at 8 L min^{-1}. Then, ions were analyzed in the range 50–1500 *m/z* using collision ion dissociation (CID) and autoMS/MS in positive or negative modes with information-dependent acquisition (IDA). Thus, compounds were analyzed in simple MS positive or negative modes without quadrupole fragmentation at a 2-Hz frequency acquisition rate, then top-intensity ions (>5000 counts in single MS (MS1)) were individually selected by the quadrupole in a window of 10 Da and fragmented in a collision cell (MS2) with a selective exclusion window of 30 s (except for ions presenting a count-intensity increase superior to 3 times) with selective ion collision set between according to ion intensity and mass (10–50 eV ramp; 50/50% time window; 2–8 Hz adapted according to ion count intensity), in consecutive cycle times of 2.5 s. The resulting ions of the fragmentation of their respective parent were transferred and detected.

MS data were processed using MetaboScape 4.0 software (Bruker, Bremen, Germany) for recalibration of each sample analysis (<1 ppm, according to sodium formate internal standard), peak detection and selection of ions whose intensity was greater than 5000 counts in at least 10% of the set of samples and a minimal RT-correlation coefficient set at 0.7. Different states of charge (1+, 2+, and 3+ or 1−, 2−, and 3−) and classical adducts were grouped together and the peak area was determined in order to generate a unique global data matrix containing semi-quantification results for each metabolite in all analyzed samples' peak for each analyte (characterized by the respective mean mass of its neutral form and its corresponding retention time). Data QC and Blank samples (injected every six samples and every triplicate, respectively) were examined to ensure the reproducibility and robustness of the whole data series. First analyte annotations were attempted according to respective ion mass (<2 ppm) and isotopic pattern (<20 msigma) by automatic match with the CyanometDB 1.0 database which contains over 2100 chemical formulas of known specialized metabolite produced by cyanobacteria[121] and the natural product NPAtlas 2.0 database[122].

The files containing all the MS2 fragmentation information for each ion analyzed were exported in mgf format (DOI: 10.17632/2rnz75jpzr.1) using MetaboScape 4.0 (Bruker, Bremen, Germany) software before being used for the generation of the molecular network of structural similarity using the GNPS algorithm[123] with the MetGem software (version 1.3.6)[124]. Additional metabolite annotations were also made from MS2 data by comparing fragmentation profiles with the GNPS algorithm (with parameters set at: *m/z* tolerance = 0.02 Da; minimum matched peaks = 4; topK = 10; minimal cosine score = 0.70) using GNPS, NIH, MS-DIAL and EMBL public and generalist spectral databases. Global annotation confidence score was provided according to Schymanski and coworkers recommandations[125].

Medaka fish embryo larvae toxicological tests

We have complied with all relevant ethical regulations for animal use. All fish exposures to different *Microcystis* extracts were conducted on freshly fertilized eggs of *Oryzias latipes* (Cab lineage) at the Medaka fish facility of the Muséum National d'Histoire Naturelle (MNHN - Paris, France – SIRET 1800441740019 and agreement number N°C75-05-11), under standard controlled conditions and in accordance with relevant European animal protection laws (European Directive 2010/63/EU)[126]. Embryos and larvae were exposed to similar amounts of *Microcystis* culture extracts from strains belonging to distinct genotypes and/or genospecies, including 2 strains from genotypes d2 (PMC 728.11 and PMC 815.12), 2 from genotype i1 (PMC 823.12 and PMC 826.12), 2 others from genotype i3 (PMC 830.12 and PMC 831.12), 2 from genotype i5 (PMC 727.11 and PMC 810.12), 2 strains from genospecies I (PCC 9804 and NIES 98) and 2 from genotype k2 (PMC 1317.21 and PMC 1323.21) (for genospecies and genotype correspondence see Fig. 1).

