Marine Genomics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Review

Exploring cyanobacterial genomes for natural product biosynthesis pathways

Melinda L. Micallef [a,1], Paul M. D'Agostino [a,b,1], Bakir Al-Sinawi [b], Brett A. Neilan [b], Michelle C. Moffitt [a,*,1]

[a] School of Science and Health, University of Western Sydney, Campbelltown, NSW 2560, Australia
[b] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Kensington, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 8 September 2014
Received in revised form 22 November 2014
Accepted 23 November 2014
Available online xxxx

Keywords:
Cyanobacteria
Genome mining
Natural products
Bioinformatic tools
Heterologous expression
Mass spectrometry
Gene cluster

ABSTRACT

Cyanobacteria produce a vast array of natural products, some of which are toxic to human health, while others possess potential pharmaceutical activities. Genome mining enables the identification and characterisation of natural product gene clusters; however, currently the number of cyanobacterial genomes remains low compared to other phyla. There has been a recent effort to rectify this issue by increasing the number of sequenced cyanobacterial genomes. This has enabled the identification of biosynthetic gene clusters for structurally diverse metabolites, including non-ribosomal peptides, polyketides, ribosomal peptides, UV-absorbing compounds, alkaloids, terpenes and fatty acids. While some of the identified biosynthetic gene clusters correlate with known metabolites, genome mining also highlights the number and diversity of clusters for which the product is unknown (referred to as orphan gene clusters). A number of bioinformatics tools have recently been developed in order to predict the products of orphan gene clusters; however, in some cases the complexity of the cyanobacterial pathways makes the prediction problematic. This can be overcome by the use of mass spectrometry-guided natural product genome mining, or heterologous expression. Application of these techniques to cyanobacterial natural product gene clusters will be explored.

Contents

* Corresponding author at: School of Science and Health, Building 21 Campbelltown Campus, University of Western Sydney, Locked Bag 1797, Penrith, NSW, 2751, Australia. Tel.: +61 2 4620 3521; fax: +61 2 4620 3025.
E-mail address: m.moffitt@uws.edu.au (M.C. Moffitt).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.margen.2014.11.009
1874-7787/

1. Introduction

Cyanobacteria are photosynthetic bacteria which inhabit a vast range of ecosystems and display an immense evolutionary history. They are capable of producing a vast array of natural products which are likely to play a role in ecological or biological functions (Singh et al., 2005; Skulberg, 2000; Kehr et al., 2011; Méjean and Ploux, 2013). Cyanobacterial secondary metabolites have a significant impact on human health, both directly through the contamination of drinking water reservoirs by toxigenic strains (Dittmann et al., 2013; Dittmann and Wiegand, 2006), and indirectly through their pharmaceutical potential.

Synechocystis sp. PCC 6803 was the first cyanobacterium to undergo full genome sequencing (Kaneko et al., 1996). While the number of cyanobacterial genomes remains low compared to other phyla, there has been a recent effort to rectify this issue by increasing the coverage of sequenced genomes within the cyanobacterial phylum (Shih et al., 2013). This recent sequencing project is known as the CyanoGEBA (Genomic Encyclopedia of Bacteria and Archaea) data set. This has enabled the identification of biosynthetic gene clusters for a number of known metabolites and gives clues to their distribution across the phylum. These genome sequencing efforts have also highlighted that cyanobacteria encode for many more natural product gene clusters than identified natural products. A number of bioinformatics tools have recently been developed in order to predict the structure of secondary metabolites produced by "orphan" gene clusters for which the product is unknown. Here, we explore the suitability of these tools to characterise cyanobacterial natural product gene clusters. Lastly, we explore the possibility of characterising orphan gene clusters via heterologous expression.

2. Cyanobacterial genomics

2.1. Current status of publically available genome sequences

As of June 2014, there are currently 208 cyanobacterial genome sequences publically available in GenBank and the Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) databases (Supplementary Table 1). The genomes discussed in this review are published or the publically available genome was used to identify and characterise a biosynthetic gene cluster (Table 1). The most recent review on this topic, reporting the smallest sequenced complete cyanobacterial genome, belonged to the marine cyanobacterium UCYN-A, with a genome size of 1.44 Mb, while the largest genome belonged to *Nostoc punctiforme* ATCC 29122, with a genome size of 9.05 Mb (Hess, 2011). Since then, the number and diversity of sequenced cyanobacterial genomes have drastically increased broadening our current knowledge and understanding of cyanobacterial genomes. Thus, with a genome size of 12.07 Mb, *Scytonema hofmanni* PCC 7110 is now the largest sequenced cyanobacterial genome, and encodes 12,356 genes (Dagan et al., 2013).

Cyanobacteria are classified into five subsections based on their morphology. The subsections I and II cyanobacteria are unicellular species, which are differentiated by their ability to reproduce through binary or multiple fissions. Subsections III–V are multicellular species. Cyanobacteria classified into subsection III have vegetative cells, while subsections IV and V are differentiated by their ability to reproduce in false or true-branching filaments (Rippka et al., 1979). Prior to the publication of the CyanoGEBA data set in 2013, there were no publically available genomes from subsection II cyanobacteria, and the subsection V cyanobacteria were underrepresented. The CyanoGEBA data set significantly increased the number and diversity of sequenced cyanobacterial genomes, reporting the whole genome shotgun sequencing of an additional 54 strains (Shih et al., 2013). This data set doubled the number of cyanobacterial genomes from subsections III and IV, and currently contains the only genomes available from the subsection II cyanobacteria. Comparing the number of cyanobacterial genomes sequenced, the subsection I cyanobacteria are overrepresented compared to the other cyanobacterial groups. In this review, we have also included analysis of the melainabacteria. Although non-photosynthetic, the melainabacteria were first classified as a sister phyla of the cyanobacteria (Di Rienzi et al., 2013). More recent phylogenetic analysis, including whole genome phylogeny, has proposed that the melainabacteria are a class within an expanded phylum of cyanobacteria (Soo et al., 2014). Although this non-photosynthetic class of cyanobacteria has only been recently described, there are currently 14 cyanobacterial strains sequenced, either to completion or to draft stage (Di Rienzi et al., 2013; Soo et al., 2014). Genomes of symbiotic cyanobacteria have also been analysed. These include, a sponge symbiont and three tunicate symbionts belonging to subsection I (Gao et al., 2014; Donia et al., 2011). Another two of the 28 genomes from subsection IV are diatom symbionts, while one is a water-fern symbiont (Hilton et al., 2013; Ran et al., 2010). This demonstrates the diverse range of cyanobacterial genome type and sizes currently available.

Table 1
Summary of publically available cyanobacterial genomes, including number, size range, average size and average %GC.

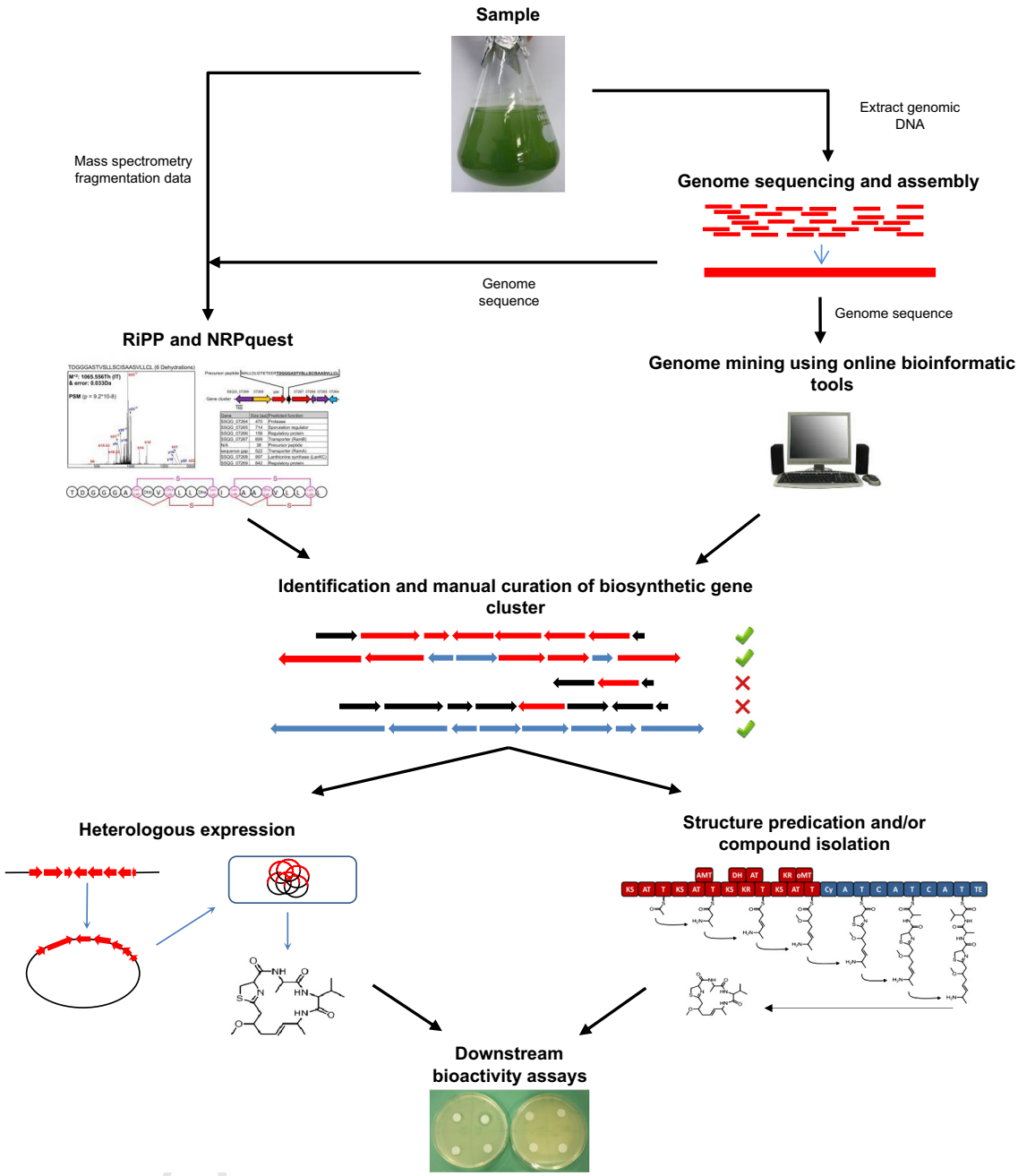
Subsection	Number of publically available genomes	Range of genome sizes (Mb) (average size)	Average %GC
I	105	0.1–8.36 (3.63)	44.8
II	6	4.99–7.39 (6.05)	39.9
III	44	4.68–9.42 (6.42)	46.5
IV	28	2.21–12.07 (6.44)	40.1
V	11	4.89–8.4 (6.90)	41.1
Melainabacteria	14	1.19–5.49 (2.25)	34.9

2.2. Natural product biosynthetic pathways identified in cyanobacterial genomes

In this section of the review, we have strictly reviewed the literature that reports the presence or absence of natural product biosynthetic pathways in cyanobacterial genomes. Additional bioinformatics screening of genomes beyond those already published was outside the scope of this review. We have strictly focused on genome-based screening studies.

