

RESEARCH ARTICLE

Open Access



Genome mining for natural product biosynthetic gene clusters in the Subsection V cyanobacteria

Melinda L. Micallef¹, Paul M. D'Agostino^{1,2}, Deepti Sharma³, Rajesh Viswanathan³ and Michelle C. Moffitt^{1*}

Abstract

Background: Cyanobacteria are well known for the production of a range of secondary metabolites. Whilst recent genome sequencing projects has led to an increase in the number of publically available cyanobacterial genomes, the secondary metabolite potential of many of these organisms remains elusive. Our study focused on the 11 publically available Subsection V cyanobacterial genomes, together with the draft genomes of *Westiella intricata* UH strain HT-29-1 and *Hapalosiphon welwitschii* UH strain IC-52-3, for their genetic potential to produce secondary metabolites. The Subsection V cyanobacterial genomes analysed in this study are reported to produce a diverse range of natural products, including the hapalindole-family of compounds, microcystin, hapalasin, mycosporine-like amino acids and hydrocarbons.

Results: A putative gene cluster for the cyclic depsipeptide hapalasin, known to reverse *P*-glycoprotein multiple drug resistance, was identified within three Subsection V cyanobacterial genomes, including the producing cyanobacterium *H. welwitschii* UH strain IC-52-3. A number of orphan NRPS/PKS gene clusters and ribosomally-synthesised and post translationally-modified peptide gene clusters (including cyanobactin, microviridin and bacteriocin gene clusters) were identified. Furthermore, gene clusters encoding the biosynthesis of mycosporine-like amino acids, scytonemin, hydrocarbons and terpenes were also identified and compared.

Conclusions: Genome mining has revealed the diversity, abundance and complex nature of the secondary metabolite potential of the Subsection V cyanobacteria. This bioinformatic study has identified novel biosynthetic enzymes which have not been associated with gene clusters of known classes of natural products, suggesting that these cyanobacteria potentially produce structurally novel secondary metabolites.

Background

The phylum cyanobacteria consists of photosynthetic bacteria that are known to survive in a range of environments, and exhibit diverse morphology. The Subsection V cyanobacteria morphologically appear as true-branching filaments capable of forming heterocysts (specialised N₂ fixing cells), akinetes (cyst-like resting cells) and hormogonia (differentiated motile trichomes), making them one of the most morphologically advanced groups of cyanobacteria [1].

Cyanobacteria are prolific producers of secondary metabolites [2–4]. In particular, the Subsection V cyanobacteria are well known for the production of the

hapalindole-family of compounds, a group of structurally related indole alkaloids consisting of hapalindoles, welwitindolinones and fisherindoles which display a broad range of bioactivities [5, 6]. Other metabolites isolated from Subsection V cyanobacteria include the hepatotoxin microcystin [7–9], the cyclic peptide, hapalasin [10, 11], hydrocarbons [12], fischerellin A and B [13–15], the cyclic peptide westiellamide [16], the aromatic compounds ambigol A, B, C and 2,4-dichlorobenzoic acid [17, 18], the alkaloid tjipanazole D [18], the depsipeptide stigonemapeptin [19], the hexapeptide hapalocyclamide [20], the antimicrobial compound parsiguine [21], and the long chain polyunsaturated fatty acid γ -Linolenic acid [22].

The majority of cyanobacterial natural products are non-ribosomal peptides, polyketides or hybrid peptide-polyketide compounds [23–26]. Nonribosomal peptides

* Correspondence: m.moffitt@uws.edu.au

¹School of Science and Health, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia

Full list of author information is available at the end of the article

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 [27]. Similarly, polyketides are biosynthesised by polyketide synthases (PKS), which assemble polyketides from acyl-CoA in a sequential manner [27]. Each NRPS or PKS module contains a series of domains; a minimum NRPS module consists of condensation (C) domain for catalysing peptide bond formation, an adenylation (A) domain for selection of the substrate, and a peptide carrier protein (PCP) domain. The amino acid selected and incorporated by the A domain can be predicted through the ten critical amino acids comprising the A domain binding pocket. Similarly to the NRPS modules, a minimum PKS module consists of a ketosynthase (KS) domain, an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain, respectively. However, additional auxiliary domains, also known as tailoring domains, may also be present within each module, which creates structural diversity within the encoded natural product. Examples of NRPS auxiliary domains include epimerisation (E) domains, *N*-methyltransferase (NM) domains and heterocyclisation domains. PKS auxiliary domains include the β -ketoreductase (KR) domain, a dehydrogenase (DH) domain, enoyl-reductase (ER) domain, *O*-methyltransferase (MT) domains and *C*-methyltransferase (CM) domains [28, 29]. Furthermore, theioesterase (TE) domains and reduction domains are encoded in the final module for

chain termination and release of the polypeptide or polyketide natural product. The genes encoding natural product biosynthesis are generally clustered together on the genome [30, 31], aiding in structural prediction of the metabolite based on bioinformatics.

Another major class of cyanobacterial natural products are ribosomally-synthesised and post translationally-modified peptides, known as RiPPs, which are biosynthesised by post-ribosomal peptide synthesis (PRPS) [32]. The RiPP is encoded within the core peptide/region of the structural gene, which is then post-translationally modified into the final RiPP [32–35]. Other cyanobacterial natural products include the UV-absorbing compounds (mycosporine-like amino acids (MAAs) and scytonemin), terpenes and hydrocarbons.

The discovery of the diverse range and sources of these natural products has been aided by recent genome sequencing efforts. Prior to 2013, the Subsection V cyanobacteria were significantly under-represented compared to the other cyanobacterial subsections in terms of the number of sequenced genomes [36]. Two recent cyanobacterial sequencing projects aimed at increasing the number of Subsection V genomes led to a significant increase in the number of publically available genomes [1, 37]. Currently, there are 11 Subsection V cyanobacterial genomes publically available (Table 1), specifically *Fischerella* sp. PCC 9339, *Fischerella* sp. PCC 9431, *Fischerella* sp. JSC-11, *Fischerella* sp. PCC 9605, *Fischerella muscicola* PCC 7414, *Fischerella muscicola* SAG 1427–1, *Fischerella thermalis* PCC 7521, *Mastigocladopsis*

Table 1 Comparison of sequenced Subsection V cyanobacterial genomes

Cyanobacterial genome	Genome size (Mb)	GC content (%)	CDS	Number of scaffolds	Environmental data	Reference
<i>Fischerella</i> sp. PCC 9339	8.01	40.12	6,720	13	N/A	[37]
<i>Fischerella</i> sp. PCC 9431	7.17	40.17	6,104	8	N/A	[37]
FS JSC-11	5.38	41.05	4,671	34	N/A	N/A
<i>Fischerella</i> sp. PCC 9605	8.08	42.60	7,060	12	Limestone, Jerucham, Har Rahama, Israel	[37]
<i>F. muscicola</i> PCC 7414	6.90	41.26	6,060	270	Thermal springs, N/A	[1]
<i>F. muscicola</i> SAG 1427-1	7.36	40.32	6,533	524	Soil, Allahabad, India	[1]
<i>F. thermalis</i> PCC 7521	5.44	41.02	4,660	141	Thermal springs, Yellowstone National Park, USA	[1]
<i>M. repens</i> PCC 10914	6.46	43.52	5,846	3	Soil, Tarragona, Spain	N/A
<i>M. testarum</i> BC008	15.87	48.21	13,458	975	N/A	N/A
<i>C. fritschii</i> PCC 6912	7.75	41.48	6,836	161	Soil, Allahabad, India	[1]
<i>Chlorogloeopsis</i> sp. PCC 9212	7.65	41.50	6,717	188	Thermal springs, Orense, Spain	[1]
<i>W. intricata</i> UH strain HT-29-1	7.05	40.13	6,016	220	Soil, Moen Island, Truk Atoll, Caroline Islands	This study
<i>H. welwitschii</i> UH strain IC-52-3	7.27	40.21	6,139	169	Freshwater, Australian Institute of Marine Sciences, Queensland	This study
Average:	7.72	41.66	6,678	209.08		

repens PCC 10914, *Chlorogloeopsis fritschii* PCC 6912, *Chlorogloeopsis* sp. PCC 9212 and *Mastigocoleus testarum* BC008.

Previous genome mining has reported a preliminary overview of NRPS/PKS, PRPS and terpene genes from the five Subsection V cyanobacterial genomes sequenced by Shih et al. [37]. Additionally an in-depth analysis of NRPS/PKS gene cluster families, showing that the percentage of genome devoted to these gene clusters is higher in the Subsection V than other cyanobacterial subsections, has recently been reported by Calteau et al. [38].

Recently, the welwitindolinone (*wel*) gene cluster and the ambiguine (*amb*) gene cluster were identified from *H. welwitschii* UTEX B1830 (also known as *Fischerella* sp. PCC 9431) and *F. ambigua* UTEX 1903, respectively [39, 40]. Furthermore, the hapalindole (*hpi*), *wel* and *amb* gene clusters were also identified from the publically available *Fischerella* sp. PCC 9339, *Fischerella* sp. PCC 9431 and *F. muscicola* SAG 1427–1 genomes, and the recently sequenced *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3, *Fischerella* sp. ATCC 43239 and *F. ambigua* UTEX 1903 genomes [41]. The other biosynthetic gene clusters reported from the Subsection V cyanobacteria are a microcystin (*mcy*) gene cluster from *Fischerella* sp. PCC 9339 [37], MAA (*mys*) gene clusters from *C. fritschii* PCC 6912 [42], *Fischerella* sp. PCC 9339 [37] and *Fischerella* sp. PCC 9431 [38], and the fatty acyl ACP reductase (FAAR) and aldehyde deformylating oxygenase (ADO) pathway for hydrocarbon biosynthesis in all the Subsection V cyanobacterial genomes [12, 43].

It was the aim of this study to provide a complete overview of the diversity and distribution of secondary metabolite biosynthesis for all 11 publically available Subsection V cyanobacteria, in addition to two genomes sequenced by our research group, *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3. We present both known and orphan gene clusters belonging to the NRPS/PKS, PRPS (specifically cyanobactin, microviridin and bacteriocin gene clusters), UV-absorbing (MAA and scytonemin), hydrocarbon and terpene classes of natural products. This is the first study that includes identification and analysis of gene clusters from all natural product structural classes in the genomes of Subsection V cyanobacteria. Based on our analysis, we propose that the Subsection V cyanobacteria have the potential to produce a number of novel metabolites for which their structure and bioactivities have not yet been identified.

Methods

Genome data

The genomes of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 were obtained from genomic DNA (gDNA) extracted, sequenced, assembled and annotated as described in Micallef et al. [41]. Briefly,

cyanobacterial cultures of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 were obtained from the University of Hawaii cyanobacterial culture collection. gDNA was extracted as described in Morin et al. [44] and additional polysaccharides were removed as described in Wilson [45]. gDNA was sequenced, assembled and annotated by BGI (Beijing Genome Institute, China) using Illumina sequencing technology and Glimmer v3.0. The 11 publically available genomes were obtained from the Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) database and the National Centre for Biotechnology Information (NCBI).

PCR and sequencing reactions

PCR was used to identify the A-KR didomain of the hapalysin gene cluster in *H. welwitschii* UH strain IC-52-3 and to close any gaps in the nucleotide sequence of orphan NRPS/PKS gene clusters from *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 genomes. Comparison of orphan NRPS/PKS gene clusters identified from these genomes were compared, and any potential gaps in the nucleotide sequences (N's) were identified and targeted for sequencing. A 50 μ L PCR reaction mixture contained 10 pmol of specific forward and reverse primer or 50 pmol of degenerate primers (Additional file 1) (Geneworks, Australia), 1 \times PCR Buffer (KAPA Biosystems), 2.5 mM MgCl₂, 1 pmol dNTPs (Fisher Biotec), 1 U of KapaTaq polymerase (KAPA Biosystems) and 50 ng of gDNA template. *Pfu* DNA polymerase (Sigma) was used in addition to KapaTaq at a ratio of 1:10 (v/v). Hotstart PCR was performed by first heating the samples to 95 °C. Thermal cycling was then performed with a 5 min denaturation cycle at 95 °C, followed by 30 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min per 1 kb. Thermal cycling was concluded with a final extension at 72 °C for 7 min. PCR products were visualized in 1 % agarose gels in TAE buffer and single bands were gel extracted and purified using the QIAquick spin gel extraction kit (QIAGEN). Single sequencing reactions were submitted to the Ramaciotti Centre for Genomics at the University of New South Wales.

