Mining of Cyanobacterial Genomes Indicates That Plasmids Are Involved in the Production of Natural Products

Rafael Popin
University of Helsinki

Danillo Alvarenga
University of Helsinki

Raquel Castelo-Branco
University of Porto

David Fewer
University of Helsinki

Kaarina Sivonen (✉ kaarina.sivonen@helsinki.fi)
University of Helsinki

Research Article

Keywords: Cyanobacteria, Microbial, biosynthetic

DOI: https://doi.org/10.21203/rs.3.rs-121751/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background** Microbial natural products have unique chemical structures and diverse biological activities. Cyanobacteria commonly possess a wide range of biosynthetic gene clusters to produce natural products. Several studies have mapped the distribution of natural product biosynthetic gene clusters in cyanobacterial genomes. However, little attention has been paid to natural product biosynthesis in plasmids. Some genes encoding cyanobacterial natural product biosynthetic pathways are believed to be dispersed by plasmids through horizontal gene transfer. Thus, we examined complete cyanobacterial genomes to assess if plasmids are involved in the production and dissemination of natural products by cyanobacteria.

**Results** The 185 analyzed genomes possessed 1 to 42 gene clusters and an average of 10. In total, 1816 biosynthetic gene clusters were found. Approximately 95% of these clusters were present in chromosomes. The remaining 5% were present in plasmids, from which homologs of the biosynthetic pathways for aeruginosin, anabaenopeptin, ambiguine, cryptophycin, hassallidin, geosmin, and microcystin were manually curated. The cryptophycin pathway was previously described as active while the other gene cluster include all genes for biosynthesis. Approximately 12% of the 424 analyzed cyanobacterial plasmids contained homologs of genes involved in conjugation. Large plasmids, previously named as “chromids”, were also observed to be widespread in cyanobacteria. Sixteen cryptic natural product biosynthetic gene clusters and geosmin biosynthetic gene clusters were located in those mobile plasmids.

**Conclusion** Homologues of genes involved in the production of toxins, protease inhibitors, odorous compounds, antimicrobials, antitumors, and other unidentified natural products are located in cyanobacterial plasmids. Some of these plasmids are predicted to be conjugative. The present study provides in silico evidence that plasmids are involved in the distribution of natural product biosynthetic pathways in cyanobacteria.

Background

Microbial natural products originate in secondary metabolism and exhibit a wide diversity of chemical structures and biological activities [1]. These metabolites can act as antibiotics, anticancer agents, antivirals, toxins, and find applications such as enzyme inhibitors, polymers or surfactants [2]. The enzymes involved in the biosynthesis of natural products are commonly encoded in biosynthetic gene clusters located in contiguous stretches of DNA known as biosynthetic gene clusters (BGCs) [3–5]. BGCs usually include genes for core biosynthesis and tailoring enzymes and regulatory and resistance genes [6, 7]. Among the accessory enzymes, 4-phosphopantetheinyl transferases (PPTs) play a major role in the biosynthesis of several natural products through the conversion of inactive apo-proteins into their active holo-forms [8–10].

Understanding of the genetic diversity and distribution of BGCs has greatly increased due to the enormous expansion in the number of sequenced bacterial genomes in the last decade [3, 11]. Genome mining, which uses bioinformatics techniques to identify genes encoding enzymes possibly involved in biosynthesis of natural products, has promoted the discovery of novel compounds [12–14]. The Cyanobacteria are among several phyla of bacteria that are commonly explored through these techniques [15].

Since the genome of *Synechocystis* sp. PCC 6803 was sequenced in 1996 [16], the number of complete cyanobacterial genomes deposited in NCBI GenBank has slowly increased in comparison to other bacteria [17]. Despite their underrepresentation in public databases, cyanobacterial genomes were successfully investigated from evolutionary, ecological, and taxonomic perspectives [18]. Cyanobacteria are recognized as a source of diverse natural products with applications in pharmacology, biotechnology, and bioenergy production [19–21]. A considerable portion of these molecules are non-ribosomal peptides and polyketides produced by non-ribosomal peptide synthetase (NRPSs) and
polyketide synthase (PKSs), respectively [22, 23]. Other classes of broadly distributed cyanobacterial natural products include, among others, ribosomally synthesized and post-translationally modified peptides (RiPPs), alkaloids, pigments, and terpenoids [24–27]. Genome mining of cyanobacterial genomes has helped unravel the diversity of BGCs involved in the production of various natural products [15, 28, 29].

Cyanobacterial natural product BGCs are mostly concentrated in the genomes of late-branching cyanobacteria, mainly in the orders Oscillatoriales and Nostocales, although these BGCs are found in almost all cyanobacterial genomes [22, 30, 31]. Several studies have attempted to map the distribution of BGCs in these organisms [22]. However, little attention has been given in cyanobacterial studies on whether these BGCs are located in chromosomes or plasmids [22, 29–32]. Horizontal gene transfer (HGT) events are linked to the dissemination of many cyanobacterial natural product BGCs, including toxins that belong to the cylindrospermopsin, microcystin, anatoxin-a, and saxitoxin families [33–35].

Plasmids play a key role in HGT, and conjugation is one of the processes that can transfer genetic material [36, 37]. The most frequent mechanism of DNA conjugation in gram-negative bacteria involves a relaxome, which includes a relaxase and a type IV coupling protein (T4CP) encoded by mobility genes (MOB), and a transferosome assembled by a type IV secretion system (T4SS) that is encoded by mating pair formation genes (MPF) [38, 39]. During conjugation, the relaxase cleaves and covalently binds itself to the transferring DNA on a site called oriT [40]. The T4SS is believed to then act as a secretor protein by transferring DNA and the relaxase to the recipient cell [41]. For this purpose, the T4CP recognizes, energizes, and delivers the nucleoprotein to the T4SS [42]. Plasmids encoding these three components are called self-transmissible or conjugative, while mobilizable plasmids usually encode just the MOB and a T4SS and are transmitted only in the presence of a helper conjugative plasmid [40]. Although the majority of the cyanobacterial plasmids were found to lack all the necessary genes to be conjugative [39], no concomitant analysis of the presence of BGCs in plasmids and their mobility is currently available for cyanobacteria.

