PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

(GENOMIC CHARACTERIZATION OF *CYLINDROSPERMOPSIS RACIBORSKII* A CYANOBACTERIUM ABLE TO PRODUCE BIOACTIVE COMPOUNDS)

(JUAN JOSÉ FUENTES VALDÉS)

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences.

Advisor:

DANIEL GARRIDO AND MÓNICA VÁSQUEZ

Santiago de Chile (Jan. 2019)

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

(GENOMIC CHARACTERIZATION OF *CYLINDROSPERMOPSIS RACIBORSKII* A CYANOBACTERIUM ABLE TO PRODUCE BIOACTIVE COMPOUNDS)

(JUAN JOSÉ FUENTES VALDÉS)

Members of the Committee:

(DR. DANIEL GARRIDO)

(DR. MONICA VASQUEZ)

(DR. LORETO VALENZUELA)

(DR. JUAN CARLOS FERNANDEZ)

(JUAN DE DIOS ORTUZAR)

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences.

Santiago de Chile (Jan. 2019)

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

CARACTERIZACIÓN GENÓMICA DE LA CIANOBACTERIA CYLINDROSPERMOPSIS RACIBORSKII PRODUCTORA DE COMPUESTOS BIOACTIVOS

Tesis enviada a la Dirección de Investigación y Postgrado en cumplimiento parcial de los requisitos para el grado de Doctor en Ciencias de la Ingeniería.

JUAN JOSÉ FUENTES VALDÉS

RESUMEN

Las cianobacterias son microorganismos esenciales para la vida, ya que, desarrollan el proceso de fotosíntesis y fijación de nitrógeno atmosférico. Los compuestos producidos por estos microorganismos tienen un potencial biotecnológico destacable, sin embargo, también son responsables de generar intoxicaciones graves producto del consumo de agua contaminadas, como es el caso de la intoxicación por la toxina saxitoxina. Avances en genómica y bioinformática han facilitado el descubrimiento de nuevos compuestos además

de ayudar a comprender su evolución cianobacteriana y diseminación en el medio ambiente. En esta tesis, mediante un enfoque molecular se buscó e identificó un grupo de compuestos bioactivos, además de proporcionar una caracterización molecular y química de las moléculas identificadas. El primer objetivo fue obtener las secuencias genómicas de aislamientos de *Cylidrospermopsis raciborskii* australianos y desarrollar un análisis de genómica comparativa, para comprender su relación genética con la especie brasileña *Raphidiopsis brookii*, con un enfoque especial en la producción de saxitoxina. En general, observamos una fuerte correlación biogeográfica entre las cepas analizadas. La comparación genómica del grupo *C. raciborskii* y *R. brookii* mostró un alto grado de similitud en términos globales, sin embargo, se observaron diferencias significativas en términos de la biosíntesis de metabolitos secundarios. En segundo lugar, los clústeres de genes para metabolitos secundarios identificados en los genomas de *Cylindrospermopsis*, se estudiaron a nivel molecular. Se determinó el potencial de *C. raciborskii* para producir compuestos activos del tipo hassallidinas y saxitoxinas. Usando LC-MS / MS determinamos el perfil de hsasallidina en *C. raciborskii* CS-505, e identificamos una nueva variante de hassallidina en la cepa CS-509. Curiosamente, no logramos detectar hassallidina en extractos de *C. raciborskii* CS-508, a pesar de la presencia del cluster de genes aparentemente completo. Esto podría explicarse por mecanismos reguladores o mutaciones puntuales. Finalmente, determinamos el perfil de saxitoxina en la cepa *C. raciborskii* MVCC14, que resultó ser el mismo perfil identificado en *R. brookii* D9. *Cylindrospermopsis raciborskii* y *Raphidiopsis brookii* se clasifican morfológicamente como especies diferentes, mediante análisis de genómica comparativa los datos sugieren

que ambos podrían ser miembros de la misma especie. En resumen, este trabajo proporcionó información importante referente a genomas cianobacterianos, determinando las estructuras genéticas responsables de la producción de metabolitos secundarios de compuestos de importancia biotecnológica como hassallidina y saxitoxina.

Miembros de la Comisión de Tesis Doctoral
(Beatriz Díez)
(Loreto Valenzuela)
(Juan Carlos Fernández)
(Juan de Dios Ortuzar)
Santiago, ene, 2019

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

GENOMIC CHARACTERIZATION OF *CYLINDROSPERMOPSIS RACIBORSKII* A CYANOBACTERIUM ABLE TO PRODUCE BIOACTIVE COMPOUNDS

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences by

JUAN JOSÉ FUENTES VALDÉS

ABSTRACT

Cyanobacteria are essential environmental microorganisms able to develop photosynthesis and in some cases fix atmospheric nitrogen. Furthermore, toxins produced by these microbes have an interesting biotechnological projection but are also responsible for severe water-borne diseases, such as saxitoxin intoxication. Genomic advances and bioinformatics could facilitate the discovery of novel toxins and help to understand their evolution and spread in the environment. In this thesis, these molecular approaches were applied to search and identify capacity bioactive compounds in cyanobacteria, and to provide their molecular and chemical characterization. The first aim was to obtain the genomic sequences of Australian *Cylidrospermopsis raciborskii* isolates and develop a comparative genomics pipeline, to understand their genetic relatedness to Brazilian species *Raphidiopsis brookii*, with a particular focus on saxitoxin production. In general, we observed a robust biogeographical correlation between the strains analyzed. Genome comparison of the *C. raciborskii* group and *R. brookii* revealed high global similarities, but significant differences regarding secondary metabolites biosynthesis. Secondly, biosynthetic gene clusters for secondary metabolites were identified in *Cylindrospermopsis* genomes, and

they were studied at a molecular level. The capacity of *C. raciborskii* to produce active hassallidins and saxitoxins compounds were also determined. Using LC-MS/MS we determined the hassallidin profile of *C. raciborskii* CS-505, and we identified a new variant of hassallidin in strain CS-509. Interestingly, we were not able to detect hassallidin in *C. raciborskii* CS-508 extracts, despite the presence of an apparently complete gene cluster containing the codification responsible for the machinery for its production. The absence of hassallidin in this strain remains unclear; however, could be the consequence of either regulatory mechanisms or point mutations, and is an open question to further studies. Finally, we determined the saxitoxin profile in *C. raciborskii* strain MVCC14, which resulted in being the same as the one identified in *R. brookii* D9. Notwithstanding the morphological classification of *Cylindrospermopsis raciborskii* and *Raphidiopsis brookii* categorizes these microorganisms as different species, the genomic analysis suggests that both could be members of the same species. In summary, this work provides critical information for cyanobacterial genomes, determining the genetic structures responsible for secondary metabolite production of compounds of biotechnological importance (hassallidin) or environmental (saxitoxin).

Members of the Doctoral Thesis Committee:
(Beatriz Díez)
(Loreto Valenzuela)
(Juan Carlos Fernández)
(Juan de Dios Ortuzar)
Santiago, Jan. 2019

II

List of Papers

This thesis is based on the following papers, referred to in the text by their Roman numerals.

I Juan J. Fuentes-Valdés, Alvaro M. Plominsky, Eric E. Allen, Javier Tamames and Mónica Vásquez.
Complete genome of a Cylindrospermopsin producing cyanobacterium *Cylindrospermopsis raciborskii* CS505, containing a circular chromosome and a single extrachromosomal element. *Genome Announcements.* doi: 10.1128/genomeA.00823-16 (published).

II Juan J Fuentes-Valdés, Katia Soto-Liebe, Danilo Pérez-Pantoja, Javier Tamames, Lucy Belmar, Carlos Pedros-Alió, Daniel Garrido and Mónica Vásquez.
Draft genome sequences of *Cylindrospermopsis* strains CS-508 and MVCC14, isolated from freshwater bloom events in Australia and Uruguay. *Standards in Genomic Sciences* (in press)

III Juan J. Fuentes-Valdés, Katia Soto, Bernd Krock, Daniel Garrido and Mónica Vásquez.
LC/MS Characterization of the antifungal Hassallidin, identified in Cylindrospermopsis raciborskii. *Toxicon* (In preparation)

IV Juan J. Fuentes-Valdés, Katia Soto-Liebe, Bernd Krock, Nicole Trefault, Daniella Spooner-Lagos, Mónica Vásquez.
Genetic context of sxt genes in producing cyanobacteria reveals evolutionary adaptations for their maintenance and spread. *Molecular Genetics and metabolism* (In preparation)

Contents

1. General Introduction

1.1. Cyanobacterial organisms.

Cyanobacteria represent a morphologically and metabolically diverse group of bacteria colonizing a wide range of aquatic and terrestrial environments. They could appear in unicellular, filamentous or multicellular forms. These organisms are remarkably unique among prokaryotes, due to their ability to perform oxygenic photosynthesis. Interestingly, this capacity during evolution was transferred to photosynthetic eukaryotes, and therefore cyanobacteria are recognized to be the chloroplast ancestors (Reyes-Prieto et al., 2007; Moreira et al., 2000). Cyanobacteria are presumably the first oxygen-evolving photosynthetic organisms that appeared during the Precambrian era, gradually modifying the atmosphere that created an ideal environment for the evolution of Eukarya (Kopp et al., 2005; Schirmeister et al., 2015). Regarding photosynthesis transforms efficiently solar energy into organic carbon, cyanobacteria are currently studied for biomass and biofuels production (Ruffing 2011).

Moreover, certain cyanobacteria are also able to fix nitrogen, a central metabolism in the Nitrogen cycle, which is inhibited by oxygen. As cyanobacteria produce oxygen by photosynthesis, diazotrophic cyanobacteria carry out nitrogen fixation either by cell differentiation, forming microaerophilic specialized cells called heterocysts (multicellular cyanobacteria), or temporally coordinating this process with photosynthesis (unicellular and non-heterocystous filamentous cyanobacteria). The adaptation process to extreme environments that started 2.45 billion years ago, possibly triggered the development of

molecular strategies related to the synthesis of compounds that have probably allowed cyanobacteria to colonize a wide range of niches (Schirmmeister et al., 2015).

1.1.1. Cyanobacteria taxonomy.

Cyanobacteria have usually been classified according to their morphology and division patterns into five subsections (Rippka et al., 1979): subsections I and II comprise unicellular cyanobacteria, and subsections III to V include multicellular cyanobacteria. Subsection III is composed of filamentous non-heterocystous representatives, while subsection IV and V comprise unramified and ramified heterocyst-forming cyanobacteria respectively.

Recently, advances in genome sequencing, analysis, and phylogenetic reconstruction have allowed evaluating the phylogeny of this group by analyzing complete genomes and considering orthologous datasets (Figure 1) (Schirmmeister et al., 2015).

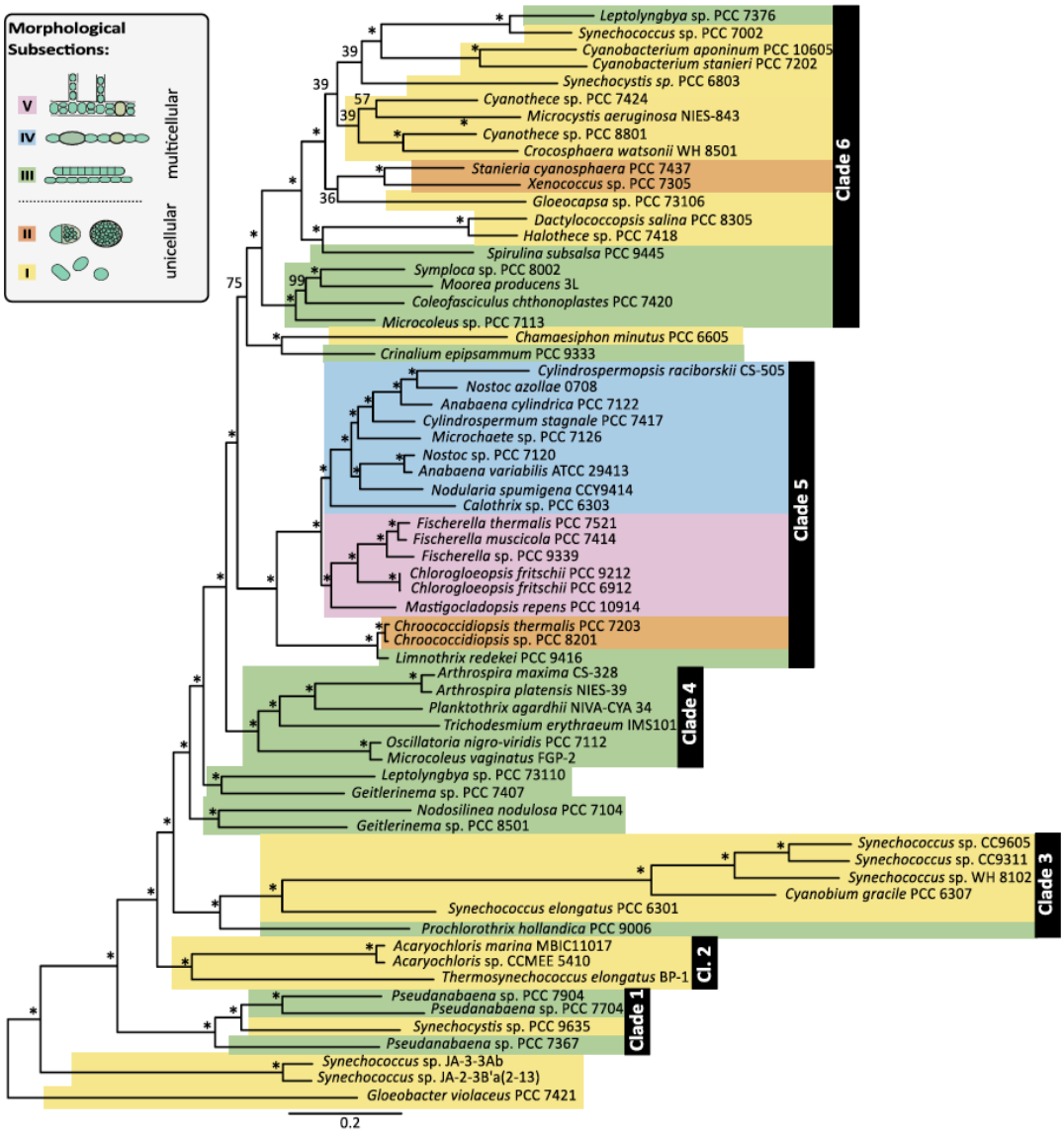

Figure 1. Phylogenomic maximum likelihood tree of 65 cyanobacterial taxa based on a supermatrix comprised of 756 concatenated protein sequences (197.761 amino acid sites). Maximum likelihood bootstrap support for clades is indicated at respective branches. Asterisks indicate 100% similarity calculated from 1000 bootstrap resampling. Cyanobacterial taxa are color-coded. Unicellular taxa belonging to morphological subsections I and II are displayed in yellow and orange, respectively, whereas multicellular

cyanobacterial taxa belonging to subsections III, IV and V are shown in green, blue and pink, respectively (Schirmmaeister et al., 2015).

1.2. Bioactive compounds from cyanobacteria.

The ecological relevance of cyanobacteria as primary producers as one of the principal organisms fixing-CO_2 is well established (Kumar et al., 2010). However, under certain circumstances, cyanobacteria could also be detrimental to human health and the environment. Some cyanobacteria are major component of Harmful Algal Blooms (HAB), an excessive proliferation of microalgae induced for certain environmental eutrophication (Moss 1988). Moreover, the adverse effects of cyanobacterial HABs are exacerbated because some of them could be toxic to humans or surrounding ecosystems.

Cyanobacteria can produce diverse bioactive compounds. These are classified as secondary metabolites and can deploy a wide range of structural variants. As secondary metabolites, they are not essential for the central metabolism in cyanobacteria. However, some bioactive compounds are constitutively produced (Dangy et al., 2017).

An increasing number of cyanobacterial secondary metabolites are classified as cyanotoxins, a potent type of toxins. Cyanotoxins are generally released to bodies of water when cell blooms collapse. According to the effects they exert, cyano-compounds are classified into dermatotoxins, cytotoxins, hepatotoxins, and/or neurotoxins (Sivonen and Jones, 1999). Many of these compounds are highly toxic to animals including humans. Intoxication with cyano-compounds can in extreme cases trigger death due to

multisystemic failure (Hurley et al., 2014). Despite the effort to elucidate the functionality and relevance of their biosynthesis, their biological role remains mostly unknown.

There are several secondary metabolites produced by cyanobacteria that are used by the pharmacological industry. For example, borophycin and apratoxin are used in complementary anticancer treatments (Vijayakumar et al., 2015). In general, cyanobacteria produce a wide range of bioactive secondary metabolites, suggesting that these microorganisms are a source of new chemicals that can potentially fulfill a biotechnological role in the development of new drugs. Currently, there are some examples of cyanobacterial compounds that are used as antivirals molecules like sulfoglycolipids and lectins (Nyok-Sean et al., 2015; Ram et al., 2017). Besides, there is a diverse spectrum of molecules produced by cyanobacteria used as analgesics or that provided therapeutics solutions for pain control (Vijayakumar et al., 2015; Chorny et al., 2009).

1.2.1. Hassallidins.

Hassallidins (HASs) are glycosylated lipopeptides with antifungal activity. They were first isolated from the epilithic cyanobacterium *Hassallia* sp. (Neuhof et al., 2005; Neuhof et al., 2006). Additional producing HASs species are *Anabaena* sp., *Cylindrospermopsis raciborskii*, *Aphanizomenon gracile*, *Nostoc* sp., *Planktothrix serta* and *Tolypothrix* sp. (Vestola et al., 2014; Pancrace et al., 2017). Structural variants of HASs could differ in the presence of sugar moieties (pentose, deoxyhexose, hexose, acetylated hexose or N-acetylhexosamine amino acid ring composition (Shishido et al., 2015), and the acyl side chain depending on the carbon number (Vestola et al., 2014).

The antifungal effect found of HASs has been shown active against *Candida* sp. and *Cryptococcus neoformans*. The toxicity mechanisms of HASs, however, are not precise (Neuhof et al., 2005; Neuhof et al., 2006).

1.2.2. Saxitoxins.

Saxitoxins (SXTs) are considered one of the most potent cyanotoxins, causing death in mammals by a rapid respiratory paralysis (Bricelj et al., 1990). STXs are a family of tricyclic tetrahydropurine alkaloids with a neurotoxic activity that acts by blocking voltage-gated sodium channels (Narahashi et al., 1967; Llewellyn L.E. 2006).

Saxitoxin (STX) is considered the basic structure of this group, but certain modifications produce more than 30 structural analogs (Llewellyn, 2006). STX is the most toxic analog among this group due to its highest affinity for the voltage-gated sodium channel. Its toxicity measured as LD50 (lethal dose, 50%) in mice is 10 μg kg-1, being 1000-fold more potent than cyanide (Batoréu et al., 2005).

1.3. Biosynthetic pathways of secondary metabolites with bioactive potential.

Three types of enzymes complexes could carry out the synthesis of bioactive compounds in cyanobacteria: (i) non-ribosomal peptide synthetases (NRPS), (ii) polyketide synthases (PKS), and (iii) a hybrid system of both enzymes complexes (Tan 2007; Uzair et al., 2012). To date, gene clusters participating in the synthesis of cyanotoxins like saxitoxin and

cylindrospermopsin have been described (Tillett et al., 2000; Moffit and Neilan, 2004; Kellman et al., 2008; Mihali et al., 2008; Mejean et al., 2010).

Based on several sequenced cyanobacterial genomes, it has been possible to identify and describe the biosynthetic pathways of several secondary metabolites of commercial relevance (Calteau et al., 2014).

1.3.1. Biosynthetic pathways of NRPS and PKS.

Peptides are synthesized by a mechanism of "thiotemplates" by enzymes possessing multiple subunits (Figure 2A). The length of these peptides is usually 3 to 15 amino acids, and their structural diversity is larger compared to peptides synthesized by ribosomes, due to the incorporation of modified amino acids (Bonnefond et al., 2011). In contrast to the structural diversity of their products, the synthetic pathway catalyzed by NRPSs is highly conserved. These multifunctional enzyme complexes are able to carry out several dozens of reactions in a highly coordinated way (Bonnefond et al., 2011). All the multi-enzymatic thiotemplate systems are composed of amino-acid activating domains that catalyze the adenylation of the constituent amino acids and the formation of "thioesters". The modular multidomain organization responsible for the elongation of the polypeptide chain consists of an adenylation domain (A), a peptide carrier domain (PCP) also called thiolation (T) and a condensation domain (C). In the last module of the synthesis mechanism, a termination/thioesterase (TE) domain is found, which catalyzes the release of the peptide in a linear, cyclic or cyclic-branched forms, (Figure 2A; Kehr et al., 2011). Analysis of the

sequences of the NRPS genes has allowed us to develop a predictive model that associates the nucleotide sequence and the recognized amino acid (Weber et al., 2015).