Bioinformatic software

All nucleotide sequences obtained from Illumina genome sequencing, annotated open reading frames, sequencing results from PCR products, and nucleotide sequences of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 genomes were organised and visualised using Geneious Version 6.1.7 created by Biomatters (available from <http://www.geneious.com/>). The 11 publically available Subsection V cyanobacterial genomes were downloaded from either the NCBI repository or the DOE Joint Genome Institute (JGI) server and visualised using Geneious Version 6.1.7.

Alignments of nucleotide sequences of individual genes and gene clusters were performed using Geneious alignment with default settings. For protein alignments, Clustal Omega (Version 1.2.1) was used with default settings, except the order of the aligned sequences was changed from aligned to input [46].

Genome mining of Subsection V cyanobacteria

Putative secondary metabolite gene clusters were originally identified using antiSMASH version 2.0 [47] with default settings. Annotations were refined manually using CDsearch and BLASTp (Basic Local Alignment Search Tool) to identify conserved domains [48, 49]. Each biosynthetic gene cluster was first categorised into the type of natural product encoded within the gene cluster. Comparative genomics identified homologous clusters in the genomes of the Subsection V cyanobacteria, and their gene organisation was compared. Homologous gene clusters in multiple genomes were also identified using the COG homology search tool in IMG JGI. NRPS/PKS gene clusters were analysed to determine domain structure and those reported have standard domain organization [27] as seen in the majority of other cyanobacterial NRPS/PKS gene clusters. Gene clusters located at the edge of contigs, encoding incomplete modules or containing gaps (N's) in nucleotide sequence were considered to be incomplete, unless stated otherwise.

Bioinformatic analysis of NRPS/PKS gene clusters

The domain organisation of NRPS and PKS gene clusters identified by antiSMASH was further analysed using the NRPS/PKS database [50]. The A domain substrate specificity for NRPS enzymes was predicted using NRSPredictor2 [51]. Furthermore, NaPDoS was used to identify C and KS domains [52].

Bioinformatic analysis of PRPS gene clusters

Cyanobactin, microviridin and bacteriocin biosynthetic gene clusters were first identified using antiSMASH [47]. Then, BLASTp [49] was used to determine potential false positives (cyanobactin and microviridin) or additional gene clusters (bacteriocin). In order to identify putative bacteriocin HetP-type and DUF37-type precursors in the Subsection V cyanobacterial genomes, BLASTp was utilised using previously identified precursors (Ava_0198 was used to identify HetP-type precursors and Ava_4222 was used to identify DUF37-type precursors). Protein alignments of precursor peptides were performed using Clustal Omega [46]. N11P and NHLP-type precursors were identified using antiSMASH. The bacteriocin gene clusters were then manually divided into the seven groups previously described by Wang et al. [34]. The seven groups are separated based on the presence of specific genes and domains. There are two types of ABC transporter

genes which encode C39 peptidases in bacteriocin gene clusters. The short type, which contain a C39 peptidase, an ABC transmembrane and an ATP-binding cassette domain, are found in groups III, IV and V. The long type, which contain an extra N-terminal nucleotide binding domain (CAP_ED), are found in group I, II and VI. The group V bacteriocin gene clusters also contain a gene encoding a bimodular protein containing only two CAP_ED domains. Groups IV and V also encode an additional ABC transporter with only an ABC transmembrane and an ATP-binding cassette domain. The group VII bacteriocin gene clusters contain a unique ABC transporter gene, which appears to be a fusion of a short and long type ABC transporter without the C39 peptidase domain in the long type. A HlyD protein, containing a type_1_hlyD domain, is found in every group, whilst a SurA protein, containing a rotamase domain, is found in groups I, II and VI. Bacteriocin gene clusters which could not be classified into these groups were labelled as unclassified.

Bioinformatic analysis of mycosporine-like amino acid gene clusters

To identify the *mys* gene cluster encoding MAA biosynthesis in the Subsection V cyanobacterial genomes, PCC9339DRAFT_04157 (encoding 3DHQS) of the previously identified MAA gene cluster in *Fischerella* sp. PCC 9339 [37] was used to perform a BLASTp search in the JGI/IMG database. Any positive matches were then manually searched for the essential O-MT, ATP-Grasp and NRPS-like/ATP-ligase enzymes downstream of the 3DHQS gene.

Bioinformatic analysis of hydrocarbon, terpenes and alkaloid gene clusters

The hydrocarbon biosynthetic gene cluster was identified using known genes (HT291_00281 and HT291_02280) encoding the FAAR/ADO pathway in the Subsection V cyanobacteria [12]. The sulfotransferase domain, characteristic of the olefin synthase (OLS) pathway, was used in BLASTp analysis. Terpene biosynthetic gene clusters were identified using antiSMASH, and grouped based on the presence of common genes, and compared to known terpene and squalene biosynthetic gene clusters.

Nucleotide sequence accession numbers

The draft genomes of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 are available from the US DOE JGI IMG server under the Taxon ID 2529292565 https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=-TaxonDetail&page=taxonDetail&taxon_oid=2529292565 and 2529292566 https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=-TaxonDetail&page=taxonDetail&taxon_oid=2529292566, respectively.

Results and discussion

Genome characteristics of *Westiella intricata* UH strain HT-29-1 and *Hapalosiphon welwitschii* UH strain IC-52-3

W. intricata UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 were chosen as candidates for analysis of secondary metabolite gene clusters, based on their potential to produce natural products of interest, specifically the hapalindole-type of natural products and hapalysin. Therefore, we conducted draft genome sequence analysis of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 not only to improve the genomic coverage of the Subsection V cyanobacteria, but also to identify and compare the secondary metabolite potential of these cyanobacteria. The genome sequence of *W. intricata* UH strain HT-29-1 was assembled into 220 scaffolds encoding 6,086 coding sequences (CDS), whilst that of *H. welwitschii* UH strain IC-52-3 was assembled into 169 scaffolds encoding 6,209 CDS. The genome size of *W. intricata* UH strain HT-29-1 was determined to be 7.05 Mb, whilst that of *H. welwitschii* UH strain IC-52-3 was slightly larger at 7.27 Mb. The GC content for *W. intricata* UH strain HT-29-1 was 40.13 %, whilst that of *H. welwitschii* UH strain IC-52-3 was highly similar with a GC content of 40.21 %. The sizes of the Subsection V cyanobacterial genomes sequenced range from 5.38 to 15.87 Mb, with an average of 7.9 Mb. The genome size of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 is slightly below this average, but well within the range of the Subsection V cyanobacterial genomes. Similarly, the number of CDS from the Subsection V cyanobacteria range from 4,671 to 11,113. Both *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 encode just below the average number of CDS, but within the range of the currently available genome sequences.

A survey of 102 housekeeping genes, previously identified as nearly universal in bacteria [53], were all identified in both *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 draft genomes (Additional file 2). The identification of all 102 housekeeping genes within both cyanobacterial genomes suggests a near complete genome. The largest percentage of functionally categorised genes in both *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 (based on cluster of orthologous groups categories, COGs) appears to be those involved in cell wall/membrane/envelope biogenesis (7 %), amino acid transport and metabolism (6.7 %) and signal transduction mechanisms (6 %) (Additional file 2).

Secondary metabolite biosynthetic gene clusters

A diverse range of secondary metabolite biosynthetic gene clusters were identified from the genomes of the

Subsection V cyanobacteria. These range from NRPS/PKS biosynthetic gene clusters, including a putative hapalysin (*hap*) gene cluster and RiPPs, including microviridin, cyanobactin and bacteriocin gene clusters. Additional gene clusters identified from the Subsection V cyanobacteria include the UV-absorbing compounds (MAA and scytonemin), hydrocarbons, terpenes, and the alkaloids (including the hapalindoles and welwitindolinones). Each Subsection V cyanobacterium dedicates between 0.87–3.5 % of the genome to secondary metabolite biosynthesis (Table 2) (however, this may be an underestimation as orphan NRPS/PKS gene clusters which appeared to be incomplete were not included in this calculation). *W. intricata* UH strain HT-29-1 and *Fischerella* sp. PCC 9339 dedicate the highest percentage of their genome to secondary metabolite biosynthesis (~3.5 %), followed closely by *H. welwitschii* UH strain IC-52-3 and *Fischerella* sp. PCC 9431 (~3.3 %). However, *M. testarum* BC008 dedicate the least percentage of its genome to secondary metabolite biosynthesis (~0.87 %) (Table 2).

NRPS/PKS biosynthetic gene clusters with a known product

Hapalysin gene cluster

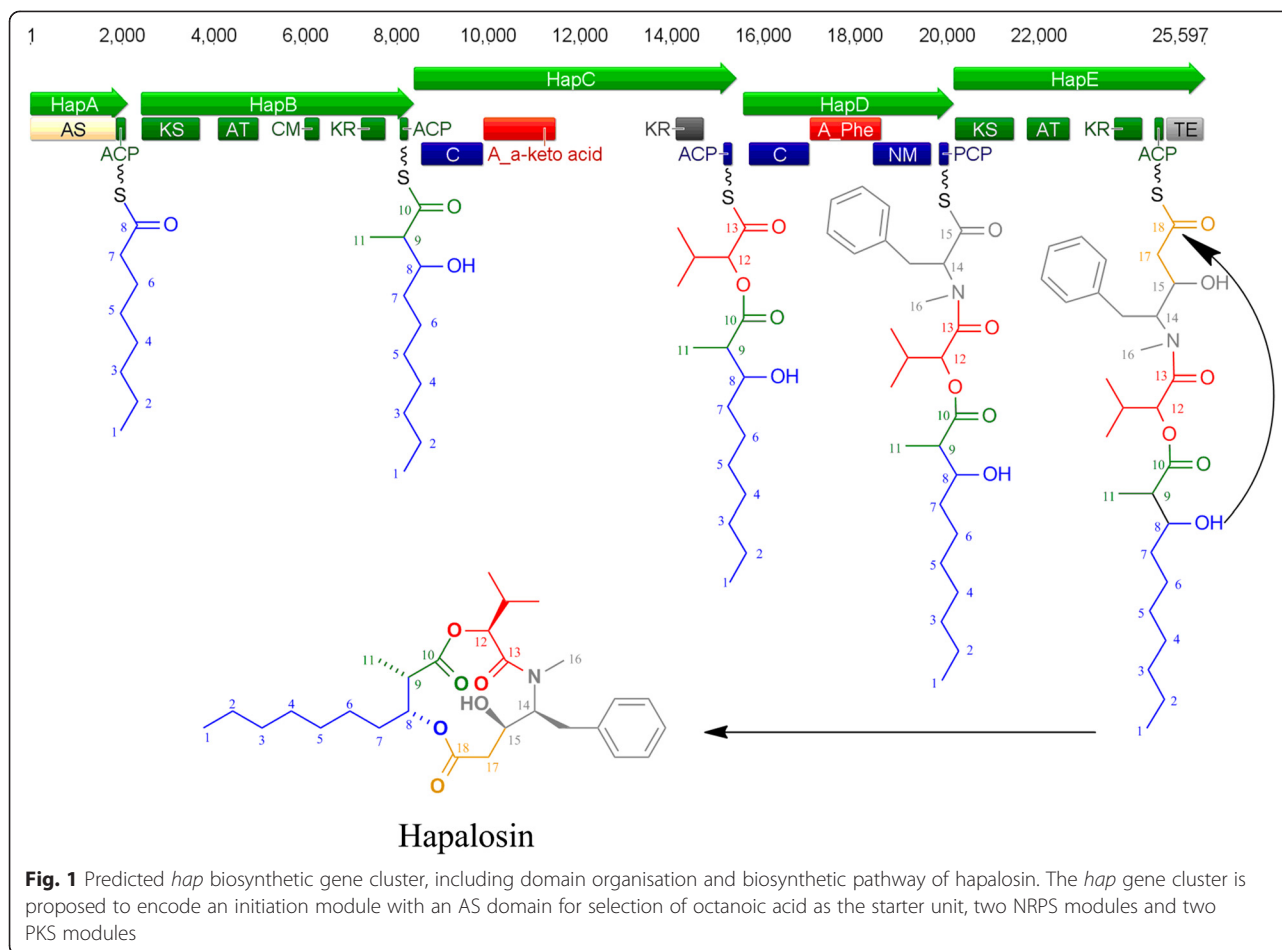
A candidate gene cluster for the biosynthesis of hapalysin was identified in three Subsection V cyanobacterial genomes, specifically *H. welwitschii* UH strain IC-52-3, *W. intricata* UH strain HT-29-1 and *Fischerella* sp. PCC 9431, although hapalysin has only been reported from *H. welwitschii* UH strain IC-52-3 [10]. Prior to genome sequencing, a degenerate primer approach (Additional file 1) targeting conserved regions of A-KR didomains in *H. welwitschii* UH strain IC-52-3 [54] was utilised as bait to identify the *hap* gene cluster within the *H. welwitschii* UH strain IC-52-3 genome, and then within the remaining Subsection V cyanobacterial genomes. Thus, the complete putative *hap* biosynthetic gene cluster was identified within *H. welwitschii* UH strain IC-52-3, *W. intricata* UH strain HT-29-1 and *Fischerella* sp. PCC 9431 genomes on a single scaffold. The gene cluster is proposed to be ~25.6 kb in length, comprising five genes, *hapA-E* (Additional file 3). Overall, there is greater than 99.2 % amino acid sequence similarity between the three gene clusters, suggesting the proteins are homologous and likely perform the same function in all strains. The overall genetic architecture and domain organisation is consistent with the proposed biosynthesis of hapalysin (Fig. 1).

