Peer-reviewed version available at *Toxins* **2019**, *11*, 498; doi:10.3390/toxins11090498

Article

Insights into the Diversity of Secondary Metabolites of *Planktothrix* Using a Biphasic Approach Combining Global Genomics and Metabolomics

Sandra Kim Tiam[1], Muriel Gugger[2], Justine Demay[1], Séverine Le Manach[1], Charlotte Duval[1], Cécile Bernard[1] and Benjamin Marie[1]*

[1] Muséum National d'Histoire Naturelle, UMR 7245, CNRS, MNHN Molécules de Communication et Adaptation des Micro-organismes (MCAM), équipe "Cyanobactéries, Cyanotoxines et Environnement", 12 rue Buffon - RDC bâtiment de cryptogamie - CP 39, 75231 Paris Cedex 05, France.
[2] Institut Pasteur, Collection des Cyanobactéries, 28 rue du Dr Roux, 75724 Paris Cedex 05, France.
* Correspondence should be addressed to: benjamin.marie@mnhn.fr

Abstract: Cyanobacteria are an ancient lineage of slow-growing photosynthetic bacteria and a prolific source of natural products with diverse chemical structures and potent biological activities and toxicities. The chemical identification of these compounds remains a major bottleneck. Strategies that can prioritize the most prolific strains and novel compounds are of great interest. Here, we combine chemical analysis and genomics to investigate the chemodiversity of secondary metabolites based on their pattern of distribution within some cyanobacteria. *Planktothrix* being a cyanobacterial genus known to form blooms worldwide and to produce a broad spectrum of toxins and other bioactive compounds, we applied this combined approach on four closely related strains of *Planktothrix.*
The chemical diversity of the metabolites produced by the four strains was evaluated using an untargeted metabolomics strategy with high-resolution LC-MS. Metabolite profiles were correlated with the potential of metabolite production identified by genomics for the different strains. Although, the *Planktothrix* strains present a global similarity in term biosynthetic cluster gene for *microcystin*, *aeruginosin* and *prenylagaramide* for example, we found remarkable strain-specific chemo-diversity. Only few of the chemical features were common to the four studied strains. Additionally, the MS/MS data were analyzed using Global Natural Products Social Molecular Networking (GNPS) to identify molecular families of the same biosynthetic origin. In conclusion, we present an efficient integrative strategy for elucidating the chemical diversity of a given genus and link the data obtained from analytical chemistry to biosynthetic genes of cyanobacteria.

Keywords: Cyanobacteria; secondary metabolite; genome mining; molecular networking.

1. Introduction

Cyanobacteria are an ancient lineage of bacteria that colonize a broad range of habitats, from soil to oceans, and play important roles in the global nitrogen and carbon cycles [1]. They are also well known for the production of toxins, and to form toxic blooms in fresh water bodies around the world, posing a threat to human health [2-4]. However, it has become clear that both marine, terrestrial and freshwater cyanobacteria can produce a large set of natural products in addition to toxic compounds, many of which exhibit remarkable biological activities potentially involved in various ecological or physiological processes [5-6].

The known chemical diversity of cyanobacterial natural products includes over 1,100 secondary metabolites [7]. Isolated by means of traditional bioactivity-guided screening techniques [8], these compounds present also promising therapeutic potential, with anticancer, multidrug-reversing, antifungal, antibacterial, anti-inflammatory, antiviral, and potent enzyme-inhibiting bioactivities [9-10]. These molecules can be respectively assigned to a diverse panel of structural classes, including peptides, polyketides, alkaloids, lipids, and terpenes [10-11]. These complex metabolites are commonly synthesized through on enzymatic assembly lines of non-ribosomal and ribosomal biosynthetic pathways namely the non-ribosomal peptide synthetase (NRPS) and the polyketide synthase (PKS) enzyme systems [12], or the ribosomally synthesized peptides and post-translationally modified pathways (RiPP) [13].

Genome-mining studies based upon the recent increase of genome sequences available in public databases have demonstrated unexpected diversity and greatly expanded the known distribution of these biosynthetic gene clusters across cyanobacteria [14]. Indeed, genome mining demonstrates that these biosynthetic gene clusters typically encode a large range of auxiliary enzymes that tailor the structure of the secondary metabolites and greatly increase the chemical diversity of the products [15-16]. Hence, the huge variety in genetic organization and various associated tailoring enzymes signify that the currently established chemical diversity remains an underestimation of the achievable potential of PKS and NRPS pathways.

Improvements of predictive bioinformatic tools combining established knowledge of secondary metabolism and hidden Markov model-based algorithms such as ClustScan [17], ClusterFinder [18] or AntiSMASH [19] have resulted in a more facile identification of core and tailoring enzymes of these pathways, as well as other secondary metabolite gene clusters. These approaches give us also the opportunity to detected gene clusters corresponding to orphan of known product. It now opens new windows for the description of biosynthetic pathways and their respective natural products.

A great perspective for compound identification is expected by combining genome mining with high throughput global analytical methods. Metabolomics, that aim at providing a holistic investigation of the chemical diversity, bring new highlights on natural products research. The main bottleneck of this approach remains in the unambiguous identification of the detected compounds. Based on approaches combining high-resolution mass spectrometry (HRMS) coupled to ultra-high-performance liquid chromatography (UHPLC), up-to-date metabolite profiling platforms can rapidly generate accurate structural information for hundreds of metabolites in crude natural extracts. In addition, the development of the Global Natural Products Social molecular networking approach (such as GNPS) brought opportunities to integrate MS/MS data providing a very powerful tool for molecular annotation [20]. This process compares individual MS/MS fragmentation patterns of each analyte, then constitutes

molecular families of structural similarity features, potentially sharing the same biosynthetic origin. This enables the comparison of a high number of samples at once aiding dereplication and tentative structural characterization, and constitutes a promising tool especially when integrated with genome information [21-22].

Here, we investigated the metabolite diversity of four monoclonal strains of *Planktothrix,* a prominent freshwater bloom-forming cyanobacteria in lakes and reservoirs producing a large set of various bioactive metabolites [*e.g.* 23]. To be the most exhaustive as possible, we develop an integrated global metabolomic and genomic approach, taking advances of genes and chemical feature analyses with publicly available databases and prediction tools.

2. Results and discussion

2.1 Biosynthetic gene cluster approach based on genome analysis

The four strains of *Planktothrix* were selected on their green (PCC 10110, PCC 7805 and NIVA CYA 126/8) *versus* red (PPC 7821) morphotypes and their characteristics to produce (PCC 10110, PCC 7821 and NIVA CYA 126/8) or not (PCC 7805) microcystins (Table S1). The global genome of PCC 10110 compared to the ones of PCC 7805, PPC 7821, and NIVA CYA 126/8 [24-25] shared a synteny value above 80%, testifying of the global genome conservation between those strains. The clustering of the four strains revealed a close relationship between the two strains PCC 7805 and NIVA 126/8, while the PCC 7821 and the more recently isolated strain from France PCC 10110 were more distantly related. No valuable distinction between the red/green (*i.e. P. rubescens/agardhii*) morphotypes was observed (Figure S1), but the clustering confirms the separation of the planktonic *Planktothrix* from the benthic *Planktothrix* species as previously observed [25]. Overall, the genomic comparison between these four strains highlights their global gene content similarities (Figure S2).

The four *Planktothrix* genomes contain distinct backgrounds in term of biosynthetic gene clusters (Figure 1). These clusters represent 1-2% of the whole genome size. In total, three genes clusters coding for three RiPPs and eight genes cluster coding for NRPS and/or PKS were found in these four genomes (Figure 1). Seven of these encode the synthesis of already described metabolite families, including prenylagaramides (*pag*), microviridins (*mdn*), anabaenopeptins (*apt*), cyanopeptolins (*oci/mcn*), aeruginosins (*aer*), microcystins (*mcy*) and microginins (*mic*). The shared gene clusters of these known compounds were highly conserved (between 92.0 and 99.4% of amino-acid sequences identity – Figure 1). In all four strains, the *apt, mic*, and *oci* genes appeared to be co-localised in a ~70 kb genomic island, as it had been previously showed in *Planktothrix rubescens* NIVA CYA 98 [25,26]. In addition to the seven gene clusters corresponding to already described metabolite families (i.e. *pag*, *mdn*, *apt*, *oci*, *aer*, *mcy* and *mic*), four other gene clusters encode enzymes for so-far unidentified products (arbitrarily called R1, PNL1, PNL2 and PNL3). Their respective sequences, although presenting the signature of RiPP, NRPS and/or PKS genes identified by AntiSMASH [27], do not exhibit significant similarity in the database.