Initially, freshly collected eggs (n_{total} = 480) were incubated in 24-well exposure plates (filled with 1 mL Yamamoto's medium and 4 embryos per well) with extracts of the different *Microcystis* strains (14 mg of extracted *Microcystis* cells per mL in 1% methanol final concentration, n = 20, solvent being used as negative control). The medium containing the extracts was completely replaced after 24 h of incubation. Subsequently, the survival rate, developmental arrest and occurrence of head or tail malformations were individually observed after 48 h of exposure. Additionally, other freshly fertilized embryos were maintained in 6-cm Petri dishes filled with Yamamoto's medium, which was renewed daily until egg hatching (8 dpf). The larvae (n_{total} = 432) were then collected and incubated for 24 h in 12-well exposure plates (filled with 3 mL Yamamoto's medium and 3 larvae per well) with extracts of different *Microcystis* strains (14 mg of extracted *Microcystis* cells per mL in 1% methanol final concentration, n = 12, solvent being used as negative control). The larvae were monitored and the survival rate was measured after 24 h of incubation. At the end of the procedure, all exposed embryos and larvae were euthanized in 0.5% tricaine methane sulfonate (MS-222; CAS n°886-86-2 Sigma, St. Louis, MO) buffered with 0.1% $NaHCO_3$. It is worth noticing that acceptable low larvae mortality (<10%) occurred with the control condition in accordance with OECD 236 test guidelines (https://www.oecd.org/en/publications/2013/07/test-no-236-fish-embryo-acute-toxicity-fet-test_g1g34036.html).

Data treatment and analyses

The MetaboAnalyst 5.0 platform[127] was utilized for data matrix normalization according to *Pareto* scaling involving mean-centering, followed by

division by the standard deviation for the different metabolite matrices. Metabolite correlation within the strains was calculated using Spearman's rank and visualized on a heatmap with hierarchical clustering in MetaboAnalyst 5.0. A clade based on the presence or absence of coding gene sequences was generated on a dendrogram employing Jaccard's similarity/dissimilarity index, Ward reconstruction, and Silhouette analysis to determine the optimal number of clusters using R software (version 1.4.1103)[117].

Statistics and reproducibility

The stability of *Microcystis* metabolomes has been investigated on three successive 28-day cultures with 59 PMC strains and on the 28-day monitoring of 3 strains (namely, PMC 810.11, 826.11 and PCC 7806) sampled every 2 or 3 days, grown in triplicated vessels. The metabolome LC-MS investigation has been performed in triplicate analyses. All conditions of the toxicological tests have been performed on 20 embryos and 12 larvae.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw data of new *Microcystis* genomes reported in this study (58 and 13 from PMC and PCC, respectively) have been deposited in the ENA database and are available under the accession numbers PRJEB101697 and PRJEB105414, respectively. The accession numbers of the assembled genomes corresponding to these individual strains are listed in Supplementary Data S1. Figure 4 combines AntiSmash and metabolomic data that are available on Supplementary Data S2, the data used for Fig. 7B are available on Supplementary Data S4. Whole metabolomics dataset can be found on Mendeley (DOI: 10.17632/2rnz75jpzr.1). Raw data are available upon request.

Received: 6 May 2025; Accepted: 15 January 2026;
Published online: 23 January 2026