Thus, the present study screened 184 complete genomes publicly available in the GenBank database [43] from the phylum Cyanobacteria and one from Candidatus Melainabacteria, a phylum that is closely related to cyanobacteria [44]. The compartmentalization of natural product BGCs and the key enzymes PPTs in chromosomes and plasmids were investigated. Moreover, the mobility of plasmids and the phylogeny of known BGCs were predicted. We found evidence that plasmids are involved in the production of several natural products and HGT of BGCs in Cyanobacteria.

Results

According to the latest proposed system of cyanobacteria, approximately 37% of the analyzed genomes belong to the order Synechococcales (mainly Synechococcus and Prochlorococcus), followed by Nostocales (26% of genomes; the genus Nostoc alone was approximately 11% of the total dataset) (Fig. 1). The remaining 37% of the genomes were distributed in the orders Gloeobacterales, Gloeomargaritales, Synechococcales, Pleurocapsales, and Chroococcidiopsidales. No representative genome of the order Spirulinales was analyzed here due to unavailability.

General features of the evaluated genomes

From the 52 genera represented in the retrieved dataset, 27 included more than one genome. Thus, these genera were used for the calculation of averages and standard deviation of genomic characteristics (Table 1). These cyanobacterial genomes consisted of up to two chromosomes and 14 plasmids (Table S1). Chromosomes consisted of 97–100% of genomic DNA, while plasmids represented up to 3% (Table S2). Genome sizes ranged from 1.65 to 12.05 Mb, GC content from 30.8–68.7%, and the number of genes varied from 1816 to 11674 (Table S1). The number of BGCs in chromosomes ranged from 3 to 42; up to five were found encoded in plasmids.
Table 1

Genome statistics of cyanobacterial genera with more than one complete sequence in NCBI GenBank [43]. Averages and standard deviations of genome size, GC content, number of genes, BGCs in the chromosomes and plasmids, and the total number of BGCs in the genome were calculated. Gen = genome, Pld = plasmid, BGC = biosynthetic gene cluster, Chr = Chromosome.

| Order          | Genus          | No. Gen | Pld/Gen | Size (Mb)       | GC (%)   | Genes          | No. BGCs Chr | No. BGCs Pld | Total BGCs |
|----------------|----------------|---------|---------|----------------|----------|----------------|--------------|--------------|------------|
| Gloeobacterales| Gloeobacter     | 2       | 0       | 4.69 ± 0.04    | 61.3 ± 1.1| 4497 ± 21     | 5 ± 3        | 0            | 7 ± 1      |
| Synechococcales| Cyanobium       | 2       | 0       | 3.18 ± 0.23    | 68.7 ± 0.1| 3227 ± 251    | 8 ± 1        | 0            | 8 ± 1      |
|                | Leptolyngbya    | 6       | 2 ± 2   | 6.37 ± 0.85    | 47.9 ± 4.1| 5890 ± 965    | 10 ± 4       | 1 ± 1        | 11 ± 5     |
|                | Prochlorococcus | 17      | 1 ± 0   | 1.83 ± 0.29    | 34.8 ± 6.3| 2018 ± 327    | 6 ± 5        | 1 ± 0        | 6 ± 5      |
|                | Pseudanabaena   | 2       | 4 ± 4   | 5.28 ± 0.54    | 44.2 ± 2.8| 4510 ± 789    | 3 ± 1        | 1 ± 0        | 4 ± 3      |
|                | Synechococcus   | 28      | 0 ± 1   | 2.79 ± 0.57    | 57.2 ± 5.4| 2847 ± 452    | 6 ± 4        | 0            | 5 ± 4      |
|                | Synechocystis   | 8       | 2 ± 3   | 3.72 ± 0.18    | 47.6 ± 0.2| 3465 ± 200    | 3 ± 0        | 0            | 3 ± 0      |
|                | Thermosynechococcus | 5    | 0       | 2.55 ± 0.06    | 53.7 ± 0.3| 2543 ± 67     | 3 ± 0        | 0            | 3 ± 0      |
| Oscillatoriales| Arthrospira     | 3       | 0       | 6.47 ± 0.35    | 44.4 ± 0.3| 8118 ± 3081   | 3 ± 1        | 0            | 3 ± 1      |
|                | Moorea          | 2       | 2 ± 1   | 9.55 ± 0.23    | 43.6 ± 0.1| 7748 ± 31     | 42 ± 0       | 0            | 42 ± 0     |
|                | Oscillatoria    | 2       | 4 ± 2   | 8.04 ± 0.33    | 46.7 ± 1.3| 6479 ± 590    | 8 ± 1        | 1 ± 0        | 9 ± 2      |
|                | Planktothrix    | 2       | 5 ± 1   | 5.07 ± 0.02    | 39.6 ± 0.1| 4538 ± 13     | 6 ± 4        | 2 ± 0        | 11 ± 1     |
|                | Synechococcus   | 7       | 5 ± 1   | 3.14 ± 0.12    | 49.2 ± 0.1| 3183 ± 117    | 3 ± 0        | 1 ± 1        | 4 ± 1      |
| Chroococcales  | Cyanobacterium  | 2       | 1 ± 1   | 3.23 ± 0.10    | 38.2 ± 0.7| 3020 ± 251    | 4 ± 1        | 0            | 4 ± 1      |
| Order       | Genus        | No. Gen | Pld/Gen | Size (Mb) | GC (%) | Genes No. BGCs | Chr No. BGCs | Pld | Total BGCs |
|-------------|--------------|---------|---------|-----------|--------|----------------|--------------|-----|------------|
|             | Gloeothecae  | 2       | 6 ± 0   | 7.20 ± 0.91 | 39.2 ± 1 | 6406 ± 795 | 8 ± 1       | 5 ± 4 | 13 ± 2     |
|             | Geminocystis | 3       | 8 ± 4   | 4.24 ± 0.20  | 33.3 ± 1.0 | 3831 ± 270 | 4 ± 1       | 0    | 4 ± 1      |
|             | Microcystis  | 7       | 0       | 5.19 ± 0.65  | 42.5 ± 0.3  | 5243 ± 752 | 10 ± 1     | 0    | 10 ± 1     |
|             | Rippkaea     | 2       | 4 ± 1   | 4.80 ± 0.01  | 39.8 ± 0.0  | 4540 ± 30 | 8 ± 0       | 0    | 8 ± 0      |
| Pleurocapsales | Stanieria    | 2       | 3 ± 3   | 5.50 ± 0.06  | 36.4 ± 0.2  | 4948 ± 30 | 11 ± 2     | 1 ± 0 | 12 ± 3     |
| Nostocales  | Anabaena     | 5       | 3 ± 3   | 6.44 ± 0.83  | 39.1 ± 1.3  | 5713 ± 558 | 14 ± 3     | 1 ± 1 | 14 ± 3     |
|             | Calothrix    | 11      | 4 ± 3   | 8.70 ± 2.11  | 40.2 ± 1.5  | 7273 ± 1957 | 15 ± 6   | 1 ± 0 | 15 ± 6     |
|             | Cylindrospermum | 2       | 4 ± 1   | 7.66 ± 0.06  | 42.1 ± 0.1  | 6574 ± 65 | 21 ± 3     | 3 ± 2 | 24 ± 1     |
|             | Dolichospermum | 2       | 2 ± 2   | 5.41 ± 0.33  | 38.2 ± 0.0  | 5032 ± 349 | 11 ± 4     | 0    | 11 ± 4     |
|             | Fischerella  | 3       | 5 ± 4   | 6.54 ± 1.00  | 40.5 ± 0.7  | 5547 ± 899 | 13 ± 1     | 5 ± 0 | 15 ± 3     |
|             | Nodularia    | 2       | 1 ± 1   | 5.43 ± 0.05  | 41.2 ± 0    | 4866 ± 42 | 11 ± 1     | 0    | 11 ± 1     |
|             | Nostoc       | 20      | 5 ± 3   | 7.88 ± 1.22  | 41 ± 0.8    | 6824 ± 1096 | 16 ± 5   | 2 ± 2 | 18 ± 6     |
|             | Scytonema    | 2       | 6 ± 2   | 9.81 ± 0.06  | 43.6 ± 0.2  | 8176 ± 76 | 25 ± 1     | 2 ± 1 | 27 ± 2     |
|             | Trichormus   | 2       | 4 ± 1   | 7.29 ± 0.25  | 41.4 ± 0    | 6147 ± 291 | 14 ± 1     | 2 ± 1 | 15 ± 1     |

