Physiogenomics of *Cylindrospermopsis raciborskii* and *Raphidiopsis brookii* (Cyanobacteria) with Emphasis on Evolution, Nitrogen Control and Toxin Biosynthesis

Dissertation submitted by

Karina Susana Stucken Marin

In partial fulfillment of the requirements for

Doktors der Naturwissenschaften (Dr. rer. Nat.)

In the Faculty of Biology and Chemistry (Fachbereich 2)
University Bremen
Germany

Bremen, May 2010

TABLE OF CONTENTS

I. AKNOWLEDGEMENTS

Somebody said that in science, we do what we like (want) and we are paid anyway! I want to thank my beloved supervisors Allan Cembella and Mónica Vásquez, first of all, to give me that chance, of working in the subject I like and enjoy. Second, I want to thank them for trust me the way they did and because they were not just supervisors, they were always close to me. Profe Mónica, although being far away, I still had the support and presence of you, we could discuss from experimental designs to final discussions, after of course, catching up in how our lives were going. You support went beyond the role as supervisor and I can say that in the five years I have known you I have growth not only as a scientist but as a better person.

Uwe, I want to thank you for your unconditional support and always good disposition. Your constant energy, positivism and happiness kept me motivated; after meeting with you I could see always the bright side of things, there were no bad results, there were more interesting!.

I want to thank to all present and past members of the AG Cembella (Nina, Ines, Sylke, Stefan, Michael, Haiyan, Philip, Aboli, Tilman, Sára, Eva, Bernd, Urban, Annegret, Wolfgang and Andrea) and the people that crossed my life within the years I spent in AWI, office mates, students, and short-time visitors. We shared coffees, barbecues, cocktails or maybe just relaxing talks, you all brought warmness to the cold times in Bremerhaven.

During my PhD I had the opportunity to maintain the contact and work together with the Microbial ecology and environmental toxicology group in Chile. All of you Katia, Alejandro, Juanjo, Nicole, Alvaro, Alejandra and Nathalie, were always available to discuss new results and solve any problem that could appear. You were also there to answer my phone calls when I just needed to talk, thank you guys for being there!

As part of the MarMic program in the MPI for Marine Microbiology in Bremen, I had the opportunity to meet other foreign students in a friendly environment like our yearly retreats or in of scientific courses. I was able to teach other MarMic students, Hannah and Sandra, who I thank for their enthusiasm in learning and the help in the progress of my PhD.

The genome sequencing and bioinformatic analyses would have not been possible without the help of Gernot Glöckner and the people form Fritz-Lippmann Institute in Jena. I want to thank them and also Rodrigo Gutierrez, for introducing me to the R software and for his contribution with the microarray data analysis.

Also special thanks to Sven Kranz and Torben Gentz, for their assistance in the acetylene reduction assays.

To my family which I love and miss, Mamá, Papá, Sonia, Iván, Maca, Fran, Claudia, Sebi y Nico, I thank you for all your love, without you there would be no motivation for continuing my career.

Ronny, mi amor, I started in Germany a life with you, and it has been beautiful to have you at my side. To be able to go home after a tiring day and knowing you will be there, that you will comfort me and make me laugh, were all dreams that came true when you came to Germany. I love you and thank you for loving me and sharing your life with me. Also thanks to your wonderful family for receiving me with their hearths.

II. ABSTRACT

This thesis presents an integrated study of evolutionary relationships linked to nitrogen metabolism and toxin production in filamentous cyanobacteria. The first aim was to use a comparative genomics approach based upon fully sequenced genomes to uncover the genotypic features underlying the phenotypic differences between the toxigenic cyanobacteria *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9. Although the two species have been assigned to different genera, the presence of heterocysts (specialized cells for dinitrogen (N_2) fixation) in *C. raciborskii* is the only distinguishing morphological characteristic between these species. Furthermore, phylogenetic analysis positioned both *C. raciborskii* and *Raphidiopsis* spp. within the same monophyletic branch. *C. raciborskii* CS-505 isolated from Australia and *R. brookii* D9 from a toxic bloom in Brazil, produce the hepatotoxin cylindrospermopsin (CYN) and neurotoxins associated with paralytic shellfish poisoning (PSP), respectively. Based on the genetic and morphological similarities and in the fact that N_2 fixation and heterocyst development separates these species, the effect of nitrogen (N) regime on physiology and gene expression of cell differentiation and of toxin biosynthesis/regulation was compared for the species.