References

1. Zhang, X., Liu, X., Yang, F. & Chen, L. Pan-genome analysis links the hereditary variation of *Leptospirillum ferriphilum* with its evolutionary adaptation. *Front. Microbiol.* **9**, 577 (2018).
2. Caputo, A., Fournier, P. E. D. & Raoult, D. Genome and pan-genome analysis to classify emerging bacteria. *Biol. Direct* **14**, 5 (2019).
3. Garner, R. E. et al. A genome catalogue of lake bacterial diversity and its drivers at continental scale. *Nat. Microbiol.* **8**, 1920–1934 (2023).
4. Zhang, X., Xiao, L., Liu, J., Tian, Q. & Xie, J. Trade-off in genome turnover events leading to adaptive evolution of Microcystis aeruginosa species complex. *BMC Genom.* **24**, 462 (2023).
5. Harke, M. J. et al. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae* **54**, 4–20 (2016).
6. Dick, G. J. et al. The genetic and ecophysiological diversity of Microcystis. *Environ. Microbiol.* **23**, 7278–7313 (2021).
7. Mantzouki, E., Visser, P. M., Bormans, M. & Ibelings, B. W. Understanding the key ecological traits of cyanobacteria as a basis for their management and control in changing lakes. *Aquat. Ecol.* **50**, 333–350 (2016).
8. Pineda-Mendoza, R. M. et al. Seasonal changes in the bacterial community structure of three eutrophicated urban lakes in Mexico City, with emphasis on *Microcystis* spp. *Toxicon* **179**, 8–20 (2020).
9. Wu, Y. et al. Seasonal dynamics of water bloom-forming Microcystis morphospecies and the associated extracellular microcystin concentrations in large, shallow, eutrophic Dianchi Lake. *J. Environ. Sci.* **26**, 1921–1929 (2014).
10. Jackrel, S. L. et al. Genome evolution and host-microbiome shifts correspond with intraspecific niche divergence within harmful algal bloom-forming Microcystis aeruginosa. *Mol. Ecol.* **28**, 3994–4011 (2019).
11. Komárek, J. & Komárková, J. Review of the European Microcystis morphospecies (Cyanoprokaryotes) from nature. *Fottea* **2**, 1–24 (2002).
12. Lepère, C., Wilmotte, A., & Meyer, B. Molecular diversity of Microcystis strains (Cyanophyceae, Chroococcales) based on 16S rDNA sequences. *Syst. Geogr. Plants* **70**, 275–283 (2000).
13. Wu, Z. X., Gan, N. Q. & Song, L. R. Genetic diversity: geographical distribution and toxin profiles of Microcystis strains (Cyanobacteria) in China. *J. Integr. Plant Biol.* **49**, 262–269 (2007).
14. Haande, S. et al. Diversity of Microcystis aeruginosa isolates (Chroococcales, Cyanobacteria) from East-African water bodies. *Arch. Microbiol.* **188**, 15–25 (2007).
15. Yoshida, M. et al. Intra-specific phenotypic and genotypic variation in toxic cyanobacterial Microcystis strains. *J. Appl. Microbiol.* **105**, 407–415 (2008).
16. Otsuka, S. et al. A proposal for the unification of five species of the cyanobacterial genus Microcystis Kützing ex Lemmermann 1907 under the rules of the Bacteriological Code. *Int. J. Syst. Evolut. Microbiol.* **51**, 873–879 (2001).
17. Frangeul, L. et al. Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genom.* **9**, 1–20 (2008).
18. Humbert, J. F. et al. A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. *PloS ONE* **8**, e70747 (2013).
19. Gaëtan, J. et al. Widespread formation of intracellular calcium carbonates by the bloom-forming 800 cyanobacterium Microcystis. Environ. Microbiol. **25**, 751–765 (2023).
20. Chen, M., Xu, C., Wang, X., Wu, Y. & Li, L. Nonribosomal peptide synthetases and nonribosomal cyanopeptides synthesis in Microcystis: a comparative genomics study. *Algal Res.* **59**, 102432 (2021a).
21. Pérez-Carrascal, O. M. et al. Single-colony sequencing reveals microbe-by-microbiome phylosymbiosis between the cyanobacterium *Microcystis* and its associated bacteria. *Microbiome* **9**, 1–21 (2021).
22. Cai, H., McLimans, C. J., Beyer, J. E., Krumholz, L. R. & Hambright, K. D. Microcystis pangenome reveals cryptic diversity within and across morphospecies. *Sci. Adv.* **9**, eadd3783 (2023).
23. Pérez-Carrascal, O. M. et al. Coherence of Microcystis species revealed through population genomics. *ISME J.* **13**, 2887–2900 (2019).
24. Willis, A. & Woodhouse, J. N. Defining cyanobacterial species: diversity and description through genomics. *Crit. Rev. Plant Sci.* **39**, 101–124 (2020).
25. Yancey, C. E. et al. 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 (2023).
26. Welker, M., Maršálek, B., Šejnohová, L. & Von Doehren, H. Detection and identification of oligopeptides in Microcystis (cyanobacteria) colonies: toward an understanding of metabolic diversity. *Peptides* **27**, 2090–2103 (2006).
27. Le Manach, S. et al. Global metabolomic characterizations of Microcystis spp. highlights clonal diversity in natural bloom-forming populations and expands metabolite structural diversity. *Front. Microbiol.* **10**, 791 (2019).
28. Gluck-Thaler, E. et al. The architecture of metabolism maximizes biosynthetic diversity in the largest class of fungi. *Mol. Biol. Evol.* **37**, 2838–2856 (2020).
29. Wiegand, C. & Pflugmacher, S. Ecotoxicological effects of selected cyanobacterial secondary metabolites a short review. *Toxicol. Appl. Pharmacol.* **203**, 201–218 (2005).