**Biosynthetic potential**

A total of 1816 BGCs were identified; approximately 10 per genome were identified (Table S1). *Synechococcus* sp. JA-2-3B’a(2–13), *Candidatus* Melainabacteria MEL.A1, and *Synechococcus* sp. JA-3-3Ab had only one BGC and thus were the genomes with the lowest number of BGCs. In contrast, *Moorea producens* JHB and *Moorea producens* PAL-8-15-08-
1 had 42 BGCs each (Fig. S1). Nostocales genomes were among those with the highest average number of BGCs. The number of BGCs appears to correlate with the genome size (Fig. 2).

Most BGCs were identified in chromosomes (1719) and were 95% of the total (1818). From these, RiPPs were the most widespread class of BGC products (526 representatives, approximately 31% of the chromosomal BGCs). Terpenes were the second most widespread products (470 representatives, approximately 27% of the BGCs in chromosomes) and were absent only in Arthrospira platensis C1, Candidatus Melainabacteria bacterium MEL.A1, and Nostoc sphaeroides CCNUC1. PKS was the least frequent class, with only 49 representatives.

A total of 424 plasmids were identified from the 185 genomes, of which 73 (approximately 17%) plasmids had at least one BGC (Table S1). Ninety-nine BGCs were found in cyanobacterial plasmids. While most plasmids encoded only one BGC, Gloeothece verrucosa PCC 7822 Cy782201 (0.88 Mb) encoded five BGCs (Figure S2).

Hybrid NRPS/PKS corresponded to more than half of the natural product BGCs located on the plasmid (26 BGCs) and were the most common. NRPS consisted of 20 representatives, followed by bacteriocin (15 representatives). In contrast to chromosomes, terpenes were one of the least frequent products with BGCs encoded on plasmids.

**Known biosynthetic pathways on plasmids**

Several of the analyzed complete genomes revealed large plasmids (here considered as > 500 kb), with sizes reaching a maximum of approximately 2.5 Mb in Stanieria cyanosphaera PCC 7437 plasmid pSTA7437.02 (Table S2). These large replicons contained one to five BGCs, which is greater than the remainder of the analyzed plasmid that presented a maximum of three. The pathways of the known natural products, including ambiguine (Fischerella sp. NIES-4106 plasmid2, 550 kb), anabaenopeptin (Scytonema sp. HK-05 plasmid1, 831 kb), geosmin (Nostoc linckia NIES-25 plasmid1, 1.78 Mb), and microcystin (Fischerella sp. NIES-4106 plasmid1, 550 kb) were identified in these large plasmids (Fig. S3, Table S3). Other known BGCs were located in smaller plasmids; for example, i.e., the aeruginosin gene cluster was present in Cylindrospermum sp. NIES-4074 plasmid1 (340 kb); anabaenopeptin in Gloeothece citiformis PCC 7424 plasmid pP742401 (328 kb) and Gloeothece verrucosa PCC 7822 plasmid Cy782201 (879 kb), cryptophycin in Nostoc sp. ATCC 53789 plasmid pNSP_c (219 kb), geosmin in Nostoc sp. NIES-2111 plasmid2 (320 kb), and hassallidin in Aulosira laxa NIES-50 plasmid1 (292 kb) and Tolypothrix tenuis PCC 7101 plasmid1 (292 kb).