Cyanobacteria produce a range of different natural product classes, including peptides, polyketides, alkaloids, fatty acids, terpenes and UV-absorbing compounds (Fig. 1). Many of the biosynthetic pathways which encode these natural products discussed below were originally identified through more traditional sequence analysis. These include biosynthetic gene clusters for the commonly occurring cyanotoxins, such as microcystin, and other metabolites of interest such as barbamide and the discovery of these sequences were reviewed previously by (Méjean and Ploux, 2013) and (Dittmann et al., 2013) and therefore are not described in this review. Subsequent genome sequence analysis has demonstrated the distribution of pathways across species and genera and has provided insights into their evolution.

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009



Q1 **Fig. 1.** Representative cyanobacterial natural products from each structural class. Microcystin is represented from the NRPS/PKS pathways; patellamide A is represented from the PRPS pathway; shinorine is represented from the UV-absorbing compounds; saxitoxin is represented from the alkaloids; geosmin is represented from the terpenes and heptadecane is a representative of the fatty acids (hydrocarbons).

2.2.1. Nonribosomal peptide synthetase, polyketide synthase and hybrid pathways

The majority of cyanobacterial natural products are non-ribosomal peptides, polyketides or hybrid peptide–polyketide compounds. Nonribosomal peptides are biosynthesised by nonribosomal peptide synthetases (NRPS), multifunctional enzyme complexes which assemble either proteinogenic or nonproteinogenic amino acids into the final peptide structure in an assembly line fashion (Fischbach and Walsh, 2006). Similarly, polyketides are biosynthesised by polyketide synthases (PKS), which assemble polyketides from acyl-CoA in a sequential manner (Fischbach and Walsh, 2006).

NRPS and PKS enzyme complexes are similar in their biochemical logic. An NRPS or PKS assembly line consists of several modules, with each module responsible for the selection and addition of a single substrate to the lengthening molecule chain (Marahiel et al., 1997; Konz and Marahiel, 1999; Staunton and Weissman, 2001). The order of modules (thus, the order of selected substrates) usually dictates the peptide or polyketide chain sequence, and is referred to as the co-linearity rule. Within each module are series of domains, each responsible for a specific enzymatic reaction. As a minimum module, an NRPS is made up of an adenylation (A) domain responsible for substrate selection, a peptide carrier protein (PCP) responsible for tethering the substrate to the enzymatic complex and a condensation (C) domain, responsible for catalysing peptide formation and chain elongation. Analogous to the NRPS, a minimum PKS module consists of an acyltransferase (AT) domain, acyl carrier protein (ACP) and ketosynthase domain (KS), respectively. Additionally,

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

auxillary or tailoring domains may be present within a module and act on the activated substrate to provide structural diversity (Liou and Khosla, 2003; Donadio et al., 2007). Three distinct classes of PKSs are involved in polyketide biosynthesis. Type I PKSs are large multifunctional enzyme complexes, containing non-interactively-acting domains, while type II PKSs contain a single interactively-acting domain. Type III PKSs, however, are homodimeric enzymes with interactively-acting domains (Bode and Müller, 2005). While type I PKS genes are common among cyanobacteria, type II and III PKS genes have been rarely identified.

A number of NRPS/PKS/hybrid biosynthetic gene clusters have also been identified through whole genome sequencing. While some biosynthetic gene clusters have been identified in single genomes across one or two genera, some gene clusters have been identified from multiple genomes, across multiple genera and species (Supplementary Table 2). Thus, the evolution and biological function of these pathways and corresponding secondary metabolites have been of major interest to researchers but to date largely remain enigmatic.

The biosynthetic gene cluster for the cyclic heptapeptide toxin microcystin has been identified in 19 publically available cyanobacterial genomes. The microcystin (*mcy*) gene cluster, which encodes two PKS, three NRPS and a hybrid NRPS/PKS gene, has been identified in nine *Microcystis aeruginosa* genomes, eight *Planktothrix* genomes, and in *Anabaena* sp. strain 90 and *Fischerella* sp. PCC 9339 (Kaneko et al., 2007; Frangeul et al., 2008; Humbert et al., 2013; Fiore et al., 2013; Tooming-Klunderud et al., 2013; Rounge et al., 2009; Wang et al., 2012; Shih et al., 2013). The *mcy* gene cluster in *M. aeruginosa* PCC 9717, however, is incomplete (Humbert et al., 2013). As expected, the *mcy* gene cluster was absent from the non-toxic *M. aeruginosa* TAIHU98 genome (Yang et al., 2013). Nodularin, a related cyclic pentapeptide toxin, is encoded by two NRPS, one PKS and two hybrid NRPS/PKS genes. This biosynthetic gene cluster has been identified from the genome of *Nodularia spumigena* CCY9414 (Voß et al., 2013).

The biosynthetic gene cluster for the cytotoxic alkaloid cylindrospermopsin encodes four PKS and one hybrid NRPS/PKS genes, in addition to an amidinotransferase, amidohydrolase, tailoring, transport and regulation enzymes. The gene cluster encoding the biosynthesis of cylindrospermopsin has been identified in three publically available cyanobacterial genomes, specifically *Cylindrospermopsis raciborskii* CS-505, *C. raciborskii* CS-506 and *Oscillatoria* sp. PCC 6506 (Sinha et al., 2014; Stucken et al., 2010; Méjean et al., 2010; Mazmouz et al., 2010).

Whole genome sequencing has led to the identification of a single biosynthetic gene cluster for the neurotoxins anatoxin-a and homoanatoxin-a. The 29 kb gene cluster has been identified in two cyanobacterial genomes, specifically *Oscillatoria* sp. PCC 6505 and *Anabaena* sp. strain 37 (Méjean et al., 2010; Rantala-Ylinen et al., 2011).

The biosynthesis of curacin a, an antiproliferative and cytotoxic compound, is encoded by a large (64 kb) gene cluster containing a single NRPS gene and multiple PKS genes. Barbamide, a chlorinated lipopeptide with potent molluscicidal activity, is biosynthesised by a 26 kb cluster encoding five NRPS, one PKS, and additional tailoring enzymes. Both of these gene clusters have been identified in the genomes of *Moorea producens* 3L (formally *Lyngbya majuscula* 3L) and *Lyngbya aestuarii* BL J (Jones et al., 2011; Kothari et al., 2013). Furthermore, a putative biosynthetic gene cluster for the lipopeptide carmabin has also been identified from *M. producens* 3L (Jones et al., 2011).

The protease inhibitors anabaenopeptilides are cyclic peptides biosynthesised by a 29 kb gene cluster containing three PKS genes and a halogenase gene. The biosynthetic gene cluster for anabaenopeptilides was identified within *Anabaena* sp. strain 90 (Rouhiainen et al., 2000; Wang et al., 2012).

The biosynthetic gene cluster for anabaenopeptins, cyclic hexapeptides which are also protease inhibitors, was identified in eleven publically available cyanobacterial genomes. The five NRPS genes comprising the anabaenopeptin gene cluster were identified in eight *Planktothrix* strains, two *M. aeruginosa* strains, and in the *Anabaena* sp. strain 90, *N. spumigena* CCY9414 and *N. punctiforme* PCC 73102 genome (Tooming-Klunderud et al., 2013; Rounge et al., 2009; Humbert et al., 2013; Rouhiainen et al., 2010; Voß et al., 2013).

The aeruginosin class of peptides is biosynthesised by NRPS/PKS biosynthetic gene clusters, and includes the aeruginosins, aeruginosides, microginins and spumigins. The 34 kb aeruginosin biosynthetic gene cluster has been identified in 19 publically available cyanobacterial genomes, including eight *Planktothrix* and eleven *M. aeruginosa* genomes (Tooming-Klunderud et al., 2013; Rounge et al., 2009; Frangeul et al., 2008; Humbert et al., 2013; Fiore et al., 2013; Yang et al., 2013). The biosynthetic gene cluster for spumigin biosynthesis has been identified in a single genome (*N. spumigena* CCY9414) (Fewer et al., 2009; Voß et al., 2013). Furthermore, the 23 kb microginin biosynthetic gene cluster, which encodes three NRPS/PKS/hybrid genes and an ABC transporter gene, has been identified in five cyanobacterial genomes (three *Microcystis* and two *Planktothrix* genomes) (Humbert et al., 2013; Tooming-Klunderud et al., 2013).

The cyanopeptolin biosynthetic gene cluster has been identified in a range of cyanobacterial genomes, including ten *M. aeruginosa* genomes (although two are incomplete), and eight *Plantothrix* genomes (Kaneko et al., 2007; Frangeul et al., 2008; Humbert et al., 2013; Tooming-Klunderud et al., 2013; Rounge et al., 2009). The cyanopeptolin gene cluster, however, was absent from the *M. aeruginosa* TAIHU98 genome (Yang et al., 2013).

Biosynthetic genes for the antifungal cyclic lipopeptide hassallidin pathway have been identified in two cyanobacterial genomes, specifically *Anabaena* sp. strain 90 and *C. raciborskii* CS-505 (Wang et al., 2012; Vestola et al., 2014).

While type II and III PKS genes are rarely identified in cyanobacteria, recent analysis of *Microcystis* genomes identified PKSmod/PKSIII genes within four genomes, and PKSI interactive genes within two *Microcystis* and the *N. spumigena* CCY9414 genomes (Frangeul et al., 2008; Humbert et al., 2013; Voß et al., 2013). This highlights the importance of genome screening to detect new and unusual natural product biosynthetic gene clusters.