30. Babica, P., Blaha, L. & Marsalek, B. Exploring the natural role of microcystins—a review of effects on photoautotrophic organisms. *J. Phycol.* **42**, 9–20 (2006).
31. Schatz, D. et al. Ecological implications of the emergence of non-toxic subcultures from toxic *Microcystis* strains. *Environ. Microbiol.* **7**, 798–805 (2005).
32. Leão, P. N., Vasconcelos, M. T. & Vasconcelos, V. M. Allelopathy in freshwater cyanobacteria. *Crit. Rev. Microbiol.* **35**, 271–282 (2009).
33. Simon, R. D. Cyanophycin granules from blue-green alga Anabaena cylindrica— Reserve material consisting of copolymers of aspartic acid and arginine. *Proc. Natl. Acad. Sci. USA* **68**, 265–267 (1971).
34. Garcia Pichel, F., Sherry, N. D. & Castenholz, R. W. Evidence for an ultraviolet sunscreen role of the extracellular pigment scytonemin in the terrestrial cyanobacterium Chlorogloeopsis sp. *Photochem. Photobio.* **56**, 17–23 (1992).
35. Itou, Y., Okada, S. & Murakami, M. Two structural isomeric siderophores from the freshwater cyanobacterium Anabaena cylindrica (NIES-19). *Tetrahedron* **57**, 9093–9099 (2001).
36. Nagle, D. G. & Paul, V. J. Production of secondary metabolites by filamentous tropical marine cyanobacteria: ecological functions of the compounds. *J. Phycol.* **35**, 1412–1421 (1999).
37. Combes, A., Dellinger, M., Cadel-six, S., Amand, S. & Comte, K. Ciliate Nassula sp. grazing on a microcystin-producing cyanobacterium (Planktothrix agardhii): impact on cell growth and in the microcystin fractions. *Aquat. Toxicol.* **126**, 435–441 (2013).
38. Leflaive, J. P. & Ten-Hage, L. Algal and cyanobacterial secondary metabolites in freshwaters: a comparison of allelopathic compounds and toxins. *Freshw. Biol.* **52**, 199–214 (2007).
39. Sharif DI, Gallon J, Smith CJ, Dudley E. Quorum sensing in Cyanobacteria: N-octanoyl-homoserine lactone release and response, by the epilithic colonial cyanobacterium Gloeothece PCC6909. *ISME J*. **2**, 1171–1182 (2008).
40. Krumbholz, J. et al. Deciphering chemical mediators regulating specialized metabolism in a symbiotic Cyanobacterium. *Angew. Chem. Int. Ed.* **61**, e202204545 (2022).
41. Welker, M. et al. Seasonal shifts in chemotype composition of Microcystis sp. communities in the pelagial and the sediment of a shallow reservoir. *Limnol. Oceanogr.* **52**, 609–619 (2007).
42. Kust, A. et al. Insight into unprecedented diversity of cyanopeptides in eutrophic ponds using an ms/ms networking approach. *Toxins* **12**, 561 (2020).
43. Lifshits, M. & Carmeli, S. Metabolites of Microcystis aeruginosa bloom material from Lake Kinneret, Israel. *J. Nat. Prod.* **75**, 209–219 (2012).
44. Pearson, L. A., Crosbie, N. D. & Neilan, B. A. Distribution and conservation of known secondary metabolite biosynthesis gene clusters in the genomes of geographically diverse Microcystis aeruginosa strains. *Mar. Freshw. Res.* **71**, 701–716 (2019).
45. McDonald, K., DesRochers, N., Renaud, J. B., Sumarah, M. W. & McMullin, D. R. Metabolomics reveals strain-specific cyanopeptide profiles and their production dynamics in *Microcystis aeruginosa* and M. flos-aquae. *Toxins* **15**, 254 (2023).
46. Kleigrewe, K. et al. Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from cyanobacteria. *J. Nat. Prod.* **78**, 1671–1682 (2015).
47. Kim Tiam, S. et al. Insights into the diversity of secondary metabolites of Planktothrix using a biphasic approach combining global genomics and metabolomics. *Toxins* **11**, 498 (2019).
