Marine Genomics 21 (2015) 1–12

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,*]

[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 5 December 2014

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

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, the current 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 bioinformatic 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 bioinformatic 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 has 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).

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

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 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.

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.

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