hapA encodes a protein with the domain organisation AS-ACP, which displays high similarity to HctA and JamA from the hectochlorin (*hct*) and jamaicamide (*jam*) biosynthetic gene clusters, respectively [55, 56]. HctA is proposed to activate free hexanoic acid for

Table 2 Size and percentage of secondary metabolite biosynthetic gene clusters identified from the Subsection V cyanobacteria

Subsection V cyanobacterium	Type of cyanobacterial gene cluster and size (kb) ^Q											Percentage of genome
	NRPS/PKS/Hybrid			RiPPs			UV-absorbing compounds		Alkaloids			
	Hapalysin	Microcystin	Orphan NRPS PKS (number of orphan clusters)	Microviridin	Cyanobactin	Bacteriocin	MAA	Scytonemin	Hydrocarbon	Terpene	Hapalindoles/welwitindolinones	
<i>W. intricata</i> UH strain HT-29-1	25.6	-	42.2 (1)	-	11.7	82.5	7.4	-	2.3	19	59.3	3.5
<i>H. welwitschii</i> UH strain IC-52-3	25.6	-	56.6 (2)	-	-	71.2	7.4	-	2.3	19	55.8	3.3
<i>Fischerella</i> sp. PCC 9431	25.8	-	57.5 (2)	-	-	66.5	7.4	-	2.3	19	57.1	3.3
<i>Fischerella</i> sp. PCC 9339	-	67.5	75.5 (4)	8.8	-	65.7	6.6	-	2.2	13.6	44.9	3.5
<i>F. muscicola</i> SAG 1427-1	-	-	-	-	-	57.6	6.5	-	2.2	17.6	25.1	1.5
<i>C. fritschii</i> PCC 6912	-	-	55.4 (2)	-	-	56.9	8.1	-	1.9	7.9	-	1.7
<i>Chlorogloeopsis</i> sp. PCC 9212	-	-	55.4 (2)	-	-	56.9	8.1	-	1.9	7.9	-	1.7
<i>M. testarum</i> BC008	-	-	43.1 (2)	-	-	74.1	8.3	-	2.0	9.8	-	0.87
<i>M. repens</i> PCC 10914	-	-	63.3 (4)	-	-	40.8	-	32.9	1.9	6.6	-	2.3
<i>Fischerella</i> sp. PCC 9605	-	-	38.3 (3)	-	-	103.2	-	-	2.0	13.9	-	2.0
<i>F. muscicola</i> PCC 7414	-	-	-	-	-	52.7	-	-	2.2	13.6	-	1.0
<i>Fischerella</i> sp. JSC-11	-	-	-	-	-	56.3	-	-	2.2	13.5	-	1.3
<i>F. thermalis</i> PCC 7521	-	-	-	-	-	45.7	-	-	2.2	13.5	-	1.1

^QTotal size of all genes clusters within each category



initiation of hectochlorin biosynthesis, and JamA has been shown to activate 5-hexanoic acid [55, 56]. Based on the structure of hapalosin, HapA is likely to activate free octanoic acid forming C1-C8 (Fig. 1).

hapB, encoding a PKS module with the domain organisation KS-AT-CM-KR-ACP, likely incorporates malonyl-CoA to produce C9-C10 of hapalosin, followed by methylation of C9 to produce C11 *via* the CM domain, and reduction of the carbonyl group of C8 to a hydroxyl group by the KR domain. Next, *hapC* displays high sequence similarity to HctE and HctF of the *hct* gene cluster, NpnA of the nostophycin gene cluster and CrpD of the cryptophycin gene cluster [55, 57, 58]. HctE, HctF, NpnA and CrpD all contain a rare A-KR didomain within a NRPS module. These KR domains are found embedded between the core motifs A8 and A9 of the A domain [59], and within the A domain binding pocket, the invariable Asp235 of amino acid incorporating A domains has been replaced by Val235. This replacement was identified in HapC in all three gene clusters (Additional file 3). The A domain of HctE and HctF has been shown to incorporate 2-oxoisovaleric acid through ATP-PPi exchange assays [55]. The A domain binding pocket of HapC is

identical to HctE and HctF, and along with the structure of hapalosin, suggests an identical substrate is incorporated (Additional file 3). Therefore, *hapC*, encoding an NRPS module with the domain organisation C-A-KR-PCP is proposed to incorporate 2-oxoisovaleric acid, which is then reduced to 2-hydroxyisovaleric acid, producing C12 and C13 of hapalosin.

The NRPS HapD has the domain organisation C-A-NM-PCP. Analysis of the A domain binding pocket suggests phenylalanine is selected and activated by HapD (Additional file 3), producing C14 and C15 of hapalosin. Subsequent *N*-methylation of phenylalanine by the NM domain leads to the biosynthesis of C16 methyl substituent.

hapE encodes a single PKS module with the domain organisation KS-AT-KR-ACP-TE, which is identical to the PKS organisation of JamP, which displays the highest amino acid sequence similarity with HapE. Analysis of the AT domain of HapE suggests malonyl-CoA is incorporated, forming C17 and C18 of hapalosin. Subsequent reduction of the carbonyl group of C15 produces the hydroxyl group. The TE domain is proposed to break the

thioester bond connecting the hapalysin chain to the ACP domain, enabling the oxygen of C8 to attack C18, forming the cyclised final structure. Together, the domain architecture and prediction of substrates is consistent with the structure of hapalysin.

The complete NRPS/PKS biosynthetic gene cluster for hapalysin biosynthesis is the first report of a biosynthetic gene cluster for this natural product. The *hap* gene cluster demonstrates similar domain architecture to the *hct* gene cluster from *L. majuscula* [55]. The first three modules are almost identical, with the exclusion of a halogenase domain from *hctB*, and a CM domain from *hctD*. The identification of the *hap* biosynthetic gene cluster in three of the Subsection V cyanobacteria was surprising, as only *H. welwitschii* UH strain IC-52-3 was reported to produce hapalysin. The conservation of the *hap* gene cluster across three genera suggests there is an unknown evolutionary significance/benefit to the organism, similar to the conservation of the saxitoxin gene cluster within cyanobacteria [60].

Microcystin gene cluster

The only known NRPS/PKS gene cluster previously identified from the Subsection V cyanobacterial genomes is the *mcy* gene cluster from *Fischerella* sp. PCC 9339. According to Shih et al. [37], the 67.5 kb gene cluster encodes *mcyA-I*, and there is an additional PKS gene with partial sequence similarity to *npnA* from the nostophycin gene cluster from *Nostoc* sp. 152 [57], and *mcyB* is located on the border of two contigs. The remaining Subsection V cyanobacterial genomes were screened for a *mcy* and other toxin biosynthetic gene clusters, however, no other toxin biosynthetic gene clusters were identified.

Orphan NRPS/PKS biosynthetic gene clusters

A total of 103 orphan NRPS/PKS/Hybrid gene clusters were identified from the Subsection V cyanobacterial genomes which produce unknown products. Analysis of the NRPS and PKS domain composition enabled these gene clusters to be divided into those that likely encode NRPS/PKS natural products and those which appear to be incomplete gene clusters. A total of 17 gene clusters were identified from nine of the Subsection V cyanobacterial genomes which are potentially complete and have been given arbitrary cluster identification numbers (Additional file 4). Four of these gene clusters were identified in more than one cyanobacterial strain. These gene clusters were then compared with the remaining identified gene clusters. Two incomplete gene clusters share homology with a complete gene cluster from another genome (Additional file 5). The remaining gene clusters were categorised as incomplete, either encoding incomplete modules, gaps (N's) in nucleotide

sequence, or located on contig borders (Additional file 5). These incomplete gene clusters will not be discussed.

Three NRPS gene clusters (Cluster 1–3) were identified from four Subsection V cyanobacterial genomes which lack a C domain in the initiation module. There are four additional gene clusters (Clusters 14–17) which also contain genes encoding NRPS modules, however, the first module also encodes a C domain. It is unclear if these gene clusters are incomplete (missing an initiation module) or if these gene clusters encode an NRPS natural product.

The remaining 10 gene clusters which likely encode a NRPS/PKS natural product all encode an initiation module containing an AS domain which likely selects for a fatty acid molecule as the starter unit for biosynthesis, such as those observed in the *jam*, *hct* and *hap* gene clusters. AS domains have only been identified as initiation modules in these three NRPS/PKS gene clusters within cyanobacteria. This high proportion of AS domains within initiation modules within the Subsection V cyanobacteria is therefore unique.

Overall, prediction of the substrate selected and incorporated by the A domain was difficult due to the low similarity of the A domain binding pocket to known amino acids in the NRSPredictor2 database. The nearest neighbour for each A domain binding pocket is reported, and the percentage identity to nearest signature for each A domain binding pocket is provided in Additional file 4. However, due to this low similarity for some A domains, and the unknown length of the proposed fatty acid substrate incorporated by the AS domain, the predicted structure of the encoded natural product cannot be proposed.

The majority of the orphan NRPS/PKS gene clusters analysed in this study included genes encoding tailoring enzymes which are not typically identified within NRPS/PKS gene clusters. Four NRPS/PKS gene clusters identified from the Subsection V cyanobacterial genomes (Cluster 6–9) contained genes encoding for dioxygenases and/or glycosyltransferases. The dioxygenase genes identified within these gene clusters display similarity to α -ketoglutarate-dependent, taurine dioxygenases. These enzymes catalyse the hydroxylation of taurine to produce sulphite and aminoacetaldehyde [61]. The dioxygenase genes precede NRPS genes within each orphan NRPS/PKS gene cluster, and are possibly involved in hydroxylation of the selected amino acid. Only Cluster 8 encodes a glycosyltransferase gene within an orphan NRPS/PKS gene cluster, which is likely to be involved in the addition of a sugar moiety on the encoded natural product. Glycosylated natural products display a wide range of bioactivities including insecticidal [62] and antitumor [63] activity, amongst others [64].

There are two gene clusters identified from the Subsection V cyanobacterial genomes (Cluster 10 and 11) which encode fatty acid desaturase genes within an NRPS/PKS gene cluster. These genes are located downstream from the AS domain, suggesting the fatty acid incorporated by the AS domain may be desaturated in the final natural product. In Cluster 3, a gene encoding a fatty acid desaturase is located upstream from the NRPS gene cluster; however, as this gene cluster does not encode an AS domain, the function of this enzyme remains unknown. The remaining orphan NRPS/PKS gene clusters contain genes encoding hypothetical proteins (Clusters 11–13), aspartate racemase (Cluster 9) and asparagine synthase and racemase (Cluster 17). However, the effect of these proteins on the encoded natural product cannot be predicted at this time.

Genome mining orphan NRPS/PKS gene clusters from the Subsection V cyanobacteria has uncovered novel biosynthetic gene clusters with unique genes, however, screening for the products themselves or characterisation of the enzymatic pathways is necessary to determine if the gene clusters are functional, or remnants of evolution. If functional, these gene clusters, which encode new enzymes within NRPS/PKS gene clusters, may perform new biosynthetic reactions, which may potentially produce natural products with enhanced or new bioactivities. The ability to successfully express cyanobacterial gene clusters in *Streptomyces* and *E. coli* hosts [65–71] means these gene clusters could potentially be characterised, and the bioactivity (if any) could eventually be determined.

There is also the potential to identify new products from previously identified biosynthetic gene clusters, such as the *hap* or *mcy* gene clusters. Characterisation of the barbamide gene cluster from *M. producens* in *Streptomyces venezuelae* [65] led to the production of the previously unidentified 4-*O*-demethylbarbamide, which was found to be more potent as a molluscicidal agent than barbamide. Therefore, characterisation of these biosynthetic gene clusters has the potential to lead to the identification of natural product analogues with enhanced bioactivities.

PRPS biosynthetic gene clusters

The RiPPs are a growing class of natural products which are being increasingly recognised within cyanobacteria. Three distinct classes of RiPPs were identified from the Subsection V cyanobacteria in this study, specifically cyanobactins, microviridins and bacteriocins.