A 16S rRNA phylogenetic analysis (Fig. 3A) was compared with phylogenetic trees built with manually curated known BGCs found in plasmids and chromosomes (Fig. 3B-E). The geosmin BGC tree shows that plasmid BGCs appear to share recent ancestors with chromosomal BGCs but tend to form their own cluster that are incongruent with 16S phylogeny. Thus, these 2 BGCs from Nostoc plasmids might have been transferred through HGT and possibly face different evolutionary pressure than geosmin BGCs present in chromosomes. In contrast, there is no evidence of HGT in hassallidin and anabaenopeptin BGC trees. Insufficient information is available for the microcystin BGC, as a single cluster was found in plasmids (Fig. 3D).

**Distribution of 4-phosphopantetheinyl transferases**

A total of 193 PPTs were found (Table S4). From the 185 complete genomes analyzed here, 155 had at least one copy of PPT homologs (approximately 84%). The majority (148) encoded only one enzyme, while 6 genomes encoded two enzymes and 1 genome (Halomicronema hongdechloris C2206) had three copies of the gene. The size of these enzymes ranged from 137 (one of the two copies in Chroococcidiopsis thermalis PCC 7203) to 339 aa (one of the three copies in Halomicronema hongdechloris C2206). However, approximately 90% (147 enzymes) of the enzymes ranged between 200 to 280 aa. The genome of Acaryochloris marina MBIC11017 included a PPT in the genome and another in
plasmid pREB1. This translated enzyme sequence was more similar to an AcpS-type PPT gene than a Spf-type from *Bacillus subtilis*. The remaining 162 PPTases in the cyanobacterial genomes were likely Spf-type.

**Homologs of proteins involved in conjugation**

The 424 plasmids were searched for the presence of the relaxase gene homologs VirB4 and VirD4 encoded in *Nostoc* sp. PCC 7120 pCC7120alpha (Fig. 4). The presence of these three genes was used to predict the transmissibility of the plasmids (Table S5).

Only 23 plasmids from Nostocales (genera *Calothrix, Cylindropermum, Fischerella, Nodularia, Nostoc, Trichormus*) and 1 from Oscillatoriales (*Microcoleus*), or approximately 6% of the analyzed cyanobacterial plasmids, encoded the three proteins and are possibly conjugative. Another plasmid from Chroococcales (*Gloeocapsa*), 24 from Nostocales (*Anabaena, Aulosira, Calothrix, Nostoc, Trichormus, Tolypothrix*), and 1 from Oscillatoriales (*Crinalium*), or approximately 6% of the evaluated cyanobacterial plasmids, were predicated to be mobilizable as they encoded the relaxase but not Virb4 and VirD4. The remaining 375 (88%) plasmids encoded either Virb4 or VirD4 only and thus are likely immobile according to the current model. Among the plasmids encoding known natural products, only *Nostoc linckia* NIES-2111 plasmid2, which has the geosmin BGC, is predicted to be conjugative (Table S6).

**Phylogenomics**

Overall, later-branching cyanobacteria from Nostocales and also Oscillatoriales and Chroococcales had more natural product BGCs than other orders (Fig. 5). The gene clusters encoded by these three orders were from all analyzed classes. In contrast, early branching cyanobacteria, especially those from Gloeobacterales, Synechococcales, and Gloeomargaritales, had fewer natural product BGCs (mainly terpenes and RiPPs) than those of the other analyzed orders. This pattern of distribution also applied to mobile plasmids. PPTs were distributed in all analyzed cyanobacterial orders.

**Discussion**

Based on the model gram-negative bacteria *Escherichia coli*, cyanobacteria are commonly assumed to be monoploid [50]. However, these organisms can become oligoploid during rapid growth [51, 52]. Cyanobacteria may also contain several chromosome copies throughout their life cycles [50, 53, 54]. Previous studies reported polyploidy in several genera, such as *Anabaena (Nostoc)* [55], *Synechococcus* [56], and *Synechocystis* [57]. Here, *Nostoc sphaeroides* CCNUC1, *Crocosphaera subtropica* ATCC 51142 [58] and *Anabaena (Dolichospermum)* sp. 90 [59] were found to possess extra chromosomes. In contrast, large plasmids, or previously named “chromids” (considered here as > 500 kb), occurred in several genomes and were found in 11 genomes from four of the six analyzed cyanobacterial orders.

Chromids are large, plasmid-like replicons that were previously found in approximately 10% of bacterial genomes [48]. Chromids possess replication systems that are similar to plasmids and can carry essential genes for cell viability [60]. One of the proposed functions of these large replicons is to increase genome plasticity through the rapid acquisition or loss of genes by HGT [61]. Here, chromids occurred in approximately 15% of the analyzed cyanobacterial genomes, and therefore seem to be more widespread in cyanobacteria than in other phyla [48].

*Nostoc* sp. strain ATCC 53789 is a known producer of the antiproliferative cytotoxin cryptophycin, which is encoded in a plasmid [62, 63]. Other plasmidial BGCs found here, such as the hepatotoxin microcystin, antifungal hassallidin, and odorous terpenoid geosmin, contained all the core genes and are possibly functional [64–66]. Consistent with our results, plasmids have been previously shown to contain genes encoding RiPPs and are associated with the products of
these toxic and odorous compounds [29, 67]. Toxins produced by other bacteria, such as botulinum from *Clostridium botulinum* and cereulide from *Bacillus cereus*, are also found on plasmids [68, 69]. In the case of the botulinum toxin, HGT of the botulinum gene cluster by conjugative plasmids < 200 kb is likely [70].