Synthetic pathways	RiPP			NRPS and/or PKS							
Planktothrix strains / clusters	*Prenylagaramide (pag)*	*Unknown R1*	*Microviridin (mdn)**	*Anabaenopeptin (apt)**	*Cyanopeptolin (mcn/oci)**	*Aeruginosin (aer)*	*Microcystin (mcy)*	*Microginin (mic)*	*Unknown PNL1*	*Unknown PNL2*	*Unknown PNL3*
PCC 7821	Ref	_	Ref	Ref	Ref	Ref	Ref[g]	Ref[m]	_	Ref	_
PCC 7805	92.1	_	92.0	Partial	96.6	97.0[h]	_	_	Ref[m]	_	_
NIVA CYA 126/8	91.7	Ref	94.7	92.2	97.1	96.3	99.4	_	_	_	Ref
PCC 10110	98.3	_	96.4	98.2	98.7	98.3	99.0	_	_	_	_

Figure 1. Distribution and diversity of the gene clusters involved in the biosynthesis of natural products of the studied *Planktothrix*. The gene clusters involved in the biosynthesis of natural products of the studied *Planktothrix* comprise ribosomally synthesized and post-translationally modified peptides

(RiPPs), non-ribosomal peptide synthetase (NRPS)/polyketide synthase (PKS). Similar genomic islands containing biosynthetic gene clusters of *apt*, *oci* and *mdn* were found in both strains. Genetic distances are calculated according to reference with PCC 7821 strain and are expressed as percentage of amino acid identity. h: halogenase-containing cluster; s: shorter *mcyA* sequence; m: sequence presents on megaplasmid; *: cluster from the same genetic island.

Overall, this suggests that various secondary metabolites could be commonly produced by the four strains (as six out of the 11 gene clusters were found in more than one strain) but also that some secondary metabolites will be strain specific (since the five remaining clusters were found in only one strain). Note that the partially *apt* sequences of the PCC 7805 strain is very likely not functional. Interestingly, the *aer* cluster of the PCC 7805 strain contains a sequence encoding for a halogenase, suggesting the production of halogenated aeruginosin variants by this strain.

The genus *Planktothrix* had been already known for the production of bioactive peptides, including the protein phosphatase inhibiting microcystins, the protease inhibiting cyanopeptolins, aeruginosins, and microviridins, among others [19,28]. A better understanding of the production of these cyanobacterial secondary metabolites has been progressively supported by the elucidation of the synthesis pathways for all the main peptide families: microcystins [29], aeruginosins [30], anabaenopeptins [31], cyanopeptolins [15], microviridins [32], and prenylagaramides [16]. Individual cyanobacterial clones generally produce a limited number of peptides families. The capacity of production of these peptides families depend of the presence/absence of the corresponding gene clusters [7,28]

In the last decade, studies have been focused on single peptide families [*e.g.* 30] and since more recently global approaches have been undertaken in order to characterize in an exhaustive way the secondary metabolites production potential for a given cyanobacteria strain. Indeed, in last years, various cyanobacterial genome projects have been initiated in order to detect the occurrence of metabolite biosynthetic pathways. Such recent genomic analyses based on the identification of biosynthetic gene clusters using predictive software are now revealing the extent of the genetic diversity for natural product biosynthesis potential from cyanobacterial genomes, such as *Planktothrix*.

2.2 Insights of the molecular networking for the characterization the chemical diversity of the Planktothrix *secondary metabolite variants*

In order to identify *Planktothrix* secondary metabolites, we performed a molecular network based on the global fragmentation pattern profile generated by high resolution tandem mass spectrometry after UHPLC separation of metabolites in the four strains (Figure S3). Molecular networking utilizes MS/MS data to sort parent ions based on their structural similarity according to their respective fragmentation profiles. Indeed, the secondary ion mass fragmentation data relates directly to molecular structure because chemical bonds break on the basis of bond strength, that rely on respective residue structures. The spectral networking of the GNPS algorithm uses the normalize intensity of all fragment ions as independent axes in order to construct multidimensional vectors specific of each spectrum, and to finally compare their similarities using a cosine function and to visualize these relation-ships between different parent ion masses on a plot diagram representation [33-34].

The resulting network of the four *Planktothrix* metabolomes were obtained from 2,360 analytes and 30 reference compounds and contains 49 clusters of three or more analytes, regrouping 405 analytes and 69 clusters of two analytes (Figure 2), whereas 429 analytes remain as singles. The molecular clusters

were further annotated using the reference compounds and matches with components from libraries publically available on GNPS platform, such as MassBank, GNPS/NIST14 or EMBL. The nodes grouped in the same molecular clusters (clouds of nodes, linked together according to cosine score >0.6) exhibit similar fragmentation patterns and are specific of the structure of single chemical families. We were thus able to annotate larger clusters according to the fragmentation pattern similarity of their analytes. Some clusters are constituted by ions of primary metabolites such as two clusters of carbohydrates and several single clusters representing for cyclic alkaloids, small alkaloids, phospholipids, and di- and/or tri-peptides. Several clusters were containing secondary metabolites such as seven clusters for cyanopeptolins, three clusters for microcystins, two clusters representing anabaenopeptins, aeruginosins, prenylagaramides and microginins, one for microviridins (Figure 2). The occurrences of these metabolites in the four *Planktothrix* strains is almost perfectly matching with the prediction of presence/absence of their respective gene clusters from genomes (Figure 1). This results shows that the large majority of secondary metabolites predicted based on the presence of gene clusters were detected by LC-MS/MS analysis (Figure 3).