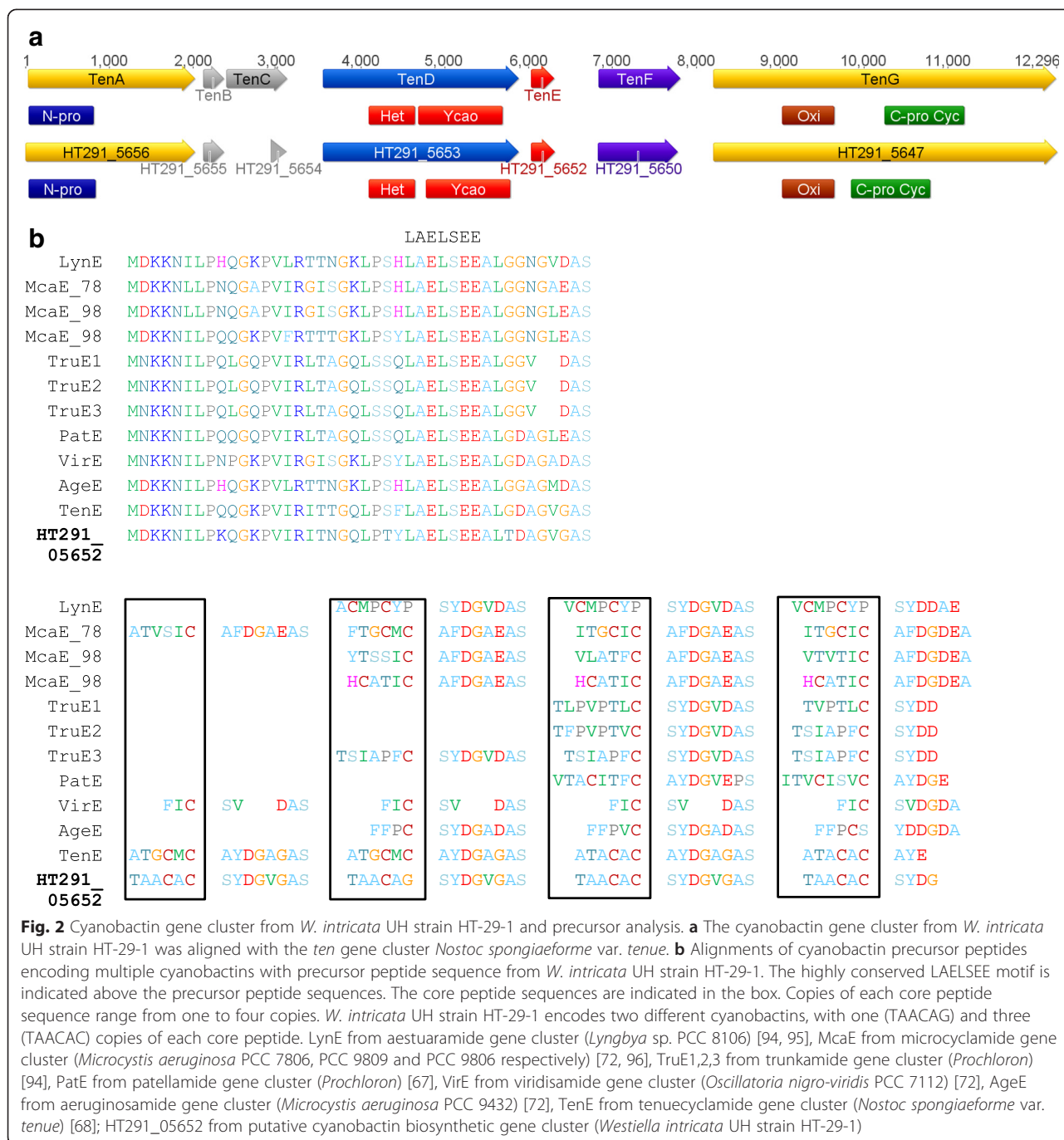
Cyanobactin biosynthetic gene cluster

Cyanobactins, N-C terminally cyclised peptides, are often prenylated or contain heterocyclised cysteine, serine or threonine residues, however linear cyanobactins have

recently been discovered [72, 73]. A single cyanobactin biosynthetic gene cluster was identified from the Subsection V cyanobacteria, specifically from the genome of *W. intricata* UH strain HT-29-1 (Fig. 2a). The putative 11.7 kb cyanobactin biosynthetic gene cluster from *W. intricata* UH strain HT-29-1 encodes seven genes, which share homology to the tenuencyclamide (*ten*) gene cluster from *Nostoc spongiaeforme* var. *tenue* [68]. The cyanobactin gene cluster from *W. intricata* UH strain HT-29-1 appears to be intact, although a truncated homologue of the non-essential *tenC* gene is present. Within the putative *W. intricata* UH strain HT-29-1 cyanobactin gene cluster, HT291_05652 demonstrates homology with TenE, the precursor peptide for tenuencyclamide biosynthesis. Protein alignments of HT291_05652 with other known cyanobactin precursors revealed the presence of the highly conserved LAELSEE motif in the leader sequence, and the presence of four core peptide sequences. Comparison of the four core peptide sequences revealed one (TAACAG) and three (TAACAC) copies of the core peptide sequences, suggesting two different cyanobactins are biosynthesised by *W. intricata* UH strain HT-29-1 (Fig. 2b).

The remaining genes in the *W. intricata* UH strain HT-29-1 cyanobactin gene cluster are required for modifications of the precursor peptide. HT291_05656 and HT291_05647, which are homologous to TenA and TenG, respectively, are predicted to cleave the N- and C-terminus of the precursor peptide. Furthermore, PatG (homologous to TenG) has been shown to be responsible for macrocyclisation to form the final cyclic peptide [74], suggesting the final peptide from *W. intricata* UH strain HT-29-1 may also be cyclised.

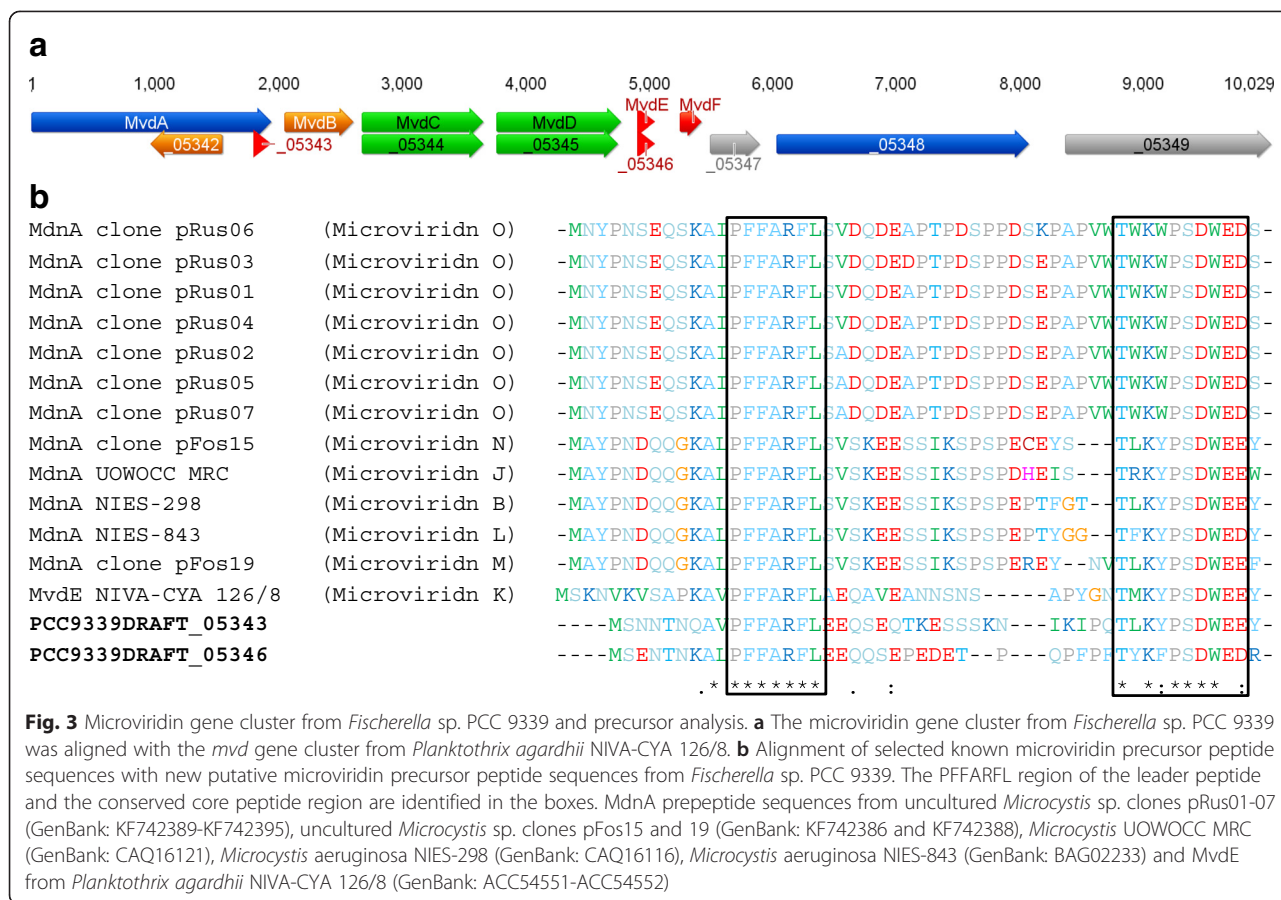
Tenuencyclamides contain heterocyclised amino acids [75], which are proposed to be catalysed by TenD and the oxidase domain of TenG [68]. HT291_05653 is homologous to TenD, whilst HT291_05647 is homologous to TenG and contains an oxidase domain, suggesting the tyrosine and cysteine residues in the core peptide sequences may be heterocyclised in the final product. Furthermore, a homologous protein to the putative prenyltransferase TenF was identified in the cyanobactin gene cluster from *W. intricata* UH strain HT-29-1 (HT291_05650). The presence of this protein encoded in the cyanobactin gene cluster from *W. intricata* UH strain HT-29-1 suggests tyrosine may also be prenylated in the final cyanobactin product. The single precursor gene from *W. intricata* UH strain HT-29-1 encoding two different cyanobactins both appear to be novel, even though over 100 cyanobactin variants are known [76]. Recent genome mining of 126 cyanobacterial genomes revealed approximately 24 % of cyanobacterial strains encoded a cyanobactin



biosynthetic gene cluster [72], however, many of these gene clusters are predicted to be non-functional [72], although active versions have been identified within closely related strains [77]. The identification of the cyanobactin gene cluster within *W. intricata* UH strain HT-29-1 encoding two novel cyanobactins contributes to the structural diversity of cyanobactins within cyanobacteria, whilst highlighting the diversity of natural products encoded within the Subsection V cyanobacteria.

Microviridin biosynthetic gene cluster

The microviridins are a family of *N*-acetylated tricyclic depsipeptides which contain a rare cage-like architecture [78, 79]. The peptide sequence is encoded within the ribosomal precursor peptide, which undergoes post-translational modifications including macrocyclisation by ATP grasp-type ligases [80, 81]. Genome mining of the Subsection V cyanobacterial genomes revealed only *Fischerella* sp. PCC 9339 encoded a microviridin (*mvd*) gene cluster (Fig. 3a). The novel eight gene biosynthetic



cluster displays high sequence similarity with the *mvd* gene cluster from *Planktothrix agardhii* NIVA-CYA 126/8 [80]. The 8.8 kb *mvd* gene cluster from *Fischerella* sp. PCC 9339 encodes two precursor peptides (PCC9339DRAFT_05343 and PCC9339DRAFT_05346). Protein alignments with other known *mvd* precursor peptide sequences revealed the *Fischerella* sp. PCC 9339 peptide sequences were novel, therefore, the encoded microviridin variant could not be proposed (Fig. 3b). The putative *mvd* gene cluster from *Fischerella* sp. PCC 9339 encodes two cyclisation proteins for amide and ester bond formation (homologous to MvdC and MvdD), as well as an ABC transporter (homologous to MvdA). Furthermore, the putative *mvd* gene cluster from *Fischerella* sp. PCC 9339 encodes a protein similar to GCN5-related *N*-acetyltransferase proteins. The microviridin gene cluster from *M. aeruginosa* NIES 298 also encodes a GCN5-related *N*-acetyltransferase, suggesting the microviridin variant produced by *Fischerella* sp. PCC 9339 may be acetylated [82]. Finally, the putative *mvd* gene cluster from *Fischerella* sp. PCC 9339 also encodes a protein belonging to the GUN4 superfamily and a hypothetical protein. A putative *mvd* biosynthetic gene cluster from *N. spumigena* CCY 9414 also encodes a protein belonging to the GUN4 superfamily [80],

however, the function and/or involvement of these proteins in microviridin biosynthesis remains unknown.