Natural products synthesized by polyketide-type routes (PKS) are generated from sequential condensations of short chain carboxylic acids. Many polyketides are clinically useful drugs (Chorny et al., 2009). Polyketide biosynthesis is catalyzed by polyketide synthases. These types of PKS enzymes synthesize a diverse range of molecules depending on the precursor. A typical PKS module consists minimally of an acyltransferase (AT) domain to extend the selected unit and transfer an acyl group; by an Acyl-carrier (ACP) protein and a ketoacyl synthetase (KS) domain for the decarboxylation condensation between the aligned acyl esters, thereby releasing the nascent polyketide chain (Figure 2B) (Kehr et al., 2011).

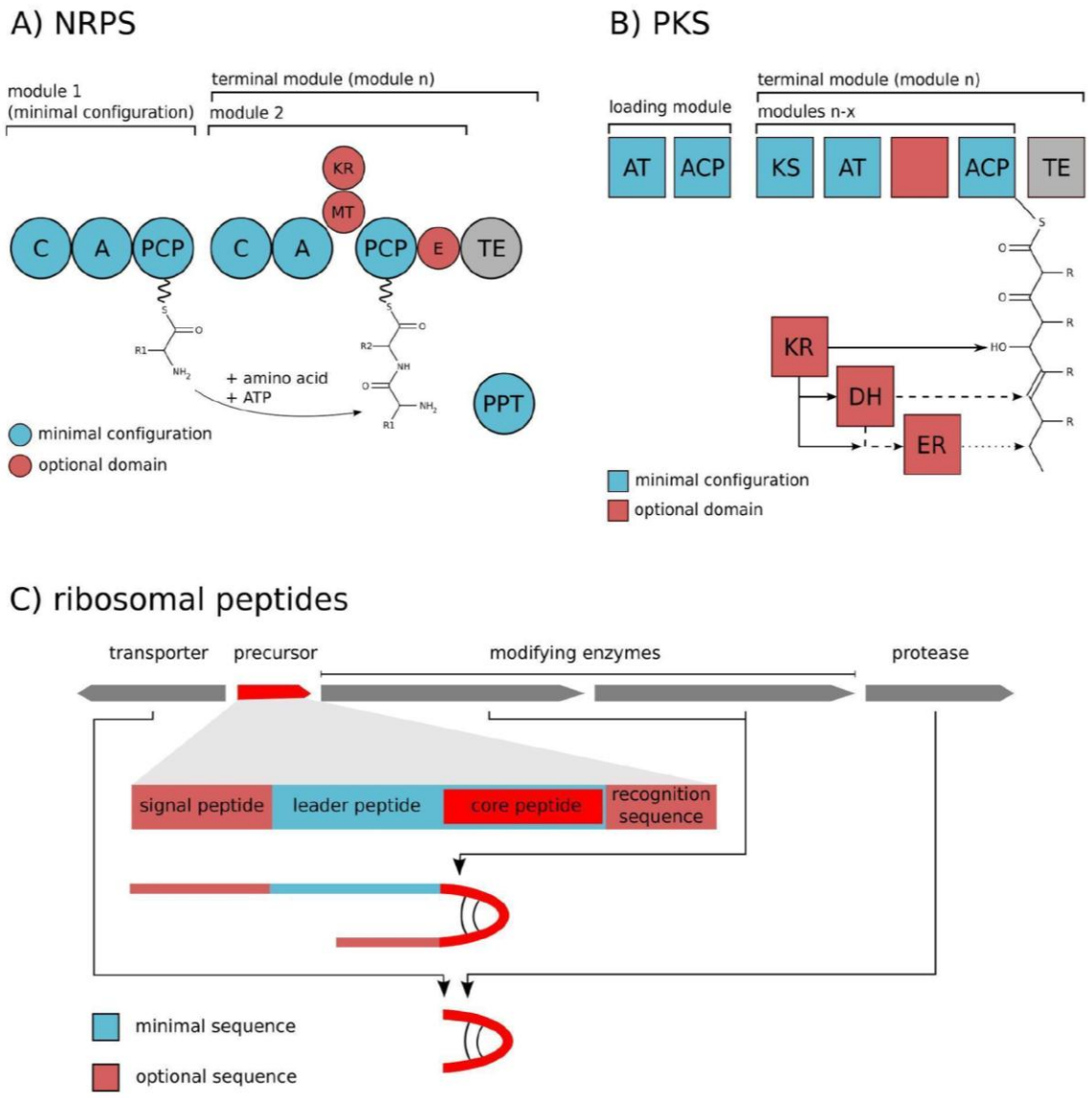

Figure 2. Schematic representation of the enzymatic domains for peptide synthesis by A) Non-Ribosomal Peptide Synthetases (NRPS), B) Polyketide Synthetases (PKS) and C) Ribosomal peptides. Abbreviations: C: domain Condensation; A: Adenylation domain; PCP: carrier proteins Peptidyl, MT: Methyltransferase; E: Epimerase, AT: Acyltransferase; ACP: Acilos, KS: Ketosynase transporter protein; KR: Ketoreductase; DH: Dehydratase; ER: Enoil Reductase, TE: Thioesterase (Kehr et al., 2011).

1.4. *Cylindrospermopsis raciborskii*.

Cylindrospermopsis raciborskii is a fascinating cyanobacterial model since it can synthesize several secondary metabolites. As mentioned *C. raciborskii* is a cyanobacterium responsible for algal blooms that cause severe problems because of the wide variety of toxic metabolites it produces (Figure 3) (Padisák, 1997). *Cylindrospermopsis* sp. was first described in Java (Indonesia) in 1912 and was morphologically characterized in 1972 by Seenayya and Subba-Raju (Woloszynska, 1912), as Gram-negative-like, cylindrical filament organisms able to fix nitrogen. To date, this species has been characterized as an STX producer (Lagos et al., 1999, Bouvy et al., 1999). It also produces cylindrospermopsin (CYL), a toxin which toxicity mechanism is not clear, but it has been shown to inhibit phosphatase activity in hepatocyte cells (Li et al., 2001a, Wiedner et al., 2007). Toxin and non-toxic strains of *C. raciborskii* are spread out worldwide. A particular characteristic of *Cylindrospermopsis* among other cyanobacterial toxigenic organisms is that it does not form dense blooms near the top of the water. Instead, this species tends to localize below

the water surface generally between one to three meters deep (Lamiot F, 2002).

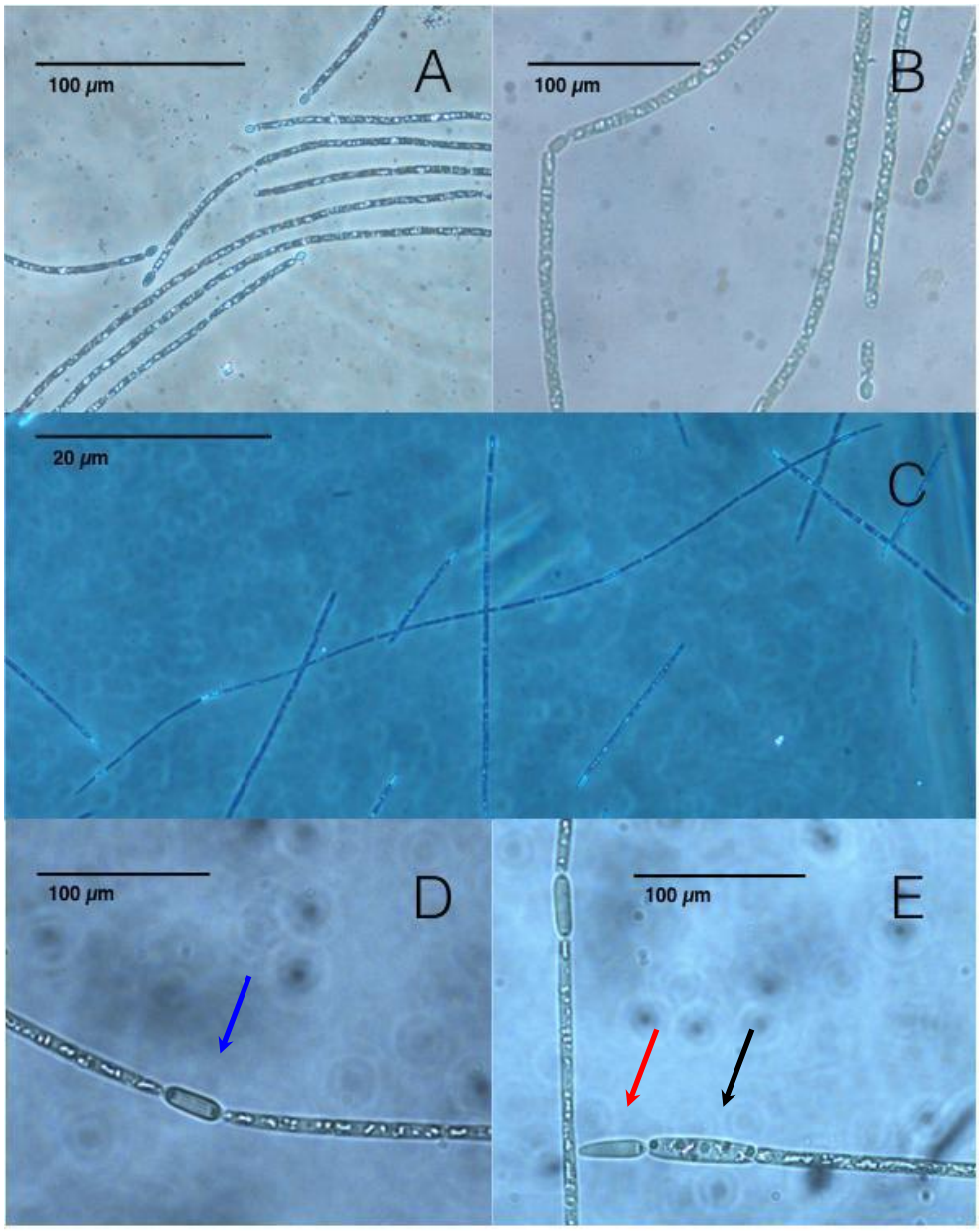

Figure 3. Light microscopy image of *C. raciborskii* CS-508 (A-B) and *C. raciborskii* MVCC14 (C-E). Note morphology similarities of both cell types: the additional intercalated heterocyst like (blue arrow) in MVCC14 instead of just terminal structures (red

arrow), also it possible to observe a resistance cell akinete (black arrow), structure that is not commonly found in this species.

1.5. Advances in cyanobacterial genomics.

Our group recently released the complete sequence genome draft of *Cylindrospermopsis raciborskii* strain CS-505, a cylindrospermopsin-producing strain (Stucken et al., 2010). This study was the first filamentous terminal heterocyst-forming cyanobacteria with a sequenced genome (Stucken et al., 2010). Deposition of complete genome sequences in public databases has allowed determining co-localization of genetic localization (synteny), between different species at a genomic level. As of today, there are 488 cyanobacterial genome sequences publicly available in GenBank and the Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) databases (Table 1). Cyanobacterial genome analyses have shown considerable genetic differences within this phylum, for example in their genome size (~1.4–9.1 Mbp), G+C content (31–63%), number of protein-coding genes (1214–8446) and nucleotide coding proportion (52–94%) (Larsson et al., 2011).

Table 1. Summary of publically available cyanobacterial genomes from cyanobacterial phyla and the number of all genomes found according to order classification (Wattam et al., 2017).

Taxonomy	Phylum	Clase	order	N°
	Cyanobacteria			488
		Gloeobacteria		
			Gloeobacterales	3
		Cyanophyceae		
			Nostocales	83
			Oscillatoriales	66
			Pleurocapsales	5
			Spirulinales	2
			Synechococcales	273
			Chroococcales	48
			Chroococcidiopsidales	2

2. Work Hypothesis

Genomic analyses coupled with functional characterization allows the identification of secondary metabolites in *Cylindrospermopsis raciborskii* with biotechnological potential

3. General and Specific Objectives

3.1. General objective

The general objective of this thesis was to obtain and characterize the complete genome sequence of three cyanobacterial strains of *Cylindrospermopsis raciborskii* to determine the biotechnological capacity of secondary metabolites produced by these strains.

3.2. Specific objectives

1. To analyze and compare the most relevant genetic aspects present in the complete genome sequences of the cyanobacterial *Cylindrospermopsis raciborskii* strains CS-505, CS-508, and MVCC14.
2. To characterize the genomic regions related to the synthesis of secondary metabolites present in each *Cylindrospermopsis raciborskii* genome.
3. To characterize the secondary metabolites of the NRPS and PKS type produced by the *Cylindrospermopsis raciborskii* strains.

4. Methodology

DNA extraction and genome sequencing process

To characterize the complete genomes of *Cylindrospermopsis raciborskii* CS-505, CS-508, and MVCC14, cyanobacterial cultures were grown using MLA medium (Bolch and Blackburn 1996), with 12:12 h light:dark cycles at 25 °C (Castro et al., 2004). Total DNA extractions were carried out using 100 mL of exponential growth culture, obtaining approximately 1 g of wet cell pellet. DNA purification was conducted by standard CTAB protocol (Ausubel et al., 1992). Total cell pellets were mechanically disrupted and resuspended in 500 µL of CTAB buffer and incubated at 55 °C for one h under constant mixing. The DNA was purified using 500 µL phenol/chloroform/isoamyl alcohol (25:24:1) and centrifuged at 8000 x g for 7 min. DNA was precipitated using isopropanol/ammonium acetate (0.54 vol cold isopropanol, 0.08 vol ammonium acetate 7.5 M). Finally, DNA was washed with 70% and then with 90% ethanol and resuspended in 50 µL of pure water. DNA extraction was visualized using red gel staining in a 1% agarose gel under UV light.

In a first stage, the full genome of *C. raciborskii* CS-505 strain was sequenced using the Pacific Biosciences (PacBio) RS II Single-Molecule Real-Time (SMRT), whole-genome sequencing system. Genome reads were assembled using the Hierarchical Genome Assembly Process (HGAP), implemented in the PacBio SMRT Analysis software suite (version 2.2.0). The complete genome and plasmid were curated using GenomeMatcher.

C. raciborskii CS-508 and MVCC14 genome sequences were obtained by a shotgun strategy using Illumina MiSeq sequencing technology in MedicalBioFilm Research and Testing Laboratory, USA. Quality control checks were performed on the raw FASTQ data

using FastQC (version 0.10.1). Sequencing reads were trimmed for sequencing adaptors using Trimmomatic (version 0.32) and the quality filtering and trimming were done by Prinseq-lite (version 0.20.4). Briefly, reads were trimmed for 'N' characters and low-quality nucleotides (Phred score cutoff of 24); and the reads with an average Phred score below 29 and shorter than 80 nt were discarded. A de novo assembly strategy involving multiple algorithms and merging of the individual assemblies was performed. Assemblies by IDBA, SPADes, VELVET and ABYSS algorithms were generated by using the platform MIX software to improve draft assembly by reducing contig fragmentation. Contigs shorter than 1000 bp were discarded.

The gene annotation process of the three draft genomes generated was conducted using the RAST Server 2.0. Predicted coding sequences were extracted from the RAST platform. The homology was evaluated by BLASTp scan, with each predicted ORF as a query against the complete bacterial database.

The homology relationships between genomes of each strain were assessed by GET_HOMOLOGUES (Contreras-Moreira and Vinuesa 2013). Considering a dataset composed of five *C. raciborskii* and one *R. brookii* strains, representing the orthologous genes for each genome.

In a second stage, hassallidin (HAS) extraction was carried out by harvesting one gram of lyophilized *C. raciborskii* pellet from CS-505, CS-508, and CS-509 strains. Pellets were extracted in pure ethanol and posterior cleaning by filtering with solid phase extraction cartridges (LC-8 SPE). The chemical fingerprinting analysis was performed on an LC-MS/MS equipment with a triple-quadrupole mass spectrometer (API 4000 Q Trap,

AB-Sciex. Finally, antimicrobial assays were carried out in solid Mueller-Hinton media to evaluate antifungal capacity.

C. raciborskii MVCC14 chemical extraction for saxitoxin (STX) analysis, was performed by centrifugation at 16,000 x g for 15 min to yield a cell pellet. Total cell pellet was dried in speed vacuum (DNA Speed Vac, SAVANT mod. DNA 110-230) and was extracted in 300 μL of 0.05 M acetic acid. Samples were disrupted in an ultrasonic cell disruptor three times (20 s each one). The extracts were then centrifuged at 5000 x g for 10 min, filtered through a 0.45 μm membrane filter and chromatographically analyzed by LC-MS using ABI-SCIEX-4000 Q Trap, triple quadrupole mass spectrometer equipped with a TurboSpray interface coupled to an Agilent model 1100 LC.

5. Results

Cylindrospermopsis raciborskii genome sequencing and analysis.

The first genome draft of *C. raciborskii* CS-505 was obtained in 2010 (Stucken et al., 2010). Taking advantage of new sequencing technologies we were able to obtain its fully assembled genome. Moreover, we were able to identify an extra-chromosomal element in *C. raciborskii* CS-505 not previously characterized and apparently exclusive of this cyanobacteria.

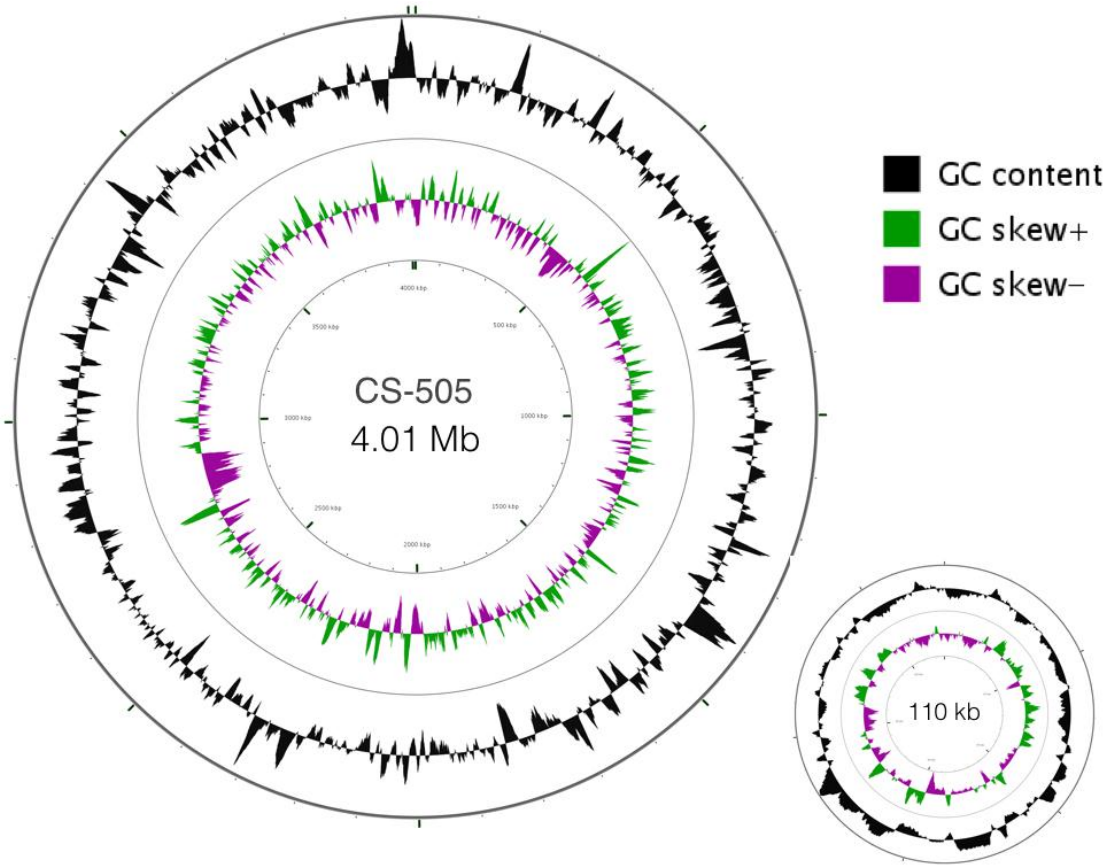

Figure 4. Circular maps showing *C. raciborskii* CS-505 draft genome of its circular chromosome and extrachromosomal element. Genome sizes are indicated in the center of each displayed map. The map explained from the outer to inner circles, each one represents a main scaffold, with the color code representing GC content in black and asymmetric nucleotide composition by GC skew index in green (positive, leading strand), and pink (negative, lagging strand).

The genome assembly of *C. raciborskii* strain CS-505 (Figure 4), resulted in six scaffolds with a total length of 4,159,260 bp. The average length of the scaffolds was 693,210 bp, with the longest and shortest scaffolds sizes of 4,011,384 bp and 2,519 bp, respectively. CheckM analysis (Parks et al., 2014) indicated genome completeness of 99.85%, a contamination level of 0.22% and no strain heterogeneity identified.

We obtained a total of 8,308,910 paired-end reads for *C. raciborskii* CS-508 strain and 28,711,437 paired-end reads for MVCC14 strain. The final assembly resulted in 163 contigs for CS508 and 99 contigs for MVCC14, accounting for 3,558,956 bp and 3,594,524 bp, respectively. CheckM analysis indicated genome completeness of 97.57% for CS-508 and 96.29% for MVCC14 (Figure 5). After bioinformatics evaluation, we were not able to identify any non-chromosomal genetic material. Insights from the genome sequence are described in Table 2.