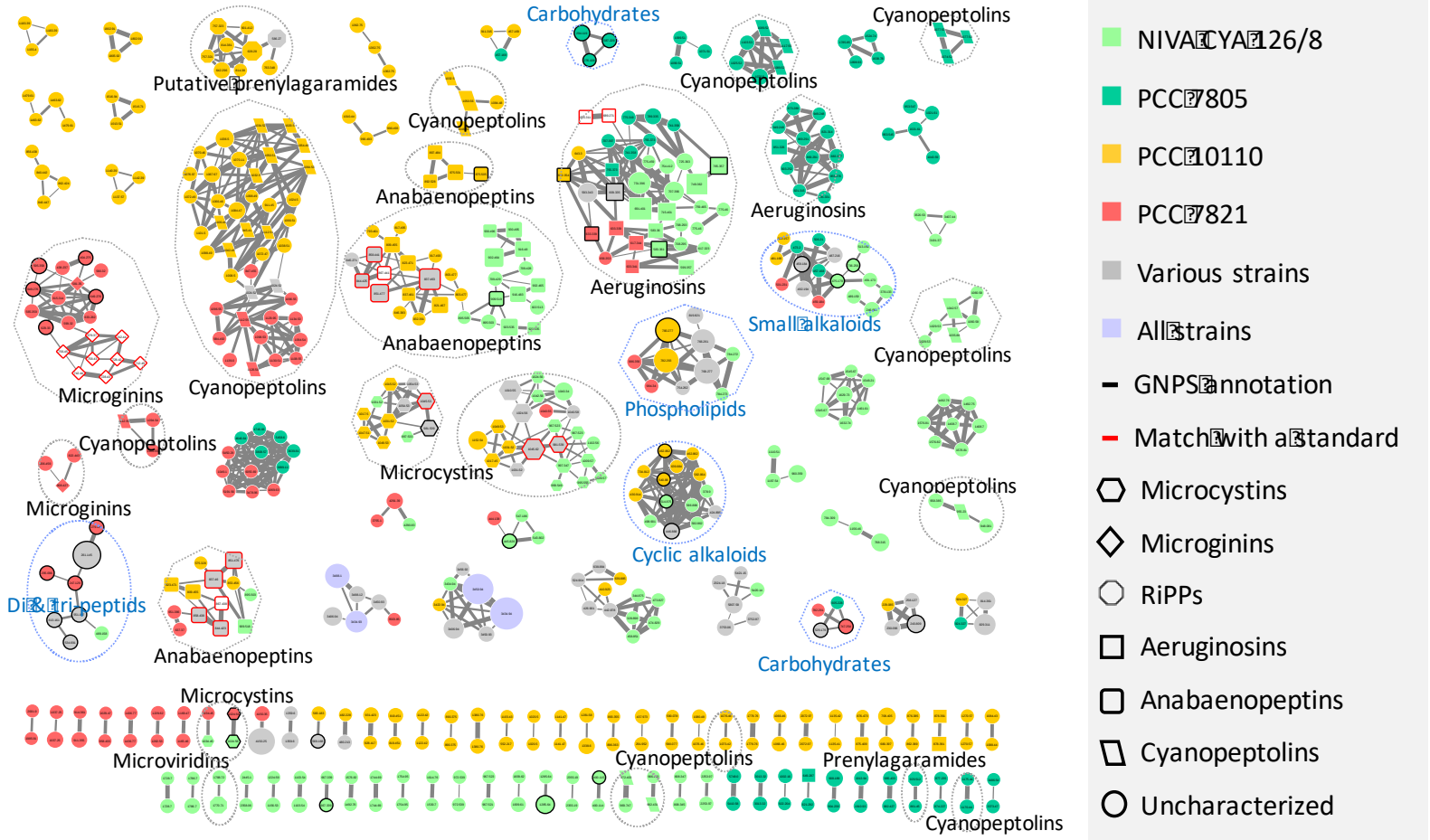


Figure 2. Molecular network generated from MS/MS spectra of the four *Planktothrix* strains using GNPS tool (all data and results are freely available on the GNPS server at the address http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=98e54f0fa2a84efeb82efa0d24e4d974). The GNPS algorithm compares all MS/MS spectra by aligning them one by one, grouping identical molecules (presenting identical mass and fragmentation pattern) and assigning cosine score ranking from 0 to 1 to each alignment, allowing network reconstruction of the link between each molecule according to the cosine score links between all molecules (cosine score significance threshold set to 0.6). Analytes whom individual masses match with known secondary metabolites from cyanobacteria are indicated with specific shapes. Correspondences with standard molecules from cyanobacterial similarly analysed or with components from the fragmentation pattern library available from the GNPS server are indicated by heavy red and black lines, respectively. Only molecular clusters of at least 2 nodes are represented.

Synthetic pathways	RiPP		NRPS and/or PKS					Others			?			
Planktothrix strains / Molecular family	Prenylagaramide	microviridin	Anabaenopeptin	Cyanopeptolin	Aeruginosin	Microcystin	microginin	Phospholipid	small alkaloid	Carbohydrate	Unknown	Unknown	Unknown	Unknown
PCC 7805	✔	N.D.	✘	✔	✔[h]	✘	✘	✔	✔	✔	✔	✘	✘	✘
PCC 7821	✔	N.D.	✔	✔	✔	✔	✔	✔	✔	✔	✘	✔	✘	✘
PCC 10110	✔	N.D.	✔	✔	✔	✔	✘	✔	✔	✔	✘	✘	✔	✘
NIVA CYA 126-8	N.D.	✔	✔	✔	✔	✔	✘	✔	✔	N.D.	✘	✘	✘	✔

Figure 3. Distribution and diversity of metabolite produced by the four *Planktothrix* strains according to metabolomic dataset investigated with GNPS (Figure 2). Green "v" and red cross indicate when the gene cluster or the metabolite are present, or not, respectively. "?" shows when metabolites were identified with no strong certainty, (because of the lack of standard and no detection of related gene cluster in the genome) and "N.D." when no metabolite could have been detected, although its specific gene is observed in the respective genome.

With contrast, prenylagaramides and microviridins (both being cyclic peptides produced by RiPP pathways) were detected in PCC 7805, PCC 7821 and PCC 10110, and in NIVA CYA 126/8 strains respectively (Figure 3), when the genome analysis predicted a potential synthesis of both families for all the four studied strains (Figure 1). We assume that the identification of such cyclic peptides derived from RiPP gene cluster machineries using MS/MS-based molecular network presents here the main limitation of such global investigation of the *Planktothrix* metabolite diversity. Indeed, this approach is established on molecular structural similarity of components, and for RiPP cyclic peptides especially, a single mutation in their pro-peptide sequence (*e.g.* INDELs) is susceptible to induce a complete structural modification of the synthetized peptide, making it not suitable for the identification of new variant by the GNPS clustering.

However, mass spectrometry based-molecular networking represents a clearly valuable tool for the description of the whole molecular diversity of the cyanobacterial metabolites produced by different strains [35-37]. It takes full advances of the capabilities of modern mass spectrometer based analytical solutions including the high sensitivity, the high resolution of the molecular mass, the accuracy of the isotopic pattern, the chromatographic retention time, and the fragmentation pattern, together with the use of reference compound libraries for automatic structural identity or analogy search [20,38-39]. Although, no ionization technique is universal, the electrospray ionization (ESI) on the positive mode appears to effectively ionize a wide range of structural classes providing good coverage of the global cyanobacterial metabolome [40].

Cyanobacteria, such as *Planktothrix*, typically produce multiple variants of the same metabolite family [19,40,41]. Much of the metabolite chemical variations can be attributed to a lack of specificity of the NRPS biosynthetic machinery. The production of different variants may indeed be attributed to gene mutations (as for example the gain and loss of genes coding for tailoring enzymes or a modification of the active site inducing a modification of the enzyme substrate affinity) or to change in the availability of the substrates within the cell [*e.g.* 42]. It has been proposed that the synthesis of multiple variants of a metabolite family could allow a larger bioactivity panel and confer some plasticity and adaptive advantages for the producing organism [43]. For example, within monospecific blooms of *Planktothrix*, environmental factors drive the dynamics of microcystin and non-microcystin producing strains [44-45]. However, the evolutive and adaptive outcomes of this structural and functional diversification remains poorly investigated [28,46].

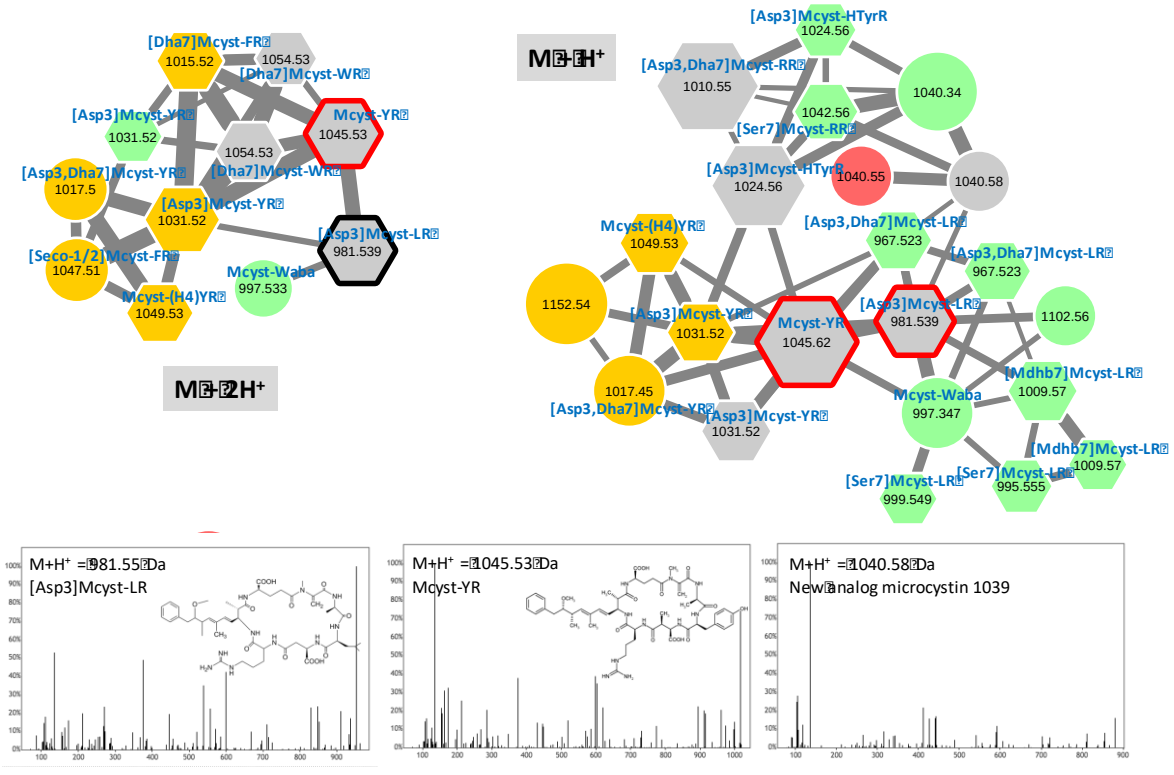
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2019 doi:10.20944/preprints201906.0310.v1

Peer-reviewed version available at Toxins 2019, 11, 498; doi:10.3390/toxins11090498

Overall, our results demonstrate the efficiency of the GNPS based analysis of the global chemical diversity of *Planktothrix* metabolites. In the following sections, we choose to mainly focus our attention on the case of the variant diversity observed within microcystin, anabaenopeptin, and aeruginosin clusters.