A number of orphan NRPS/PKS/hybrid biosynthetic gene clusters have been identified from publically available cyanobacterial genomes. Analysis of individual genomes has identified 29 orphan NRPS/PKS/hybrid biosynthetic gene clusters from 13 cyanobacterial genomes (Kaneko et al., 2007; Stucken et al., 2010; Sinha et al., 2014; Jones et al., 2011; Humbert et al., 2013; Voß et al., 2013; Méjean et al., 2010; Rounge et al., 2009).

2.2.2. Ribosomal peptides

Cyanobacteria are also known to produce a group of ribosomally-synthesised and post translationally modified peptides, known as RiPPs (Arnison et al., 2013). The biosynthetic pathway of these compounds is referred to as post-ribosomal peptide synthesis (PRPS). The RiPPs produced by PRPS are usually encoded within the structural gene, which produces the precursor peptide. The precursor peptide is comprised of a signal sequence, a leader peptide, a core peptide and a recognition sequence. The core peptide/region of the precursor peptide is then post-translationally modified to produce the final RiPP (Arnison et al., 2013). A range of RiPP families have been identified from cyanobacterial genomes, including the bacteriocins, microviridins, and cyanobactins (Supplementary Table 3).

The bacteriocins are one family of RiPPs which are identified by a double-glycine motif in the leader sequence. A major class of the bacteriocins is the lanthipeptides (lanthionine-containing peptides). Of the 131 cyanobacterial genomes analysed, 115 genomes contain a bacteriocin gene cluster (Lee et al., 2008; Begley et al., 2009; Li et al., 2010; O'Sullivan et al., 2011; Wang et al., 2012; Fiore et al., 2013; Shih et al., 2013; Lima et al., 2014), including 8 genomes which are reported by Shih et al. (2013) to possess a bacteriocin cluster, although, those same clusters could not be identified by Wang et al. (2011). Q6

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

Another family of the RiPPs is the microviridins, N-acetylated tri- and tetra-decapeptides. Microviridins are known to be inhibitors of proteases. The biosynthetic gene cluster for microviridin biosynthesis contains five genes encoding the precursor peptide, two cyclisation proteins, an acetyltransferase and an ABC transporter gene. The microviridin biosynthetic gene cluster has been identified in 26 cyanobacterial genomes, specifically from the genera *Microcystis*, *Nodularia*, *Nostoc*, *Anabaena* and *Planktothrix* (Philmus et al., 2008; Rounge et al., 2009; Fiore et al., 2013; Humbert et al., 2013; Tooming-Klunderud et al., 2013; Voß et al., 2013). However one gene cluster from *Microcystis* is incomplete, while the *Nostoc* (*Anabaena*) sp. PCC 7120 cluster is inactive (Humbert et al., 2013).

Cyanobactins are small cyclic RiPPs which consist solely of proteinogenic amino acids. Cyanobactins display a range of bioactivities, including cytotoxic and multiple-drug reversing activities (Ishida et al., 2000; Ogino et al., 1996). The cyanobactin biosynthetic gene cluster is comprised of a precursor peptide, two genes encoding proteases, and two short hypothetical proteins. Cyanobacterial genomes have been extensively screened for the cyanobactin gene cluster (142 strains), and has been identified in 32 genomes (Schmidt et al., 2005; Sudek et al., 2006; Donia et al., 2008; Leikoski et al., 2010; Donia and Schmidt, 2012; Fiore et al., 2013; Humbert et al., 2013; Shih et al., 2013; Voß et al., 2013). Donia and Schmidt (2012) identified cyanobactin genes in *Oscillatoria* sp. PCC 6506; however, Shih et al. (2013) did not identify this cluster. The 32 genomes containing a cyanobactin gene cluster include those from *Oscillatoria* sp. PCC 6506. Furthermore, the cyanobactin gene clusters from *N. spumigena* CCY9414, *Calothrix* sp. PCC 7103, *Cyanothece* sp. PCC 7822, *Leptolyngbya* sp. PCC 7376 and *M. aeruginosa* NIES-843 may be non-functional (Leikoski et al., 2013). A biosynthetic gene cluster for oscillatorin has been identified in eight *Planktothrix* cyanobacterial genomes (Rounge et al., 2009; Tooming-Klunderud et al., 2013).

2.2.3. The UV-absorbing compounds scytonemin and mycosporine-like amino acids

Currently, cyanobacteria are known to produce only two different types of UV-absorbing molecules to protect the cell against either UVA or UVB radiation. One type of UV-absorbing compound which is produced by cyanobacteria is scytonemin, which is biosynthesised in response to UVA (315–400 nm wavelength) stress (Sorrels et al., 2009; Soule et al., 2009a). The 6 kb scytonemin biosynthetic gene cluster (*scy*) has been identified in six publically available cyanobacterial genomes (Supplementary Table 3 (Soule et al., 2009b; Kothari et al., 2013; Voß et al., 2013).

The other type of UV-absorbing molecule produced by cyanobacteria is the mycosporine-like amino acids (MAA), which are biosynthesised in response to UVB (280–315 nm wavelength) stress (Sinha and Häder, 2008). A number of MAAs have been identified from a range of cyanobacteria (Sinha and Häder, 2008); however, the biosynthesis of these compounds has only been characterised for the MAA shinorine. Three genes are required for the biosynthesis of the precursor compound for shinorine biosynthesis, mycosporine-glycine (Balskus and Walsh, 2010). In cyanobacteria, two alternative biosynthetic routes lead to shinorine biosynthesis from mycosporine-glycine. The first requires a NRPS-like enzyme and the other requires an ATP-Grasp ligase (Balskus and Walsh, 2010; Gao and Garcia-Pichel, 2011). The MAA (*mys*) gene cluster containing all four genes (with either the NRPS-like enzymes or the ATP-Grasp ligase) has been identified in 14 publically available cyanobacterial genomes (Supplementary Table 3) (Balskus and Walsh, 2010; Gao and Garcia-Pichel, 2011; Shih et al., 2013; Voß et al., 2013; Kothari et al., 2013). Interestingly, there are two cyanobacterial genomes in which both the *mys* and *scy* biosynthetic gene clusters have identified (*Cyanothece* sp. PCC 7424 and *N. spumigena* CCY9414) (Balskus and Walsh, 2010; Soule et al., 2009b; Voß et al., 2013). The *L. aestuarii* BL J genome has been reported to encode both biosynthetic gene clusters; however, the fourth gene required for MAA biosynthesis is present, but not clustered with the other genes in the genome (Kothari et al., 2013).

2.2.4. Alkaloids and terpenes

There are two major alkaloids biosynthesised by cyanobacteria. The first, saxitoxin, is a neurotoxic alkaloid responsible for paralytic shellfish poisoning (Wiese et al., 2010; D'Agostino et al., 2014). Saxitoxin is the most potent member of the paralytic shellfish toxin family of alkaloids, consisting of over 50 different analogues. Unlike NRPS and PKS biosynthesis, the saxitoxin biosynthetic gene cluster encodes a series of monofunctional enzymes responsible for building the saxitoxin core and subsequent tailoring reactions (D'Agostino et al., 2014). To date, the saxitoxin gene cluster has been characterised from five species of cyanobacteria, but only one has been identified by whole genome analysis, specifically *Raphidiopsis brookii* D9, which spans 25.7 kb, and putatively contains the minimal genes required for saxitoxin biosynthesis (Stucken et al., 2010). Further, whole genome sequencing has also allowed for comparative studies between the saxitoxin-producing and non-toxic strains of the Australian species *Anabaena circinalis*. A whole genome sequencing approach provided a database to allow D'Agostino et al. (2014) to compare the saxitoxin-producing *A. circinalis* AWC131C and the non-toxic *A. circinalis* AWQC310F at the proteomic level.

The other major group of alkaloids biosynthesised by cyanobacteria is the hapalindole-family of natural products, a group of isoprenoid-indole alkaloids, found exclusively in the subsection V cyanobacteria. Recently, biosynthetic gene clusters encoding the biosynthesis of the hapalindoles, ambiguines and welwitindolinones were reported from the genomes of three publically available *Fischerella* strains (Micallef et al., 2014).

Geosmin and 2-methylisoborneol (MIB) are volatile terpenes produced by cyanobacteria. Geosmin is biosynthesised by a single enzyme, while MIB requires both a methyltransferase and a monoterpene cyclase. Terpene biosynthesis has been investigated in 129 publically available cyanobacterial genomes, and 127 were found to contain genes involved in terpene biosynthesis (Supplementary Table 3) (Shih et al., 2013; Lima et al., 2014; Fiore et al., 2013; Englund et al., 2014).

2.2.5. Fatty acids (hydrocarbons)

There are two separate pathways for hydrocarbon biosynthesis from fatty acids in cyanobacteria.

The first pathway requires two genes for the conversion of fatty acids to alkanes, a fatty acyl-ACP reductase (FAAR) and an aldehyde deformylating oxygenase (ADO), which comprise the FAAR/ADO pathway. This pathway was originally identified in cyanobacterial genomes by Schirmer et al. (2010). A total of 120 out of 135 publically available cyanobacterial strains have been found to encode these two genes within their genomes (Supplementary Table 3) (Schirmer et al., 2010; Coates et al., 2014; Kothari et al., 2013). Klähn et al. (2014) recently identified the FAAR/ADO genes in 163 out of 181 publically available cyanobacterial genomes; however, the genes were only clustered in 138 cyanobacterial genomes. The second biosynthetic pathway for hydrocarbon biosynthesis in cyanobacteria is the olefin-producing (OLS) pathway. The OLS pathway requires a single PKS gene to convert the fatty acid to an alkene. By analysing cyanobacterial genomes currently available, this pathway has been identified in 15 publically available genomes by Coates et al. (2014) (Supplementary Table 3).

3. Bioinformatic characterisation of natural product biosynthetic gene clusters

The enormous accumulation of genome sequencing data has meant that many natural product researchers have turned to *in silico* approaches to quickly assess the natural product potential of a newly sequenced organism. In light of this, several bioinformatic tools have been developed. Generally, the bioinformatics tool used is dependent on the specificity of the search, that is, if the investigator is attempting to identify all possible gene clusters within an organism, a specific gene cluster of interest, or a putative substrate incorporated by a specific domain. Once the programme has completed the search, it is essential

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

for the user to manually check the output to determine if the identified gene clusters are likely to be genuine and to identify potential tailoring enzymes that are likely to be part of the biosynthetic gene cluster. Finally, a newly identified gene cluster can be analysed to predict the natural product structure or may be used for heterologous expression. Eventually, it is hoped that natural products identified by genome mining will be tested and utilised for their bioactivive and beneficial properties (Fig. 2).