Figure 5. Circular map showing *C. raciborskii* CS-508 and MVCC14 draft genomes circular chromosome. Genome sizes are indicated in the center of the displayed map, composed by 163 and 99 contigs, respectively. The map explained from the outer to inner circles, each one represents a main scaffold, with the color code representing Coding DNA Sequence and tRNA in blue and orange color, GC content in black and asymmetric nucleotide composition by GC skew index in green (positive, leading strand), and pink (negative, lagging strand).

Genome characteristics *C. raciborskii* strains					
	CS-505	**CS-506**	**CS-508**	**CS-509**	**MVCC14**
Size (Mb)	4.1	4.1	3.5	4.0	3.6
Contig nº	6	4,887	193	2,891	99
G + C content (%)	42.9	41.1	43.0	40.7	44.0
N50 (bp)	NA	25	62,252	56,411	150,437
N90 (bp)	NA	412	1,027	268	1,035
Largest scaffold (bp)	4,011,384	67,497	193,915	188,708	299,478
CDS*	4,044	3,268	3,415	3,416	3,754
Unique CDS*	181	176	140	101	128
Common CDS*	2,767	2,767	2388	2,767	2,397
RNAs	51	52	42	46	56

Table 2. Genomic features of *C. raciborskii* genomes sequenced available.

Genome comparison among *Cylindrospermopsis* strains.

Given the geographical coordinates of the five *Cylindrospermopsis* strains under study (CS strains from Australia and MVCC14 from Uruguay), we could expect that they might show sequence divergence. The five genomes analyzed were found to have a relatively low average nucleotide identity (minimum of 93.73%), (Table 3), considering a threshold value of 95% to be recognized as members of the same species. Sequenced genomes were also compared to a previously published *R. brookii* D9 genome, a phylogenetically related species. According to the ANI value, *C. raciborskii* MVCC14 and *R. brookii* D9 share a 97.73% of nucleotide identity, suggesting that they could be considered a member of the same species or subspecies.

Shared orthologous genes between species and strain-specific genes within a given *Cylindrospermopsis* and *Raphydiopsis* species were examined. We found 1,238 orthologs

shared in the six tested strains, which were defined as the core genome dataset (Figure 6). From the number of unique genes presents in the orthologous groups, we estimated a gene content dissimilarity quantitative index between microbial strains. As well as ANI value cutoffs of 0.88 in this case (dissimilarity between *C. raciborskii* CS-505 and *R. brookii* D9), can be used for species delimitation (Figure 6). The dissimilarity index between *C. raciborskii* MVCC14 and *R. brookii* D9, interestingly, demonstrated that cyanobacteria from different species, but isolated from closely surrounding regions, could have so highly genomic similarity, even to be considered as members of the same species.

Table 3. Average nucleotide identity (ANI) values for the sequenced *C. raciborskii* and *R. brookii* strains.

	Cylindrospermopsis CS-505	Cylindrospermopsis CS-506	Cylindrospermopsis CS-508	Cylindrospermopsis CS-509	Cylindrospermopsis MVCC14	*R. brookii* D9
Cylindrospermopsis CS-505	—	99.31	99.73	99.77	93.73	93.26
Cylindrospermopsis CS-506	99.31	—	99.39	99.32	93.45	92.85
Cylindrospermopsis CS-508	99.73	99.39	—	99.76	93.82	93.25
Cylindrospermopsis CS-509	99.77	99.32	99.76	—	93.80	93.23
Cylindrospermopsis MVCC14	93.73	93.45	93.82	93.80	—	97.17
R. brookii D9	93.26	92.85	93.25	93.23	97.17	—

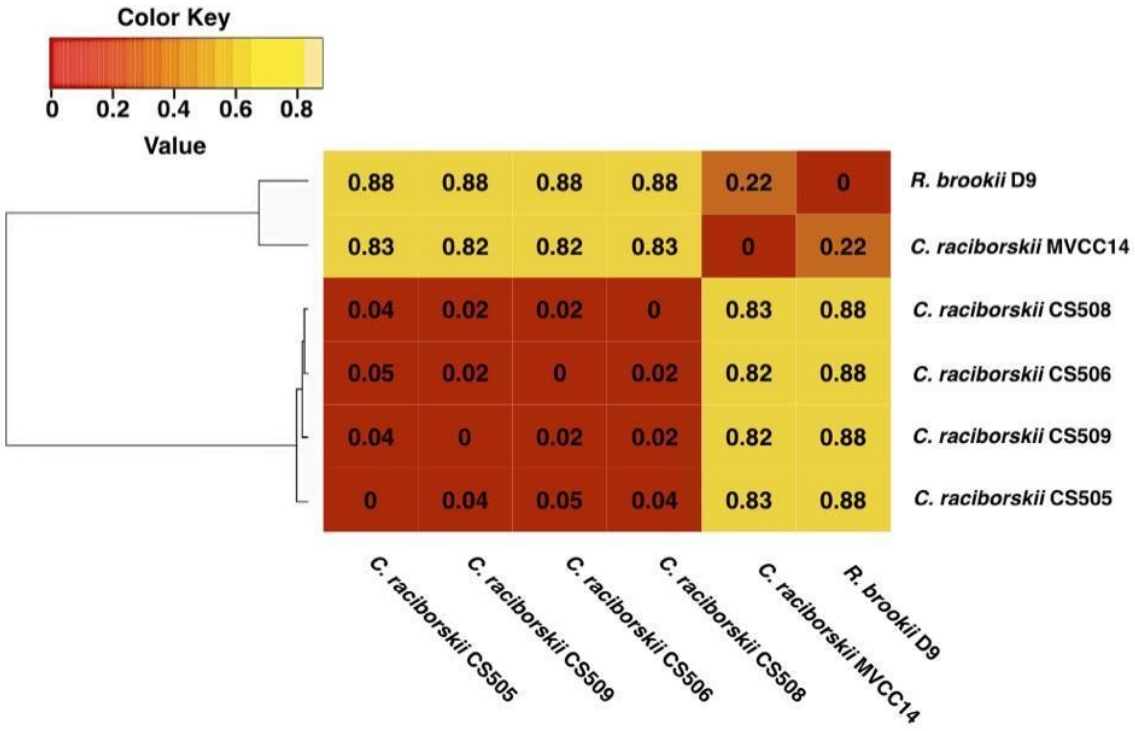

Figure 6. Dissimilarity heat maps *C. raciborskii* and *R. brookii* orthologous gene core. The dissimilarity matrix was constructed with 1238 representative shared genes based on BBDH analysis.

Uncovering the secondary metabolite capacity of the genus *Cylindrospermopsis*.

A comparative genomic analysis (Figure 7) confirmed that CS-508 and CS-509 strains had the capacity to produce secondary metabolites since discrete gene clusters for HAS biosynthesis were identified (Figure 7). Moreover, both genomes contained a similar HAS gene cluster supported by the findings of genomic synteny. We were able to detect

the presence of hassallidin in CS-505 and CS-509 supernatants. Hassallidin detection was negative in CS-508 cultures, according to LC-MS/MS analysis (Figure 8).

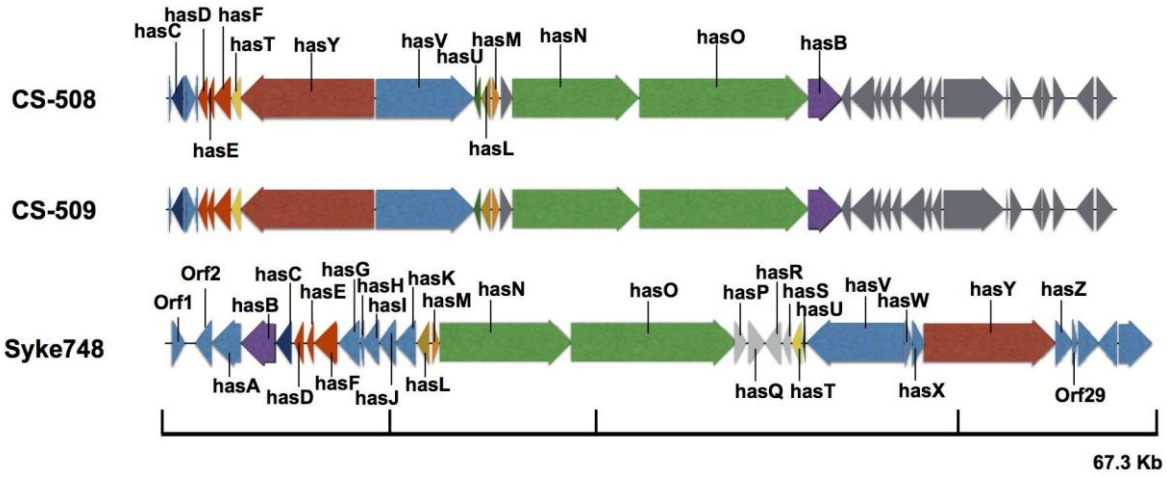

Figure 7. Schematic representation of hassallidin (has) gene clusters of *C. raciborskii* strains CS-508, CS-509 and two strains of *Anabaena* sp. 90 and Syke748. Genes in color type are in synteny based on Syke748 has gene cluster.

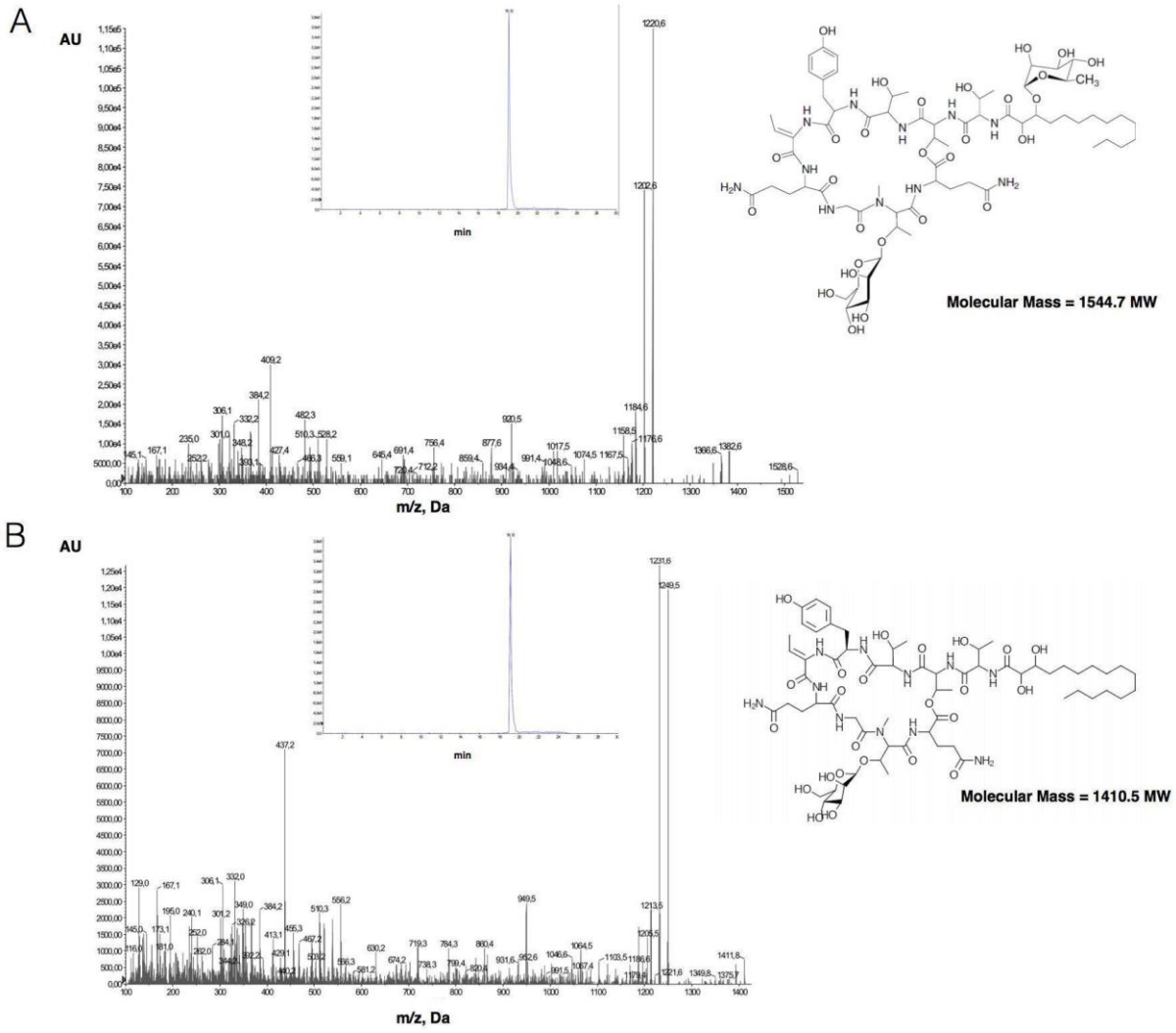

Figure 8. LC-MS/MS of A: *C.raciborskii* strain CS-505 and B: *C. raciborskii* strain CS-509. Each upper figure shows the retention time and fragmentation fingerprinting of every cyanobacteria cell extract; figures at the right side indicate the hassallidin structure identified.

Saxitoxin Gene Cluster of SXT in *C. raciborskii* MVCC14.

By comparative genomic analysis against a previous characterized STX gene cluster in *C. raciborskii* T3 and *R. brookii* D9, we confirmed that MVCC14 produce STX (Figure 9), as determined by chromatography. In this genome, we identified a gene cluster 100% identical to the PKS-like complex described in *R. brookii* D9 responsible of the synthesis of STX, a paralytic biotoxin produced by marine dinoflagellates and freshwater cyanobacteria. The *sxt* gene cluster found in MVCC14 has a similar distribution and toxin profile to *R. brookii* D9. Finally, we did not find NRPS sequences in the MVCC14 genome.

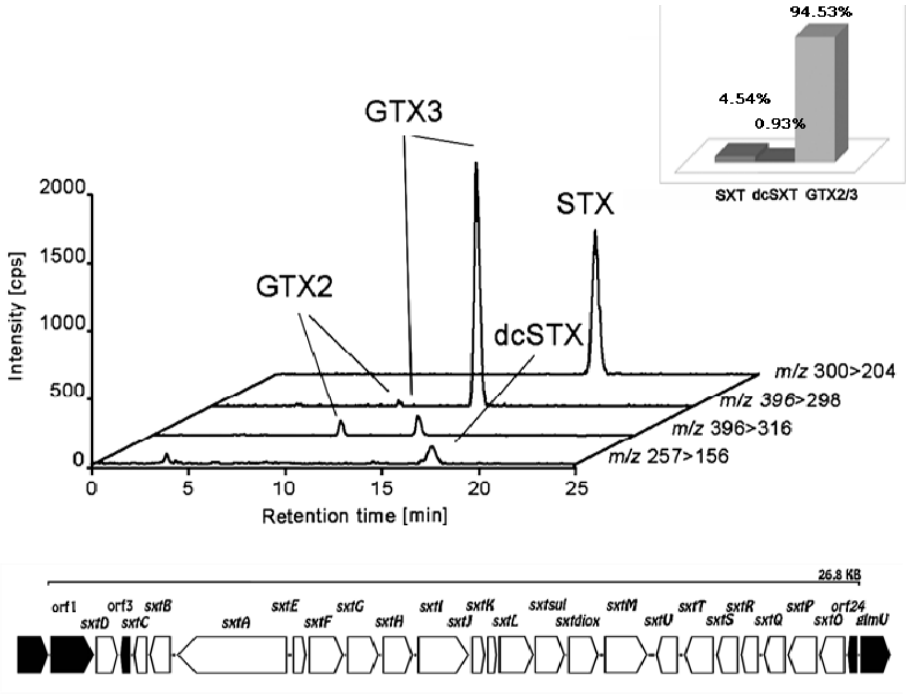

Figure 9. LC-MS/MS analysis of PSP toxin content in *C. raciborskii* MVCC14. (Upper) Toxin content (pmol mg1 of fresh weight). (Down) LC-MS/MS spectrum of PSP toxin components of a cellular extract.

6. Conclusions

The results presented in this thesis show the relevance of predefining the capacity molecules that could be produced by a highly dynamic and metabolically active environmental microorganism. We were also able to identify key genes related with bioactive metabolites biosynthesis.

The most significant findings are presented below:

- Sequencing a genome of a non-toxic strain CS-508 allowed us to identify a gene cluster related to biosynthesis of the antifungal hassallidin. In addition, we were able to sequence a complete genome of the toxic strain MVCC14, identifying the functional *sxt* gene cluster related to SXT production. Finally, re-sequencing of CS-505 strain allowed us to identify a capacity extra-chromosomal element. Based on bioinformatics analysis we developed a full genome comparison that could help to understand the phylogenetic relationship between strain and location of isolation.
- Application of gene orthology dissimilarity to highly similar microbial species suggested it as an effective approach in classifying closely related microorganisms into subgroups or subspecies.
- Finally, we determined that *C. raciborskii* CS-509 is capable of releasing HASs into the extracellular medium, on the contrary to CS-508 strain. The latter containing a complete *has* gene cluster, but we were not able to detect hassallidin in the cell extracts or extracellular media contents of the strain. These results correlate

with the presence of a highly conserved but partially rearrangement has gene cluster identified in the toxic strain CS-505. If production of HASs confers any enhanced fitness or adaptability to an ecosystem, so far, is no clear how this molecule could be useful.

7. Future directions

The research presented would be useful to identify requirements and capacity candidates to improve biotechnological identification of new molecules. This work generates three types of future research questions.

- Our LC-MS/MS results showed ion fragmentation patrons for a non-identified compound when we characterized the profiles of non-toxic strain CS-508 cell extracts. These results were not included in the present work because they are not related or could be predicted by bioinformatics. Such characterization and purification would allow evaluating a capacity biological function capacity would give a commercial value for this strain.
- Our results indicated that while a gene cluster for hassallidin biosynthesis is present in CS-508 strain, the cognate bioactive compound was absent. A contrasting discovery from CS-505 and CS-509 strains, which have similar gene clusters and produce hassallidin. The absence of hassallidin in CS-508 extracts could be explained by post-transcriptional modifications in the biosynthesis cluster or point mutations. Further search would look deeply into the cyanobacterial transcriptome, or under what conditions hassallidin production is favored. This results would have biotechnological implications.
- Finally, further studies will look at biotechnological aspects of metabolite production. A study of metabolite production efficiency during bacterial growth, in

batch cultures, optimizing light conditions, media, and agitation, would render larger quantities of hassallidins.

8. References

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1992) Current protocols in molecular biology. New York: Greene Publishing Association; Wiley-Interscience. v.1

Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stöckel, J., Liberton, M., Min, H., and Sherman, L. A. (2011). Novel Metabolic Attributes of the Genus Cyanothece, Comprising a Group of Unicellular Nitrogen-Fixing Cyanobacteria, *2*(5): 1–10

Batoréu, M.C.C., Dias, E., Pereira, P. and Franca, S., (2005). Risk of human exposure to paralytic toxins of algal origin.Environ. Toxicol. Phar. 19, 401-406

Bolch, C.J.S. and Blackburn, S.I. (1996). Isolation and purification of Australian isolates of the toxic cyanobacterium *Microcystis aeruginosa*. Journal of Applied Phycology, 8:5–13

Bonnefond, L., Arai, T., Sakaguchi, Y., Suzuki, T., Ishitani, R. and Nureki, O. (2011) Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. Proc Natl Acad Sci U S A108(10):3912–3917.