2.2.1 Microcystins

Some of the cyanobacterial metabolites, such as microcystins, are subject to intensive monitoring due to their role in animal toxicoses [3]. Microcystins are cyclic hepta-peptides that have been firstly described from *Microcystis* and *Planktothrix* (Review in [19,44,47]). They are characterized by the presence of a non-proteinaceous amino acid in position 5 (Adda), 2 amino acid derived from Asp and Glu in position 3 and 6, respectively, and 2 very variable positions (2 and 4), that serve as reference to the name of the variant. Above 240 different variants have been described so far [48], 215 being references in the microcystin database Toxinmasslist_com_v15b (Miles 2018). Two main microcystin molecular clusters were highlighted in the GNPS network. Their identification as microcystin variants was confirmed by their direct identification of two of them according to microcystin standards ([Asp3]-Mcyst-LR and Mcyst-YR) (Figure 4). In these two microcystin clusters, other various ions presented a match between their respective mass and those of microcystin variants previously described [49], while the others, not presenting any match with known microcystin variants correspond to potential new microcystin variants (*e.g.* the *m/z* 1040.58 node). For these potentially new microcystin variants, the observation of their MS/MS spectra shows that they present similar fragmentation patterns with other known microcystins (Figure 4). Globally, the three microcystin-producing strains studied here (*Planktothrix* PCC 10110, PCC 7821 and NIVA CYA 126/8) exhibit only few common variants (in grey), comprising principally MC-YR, [Asp3]-Mcyst-YR, [Asp3]-Mcyst-LR, among others.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2019 doi:10.20944/preprints201906.0310.v1

Peer-reviewed version available at Toxins 2019, 11, 498; doi:10.3390/toxins11090498

Figure 4. Two main microcystin clusters highlighted by the GNPS analysis based on the MS/MS CID fragmentation spectra obtained from the four *Planktothrix* strains. Analytes whom individual masses match with known microcystins are indicated by hexagons. Correspondences with standard microcystins similarly analysed or with components whom fragmentation pattern are available in library of the GNPS server are indicated by heavy red and black lines, respectively. The colour code is similar as those shown on figure 2 caption. Example of MS/MS spectra and chemical structures are shown for [Asp3]Mcyst-LR, Mcyst-YR and a potential new MC variant exhibiting a *m/z* of 1040.58 Da. Note that $(M+H)^+$ and $(M+2H)^{2+}$ ions are grouped in two distinct molecular clusters.

2.2.2 Anabaenopeptins

Anabaenopeptins constitutes a very diverse family of cyclic hexa-peptides that have been described from *Microcystis*, *Planktothrix*, *Anabaena*, *Aphanizomenon* and *Nostoc* (see review in [50], and above 75 different variants have been described so far [40]. Except for the D-Lys (position 2) that is linked to the carboxylic group of the amino acid placed in position 6, all other positions are variable allowing a large structural diversity of the family which molecules exhibit masses between 750 and 950 Da [51]. Two main anabaenopeptin clusters ($M+H^+$ and $M+2H^{2+}$ clusters) were highlighted in the GNPS network according to the formal identification of six standard molecules (anabaenopeptin A, B, F, G, NZ857 and ferentoic acid A) analyses in parallel of the four *Planktothrix* extracts with the same protocol (Figure 5). Other components of these molecular clusters correspond to ions presenting a match with the mass of other anabaenopeptin variants previously described [40], or for above 1/3 of them to compounds that very likely correspond to potentially new anabaenopeptin variants.

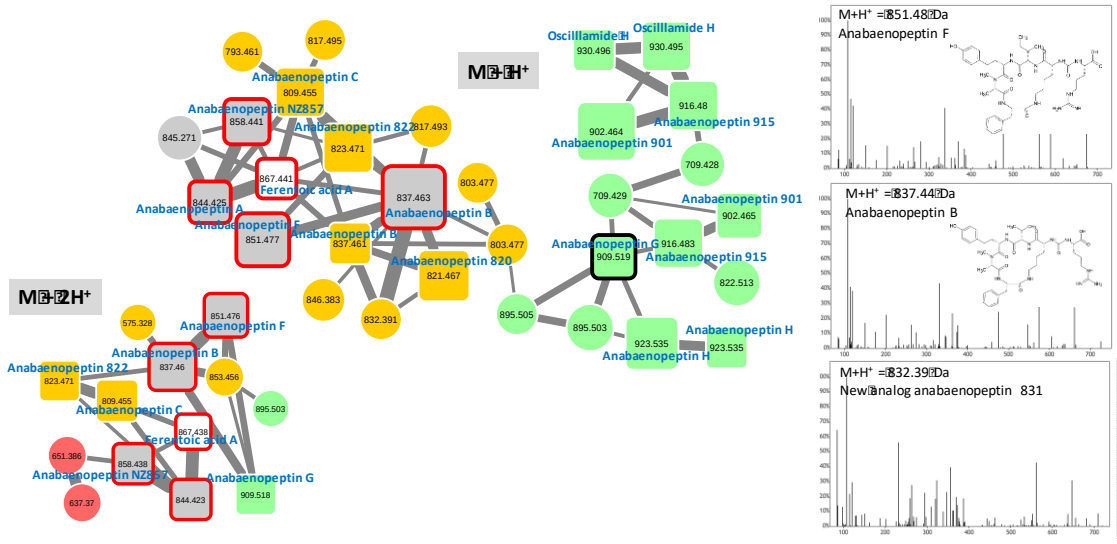


Figure 5. Two anabaenopeptin clusters highlighted by the GNPS analysis based on the MS/MS CID fragmentation spectra obtained from the four *Planktothrix* strains. Analytes whom individual masses match with known anabaenopeptins are indicated by squares with rounded corners. Correspondences with standard anbaenopeptins similarly analysed or with components whom fragmentation pattern are available in library of the GNPS server are indicated by heavy red and black lines, respectively. The colour code is similar as those shown on figure 2 caption. Example of MS/MS spectra and chemical structures are shown for anabaenopetin F, anabaenopetin A and a potential new anabaenopeptin variant exhibiting a *m/z* of 832.539 Da. Note that $(M+H)^+$ and $(M+2H)^{2+}$ ions are grouped in two distinct clusters.

2.2.3 *Aeruginosins and halogenation*

Aeruginosins constitute a linear tetra-peptide family that represent above 94 different variants that have been described so far [40]. Their MS/MS fragmentation patterns are often characterized by the presence of a Choi fragment (immonium with 140.109 *m/z*) and other recurrent fragments from Hpla or Pla residues [19]. Their composition is rather variable and the component of this family exhibit masses comprised between 430 and 900 Da [51]. The molecular network obtained from the four *Planktothrix* strains exhibits two aeruginosin clusters (Figure 6) that were highlighted by the presence of seven standard molecules. Other components of these molecular clusters correspond to ions presenting a mass match with other variants of aeruginosin previously described [40], or for 58% of all this compounds to potential new aeruginosin variants. Interestingly, we notice that several aeruginosin variants of the PCC 7805 (shown in dark green in Figure 6A) present characteristic isotopic pattern of mono- or di-chlorination, as recently illustrated for this strain by Briand and co-workers [52]. These observations are in perfect agreement with genomic analysis that detects a non-heme iron O_2-dependent halogenase, potentially involved in the halogenation of biosynthetic products in the *aer* cluster of *Planktothrix* PCC 7805 strain (Figure 1).