Bacteriocin biosynthetic gene clusters

Bacteriocins are another major class of RiPPs identified from cyanobacteria. Bacteriocins are encoded within short ribosomally produced precursor peptides, which contains both the core peptide and leader peptide [83]. The leader peptide is then cleaved by a peptidase domain during maturation [84]. The core peptide sequence can then undergo further post-translational modification, including macrocyclisation, dehydration, heterocyclization, as well as lanthionine formation, to produce the final RiPP [85, 86].

From the 13 Subsection V cyanobacterial genomes analysed, a total of 116 bacteriocin gene clusters were identified (Table 3). The gene clusters were organised into the seven groups according to Wang et al. [34] (Fig. 4) based on the presence of the genes in the gene cluster, however, this study identified nine gene clusters which were unable to be classified (Table 3).

The group I were the most abundant type identified from the Subsection V cyanobacterial genomes (Table 3). Six of the group I bacteriocin gene clusters contained a

Table 3 Summary of bacteriocin biosynthetic gene clusters identified from Subsection V cyanobacteria, separated into groups according to Wang et al., [34]

Organism	I	II	III	IV	V	VI	VII	Unclassified	Total
<i>W. intricata</i> UH strain HT-29-1	3	2	1	1	1	1	0	0	9
<i>H. welwitschii</i> UH strain IC-52-3	3	2	1	2	1	1	0	0	10
<i>Fischerella</i> sp. PCC 9339	3	2	1	0	1	1	0	1	9
<i>Fischerella</i> sp. PCC 9605	4	2	2	1	1	1	0	1	12
<i>Fischerella</i> sp. PCC 9431	4	1	1	1	1	1	0	0	9
<i>Fischerella</i> sp. JSC-11	3	1	2	1	0	1	0	0	8
<i>F. muscicola</i> SAG 1427-1	3	1	1	1	1	1	0	0	8
<i>F. muscicola</i> PCC 7414	3	1	2	0	0	1	0	1	8
<i>F. thermalis</i> PCC 7521	3	0	2	1	0	1	0	0	7
<i>Chlorogloeopsis</i> sp. PCC 9212	2	0	3	2	0	1	0	2	10
<i>C. fritschii</i> PCC 6912	2	0	3	2	0	1	0	2	10
<i>M. testarum</i> BC008	3	2	2	1	1	0	0	1	10
<i>M. repens</i> PCC 10914	2	1	0	1	0	1	0	1	6
Total:	38	15	21	14	7	12	0	9	116

gene encoding M41 peptidases, which were not previously identified in other bacteriocin gene clusters. Genes encoding M16 peptidases (characteristic of the group II bacteriocin gene clusters) were only identified from the genomes of *M. repens* PCC 10914 and *Fischerella* sp. PCC 9605 (Additional file 6).

The second most abundant group of bacteriocin gene clusters was group III (Table 3). The group IV bacteriocin gene clusters, which contain the LanM-type genes, were identified in 11 of the Subsection V cyanobacterial genomes. The single LanM gene was identified in six genomes however, in two of these genomes (*Chlorogloeopsis* sp. PCC 9212 and *C. fritschii* PCC 6912), a single type 2 lantibiotic gene was identified clustered with the LanM gene (Additional file 6). The group IV bacteriocin gene cluster from *Fischerella* sp. PCC 9605 encodes an S8 peptidase, which according to Wang et al. [34], has only been previously identified in group VI bacteriocin gene clusters. No group VII bacteriocin gene clusters, containing the fused ABC transporter, were identified from the Subsection V cyanobacterial genomes.

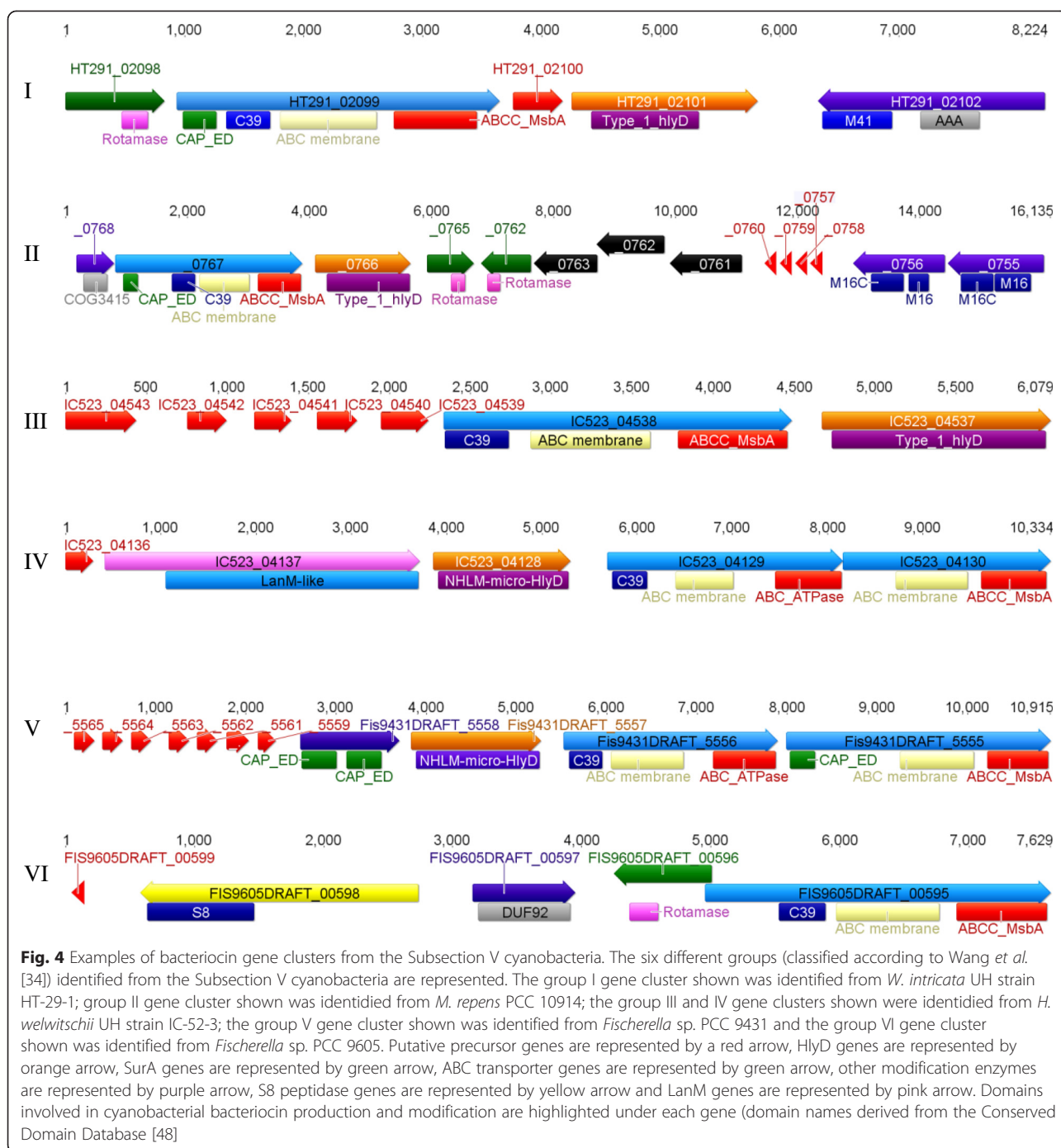
Within each of the Subsection V cyanobacterial genomes, a single putative HetP-type precursor was identified from a group I bacteriocin gene cluster. Protein alignments of these putative precursors with previously identified HetP type precursors from Wang et al. [34] identified the leader peptide cleavage motif (KIXDLXYLEX₁₀GG) (Additional file 6).

The putative DUF37-type precursors were also identified from all the Subsection V cyanobacterial genomes, however, none of these genes were clustered with other

bacteriocin biosynthetic genes. Protein alignments of these putative precursors with previously identified DUF37 precursors identified by Wang et al. [34] enabled the identification of a putative leader peptide cleavage motif from the Subsection V cyanobacterial precursors (Additional file 6). These precursors do not demonstrate sequence similarity to any known proteins in the NCBI database, and therefore their function cannot be predicted at this time.

The N11P-type precursor peptide was identified in all of the Subsection V cyanobacterial genomes. However, the double glycine motif for peptide cleavage was unable to be identified in any of the precursor peptide sequences (Additional file 6). The NHLP-type precursor was only identified from *Fischerella* sp. PCC 9605. This putative precursor peptide was aligned with selected NHLP precursor peptides from *N. punctiforme* PCC 73102, and a conserved region near the peptide cleavage site (double glycine motif) was identified [87] (Additional file 6).

The wide distribution and range of bacteriocin gene clusters identified from the Subsection V cyanobacteria is consistent with previous reports from other cyanobacteria. Recent genome mining studies by Wang et al. [34] identified 145 bacteriocin gene clusters from 43 cyanobacteria, and Shih et al. [37] identified 358 bacteriocin gene clusters from 106 cyanobacterial genomes. Shih et al. [37] identified 23 bacteriocin (including lantipeptide) gene clusters from the five Subsection V cyanobacteria analysed, however, in this study, we identified 38 bacteriocin gene clusters from the same genomes. Wang et al. [34] identified M16 and S8 peptidase domains located within bacteriocin



gene clusters. In this study, M41 peptidase domains, also known as FtsH peptidases, were identified within or adjacent to bacteriocin gene clusters encoded within the Subsection V cyanobacterial genomes. These membrane-anchored ATP-dependent peptidases suggest other cleavage sites, in addition to the double glycine motif, are located on the precursor peptides. Interestingly, the double glycine motif for peptide cleavage was unable to be identified within the N11P-type precursor peptide sequences,

suggesting these precursor peptides may be nonfunctional. Out of the 51 putative precursors with similarity to known precursor types identified from the Subsection V cyanobacterial genomes, only two of these putative precursors were identified using antiSMASH; the remaining precursor peptides were identified using the BLAST program within the IMG JGI database. However, almost every bacteriocin gene cluster identified from the Subsection V cyanobacteria also encoded a large number of short

peptide sequences either within or located at the ends of the bacteriocin gene clusters, in addition to the known precursor types. These short peptide sequences may encode precursor peptides for bacteriocin biosynthesis. Future functional characterisation of these putative precursor peptides will determine if these sequences are part of the bacteriocin gene clusters identified from the Subsection V cyanobacteria.

Mycosporine-like amino acid (MAA) and scytonemin biosynthetic gene clusters

Cyanobacteria are known to produce two different types of UV-absorbing molecules, MAAs and scytonemin, in order to protect the cells against either UV-B or UV-A radiation, respectively [88–92]. The *mys* biosynthetic gene cluster has previously been identified from three Subsection V cyanobacterial genomes, specifically *Fischerella* sp. PCC 9339 [37], *Fischerella* sp. PCC 9431 [38] and *C. fritschii* PCC 6912 [42]. In this study, we identified an additional five genomes harbouring the *mys* gene cluster, including *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3, *M. testarum* BC008, *F. muscicola* SAG 1427–1 and *Chlorogloeopsis* sp. PCC 9212 (Fig. 5). All eight *mys* gene clusters identified contain the three genes required for the biosynthesis of mycosporine-glycine, the precursor compound for shinorine biosynthesis. All eight gene clusters also encode an NRPS-like enzyme proposed to activate serine (Additional file 7) and suggests all eight Subsection V cyanobacteria are capable of biosynthesising the MAA shinorine, identical to the pathway in *A. variabilis* ATCC 29413, as reported by Balskus and Walsh [88]. However, the domain organisation of the NRPS-like enzyme varies between the cyanobacterial strains. The majority of the identified gene clusters encode the domain organisation A-PCP-TE (Fig. 5a), identical to the characterised NRPS-like enzyme from *A. variabilis* ATCC 29413. However, three cyanobacterial strains (*C. fritschii* PCC 6912, *Chlorogloeopsis* sp. PCC 9212 and *M. testarum* BC008) all also encode a C domain at the N-terminus (Fig. 5b). An additional gene was identified downstream from the NRPS-like enzyme in three Subsection V cyanobacterial strains, specifically in *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3 and *Fischerella* sp. PCC 9431 (Fig. 5c). The encoded protein belongs to the prephenate/arogenate dehydrogenase family, which catalyse the conversion of prephenate to tyrosine in the shikimate pathway. While it is known that a wide variety of MAA analogues are biosynthesised, the effect of this fifth encoded gene (if any) remains unknown. However, transcription analysis revealed that this fifth gene is co-transcribed with the *W. intricata* UH strain HT-29-1 *mys* gene cluster in response to UV-B exposure (Micallef, ML. unpublished data).