Bouvy, M., Molica, R.J.R., De Oliveira, S., Marinho, M. and Beker B (1999) Dynamics of a toxic cyanobacterial bloom (Cylindrospermopsis raciborskii) in a shallow reservoir in the semi-arid region of Northeast Brazil. Aquat Microb Ecol 20:285–29

Bricelj, VM., Lee, JH., Cembella, AD., and Anderson, DM. (1990). Uptake kinetics of paralytic shellfish toxins from the dinoflagellate *Alexandrium fundyense* in the mussel *Mytilus edulis.* Mar Ecol: Prog Ser, 63:177–188

Calteau, A., Fewer, D. P., Latifi, A., Coursin, T., Laurent, T., Jokela, J., … Gugger, M. (2014). Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics*, *15*(1), 977

Castro, D., Vera, D., Lagos, N., García, C., and Vásquez, M. (2004). The effect of temperature on growth and production of paralytic shellfish poisoning toxins by the cyanobacterium *Cylindrospermopsis raciborskii* C10. Toxicon: official journal of the International Society on Toxicology, 44(5), 483–9

Cembella, AD. (1998). Ecophysiology and metabolism of paralytic shellfish toxins in marine microalgae. In: Anderson DM, AD Cembella & GM Hallegraeff (eds). Physiological ecology of Harmful Algal Blooms, 381-403

Chorny, M. and Levy, R. (2009) Site-Specific Analgesia with Sustained Release Liposomes. PNAS, 106(17):6891–92

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. *Applied and Environmental Microbiology*, *79*(24):7696–7701

Dangi, S. K., Dubey, S., and Bhargava, S. (2017). Cyanobacterial Diversity : A Capacity Source of Bioactive Compounds, 4(1): 19–21

Hurley, W., Wolterstorff, C., MacDonald, R., and Schultz, D. (2014). Paralytic Shellfish Poisoning: A Case Series. Western Journal of Emergency Medicine, 15(4): 378–381

Kellmann, R., Mihali T. K., Jeon, Y. J., Pickford, R., Pomati, F. and Neilan, B. A., (2008) Biosynthetic intermediate analysis and functional homology reveal a putative saxitoxin gene cluster in cyanobacteria. Appl. Environ. Microbiol. 74: 4044 - 4053

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

Kopp, R.E., Kirschvink, J.L., Hilburn, I.A. and Nash, C.Z. (2005). The Paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. Proc. Natl. Acad. Sci. USA 102: 11131–11136

Kumar, A., Ergas, S., Yuan, X., Sahu, A., Zhang, Q., Dewulf, J., Malcata F.X. and Langenhove, H. (2010). Enhanced CO2 Fixation and Biofuel Production via Microalgae: Recent Developments and Future Directions. *Trends in Biotechnology* 28(7):371–80

Lagos, N., Onodera, H., Zagatto, P., Andrinolo, D., Azevedo, S. and Oshima, Y. 1999. The first evidence of paralytic shellfish toxins in the freshwater cyanobacterium *Cylindrospermopsis raciborskii*, isolated from Brazil. Toxicon 37 1359-1373

Lamiot, F., (2002) Fourmilière st amand 2002 vue. Wood Ants, (north of France)

Lau, N., Matsuim M. and Abdullah, AA. (2015) Cyanobacteria : Photoautotrophic Microbial Factories for the Sustainable Synthesis of Industrial Products. BioMed Research International, 2015, Article ID 754934

Larsson, J., Nylander, J. A. A., and Bergman, B. (2011). Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. BMC Evolutionary Biology, 11:187

Li, R., Carmichael, W.W., Brittain, S., Eaglesham, G.K., Shaw, G.R. and Watanabe, M.M. (2001) First report of the cyanotoxins cylindrospermopsin and deoxycylindrospermopsin from *Raphidiopsis curvata* (cyanobacteria). J. Phycol. (37), 1121–1126

Llewellyn, L.E. (2006) Saxitoxin, a toxic marine natural product that targets a multitude of receptors. Nat Prod Rep, 23: 200–222

Méjean, A., Mazmouz, R., Mann, S., Calteau, A., Médigue, C. and Ploux, O. (2010) The genome sequence of the cyanobacterium *Oscillatoria* sp. PCC 6506 reveals several gene clusters responsible for the biosynthesis of toxins and secondary metabolites. J Bacteriol. 192(19):5264-5.

Mihali, T., Kellmann, R., Muenchhoff, J., Barrow, K. and Neilan B. (2008) Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. Applied and environmental microbiology 74: 716-722

Moffitt, M.C. and Neilan, B.A. (2004) Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. Appl Environ Microbiol. 70(11):6353-62

Moreira, D.L., Guyader, H. and Philippe, H. (2000) The origin of red algae and the evolution of chloroplasts. Nature. 4;405(6782):69-72

Moss, B., (1988). Ecology of Fresh Waters. Blackwell, Oxford, 417

Narahashi, T., Haas, H.G. and Therrien, E.F. (1967). Saxitoxin and tetrodotoxin: comparison of nerve blocking mechanism. Science, 157(3795) 1441-1442

Narechania, A., Baker, R. H., Sit, R., Kolokotronis, S.-O., DeSalle, R., and Planet, P. J. (2012). Random Addition Concatenation Analysis: a novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. Genome Biol. Evol, 4: 30–43

Neuhof, T., Schmieder, P., Preussel, K., Dieckmann, R., Pham, H., Bartl, F. and von Döhren, H. (2005). Hassallidin A, a glycosylated lipopeptide with antifungal activity from the cyanobacterium *Hassallia* sp. J. Nat. Prod, 68: 695–700

Neuhof, T., Schmieder, P., Seibold, M., Preussel, K., and von Döhren, H. (2006). Hassallidin B—Second antifungal member of the Hassallidin family. Bioorg. Med. Chem. Lett, 16:4220–4222

Nyok-Sean, L., Minami, M. and Amirul, A. (2015) Cyanobacteria: Photoautotrophic Microbial Factories for the Sustainable Synthesis of Industrial Products," BioMed Research International, 2015:9

Padisák J. and Istvánovics V. (1997) Differential response of blue-green algal groups to phosphorus load reduction in a large shallow lake: Balaton, Hungary. Verh. Int. Ver. Theor. Angew. Limnol. 26, 574–580

Pancrace, C., Barny, M., Ueoka, R., Calteau, A., Scalvenzi, T., Pédron, J., … Gugger, M. (2017). Insights into the Planktothrix genus : Genomic and metabolic comparison of benthic and planktic strains. *Nature Publishing Group*, 1–10

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W. 2014. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research, 25: 1043-1055

Ram, S., Amandeep, W., Pratibha, Jasvirinder, S. and Davinder, S. (2017) New cell surface lectins with complex carbohydrate specificity from cyanobacteria. *IJEB*, 55(07);514-522

Reyes-Prieto, A., Weber, AP., and Bhattacharya, D. (2007) The origin and establishment of the plastid in algae and plants. Annu Rev Genet. 41:147-68

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

Ruffing, A.M. (2011) Engineered cyanobacteria: teaching an old bug new tricks. Bioeng Bugs. 2(3):136-49

Sarup, R., Kaur, A., Singh, J., Pal, D., and Kennedy, JF. (2017) Cyanobacterial lectins characteristics and their role as antiviral agents. International Journal of Biological Macromolecules,102:475–496

Schirrmeister, B. E., Gugger, M., and Donoghue, C. J. P. (2015). Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. Palaeontology, 58(5): 769–785

Shishido, T.K., Humisto, A., Jokela, J., Liu, L., Wahlsten, M., Tamrakar, A., Fewer, D.P., Permi, P., Andreote, A.P.D., Fiore, M.F. and Sivonen, K. (2015). Antifungal Compounds from Cyanobacteria. *Mar. Drugs*, 13: 2124-2140

Sivonen, K. and Jones, G. (1999) Cyanobacterial Toxins. In: Chorus I, Bartram J, eds. Toxic Cyanobacteria in Water: A guide to their public health consequences, monitoring and management. London: E & FN Spon. pp 41–111

Stucken, K., John, U., Cembella, A., Murillo, AA., Soto-Liebe, K., Fuentes, JJ., Friedel, M., Plominsky, A., Vásquez, M. and Glöckner, G. (2010). The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications. PLoS ONE 5(2): e9235

Tan, L.T. (2007) Bioactive natural products from marine cyanobacteria for drug discovery. Phytochemistry. 68(7):954-7

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T. and Neilan, B.A. (2000) Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. Chemistry & Biology. 7: 753-764

Uzair, B., Tabassum, S., Rasheed, M. and Firdous-Rehman, S. (2012). Exploring Marine Cyanobacteria for Lead Compounds of Pharmaceutical Importance. The ScientificWorld Journal. Article ID 179782, 10

Vestola, J., Shishido, T. K., Jokela, J., Fewer, D. P., Aitio, O., Permi, P., and Wahlsten, M. (2014). widespread among cyanobacteria and are the end-product of a nonribosomal pathway, PNAS, 1909–1917

Vijayakumar, S, and Menakha, M. (2015). Pharmaceutical applications of cyanobacteria. A review. J Acute Med, 5(1):15–23

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015) antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.*43, W237–W243

Wiedner, C., Visser, P.M., Fastner, J., Metcalf, J., Codd, G.A. and Mur, L. (2003) Effects of Light on the Microcystin Content of *Microcystis* Strain PCC 7806. Applied and Environmental Microbiology 1475–1481

Woloszynska, J. (1912) Das Phytoplankton einiger Javanian Seen mit Berücksichtigung des Sawa-Planktons. *Bull. Int. Acad. Sci. Cracoviae Ser. B*6, 649–709.

Yunes, J., S., De La Rocha, S., Giroldo, D., Silveira, S.B., et al. (2009) Release of carbohydrates and proteins by a subtropical strain of *Raphidiopsis brookii* (cyanobacteria) able to produce saxitoxin at three nitrate concentrations. J Phycol, 45:585–591

CHAPTER I

Complete genome of a Cylindrospermopsin-producing cyanobacterium *Cylindrospermopsis raciborskii* CS505, containing a circular chromosome and a single extrachromosomal element

Complete genome of a Cylindrospermopsin-producing cyanobacterium *Cylindrospermopsis raciborskii* CS505, containing a circular chromosome and a single extrachromosomal element

Juan J. Fuentes-Valdés,[a,b] Alvaro M. Plominsky,[c,d] Eric E. Allen,[e,f] Javier Tamames[g] and Mónica Vásquez[b]

Departament of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile[a], Department of Molecular Genetics and Microbiology, Faculty of Biological Sciences, Pontificia Universidad Católica de Chile, Santiago, Chile[b], Department of Oceanography, Universidad de Concepción, P.O. Box 160-C, Concepción, Chile[c], Instituto Milenio de Oceanografía, Universidad de Concepción, Concepción, Chile[d], Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, 92093-0202 USA[e], Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA[f], Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain[g].

Email addresses:

JJFV: jsfuente@uc.cl

AMP: aamunoz1@uc.cl

EEA: eallen@ucsd.edu

JT: jtamames@cnb.csic.es

MV: mvasquez@bio.puc.cl

GENOME ANNOUNCEMENT

Cylindrospermopsis raciborskii is defined as a planktonic nitrogen-fixing freshwater cyanobacterium (1). Strain CS505 (CSIRO culture collection), was isolated from Solomon Dam, North Queensland, Australia (2) and characterized based on its production of the hepatotoxin cylindrospermopsin (CYL), a potent protein synthesis inhibitor (3). In 2010, the draft genome of this strain was analyzed by a combination of 454 and Sanger sequencing, yielding 95 scaffolds with a total length of 3,879,017 bp (4). The cluster associated with the synthesis of CYL toxin was identified, as it had been previously described (5). The strain was grown in MLA liquid media (6) at 25 - 28ºC under 12:12-h light/dark with a photon flux density of 40 µmol m^{-2} s^{-1} with no aeration.

In this work, the full genome of *C. raciborskii* CS505 strain was sequenced using the Pacific Biosciences (PacBio) RS II Single-Molecule Real-Time (SMRT) whole-genome sequencing system and assembled using the hierarchical genome assembly process (HGAP) implemented in the PacBio SMRT Analysis software suite (version 2.2.0). The assembly resulted in 6 scaffolds with a total length of 4,159,260 bp. The average length of the scaffolds was 693,210 bp, with the longest and shortest scaffolds sizes of 4,011,384 bp and 2,519 bp, respectively. CheckM analysis (7) indicated genome completeness of 99.85% with a contamination level of 0.22% and no strain heterogeneity identified. The complete genome and plasmid were annotated using RAST (8) and curated using GenomeMatcher (9).

Nucleotide sequence accession numbers. The complete genome sequence of *Cylindrospermopsis raciborskii* CS505 has been deposited in NCBI (BioProject ID: PRJNA323609).

ACKNOWLEDGMENTS

This work was supported by Grant Fondecyt Regular 1131037, 1161232; FONDECYT postdoctorate 3140191.

Address correspondence to Mónica Vásquez, mvasquez@bio.puc.cl.

REFERENCES

1. Woloszynska J. 1912. Das Phytoplankton einiger javanischer Seen mit Berücksichtigung des Sawa-Planktons, Bull. Int. Acad. Sci. Cracovie, Ser. B. 6:649–709.
2. Saker M.L. 2014 Cyanobacterial blooms in tropical north Queensland water bodies. Ph.D. Thesis, James Cook University, Townsville, Australia. 2000.3.↲ Nagata Y, Senbongi J, Ishibashi Y, Sudo R, Miyakoshi M, Ohtsubo Y, Tsuda M.
3. Griffiths DJ, Saker M.L. 2003 The Palm Island mystery disease 20 years on: a review of research on the cyanotoxin cylindrospermopsin, Environ. Toxicol. 18:78–93.
4. Stucken K, Uwe J, Cembella A, Murillo AA, Soto-Liebe K, Fuentes-Valdés JJ, Friedel M, Plominsky AM, Vásquez M, Glöckner G. 2010. The Smallest Known Genomes of Multicellular and Toxic Cyanobacteria: Comparison, Minimal Gene Sets for Linked Traits and the Evolutionary Implications. PLoS ONE 5(2):e9235.

5. Mihali TK, Kellmann R, Muenchhoff J, Barrow KD, Neilan BA. 2008. Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis, Appl. Environ. Microbiol. 74:716–722.
6. Castro D, Vera D, Lagos N, García C, Vásquez V. 2004. The effect of temperature on growth and production of paralytic shellfish poisoning toxins by the cyanobacterium Cylindrospermopsis raciborskii C10, Toxicon 44:483–489.
7. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2014. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research, 25: 1043-1055.
8. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 9:75.
9. Ohtsubo Y, Ikeda-Ohtsubo W, Nagata Y, Tsuda M. 2008. GenomeMatcher: a graphical user interface for DNA sequence comparison. BMC Bioinformatics. 9:376.

CHAPTER II

Draft genome sequences of *Cylindrospermopsis raciborskii* strains CS-508 and MVCC14, isolated from freshwater bloom events in Australia and Uruguay

Draft genome sequences of *Cylindrospermopsis raciborskii* strains CS-508 and MVCC14, isolated from freshwater bloom events in Australia and Uruguay

Authors

Juan J Fuentes-Valdés[1,2], Katia Soto-Liebe[1], Danilo Pérez-Pantoja[3], Javier Tamames[4], Lucy Belmar[1], Carlos Pedro-Alió[4], Daniel Garrido[2] and Mónica Vásquez[1*]

Institutional Affiliation

[1]Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Santiago, 8331150, Chile

[2]Department of Chemical and Bioprocess Engineering, Pontificia Universidad Católica de Chile, Santiago, 7820436, Chile

[3]Programa Institucional de Fomento a la Investigación, Desarrollo e Innovación Universidad Tecnológica Metropolitana, Santiago, 8940577, Chile

[4]Systems Biology Program, CNB, CSIC, Calle Darwin 3, 28049, Madrid, Spain

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

Corresponding author

[*]Correspondence: mvasquez@bio.puc.cl

Keywords:

Cylindrospermopsis, bloom, cyanobacteria, environmental toxicity, non-ribosomal peptide-synthetase, polyketide synthases.

Abbreviations

Cylindrospermopsis: *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya & Subba Raju

Introduction

Cyanobacterial bloom-forming species are a persistent global problem [1,2]. *Cylindrospermopsis raciborskii* is a species responsible for algal blooms that cause serious problems because of the wide variety of toxic compounds that it produces [3,4]. Animal consumption of contaminated water with toxic metabolites produces symptoms associated with dermal rash, neural disturbance, hepatic and digestive disorder, and in some cases causing death [4,5]. *C. raciborskii* was first described in Java (Indonesia) in 1912 [6], and was morphologically characterized in 1972 by Seenayya and Subba-Raju [7] as a Gram-negative-like, cylindrical filament able to fix nitrogen. To date, this species has been characterized as a producer of saxitoxin (STX), a neurotoxin able to block voltage-dependent mammalian sodium channels [8]. It also produces cylindrospermopsin (CYL), a toxin related with phosphatase metabolic inhibition in hepatocyte cells [9]. Recently, an anti-fungal glycolipopeptide affecting the plasma membrane integrity of *Candida albicans* cells, classified as hassallidin (HAS), has also been identified [10–12].

In order to understand the mechanisms responsible for the synthesis of these toxins, representative strains of this species have been characterized both genetically and chromatographically [13]. To date, Australian isolates have been characterized as CYL producers (CS-505 and CS-506), HAS producers (CS-505 and CS-509) and as non-toxin producers (CS-508) (unpublished data). In addition, the Uruguayan strain MVCC14 has been described as an STX producer [14]. Moreover, a Brazilian isolate *Raphidiopsis brookii* D9, a species phylogenetically closely related to *C. raciborskii* (Fig 1), has also been reported as an STX producer [15–17]. The complete genome of *C. raciborskii* CS-505

and draft genomes of strains CS-506, CS-509, and *R. brookii* D9 are currently available [16,18].

To provide further data to better understand the genomics and physiology of *C. raciborskii*, including its high capacity for dispersal, we performed a genome sequence analysis of Australian strain CS-508 and Uruguayan strain MVCC14, including gene annotation using the Clusters of Orthologous Group (COG) database [19]. Moreover, we also conducted a comparative genome analysis of five *C. raciborskii* strains: CS-505, CS-506, CS-508, CS-509, and MVCC14, in addition to *R. brookii* D9 to identify common genes.

Organism Information

Classification and features

C. raciborskii is a relevant environmental species causing harmful blooms in freshwater environments, with certain strains synthesizing toxins.

C. raciborskii species were initially described as microorganisms growing in the tropics, although, they have been reported in temperate freshwaters [20]. As previously described [21], the cells belonging to the genus *Cylindrospermopsis* could either be cylindrical filaments with terminal nitrogen fixation structures (heterocysts) (Fig. 1A-E) or resistant cells (akinetes). Both structures could be differentiated under nutrient-deficient culture media. In heterocyst-forming cyanobacteria, heterocysts are distributed in semi-regular intervals along the filament or only in the terminal position. The presence of intercalated

heterocysts in *C. raciborskii* has rarely been observed and, thus, has been described as a species with terminal heterocysts [22]. However, we observed intercalated heterocysts in strain MVCC14 under nitrogen starvation and different nitrogen conditions (Fig 1C-E). The distribution of the heterocysts along the filament has been the subject of research by comparing genetic and physiological traits between *Cylindrospermopsis* and *Anabaena*, as models of differential patterns [23,24]. *Anabaena* sp. PCC7120 differentiates heterocysts after every 8 to 12 vegetative cells under nitrogen deprivation [23,24]. We were able to observe heterocysts more frequently in some filaments; however, regularity between heterocyst cells was every 30 neighboring vegetative cells approximately (30 ± 7.4, average ± SD).

This is the first report showing the transient presence of intercalary heterocyst in this *C. raciborskii* strain, and further research should help to understand the genetic control that regulates this sporadic distribution of heterocysts in this *C. raciborskii* strain.

Despite their very similar morphology, *C. raciborskii* and *R. brookii* have been classified as different species because the latter is unable to fix nitrogen and does not develop heterocysts (*e.g.* [25]). Here, the maximum likelihood phylogenetic tree of 16S-rRNA gene sequences shows that *R. brookii* and *C. raciborskii* strains constitute a statistically well-supported monophyletic clade (Fig. 2 and Suppl. Fig. 1). This clade comprises sequences sharing ≥ 98% of similarity and show low evolutionary rate within the clade.

Despite this, it is possible to identify some sub-clusters with a certain coherent phylogeographical distribution as was previously described [26,27]. For example, the sub-cluster comprising strains exclusively from South America (*R. brookii* D9, *C. raciborskii* MVCC14 and T3) is segregated with a well-supported statistical value (Fig. 2, Suppl. Fig. 2 and 4). Phylogenetic analyses from other phylogenetic markers also displayed the monophyletic nature among *R. brookii* and *C. raciborskii* strains (Suppl. Fig. 2 - 5). This is congruent with a previous study of phylogenetic relationships inferred from several conserved genes, which postulate that *Cylindrospermopsis* and *Raphidiopsis* representatives should be congeners [28]. However, to assess the taxonomic classification of these microorganisms further phylogenetic analyses (*e.g.*, global genome comparisons) or more complete physiological descriptions are required.

Genome sequencing information

Genome project history

Strains CS-508 and MVCC14 were selected for sequencing based on their phylogenetic relationship between strains from South America and Australia. Sequenced draft genomes were annotated using RAST [29] The CS508 Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession MBQX00000000. The version described here is MBQX01000000. MVCC14 Whole Genome Shotgun Project has been

deposited under the accession ID MBQY00000000. The version described in this paper is version MBQY01000000. A summary of the project information is shown in Table 2.

Growth conditions and genomic DNA preparation

C. raciborskii cultures were grown in MLA medium [30], under 12:12 light:dark cycles at 25 °C. Total DNA extractions were carried out using 100 mL of exponential growth culture, obtaining approximately 1 g of wet cell pellet. DNA purification was conducted by standard CTAB protocol [31]. Total cell pellets were mechanically disrupted and resuspended in 500 μL of CTAB buffer and incubated at 55 °C for 1 h under constant mixing. The DNA was purified using 500 μL phenol/chloroform/isoamyl alcohol (25:24:1) and centrifuged at 8000 x *g* for 7 min. DNA was precipitated using isopropanol/ammonium acetate (0.54 vol cold isopropanol, 0.08 vol ammonium acetate 7.5 M). Finally, DNA was washed with 70% and then with 90% ethanol and resuspended in 50 μL of pure water. DNA extraction was visualized using red gel staining in a 1% agarose gel under UV light.