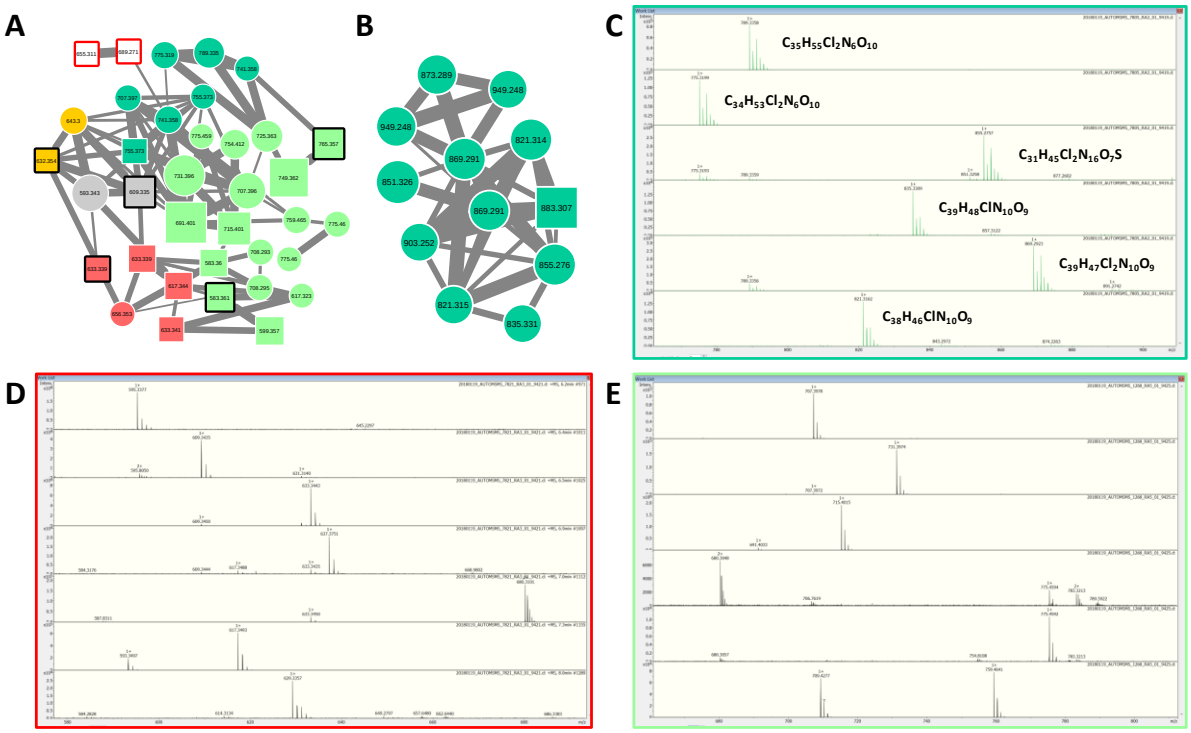


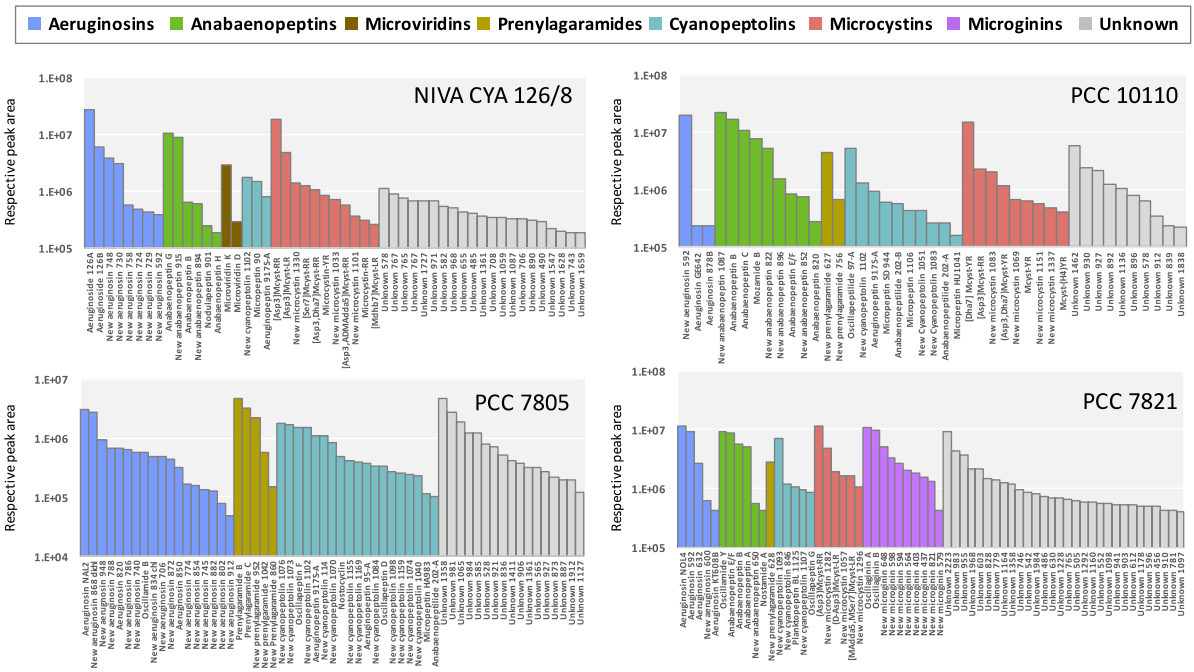
Figure 6. Two aeruginosin clusters highlighted by the GNPS analysis based on the MS/MS CID fragmentation spectra obtained from the four *Planktothrix* strains. Analytes whom individual masses match with known aeruginosins are indicated as squares with sharped corners (A-B). Correspondences with standard anbaenopeptins similarly analysed or with components whom fragmentation pattern are available in library of the GNPS server are indicated by heavy red and black lines, respectively. The colour code is similar as those shown on figure 2. Selection of isotopic pattern of various aeruginosins observed in the strains PCC 7805 (C), PCC 7821 (D), and PCC 10110 (E), respectively. Note that these isotopic patterns indicate the presence of mono- or di- chlorinations, but exclusively for aeruginosins observed in the PCC 7805 strain.

315

316 *2.3 Global capability to annotate cyanobacteria metabolites*

317 On the first hand, the potential of secondary metabolite production of the four *Planktothrix* strains was
318 investigated by genome mining approach using RiPP, NRPS and/or PKS gene cluster search software
319 AntiSmash 4.0 (Figure 1). This investigation suggests that various metabolite families (*i.e.* microcystins,
320 anabaenopeptins, aeruginosins, cyanopeptolins, microviridins, microviridins, prenylagaramides and
321 cyanobactins) could be synthetized by these different strains. On the other hand, the molecular diversity
322 of the produced metabolites was investigated by high resolution mass spectrometry based analyses. In
323 this approach, molecules could be identified according to their molecular formula (estimated according
324 to their accurate mass and isotopic pattern), their retention time, and the presence of qualifying ions in
325 their MS/MS fragmentation spectra. Such direct identifications (referred here as "gold" annotations) are
326 supported by a specific database of analytical standard molecules. The whole analysis of MS/MS spectra
327 by GNPS molecular networking gave the opportunity to propose even more annotations for the
328 molecules whom analytical standards are lacking (that is the case of most of the cyanobacterial
329 metabolites). Indeed, the analytes that belong to a molecular cluster that present a match with a standard
330 can be annotated by extension, as "silver" annotation, when the molecule presents a mass that
331 corresponds to an already known cyanobacterial metabolite [40] or as "bronze" annotation when no
332 known metabolite corresponds, suggesting this analyte concerns a potentially new variant. Following
333 these different criteria of identification, the 40-60 most intensive ions analysed by mass spectrometry
334 have been manually annotated (Figure 7; Supplementary data S1-4). Taking together, these two
335 complementary approaches lead to the annotation of most of the main metabolites of the four
336 investigated *Planktothrix* strains. Indeed, with a good level of certainty (gold, silver or bronze
337 annotations), up to 77% of the above 40-60 most intense metabolite could have been annotated, when
338 only less than 10% could be directly annotated thanks to analytical standards only (gold annotation).

339



340

341 **Figure 7.** Main secondary metabolites detected by mass spectrometry in the four *Planktothrix* strains and
342 their respective annotations (supplementary data S1-4) according to their direct identification thanks to

analytical standards or their belonging to GNPS molecular clusters that could have been annotated (Figure 2).