Only the plasmid pCC7120α from *Nostoc* sp. PCC7120 has been reported to be transmissible [71]. Nevertheless, our results indicate that other cyanobacterial plasmids are possibly conjugative. A previous study using automatic annotation found no homologs of the T4SS in cyanobacteria and hypothesized that an unknown mechanism of conjugation could be present in these organisms [72]. It is currently unclear whether cyanobacterial plasmids are predominantly immobile, unlike in other bacterial phyla, due to the reduced availability of cyanobacterial genome sequences [72].

NRPSs and PKSs are constituted by multi-domains that have specific functions in the biosynthetic pathways of polyketides and non-ribosomal peptides [73, 74]. While the core module of an NRPS consists of at least the adenylation, condensation, and peptidyl carrier protein modules, acyltransferase, acyl carrier protein, and a ketoacyl synthase are the core domains of a PKS [75, 76]. Thus, carrier proteins, such as the PPTs, are essential for the biosynthesis of these natural products. Two main families of PPTs are known, namely AcpS-type PPTs, which are involved in activating carrier proteins involved in the primary metabolism, and Sfp-like PPTs, which are involved in secondary metabolism pathways [9, 77].

In cyanobacteria, only one copy of Sfp-like PPTs had previously been found in 29 different genomes [78]. However, the present study revealed that some cyanobacterial genomes can encode up to three different PPTs. Other bacteria also contain multiple copies of these enzymes in their genomes [10, 79]. Interestingly, the fact that a plasmid from *Acaryochloris marina* MBIC11017 was the only representative of an AcpS-like PPT indicates that this enzyme could possibly be transferred horizontally together with BGCs. Consistent with our results, plasmids from other bacterial phyla have also been found to encode PPTs [80, 81].

RiPPs gene clusters were located in almost all analyzed genomes. These molecules are products of post-translational modification of ribosomally synthesized precursor peptides [82]. Thus far, over 20 families of compounds that possess unique chemical features have been proposed [82]. Cyanobacteria encode the machinery to produce several RiPPs, including cyanobactins [83], lanthipeptides [84], lassopeptides [85], and microviridins [86]. Although cyanobactin BGCs are widespread in cyanobacteria and initially received the most attention, other RiPPs from cyanobacteria are also being explored [29, 84]. Considering that automated tools are being improved to better predict genes involved in the biosynthesis of these compounds, future studies may expand the known repertoire of RiPPs produced by cyanobacteria [47, 87].

Although terpenes are commonly isolated from plants and fungi, genes involved in their biosynthesis are widely found in bacterial genomes [88]. These compounds are essential in primary metabolism, such as for photosynthesis and respiration, but also have roles as secondary metabolites [89]. This could explain why genes encoding enzymes involved in the biosynthesis of terpenes are present in cyanobacterial genomes [90]. In cyanobacteria, geosmin and 2-methylisoborneol are widely studied terpenes as they are odorous metabolites that impact drinking water quality [64, 91, 92]. Nevertheless, the repertoire of terpenes produced by cyanobacteria is possibly larger than currently known, as various cryptic terpene synthases are found in their genomes [30, 88].

**Conclusion**

The availability of complete genomes has allowed mapping of BGCs in plasmids and the detection of known pathways of toxins (microcystin), odorous metabolites (geosmin), protease inhibitors (anabaenopeptin, aeruginosin),
antimicrobial compounds (ambiguine and hassalidin), and antitumor (cryptophycin) compounds. This is new in silico evidence that plasmids are involved in the biosynthesis of diverse natural products. Cyanobacterial plasmids also appear to be involved in the dissemination of BGCs by HGT in cyanobacteria. The likelihood of mobility of natural product BGCs seems to be higher in certain orders with larger genomes, particularly from Nostocales. Thus, future research should investigate potential transmission of BGCs between cyanobacteria in vivo. If possible, the transmission of BGCs among cyanobacteria would present new biotechnological opportunities but also environmental and economic risks. Cyanobacteria, which are believed to be harmless, could acquire genes for toxin biosynthesis.

Methods

Cyanobacterial genomes

“Cyanobacteria/Melainabacteria group” genomes deposited between 27 July 2001 and 14 January 2020 in the NCBI GenBank [43] at the “Complete” and “Chromosome” assembly level were analyzed. Altogether, they included 184 genomes from the phylum Cyanobacteria and 1 genome from Candidatus Melainabacteria (Table S1). The statistics of the genome assemblies were obtained from GenBank. Averages, standard deviations, and boxplot and scatter graphs were generated using Microsoft Excel v16.0.6742.2048 (Microsoft, Redmond, WA, USA).

Identification of natural product pathways and other proteins of interest

Gene clusters involved in secondary metabolite pathways were automatically annotated with antiSMASH v5.1.1 [47]. Manual annotation and curation were performed in the program Artemis v18.1.0 [93], and sequences were compared against the NCBI GenBank database using BLASTp [94]. For the manual identification of BGCs in the plasmids and the gene cluster involved in plasmid mobility, the parameters of e-value \( \leq 1 \times 10^{-20} \), identity \( \geq 60\% \) were used for assigning orthologs. Identification of relaxase, VirB4, and VirD4 involved a wide diversity of strains and thus the parameters e-value \( \leq 1 \times 10^{-20} \), identity \( \geq 20\% \) were used. Plasmid representations were generated using the standard parameters of the BLAST analysis in the server Gview [95] and BRIG v0.95 [96]. The program Inkscape v0.92 was used for drawing BGCs (https://inkscape.org/).

Phylogenetic analyses

The phylogenetic analyses of the concatenated genes from the BGCs and 16S rRNA were created with 5 000 000 generations in MrBayes 3.2.7a [97]. The best substitution model for each gene in the BGCs was predicted using BIC calculation in jModelTest v2.1.10 [98]. The same method was used for the 16S rRNA phylogenetic analyses. The model HKY + I + G was predicted as the best. Tree visualization was performed with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and edited with Inkscape 1.0 (https://inkscape.org/). Cyanobacterial order was assigned according a polyphasic taxonomic system [45].