Lastly, the scytonemin (*scy*) biosynthetic gene cluster was only identified within the *M. repens* PCC 10914 genome (Fig. 5d). The 32.9 kb gene cluster is similar to the *scy* gene cluster from *N. spumigena* CCY9414 [91], however the gene cluster from *M. repens* PCC 10914 encodes a transposase downstream of *scyF*. The presence of the *scy* gene cluster within the *M. repens* PCC 10914 genome suggests this organism is capable of biosynthesising scytonemin in order to protect the cells against UV-A radiation.

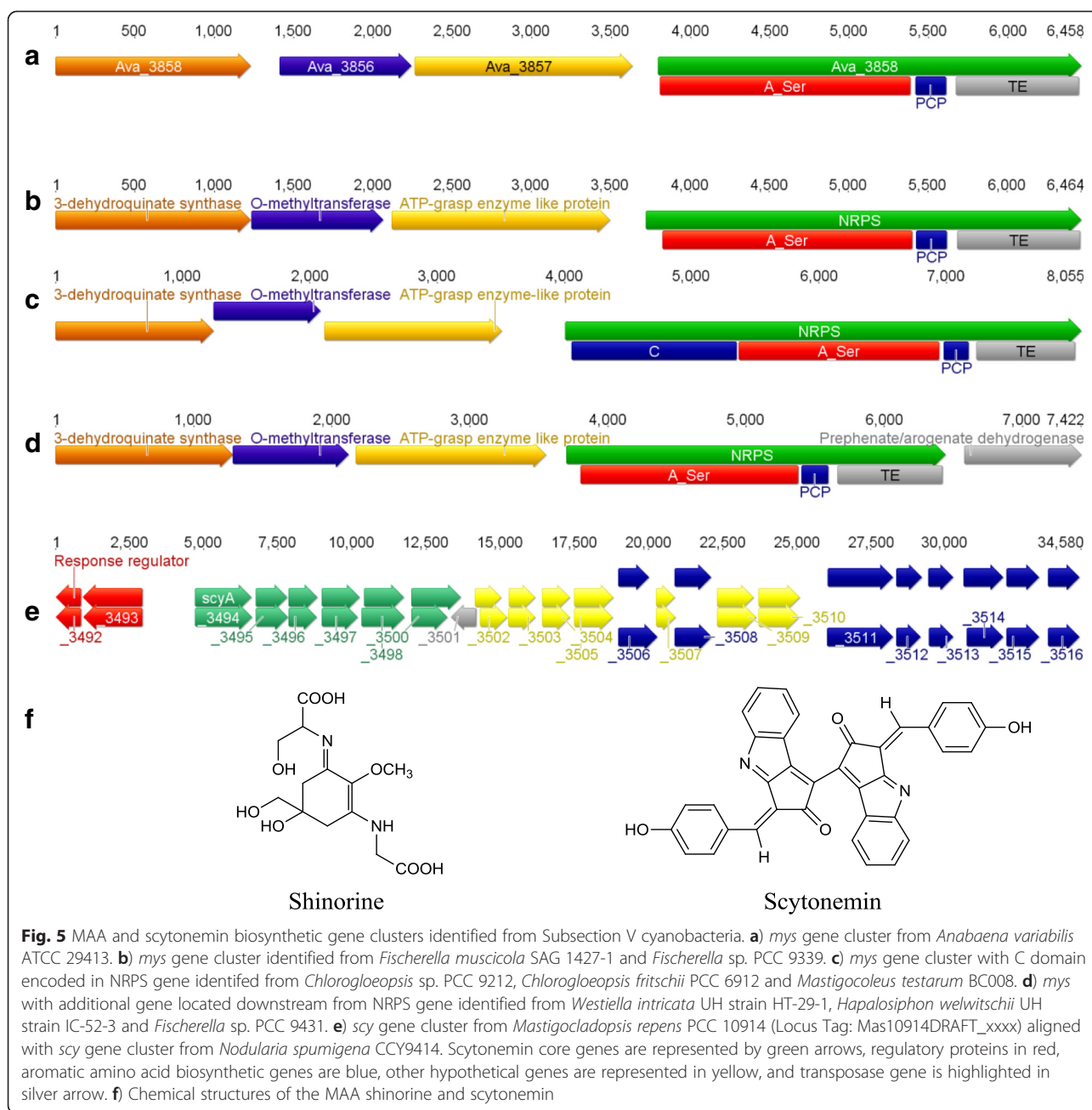
Identification of hydrocarbon biosynthetic gene clusters

Cyanobacteria have the unique ability to produce hydrocarbons from fatty acids, including heptadecane and methylheptadecane, which have potential diesel fuel applications [12]. The hydrocarbon biosynthetic gene cluster contains two genes, FAAR and ADO, for the biosynthesis of alkanes from fatty acids [69]. The FAAR/ADO pathway has previously been identified in all 13 Subsection V cyanobacterial genomes (Fig. 6a) [12, 43, 69]. All the Subsection V cyanobacterial strains analysed contain only the FAAR/ADO gene cluster; the OLS pathway was not identified from any Subsection V cyanobacterial genomes, which is consistent with the observation that all analysed cyanobacterial genomes encode only one type of hydrocarbon biosynthetic gene cluster [12].

Hydrocarbon composition has been characterised for four of the Subsection V cyanobacteria. Whilst *W. intricata* UH strain HT-29-1 is only able to biosynthesise heptadecane, *H. welwitschii* UH strain IC-52-3, *F. muscicola* PCC 7414 and *C. fritschii* PCC 6912 are capable of biosynthesising heptadecanes and methylheptadecanes [12]. Heptadecane is the most commonly observed hydrocarbon from cyanobacteria, whilst branched alkanes, such as methylheptadecanes, are also only observed from cyanobacteria encoding the FAAR/ADO pathway [12]. Since hydrocarbon biosynthetic gene clusters have been identified from a broad range of cyanobacteria, it has been suggested that an unknown selective pressure is forcing these organisms to maintain either pathway within their genomes [12].

Identification of terpene biosynthetic gene clusters

A number of terpene biosynthetic gene clusters were identified from the Subsection V cyanobacteria (Fig. 6b-h) (Additional file 8). The 5.4 kb gene cluster encoding the biosynthesis of geosmin (encoding geosmin synthase and two cyclic nucleotide-binding domain proteins) was identified in four of the 13 Subsection V cyanobacterial genomes, specifically *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3, *Fischerella* sp. PCC 9431 and *F. muscicola*



SAG 1427-1 (Fig. 6b). The gene encoding geosmin synthase displays approximately 90 % amino acid sequence similarity with geosmin synthase from *N. punctiforme* PCC 7120 [71]. Characterisation of the geosmin synthase from *N. punctiforme* PCC 73102 by Giglio et al. [71] revealed that the protein catalyses the biosynthesis of geosmin, as well as germacradienol, germacrene D, octalin and (*E*)-nerolidol. The gene cluster encoding 2-methylisoborneol, however, was not identified from any of the Subsection V cyanobacterial genome sequences.

A 8.2 kb sesquiterpene biosynthetic gene cluster, encoding three proteins (a sesquiterpene synthase, a cytochrome p450 and a putative hybrid two-component protein), was identified in four Subsection V cyanobacterial genomes (*Fischerella* sp. JSC-11, *F. thermalis* PCC 7521, *F. muscicola* PCC 7414 and *Fischerella* sp. PCC 9605) (Fig. 6c). This gene cluster demonstrates approximately 78 % amino acid sequence similarity with the 8a-*epi*- α -selinene biosynthetic gene cluster from *N. punctiforme* PCC 7120 [70]. However, in four additional cyanobacterial genomes, a similar

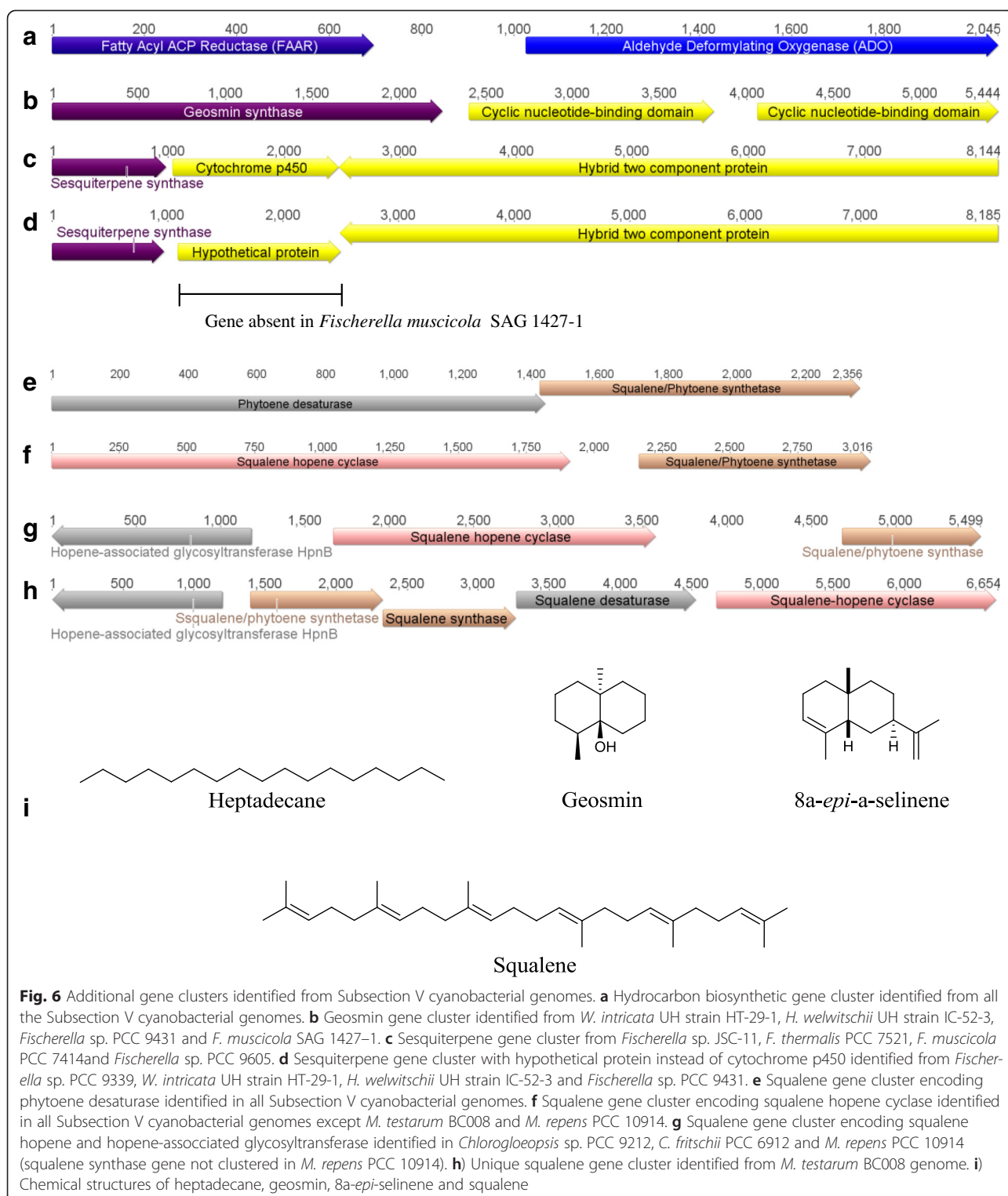


Fig. 6 Additional gene clusters identified from Subsection V cyanobacterial genomes. **a** Hydrocarbon biosynthetic gene cluster identified from all the Subsection V cyanobacterial genomes. **b** Geosmin gene cluster identified from *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3, *Fischerella* sp. PCC 9431 and *F. muscicola* SAG 1427-1. **c** Sesquiterpene gene cluster from *Fischerella* sp. JSC-11, *F. thermalis* PCC 7521, *F. muscicola* PCC 7414 and *Fischerella* sp. PCC 9605. **d** Sesquiterpene gene cluster with hypothetical protein instead of cytochrome p450 identified from *Fischerella* sp. PCC 9339, *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3 and *Fischerella* sp. PCC 9431. **e** Squalene gene cluster encoding phytoene desaturase identified in all Subsection V cyanobacterial genomes except *M. testarum* BC008 and *M. repens* PCC 10914. **f** Squalene gene cluster encoding squalene hopene cyclase identified in all Subsection V cyanobacterial genomes. **g** Squalene gene cluster encoding squalene hopene and hopene-associated glycosyltransferase identified in *Chlorogloeopsis* sp. PCC 9212, *C. fritschii* PCC 6912 and *M. repens* PCC 10914 (squalene synthase gene not clustered in *M. repens* PCC 10914). **h** Unique squalene gene cluster identified from *M. testarum* BC008 genome. **i** Chemical structures of heptadecane, geosmin, 8a-epi-selinene and squalene

sesquiterpene biosynthetic gene cluster was identified, in which a gene encoding a hypothetical protein related to 2-polyprenyl-6-methoxyphenol hydroxylase and FAD-dependent oxidoreductases (COG0654) was identified in place of the gene encoding a cytochrome

p450 (*Fischerella* sp. PCC 9339, *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3 and *Fischerella* sp. PCC 9431) (Fig. 6d). The presence of this alternative gene in the sesquiterpene gene cluster prevents the prediction of the encoded natural

product. In *F. muscicola* SAG 1427–1, only genes encoding sesquiterpene synthetase and a hybrid two-component protein were identified.