3.1. Genome mining tools

A simple and common approach for gene cluster discovery utilises a user defined query sequence of a homologous protein to screen a newly sequenced genome using the BLAST suite of homology search algorithms (e.g. tblastn, blastp, blastx blastn) (Altschul et al., 1997). This method allows the user to define the specificity of the query sequence to search for a specific or broad range of enzymes belonging to a particular class of natural product gene clusters. For example, the hypothetical protein encoded by *scyC* has only been found within the genomes of scytonemin-producing cyanobacteria and may be used to specifically identify the scytonemin gene cluster (Balskus et al., 2011), while an A or KS domain may be used to screen for any NRPS or PKS biosynthesis pathway, respectively. However, BLAST is not ideal because the query sequence cannot contain degeneracy and will only provide the closest hits based on sequence identity. Thus, if searching for a broad range of enzymes from a specific class, methods that utilise sequence independent models, such as hidden Markov models (HMMs) should be used.

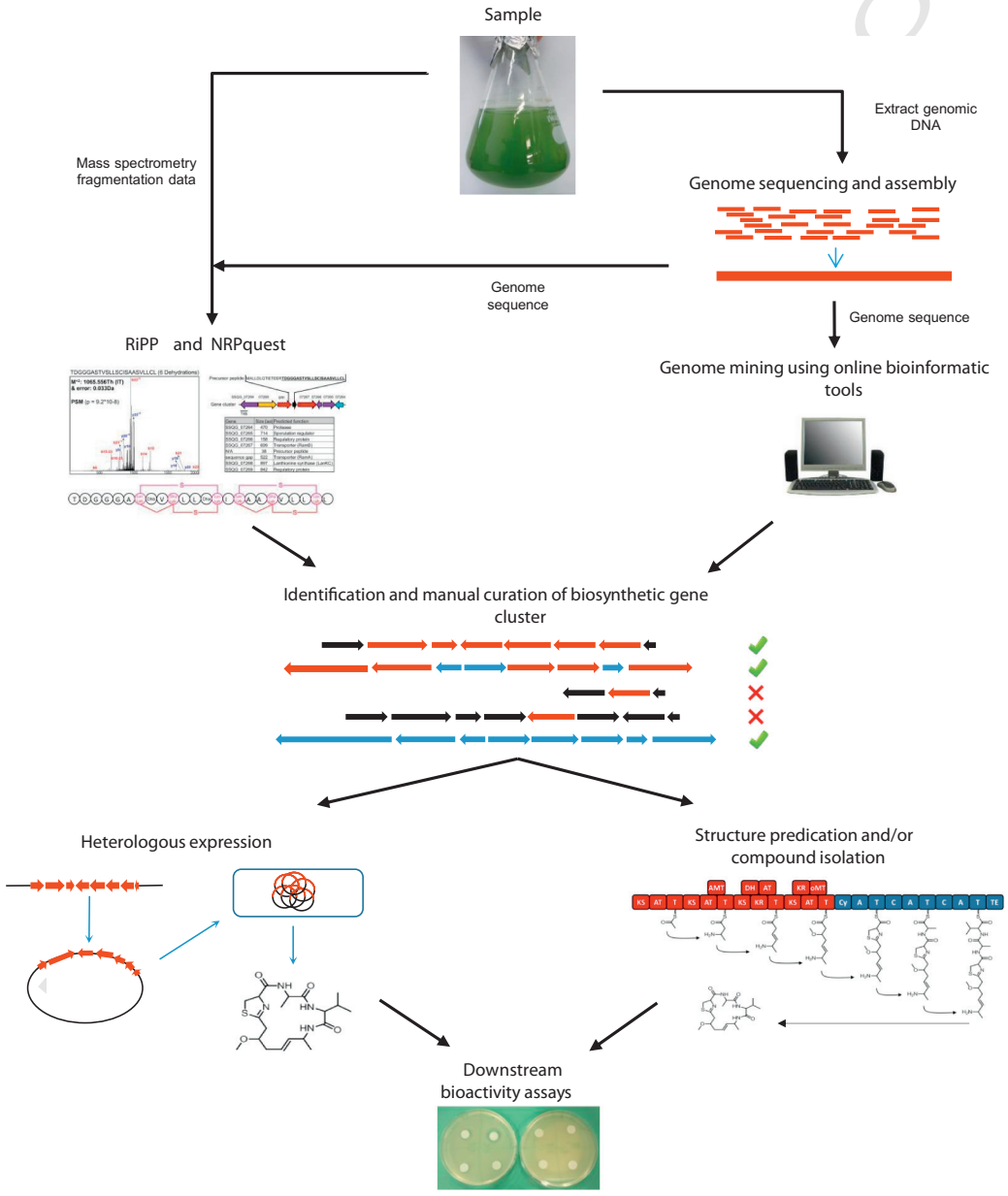


Fig. 2. Workflow of genome mining and compound elucidation from sample to bioactive assays. The new sample (isolate or metagenome) firstly undergoes DNA extraction and metabolomic analysis. The genome is then sequenced and assembled, with the genome and metabolome results are submitted into RiPPquest and RNPquest. Alternatively, the assembled genome can be mined using several bioinformatic programmes such as antiSMASH or ClusterFinder. Once a list of candidate gene clusters has been identified, they can be manually curated to exclude false positive results. In some cases, the newly identified cluster could be used to predict the structure of the final natural product and may help with compound isolation. Additionally, the gene cluster may be heterologously expressed. Finally, the isolated compound can be used for downstream bioactivity assays. The figure depicting RiPPquest and NRPquest has been amended from Mohimani et al. (2014a,b).

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

HMMs use multiple sequences and statistical analysis to identify homologous proteins and has provided a basis for the development of several sequence independent bioinformatic tools. Each *in silico* tool has specific traits and objectives, and the user should choose the type or combination of tools based on their specific requirements for optimal results. Two tools have been developed that have the advantage of screening a wide range of gene cluster classes. Firstly, the most commonly used tool for analysing the secondary metabolite potential of sequenced genomes is the "Antibiotics and secondary metabolite analysis shell" (antiSMASH) (Blin et al., 2013; Medema et al., 2011). antiSMASH utilises a web-user interface and attempts to identify up to 24 clusters classes. It is extremely user friendly given that a query can be submitted as soon as a newly sequenced genome has been assembled. Secondly, the tool ClusterFinder has only been recently developed; thus, the interest of the wider genome mining community to utilise this tool is still unknown (Cimermancic et al., 2014). ClusterFinder is based solely on Pfam domain frequencies obtained from 732 manually curated gene clusters and attempts to identify both known and unknown classes of molecules (Cimermancic et al., 2014). Some disadvantages of ClusterFinder compared to antiSMASH include the required command line interface experience and the need for the genome of interest to have been previously uploaded to JGI, annotated, and manually curated to obtain the required input file. A results based comparison of antiSMASH and ClusterFinder revealed that antiSMASH excelled in identifying previously characterised clusters and domains, while ClusterFinder identified considerably more unannotated clusters putatively responsible for small molecule biosynthesis (Cimermancic et al., 2014). Clearly, the best approach for the genome miner interested in identifying a wide range of natural product cluster classes is to use a combination of both tools.

With the exception of antiSMASH and ClusterFinder, the majority of bioinformatic genome mining tools have been designed to specifically identify the NRPS, PKS or hybrid NRPS/PKS class of gene clusters (for reviews with a non-cyanobacterial focus see Bachmann and Ravel (2009) and Boddy (2014)). This is due to the highly conserved and well characterised nature of NRPS/PKS pathways in addition to their usual clustering in the genome. A method developed by Kalaitzis et al. (2009) utilised a Pfam approach to search a genome for replicated genomic features essential to NRPS/PKS and hybrid biosynthetic pathways such as PF00501 (AMP-binding enzyme), PF00608 (condensation domain), and PF00550 (phosphopantetheine attachment site), and PF00109/PF02801 (β-ketoacyl synthase N- and C-terminal domains). Other programmes such as NP.searcher (Li et al., 2009), CLUSEAN (Weber et al., 2009), ClustScan (Starcevic et al., 2008) and ASMPKS (Tae et al., 2007) are capable of searching entire genomes. NaPDoS (Natural Product Domain Seeker) also screens complete genomes, but instead of identifying whole gene clusters, this tool specifically targets all C and KS domains within a query genome (Ziemert et al., 2012). NaPDoS then analyses the identified sequences by BLAST and phylogeny to gauge if the domain may belong to a unique cluster and thus may be responsible for a unique natural product. The user must manually link each identified domain with its corresponding gene cluster to determine the number of gene clusters present.

Once a complete gene cluster has been identified, the NRPS/PKS Analysis Website (see 2metdb for the standalone version) hosted by the University of Maryland (Bachmann and Ravel, 2009) is tailored toward investigating the domain type and order encoded within an input NRPS/PKS protein query sequence. Data of the domain architecture encoded by a single pathway can be combined with substrate specificity tools to generate a putative metabolite structure. To obtain these predictions, substrate specificity tools are often embedded within the natural product identification pipelines, such as antiSMASH and the NRPS/PKS Analysis Website. The aim of embedding substrate specificity tools into these pipelines is to allow the user to identify a cluster, determine the gene, module and domain organisation and the proposed substrate of each module. The embedded substrate specificity pipelines are based on stand-alone prediction tools that were developed after Challis et al. (2000) identified the A domain substrate binding pocket. Some stand-alone substrate prediction tools include: NRPSpredictor2 (Röttig et al., 2011), NRPSsp (Prieto et al., 2012), NRPS-PKS-substrate-predictor (Khayatt et al., 2013), SBSPKS (Sequence Based Sequence Analysis of Polyketide Synthases) (Anand et al., 2010), and MAPSI (Management and Analysis for Polyketide Synthase type I) (Tae et al., 2009) to further characterise a gene cluster. These programmes will specifically recognise the sequence of substrate incorporating domains (A for NRPS or AT for PKS) and will attempt to provide a putative substrate based on the binding pocket sequence. For example, once a NRPS cluster has been identified, substrate specificity tools can be utilised to predict the incorporated amino acids. The co-linearity rule can then be used to elucidate the structure and determine if the gene cluster may encode a novel molecule. Structural information will greatly aid the discovery efforts via the extraction, isolation, high performance liquid chromatography (HPLC) and liquid chromatography mass spectrometry (LCMS) processes. Many of the NRPS/PKS cluster analysis tools include inbuilt substrate specificity software. Unfortunately, many of the predictions are not consistent across different bioinformatic tools. Thus, it is recommended that a range of tools be utilised to increase the strength of substrate prediction.