Genome sequencing revealed a large set of genes (2,539) shared between CS-505 and D9 with an average nucleotide identity >90%, corresponding to conspecific isolates that have followed diverging evolutionary pathways. These genomes are the smallest known for filamentous cyanobacteria. Nevertheless, these strains are capable of cell differentiation within filaments, which allowed to determination of minimal gene sets for filament formation, N_2 fixation and heterocyst differentiation. The gene organization (synteny) between genomes indicated that the small genome (3.2 Mb) of D9 was structured by gene loses, whereas the numerous plasticity elements of the somewhat larger genome (3.9 Mb) of CS-505 may be associated with genome reduction, and allow for acquisition of new genes via horizontal gene transfer (HGT).

Although D9 does not develop heterocysts, several genes related with this process are encoded in the genome and some genes are even induced under N-deprivation, as revealed by whole genome microarray hybridization. These genes may therefore have been recycled for general metabolism or to fulfill roles in other cell differentiation processes, such as akinete development, cell division, or cell death. The process of heterocyst differentiation in CS-505,

unlike in the model cyanobacterium *Anabaena* sp. PCC 7120, occurs in terminal cells and in the presence of combined N. Physiological and transcriptional analysis revealed that while N-metabolism genes are promptly regulated under N-deprivation, the cascade of transcriptional responses leading to heterocyst differentiation starts between 24-48 h, after the induction of *ntcA*. In correlation, new heterocysts are formed within 48-72 h. These late responses contrast with those of other heterocystous cyanobacteria, suggesting an efficient mechanism of N- storage and transport along the filament in *C. raciborskii* CS-505.

Finally, examination of the N-dependent production and regulatory pathways of CYN and PSP toxins (PST) showed that in early exponential phase, growth inhibition arrests PST- and slows CYN-production. The pattern of biosynthetic gene regulation suggested that additional mechanisms of N-control and post-transcriptional mechanisms are involved in toxin production. The role of repetitive elements and transposases in gene regulation may be critical but this aspect requires further investigation.

In summary, these physiogenomic studies substantially contributed to understanding evolution, and processes of cell differentiation and toxin production, in these two closely related cyanobacteria. The small and likely streamlined genomes recommend these strains as model systems for further investigation of related processes in filamentous cyanobacteria.

III. ZUSAMMENFASSUNG

Die vorliegende Arbeit umfasst eine integrierte Studie, in der die evolutiven Beziehungen mit dem Stickstoffmetabolismus und der Toxinproduktion in filamentösen Cyanobakterien verknüpft werden. Das Hauptziel dieser Arbeit bildet die vergleichende Genomik der vollständig sequenzierten Genome zweier toxischen Cyanobakterien, *Cylindrospermopsis raciborskii* CS-505 und *Raphidiopsis brooki* D9, um genotypische Merkmale zu entdecken, die den phenotypischen Unterschieden zugrunde liegen. Die beiden verwendeten Cynanobakterien-Arten sind zwar unterschiedlichen Gattungen zugeordnet, dass Vorhandensein von Heterozysten (spezialisierte Zellen zur Fixierung molekularen Stickstoffs) ist jedoch das einzige morphologische Merkmal, das die beiden Arten unterscheidet. Des Weiteren positionierten phylogenetische Untersuchungen beide Arten in dieselbe monophyletische Gruppe. Sowohl *C. raciborskii* CS-505, isoliert in Australien, als auch *R. brookii* D9, isoliert aus einer toxischen Algenblüte in Brasilien, produzieren das Heptatoxin Cylindrospermopsin (CYN) und Neurotoxine assoziiert mit der paralytischen Form der Muschelvergiftungen (PSP, Paralytic Shellfish Poisioning). Die genetischen und morphologischen Ähnlichkeiten sowie die Tatsache, dass die Fähigkeit der molekularen Stickstofffixierung und die damit verbundende Bildung von Heterozysten diese beiden Arten unterscheidet, bildeten die Basis für die vergleichenden physiologischen Untersuchungen des Stickstoffhaushalts und dessen Effekte auf die Genexpression, die Zelldifferenzierung sowie auf die Biosynthese und die Regulation der Toxinproduktion in den beide Arten.