48. Yancey, C. E. et al. Metabologenomics reveals strain-level genetic and chemical diversity of *Microcystis* secondary metabolism. *mSystems* **9**, e0033424 (2024).
49. Ferrinho, S., Connaris, H., Mouncey, N. J., & Goss, R. J. Compendium of metabolomic and genomic datasets for Cyanobacteria: mined the gap. *Water Res.* **256**, 121492 (2024).
50. Arevalo, P., VanInsberghe, D., and Polz, M.F. A reverse ecology framework for Bacteria and Archaea. in *Population Genomics: Microorganisms*, 77–96 (Springer, 2018).
51. Chen, M. Y. et al. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J.* **15**, 211–227 (2021b).
52. Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. USA* **102**, 2567–2572 (2005).
53. Chang, H. Y., Yen, H. C., Chu, H. A. & Kuo, C. H. Population genomics of a thermophilic cyanobacterium revealed divergence at subspecies level and possible adaptation genes. *Bot. Stud.* **65**, 35 (2024).
54. Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D. H. & Soo, R. M. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J.* **15**, 1879–1892 (2021).
55. Ribeiro, K. F., Ferrero, A. P., Duarte, L., Turchetto-Zolet, A. C. & Crossetti, L. O. Comparative phylogeography of two free-living cosmopolitan cyanobacteria: insights on biogeographic and latitudinal distribution. *J. Biogeogr.* **47**, 1106–1118 (2020).
56. Tromas, N. et al. Niche separation increases with genetic distance among bloom-forming cyanobacteria. *Front. Microbiol.* **9**, 438 (2018).
57. van Gremberghe, I. et al. Lack of phylogeographic structure in the freshwater cyanobacterium Microcystis aeruginosa suggests global dispersal. *PloS ONE* **6**, e19561 (2011).
58. Halary, S. et al. Intra-population genomic diversity of the bloom-forming cyanobacterium, Aphanizomenon gracile, at low spatial scale. *ISME Commun.* **3**, 57 (2023).
59. Doré, H. et al. Differential global distribution of marine picocyanobacteria gene clusters reveals distinct niche-related adaptive strategies. *ISME J.* **17**, 720–732 (2023).
60. Xu, Y., Leung, S. K., Li, T. M. & Yung, C. C. Hidden genomic diversity drives niche partitioning in a cosmopolitan eukaryotic picophytoplankton. *ISME J.* **18**, wrae163 (2024).
61. Shih, P. M. et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. USA* **110**, 1053–1058 (2013).
62. Terlouw, B. R. et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **51**, D603–D610 (2023).
63. Gavriilidou, A. et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7**, 726–735 (2022).
64. Rodríguez-Gijón, A. et al. Linking prokaryotic genome size variation to metabolic potential and environment. *ISME Commun.* **3**, 25 (2023).
65. Dehm, D. et al. Unlocking the spatial control of secondary metabolism uncovers hidden natural product diversity in Nostoc punctiforme. *ACS Chem. Biol.* **14**, 1271–1279 (2019).
66. Dienst, D., Wichmann, J., Mantovani, O., Rodrigues, J. S. & Lindberg, P. High density cultivation for efficient sesquiterpenoid biosynthesis in Synechocystis sp PCC 6803. *Sci. Rep.* **10**, 5932 (2020).
67. Núñez-Montero, K. et al. Genomic and metabolomic analysis of Antarctic bacteria revealed culture and elicitation conditions for the production of antimicrobial compounds. *Biomolecules* **10**, 673 (2020).
68. Roussel, T. et al. Limnospira (Cyanobacteria) chemical fingerprint reveals local molecular adaptation. *Microbiol. Spectr.* **13**, e01901–24 (2025).
69. Le Manach, S. et al. Physiological effects caused by microcystin-producing and non-microcystin producing *Microcystis aeruginosa* on medaka fish: a proteomic and metabolomic study on liver. *Environ. Pollut.* **234**, 523–537 (2018).