Two additional terpene gene clusters were identified from the Subsection V cyanobacterial genomes using antiSMASH. Genes encoding a phytoene/squalene synthetase and a phytoene desaturase (COG3349) were clustered together in all Subsection V cyanobacterial genomes (Fig. 6e). The other terpene gene cluster found in 11 Subsection V cyanobacterial genomes encodes a phytoene/squalene synthetase and a squalene-hopene cyclase (Fig. 6f). The encoded phytoene/squalene synthetases and squalene-hopene cyclase from the Subsection V cyanobacteria demonstrate approximately 80 % amino acid sequence similarity to the recently characterised squalene synthase and squalene hopene cyclase from *Synechocystis* sp. PCC 6803 [93]. In three genomes (*Chlorogloeopsis* sp. PCC 9212, *C. fritschii* PCC 6912 and *M. repens* PCC 10914), a hopene-associated glycosyltransferase *hpnB* (COG1215) gene was identified downstream from the squalene-hopene cyclase gene (Fig. 6g). The gene cluster from *M. repens* PCC 10914, however, does not contain the phytoene/squalene synthetase gene, although the gene is located within the genome.

A 6.6 kb terpene gene cluster from *M. testarum* BC008 is distinct from the other Subsection V cyanobacterial terpene gene clusters (Fig. 6h). In *M. testarum* BC008, the terpene gene cluster contains a hopene-associated glycosyltransferase *hpnB* (COG1215) gene, two phytoene/squalene synthetase genes, a phytoene desaturase gene, and a squalene-hopene cyclase gene. Furthermore *M. testarum* BC008 also encodes an additional squalene synthetase gene in the genome.

Identification of hapalindole and welwitindolinone biosynthetic gene clusters

A major class of secondary metabolites produced exclusively, to date, by the Subsection V cyanobacteria are the hapalindole family of natural products. A putative gene cluster encoding the biosynthesis of the hapalindoles and welwitindolinones has been published separately by our research group and Hillwig et al. [39–41]. Briefly, the presence of genes encoding methyltransferases (*welM1-3*), specific oxygenases (*welO11-19*) and regulation proteins (*welR3*) within the gene cluster suggests *W. intricata* UH strain HT-29-1, *H. welwitschii* UH strain IC-52-3, *Fischerella* sp. PCC 9431 and *F. muscicola* SAG 1427–1 encodes the *wel* gene cluster [41]. Characterisation of *WelM1*, responsible for the biosynthesis of *N*-methylwelwitindolinone C isothiocyanate, confirmed the *wel* gene cluster is responsible for welwitindolinone biosynthesis [39]. The gene cluster from *Fischerella* sp. PCC 9339 lacks the gene *ambP3* (encoding the

prenyltransferase responsible for ambiguine biosynthesis) and the genes characteristic of the *wel* gene cluster, suggesting the gene cluster from *Fischerella* sp. PCC 9339 encodes the biosynthesis of the hapalindoles [41].

Conclusion

In this study, all 11 publically available Subsection V cyanobacterial genomes, together with our draft genomes of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 were mined for their genetic potential to produce secondary metabolites. In this study, we were able to identify a putative gene cluster for hapalosin from the producing organism *H. welwitschii* UH strain IC-52-3. However, through genome mining, additional Subsection V cyanobacteria were identified to also encode the *hap* gene cluster. This study has also identified a wide range of biosynthetic gene clusters from the Subsection V cyanobacteria, including orphan NRPS/PKS gene clusters, PRPS gene cluster (including cyanobactin, microviridin and bacteriocin gene clusters), MAA and scytonemin gene clusters for UV-absorbing compounds and terpene gene clusters for geosmin, sesquiterpene and squalene biosynthesis. Through genome mining, the distribution and diversity of secondary metabolite biosynthesis in the Subsection V cyanobacteria has been revealed.

Availability of supporting data

The draft genome sequences of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 reported in this article were deposited as data sets in the US Department of Energy (DOE) Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) repository, under the Taxon ID 2529292565 [https://img.jgi.doe.gov/cgi-bin/er/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=2529292565] and 2529292566 [https://img.jgi.doe.gov/cgi-bin/er/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=2529292566], respectively. All other data was obtained from the US DOE JGI IMG database with the following Taxon ID *Fischerella* sp. PCC 9339 (2516653082), *Fischerella* sp. PCC 9431 (2512875027) *Fischerella* sp. JSC-11 (2505679024), *Fischerella* sp. PCC 9605 (2516143000), *F. muscicola* PCC 7414 (2548876996), *F. muscicola* SAG 1427–1 (2548876995), *F. thermalis* PCC 7521 (2548876998), *M. repens* PCC 10914 (2517093042), *C. fritschii* PCC 6912 (2551306142) and *Chlorogloeopsis* sp. PCC 9212 (2548877023). *M. testarum* BC008 genomic data was obtained from the National Centre for Biotechnology Information (NCBI) with GenBank Assembly Accession 000472885.1.

Additional files

Additional file 1: Specific and degenerative primers used to close gaps in orphan NRPS/PKS gene sequences and identify A-KR didomain. (DOCX 16 kb)

Additional file 2: *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3 102 housekeeping genes, COGs. (XLSX 118 kb)

Additional file 3: Table S1. Hapalysin biosynthetic gene cluster, Table S2: Adenylation domain binding pockets from *hap* gene cluster (DOCX 22 kb)

Additional file 4: Complete orphan NRPS/PKS gene clusters identified from the Subsection V cyanobacterial genomes. The genomes in which the orphan gene clusters were identified are stated to the left of each cluster. Fragments of gene clusters, or differences from the gene cluster shown are noted in the image. NRPS/PKS genes are represented by green arrows. Additional genes that may be involved in biosynthesis are represented by blue arrows. Genes encoding hypothetical proteins and transposases are represented by silver arrows. (PDF 307 kb)

Additional file 5: Hapalysin and orphan NRPS/PKS gene clusters. (DOCX 94 kb)

Additional file 6: Bacteriocin gene cluster and precursor analysis. (DOCX 103 kb)

Additional file 7: A domain binding pocket of NRPS-like enzyme identified in *mys* gene clusters from Subsection V cyanobacteria. (DOCX 18 kb)

Additional file 8: Terpene gene clusters. (DOCX 20 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MCM, MLM and RV designed the overall project. MLM and MCM sequenced the genomes of *W. intricata* UH strain HT-29-1 and *H. welwitschii* UH strain IC-52-3. MLM, PMD and DS contributed to identification and functional assignment of the gene clusters. MLM wrote the manuscript. MLM, PMD, DS, RV and MCM edited the final version of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Prof. Thomas Hemscheidt and Dr Benjamin Philmus assisted with providing University of Hawaii strains. RV and DS were funded by Case Western Reserve University. MCM and MLM were funded by the University of Western Sydney HDR Scholarship and RTS funding and the Australian Research Council, Discovery Project DP0880264.

Author details

¹School of Science and Health, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia. ²School of Biotechnology and Biomolecular Sciences, University of New South Wales, Kensington, NSW 2052, Australia. ³Department of Chemistry, Case Western Reserve University, 2740 Millis Science Center, Adelbert Road, Cleveland, OH 44106, USA.

Received: 19 February 2015 Accepted: 17 August 2015

Published online: 03 September 2015

References

- Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, et al. Genomes of Stigonematalean cyanobacteria (Subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol Evol.* 2013;5(1):31–44.
- Namikoshi M, Rinehart KL. Bioactive compounds produced by cyanobacteria. *J Ind Microbiol Biotechnol.* 1996;17(5–6):373–84.
- Burja AM, Banaigs B, Abou-Mansour E, Grant Burgess J, Wright PC. Marine cyanobacteria—a prolific source of natural products. *Tetrahedron.* 2001;57(46):9347–77.
- Welker M, Döhren H. Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS Microbiol Rev.* 2006;30(4):530–63.
- Bhat V, Dave A, MacKay JA, Rawal VH: Chapter Two - The chemistry of hapalindoles, fischerindoles, ambiguines, and welwitindolinones. In: *The Alkaloids: Chemistry and Biology*. Edited by Hans-Joachim K, vol. Volume 73: New York, Academic Press; 2014: 65–160.
- Moffitt MC, Burns BP: Hapalindole family of cyanobacterial natural products: structure, biosynthesis and function. In: *Handbook on Cyanobacteria* Edited by Gault PM, Marler HJ: New York, Nova Science Publishers, Inc; 2009
- Prinsep MR, Caplan FR, Moore RE, Patterson GML, Honkanen RE, Boynton AL. Microcystin-LA from a blue-green alga belonging to the Stigonematales. *Phytochemistry.* 1992;31(4):1247–8.
- Fiore MF, Genuário DB, da Silva CSP, Shishido TK, Moraes LAB, Neto RC, et al. Microcystin production by a freshwater spring cyanobacterium of the genus *Fischerella*. *Toxicon.* 2009;53(7–8):754–61.
- Cirés S, Alvarez-Roa C, Wood SA, Puddick J, Loza V, Heimann K. First report of microcystin-producing *Fischerella* sp. (Stigonematales, Cyanobacteria) in tropical Australia. *Toxicon.* 2014;88:62–6.
- Stratmann K, Burgoyne DL, Moore RE, Patterson GML, Smith CD. Hapalysin, a cyanobacterial cyclic depsipeptide with multidrug-resistance reversing activity. *J Org Chem.* 1994;59(24):7219–26.
- Kim H, Kronic A, Lantvit D, Shen Q, Kroll DJ, Swanson SM, Orjala J: Nitrile-containing fischerindoles from the cultured cyanobacterium *Fischerella* sp. *Tetrahedron* 2012;68(15):3205–3209.
- Coates RC, Podell S, Korobeynikov A, Lapidus A, Pevzner P, Sherman DH, et al. Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. *PLoS ONE.* 2014;9(1), e85140.
- Hagmann L, Jüttner F. Fischerellin A, a novel photosystem-II-inhibiting allelochemical of the cyanobacterium *Fischerella muscicola* with antifungal and herbicidal activity. *Tetrahedron Lett.* 1996;37(36):6539–42.
- Gross EM, Wolk CP, Jüttner F. Fisherellin, a new allelochemical from the freshwater cyanobacterium *Fischerella muscicola*. *J Phycol.* 1991;27(6):686–92.
- Papke U, Gross EM, Francke W. Isolation, identification and determination of the absolute configuration of fischerellin B. A new algicide from the freshwater cyanobacterium *Fischerella muscicola* (Thuret). *Tetrahedron Lett.* 1997;38(3):379–82.
- Prinsep MR, Moore RE, Levine IA, Patterson GML. Westiellamide, a bistratamide-related cyclic peptide from the blue-green alga *Westiellopsis prolifica*. *J Nat Prod.* 1992;55(1):140–2.
- Falch BS, Koenig GM, Wright AD, Sticher O, Ruegger H, Bernardinelli G. Ambigol A and B: new biologically active polychlorinated aromatic compounds from the terrestrial blue-green alga *Fischerella ambigua*. *J Org Chem.* 1993;58(24):6570–5.
- Wright AD, Papendorf O, König GM. Ambigol C and 2,4-dichlorobenzoic acid, natural products produced by the terrestrial cyanobacterium *Fischerella ambigua*. *J Nat Prod.* 2005;68(3):459–61.
- Kang H-S, Kronic A, Orjala J. Stigonemapeptin, an Ahp-containing depsipeptide with elastase inhibitory activity from the bloom-forming freshwater cyanobacterium *Stigonema* sp. *J Nat Prod.* 2012;75(4):807–11.
- Koodkaew I, Matsuyama S, Sunohara Y, Matsumoto H. Hapalocyclamide: a novel phytotoxic hexapeptide of the cyanobacterium *Hapalosiphon* sp. *Tetrahedron Lett.* 2012;53(8):977–9.
- Ghasemi Y, Yazdi MT, Shafiee A, Amini M, Shokravi S, Zarrini G. Parsiguine, a novel antimicrobial substance from *Fischerella ambigua*. *Pharm Biol.* 2004;42(4–5):318–22.
- Asthana S, Srivastava A, Kayastha A, Nath G, Singh S. Antibacterial potential of γ -linolenic acid from *Fischerella* sp. colonizing Neem tree bark. *World J Microbiol Biotechnol.* 2006;22(5):443–8.
- Ehrenreich IM, Waterbury JB, Webb EA. Distribution and diversity of natural product genes in marine and freshwater cyanobacterial cultures and genomes. *Appl Environ Microbiol.* 2005;71(11):7401–13.
- Kehr J-C, Gatte Picchi D, Dittmann E. Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes. *Beilstein J Org Chem.* 2011;7:1622–35.
- Dittmann E, Neilan B, Börner T. Molecular biology of peptide and polyketide biosynthesis in cyanobacteria. *Appl Microbiol Biotechnol.* 2001;57(4):467–73.
- Christiansen G, Dittmann E, Ordorika LV, Rippka R, Herdman M, Börner T. Nonribosomal peptide synthetase genes occur in most cyanobacterial genera as evidenced by their distribution in axenic strains of the PCC. *Arch Microbiol.* 2001;176(6):452–8.
- Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev.* 2006;106(8):3468–96.