3.1.1. Challenges of using these tools for cyanobacteria

The bioinformatic tools developed for analysing natural product gene clusters are useful for the analysis of genome sequences. However, in general, these tools have been designed based on actinomycete pathway architecture and in many cases, cyanobacterial gene clusters prove to be challenging for these programmes, requiring additional manual analysis by the researcher.

Cyanobacteria have been observed to contain many common natural product pathways that are uncommon in other phyla, such as the UV sunscreen molecules scytonemin and the MAAs. While the biosynthetic genes are common in cyanobacteria, these clusters encode several monofunctional enzymes that are difficult to identify using current software pipelines, such as antiSMASH. The scytonemin cluster is unable to be identified by antiSMASH and the detection of MAA clusters using bioinformatics programmes is dependent on the genetic architecture of the cluster in a particular organism. Not all MAA gene clusters contain a NRPS gene (Gao and Garcia-Pichel, 2011). However, these programmes have difficulty detecting other MAA biosynthesis genes. Therefore, MAA clusters that do not harbour the NRPS gene would not be detected using current bioinformatics software.

Structural elucidation of NRPS/PKS molecules based on sequence data is an extremely difficult process. While prediction algorithms work well for common and highly characterised domains, the identification of more natural product clusters has led to the discovery of many unusual domain architectures. This is particularly true in cyanobacteria, where several domain types, particularly initiation modules (Moore and Hertweck, 2002), cannot be recognised using the current prediction software. Often, if an A or C domain is specific for an unusual substrate, bioinformatics NRPS domain recognition fails. This is due to the limited number of domains present in the database and because many domains do not conform to known rules about NRPS/PKS. Biosynthesis of the cyanobacterial toxin microcystin initiates with the A domain of McyG, thought to activate phenylacetate based on structure (Tillett et al., 2000). However, *in vitro* studies have found that phenylpropanoids are preferentially activated (Hicks et al., 2006). Additionally, the A-KR di-domain encoded in HctE and HctF of the hectochlorin pathway is difficult to bioinformatically analyse using the substrate specificity code because these A-KR di-domains activate α-hydroxy acid instead of an amino acid (Magarvey et al., 2006; Ramaswamy et al., 2007). Where substrate specificity prediction fails, putative substrates can be suggested based on manually interrogating genomic flanking regions for substrate-biosynthesis genes. However, experimental structural validation is always needed to identify the activated substrate and clusters harbouring these domains cannot be analysed solely using NRPS

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

substrate prediction tools. The rare nature of these domains ensures that a combination of manual analysis and bioinformatic tools is needed for effective genome mining. An increase in the number of identified and characterised cyanobacterial natural product biosynthetic gene clusters is needed to improve the accuracy of bioinformatic predictions in the future.

Shih et al. (2013) and Wang et al. (2014) recently analysed the currently available cyanobacterial genomes and predicted the biosynthetic potential of these genomes to encode NRPS/PKS/hybrid gene clusters. While similar methods were used to identify these gene clusters, there is a large discrepancy between the number and type of NRPS/PKS/hybrid gene clusters between publications (Supplementary Tables 2 and 4). For example, Shih et al. (2013) predicted *Prochlorococcus marinus* str. MIT 9313 encodes one PKS gene cluster, while Wang et al. (2014) did not detect a PKS gene cluster within the same genome. Furthermore, Jones et al. (2014) analysed the orphan NRPS/PKS gene clusters encoded within the *M. producens* 3L genome and identified five orphan NRPS/PKS gene clusters in addition to the NRPS/PKS/hybrid curacin A, barbamide and carmabin gene clusters. However, Shih et al. (2013) predicted that 14 gene clusters were encoded within this cyanobacterial genome, including both known and orphan gene clusters. The differences in the number of putative biosynthetic gene clusters identified between these studies highlight several important issues in cyanobacterial genome mining, including inconsistent methodologies/cut-offs used to distinguish a biosynthetic gene cluster. Furthermore, the difficulty in experimentally validating cyanobacterial gene clusters has meant that researchers must rely solely on bioinformatic predictions, limiting functional knowledge of the genes and domains essential for the biosynthesis of cyanobacterial natural products. To overcome this, increased genome sequencing and the number of identified clusters will lead to increased confidence in bioinformatic tools. Additionally, manually curating all identified gene clusters for the required domains should always be performed to minimise the number of false positive identifications. Lastly, breakthroughs in the heterologous expression of cyanobacterial gene clusters will provide experimental evidence for genuine gene clusters which can then be further used to strengthen bioinformatic predictions.

3.2. MS-guided peptide natural product genome mining

The connection between natural product structures (chemotype) and their encoding biosynthetic gene clusters (genotype) has led to the development of MS-guided genome mining approaches, referred to as natural product peptidogenomics (NPP) (Kersten et al., 2011). This method has been spear-headed by the Dorrestein laboratory and is capable of detecting both ribosomal peptides and non-ribosomal peptides. Firstly, mass spectrometry is used to fragment peptides. The fragmentation patterns (mass shifts based on m/z) define the potential amino acid residues present within a ribosomal or non-ribosomal peptide. The mass shifts are converted into a search tag which are searchable in the genome mining query for ribosomal peptides (six-frame translation of the genome of interest) or non-ribosomal peptides (substrate specificity predicted peptide obtained from NRPS clusters within the genome of interest). With regards to the ribosomal peptides, the search tag is linked to the *C*-terminal half of the <100 aa precursor peptide that clusters with ribosomal peptide biosynthetic genes.

Recently, the NPP method was further developed into two platforms. RiPPquest and NRPquest are the first genome mining platforms for the discovery of ribosomal peptides and non-ribosomal peptides, respectively (Mohimani et al., 2014a,b). RiPPquest specifically targets peptides with poor fragmentation and a high number of modifications, such as the lanthipeptides. By using the RiPPquest workflow, Mohimani et al. (2014a) was able to discover several known *Streptomyces* lanthipeptides in addition to the novel class III lanthipeptide informatipeptin.

To complement RiPPquest, NRPquest (www.cyclo.ucsd.edu) can be utilised to search for bacterial non-ribosomal peptides. As input, NRPquest uses a sequenced genome and MS spectral files then follows a series of steps for non-ribosomal peptide identification. NRPquest utilises NRPSpredictor2 to predict putative non-ribosomal peptides encoded within a bacterial genome to provide a database of putative non-ribosomal peptides. Spectral data is then matched against the non-ribosomal peptide database and the statistical significance is calculated. Because of the large number of possible unusual substrates activated by A domains, NRPquest allows for a maximum of two blind modifications (mismatches). However, using this approach, nearly one hundred non-ribosomal peptides were identified from two strains of *Streptomyces* and two strains of *Bacillus* by Mohimani et al. (2014b), including several novel analogues of previously known ribosomal peptides. As the final step in the pipeline, both RiPPquest and NRPquest use molecular networking, a MS spectral visualisation tool that connects nodes (MS/MS spectra) and group into areas of analogous molecules (Watrous et al., 2012). Molecular networking has become a valuable tool itself, and in this case is used to increase the reliability of peptide identification and to help identify families of compounds based on structural and spectral similarities.

The advances in mass spectrometry for secondary metabolite identification and structural elucidation promise an interesting future for natural product discovery (Bouslimani et al., 2014). However, this approach has yet to be utilised for natural product discovery in cyanobacteria. It is well known that cyanobacteria are highly prevalent in the biosynthesis of NRPS and ribosomal peptides (Kehr et al., 2011; Goto et al., 2010). Thus, coinciding with the increase in cyanobacterial genome data, it is perfect timing for researchers to apply NPP to the discovery of cyanobacterial natural products. However, there are several issues to note when using an NPP approach. This method relies on detection of the biosynthesised peptide via MS. Gene clusters that are not expressed under the selected fermentation conditions will not be detected. Furthermore, the current method has difficulty elucidating hybrid NRPS/PKS, where the major constituents are derived from PKS, and will remain an issue until proper fragmentation rules are established (Kersten et al., 2011). Lastly, biosynthetic gene clusters are predicted based on bioinformatics and some clusters will be false positive identifications, particularly in the case of ribosomal peptides. This could be rectified manually but will be a subjective decision based on the individual researcher.

4. Heterologous expression

Heterologous expression of natural products is a growing field that has traditionally been applied to the production of compounds originating from *Streptomyces* species. Heterologous expression of natural product pathways is of particular interest as a tool for the production of unknown natural products from "orphan" gene clusters. Despite lagging behind *Streptomyces*, there has been some recent success in producing cyanobacterial peptides in a heterologous host, which may be employed for the characterisation of orphan gene clusters in the future.

4.1. Nonribosomal peptide and polyketide expression

The first cyanobacterial polyketide/non-ribosomal peptide to be expressed in a heterologous host was 4-O-demthylbarbamide from *M. producens* in *Streptomyces venezuelae* (Kim et al., 2012). The entire 26 kb barbamide gene cluster, comprising 12 open reading frames (*barA-barK*) was cloned into the pYJ1614 expression vector, and introduced into *S. venezuelae* DHS2001. Although barbamide was not detected in the organic extract, a naturally occurring derivative was produced by *S. venezuelae* DHS2001. 4-O-demethylbarbamide was found to be several-fold more potent as a molluscicidal agent than barbamide. This study demonstrated the first successful expression of a cyanobacterial PKS/NRPS gene cluster in a heterologous host.

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

Lyngbyatoxin is a cyanotoxin that was first isolated from cyanobacterial blooms dominated by *M. producens* (Cardellina et al., 1979). The biosynthetic pathway of lyngbyatoxin contains four open reading frames spanning over 11.3 kb (*ltxA-D*) (Edwards and Gerwick, 2004). The heterologous expression of lyngbyatoxin was first attempted in *Streptomyces coelicolor* (Jones et al., 2012). The study attempted to determine the suitability of *S. coelicolor* as a production host, and was successful in expressing proteins LtxB and LtxC individually, as well as showing functionality of LtxC in an *in vitro* assay. However, no lyngbyatoxin was produced when the entire biosynthetic pathway was overexpressed due to early transcriptional termination of *ltxA*.

A more successful attempt showed the production of lyngbyatoxin in an *Escherichia coli* host (Ongley et al., 2013b). *E. coli* GB2005-MtaA carrying a heterologous myxobacterial phosphopantetheinyl transferase was chosen as a host due to its similar GC content to that of the cyanobacterial genes. Heterologous expression of the *ltx* gene cluster resulted in a mixture of both lyngbyatoxin and indolactam-V, an intermediate in the pathway.