Genome sequencing and assembly

Both genomes were obtained by a shotgun strategy using Illumina MiSeq sequencing technology. A total of 8,308,910 paired-end reads were obtained for CS-508 strain and 28,711,437 paired-end reads for MVCC14 strain. Quality control checks were performed on the raw FASTQ data using FastQC (version 0.10.1) [32]. Sequencing reads were trimmed for sequencing adaptors using Trimmomatic (version 0.32) [33]. The quality filtering and trimming were done by Prinseq-lite (version 0.20.4) [34]. Briefly, reads were

trimmed for 'N' characters and low-quality nucleotides (Phred score cutoff of 24) and then any read with an average Phred score below 29 and shorter than 80 nt was discarded. A *de novo* assembly strategy involving multiple algorithms and merging of the individual assemblies was performed. Assemblies by IDBA [35], SPADes [36], VELVET [37] and ABYSS [38] algorithms were generated by using the platform MIX software [39] to improve draft assembly by reducing contig fragmentation. Contigs shorter than 1000 bp were discarded. The final assembly resulted in 163 contigs for CS508 and 99 contigs for MVCC14, accounting for 3,558,956 bp and 3,594,524 bp, respectively. CheckM analysis [40] indicated genome completeness of 97.57% for CS-508 and 96.29% for MVCC14.

Genome annotation

The gene annotation process was conducted using the RAST Server 2.0 [29]. Predicted coding sequences were extracted from RAST platform, and homology was evaluated by BLASTp scan, with each predicted ORF as a query against the complete bacterial database.

Genome Properties

C. raciborskii CS-508 and MVCC14 draft genomes have a GC% content of 43 and 44 respectively (Table 3), containing 3202 and 3560 ORFs each. Table 4 shows the COG distribution of the corresponding genes. A high number of these encode metabolic proteins (COG codes R, S, M, C, E, P, O, H and T). Interestingly, no genes for the "RNA processing and modification" category were found in any genome. This has been observed in another cyanobacterial genome [41] and could be explained by a genetic divergence of

these cyanobacteria. Approximately 22% (CS-508) and 26% (MVCC14) of the total coding genes were not classified in any COG category.

Insights from the genome sequence

Photoautotrophic metabolic pathways were reconstructed in CS-508 and MVCC14 draft genomes, based on the predicted metabolic pathways in previously sequenced genomes of *C. raciborskii* [16,18]. Nitrogen metabolic systems related to ammonium, nitrate and nitrite acquisition genes, as well as heterocyst differentiation and nitrogen fixation, were identified in both genome drafts.

Sequenced genomes were compared to previously published *C. raciborskii* and *R. brookii* genomes. We determined the average nucleotide identity (ANI) in these genomes by a two-way comparison analysis (Table 5), using the inference tool ANI calculator [20]. The percentage of shared genes between strains ranged from 93.23% to 99.77%. According to the ANI value, the complete group, *C. raciborskii* and *R. brookii* could be considered as members of the same species, considering a threshold value of 95% [42].

We identified four genes encoding a non-ribosomal peptide synthetase (NRPS) complex in the CS-508 genome related to the hassallidin biosynthesis. We found in CS-508 the same gene cluster as in the hassallidin producers CS-509, CS-505 and *Anabaena* SYKE748A [10,16,18], with no evidence of mutations in the hassallidin cluster. Surprisingly, we were not able to detect the presence of hassallidin in CS-508 cultures, according to LC-MS/MS analysis (unpublished results). In the MVCC14 draft genome, we identified a group of

genes related to STX biosynthesis. STX is a paralytic biotoxin produced by marine dinoflagellates and freshwater cyanobacteria [14]. The *sxt* gene cluster found in MVCC14 has a similar distribution and toxin profile to *R. brookii* D9 [16]. We did not find NRPS sequences in the MVCC14 genome.

Conclusions

To understand the genomics of the toxin-producing, bloom-forming *C. raciborskii*, this work presents two drafts of sequenced genomes from the non-toxic Australian strain CS-508 and the Uruguayan neurotoxin-producer strain MVCC14. An NRPS gene cluster related with hassallidin production was identified in CS-508, and PKS-like set of genes related with STX production was identified in the genome of the MVCC14 strain. Considering the 16S rRNA gene phylogenetic analysis and genome level comparison, we identified phylogeographical segregation of the *C. raciborskii* and *R. brookii* strains retrieved from South America. Disregarding the metabolic differences concerning the nitrogen fixation, these results suggest *R. brookii* D9 and *C. raciborskii* MVCC14 are closely related at genome level, which could lead to new research to corroborate the *Cylindrospermopsis/Raphidiopsis* clade as one comprised by two genera or by a single genus with different species.

Authors' contributions

JJF, KS carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. DG, DPP participated in designing and performing the

bioinformatics analysis. MV, CPA conceived the study and participated in the draft of the manuscript. LB conducted the 16S rRNA gene phylogenetic analysis. All authors read and approved the final manuscript. The authors declare that they have no competing interests.

Acknowledgments

This work was financed by the following grants: Fondecyt regular 1131037, 1161232, Fondecyt de Iniciación 11130518 and JJF Ph.D. Conicyt Fellowship 21120837, CTM2013-48292-C3-1-R from the Spanish Ministry of Economy and Competitiveness; KS was financed by postdoctoral Fondecyt Nº 3130681, LB was funded by Postdoctoral Fondecyt N° 3140330. We thanked Karla del Rio for strain cultivation and DNA extraction and Dr. Sylvia Bonilla for kindly providing the MVCC14 *C. raciborskii* strain.

References

1. Carmichael WW. Health Effects of Toxin-Producing *Cyanobacteria*: "The CyanoHABs." Hum. Ecol. Risk Assess. An Int. J. 2001;7:1393–407.

2. Paerl HW, Paul VJ. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. 2012;46:1349–63.

3. Hawkins PR, Runnegar MT, Jackson AR, Falconer IR. Severe Hepatotoxicity Caused by the Tropical *Cyanobacterium* Supply Reservoir. Appl Env. Microbiol. 1985;50:1292–5.

4. Hawkins PR, Chandrasena NR, Jones GJ, Humpage AR, Falconer IR. Isolation and toxicity of *Cylindrospermopsis raciborskii* from an ornamental lake. Toxicon, 35(3), 341-346. Toxicon. 1997;35:341–6.

5. Briand J-F, Jacquet S, Bernard C, Humbert J-F. Health hazards for terrestrial vertebrates from toxic cyanobacteria in surface water ecosystems. Vet. Res. 2003;34:361–77.

6. Wołoszyńska J. Das Phytoplankton einiger javanischer Seen, mit Berücksichtigung des Sawa-Planktons. Imprimerie de l'Université; 1912.

7. Seenayya G, Raju NS. On the ecology and systematic position of the alga known as *Anabaenopsis raciborskii* (Wolosz.) Elenk. and a critical evaluation of the forms described under the genus *Anabaenopis*. Int. Symp. Taxon. Biol. Bluegreen Algae, 1st, Madras, 1970. Pap. 1972.

8. Lagos N, Onodera H, Zagatto PA, Andrinolo D, Azevedo SMF., Oshima Y. The first evidence of paralytic shellfish toxins in the freshwater cyanobacterium *Cylindrospermopsis raciborskii*, isolated from Brazil. Toxicon. 1999;37:1359–73.

9. Griffiths DJ, Saker ML. The Palm Island mystery disease 20 years on: A review of research on the cyanotoxin cylindrospermopsin. Environ. Toxicol. 2003;18:78–93.

10. Vestola J, Shishido TK, Jokela J, Fewer DP, Aitio O, Permi P, et al. Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. Proc. Natl. Acad. Sci. U. S. A. 2014;111: E1909-17.

11. Neuhof T, Schmieder P, Preussel K, Dieckmann R, Pham H, Bartl F, et al. Hassallidin A, a glycosylated lipopeptide with antifungal activity from the cyanobacterium Hassallia sp. J. Nat. Prod. 2005;68:695–700.

12. Neuhof T, Schmieder P, Seibold M, Preussel K, von Döhren H. Hassallidin B - Second antifungal member of the Hassallidin family. Bioorganic Med. Chem. Lett. 2006;16:4220–2.

13. Dahlmann J, Budakowski WR, Luckas B. Liquid chromatography-electrospray ionization-mass spectrometry-based method for the simultaneous determination of algal and cyanobacterial toxins in phytoplankton from marine waters and lakes followed by tentative structural elucidation of microcystins. J. Chromatogr. A. 2003;994:45–57.

14. Piccini C, Aubriot L, Fabre A, Amaral V, González-Piana M, Giani A, et al. Genetic and eco-physiological differences of South American *Cylindrospermopsis raciborskii* isolates support the hypothesis of multiple ecotypes. Harmful Algae. 2011;10:644–53.

15. Stucken K, Murillo AA, Soto-Liebe K, Fuentes-Valdés JJ, Méndez MA, Vásquez M. Toxicity phenotype does not correlate with phylogeny of *Cylindrospermopsis raciborskii* strains. Syst. Appl. Microbiol. 2009;32:37–48.

16. Stucken K, John U, Cembella A, Murillo A a, Soto-Liebe K, Fuentes-Valdés JJ, et al. The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications. PLoS One. 2010;5:e9235.

17. Fuenzalida L. Genetic and physiologic studies in the cyanobacterium *Cylindrospermopsis raciborskii*, a PSP-toxin producer. [Santiago]: Universidad de Chile; 2005.

18. Sinha R, Pearson LA, Davis TW, Muenchhoff J, Pratama R, Jex A, et al. Comparative genomics of *Cylindrospermopsis raciborskii* strains with differential toxicities. BMC Genomics. 2014;15:83.

19. Tatusov RL, Koonin E V., Lipman DJ. A Genomic Perspective on Protein Families. Science (80-.). 1997;278:631–7.

20. Sukenik A, Hadas O, Kaplan A, Quesada A. Invasion of *Nostocales* (cyanobacteria) to subtropical and temperate freshwater lakes - physiological, regional, and global driving forces. Front. Microbiol. 2012;3:1–9.

21. Padisák J. *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. Arch. Für Hydrobiol. Suppl. Monogr. Beitrage. 1997. p. 563–93.

22. Chonudomkul D, Yongmanitchai W, Theeragool G, Kawachi M, Kasai F, Kaya K, et al. Morphology, genetic diversity, temperature tolerance and toxicity of *Cylindrospermopsis raciborskii* (*Nostocales*, *Cyanobacteria*) strains from Thailand and Japan. FEMS Microbiol. Ecol. 2004;48:345–55.

23. Plominsky ÁM, Larsson J, Bergman B, Delherbe N, Osses I, Vásquez M. Dinitrogen fixation is restricted to the terminal heterocysts in the invasive cyanobacterium *Cylindrospermopsis raciborskii* CS-505. PLoS One. 2013;8:e51682.

24. Muñoz-García J, Ares S. Formation and maintenance of nitrogen-fixing cell patterns in filamentous cyanobacteria. Proc. Natl. Acad. Sci. 2016;113:201524383.

25. Mohamed ZA. First report of toxic *Cylindrospermopsis raciborskii* and *Raphidiopsis mediterranea* (Cyanoprokaryota) in Egyptian freshwaters. FEMS Microbiol. Ecol. 2007;59:749–61.

26. Gugger M, Molica R, Le Berre B, Dufour P, Bernard C, Humbert J-F. Genetic Diversity of *Cylindrospermopsis* Strains (*Cyanobacteria*) Isolated from Four Continents. Appl. Environ. Microbiol. 2005;71:1097–100.

27. Haande S, Rohrlack T, Ballot A, Røberg K, Skulberg R, Beck M, et al. Genetic characterization of *Cylindrospermopsis raciborskii* (*Nostocales*, *Cyanobacteria*) isolates from Africa and Europe. Harmful Algae. 2008;7:692–701.

28. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 2011;12:444.

29. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

30. Castro D, Vera D, Lagos N, García C, Vásquez M. The effect of temperature on growth and production of paralytic shellfish poisoning toxins by the cyanobacterium *Cylindrospermopsis raciborskii* C10. Toxicon. 2004;44:483–9.

31. Richards E, Reichardt M, Rogers S. Preparation of Genomic DNA from Plant Tissue. Curr. Protoc. Mol. Biol. John Wiley & Sons, Inc.; 2001. p. I:2.3:2.3.1–2.3.7.

32. Andrews S. FastQC: A quality control tool for high throughput sequence data [Internet]. Babraham Bioinforma. 2010 [cited 2016 Dec 2]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

33. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

34. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.

35. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA - A practical iterative De Bruijn graph *De Novo* assembler. The 14th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2010). 2010;6044:426–40.

36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 2012;19:455–77.

37. Zerbino DR, Birney E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.

38. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. Genome Res. 2009;19:1117–23.

39. Soueidan H, Maurier F, Groppi A, Sirand-Pugnet P, Tardy F, Citti C, et al. Finishing bacterial genome assemblies with Mix. BMC Bioinformatics. 2013;14: S16.

40. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes 5. Genome Res. 2015;25:1043–55.

41. Cheevadhanarak S, Paithoonrangsarid K, Prommeenate P, Kaewngam A, Musigkain A, Tragoonrung S, et al. Draft genome sequence of *Arthrospira platensis* C1 (PCC9438). Stand. Genomics Sci. 2012;6:43–53.

42. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 2007;57:81–91.

43. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequences (MIGS) specification. Nat Biotechnol. 2008;26:541–7.

44. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. 1990;87:4576–9.

45. Caumette P, Brochier-Armanet C, Normand P. Taxonomy and Phylogeny of Prokaryotes. Environ. Microbiol. Fundam. Appl. Dordrecht: Springer Netherlands; 2015. p. 145–90.

46. Meeks JC, Elhai J. Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. Microbiol. Mol. Biol. Rev. 2002;66:94–121; table of contents.

47. Saker ML, Griffiths DJ. The effect of temperature on growth and cylindrospermopsin content of seven isolates of *Cylindrospermopsis raciborskii* (*Nostocales*, Cyanophyceae) from water bodies in northern Australia. Phycologia. 2000;39:349–54.

48. Moisander PH, McClinton E, Paerl HW. Salinity effects on growth, photosynthetic parameters, and nitrogenase activity in estuarine planktonic cyanobacteria. Microb. Ecol. 2002;43:432–42.

49. Saker ML. Cyanobacterial blooms in tropical north Queensland water bodies. [Townsville, Australia]: James Cook University; 2000.

50. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat. Genet. 2000;25:25–9.

51. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

52. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 2010;59:307–21.

53. Castenholz RW. General characteristics of the cyanobacteria. In: Boone DR, Castenholz RW, editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer; 2001:474–87.

54. Saker ML, Neilan BA. Varied diazotrophies, morphologies, and toxicities of genetically similar isolates of *Cylindrospermopsis raciborskii* (Nostocales, Cyanophyceae) from northern Australia Appl. Environ. Microbiol., 2001;67:1839–1845.

55. Vidal L, Kruk C. *Cylindrospermopsis raciborskii*(Cyanobacteria) extends its distribution to Latitude 34°53′S: taxonomical and ecological features in Uruguayan eutrophic lakes PANAMJAS, 2008;3:142–151.

Table 1

Table 1a. Classification and general features of *C. raciborskii* strain CS-508 according to MIGS designation [43]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [44]
		Phylum *Cyanobacteria*	TAS [53]
		Class *Cyanophyceae*	
		Order *Nostocales*	
		Family *Aphanizomenonacea*	
		Genus *Cylindrospermopsis*	TAS [6]
		Species *Cylindrospermopsis raciborskii*	
		Strains: CS508	TAS [54]
	Gram stain	*Negative*	TAS [7]
	Cell shape	*Filaments*	
	Motility	*Temporary-motile (Hormogonia)*	
	Sporulation	*None*	TAS [45]

	Temperature range	*Mesophile*	TAS [6]
	Optimum temperature	*25°C*	TAS [46]
	pH range; Optimum	*pH 7.50-9.21; pH 8.33*	
	Carbon source	*Autotroph*	TAS [21]
MIGS-6	Habitat	*Freshwater*	TAS [47]
MIGS-6.3	Salinity	*0.4% NaCl (maximum)*	IDA
MIGS-22	Oxygen requirement	*Aerobic*	NAS
MIGS-15	Biotic relationship	*free-living*	NAS
MIGS-14	Pathogenicity	*non-pathogen*	TAS [48]
MIGS-4	Geographic location	*Isolated Solomon Dam, Australia*	NAS
MIGS-5	Sample collection	*1999*	NAS
MIGS-4.1	Latitude	*-18.7241*	IDA
MIGS-4.2	Longitude	*146.5938*	TAS [49]
MIGS-4.4	Altitude	*Unknown*	TAS [49]

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not

directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [50]

Table 1b. Classification and general features of *C. raciborskii* strain MVCC14 according to MIGS designation [43]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [44]
		Phylum *Cyanobacteria*	TAS [53]
		Class *Cyanophyceae*	
		Order *Nostocales*	
		Family *Aphanizomenonacea*	
		Genus *Cylindrospermopsis*	TAS [6]
		Species *Cylindrospermopsis raciborskii*	
		Strains: MVCC14	TAS [55]
	Gram stain	*Negative*	

	Cell shape	*Filaments*	TAS [7]
	Motility	*Non-motile*	
	Sporulation	*None*	TAS [45]
	Temperature range	*Mesophile*	TAS [6]
	Optimum temperature	*25°C*	TAS [46]
	pH range; Optimum	*pH 7.50-9.21; pH 8.33*	
	Carbon source	*Autotroph*	TAS [21]
MIGS-6	Habitat	*Freshwater*	TAS [47]
MIGS-6.3	Salinity	*0.4% NaCl (maximum)*	IDA
MIGS-22	Oxygen requirement	*Aerobic*	NAS
MIGS-15	Biotic relationship	*free-living*	NAS
MIGS-14	Pathogenicity	*Saxitoxin (STX)*	TAS [48]
MIGS-4	Geographic location	*Isolated Laguna Blanca, Uruguay*	NAS
MIGS-5	Sample collection	*Unknown*	NAS
MIGS-4.1	Latitude	*-34.8984*	TAS [14]
MIGS-4.2	Longitude	*-54.8369*	TAS [14]

MIGS-4.4	Altitude	*Unknown*	NAS

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [50]

Table 2

Project Information. Authors must provide the requested data, in conformance with the MIGS standard. Supply the information for column 3. Do not modify the order of columns or rows.

MIGS ID	Property	Term (for CS508)	Term (for MVCC14)
MIGS 31	Finishing quality	High- Quality Draft	High- Quality Draft
MIGS-28	Libraries used	Illumina	Illumina
MIGS 29	Sequencing platforms	Illumina HiSeq2000	Illumina HiSeq2000
MIGS 31.2	Fold coverage	20x	20x
MIGS 30	Assemblers	IDBA, SPADES, VELVET, and ABYSS	IDBA, SPADES, VELVET, and ABYSS
MIGS 32	Gene calling method	Rast	Rast
	Locus Tag	CYL_CS508	CYL_MVCC14
	GenBank ID	MBQX00000000	MBQY00000000
	GenBank Date of Release	November 01, 2016	November 01, 2016
	GOLD ID	Gs0120410	Gs0121371
	BIOPROJECT	PRJNA327084	PRJNA327088

	Source Material Identifier	Freshwater	Freshwater
MIGS 13	Project relevance	Environment	Environment

Table 3 Genome statistics of CS-508 (A) and MVCC14 (B).

Attribute	A Value	A % of Total	B Value	B % of Total
Genome size (bp)	3,558,956	100	3,594,524	100
DNA coding (bp)	3,039,246	85.34	3,074,946	85.55
DNA G+C (bp)	1,530,351	43	1,581,591	44
DNA scaffolds	163	100	99	100
Total genes	3,344	100	3,616	100
Protein coding genes	3302	98.74	3560	98.45
RNA genes	42	1.26	56	1.55
Pseudo genes	-	-	-	-
Genes in internal clusters	-	-	-	-
Genes with function prediction	2,247	67.19	2,337	64.63
Genes assigned to COGs	1,747	56.16	1,796	55.55
Genes with Pfam domains	2,656	79.43	2,800	77.43
Genes with signal peptides	71	2.12	63	1.74
Genes with transmembrane helices	255	7.63	748	20.66

				80
CRISPR repeats	7	-	9	-

Table 4. Number of genes associated with general COG functional categories.