By leading to the global annotation of most of the secondary metabolites (*i.e.* RiPP, NRPS/PKS), the present work illustrates the strength and the promising perspective that are offered by such joint genomic and metabolomic investigation of the chemical richness supported by clonal cultures of micro-organisms, such as cyanobacteria [53], heterotrophic bacteria [54] or fungi [55].

3. Conclusions

Combining novative global approaches, such as mass spectrometric based molecular networking and bioinformatics investigation of genomes, can help to identify interesting targets for chemical isolation. This constitutes powerful and orthogonal means for the novel natural product discovery and the in-depth strain characterization [21,53,56-57]. Such approaches can now be used in order to shed light on both hot spots strains presenting specific chemical diversity and specific molecular family that aim at being further structurally and biologically characterized.

When genome mining investigation through biosynthetic gene cluster search provides basic information on the metabolite molecular families that might be produced by an organism [58], the mass spectrometric molecular networking constitutes a remarkable tool for the direct identification of structural analogs within a set of chemical extracts [40]. We thus recommend to perform a first screening with genome mining in order to select the cyanobacteria strains of interest based on their secondary metabolites production potential. In a second step, the effective production of secondary metabolites can be characterized in order to verify the prediction based on genome mining. For that commercial standards, databases on fragmentation mass spectra and molecular networking analyses can be used together in order to obtain a precise description of the secondary metabolite diversity.

In the near future, we expect an even deeper global metabolome characterization thanks to chemo- and bio-informatic tools that are currently in development and aim at better predicting the structure of novel analogs by *in silico* MS/MS fragmentation [59], the *de-novo* sequencing of circularized and modified peptides [60] or the structure prediction of biosynthetic products based on cluster gene sequences [27].

4. Material & methods

***Planktothrix* strains and culture.** *Planktothrix agardhii* PCC 10110 isolated from a 2001 bloom sample in Paris suburban area is a green pigmented strain producing microcystins. *P. agardhii* NIVA-CYA126/8, green pigmented, and *Planktothrix rubescens* PCC 7821, red pigmented, were both microcystins-producer and originated from Nordic lakes. *Planktothrix agardhii* PCC 7805 isolated from a temperate lake in the Netherlands does not produce microcystins. The four strains were grown in BG11 media at 18 °C for PCC 7821, and at 22 °C for the three others. The cultures were maintained in 2-L Erlenmeyer flasks with a photon flux density of 6 µmol. m^{-2}. s^{-1} and a 13:11 h light:dark cycle. Fresh cultures were inoculated every four weeks to promote optimal growth.

Description of the DNA isolation, sequencing and assembling methods. The genome of *Planktothrix agardhii* PCC 10110 was obtained from a 40-mL culture. Nucleic acid extraction of cyanobacterial cells to obtain DNA were carried out as previously described [61]. Genome sequencing was performed by the Mutualized Platform for Microbiology at Institut Pasteur. The whole genome sequencing was carried out using the Nextera XT DNA sample preparation kit (Illumina) for 2x150 bps paired-ends

reads (insert size ~300 bps). All sequenced paired-ends reads were clipped and trimmed with AlienTrimmer2 (v. 0.4.0), and subjected to a sequencing error correction with Musket3 (v. 1.1) as well as a digital normalization procedure with khmer4 (v. 1.3). For each sample, remaining processed reads were assembled with SPAdes5 (v. 3.7.0). The genome was further integrated in the MicroScope platform v3.12.2 (http://www.genoscope.cns.fr/agc/microscope) similarly as the genomes of PCC 7805, PCC 7821 and NIES 126/8 [24,25,62].

***In-silico* analyses in MicroScope**. The clustering of the strain PCC 10110 was performed with the three other published genomes to display the genomic similarity between these strains. This clustering was computed from all-pairs distances ≤ 0.06 (≈94% ANI) that correspond to the ANI standard to define a species group. In addition, natural product gene clusters, including NRPS, PKS and RiPP genes, were identified using the antiSMASH 4.0 software [27] available through Microscope platform.

Metabolome biomass extraction and analysis by mass spectrometry. The biomass (20 mL) of the four *Planktothrix* strains cultured in triplicats were centrifuged at 4,000 rpm for 10 min. The supernatants discarded, the pellets were freeze-dried and lyophilized. The lyophilized cells were weighted then sonicated 2 min in 80% methanol with a constant ratio of 100 µL of solvent for 1 mg of dried biomass and centrifuged at 4°C (12,000 g; 5 min). Two µL of the supernatant representing the metabolite extracts were then analysed in triplicats on an Ultra high performance liquid chromatography (UHPLC Ultimate 3000, ThermoFisher Scientific) using a Polar Advances II 2.5 pore C_{18} column (Thermo®) at a 300 $\mu L.min^{-1}$ flow rate with a linear gradient of acetonitrile in 0.1% formic acid (5 to 90% in 21 min) coupled with a high-resolution mass spectrometer. The eluted metabolite contents were analysed using an electrospray ionization hybrid quadrupole time-of-flight (ESI-QqTOF) high resolution mass spectrometer (Maxis II ETD, Bruker) on positive simple MS or on positive autoMSMS mode with information dependent acquisition (IDA), on the 50-1500 *m/z* rang at 2 Hz or between 2-8 Hz speed, for MS and MS/MS respectively, according to relative intensity of parent ions, in consecutive cycle times of 2.5 s, with an active exclusion of previously analysed parents. The data were analysed with the DataAnalysis 4.4 software for internal recalibration (<0.5 ppm) and MGF export were generated from MS/MS spectra between 1 and 15 min. Metabolite annotation was attempted according to the precise mass of the molecules and their respective MS/MS fragmentation patterns with regards to an in-house database of above 700 cyanobacteria metabolites and confirmed with 36 commercially available standard molecules from the various cyanobacterial specific metabolite families (*e.g.* cyanopeptolins, aeruginosins, microginins, anabaenopeptins, aerucyclamides, microcystins, saxitoxins, anatoxins, cylindrospermopsins, …) analysed similarly in our MS platform.

Data treatment and molecular networking. Using the whole MS/MS data (converted in mgf format) obtained for the four strains taken together, a molecular network was created using the online workflow at Global Natural Products Social molecular networking (GNPS) (http://gnps.ucsd.edu) [20]. The data were then clustered with MS-Cluster with a parent mass tolerance of 1.0 Da and an MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra. Consensus spectra that contained less than two spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.6 and more than five matched peaks. Further edges between two nodes were kept in the network only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against the GNPS spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.6 and at least five

matched peaks. All results are freely available on GNPS server (http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=98e54f0fa2a84efeb82efa0d24e4d974). The clustered spectra of the network were annotated by comparing monoisotopic mass to our in-house cyanobacteria metabolite databases according to MS and MS/MS fragmentation pattern matches. Molecular networks were visualized using Cytoscape 3.2.1.

Supplementary Materials: Table S1: main characteristics of the 4 *Planktothrix* strains, Figure S1: Clustering of the of *Planktothrix* visualized on the MicroScope platform, Figure S2: Venn diagram of the pan-genomes of the 4 *Planktothrix* strains, Figure S2: A) Principal component analysis and B) Rarefaction curve of the whole MS/MS dataset presented in this study, Data S1-S4 : List of annotated metabolites for the NIVA CYA 126/8, PCC 10110, PCC 7805 and PCC 7821 strains, respectively.

Author contributions: BM, SKT, SLM, MG, and CB conceived and designed the experiments; CD performed cell cultures; BM, SKT, MG and SLM performed the analysis; BM, SKT, JD and MG treated the data. All authors wrote and reviewed the manuscript.

Fundings: This work was supported by the grant ANR-15-CE34-0002, project CYPHER, to CB, MG and BM. The mass spectra were acquired at the Plateau technique de Spectrométrie de Masse Bio-organique, UMR 7245 Molécules de Communication et Adaptation des Micro-organismes, Muséum National d'Histoire Naturelle, Paris, France.

Acknowledgements: The LABGeM (CEA/Genoscope & CNRS UMR8030), the France Génomique and French Bioinformatics Intitute national infrastructures (funded as part of Investissement d'Avenir program managed by Agence Nationale pour la Recherche contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support within the MicroScope annotation platform. MG thanks the Institut Pasteur for funding the collection Pasteur Culture of Cyanobacteria (PCC) where the PCC strains studied here are available.