4.2. Ribosomal peptides

Heterologous expression of the patellamide biosynthetic pathway was the first functional expression of a marine natural product (Schmidt et al., 2005). The 12.5 kb gene cluster from the obligate symbiont *Prochloron didemni* was expressed in *E. coli* BL21 (DE3), leading to the production of patellamide A and C.

A relative of the patellamides, trunkamide, was also produced in *E. coli* by overexpressing the 11 kb *tru* gene cluster (Donia et al., 2008). The *tru* gene cluster was further used to demonstrate the potential of RiPPs as a source of novel bioactive variants. Peptide variants were produced by introducing mutations into the *truE* gene, and then coexpressing them alongside the rest of the pathway in an *E. coli* host (Tianero et al., 2012).

4.3. Alkanes, alkenes, and terpenes

Other secondary metabolites, such as alkanes and terpenes, have also been heterologously expressed in *E. coli*. Two open reading frames (ORF) from *S. elongatus* PCC 7942 were identified as candidates for alkane production (Schirmer et al., 2010). The two ORFs were expressed in *E. coli*, resulting in the production of alkanes, alkenes, chain fatty acids, and fatty alcohols.

Three terpene synthases from *Nostoc* were investigated by heterologous expression in *E. coli* (Agger et al., 2008). The terpene synthase homolog, referred to as NS1, from *Nostoc* (*Anabaena*) sp. strain PCC 7120, produced β-elemene as the major product, along with the thermally unstable germacrene A as a minor product. NP1, a terpene synthase from *N. punctiforme* PCC 73102 (ATCC 29133), was found to produce the rare sesquiterpene 8a-*epi*-α-selinene. Similarly, the putative geosmin synthase homolog, NP2, from *N. punctiforme* PCC 73102 (ATCC 29133) was heterologously expressed in *E. coli*. While expression did not yield geosmin, germacradienol and germacrene D were produced. This is likely due to the absence of activity in the C-terminal domain which catalyses the formation of geosmin from germacradienol, as demonstrated in streptomycete geosmin synthases (Agger et al., 2008). Giglio et al. (2008) corrected the sequencing error and expressed of the full length geosmin synthase gene (NPUNMOD) from *N. punctiforme* PCC 73012, which resulted in the biosynthesis of geosmin, germacradienol, germacrene D, octalin and (*E*)-nerolidol.

4.4. Awakening orphan biosynthetic pathways

The heterologous expression of orphan pathways is an important tool in unravelling their function (Gross, 2007; Chiang et al., 2011). The recently reported transformation associated recombination (TAR) represents a simple procedure for cloning large natural product clusters (Yamanaka et al., 2014). This method facilitated the capture and expression of a transcriptionally silent 67 kb NRPS gene cluster from *Saccharomonospora* sp. CNQ490 in the expression host *S. coelicolor*, resulting in the production of taromycin A. To date, no silent cyanobacterial gene clusters have been heterologously expressed. However, the recent successes in cloning large clusters and identification of suitable hosts outlined above pave the way for future expression of cyanobacterial gene clusters.

4.5. Important factors when expressing a foreign gene cluster

Choosing an expression host is one of the main challenges of heterologous expression. There are several factors that need to be considered when choosing an appropriate host. Disparity in GC content, and ultimately codon usage, can lead to reduced or abortive expression of the target pathway (Angov et al., 2008). Therefore, the selection of a production strain with a similar GC content to the desired gene cluster is ideal. On the other hand, changing the genetic code of the pathway to alter its codon usage without changing the peptide sequence can negate this issue (Angov, 2011), while also allowing the use of a range of expression hosts.

The amenability of a strain to genetic manipulation and the availability of genetic tools are important factors when choosing a host. The growth rate of the host and conditions required for production must also be taken into account. Organisms such as *E. coli* have been studied and genetically engineered for decades, making it an attractive option for the production of natural products. Other potential hosts are the model cyanobacteria *Synechocystis* sp. PCC6803 and *S. elongatus* PCC7002 (Ongley et al., 2013a,b). Although the cyanobacterial genetic tools are not as well developed as *E. coli*, there have been great advances over the last few years in tool development (Heidorn et al., 2011; Huang et al., 2010), largely due to the increased interest of using cyanobacteria for biofuel production (Lan and Liao, 2011; Varman et al., 2013).

5. Conclusions

Recent years have seen an explosion in the number of cyanobacterial genome sequences. These have yielded significant data highlighting the presence, diversity and distribution of known natural product gene clusters, in addition to the identification of orphan gene clusters. While there are a number of bioinformatic tools that can be employed specifically for the analysis of natural product gene clusters, many of them are problematic when used for the analysis of cyanobacterial pathways. This should be taken into consideration when analysing cyanobacterial natural product gene clusters, as additional manual analysis may be required. The heterologous expression of cyanobacterial natural product pathways will be required to take advantage of their potential as sources of novel compounds. Despite the slow progress, recent successes in the cloning of large gene clusters and expression of key pathways provide a platform upon which further research and natural product discovery can be based.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2014.11.009.

Acknowledgements

We thank Rocky Chau for assistance with the preparation and editing of this article. B.A.N. was funded by the Australian Research Council. M.L.M. was funded by the University of Western Sydney HDR Scholarship. B.A. was supported by an Australian Postgraduate Award.

References

Agger, S.A., Lopez-Gallego, F., Hoye, T.R., Schmidt-Dannert, C., 2008. Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. J. Bacteriol. 190, 6084–6096.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., Mohanty, D., 2010. SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res. 38, W487–W496.

Angov, E., 2011. Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol. J. 6, 650–659.

Angov, E., Hillier, C.J., Kincaid, R.L., Lyon, J.A., 2008. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS ONE 3, e2189.

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., Cotter, P.D., Craik, D.J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P.C., Entian, K.-D., Fischbach, M.A., Garavelli, J.S., Goransson, U., Gruber, C.W., Haft, D.H., Hemscheidt, T.K., Hertweck, C., Hill, C., Horswill, A.R., Jaspars, M., Kelly, W.L., Klinman, J.P., Kuipers, O.P., Link, A.J., Liu, W., Marahiel, M.A., Mitchell, D.A., Moll, G.N., Moore, B.S., Muller, R., Nair, S.K., Nes, I.F., Norris, G.E., Olivera, B.M., Onaka, H., Patchett, M.L., Piel, J., Reaney, M.J.T., Rebuffat, S., Ross, R.P., Sahl, H.-G., Schmidt, E.W., Selsted, M.E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R.D., Tagg, J.R., Tang, G.-L., Truman, A.W., Vederas, J.C., Walsh, C.T., Walton, J.D., Wenzel, S.C., Willey, J.M., van der Donk, W.A., 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. 30, 108–160.

Bachmann, B.O., Ravel, J., 2009. Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. In: David, A.H. (Ed.), Methods Enzymol vol. 458. Academic Press, pp. 181–217.

Balskus, E.P., Walsh, C.T., 2010. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. Science 329, 1653–1656.

Balskus, E.P., Case, R.J., Walsh, C.T., 2011. The biosynthesis of cyanobacterial sunscreen scytonemin in intertidal microbial mat communities. FEMS Microbiol. Ecol. 77, 322–332.

Begley, M., Cotter, P.D., Hill, C., Ross, R.P., 2009. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. Appl. Environ. Microbiol. 75, 5451–5460.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., Weber, T., 2013. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res. 41, W204–W212.

Boddy, C., 2014. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. J. Ind. Microbiol. Biotechnol. 41, 443–450.

Bode, H.B., Müller, R., 2005. The impact of bacterial genomics on natural product research. Angew. Chem. Int. Ed. 44, 6828–6846.

Bouslimani, A., Sanchez, L.M., Garg, N., Dorrestein, P.C., 2014. Mass spectrometry of natural products: current, emerging and future technologies. Nat. Prod. Rep. 31, 718–729.

Cardellina, J., Marner, F., Moore, R., 1979. Seaweed dermatitis: structure of lyngbyatoxin A. Science 204, 193–195.

Challis, G.L., Ravel, J., Townsend, C.A., 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. 7, 211–224.

Chiang, Y.-M., Chang, S.-L., Oakley, B.R., Wang, C.C.C., 2011. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. Curr. Opin. Chem. Biol. 15, 137–143.

Cimermancic, P., Medema Marnix, H., Claesen, J., Kurita, K., Wieland Brown Laura, C., Mavrommatis, K., Pati, A., Godfrey Paul, A., Koehrsen, M., Clardy, J., Birren Bruce, W., Takano, E., Sali, A., Linington Roger, G., Fischbach Michael, A., 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158, 412–421.

Coates, R.C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D.H., Allen, E.E., Gerwick, L., Gerwick, W.H., 2014. Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. PLoS ONE 9, e85140.

D'Agostino, P.M., Song, X., Neilan, B.A., Moffitt, M.C., 2014. Comparative proteomics reveals that a saxitoxin-producing and a nontoxic strain of *Anabaena circinalis* are two different ecotypes. J. Proteome Res. 13, 1474–1484.

Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., Gould, S.B., Goremykin, V.V., Rippka, R., Tandeau de Marsac, N., Gugger, M., Lockhart, P.J., Allen, J.F., Brune, I., Maus, I., Pühler, A., Martin, W.F., 2013. Genomes of stigonematalean cyanobacteria (Subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. Genome Biol. Evol. 5, 31–44.

D'Agostino, P.M., Moffitt, M.C., Neilan, B.A., 2014. Current knowledge of paralytic shellfish toxin biosynthesis, molecular detection and evolution. In: Rossini, G.P. (Ed.), Toxins and Biologically Active Compounds from Microalgae vol. 1. CRC Press, pp. 251–280.

Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F., Ley, R.E., 2013. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to cyanobacteria. eLife 2, e01102.

Dittmann, E., Wiegand, C., 2006. Cyanobacterial toxins—occurrence, biosynthesis and impact on human affairs. Mol. Nutr. Food Res. 50, 7–17.

Dittmann, E., Fewer, D.P., Neilan, B.A., 2013. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. FEMS Microbiol. Rev. 37, 23–43.

Donadio, S., Monciardini, P., Sosio, M., 2007. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Nat. Prod. Rep. 24, 1073–1109.

Donia, Mohamed S., Schmidt, Eric W., 2012. Linking chemistry and genetics in the growing cyanobactin natural products family. Chem. Biol. 18, 508–519.