Phylogenomic analysis

A maximum likelihood tree was constructed in RAxML v8.2.12 [99] using 1000 bootstrap samples. The model LG + G + I was identified as the best fitting model by ProtTest 3.4.2 [100]. The phylogenomic placement was based on 120 bacterial single-copy conserved marker genes identified by GTDBTk 1.0.2 [101]. The tree was visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and figure editing was performed in Inkscape 1.0 (https://inkscape.org/).

List Of Abbreviations
BGC: Biosynthetic gene cluster

PPT: Phosphopantetheinyl transferase

NRPS: Nonribosomal peptide synthetase

PKS: Polyketide synthase

RiPP: Ribosomally synthesized and post-translationally modified peptide

HGT: Horizontal gene transfer

T4CP: Type IV coupling protein

MOB: Mobility gene

MPF: Mating pair formation gene

T4SS: Type IV secretion system

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and materials**

All data generated or analyzed in this study are included in this published article and its supplementary information files. The phylogenetic and phylogenomic datasets generated and/or analyzed during the current study are available in the TreeBase repository 27385, [http://purl.org/phylo/treebase/phylows/study/TB2:S27385?x-access-code=a0eb8a6420200cbeae9110ca518fe8f&format=html](http://purl.org/phylo/treebase/phylows/study/TB2:S27385?x-access-code=a0eb8a6420200cbeae9110ca518fe8f&format=html)

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

RVP was funded by the Doctoral Programme in Microbiology and Biotechnology of the University of Helsinki. DOA was supported by a São Paulo Research Foundation post-doctoral fellowship (FAPESP #2018/01563-2). RCB acknowledges financial support from Fundação para a Ciência e a Tecnologia (FCT) fellowship SFRH/BD/136367/2018. Grants from Nordforsk (NordAqua) and the Jane and Aatos Erkko Foundation to KS supported this study.

**Author contributions**

RVP, DOA, and DPF conceptualized the study. RVP, DOA, and RC performed the analyses. RVP wrote the manuscript. DPF and KS were responsible for supervision. KS managed funding acquisition. All authors participated in reviewing and
editing the manuscript and agreed to the published version of the manuscript.

Acknowledgements

Not applicable.

References

1. Woodruff H. Natural products from microorganisms. Science. 1980;208:1225–9.
2. Demain AL. Importance of microbial natural products and the need to revitalize their discovery. Journal of Industrial Microbiology & Biotechnology. 2014;41:185–201.
3. Jensen PR. Natural Products and the gene cluster revolution. Trends in Microbiology. 2016;24:968–77.
4. Osbourn A. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends in Genetics. 2010;26:449–57.
5. Stone MJ, Williams DH. On the evolution of functional secondary metabolites (natural products). Molecular Microbiology. 1992;6:29–34.
6. Nett M, Ikeda H, Moore BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes. Natural Product Reports. 2009;26:1362.
7. Daum M, Peintner I, Linnenbrink A, Frerich A, Weber M, Paululat T, et al. Organisation of the biosynthetic gene cluster and tailoring enzymes in the biosynthesis of the tetracyclic quinone glycoside antibiotic polyketomycin. ChemBioChem. 2009;10:1073–83.
8. Yang G, Zhang Y, Lee NK, Cozad MA, Kearney SE, Luesch H, et al. Cyanobacterial Sfp-type phosphopantetheinyl transferases functionalize carrier proteins of diverse biosynthetic pathways. Scientific Reports. 2017;7. doi:10.1038/s41598-017-12244-3.
9. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, et al. A new enzyme superfamily — the phosphopantetheinyl transferases. Chemistry & Biology. 1996;3:923–36.
10. Beld J, Sonnenschein EC, Vickery CR, Noel JP, Burkart MD. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. Nat Prod Rep. 2014;31:61–108.
11. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, et al. Insights from 20 years of bacterial genome sequencing. Functional & Integrative Genomics. 2015;15:141–61.
12. Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. ChemBioChem. 2009;10:625–33.
13. Corre C, Challis GL. New natural product biosynthetic chemistry discovered by genome mining. Natural Product Reports. 2009;26:977.
14. Bachmann BO, Van Lanen SG, Baltz RH. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? Journal of Industrial Microbiology & Biotechnology. 2014;41:175–84.
15. Micallef ML, D’Agostino PM, Sharma D, Viswanathan R, Moffitt MC. Genome mining for natural product biosynthetic gene clusters in the Subsection V cyanobacteria. BMC Genomics. 2015;16. doi:10.1186/s12864-015-1855-z.
16. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 1996;3:109–36.
17. Alvarenga DO, Fiore MF, Varani AM. A Metagenomic Approach to Cyanobacterial Genomics. Frontiers in Microbiology. 2017;8. doi:10.3389/fmicb.2017.00809.
18. Sciuto K, Moro I. Cyanobacteria: the bright and dark sides of a charming group. Biodiversity and Conservation. 2015;24:711–38.

19. Kehr J-C, Gatte Picchi D, Dittmann E. Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes. Beilstein Journal of Organic Chemistry. 2011;7:1622–35.

20. Angermary SA, Hellwingwerf KJ, Lindblad P, Teixeira de Mattos MJ. Energy biotechnology with cyanobacteria. Current Opinion in Biotechnology. 2009;20:257–63.

21. Burja AM, Banaigs B, Abou-Mansour E, Grant Burgess J, Wright PC. Marine cyanobacteria—a prolific source of natural products. Tetrahedron. 2001;57:9347–77.

22. Dittmann E, Gugger M, Sivonen K, Fewer DP. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. Trends in Microbiology. 2015;23:642–52.

23. Tan LT. Bioactive natural products from marine cyanobacteria for drug discovery. Phytochemistry. 2007;68:954–79.

24. Martins J, Vasconcelos V. Cyanobactins from cyanobacteria: current genetic and chemical state of knowledge. Marine Drugs. 2015;13:6910–46.