Conflict of interest: The authors declare no conflict of interest.

References

1. Whitton, B.A. In *Ecology of Cyanobacteria II, Ecology of Cyanobacteria II: Their Diversity in Space and Time*. Springer Netherlands, Dordrecht, **2012**. doi:10.1007/978-94-007-3855-3
2. Carmichael, W. Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs. *Adv. Exp. Med. Biol.* **2008** 619, 831-53. doi:10.1007/978-0-387-75865-7
3. Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M., & Visser, P. M. Cyanobacterial blooms. *Nature Reviews Microbiology* **2018** 16(8), 471.
4. Pearson, L., Mihali, T., Moffitt, M., Kellmann, R., Neilan, B. On the chemistry, toxicology and genetics of the cyanobacterial toxins, microcystin, nodularin, saxitoxin and cylindrospermopsin. *Mar. Drugs* **2010** 8, 1650-80. doi:10.3390/md8051650
5. Leflaive, J. P., & Ten-Hage, L. Algal and cyanobacterial secondary metabolites in freshwaters: a comparison of allelopathic compounds and toxins. *Freshwater Biology* **2007** 52(2), 199-214.
6. Zak, A., Kosakowska, A. Cyanobacterial and microalgal bioactive compounds-the role of

secondary metabolites in allelopathic interactions. *Oceanol. Hydrobiol. Stud.* **2016** 45, 131-143. doi:10.1515/ohs-2016-0013

7. Dittmann, E., Gugger, M., Sivonen, K., Fewer, D.P. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. *Trends Microbiol.* **2015** 23, 642-652. doi:10.1016/j.tim.2015.07.008
8. Burja, A. M., Banaigs, B., Abou-Mansour, E., Burgess, J. G., & Wright, P. C. (2001). Marine cyanobacteria—a prolific source of natural products. *Tetrahedron* **2001** 57(46), 9347-9377.
9. Demay, J., Bernard, C., Reinhart, A., Marie, B. Natural Products from Cyanobacteria: Focus on Beneficial Activities. *Marine Drugs* **2019** 17(6), 320.
10. Welker, M., Dittmann, E., Von Döhren, H. Cyanobacteria as a source of natural products. *Methods Enzymol.* **2012** 517, 23-46. doi:10.1016/B978-0-12-404634-4.00002-4
11. Jones, A., Gu, L., Sorrels, C., Sherman, D., Gerwick, W. New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. *Current Opinion in Chemical Biology* **2009** 13, 216-223.
12. Nikolouli K, Mossialos D. Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. *Biotechnol Lett.* **2012** 34(8), 1393-403.
13. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., ... & Cotter, P. D. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural product reports* **2013** 30(1), 108-160.
14. Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K.W., Han, C.S., Rubin, E.M., Eisen, J. a, Woyke, T., Gugger, M., Kerfeld, C. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **2013** 110, 1053-8. doi:10.1073/pnas.1217107110
15. Rounge, T. B., Rohrlack, T., Tooming-Klunderud, A., Kristensen, T., & Jakobsen, K. S. Comparison of cyanopeptolin genes in *Planktothrix*, *Microcystis*, and *Anabaena* strains: evidence for independent evolution within each genus. *Applied and environmental microbiology* **2007** 73(22), 7322-7330.
16. Donia, M. S., & Schmidt, E. W. Linking chemistry and genetics in the growing cyanobactin natural products family. *Chemistry & Biology* **2011** *18*(4), 508-519.
17. Starcevic, A., Zucko, J., Simunkovic, J., Long, P. F., Cullum, J., & Hranueli, D. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic acids research* **2008** 36(21), 6882-6892.
18. Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Brown, L. C. W., Mavrommatis, K., et al., Birren, B. W. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014** 158(2), 412-421.
19. Welker, M., & Erhard, M. Consistency between chemotyping of single filaments of *Planktothrix rubescens* (cyanobacteria) by MALDI-TOF and the peptide patterns of strains determined by HPLC-MS. *Journal of Mass Spectrometry* **2007** 42(8), 1062-1068.
20. Wang, Mingxun, *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* **2016** 34.8, 828-837.
21. Boudreau, P., Monroe, E., Mehrotra, S., Desfor, S., Korabeynikov, A., Sherman, D., Murray, T., Gerwick, L., Dorrestein, P., Gerwick, W. Marine cyanobacterium *Moorea producens* JHB through

519 orthogonal natural products workflows. *PLoS ONE* **2015** 10:e0133297.
520 22. Mohimani, H., & Pevzner, P. A. Dereplication, sequencing and identification of peptidic natural
521 products: from genome mining to peptidogenomics to spectral networks. *Natural product reports*
522 **2016** 33(1), 73-86.
523 23. Kurmayer, R., Blom, J. F., Deng, L., & Pernthaler, J. Integrating phylogeny, geographic niche
524 partitioning and secondary metabolite synthesis in bloom-forming *Planktothrix*. *The ISME*
525 *journal* 2015 9(4), 909.
526 24. Christiansen, G., Goesmann, A., & Kurmayer, R. Elucidation of insertion elements carried on
527 plasmids and in vitro construction of shuttle vectors from the toxic cyanobacterium *Planktothrix*.
528 *Appl. Environ. Microbiol.* **2014** 80(16), 4887-4897.
529 25. Pancrace, C., Barny, M. A., Ueoka, R., Calteau, A., Scalvenzi, T., Pédron, J., ... & Gugger, M.
530 Insights into the *Planktothrix* genus: Genomic and metabolic comparison of benthic and planktic
531 strains. *Scientific reports* **2017** 7, 41181.
532 26. Rounge, T. B., Rohrlack, T., Nederbragt, A. J., Kristensen, T., & Jakobsen, K. S. A genome-wide
533 analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix*
534 *rubescens* strain. *BMC genomics* **2009** 10(1), 396.
535 28. Rohrlack, T., Edvardsen, B., Skulberg, R., Halstvedt, C. B., Utkilen, H. C., Ptacnik, R., &
536 Skulberg, O. M. Oligopeptide chemotypes of the toxic freshwater cyanobacterium *Planktothrix*
537 can form sub-populations with dissimilar ecological traits. *Limnology and Oceanography* **2008**
538 53(4), 1279-1293.
539 27. Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Dickschat, J. S.
540 AntiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary
541 identification. *Nucleic acids research* **2017** 45(1), 36-41.
542 29. Christiansen, G., Fastner, J., Erhard, M., Börner, T., & Dittmann, E. (2003). Microcystin
543 biosynthesis in *Planktothrix*: genes, evolution, and manipulation. *Journal of bacteriology* **2003**
544 185(2), 564-572.
545 30. Ishida, K., Christiansen, G., Yoshida, W.Y., Kurmayer, R., Welker, M., Valls, N., Bonjoch, J.,
546 Hertweck, C., Börner, T., Hemscheidt, T. and Dittmann, E. Biosynthesis and structure of
547 aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-
548 hydroxyoctahydroindole moiety. *Chemistry & biology* **2007** 14(5), 565-576.
549 31. Christiansen, G., Philmus, B., Hemscheidt, T., & Kurmayer, R. Genetic variation of adenylation
550 domains of the anabaenopeptin synthesis operon and the evolution of substrate promiscuity.
551 *Journal of bacteriology*, *Journal of bacteriology* **2011** 193(15), 3822-3831.
552 32. Philmus, B., Christiansen, G., Yoshida, W. Y., & Hemscheidt, T. K. Post-translational
553 modification in microviridin biosynthesis. *ChemBioChem* **2008** 9(18), 3066-3073.
554 33. Olivon, F., Allard, P.M., Koval, A., Righi, D., Genta-Jouve, G., Neyts, J., Apel, C., Pannecouque,
555 C., Nothias, L.F., Cachet, X. and Marcourt, L. Bioactive natural products prioritization using
556 massive multi-informational molecular networks. *ACS chemical biology* **2017** 12(10), pp.2644-2651.
557 34. Yang, J.Y., Sanchez, L.M., Rath, C.M., Liu, X., Boudreau, P.D., Bruns, N., Glukhov, E., Wodtke,
558 A., De Felicio, R., Fenner, A., Wong, W.R., Linington, R.G., Zhang, L., Debonsi, H.M., Gerwick,
559 W.H., Dorrestein, P.C. Molecular networking as a dereplication strategy. J. *Nat. Prod.* **2013** *76*,
560 1686-1699. doi:10.1021/np400413s
561 35. Brito, Â., Gaifem, J., Ramos, V., Glukhov, E., Dorrestein, P.C., Gerwick, W.H., Vasconcelos,
562 V.M., Mendes, M.V. and Tamagnini, P. (2015) Bioprospecting Portuguese Atlantic coast

cyanobacteria for bioactive secondary metabolites reveals untapped chemodiversity. *Algal Research* **2015** 9, 218-226.