Donia, M.S., Ravel, J., Schmidt, E.W., 2008. A global assembly line for cyanobactins. Nat. Chem. Biol. 4, 341–343.

Donia, M.S., Fricke, W.F., Partensky, F., Cox, J., Elshahawi, S.I., White, J.R., Phillippy, A.M., Schatz, M.C., Piel, J., Haygood, M.G., Ravel, J., Schmidt, E.W., 2011. Complex microbiome underlying secondary and primary metabolism in the tunicate-*Prochloron* symbiosis. Proc. Natl. Acad. Sci. U. S. A. 108, E1423–E1432.

Edwards, D.J., Gerwick, W.H., 2004. Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. J. Am. Chem. Soc. 126, 11432–11433.

Englund, E., Pattanaik, B., Ubhayasekera, S.J.K., Stensjö, K., Bergquist, J., Lindberg, P., 2014. Production of squalene in *Synechocystis* sp. PCC 6803. PLoS ONE 9, e90270.

Fewer, D.P., Jokela, J., Rouhiainen, L., Wahlsten, M., Koskenniemi, K., Stal, L.J., Sivonen, K., 2009. The non-ribosomal assembly and frequent occurrence of the protease inhibitors spumigins in the bloom-forming cyanobacterium *Nodularia spumigena*. Mol. Microbiol. 73, 924–937.

Fiore, M.F., Alvarenga, D.O., Varani, A.M., Hoff-Risseti, C., Crespim, E., Ramos, R.T.J., Silva, A., Schaker, P.D.C., Heck, K., Rigonato, J., Schneider, M.P.C., 2013. Draft genome sequence of the Brazilian toxic bloom-forming cyanobacterium *Microcystis aeruginosa* Strain SPC777. Genome Announc. 1, e00547-13.

Fischbach, M.A., Walsh, C.T., 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. 106, 3468–3496.

Frangeul, L., Quillardet, P., Castets, A.-M., Humbert, J.-F., Matthijs, H., Cortez, D., Tolonen, A., Zhang, C.-C., Gribaldo, S., Kehr, J.-C., Zilliges, Y., Ziemert, N., Becker, S., Talla, E., Latifi, A., Billault, A., Lepelletier, A., Dittmann, E., Bouchier, C., Tandeau de Marsac, N., 2008. Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. BMC Genomics 9, 274.

Gao, Q., Garcia-Pichel, F., 2011. An ATP-Grasp ligase involved in the last biosynthetic step of the iminomycosporine shinorine in *Nostoc punctiforme* ATCC 29133. J. Bacteriol. 193, 5923–5928.

Gao, Z.-M., Wang, Y., Tian, R.-M., Wong, Y.H., Batang, Z.B., Al-Suwailem, A.M., Bajic, V.B., Qian, P.-Y., 2014. Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont "*Candidatus Synechococcus spongiarum*". mBio 5, e00079-14.

Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M.J., van der Donk, W.A., 2010. Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. PLoS Biol. 8, e1000339.

Gross, H., 2007. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. Appl. Microbiol. Biotechnol. 75, 267–277.

Heidorn, T., Camsund, D., Huang, H.H., Lindberg, P., Oliveira, P., Stensjo, K., Lindblad, P., 2011. Synthetic biology in cyanobacteria engineering and analyzing novel functions. Methods Enzymol. 497, 539–579.

Hess, W.R., 2011. Cyanobacterial genomics for ecology and biotechnology. Curr. Opin. Microbiol. 14, 608–614.

Hicks, L.M., Moffitt, M.C., Beer, L.L., Moore, B.S., Kelleher, N.L., 2006. Structural characterization of *in vitro* and *in vivo* intermediates on the loading module of microcystin synthetase. ACS Chem. Biol. 1, 93–102.

Hilton, J.A., Foster, R.A., James Tripp, H., Carter, B.J., Zehr, J.P., Villareal, T.A., 2013. Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. Nat. Commun. 4, 1767.

Huang, H.H., Camsund, D., Lindblad, P., Heidorn, T., 2010. Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. Nucleic Acids Res. 38, 2577–2593.

Humbert, J.-F., Barbe, V., Latifi, A., Gugger, M., Calteau, A., Coursin, T., Lajus, A., Castelli, V., Oztas, S., Samson, G., Longin, C., Medigue, C., de Marsac, N.T., 2013. A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. PLoS ONE 8, e70747.

Ishida, K., Nakagawa, H., Murakami, M., 2000. Microcyclamide, a cytotoxic cyclic hexapeptide from the cyanobacterium *Microcystis aeruginosa*. J. Nat. Prod. 63, 1315–1317.

Jones, A.C., Monroe, E.A., Podell, S., Hess, W.R., Klages, S., Esquenazi, E., Niessen, S., Hoover, H., Rothmann, M., Lasken, R.S., Yates, J.R., Reinhardt, R., Kube, M., Burkart, M.D., Allen, E.E., Dorrestein, P.C., Gerwick, W.H., Gerwick, L., 2011. Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. Proc. Natl. Acad. Sci. U. S. A. 108, 8815–8820.

Jones, A.C., Ottilie, S., Eustáquio, A.S., Edwards, D.J., Gerwick, L., Moore, B.S., Gerwick, W.H., 2012. Evaluation of *Streptomyces coelicolor* A3(2) as a heterologous expression host for the cyanobacterial protein kinase C activator lyngbyatoxin A. FEBS J. 279, 1243–1251.

Kalaitzis, J.A., Lauro, F.M., Neilan, B.A., 2009. Mining cyanobacterial genomes for genes encoding complex biosynthetic pathways. Nat. Prod. Rep. 26, 1447–1465.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Tabata, S., 1996. 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. 3, 109–136.

Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., Nakamura, Y., Kasai, F., Watanabe, A., Kawashima, K., Kishida, Y., Ono, A., Shimizu, Y., Takahashi, C., Minami, C., Fujishiro, T., Kohara, M., Katoh, M., Nakazaki, N., Nakayama, S., Yamada, M., Tabata, S., Watanabe, M.M., 2007. Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. DNA Res. 14, 247–256.

Kehr, J., Gatte Picchi, D., Dittmann, E., 2011. Natural product biosyntheses in cyanobacteria: a treasure trove of unique enzymes. Beilstein J. Org. Chem. 7, 1622–1635.

Kersten, R.D., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.-J., Fenical, W., Fischbach, M.A., Moore, B.S., Dorrestein, P.C., 2011. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. Nat. Chem. Biol. 7, 794–802.

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

Khayatt, B.I., Overmars, L., Siezen, R.J., Francke, C., 2013. Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. PLoS ONE 8, e62136.

Kim, E.J., Lee, J.H., Choi, H., Pereira, A.R., Ban, Y.H., Yoo, Y.J., Kim, E., Park, J.W., Sherman, D.H., Gerwick, W.H., Yoon, Y.J., 2012. Heterologous production of 4-O-demethylbarbamide, a marine cyanobacterial natural product. Org. Lett. 14, 5824–5827.

Klähn, S., Baumgartner, D., Pfreundt, U., Voigt, K., Schoen, V., Steglich, C., Hess, W.R., 2014. Alkane biosynthesis genes in cyanobacteria and their transcriptional organization. Front. Bioeng. Biotechnol. 2.

Konz, D., Marahiel, M.A., 1999. How do peptide synthetases generate structural diversity? Chem. Biol. 6, R39–R48.

Kothari, A., Vaughn, M., Garcia-Pichel, F., 2013. Comparative genomic analyses of the cyanobacterium, *Lyngbya aestuarii* BL J, a powerful hydrogen producer. Front. Microbiol. 4, 363.

Lan, E.I., Liao, J.C., 2011. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. Metab. Eng. 13, 353–363.

Lee, S.W., Mitchell, D.A., Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V., Dixon, J.E., 2008. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc. Natl. Acad. Sci. U. S. A. 105, 5879–5884.

Leikoski, N., Fewer, D.P., Jokela, J., Wahlsten, M., Rouhiainen, L., Sivonen, K., 2010. Highly diverse cyanobactins in strains of the genus *Anabaena*. Appl. Environ. Microbiol. 76, 701–709.

Leikoski, N., Liu, L., Jokela, J., Wahlsten, M., Gugger, M., Calteau, A., Permi, P., Kerfeld Cheryl, A., Sivonen, K., Fewer David, P., 2013. Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. Chem. Biol. 20, 1033–1043.

Li, M., Ung, P., Zajkowski, J., Garneau-Tsodikova, S., Sherman, D., 2009. Automated genome mining for natural products. BMC Bioinformatics 10, 185.

Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P.J., Joewono, I., Rusch, D., Chisholm, S.W., van der Donk, W.A., 2010. Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. Proc. Natl. Acad. Sci. U. S. A. 107, 10430–10435.

Lima, A.R.J., Siqueira, A.S., dos Santos, B.G.S., da Silva, F.D.F., Lima, C.P., Cardoso, J.F., JLdSG, Vianez Júnior, Dall'Agnol, L.T., McCulloch, J.A., Nunes, M.R.T., Gonçalves, E.C., 2014. Draft genome sequence of the Brazilian *Cyanobium* sp. strain CACIAM 14. Genome Announc. 2, e00669-14.

Liou, G.F., Khosla, C., 2003. Building-block selectivity of polyketide synthases. Curr. Opin. Chem. Biol. 7, 279–284.

Magarvey, N.A., Ehling-Schulz, M., Walsh, C.T., 2006. Characterization of the cereulide NRPS α-hydroxy acid specifying modules: activation of α-keto acids and chiral reduction on the assembly line. J. Am. Chem. Soc. 128, 10698–10699.

Marahiel, M.A., Stachelhaus, T., Mootz, H.D., 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. 97, 2651–2674.

Mazmouz, R., Chapuis-Hugon, F., Mann, S., Pichon, V., Méjean, A., Ploux, O., 2010. Biosynthesis of cylindrospermopsin and 7-epicylindrospermopsin in *Oscillatoria* sp. strain PCC 6506: identification of the *cyr* gene cluster and toxin analysis. Appl. Environ. Microbiol. 76, 4943–4949.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39, W339–W346.

Méjean, A., Ploux, O., 2013. A genomic view of secondary metabolite production in cyanobacteria. In: Franck, C., Corinne, C.-C. (Eds.), Adv Bot Res vol. 65. Academic Press, pp. 189–234.