25. Hillwig ML, Zhu Q, Liu X. Biosynthesis of ambiguine indole alkaloids in cyanobacterium Fischerella ambigua. ACS Chemical Biology. 2014;9:372–7.

26. Jain S, Prajapat G, Abrar M, Ledwani L, Singh A, Agrawal A. Cyanobacteria as efficient producers of mycosporine-like amino acids. Journal of Basic Microbiology. 2017;57:715–27.

27. Pattanaik B, Lindberg P. Terpenoids and their biosynthesis in cyanobacteria. Life. 2015;5:269–93.

28. Singh SP, Klisch M, Sinha RP, Häder D-P. Genome mining of mycosporine-like amino acid (MAA) synthesizing and non-synthesizing cyanobacteria: A bioinformatics study. Genomics. 2010;95:120–8.

29. Wang H, Fewer DP, Sivonen K. Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. PLoS ONE. 2011;6:e22384.

30. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, et al. Improving the coverage of the cyanobacterial phylogeny using diversity-driven genome sequencing. Proceedings of the National Academy of Sciences. 2013;110:1053–8.

31. Calteau A, Fewer DP, Latifi A, Coursin T, Laurent T, Jokela J, et al. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. BMC Genomics. 2014;15:977.

32. Ehrenreich IM, Waterbury JB, Webb EA. Distribution and diversity of natural product genes in marine and freshwater cyanobacterial cultures and genomes. Applied and Environmental Microbiology. 2005;71:7401–13.

33. Ginolhac A, Jarrin C, Robe P, Perrière G, Vogel TM, Simonet P, et al. Type I polyketide synthases may have evolved through horizontal gene transfer. Journal of Molecular Evolution. 2005;60:716–25.

34. Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics. 1996;143:1843–60.

35. Dittmann E, Fewer DP, Neilan BA. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. FEMS Microbiology Reviews. 2013;37:23–43.

36. Boto L. Horizontal gene transfer in evolution: facts and challenges. Proceedings of the Royal Society B: Biological Sciences. 2010;277:819–27.

37. Harrison E, Brockhurst MA. Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends in Microbiology. 2012;20:262–7.

38. De La Cruz F, Frost LS, Meyer RJ, Zechner EL. Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiology Reviews. 2010;34:18–40.

39. Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. Microbiology and Molecular Biology Reviews. 2010;74:434–52.
40. Garcillán-Barcia MP, Francia MV, de La Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiology Reviews. 2009;33:657–87.

41. Christie PJ. Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2004;1694:219–34.

42. Zechnel EL, Lang S, Schildbach JF. Assembly and mechanisms of bacterial type IV secretion machines. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367:1073–87.

43. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Research. 2016;44:D67–72.

44. Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC, et al. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. eLife. 2013;2. doi:10.7554/eLife.01102.

45. Komarek J, Kastovsky J, Mares J, Johansen J. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia. 2014;86:295–335.

46. Komarek J, Johansen JR, Smarda J, Strunecky O. Phylogeny and taxonomy of Synechococcus-like cyanobacteria. Fottea. 2020;20:171–91.

47. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Research. 2019;47:W81–7.

48. Harrison PW, Lower RPJ, Kim NKD, Young JPW. Introducing the bacterial “chromid”: not a chromosome, not a plasmid. Trends in Microbiology. 2010;18:141–8.

49. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, la Cruz F de, Rocha EPC. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. Nucleic Acids Research. 2014;42:5715–27.

50. Griese M, Lange C, Soppa J. Ploidy in cyanobacteria. FEMS Microbiology Letters. 2011;323:124–31.

51. Skarstad K, Steen HB, Boye E. Cell cycle parameters of slowly growing Escherichia coli B/r studied by flow cytometry. J Bacteriol. 1983;154:656–62.

52. Blank CE, Sánchez-Baracaldo P. Timing of morphological and ecological innovations in the cyanobacteria – a key to understanding the rise in atmospheric oxygen: Timing of evolutionary innovations in the cyanobacteria. Geobiology. 2010;8:1–23.

53. Sargent EC, Hitchcock A, Johansson SA, Langlois R, Moore CM, LaRoche J, et al. Evidence for polyploidy in the globally important diazotroph Trichodesmium. FEMS Microbiology Letters. 2016;363:fnw244.

54. Watanabe S. Cyanobacterial multi-copy chromosomes and their replication. Bioscience, Biotechnology, and Biochemistry. 2020;1–13.

55. Hu B, Yang G, Zhao W, Zhang Y, Zhao J. MreB is important for cell shape but not for chromosome segregation of the filamentous cyanobacterium Anabaena sp. PCC 7120: chromosome partitioning in a cyanobacterium. Molecular Microbiology. 2007;63:1640–52.

56. Chen AH, Afonso B, Silver PA, Savage DF. Spatial and temporal organization of chromosome duplication and segregation in the cyanobacterium Synechococcus elongatus PCC 7942. PLoS ONE. 2012;7:e47837.

57. Schneider D, Fuhrmann E, Scholz I, Hess WR, Graumann PL. Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes. BMC Cell Biology. 2007;8. doi:10.1186/1471-2121-8-39.

58. Welsh EA, Liberton M, Stockel J, Loh T, Elvitigala T, Wang C, et al. The genome of Cyanothce 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. Proceedings of the National Academy of Sciences. 2008;105:15094–9.
59. Wang H, Sivonen K, Rouhiainen L, Fewer DP, Lyra C, Rantala-Ylinen A, et al. Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium *Anabaena* sp. strain 90. BMC Genomics. 2012;13:613.

60. diCenzo GC, Finan TM. The divided bacterial genome: structure, function, and evolution. Microbiology and Molecular Biology Reviews. 2017;81. doi:10.1128/MMBR.00019-17.

61. diCenzo GC, Mengoni A, Perrin E. Chromids aid genome expansion and functional diversification in the family *Burkholderiaceae*. Molecular Biology and Evolution. 2019;36:562–74.