28. Liou GF, Khosla C. Building-block selectivity of polyketide synthases. *Curr Opin Chem Biol.* 2003;7(2):279–84.
29. Donadio S, Monciardini P, Sosio M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep.* 2007;24(5):1073–109.
30. Cane DE, Walsh CT, Khosla C. Harnessing the biosynthetic code: Combinations, permutations, and mutations. *Science.* 1998;282(5386):63–8.
31. Walsh CT, Fischbach MA. Natural products version 2.0: Connecting genes to molecules. *J Am Chem Soc.* 2010;132(8):2469–93.
32. 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(1):108–60.
33. Sivonen K, Leikoski N, Fewer D, Jokela J. Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol.* 2010;86(5):1213–25.
34. Wang H, Fewer DP, Sivonen K. Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. *PLoS ONE.* 2011;6(7), e22384.
35. Hemscheidt TK: Chapter two - microviridin biosynthesis. In: *Methods Enzymol.* Edited by David AH, vol. Volume 516: New York, Academic Press; 2012: 25–35.
36. Micallef ML, D'Agostino PM, Al-Sinawi B, Neilan BA, Moffitt MC. Exploring cyanobacterial genomes for natural product biosynthesis pathways. *Mar Genomics.* 2014.
37. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA.* 2013;110(3):1053–8.
38. Calteau A, Fewer D, 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(1):977.
39. Hillwig ML, Fuhrman HA, Ittiarnornkul K, Sevco TJ, Kwak DH, Liu X. Identification and characterization of a welwitindolinone alkaloid biosynthetic gene cluster in the stigonematalean cyanobacterium *Hapalosiphon welwitschii*. *ChemBioChem.* 2014;15(5):665–9.
40. Hillwig ML, Zhu Q, Liu X. Biosynthesis of ambigua indole alkaloids in cyanobacterium *Fischerella ambigua*. *ACS Chem Biol.* 2013;9(2):372–7.
41. Micallef M, Sharma D, Bunn B, Gerwick L, Viswanathan R, Moffitt M. Comparative analysis of hapalindole, ambigua and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. *BMC Microbiol.* 2014;14(1):1–18.
42. Gao Q, Garcia-Pichel F. An ATP-Grasp ligase involved in the last biosynthetic step of the iminomycesporine shinorine in *Nostoc punctiforme* ATCC 29133. *J Bacteriol.* 2011;193(21):5923–8.
43. Klähn S, Baumgartner D, Pfreundt U, Voigt K, Schoen V, Steglich C, et al. Alkane biosynthesis genes in cyanobacteria and their transcriptional organization. *Front Bioeng Biotechnol.* 2014;2.
44. Morin N, Vallaeys T, Hendrickx L, Natalie L, Wilmotte A. An efficient DNA isolation protocol for filamentous cyanobacteria of the genus *Arthrospira*. *J Microbiol Methods.* 2010;80(2):148–54.
45. Wilson K. Preparation of genomic DNA from bacteria. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, editors. *Current Protocols in Molecular Biology.* New York: John Wiley & Sons, Inc; 2001.
46. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7(1):539.
47. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, et al. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 2013;41(W1):W204–12.
48. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 2011;39(Database issue):D225–9.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
50. Bachmann BO, Ravel J. *In silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. In: *Methods Enzymol.* Edited by David AH, vol. 458: New York, Academic Press; 2009: 181–217.
51. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* 2011;39 suppl 2:W362–7.
52. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE.* 2012;7(3), e34064.
53. Puigbo P, Wolf Y, Koonin E. Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J Biol.* 2009;8(6):59.
54. D'Agostino PM. Analysis of cyanobacterial genes for the biosynthesis of the anti-cancer compound hapalasin. University of Western Sydney. 2008.
55. Ramaswamy AV, Sorrels CM, Gerwick WH. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod.* 2007;70(12):1977–86.
56. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, et al. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol.* 2004;11(6):817–33.
57. Fewer DP, Österholm J, Rouhiainen L, Jokela J, Wahlsten M, Sivonen K. Nostophycin biosynthesis is directed by a hybrid polyketide synthase-nonribosomal peptide synthetase in the toxic cyanobacterium *Nostoc* sp. strain 152. *Appl Environ Microbiol.* 2011;77(22):8034–40.
58. Magarvey NA, Beck ZQ, Golakoti T, Ding Y, Huber U, Hemscheidt TK, et al. Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from *Nostoc* cyanobionts. *ACS Chem Biol.* 2006;1(12):766–79.
59. Magarvey NA, Ehling-Schulz M, Walsh CT. 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.* 2006;128(33):10698–9.
60. Hackett JD, Wisecaver JH, Brosnahan ML, Kulis DM, Anderson DM, Bhattacharya D, et al. Evolution of saxitoxin synthesis in cyanobacteria and dinoflagellates. *Mol Biol Evol.* 2013;30(1):70–8.
61. Knauer SH, Hartl-Spiegelhauer O, Schwarzinger S, Hänzelmann P, Dobbek H. The Fe(II)/ α -ketoglutarate-dependent taurine dioxygenases from *Pseudomonas putida* and *Escherichia coli* are tetramers. *FEBS J.* 2012;279(5):816–31.
62. Ikeda H, Nonomiya T, Usami M, Ohta T, Omura S. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrocyclic avermectin in *Streptomyces avermitilis*. *Proc Natl Acad Sci.* 1999;96(17):9509–14.
63. La Ferla B, Airoidi C, Zona C, Orsato A, Cardona F, Merlo S, et al. Natural glycoconjugates with antitumor activity. *Nat Prod Rep.* 2011;28(3):630–48.
64. Kersten RD, Ziemert N, Gonzalez DJ, Duggan BM, Nizet V, Dorrestein PC, et al. Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc Natl Acad Sci USA.* 2013;110(47):E4407–16.
65. Kim EJ, Lee JH, Choi H, Pereira AR, Ban YH, Yoo YJ, et al. Heterologous production of 4-O-demethylbarbamide, a marine cyanobacterial natural product. *Org Lett.* 2012;14(23):5824–7.
66. Ongley SE, Bian X, Zhang Y, Chau R, Gerwick WH, Müller R, et al. High-titer heterologous production in *E. coli* of lyngbyatoxin, a protein kinase C activator from an uncultured marine cyanobacterium. *ACS Chem Biol.* 2013;8(9):1888–93.
67. Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, et al. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc Natl Acad Sci USA.* 2005;102(20):7315–20.
68. Donia MS, Ravel J, Schmidt EW. A global assembly line for cyanobactins. *Nat Chem Biol.* 2008;4(6):341–3.
69. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB. Microbial biosynthesis of alkanes. *Science.* 2010;329(5991):559–62.
70. Agger SA, Lopez-Gallego F, Hoye TR, Schmidt-Dannert C. Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. *J Bacteriol.* 2008;190(18):6084–96.
71. Giglio S, Jiang J, Saint CP, Cane DE, Monis PT. Isolation and characterization of the gene associated with geosmin production in cyanobacteria. *Environ Sci Technol.* 2008;42(21):8027–32.
72. Leikoski N, Liu L, Jokela J, Wahlsten M, Gugger M, Calteau A, et al. Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. *Chem Biol.* 2013;20(8):1033–43.
73. Donia M, Schmidt E: Cyanobactins—ubiquitous cyanobacterial ribosomal peptide metabolites, vol. 2: Elsevier: Oxford, UK; 2010.
74. Koehnke J, Bent A, Housen WE, Zollman D, Morawitz F, Shirran S, et al. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat Struct Mol Biol.* 2012;19(8):767–72.

75. Banker R, Carmeli S. Tenuocyclamides A – D, cyclic hexapeptides from the cyanobacterium *Nostoc spongiaeforme* var. *tenue*. *J Nat Prod*. 1998;61(10):1248–51.
76. Schmidt EW, Donia MS: Chapter 23 Cyanobactin ribosomally synthesized peptides—A case of deep metagenome mining. In: *Methods Enzymol*. Edited by David AH, vol. 458: New York, Academic Press; 2009: 575–596.
77. Leikoski N, Fewer DP, Jokela J, Alakoski P, Wahlsten M, Sivonen K. Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus *Microcystis*. *PLoS ONE*. 2012;7(8), e43002.
78. Ishitsuka MO, Kusumi T, Kakisawa H, Kaya K, Watanabe MM. Microviridin. A novel tricyclic depsipeptide from the toxic cyanobacterium *Microcystis viridis*. *J Am Chem Soc*. 1990;112(22):8180–2.
79. Shin HJ, Murakami M, Matsuda H, Yamaguchi K. Microviridins D-F, serine protease inhibitors from the cyanobacterium *Oscillatoria agardhii* (NIES-204). *Tetrahedron*. 1996;52(24):8159–68.
80. Philmus B, Christiansen G, Yoshida WY, Hemscheidt TK. Post-translational modification in microviridin biosynthesis. *ChemBioChem*. 2008;9(18):3066–73.
81. Philmus B, Guerrette JP, Hemscheidt TK. Substrate specificity and scope of MvdD, a GRASP-like ligase from the microviridin biosynthetic gene cluster. *ACS Chem Biol*. 2009;4(6):429–34.
82. Ziemert N, Ishida K, Liaimer A, Hertweck C, Dittmann E. Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. *Angew Chem Int Edit*. 2008;47(40):7756–9.
83. Oman TJ, van der Donk WA. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol*. 2010;6(1):9–18.
84. Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol*. 1995;16(2):229–40.
85. Willey JM, van der Donk WA. Lantibiotics: Peptides of diverse structure and function. *Annu Rev Microbiol*. 2007;61(1):477–501.
86. Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA. Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol*. 2010;8(3), e1000339.
87. Haft D, Basu M, Mitchell D. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biol*. 2010;8(1):70.
88. Balskus EP, Walsh CT. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science*. 2010;329(5999):1653–6.
89. Rastogi RP, Incharoensakdi A. UV radiation-induced biosynthesis, stability and antioxidant activity of mycosporine-like amino acids (MAAs) in a unicellular cyanobacterium *Gloeocapsa* sp. CU2556. *J Photochem Photobiol B: Biol*. 2014;130:287–92.
90. Sinha RP, Ambasht NK, Sinha JP, Klisch M, Häder D-P. UV-B-induced synthesis of mycosporine-like amino acids in three strains of *Nodularia* (cyanobacteria). *J Photochem Photobiol B: Biol*. 2003;71(1–3):51–8.
91. Sorrels CM, Proteau PJ, Gerwick WH. Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl Environ Microbiol*. 2009;75(14):4861–9.
92. Soule T, Garcia-Pichel F, Stout V. Gene expression patterns associated with the biosynthesis of the sunscreen scytonemin in *Nostoc punctiforme* ATCC 29133 in response to UVA radiation. *J Bacteriol*. 2009;191(14):4639–46.
93. Englund E, Pattanaik B, Ubhayasekera SJK, Stensjö K, Bergquist J, Lindberg P. Production of squalene in *Synechocystis* sp. PCC 6803. *PLoS ONE*. 2014;9(3):e90270.
94. Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, Ravel J, et al. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol*. 2006;2(12):729–35.
95. McIntosh JA, Lin Z, Tianero MDB, Schmidt EW. Aestuarinamides, a natural library of cyanobactin cyclic peptides resulting from isoprene-derived claisen rearrangements. *ACS Chem Biol*. 2013;8(5):877–83.
96. Ziemert N, Ishida K, Quillardet P, Bouchier C, Hertweck C, de Marsac NT, et al. Microcyclamide biosynthesis in two strains of *Microcystis aeruginosa*: From structure to genes and vice versa. *Appl Environ Microbiol*. 2008;74(6):1791–7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