	CS508		MVCC14		
Code	Value	%age	Value	%age	Description
J	142	4.56	143	4.37	Translation, ribosomal structure and biogenesis
A	0	0.00	0	0.00	RNA processing and modification
K	69	2.22	64	1.96	Transcription
L	88	2.83	112	3.43	Replication, recombination and repair
B	0	0.00	0	0.00	Chromatin structure and dynamics
D	21	0.68	19	0.58	Cell cycle control, Cell division, chromosome partitioning
V	0	0.00	0	0.00	Defense mechanisms
T	49	1.58	54	1.65	Signal transduction mechanisms
M	123	3.95	130	3.98	Cell wall/membrane biogenesis
N	6	0.19	5	0.15	Cell motility
U	0	0.00	0	0.00	Intracellular trafficking and secretion
O	111	3.57	111	3.40	Posttranslational modification, protein

					turnover, chaperones
C	157	5.05	163	4.99	Energy production and conversion
G	99	3.18	93	2.84	Carbohydrate transport and metabolism
E	125	4.02	123	3.76	Amino acid transport and metabolism
F	45	1.45	44	1.35	Nucleotide transport and metabolism
H	104	3.34	107	3.27	Coenzyme transport and metabolism
I	32	1.03	31	0.95	Lipid transport and metabolism
P	128	4.11	130	3.98	Inorganic ion transport and metabolism
Q	40	1.29	36	1.10	Secondary metabolites biosynthesis, transport, and catabolism
R	252	8.10	262	8.01	General function prediction only
S	156	5.01	169	5.17	Function unknown
-	1364	43.84	1473	45.06	Not in COGs

The total is based on the total number of protein-coding genes in the genome.

Table 5. Average nucleotide identity (ANI) values for the sequenced *C. raciborskii* and *Raphidiopsis brookii* strains

	Cylindrosper mopsis CS-505	*Cylindrosper mopsis* CS-506	*Cylindrosper mopsis* CS-508	*Cylindrospe rmopsis* CS-509	*Cylindrospe rmopsis* MVCC14	*R. brookii* D9
Cylindrosper mopsis CS-505	—	99.31	99.73	99.77	93.73	93.26
Cylindrosper mopsis CS-506	99.31	—	99.39	99.32	93.45	92.85
Cylindrosper mopsis CS-508	99.73	99.39	—	99.76	93.82	93.25
Cylindrosper mopsis CS-509	99.77	99.32	99.76	—	93.80	93.23
Cylindrosper mopsis MVCC14	93.73	93.45	93.82	93.80	—	97.17
R. brookii D9	93.26	92.85	93.25	93.23	97.17	—

Figure 1. **Inverted light photomicrographs of source organisms.** A-B C. raciborskii CS-508 and of C-E C. raciborskii MVCC14.

Figure 2. Maximum likelihood (ML) phylogenetic tree indicating the phylogenetic relationship of *C. raciborskii* strains. The ML tree is based on 16S rRNA gene sequences from *C. raciborskii* strains CS−508 and MVCC14 and sequences retrieved from previous reports stored in the NCBI database. These sequences were aligned using MUSCLE [51], and the phylogenetic tree was constructed with the phyML using GTR substitution model and BEST option for searching the starting tree [52]. Bootstrap support values ≥50% are indicated from 1000 bootstrap replicates. In supplemental material, a complete phylogenetic tree is reported (Suppl. Fig. 1).

CHAPTER III

Characterization of the antifungal Hassallidin in freshwater cyanobacterium *Cylindrospermopsis raciborskii.*

Short communication Toxicon

Characterization of the antifungal Hassallidin in freshwater cyanobacterium *Cylindrospermopsis raciborskii.*

Juan J. Fuentes-Valdés[a,b], Katia Soto[b], Bernd Krock[c], Daniel Garrido[a] and Mónica Vásquez [d*]

[a]Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

[b]Facultad de Ciencias de la Salud, Universidad San Sebastián, Campus Los Leones, Santiago 7510157, Chile

[c]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

[d]Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

*Corresponding author: Mónica Vásquez, Pontificia Universidad Católica de Chile, Alameda 340, Santiago, Chile, Tel.: +56 26862847, Fax: +562 6862185, E-mail: mvasquez@bio.puc.cl

Keywords: *Cylindrospermopsis raciborskii*; non-ribosomal peptide synthetases; cyclic peptides; liquid chromatography coupled to tandem mass spectrometry

Cyanobacteria are photosynthetic microorganisms that under certain circumstances cause harmful algal blooms (HABs). This phenomenon is associated with diverse physical, chemical and biotic factors, which in appropriate combination can persistently affect the environment and human activities (Pearl, 1988). Some HAB-associated cyanobacteria can produce multiple secondary metabolites, including alkaloids, polyketides, and non-ribosomal peptides. Some of these molecules have been used as bioactive compounds (Neilan et al., 1999; Namikoshi and Rinehart, 1996). The non-ribosomal peptide synthetase (NRPS) pathway is responsible for the biosynthesis of several bioactive compounds (Marahiel et al., 1997; Sieber and Marahiel, 2005). The antifungal/antimicrobial hassallidin (HAS) is an aliphatic glycosylated cyclic peptide, synthesized via NRPS enzymes. HAS was first identified in *Hassallia* sp. (Neuhof et al., 2005). This cyanobacterium produces the variants HAS-A and HAS-B, with an estimated molecular weight of 1200 and 1400 daltons, respectively (Neuhof et al., 2005; Neuhof et al., 2006). HAS-A and HAS-B have antimicrobial activity, for example affecting the cell membrane integrity of yeast *Candida albicans* (Neuhof et al., 2005). Recently, the structural variants HAS-C and HAS-D were described in cultures of freshwater cyanobacterium *Anabaena* sp. SYKE748A and the gene cluster *has* was determined to be responsible for their biosynthesis (Vestola et al., 2014).

Antimicrobial activity was variable and dependent on the cyanobacterial strain tested, which may reflect different HAS production profiles. Other HAS molecules have been detected in diverse cyanobacterial freshwater ecosystems (Vestola et al., 2014) and also structural variants of HAS classified as balticidins A-D, are produced by marine cyanobacterium *Anabaena cylindrica* Bio33 (isolated from Baltic Sea) (Bui et al., 2014).

Cylindrospermopsis raciborskii is a freshwater filamentous nitrogen-fixing cyanobacterium (Wołoszyńska, 1912). The bacterium is globally distributed (Vasconcelos et al., 2015) and responsible for toxic freshwater eutrophication (Vasconcelos et al., 2015). *C. raciborskii* is characterized by the presence of terminal heterocysts, a type of cells specialized in nitrogen fixation (Reddy and Talpasayi, 1974). Some strains produce the hepatotoxin cylindrospermopsin (CYN) (Hawkins et al., 1985) or the potent neurotoxin saxitoxin (STX) (Lagos et al., 1999). The CYN-producing strain CS-505 and the non-toxic strains CS-508 and CS-509 were isolated in 1999 from Solomon Dam, North Queensland, Australia (Saker, 2000) and subsequently the genome of strain CS-505 was published in 2010 (Stucken et al., 2010). In addition to the presence of a *cyr* cluster involved in CYN biosynthesis, C. *raciborskii* CS-505 contains a second NRPS gene cluster (Stucken et al., 2010). This cluster is closely similar to the *has* cluster described in *Anabaena* (Sinha et al., 2014; Vestola et al., 2014). These genes were also identified in the non-toxic strain CS-509 (Sinha et al., 2014).

In this study we characterized HAS biosynthesis in *C. raciborskii* CS-505, CS-508 and CS-509, using PCR amplification and genome analysis for the *has* gene cluster. We used LC-

MS/MS to identify HAS, and evaluated the antifungal effect on *C. albicans* to determine HAS bioactivity. PCR primers are shown in Table 1A. Thermal cycling was carried out using a MyCycler Bio-Rad. Chemical analyses were performed on a hybrid triple quadrupole-linear ion trap mass spectrometer (API 4000 QTrap, AB Sciex, Darmstadt, Germany) equipped with a Turbo Ion Spray interface, coupled to a liquid chromatograph (model 1100, Agilent, Waldbronn, Germany). Long Range PCR reagents and cloning kits were supplied by QIAGEN. Chromatographic commercial hassallidin B was supplied by Enzo Life Science, solid phase extraction cartridges (LC-8 SPE) by Supelco, mobile phase for chromatography analysis and cell extractions by Merck, Antimycotic and solid assay Mueller-Hinton media were from Linsan SA. All inhibition assays were performed in triplicate.

The *has* gene cluster was identified in strain CS-505, containing the four NRPS genes (*hasN*, *hasO*, *hasV* and *hasY).* Also, 13 genes present in the *has* cluster of *Anabaena* sp. 90 and *Anabaena* sp. SYKE748A were found in CS-505 (Vestola et al., 2014) (Figure 1). Amino acid identity between *has* clusters in SYKE748A and CS-505 was in average 76.5%. Absent genes in CS-505, were *hasA*, segment *hasG-K*, segment *hasP-S*, *hasW*, *hasX*, and *hasZ* (Table 1). Using the CS-505 *has* gene cluster, we designed primers to detect and obtain nucleotide sequence of the corresponding cluster in strains CS-508 and CS-509. The complete *C. raciborskii* CS-508 gene cluster shared over 99% similarity with CS-509 and only a 50% with the global *has* gene cluster described in *Anabaena* sp. 90.

Gene	%Identity *Anabaena* sp. Syke748	%Covery	E-value
hasB	76	88	0
hasC	81	98	0
hasD	79	99	0
hasE	75	83	$5*10^{-77}$
hasF	75	95	0
hasL	79	99	0
hasM	74	77	$1*10^{-72}$
hasN	69	99	0
hasO	74	99	0
hasT	83	95	0
hasU	80	88	$3*10^{-43}$
hasV	75	98	0
hasY	75	94	0

We later determined HAS production in cell extracts of *C. raciborskii* CS-505 CS-508 and CS-509 using LC-MS/MS in the full scan mode . This showed an intense ion peak derived with an HAS structural variant of 1543.7 Da (HAS 1444.7) (m/z 1443.7 [M + H]+) in CS-505 (Fig. 2) and 1411.7 Da (HAS 1411.7) (m/z 1410.7 [M + H]+) structural variants in CS-509 extract, respectively. When we analyzed extracts of strain CS-508, it was not possible to discriminate any ion fingerprint related with HAS retention time or fragmented ionization pattern. The molecular variant identified in CS-505 extracts matches with the

previously described Hassallidin profile (Vestola et al., 2014). This molecule contains nine amino acids in the peptide core (Thr, Tyr, dehydrobutyrine (Dhb), Gln, Gly, and N-MeThr residues). However, CS-505 HAS had a final Thr instead of Gln, as identified in CS-509 structure. CS-505 HAS also had a longer fatty acid chain (C18) compared against CS-509 (C17) or the hassallidin B standard (C16). Ion residues of 132 (M2) and 162 (M3) *m/z* corresponding to a pentose and hexose, respectively, were identified in the aglyconic lipopeptide molecule in CS-505. In CS-509 extracts we only detected the M3 residue, standard fragmentation, including the methylated hexose presenting an increasing glycosylation size of 180 Da fragment.

Finally, we evaluated the antifungal activity of the HAS variants. A culture of *C. albicans* was spread in Mueller-Hinton agar plates containing cyanobacterial extracts. After 24 hours incubation at 37 °C, we observed an important reduction in the growth of the yeast in the presence of CS-509 extracts (Fig. 2B) (n = 3). CS-508 extracts did not cause any inhibition on *C. albicans*. These results are consistent with the absence of HAS in CS-508 and indicate that the strain does not produce HAS under culture conditions applied in the study. The four NRPS genes and the Acyl-Carrier Protein (ACP) gene were amplified through Targeted Reverse Transcription-PCR from RNA from CS-508 and CS-509 strains (fragments between 200 to 300 bp; data not shown), suggesting gene expression of the *has* gene cluster in both isolates.

In conclusion, in this work we studied the *has* gene cluster in the non-toxic cyanobacteria *C. raciborskii* CS-508 and CS-509, and also elucidated by LC-MS/MS the

molecular structure of the secondary metabolite hassallidin produced by CS-509. We also showed that this molecule shows antimicrobial activity against *C. albicans* cells. Although we identified the complete gene cluster in strains CS-505, CS-508 and CS-509, the analytical detection of HAS, was positive only in crude extracts of CS-505 and CS-509. Interestingly, the HAS molecules produced in these *C. raciborskii* strains presented clear differences in their structures and were different from *Anabaena* strains (Vestola et al., 2014).

Acknowledgments

This work was supported by Grant Fondecyt Regular 1131037, 1161232. We also thank Paulina Merino for providing the opportunistic yeast *Candida albicans*.

Conflict of interest

None.

References

1. Paerl, H. W. 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnol. Oceanogr. 33:823-847.

2. Neilan, B. A., Dittmann, E., Rouhiainen, L., Bass, R. A., Schaub, V., Sivonen, K

and T. Börner. 1999. Nonribosomal Peptide Synthesis and Toxigenicity of Cyanobacteria. J. Bacteriol. 181:4089–4097.

3. Namikoshi, M. and K.L. Rinehart. 1996. Bioactive compounds produced by cyanobacteria. J. Indus. Microbiol. 17:373-384.

4. Sieber S. A. and M. A. Marahiel. 2005. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. Chem. Rev. 105: 715-738.

5. Marahiel M. A., Stachelhaus T. and. H. D. Mootz. 1997. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. Chem. Rev. 97: 2651-2673.

6. Woloszynska, J., 1912. Das Phytoplankton einiger Javanian Seen mit Berücksichtigung des Sawa-Planktons. Bull. Int. Acad. Sci. Cracoviae Ser. B 6, 649–709.

7. Saker, M. L., 2000. Cyanobacterial blooms in tropical north Queensland water bodies. Ph.D. Thesis, James Cook University, Townsville, Australia.

8. Antunes, J., Leão, P.N., Vasconcelos, V.M., 2015. *Cylindrospermopsis raciborskii*: review of the distribution, phylogeography, and ecophysiology of a global invasive species. Front Microbiol. 6, 473.

9. Stucken, K., J. Uwe, A. Cembella, A. A. Murillo, K. Soto-Liebe, J. J. Fuentes-

Valdés, M. Friedel, A. M. Plominsky, M. Vásquez, and G. Glöckner. 2010. The Smallest Known Genomes of Multicellular and Toxic Cyanobacteria: Comparison, Minimal Gene Sets for Linked Traits and the Evolutionary Implications. PLoS ONE 5(2):e9235.

10. Vestola, J., Shishido, T.K., Jokela, J., Fewer D.P., Aitio, O., Permi, P., Wahlsten, M., Wang, H., Rouhiainen, L., Sivonen K., 2014. Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. Proc Natl Acad Sci U S A. 111(18), E1909-17.

11. Neuhof, T., Schmieder, P., Preussel K., Dieckmann, R., Pham, H., Bartl F., von Döhren, H., 2005. Hassallidin A, a glycosylated lipopeptide with antifungal activity from the cyanobacterium *Hassallia* sp. J Nat Prod. 68(5), 695–700.

12. Neuhof, T., Schmieder, P., Seibold, M., Preussel, K., von Döhren, H., 2006. Hassallidin B Second antifungal member of the Hassallidin family. Bioorg Med Chem Lett. 16(16), 4220–4222.

13. Bui, T.H., Wray, V., Nimtz, M., Fossen, T., Preisitsch, M., Schröder, G., Wende, K., Heiden, S.E., Mundt, S., 2014. Balticidins A-D, antifungal hassallidin-like lipopeptides from the Baltic Sea cyanobacterium *Anabaena cylindrica* Bio33. J Nat Prod. 77(6), 1287-96.

14. Reddy, P.M., Talpasayi, E.R.S., 1974. Heterocyst formation in a blue-green alga, *Cylindrospermum*. Nature. 249, 493-4.

15. Sinha R, Pearson LA, Davis TW, Muenchhoff J, Pratama R, Jex A, Burford MA, Neilan BA. Comparative genomics of Cylindrospermopsis raciborskii strains with differential toxicities. BMC Genomics. 2014 Jan 29;15:83.

LEGENDS TO FIGURES

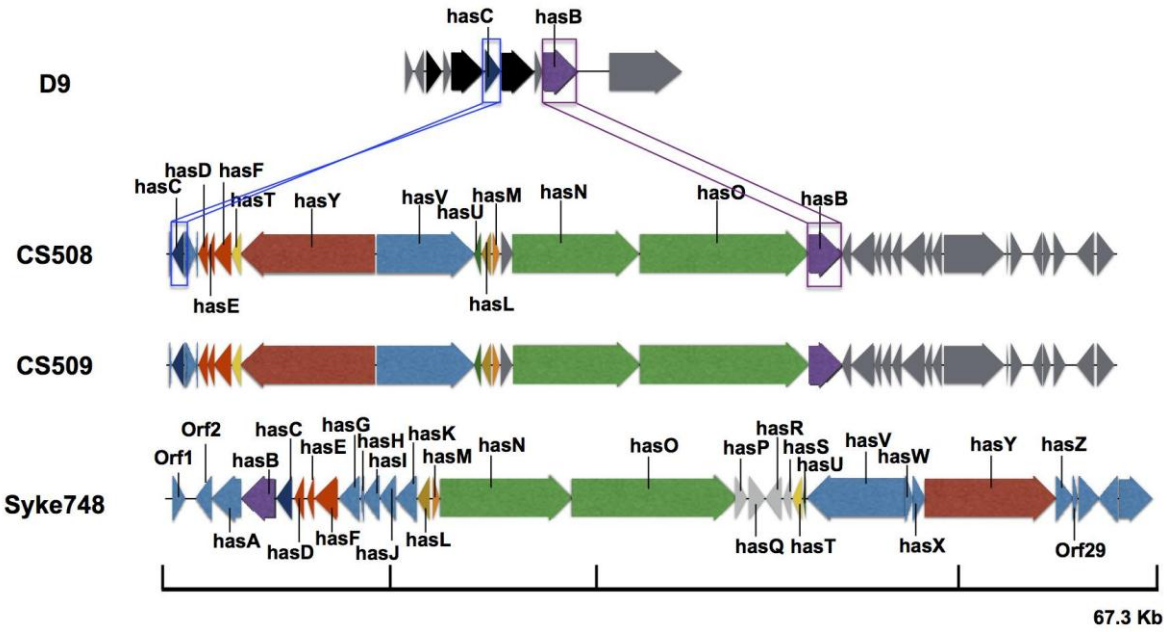

Fig. 1 Structure and genetic organization of the 37.6 kb HAS biosynthetic gene cluster from *C. raciborskii*, strain CS-505, Cs-508, CS-509 *Anabaena* sp. Syke748

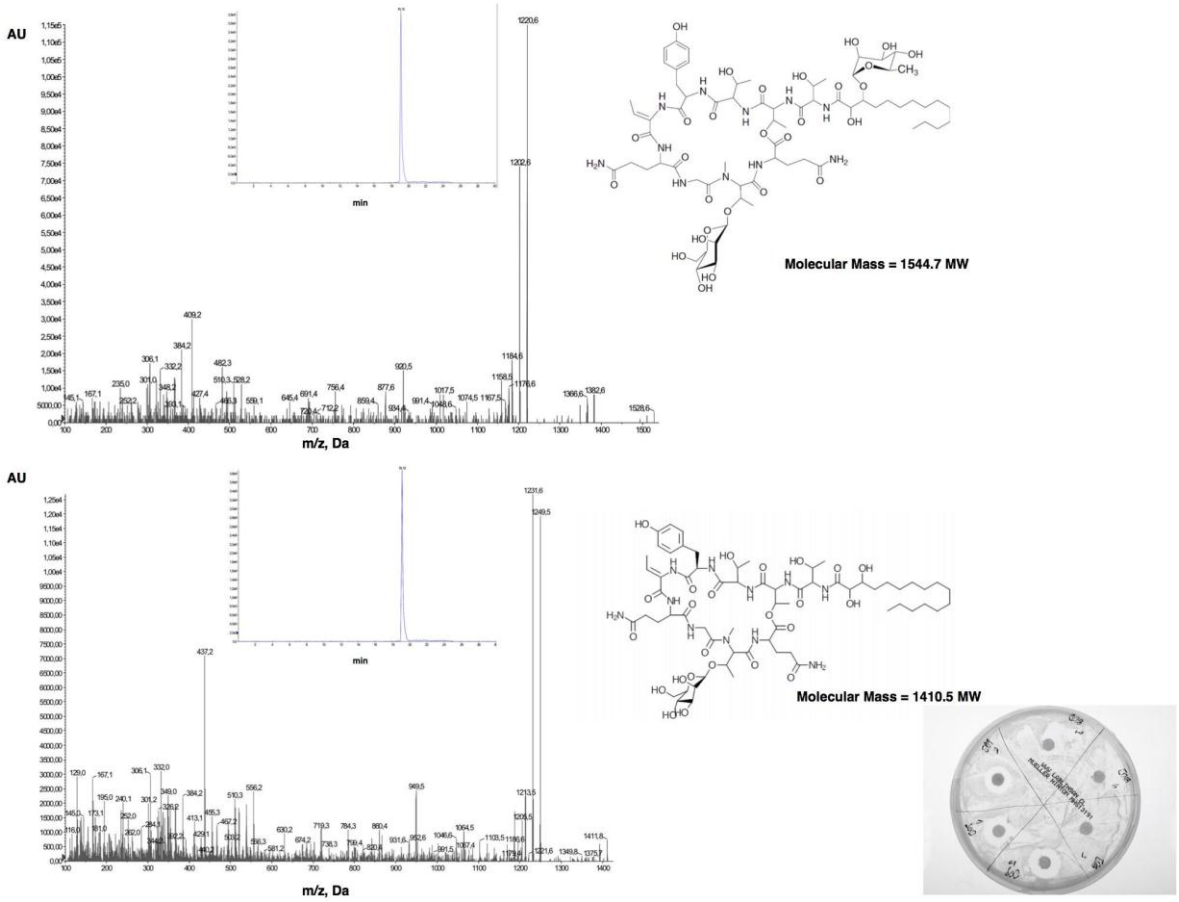

Fig. 2 Fragmentation mass ionization patron of standard and *C. raciborskii* samples

CHAPTER IV

Genetic context of *sxt* genes in producing cyanobacteria reveals evolutionary adaptations for their maintenance and spread

Genetic context of *sxt* genes in producing cyanobacteria reveals evolutionary adaptations for their maintenance and spread

(Molecular Genetics and Metabolism Journal, Elsevier)

Juan J. Fuentes-Valdés[1*], Katia Soto-Liebe[2*], Bernd Krock[3*], Nicole Trefault[4*], Daniella Spooner-Lagos[2*], Mónica Vásquez[2§]

[1] Departamento de ingeniería Química y Bioprocesos, Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Santiago, Chile

[2] Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

[3] Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

[4] Facultad de Ciencias, Universidad Mayor, Santiago, Chile

*These authors contributed equally to this work

[§] Corresponding author

Email addresses:

JJFV: jsfuente@uc.cl

KSL: kasoto@uc.cl

BK: Bernd.Krock@awi.de

NT:

DS:

MV: mvasquez@bio.puc.cl

Natural compounds are produced by a large number of microorganisms. An example of these is the Paralytic Shellfish Toxins (PSTs) also known as saxitoxins (STXs) (9). STXs are a group of alkaloid neurotoxins which block voltage-dependent sodium channels (15), this toxin encompasses a group of more than 30 analogous molecules depending on it structure substitution. There have been identified in eukaryotic marine water dinoflagellates and prokaryotic freshwater cyanobacteria (18). Despite their wide range of distribution, to date, only five cyanobacterium *sxt* gene clusters encoding for PST toxin, have been characterized in the cyanobacteria: *Aphanizomenon* sp. NH-5 (2), *Anabaena circinalis* AWQC131C (2), *Cylindrospermopsis raciborskii* T3 (3), *Lyngbya wollei* (25) and *Raphidiopsis brookii* D9 (5). Recently, fragments of *sxtA* gene have been identified in the nuclear genome of *Alexandrium fundyense* CCMP1719 and *A. minutum* CCMP113 (26). However, it is still not possible to correlate the presence of a gene cluster with the ability to produce Saxitoxins.