62. Panda D, DeLuca K, Williams D, Jordan MA, Wilson L. Antiproliferative mechanism of action of cryptophycin-52: Kinetic stabilization of microtubule dynamics by high-affinity binding to microtubule ends. Proceedings of the National Academy of Sciences. 1998;95:9313–8.

63. Tippelt A, Busche T, Rückert C, Nett M. Complete genome sequence of the cryptophycin-producing cyanobacterium *Nostoc* sp. strain ATCC 53789. Microbiology Resource Announcements. 2020;9. doi:10.1128/MRA.00040-20.

64. Suurnäkki S, Gomez-Saez GV, Rantala-Ylinen A, Jokela J, Fewer DP, Sivonen K. Identification of geosmin and 2-methylisoborneol in cyanobacteria and molecular detection methods for the producers of these compounds. Water Research. 2015;68:56–66.

65. Vestola J, Shishido TK, Jokela J, Fewer DP, Aitio O, Permi P, et al. Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. Proceedings of the National Academy of Sciences. 2014;111:E1909–17.

66. Kaebernick M, Neilan BA, Börner T, Dittmann E. Light and the transcriptional response of the microcystin biosynthesis gene cluster. Applied and Environmental Microbiology. 2000;66:3387–92.

67. Bolch CJS, Blackburn SI, Jones GJ, Orr PT, Grewe PM. Plasmid content and distribution in the toxic cyanobacterial genus *Microcystis* Kützing *ex* Lemmermann (Cyanobacteria: Chroococcales). Phycologia. 1997;36:6–11.

68. Ehling-Schulz M, Fricker M, Grallert H, Rieck P, Wagner M, Scherer S. Cereulide synthetase gene cluster from emetic *Bacillus cereus*: Structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. BMC Microbiology. 2006;6:20.

69. Skarin H, Segerman B. Horizontal gene transfer of toxin genes in *Clostridium botulinum*: Involvement of mobile elements and plasmids. Mobile Genetic Elements. 2011;1:213–5.

70. Muro-Pastor AM, Kuritz T, Flores E, Herrero A, Wolk CP. Transfer of a genetic marker from a megaplasmid of *Anabaena* sp. strain PCC 7120 to a megaplasmid of a different *Anabaena* strain. Journal of Bacteriology. 1994;176:1093–8.

71. Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. Microbiology and Molecular Biology Reviews. 2010;74:434–52.

72. Cane DE. Harnessing the biosynthetic code: combinations, permutations, and mutations. Science. 1998;282:63–8.

73. Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. Natural Product Reports. 2001;18:380–416.

74. Quadri LEN, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT. Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry. 1998;37:1585–95.
78. Copp JN, Neilan BA. The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria. Applied and Environmental Microbiology. 2006;72:2298–305.

79. Kim JH, Komatsu M, Shin-ya K, Omura S, Ikeda H. Distribution and functional analysis of the phosphopantetheinyl transferase superfamily in Actinomycetales microorganisms. Proceedings of the National Academy of Sciences. 2018;115:6828–33.

80. Liu Q, Ma Y, Zhou L, Zhang Y. Gene cloning, expression and functional characterization of a phosphopantetheinyl transferase from Vibrio anguillarum serotype O1. Archives of Microbiology. 2005;183:37–44.

81. De Lay NR, Cronan JE. Genetic interaction between the Escherichia coli AcpT phosphopantetheinyl transferase and the YejM inner membrane protein. Genetics. 2008;178:1327–37.

82. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013;30:108–60.

83. Sivonen K, Leikoski N, Fewer DP, Jokela J. Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria. Applied Microbiology and Biotechnology. 2010;86:1213–25.

84. Cubillos-Ruiz A, Berta-Thompson JW, Becker JW, van der Donk WA, Chisholm SW. Evolutionary radiation of lanthipeptides in marine cyanobacteria. Proceedings of the National Academy of Sciences. 2017;114:E5424–33.

85. Tietz J, Schwalen CJ, Patel PS, Maxton S, Blair PM, Tai H-C, et al. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nature Chemical Biology. 2017;13:470–8.

86. Ziemert N, Ishida K, Weiz A, Hertweck C, Dittmann E. Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides. Applied and Environmental Microbiology. 2010;76:3568–74.

87. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar S, et al. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Research. 2017;45:W36–41.

88. Yamada Y, Kuzuyama T, Komatsu M, Shin-ya K, Omura S, Cane DE, et al. Terpene synthases are widely distributed in bacteria. Proceedings of the National Academy of Sciences. 2015;112:857–62.

89. Breitmaier E. Terpenes. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2006. doi:10.1002/9783527609949.

90. Leao T, Castelão G, Korobeynikov A, Monroe EA, Podell S, Glukhov E, et al. Comparative genomics uncovers the prolific and distinctive metabolic potential of the cyanobacterial genus Moorea. Proceedings of the National Academy of Sciences. 2017;114:3198–203.

91. Oh H-S, Lee CS, Srivastava A, Oh H-M, Ahn C-Y. Effects of environmental factors on cyanobacterial production of odorous compounds: geosmin and 2-methylisoborneol. Journal of Microbiology and Biotechnology. 2017;27:1316–23.

92. Jüttner J. Physiology and biochemistry of odorous compounds from freshwater cyanobacteria and algae. Water Science and Technology. 1995;31. doi:10.1016/0273-1223(95)00458-Y.

93. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012;28:464–9.

94. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215:403–10.

95. Petkau A, Stuart-Edwards M, Stothard P, Van Domselaar G. Interactive microbial genome visualization with GView. Bioinformatics. 2010;26:3125–6.

96. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12. doi:10.1186/1471-2164-12-402.
97. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology. 2012;61:539–42.

98. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 2012;9:772–772.

99. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

100. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011;27:1164–5.

101. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarszewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology. 2018;36:996–1004.