Méjean, A., Mazmouz, R., Mann, S., Calteau, A., Médigue, C., Ploux, O., 2010. The genome sequence of the cyanobacterium *Oscillatoria* sp. PCC 6506 reveals several gene clusters responsible for the biosynthesis of toxins and secondary metabolites. J. Bacteriol. 192, 5264–5265.

Micallef, M., Sharma, D., Bunn, B., Gerwick, L., Viswanathan, R., Moffitt, M., 2014. Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. BMC Microbiol. 14, 1–18.

Mohimani, H., Kersten, R.D., Liu, W.-T., Wang, M., Purvine, S.O., Wu, S., Brewer, H.M., Pasa-Tolic, L., Bandeira, N., Moore, B.S., Pevzner, P.A., Dorrestein, P.C., 2014a. Automated genome mining of ribosomal peptide natural products. ACS Chem. Biol. 9, 1545–1551.

Mohimani, H., Liu, W.-T., Kersten, R.D., Moore, B.S., Dorrestein, P.C., Pevzner, P.A., 2014b. NRPquest: coupling mass spectrometry and genome mining for nonribosomal peptide discovery. J. Nat. Prod.

Moore, B.S., Hertweck, C., 2002. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. Nat. Prod. Rep. 19, 70–99.

O'Sullivan, O., Begley, M., Ross, R.P., Cotter, P., Hill, C., 2011. Further identification of novel lantibiotic operons using LanM-based genome mining. Probiotics Antimicrob. Proteins 3, 27–40.

Ogino, J., Moore, R.E., Patterson, G.M.L., Smith, C.D., 1996. Dendroamides, new cyclic hexapeptides from a blue-green alga. Multidrug-resistance reversing activity of dendroamide A. J. Nat. Prod. 59, 581–586.

Ongley, S.E., Bian, X., Neilan, B.A., Muller, R., 2013a. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. Nat. Prod. Rep. 30, 1121–1138.

Ongley, S.E., Bian, X., Zhang, Y., Chau, R., Gerwick, W.H., Müller, R., Neilan, B.A., 2013b. High-titer heterologous production in *E. coli* of lyngbyatoxin, a protein kinase C activator from an uncultured marine cyanobacterium. ACS Chem. Biol. 8, 1888–1893.

Philmus, B., Christiansen, G., Yoshida, W.Y., Hemscheidt, T.K., 2008. Post-translational modification in microviridin biosynthesis. ChemBioChem 9, 3066–3073.

Prieto, C., García-Estrada, C., Lorenzana, D., Martín, J.F., 2012. NRPSsp: non-ribosomal peptide synthase substrate predictor. Bioinformatics 28, 426–427.

Ramaswamy, A.V., Sorrels, C.M., Gerwick, W.H., 2007. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. J. Nat. Prod. 70, 1977–1986.

Ran, L., Larsson, J., Vigil-Stenman, T., Nylander, J.A.A., Ininbergs, K., Zheng, W.-W., Lapidus, A., Lowry, S., Haselkorn, R., Bergman, B., 2010. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. PLoS ONE 5, e11486.

Rantala-Ylinen, A., Känä, S., Wang, H., Rouhiainen, L., Wahlsten, M., Rizzi, E., Berg, K., Gugger, M., Sivonen, K., 2011. Anatoxin-a synthetase gene cluster of the cyanobacterium *Anabaena* sp. strain 37 and molecular methods to detect potential producers. Appl. Environ. Microbiol. 77, 7271–7278.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., Stanier, R.Y., 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. 111, 1–61.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., Kohlbacher, O., 2011. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39, W362–W367.

Rouhiainen, L., Paulin, L., Suomalainen, S., Hyytiäinen, H., Buikema, W., Haselkorn, R., Sivonen, K., 2000. Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90. Mol. Microbiol. 37, 156–167.

Rouhiainen, L., Jokela, J., Fewer, D.P., Urmann, M., Sivonen, K., 2010. Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). Chem. Biol. 17, 265–273.

Rounge, T.B., Rohrlack, T., Nederbragt, A.J., Kristensen, T., Jakobsen, K.S., 2009. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. BMC Genomics 10, 1–11.

Schirmer, A., Rude, M.A., Li, X., Popova, E., del Cardayre, S.B., 2010. Microbial biosynthesis of alkanes. Science 329, 559–562.

Schmidt, E.W., Nelson, J.T., Rasko, D.A., Sudek, S., Eisen, J.A., Haygood, M.G., Ravel, J., Haselkorn, R., 2005. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. Proc. Natl. Acad. Sci. U. S. A. 102, 7315–7320.

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.A., 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. U. S. A. 110, 1053–1058.

Singh, S., Kate, B.N., Banerjee, U.C., 2005. Bioactive compounds from cyanobacteria and microalgae: an overview. Crit. Rev. Biotechnol. 25, 73–95.

Sinha, R.P., Häder, D.-P., 2008. UV-protectants in cyanobacteria. Plant Sci. 174, 278–289.

Sinha, R., Pearson, L.A., Davis, T.W., Muenchhoff, J., Pratama, R., Jex, A., Burford, M.A., Neilan, B.A., 2014. Comparative genomics of *Cylindrospermopsis raciborskii* strains with differential toxicities. BMC Genomics 15, 83.

Skulberg, O., 2000. Microalgae as a source of bioactive molecules — experience from cyanophyte research. J. Appl. Phycol. 12, 341–348.

Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., Steen, J.A., Parks, D.H., Tyson, G.W., Hugenholtz, P., 2014. An expanded genomic representation of the phylum cyanobacteria. Genome Biol. Evol. 6, 1031–1045.

Sorrels, C.M., Proteau, P.J., Gerwick, W.H., 2009. Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. Appl. Environ. Microbiol. 75, 4861–4869.

Soule, T., Garcia-Pichel, F., Stout, V., 2009a. Gene expression patterns associated with the biosynthesis of the sunscreen scytonemin in *Nostoc punctiforme* ATCC 29133 in response to UVA radiation. J. Bacteriol. 191, 4639–4646.

Soule, T., Palmer, K., Gao, Q., Potrafka, R., Stout, V., Garcia-Pichel, F., 2009b. A comparative genomics approach to understanding the biosynthesis of the sunscreen scytonemin in cyanobacteria. BMC Genomics 10, 336.

Starcevic, A., Zucko, J., Simunkovic, J., Long, P.F., Cullum, J., Hranueli, D., 2008. 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 Res. 36, 6882–6892.

Staunton, J., Weissman, K.J., 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18, 380–416.

Stucken, K., John, U., Cembella, A., Murillo, A.A., Soto-Liebe, K., Fuentes-Valdés, J.J., Friedel, M., Plominsky, A.M., Vásquez, M., Glöckner, G., 2010. The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications. PLoS ONE 5, e9235.

Sudek, S., Haygood, M.G., Youssef, D.T.A., Schmidt, E.W., 2006. Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. Appl. Environ. Microbiol. 72, 4382–4387.

Tae, H., Kong, E.-B., Park, K., 2007. ASMPKS: an analysis system for modular polyketide synthases. BMC Bioinformatics 8, 327.

Tae, H., Sohng, J.K., Park, K., 2009. Development of an analysis program of type I polyketide synthase gene clusters using homology search and profile hidden Markov model. J. Microbiol. Biotechnol. 19, 140–146.

Tianero, M.D., Donia, M.S., Young, T.S., Schultz, P.G., Schmidt, E.W., 2012. Ribosomal route to small-molecule diversity. J. Am. Chem. Soc. 134, 418–425.

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., Neilan, B.A., 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. Chem. Biol. 7, 753–764.

Tooming-Klunderud, A., Sogge, H., Rounge, T.B., Nederbragt, A.J., Lagesen, K., Glöckner, G., Hayes, P., Rohrlack, T., Jakobsen, K.S., 2013. From green to red: horizontal gene transfer of phycoerythrin gene cluster between *Planktothrix* strains. Appl. Environ. Microbiol. 79, 6803–6812.

Varman, A.M., Xiao, Y., Pakrasi, H.B., Tang, Y.J., 2013. Metabolic engineering of *Synechocystis* sp. strain PCC 6803 for isobutanol production. Appl. Environ. Microbiol. 79, 908–914.

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009

Vestola, J., Shishido, T.K., Jokela, J., Fewer, D.P., Aitio, O., Permi, P., Wahlsten, M., Wang, H., Rouhiainen, L., Sivonen, K., 2014. Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. Proc. Natl. Acad. Sci. U. S. A. 111, E1909–E1917.

Voß, B., Bolhuis, H., Fewer, D.P., Kopf, M., Möke, F., Haas, F., El-Shehawy, R., Hayes, P., Bergman, B., Sivonen, K., 2013. Insights into the physiology and ecology of the brackish-water-adapted cyanobacterium *Nodularia spumigena* CCY9414 based on a genome-transcriptome analysis. PLoS One 8, e60224.

Wang, H., Sivonen, K., Rouhiainen, L., Fewer, D.P., Lyra, C., Rantala-Ylinen, A., Vestola, J., Jokela, J., Rantasärkkä, K., Li, Z., 2012. Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium *Anabaena* sp. strain 90. BMC Genomics 13, 613–630.

Wang, H., Fewer, D.P., Holm, L., Rouhiainen, L., Sivonen, K., 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proc. Natl. Acad. Sci. U. S. A. 111, 9259–9264.

Watrous, J., Roach, P., Alexandrov, T., Heath, B.S., Yang, J.Y., Kersten, R.D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J.M., Moore, B.S., Laskin, J., Bandeira, N., Dorrestein, P.C., 2012. Mass spectral molecular networking of living microbial colonies. Proc. Natl. Acad. Sci. U. S. A. 109, E1743–E1752.

Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D.H., Wohlleben, W., 2009. CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J. Biotechnol. 140, 13–17.

Wiese, M., D'Agostino, P.M., Mihali, T.K., Moffitt, M.C., Neilan, B.A., 2010. Neurotoxic alkaloids: saxitoxin and its analogs. Mar. Drugs 8, 2185–2211.

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., Moore, B.S., 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc. Natl. Acad. Sci. U. S. A. 111, 1957–1962.

Yang, C., Zhang, W., Ren, M., Song, L., Li, T., Zhao, J., 2013. Whole-genome sequence of *Microcystis aeruginosa* TAIHU98, a nontoxic bloom-forming strain isolated from Taihu Lake, China. Genome Announc. 1, e00333-13.

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., Jensen, P.R., 2012. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS ONE 7, e34064.

Please cite this article as: Micallef, M.L., et al., Exploring cyanobacterial genomes for natural product biosynthesis pathways, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.11.009