So far, based on the genetic composition of the gene clusters related to the production of PST in different organisms, these could be considered as a genomic mosaic. These gene clusters are composed by elements that have an apparent mobility transposition mediated gene segments of varying sizes, resulting in having different sizes rearrangements (2, 19) which do not show an apparent relation to the taxonomic distribution present in these cyanobacterial species (1). Unfortunately, it is unclear how these genes have evolved to make up the gene cluster that currently exists, some proposals that give rise to different scenarios, however, which very little information is available to support them. Something similar happens when compared the gene flanking regions of the SXT gene cluster.

In all genetic groups related to the production of toxins, we can see a different gene context. In the case of *Cylindrospermopsis* and *Raphidiopsis* genera, despite having morphogenetic similarities, the genetic composition of gene grouping and its context is relatively different. *R. brookii* D9 has at its 5' end region encoding to a protein of unknown function that is different from the genes identified at same coding region in the *C. raciborskii* T3, and something similar happens in the 3' end region where D9 presents a glmU gene, while T3 has a coding region assigned as DUF58.

Cyanobacterial blooms occur in different kind of environments, one of the species responsible for the eutrophication process in rivers and lakes is *Cylindrospermopsis raciborskii* a freshwater cyanobacterium able to fix atmospheric nitrogen. *C. raciborskii*, now regarding toxicity is characterized by having representatives responsible for the biosynthesis of STX and others Cylindrospermopsin (CYN) a hepatotoxin (27). *C. raciborskii* MVCC14 was isolated from Laguna Blanca (Uruguay). The SXT toxin profile was provided by direct competitive Enzyme-Linked Immuno Sorbent Assay (ELISA) and High-Performance Liquid Chromatography (HPLC) (STX, GTX2, and GTX3) (28).

On the other hand, *R. brookii* D9, (Brazil) corresponds to a species that cannot fix atmospheric nitrogen, nevertheless, in terms of 16S DNA molecular marker (over a 99% nucleotide identity), is phylogenetically close to *C. raciborskii* and also present the same toxin profile described for MVCC14 (28). The toxin profiles are correlated with their corresponding gene clusters (1, 5) and also associated with specific flanking neighboring genes.

In the present study we characterized the putative PST gene cluster (*sxt* genes) from *C. raciborskii* MVCC14, able to produce several PSTs analogs and we develop a liquid chromatography-tandem mass spectrometry (LC-MS/MS) for corroboration of the corresponding toxin profile. This information was used to perform a comparative phylogeny analysis based on the concatenation of multiple PST core genes described in all the PST producers and their corresponding 16S rRNA genes. We focus on the *sxt* flanking genes of all known PST cluster found in cyanobacteria, and also present evidence for mobility inside or outside the *sx*t clusters in the available genomes of the PST-producing cyanobacteria.

Materials and Methods

Morphology and cyanobacterial culturing conditions.

C. raciborskii MVCC14 graciously provided by S. Bonilla (Instituto de Investigaciones Biologicas Clemente Estable, Montevideo, Uruguay). Strain was cultured in MLA 1X media (30). The temperature in the culture chamber was 25ºC, under continuous cool-white fluorescence light at a photon flux density of 40 µE s^{-1} m^{-2}.

The Morphological studies were conducted using an Olympus BX51 light microscope and color view imaging system (Olympus, Germany).

Toxin extraction and (LC-MS/MS)

Cyanobacteria were harvested by centrifugation at 16,000 x g for 15 min to yield a cell pellet. The cell pellets were dry in speed-vac (DNA Speed Vac, SAVANT mod. DNA 110-

230) and were extracted in 300 µl of 0.05 M acetic acid. Samples were disrupted with ultrasonic cell disruptor (Microson XL ultrasonic cell disruptor, Misonix, Inc. Farmingdale) for 3 times (20 seconds for each one). The extracts were then centrifuged at 5000 x g for 10 min, filtered through a 0.45 µm membrane filter and stored at –20°C until analysis.

Mass spectrometric analyses for PSP toxins were performed according to a hydrophilic interaction liquid ion-chromatography (HILIC) method (29) with slight modifications described in Soto, et al. (1).

DNA isolation, amplification, and sequencing

In order to characterize the *sxt* gene cluster from *C. raciborskii* MVCC14, we extracted DNA according to the CTAB method (17) and PCR amplification of a complete *sxt* genes cluster and 16S rRNA using the primers shown in Table 1S. DNA fragments between ~1-4 kb were cloned, amplified and sequenced using the universal primers M13F-pUC and M13R-pUC through the Macrogen Service (Korea). Published PST gene cluster and 16S rRNA gene sequences, were obtained from the National Center for Biotechnology Information (NCBI) databases (http://www.ncbi.nlm.nih.gov/): *C. raciborskii* T3 16S rRNA gene (EU439566); *C. raciborskii* T3 *sxt* gene cluster, (DQ787200); *R. brookii* D9 16S rRNA (EU552070); *R. brookii* D9 *sxt* gene cluster, complete genome sequence (ACYB00000000); *Lyngbya wollei* 16S rRNA gene (EU603708), Lyngbya wollei *sxt* gene cluster, (EU603711); *Aphanizomenon* sp. NH-5 16S rRNA gene (AF425995); *Aphanizomenon* sp. NH-5 *sxt* gene cluster, (EU603710); *Anabaena circinalis* AWQC131C

16S rRNA gene (AF247589); *Anabaena circinalis* AWQC131C *sxt* gene cluste*r*, (DQ787201). Nucleotide sequences derived from this work were submitted to GenBank under accession number: C. raciborskii MVCC14 16S rRNA gene (XXX); and C. raciborskii MVCC14 sxt gene cluster (XXX).

Sequence, phylogeny and selection analysis

Similarity analysis of conserved sequences of the products related to Pst gene cluster was performed by comparison with homologous sequences present in the GeneBank database by BLASTN, National Center for Biotechnology Information (NCBI). The assembly and identification of open reading frames (ORF) were performed with the software Vector NTI software (Invitrogen, California, USA). The insertion sequence element analysis was performed using the online software IS Finder (https://www-is.biotoul.fr/) (36). For the phylogenetic correlation inference, conserved sequences present in the sxt gene clusters were used (that share identities between 42% in the less similar and 90% for the gene most conserve), plus the 16S rRNA gene sequences (496 bp) was used, we use only the 16S rRNA because even no more information available in the database. 15 sxt gene sequences belonging to the cluster (sxtA to sxtO) in total 14635 bp, shared by the five cyanobacterial species so far described (see single phylogeny in the supplementary material) were used. The 15 genes were concatenated sxt multiple sequence analysis were performed using the BioEdit program, alignments were manually strengthened, and the phylogeny was built using the NJ algorithm, performing 1000 bootstrap replicates for statistical support. Codon selection analysis from *sxt* shared genes was submitted into the phylogenetic analysis tools

(http://www.datamonkey.org/) (31). The selection of the phylogenetic model was using the automatic selection models tool. Detection of possible recombination sites was performed by the GARD tool (32) and the recombination points identified were evaluated for topological incongruence by the test of Kishino Hasegawa (KH) (33). Codon selection was assessed by the SLAC, FEL (34) and PARRIS (35) tools, assuming a significance level of 0.05.

Results

Morphological and toxin analysis of *C. raciborskii* MVCC14

C. raciborskii MVCC14 is a filamentous, freshwater and diazotrophic cyanobacteria (Fig. 1A). PST profile of MVCC14 was analyzed by LC-MS/MS (1), detecting the presence of GTX2/3, dcGTX2/3, dcSTX and STX analogs (Fig. 1B). The PST profiles obtained were similar to the *R. brookii* D9 (13) and MVCC14 profiles previous described (Puccini 2011). So as to corroborate the toxin profile from *C. raciborskii* MVCC14 with its corresponding *sxt* gene cluster, we sequence a 24.9 kb gene cluster of MVCC14 (Fig. 1C), encompasses 25 open reading frames (ORF) (Fig. 1C). This cluster shares a 99.8% of nucleotide identity and gene distribution with the identified *sxt* cluster in *R. brookii* D9 (*sxtD* to *sxtO*), which was proposed as the minimal core of genes necessary for GTX2/3, dcGTX2/3, STX, and dcSTX production (1), the same analogues synthesized by the *C. raciborskii* MVCC14. From the entire PST gene cluster described so far, this is the first toxin cluster with the same size and composition described in a different species. The presence of the enzyme O-

sulfotransferase codified by *sxtSUL* gene, correlated with the presence of the analogous GTX2/3 (1).

Characterization of the *sxt* gene cluster in *C. raciborskii* MVCC14.

As we can see in Figure 1C, we amplify the full sxt cluster for cyanobacteria *C: raciborskii* MVCC14. These results show that the distribution of the gene cluster MVCC14 is similar to that described in *R. brookii* D9. Where can see that the gene cluster Meeting on sxtD preserved from gene to gene sxtO including ORF24 and *glmU* gene, both apparently would have no involvement in the synthesis of toxins. Of the genes in the cluster C. raciborskii MVCC14 presents a significant degree of similarity to genes in the gene cluster C. raciborskii T3, however, as in R. brookii D9, some genes, or as sxtSul and sxtDiox, would not be present in the gene cluster of C. raciborskii T3. Something similar happens with *sxtN*, *sxtX*, sxt*W* and *sxtV* genes present in *C. raciborskii* T3 but absent in MVCC14 and D9 (Table 1). The similarity between the genes in the *sxt* gene cluster present in MVCC14 begins to decrease when they are compared with the rest of the cluster so far described. The cluster described for *Lyngbya wollei* exhibits the less similarity with MVCC14, ranging from *sxtG* with 100% of similarity within a 93% of coverage, as the more conserved gene, to *sxtI* with only a 55% of conserved nucleotide in a 93% of coverage, however, the *sxtM* gene, presents a unique behavior.

The flank 5’ end of the *sxt* gene cluster of *C. raciborskii* MVCC14 there is ORF1. Based on BLASTP and PSI-BLAST searches of the NCBI nr database, we identified this unknown domain function (DUF58) protein in diverse cyanobacterial organisms. A

BLAST hit of 100% nucleotide identity corresponds to the same flanking genes that are present in the flanking 5'end of the *sxt* gene cluster in *C. raciborskii* T3 (Table 1).

At the flanking 3' end of the *sxt* gene cluster in *C. raciborskii* MVCC14, we identified two ORFs with a 99.5% and a 99.4% nucleotide identity respectively with *glmU* and *cpcB* ORF described in 3' end sxt gene cluster from *R. brookii* D9 (Fig. 2). In the 3' downstream region of the PST gene cluster from *C. raciborskii* T3, there is a putative ORF encoding a transposase, belonging to the IS1380 family, as determined by the IS Finder database (36).

Phylogeny and selection analysis

The phylogenetic reconstructions comparing the 16S-rRNA gene and the concatenation of the genes shared among saxitoxin-producing species show a consistent trend between both reconstructions, which indicate some degree of coevolution between phylogenetically close species (Fig. 3). The 16S-rRNA gene from *C. raciborskii* MVCC14 groups into a single branch with strains T3 (100% nucleotide identity) and *R. brookii* D9 (99,98% nucleotide identity). The 16S rRNA gene sequence from *A. circinalis* AWQC131C groups with *Aphanizomenon* sp. NH-5. *L. wollei* is the most distant organism according to this analysis (Fig 3). *sxt* cluster phylogeny shows two main branches: in an upper branch, *R. brookii* D9 is closely related to *C, raciborskii* MVCC14, which is immediately near to *C, raciborskii* T3; the lower branch groups *A. circinalis* AWQC131C, A*phanizomenon* sp.NH-5 and *L. wollei* (Fig.3). Phylogenies reconstruction was performed for each shared *sxt* gene and the corresponding 16S RNA from the same cyanobacterial strains (Fig 2S). Similar results were observed, the phylogeny trees were congruent. The only exception was the gene *sxtF* from *L. wollei* when phylogeny relationship was much closer to the C/R group than the 16S

RNA tendency. Selection analysis, based on the results obtained by the analysis of recombination GARD, identified various capacity points of recombination, but only one recombination point identified en *sxtM* gene was positively validated by the Kishino Hasegawa test (Table 2). Under the criterion of SLAC, it was possible to identify a positively selected codon and up to four with negative selection. FEL found a total of 11 codons in *sxtP* positively selected and 80 codon in *sxtA* negatively selected. In the *sxtM* gene, SLAC only detected three negative selected codons, FEL detected 10 positives and 37 negative codons and PARRIS identified a positively selection with a probability of 0.0024 in the *sxtM* gene.

Discussion

C. raciborskii MVCC14, has a shape similar to that observed in *C. raciborskii* T3 morphology. Light Microscopy, show a typically filamentary structure with vegetative cells and specialized nitrogen fixation heterocysts.

We corroborated by LC-MS/MS the presence of STX toxins in the strain *C. raciborskii* MVCC14. In genetic terms, we amplify a genetic region associated with the synthesis of the toxin, which proved to be similar to that described in *R. brookii* D9; however, there are certain differences in the flanking regions of this cluster. We also conducted a phylogenetic analysis in order to establish an evolutionary relationship concerning the acquisition of *sxt* cluster by different cyanobacterial species and compare to the 16S RNA molecular marker.

The presence of genes *sxtDiox* and *sxtSul* confirm the synthesis of STX and toxins GTX3/2, in the cyanobacterial cells extract. A similar result was also found in the cyanobacterium *R. brooki* D9. Comparison of the sxt cluster flanking regions between *C. raciborskii* T3 and *R. brookii* D9 reveals the presence of a domain with unknown function, DUF58, in the 5' end of the sxt gene cluster from *C. raciborskii* MVCC14 similar to 5' end identified in *C. raciborskii* T3. On the other hand, at the 3' end of the gene cluster in *C. raciborskii* MVCC14, we can see the presence of a coding region for a bifunctional GlmU protein, a similar condition that observed in *R. brookii* D9 *sxt* flanking 3' end.

With respect to this composition in the ORF1 gene, the 5' end region of the *sxt* cluster in *R. brookii* shows three different genetic sections could be explained a capacity point of recombination (Fig. 3). Comparing present in the flanking region 5' end from *C. raciborskii* MVCC14 and T3, one can see the first 150 bp ORF comprising 1500 bp DUF58 called (section 1), in the middle region one can see a segment of 40 bp presents no known counterpart in BLASTn database, but if you can observe 10 kb upstream of sxt cluster in R. brookii D9, annotated as protein with unknown function. Section 3 corresponds to a 150 bp segment corresponds to the last gene region encoding protein glmU. It is thought that this ability was conferred to D9 while observing the internal rearrangement process of the PSP toxin gene cluster; jumping from a DUF58 associated to an ATPase in MVCC14 and T3, to a *glmU* gene context. Is it that we are observing an HGT event between these two cyanobacterial strains, considering their geographic location. These three strains were collected from nearby sources in Brazil, suggesting there could be a transference event.

The translocation of these segments, which eventually constitute the ORF1 *R. brookii* D9, is a clear example of how gene regions suffer genomic rearrangement processes. However, this is difficult to define if translocation occurred process intragenomic mediated form or intergenomic horizontal gene transfer, given the high similarity shared by the sequences *sxt* in *C. raciborskii* and *R. brookii* clade (C-R clade). It is possible that this phenomenon is related to the reduction and genomic rearrangement experienced R. brookii D9 (5) compared to its closest relative *C. raciborskii* CS505. If present in the contexts cluster A. circinalis sxt, Aphanizomenon sp. or L. wollei was not possible to show any genetic element similar to that observed in R. brookii D9.

Previous work has reported that C-R form a monophyletic group (12), also observed in our analysis. A correlation (Fig. 3) between both, 16S rRNA genes and *sxt* cluster phylogenies observed in the concatenated phylogeny, indicate an early acquisition of the entire *sxt* gene cluster. Thus, suggesting a close evolutionary relationship at the time, the *sxt* cluster was spread. This finding supports the idea about the origin of toxin biosynthetic genes, the contribution of Horizontal Gene Transfer (HGT), deletions or gene insertions from different sources in different toxic strains, or the absence of them in closely related strains.

The most probable explanation towards the conservative but widely distributed toxin gene clusters is by considering a HGT event of the entire cluster or at least part of it (3). In fact, phylogenetic analysis of entire cyanobacterial genomes has detected that over a ~50% of extended gene families have a putative history of HGT -either between

cyanobacteria and other phyla, or within cyanobacteria, or both -suggesting that HGT plays an important role in cyanobacterial evolution (4).

In reference to the similarity between the *sxt* genes present in all identified cluster, it was not possible to validate by PARRIS none of the positively selected codons, except the gene encoding the toxin transporter *sxtM*, where PARRIS recognizes an event codons purifying selection and also were identified and validate a recombination event, according to KH test, which could explain why it is possible to observe 3 copies of *sxtM* gene in the cluster of *L. wollei* (25). However, this background does not explain the high selection that exists for all of the 15 genes shared by all the *sxt* gene clusters described so far, but this indicates that at least part of it present particular selection mechanisms that are independent of other *sxt* genes.

To further look for evidence of mobility inside or outside the *sx*t clusters and in the available genomes of the PST-producing cyanobacteria, we perform two different types of analysis: one based on the presence of entire coding sequences and the other one based on the remnants sequences from transposases and the presence of inverted repeats. The exhaustive search for inverted repeats in the genome of *R. brookii* D9 was done using the web application TRDB - The Tandem Repeats Database (TRDB - The Tandem Repeats Database Nucleic Acids Research, 2006, Vol. 00, Database issue D1-D8 doi:10.1093/nar/gkl1013). In D9, an ORF with high homology to protein InsA from T3 (93% of similarity) is located at contig xxx_5, ca. 100 kb apart from the *sxt* gene cluster. A second ORF is located upstream but in the opposite direction, with 60% of identity to an

IS1 transposase gene. These two ORFs are flanked by two perfect 78 bp inverted repeats, forming a 1,251 bp complete IS (Insertion Sequence) element. Remains unclear if this IS-element is a *bona fide* insertion element or was built by a mobility event. The latter idea surges as an explanation of the apparition in the upside-down sequence of more extended inverted repeat sequences (368 bp) that cover the complete *insA* like gene and the 5´ extreme from the transposase-like gene. In *C. raciborskii* T3, the genes *insA* and *insB* are located at one extreme of the *sxt* cluster, forming the left boundary of an insertion sequence with homology to IS1 family, including the 78 bp from the inverted repeat present in the element from *R. brookii* D9. The transposase of the bacterial IS1 is usually expressed by inefficient translational frameshifting between an upstream reading frame which itself specifies a transposition inhibitor, *InsA*, and a second consecutive reading frame located immediately downstream, *InsB* (Escoubas et al., 1994). Another putative mobile element found in the D9 genome is a Toxin-Antitoxin System (TAS) -like (begin of contig xxx_37). Close to this element, there is an ORF similar to *tnpA* from the ISAcma27 from *Acaryochloris marina*, member of the IS630 family. TAS-like systems belong to the resistome domain of the prokaryotic mobilome which includes partially selfish, addictive gene cassettes involved in various aspects of stress response. The "selfish altruism" or "responsible selfishness" of TAS-like systems appears to be a defining feature of the resistome and an important characteristic of the entire prokaryotic pan-genome given that in the prokaryotic world the mobilome and the "stable" chromosomes form a dynamic continuum (Makarova, 2009).