70. Briand, E. et al. Spatiotemporal changes in the genetic diversity of a bloom-forming *Microcystis aeruginosa* (cyanobacteria) population. *ISME J.* **3**, 419–429 (2009).
71. Kuijpers, M. C. et al. Intraspecific divergence within Microcystis aeruginosa mediates the dynamics of freshwater harmful algal blooms under climate warming scenarios. *Proc. B* **292**, 20242520 (2040).
72. Puddick, J. et al. High levels of structural diversity observed in microcystins from Microcystis CAWBG11 and characterization of six new microcystin congeners. *Mar. Drugs* **12**, 5372–5395 (2014).
73. Hellweger, F. L. et al. Models predict planned phosphorus load reduction will make Lake Erie more toxic. *Science* **376**, 1001–1005 (2022).
74. Whitton, B. A. *Ecology of Cyanobacteria II: Their Diversity in Space and Time* (Springer, 2012).
75. Cirés, S. & Ballot, A. A review of the phylogeny, ecology and toxin production of bloom-forming Aphanizomenon spp. and related species within the Nostocales (cyanobacteria). *Harmful Algae* **54**, 21–43 (2016).
76. Kurmayer, R., Deng, L. & Entfellner, E. Role of toxic and bioactive secondary metabolites in colonization and bloom formation by filamentous cyanobacteria Planktothrix. *Harmful Algae* **54**, 69–86 (2016).
77. Leikoski, N. et al. Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. *Chem. Biol.* **20**, 1033–1043 (2013).
78. Figueiredo, S. A. et al. Discovery of cyanobacterial natural products containing fatty acid residues. *Angew. Chem. Int. Ed.* **60**, 10064–10072 (2021).
79. Chen, Q., Wang, L., Qi, Y. & Ma, C. Imaging mass spectrometry of interspecies metabolic exchange revealed the allelopathic interaction between *Microcystis aeruginosa* and its antagonist. *Chemosphere* **259**, 127430 (2020).
80. Kim Tiam, S. et al. The success of the bloom-forming cyanobacteria Planktothrix: genotypes variability supports variable responses to light and temperature stress. *Harmful Algae* **117**, 102285 (2022).
81. van Wichelen, J. eroen et al. Strong effects of amoebae grazing on the biomass and genetic structure of a Microcystis bloom (Cyanobacteria). *Environ. Microbiol.* **12**, 2797–2813 (2010).
82. Zhai, C. hunmei et al. The mechanism of competition between two bloom-forming Microcystis species. *Freshw. Biol.* **58**, 1831–1839 (2013).
83. Herrera, N. & Echeverri, F. Evidence of quorum sensing in Cyanobacteria by Homoserine Lactones: the origin of blooms. *Water* **13**, 1831 (2021).
84. Burberg, C., Ilić, M., Petzoldt, T. & von Elert, E. Nitrate determines growth and protease inhibitor content of the cyanobacterium Microcystis aeruginosa. *J. Appl. Phycol.* **31**, 1697–1707 (2019).
85. Dittmann, E., Gugger, M., Sivonen, K. & Fewer, D. P. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. *Trends Microbiol.* **23**, 642–652 (2015).
86. Carmichael, W. W. Cyanobacteria secondary metabolites—the cyanotoxins. *J. Appl. Bacteriol.* **72**, 445–459 (1992).
87. Demay, J., Bernard, C., Reinhardt, A. & Marie, B. Natural products from cyanobacteria: focus on beneficial activities. *Mar. Drugs* **17**, 320 (2019).
88. Vesth, T. C. et al. Investigation of inter-and intraspecies variation through genome sequencing of Aspergillus section Nigri. *Nat. Genet.* **50**, 1688–1695 (2018).
89. Briand, E. et al. Chemically mediated interactions between Microcystis and Planktothrix: impact on their growth, morphology and metabolic profiles. *Environ. Microbiol.* **21**, 1552–1566 (2019).
90. Firn, R. D. & Jones, C. G. Natural products–a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20**, 382–391 (2003).
91. Geers, A. U., Strube, M. L. & Bentzon-Tilia, M. Small spatial scale drivers of secondary metabolite biosynthetic diversity in environmental microbiomes. *Msystems* **8**, e00724–22 (2023).
92. Cabello-Yeves, P. J. et al. Elucidating the picocyanobacteria salinity divide through ecogenomics of new freshwater isolates. *BMC Biol.* **20**, 175 (2022).
93. Wang, Y. et al. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of Illumina tags. *Appl. Environ. Microbiol.* **78**, 8264–8271 (2012).
94. Salcher, M. M. Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. *J. Limnol.* **73**, 74–87 (2014).
95. Singer, D. et al. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ. Int.* **146**, 106262 (2021).
96. Wu, D., Seshadri, R., Kyrpides, N. C. & Ivanova, N. N. A metagenomic perspective on the microbial prokaryotic genome census. *Sci. Adv.* **11**, eadq2166 (2025).
97. Silva, S. G., Nabhan Homsi, M., Keller-Costa, T., Rocha, U. & Costa, R. Natural product biosynthetic potential reflects macroevolutionary diversification within a widely distributed bacterial taxon. *Msystems* **8**, e00643–23 (2023).
98. Chevrette, M. G. et al. Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat. Prod. Rep.* **37**, 566–599 (2020).
99. Doolittle, W. F. & Zhaxybayeva, O. On the origin of prokaryotic species. *Genome Res.* **19**, 744–756 (2009).
100. Li, Q. et al. A large-scale comparative metagenomic study reveals the functional interactions in six bloom-forming Microcystis-epibiont communities. *Front. Microbiol.* **9**, 746 (2018).
101. Kiledal, E. A. et al. Comparative genomic analysis of Microcystis strain diversity using conserved marker genes. *Harmful Algae* **132**, 102580 (2024).
102. López-Pérez, M., Haro-Moreno, J. M., Coutinho, F. H., Martinez-Garcia, M. & Rodriguez-Valera, F. The evolutionary success of the marine bacterium SAR11 analyzed through a metagenomic perspective. *mSystems* **5**, e00605-20 (2020).
103. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
104. Kotai, J. Instructions for preparation of modified nutrient solution Z8 for algae. *Nor. Inst. Water Res.* **11**, 5 (1972).
105. Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. Generic assignments, strain histories and properties of pure cultures of Cyanobacteria. *J. Gen. Microbiol.* **111**, 1–61 (1979).
106. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
107. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
108. von Meijenfeldt, F. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* **20**, 1–14 (2019).
109. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
110. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
111. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
112. Stamatakis, A. et al. RAxML-Light: a tool for computing terabyte phylogenies. *Bioinformatics* **28**, 2064–2066 (2012).
113. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

114. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
115. Pritchard, L. et al. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* **8**, 12–24 (2016).
116. Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. J. M. E. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* **21**, 1864–1877 (2012).
117. R Core Team R : a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org (2024).
118. Blin, K. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
119. Le Moigne, D. et al. Dynamics of the metabolome of Aliinostoc sp. PMC 882.14 in response to light and temperature variations. *Metabolites* **11**, 745 (2021).
120. Cegłowska, M., Szubert, K., Wieczerzak, E., Kosakowska, A. & Mazur-Marzec, H. Eighteen new aeruginosamide variants produced by the Baltic cyanobacterium *Limnoraphis* CCNP1324. *Mar. Drugs* **18**, 446 (2020).
121. Jones, M. R. et al. CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria. *Water Res.* **196**, 117017 (2021).
122. van Santen, J. A. et al. The Natural Products Atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res.* **50**, D1317–D1323 (2022).
123. Aron, A. T. et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **15**, 1954–1991 (2020).
124. Olivon, F. et al. MetGem software for the generation of molecular networks based on the t-SNE algorithm. *Anal. Chem.* **90**, 13900–13908 (2018).
125. Schymanski, E. L. et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).
126. Marie, B., Le Meur, M., Duval, C., Quiquand, M., Lance, E. & Duperron, S. The threat is in the details–critical gap in ecotoxicological assessment of Microcystis blooms revealed by critical distinctions of genotype effects induced on Medaka fish. *Environ. Pollut*. **387**, 127344 (2025).
127. Pang, Z. et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).

Acknowledgements

This work was supported by the ANR MC-Tox project, grant ANR CE34-SJ 11008-22 of the French Agence Nationale de la Recherche. The Paris Muséum Collection (PMC) and the Pasteur Cultures of Cyanobacteria (PCC) collection are funded by the MNHN and the Institut Pasteur, respectively. The mass spectrometry analyses were acquired at the Plateau technique de spectrométrie de masse bio-organique, Muséum National d'Histoire Naturelle, Paris, France. The authors thank the anonymous referees for providing valuable suggestions that significantly improved the quality of the manuscript.

Author contributions

S.H., M.G., and B.M. conceived and designed the experiments; C.D. isolated all new strains of the PMC; A.H., M.L.M., C.D., S.H., M.Q., and M.B. performed the analysis; A.H., J.B.P., L.M., M.D., S.H., M.D., and B.M. treated the data. All authors wrote and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-026-09599-7.

Correspondence and requests for materials should be addressed to Sébastien Halary or Benjamin Marie.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary handling editor: David Favero. A peer review file is available.