36. Ding, C., Pang, L., Liang, Z. X., Goh, K., Glukhov, E., Gerwick, W., & Tan, L. MS/MS-Based Molecular Networking Approach for the Detection of Aplysiatoxin-Related Compounds in Environmental Marine Cyanobacteria. *Marine Drugs* **2018** 16(12), 505.

37. Luzzatto-Knaan, T., Garg, N., Wang, M., Glukhov, E., Peng, Y., Ackermann, G., ... & Knight, R. Digitizing mass spectrometry data to explore the chemical diversity and distribution of marine cyanobacteria and algae. *eLife* 2017 6, e24214.

38. Allard, P.M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V.C., Roussi, F., Litaudon, M. and Wolfender, J.L. Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Analytical chemistry* **2016** 88(6), 3317-3323.

39. Allard, P. M., Genta-Jouve, G., & Wolfender, J. L. Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification. *Current opinion in chemical biology* **2017** 36, 40-49.

40. Le Manach, S., Marie, B., Duval, C., Marie, A., Djediat, C., Catherine, A., ... , Marie, B. Global metabolomic characterizations of *Microcystis* spp. highlights clonal diversity in natural bloom-forming populations and expands metabolite structural diversity. *Frontiers in microbiology* **2019** 10, 791.

41. Welker, M., Christiansen, G., & von Döhren, H. Diversity of coexisting *Planktothrix* (Cyanobacteria) chemotypes deduced by mass spectral analysis of microystins and other oligopeptides. *Archives of microbiology* **2004** 182(4), 288-298.

42. Ishida, K., Welker, M., Christiansen, G., Cadel-Six, S., Bouchier, C., Dittmann, E., Hertweck, C. and De Marsac, N.T. Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria. *Applied and environmental microbiology* **2009** 75(7), 2017-2026.

43. Haruštiaková, D., & Welker, M. Chemotype diversity in *Planktothrix rubescens* (cyanobacteria) populations is correlated to lake depth. *Environmental microbiology reports* **2017** 9(2), 158-168.

44. Briand, J. F., Jacquet, S., Flinois, C., Avois-Jacquet, C., Maisonnette, C., Leberre, B., & Humbert, J. F. (2005). Variations in the microcystin production of *Planktothrix rubescens* (Cyanobacteria) assessed from a four-year survey of Lac du Bourget (France) and from laboratory experiments. *Microbial ecology* **2005** 50(3), 418-428.

45. Yéprémian C., Gugger M.F., Briand E., Catherine A., Berger C., Quiblier C., Bernard C. Microcystin ecotypes in a perennial *Planktothrix* agardhii bloom. *Water Res.* **2007** 41(19):4446-56.

46. Kurmayer, R., Deng, L., & Entfellner, E. Role of toxic and bioactive secondary metabolites in colonization and bloom formation by filamentous cyanobacteria *Planktothrix*. *Harmful Algae* **2016** 54, 69-86.

47. Harke, M.J., Steffen, M.M., Gobler, C.J., Otten, T.G., Wilhelm, S.W., Wood, S.A., Paerl, H.W. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. *Harmful Algae* **2016** 54, 4-20.

48. Catherine, A., Bernard, C., Spoof, L., Bruno, M. Microcystins and Nodularins. In *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis* - eds J. Meriluoto, L. Spoof and G. A. Codd, **2017**. doi:10.1002/9781119068761.ch11

49. Miles, C. O. *Toxinmasslist COM v15b* [Data set]. Unpublished **2018**. https://doi.org/10.13140/RG.2.2.15794.50889

50. Entfellner, E., Frei, M., Christiansen, G., Deng, L., Blom, J., & Kurmayer, R. Evolution of anabaenopeptin peptide structural variability in the cyanobacterium *Planktothrix*. *Frontiers in microbiology* **2017** 8, 219.

609 51. Welker, M., Von Döhren, H. Cyanobacterial peptides - Nature's own combinatorial
610 biosynthesis. *FEMS Microbiol. Rev.* **2006** 30, 530-563. doi:10.1111/j.1574-6976.2006.00022.x
611 52. Briand, E., Reubrecht, S., Mondeguer, F., Sibat, M., Hess, P., Amzil, Z., & Bormans, M.
612 Chemically mediated interactions between *Microcystis* and *Planktothrix*: impact on their growth,
613 morphology and metabolic profiles. *Environmental Microbiology* **2019** 21, 1552-1566. doi:
614 10.1111/1462-2920.14490.
615 53. Moss, N. A., Bertin, M. J., Kleigrewe, K., Leão, T. F., Gerwick, L., & Gerwick, W. H. Integrating
616 mass spectrometry and genomics for cyanobacterial metabolite discovery. *Journal of industrial*
617 *microbiology & biotechnology* **2016** 43(2-3), 313-324.
618 54. Duncan, K. R., Crüsemann, M., Lechner, A., Sarkar, A., Li, J., Ziemert, N., *et al.*, Jensen, P. R.
619 Molecular networking and pattern-based genome mining improves discovery of biosynthetic
620 gene clusters and their products from *Salinispora* species. *Chemistry & biology* **2015** 22(4), 460-471.
621 55. Hautbergue, T., Jamin, E. L., Debrauwer, L., Puel, O., & Oswald, I. P. From genomics to
622 metabolomics, moving toward an integrated strategy for the discovery of fungal secondary
623 metabolites. *Natural product reports* **2018** 35(2), 147-173.
624 56. Maansson, M., Vynne, N.G., Klitgaard, A., Nybo, J.L., Melchiorsen, J., Nguyen, D.D., Sanchez,
625 L.M., Ziemert, N., Dorrestein, P.C., Andersen, M.R. and Gram, L. An integrated metabolomic
626 and genomic mining workflow to uncover the biosynthetic potential of bacteria. *Msystems* 2016
627 1(3), e00028-15.
628 57. Kleigrewe, K., Almaliti, J., Tian, I. Y., Kinnel, R. B., Korobeynikov, A., Monroe, E. A., ... &
629 Gerwick, L. Combining mass spectrometric metabolic profiling with genomic analysis: a
630 powerful approach for discovering natural products from cyanobacteria. *Journal of natural*
631 *products* 2015 78(7), 1671-1682.
632 58. Leikoski, N., Liu, L., Jokela, J., Wahlsten, M., Gugger, M., Calteau, A., Permi, P., Kerfeld, C.A.,
633 Sivonen, K. and Fewer, D.P. Genome mining expands the chemical diversity of the cyanobactin
634 family to include highly modified linear peptides. *Chemistry & Biology* **2013** 20(8), 1033-1043.
635 59. Beauxis, Y., & Genta-Jouve, G. Metwork: a web server for natural products anticipation.
636 *Bioinformatics* **2019** 35, 1795-1796.
637 60. Behsaz, B., Mohimani, H., Gurevich, A., Prjibelski, A., Fisher, M., Smarr, L., *et al.*, Pevzner, P.A.
638 *De novo* peptide sequencing reveals a vast cyclopeptidome in human gut and other
639 environment. *BioRxiv* **2019** 521872.
640 61. Humbert, J. F., Barbe, V., Latifi, A., Gugger, M., Calteau, A., Coursin, T., Lajus, A., Castelli, V.,
641 Oztas, S., Samson, G., Longin, C., Medigue, C., and de Marsac, N. T. A tribute to disorder in the
642 genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*, *PLoS One* **2013**
643 8, e70747.
644 62. Médigue, C., Calteau, A., Cruveiller, S., Gachet, M., Gautreau, G., Josso, A., Lajus, A., Langlois,
645 J., Pereira, H., Planel, R. and Roche, D. MicroScope—an integrated resource for community
646 expertise of gene functions and comparative analysis of microbial genomic and metabolic data.
647 *Brief Bioinform.* **2017** *Sep 12. doi: 10.1093/bib/bbx113.*