Mobile elements are typically clustered in specific zones from the bacterial genomes, in the chromosomes or extrachromosomal elements (REF). In contig xxx_37, we can note this typical configuration because of adjacent to the TAS-like element we can distinguish several ORFs with homology with genes codifying for transposases or resolvases. These ORFs are clustered at the beginning of the contig, and we can speculate that this is the reason why the *R. brookii* D9 genome was difficult to close.

Unfortunately, at present, there are only six *sxt* gene clusters described in an STX toxin-producing cyanobacteria. However, with this information, we can develop an inference of selective evolutionary mechanism that has occurred in a group of genes shared by all toxin-producing species, compared to the divergence between species, observe in the phylogeny reconstruction obtained by the 16S RNA. Certainly, it is necessary to use a higher number of evolutionary markers to confirm the selective phenomenon that occurs in the *sxt* gene cluster and try to understand why this is so important to keep it. In short, it is necessary to continue working on this issue, since besides this toxin is produced by algal blooms forming cyanobacteria that endanger human health.

References

1. Ausubel F, Brent R, Kingston R, Seidman J, Smith J, Struhl K. 1994. Current Protocols in Molecular Biology, Current Protocols, vol. I (Supplement 27), Wiley, New York, (pp. 2.4.1–2.4.2, Chapter II).

2. Carmichael WW, Azevedo SM, An JS, Molica RJ, Jochimsen EM, Lau S, Rinehart KL, Shaw GR, Eaglesham GK. 2001. Human fatalities from cyanobacteria: chemical and biological evidence for cyanotoxins. Environ Health Perspect. 109(7):663-8.

3. 30. Castro D, Vera D, Lagos N, Garcia C, Vásquez M. 2004. The effect of temperature on growth and production of paralytic shellfish poisoning toxins by the cyanobacterium *Cylindrospermopsis raciborskii* C10. Toxicon 44:483-489.

4. Diener M, Erler K, Christian B, Luckas B. 2007. Application of a new zwitterionic hydrophilic interaction chromatography column for determination of paralytic shellfish poisoning toxins. J. Sep. Sci. 30:1821-1826.

5. Escoubas, J. M., Lane, D., Chandler, M. 1994. Is the IS1 transposase, InsAB', the only IS1-encoded protein required for efficient transposition?, J Bacteriol. 1994 176(18): 5864–5867

6. Kellmann R, Mihali TK, Jeon YJ, Pickford R, Pomati F, Neilan BA. 2008. Biosynthetic Intermediate Analysis and Functional Homology Reveal a Saxitoxin Gene Cluster in Cyanobacteria. Appl. Envir. Microbiol. 74:4044–4053.

7. Kosakovsky Pond SL, Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21(10):2531-2533.

8. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. Molecular Biology and Evolution 23(10):1891-1901.

9. Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170-179.

10. Kosakovsky Pond SL, Frost SDW. 2005. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution 22(5):1208-1222.

11. Scheffler K, Martin DP, Seoighe C. 2006. Robust inference of positive selection from recombining coding sequences. Bioinformatics 22(20):2493-2499.

12. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34:D32-D36.

13. Soto-Liebe K, Murillo A, Krock B, Stucken K, Fuentes-Valdés JJ, Trefault N, Cembella A, Vásquez M. 2010. Reassessment of the toxin profile of *Cylindrospermopsis raciborskii* T3 and function of putative sulfotransferases in synthesis of sulfated and sulfonated PSP toxins. Toxicon. 56(8):1350-1361.

14. Soto-Liebe K, Stucken K, Méndez MA, Lagos N, Cembella A, Krock B, Vásquez M. 2008. The effects of chloramphenicol, arginine and temperature on PST-production by *Cylindrospermopsis raciborskii* strain D9. In: Moestrup, Ø., Doucette, G., Enevoldsen, H., Godhe, A., Hallegraeff, G., Lukas, B., Lundholm, N., Lewis, J., Rengefors, K., Sellner, K., Steidinger, K., Tester, P., Zingone, A. (Eds.), Proceedings of the 12th International Conference on Harmful Algae. International Society for the Study of

Harmful Algae (ISSHA) and the Intergovernmental Oceanographic Commission (IOC of UNESCO), Copenhagen, pp. 330–333.

15. Stucken K, Murillo AA, Soto-Liebe K, Fuentes-Valdés JJ, Méndez MA, Vásquez M. 2009. Toxicity phenotype does not correlate with phylogeny of Cylindrospermopsis raciborskii strains. Applied Microbiology. 32:37-48.

16. Mihali TK, Kellmann R, Neilan BA. 2009. Characterisation of the paralytic shellfish toxin biosynthesis gene clusters in *Anabaena circinalis* AWQC131C and *Aphanizomenon* sp. NH-5. BMC Biochemistry. 10:8.

17. Mihali TK., Carmichael WW, Neilan BA. 2011. A putative gene cluster from a *Lyngbya wollei* bloom that encodes paralytic shellfish toxin biosynthesis. 6(2):e14657.

18. Kellmann, R., T.K. Mihali, Y. J. Jeon, R. Pickford, F. Pomati, and B. A. Neilan. 2008. Biosynthetic Intermediate Analysis and Functional Homology Reveal a Saxitoxin Gene Cluster in Cyanobacteria. Appl. Envir. Microbiol. **74:**4044-4053.

19. Piccini C, Aubriot L, Fabre A, Amaral V, González-Piana M, Giani A, Figueredo CC, Vidal L, Kruk C, Bonilla S. 2011. Genetic and eco-physiological differences of South American *Cylindrospermopsis raciborskii* isolates support the hypothesis of multiple ecotypes. Harmful Algae 10:644-653.

20. Stucken K, Uwe J, Cembella A, Murillo AA, Soto-Liebe K, Fuentes-Valdés JJ, Friedel M, Plominsky AM, Vásquez M, Glöckner G. 2010. The Smallest Known Genomes of

Multicellular and Toxic Cyanobacteria: Comparison, Minimal Gene Sets for Linked Traits and the Evolutionary Implications. PLoS ONE 5(2):e9235

21. Stüken A, Orr RJ, Kellmann R, Murray SA, Neilan BA, Jakobsen KS. 2011. Discovery of nuclear-encoded genes for the neurotoxin saxitoxin in dinoflagellates. PLoS One. 6(5):e20096.

22. Hansen C, Strichartz G. 1980. Saxitoxin Binding to Sodium Channels of Rat Skeletal Muscles. J. Physiol. 300:89-103.

23. Llewellyn LE. 2006. Saxitoxin, a toxic marine natural product that targets a multitude of receptors. Nat Prod Rep. 23(2):200-22.

24. Onodera H, Satake M, Oshima Y, Yasumoto T, Carmichael WW. 1997. New saxitoxin analogues from the freshwater filamentous cyanobacterium *Lyngbya wollei*. Nat. Toxins 5:46-151.

25. Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle DF, Papke RT. 2006. Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. Genome Research. 16:1099-1108.

Authors' contributions

JJFV and KS designed the experiments. JJFV, BK, NT, and DS conducted the experiments. JJFV, KS, and MV analyzed and interpreted the data and drafted the manuscript. All authors revised the manuscript critically for important intellectual content, participated in the discussions and approved its final form.

Acknowledgments

We thank Annegret Müller (AWI, Bremerhaven, Germany) who confirmed the toxin profiles by LC-MS/MS, Dr. S. Bonilla (Instituto de Investigaciones Biologicas Clemente Estable, Montevideo, Uruguay) who kindly provided the MVCC14 strain, and Nathalie DelHerbe (PUC, Santiago, Chile) for technical support. Financial support for this work was provided by Grants from Chile to M.V.: Fondecyt 1050433 and 1080075, Fondef MR07I1005.

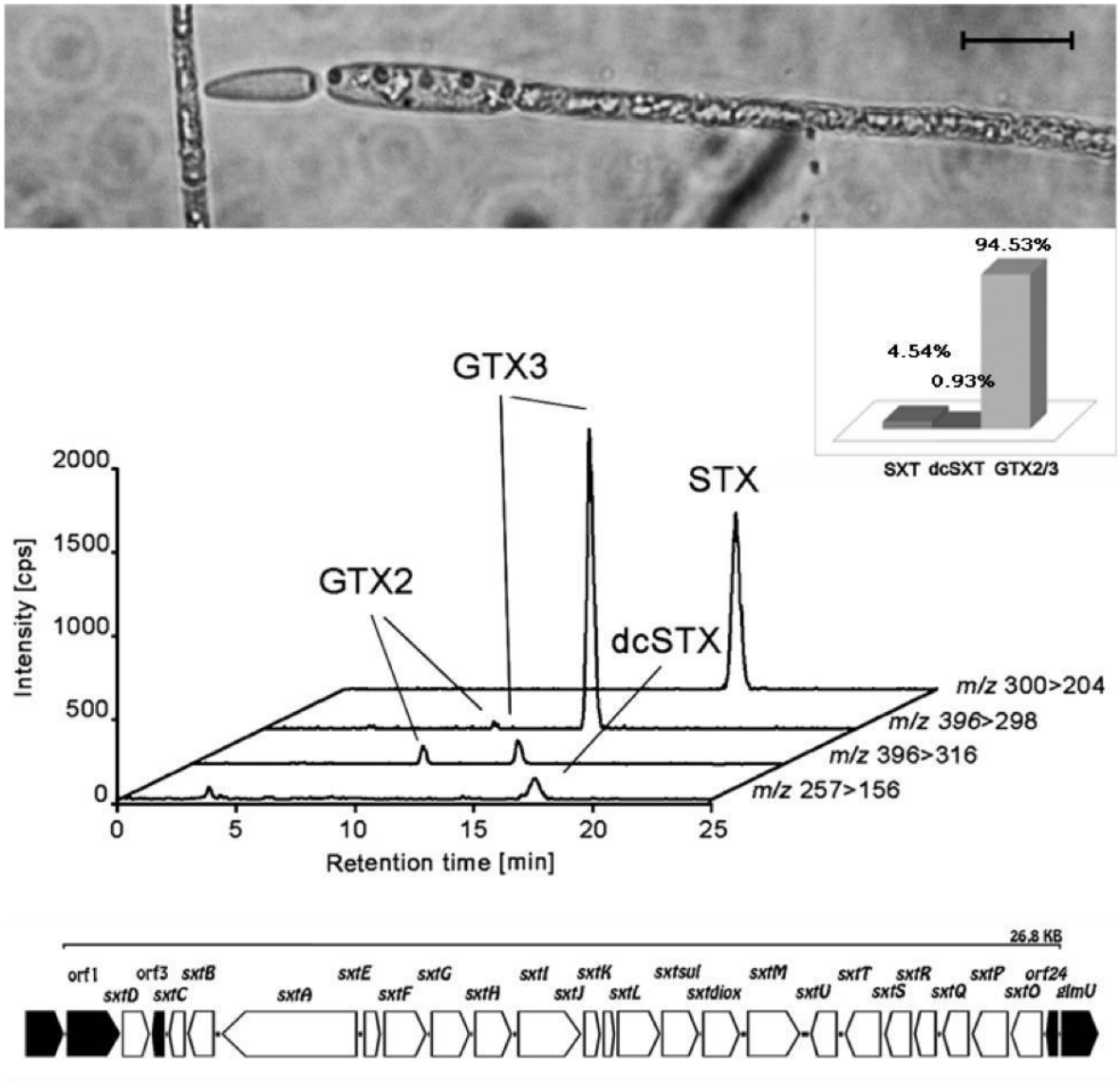

Figure 1 - A) Light Microscopy of MVCC14 and B) toxin profile by LC-MS. C) PSP-toxin cluster of MVCC14.

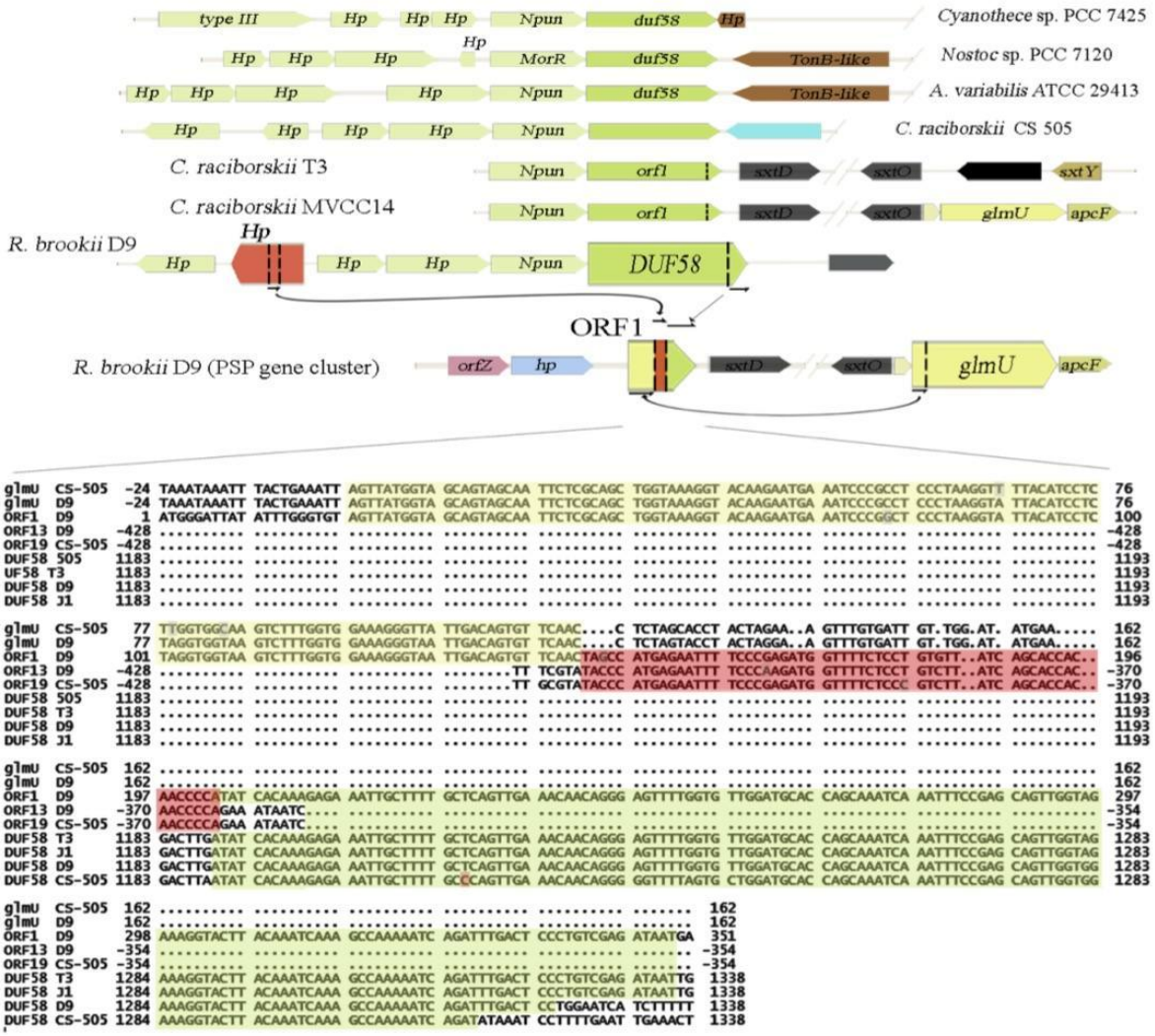

Figure 2 - Alignment of ORF1 sequences and *sxt* gene context sequences from *R. brookii* D9.

Figure 3 - Correlation between phylogenetic analysis based on 16S rRNA and sxt concatenated conserved genes.

Tables

Table 1S - Primers used and PCR fragment sizes.

Primer Pair Names	Primer Sequence	Amplified Region (bp)	Authors
14 orf1 R 14 orf1 F	CTAACACTGGTTCAATAGGT ATGATCCCAGCTAAAAAAGT	1-940	
D9 RIsxtOrf1D R 15 ATP F	TTTAATCCGTGCTGACTA GAACCTGTTAAACTGGTTAG	6-2665	
7 sxtD R 7 DUF58 fF	ATG GAA GGG AGA GCG ATT TT ATA CCT CCA GAG CGA GGA CA	868-1720	
9 sxtA R 8 sxtD F	TGG TTA CAC AAG CAA GAG CCA ATA TCA GTA CTA TTA AGA GAG TG	1637-5360	
sxtA R sxtA F	GCG GTT CCC GTT ATT CTT GC CTC CTC TTC GGT ATT GGC GG	5146-8745	
10 sxtG R 10 sxtA F	TCC TTG AGC ATC TCT CGA TAC CTC CCA GAT GTT CCA CCA TTT	8646-11599	
11 sxtI R 11 sxtG F	GAT ACC AGG GTG ATA ATT AG TTG GTG ATT GGG AAT GAG ATT	11511-15858	
12 sxtL R sxtI F	CTA TAT CAG GTG AAT AGT CCC TCT GAA ATT CGG ACT GCG GA	13010-16417	

Table 1 - Nucleotide Identity of *sxt* genes for *C. raciborskii* MVCC14, *C. raciborskii* T3, *R. brookii* D9, *Anabaena circinalis* AWQC131C, *Aphanizomenon* sp. NH-5 and *Lyngbya wollei*

MVCC14 Gene	% (Identity/Covery) to corresponding gene from strain: T3	D9	AWQC11C	NH-5	LW
ORF1	100/100	-	-	-	-
sxtD	99/100	100/100	89/100	89/100	90/100
ORF3	sxtR	100	89	89	90
sxtC	100/99	100/99	93/80	92/80	87/98
sxtB	100/99	100/100	88/100	89/100	91/100
sxtA	99/100	100/100	92/98	92/98	90/99
sxtE	100/100	99/100	89/99	89/99	91/99
sxtG	99/100	99/100	93/100	94/100	93/100
sxtH	99/100	99/100	86/100	87/100	88/100
sxtI	99/100	99/100	90/100	91/100	93/55
sxtJ	sxtR	99/100	92/100	92/100	92/100
sxtK	sxtR	99/100	92/100	92/100	92/100
sxtL	sxtR	sxtR	sxtR	sxtR	sxtR
sxtsul	-	100/99	90/99	86/99	93/99
sxtdiox	-	100/99	87/99	88/99	91/99
sxtM	91/99	99/100	89/99	85/99	91/99*
sxtU	99/99	100/99	88/99	89/99	93/99
sxtT	99/99	100/99	87/99	88/99	91/99
sxtS	99/99	100/99	90/99	86/99	93/99
sxtR	sxtR	sxtR	sxtR	sxtR	sxtR
sxtQ	99/100	99/100	92/100	92/100	92/100
sxtP	99/100	99/100	92/100	92/100	92/100
sxtO	sxtR	sxtR	sxtR	sxtR	sxtR
ORF24	-	99/100	92/100	92/100	92/100
glmU	-	99/100	92/100	92/100	92/100

Table 2 - Number of capacity recombination breakpoints and selection test results for the *sxt* genes.

Gene	No. of recombination breakpoints by GARD model	ΔAICca	Significance by KH test (P < 0.05)b	No. of codons found by positive/negative selection with: SLAC	FEL	Significance of positive selection by PARRIS (P < 0.05)
sxtA	6	2.64	NS	0/4	2/80	NS
sxtB	2	1.10	NS	0/2	1/40	NS
sxtC	0	-	-	0/3	0/12	NS
sxtD	0	-	-	0/0	1/20	NS
sxtE	1	3.04	NS	0/0	0/6	NS
sxtG	0	-	-	0/0	0/36	NS
sxtH	2	7.66	NS	0/0	5/31	NS
sxtI	1	419.75	NS	0/2	0/71	NS
sxtM	5	10.47	1	0/3	10/37	S p = 0.0024
sxtP	1	14.01	NS	0/0	11/12	NS
sxtQ	0	-	-	0/0	0/23	NS
sxtR	1	117.50	NS	0/1	0/18	NS
sxtS	2	6.99	NS	1/0	0/17	NS
sxtT	3	3.17	NS	0/1	1/44	NS
sxtU	-	-	-	0/1	0/25	NS