Durch die Sequenzierung der Genome der beiden Arten CS-505 und D9 konnte für einen großen Teil der Gene (2539) eine Übereinstimmung in der Nukleotidabfolge von über 90% festgestellt werden; dies entspricht einer Ähnlichkeit wie sie bei Stämmen einer Art mit anschließender divergierender Entwicklung vorzufinden ist. Die beiden Genome sind die bisher kleinsten bekannten Genome filamentöser Cyanobakterien. Dennoch sind beide Arten dazu befähigt, Zellen in Filamente zu differenzieren und dies ermöglichte, eine Minimalanzahl an Genen der Filamentformation zuzuordnen sowie Gene, die für die N_2-Fixierung und die Zelldifferenzierung zu Heterozysten notwendig sind, zu bestimmen. Der Vergleich der Abfolge der Gene (Synteny) der Genome zeigte, dass das vergleichsweise kleinere Genom von D9 (3,2MB) durch Gen-Verluste gekennzeichnet ist, während die zahlreichen plastischen Elemente

des vergleichsweisen größeren Genoms von CS-505 (3,9MB) auf eine Genomreduktion deuten und eine damit verbundene Aufnahme neuer Gene durch horizontalen Gentransfer (HGT).

Obwohl D9 keine Heterozysten ausbildet, sind mehrere an diesem Prozess beteiligten Gene in dem Genom zu finden und durch Microarray Hybridisierungen konnte gezeigt werden, dass diese sogar unter Stickstoffmangel exprimiert werden. Diese Gene könnten daher für allgemeine metabolische Funktionen recycelt worden sein oder sind an anderen Zelldifferenzierungsprozessen wie der Bildung von Akineten, der Zellteilung oder dem Zelltod beteiligt. Die Ausbildung von Heterozysten in CS-505 erfolgt im Gegensatz zu dem Modelcyanobakterium *Anabaena* sp. PCC 7120 in den terminalen Zellen und in Anwesenheit von fixiertem Stickstoff. In physiologischen und transkriptomischen Untersuchungen konnte gezeigt werden, dass die Regulation der Gene des N-Metabolismus direkt auf den N-Mangel folgt, während die Regulation der an der Heterozystendifferenzierung beteiligten Gene zwischen 24-48 Std. nach der Aktivierung des *ntcA* erfolgt. Übereinstimmend dazu werden die neuen Heterozysten innerhalb von 48-72 Std. geformt. Diese verspätete Reaktion steht in Kontrast zu anderen Heterozysten-bildenden Cyanobakterien und deutet daher auf einen effizienteren Mechanismus der N-Speicherung und des N-Transports entlang der Filamente in *C. raciborskii* CS-505 hin.

Die Untersuchungen zur N-abhängigen Produktion und der Regulation der Stoffwechselwege der CYN und PSP-Toxine (PST) zeigten, dass die einsetzende Wachstumshemmung in der frühen exponentiellen Phase die PST-Produktion stoppt und die CYN-Produktion reduziert. Zugehörige Genregulationsmuster deuten auf eine zusätzliche Beteiligung von N-Regulierungs- und posttranskriptionellen Mechanismen bei der Toxinproduktion hin. Die Rolle repetitiver Elemente und Transposasen an der Genregulation könnte hierfür entscheidend sein, dieser Aspekt bedarf jedoch weiterer Untersuchungen.

Zusammenfassen kann gesagt werden, dass diese Studie entscheidend zu dem Verständnis von Evolution, Zelldifferenzierungsprozessen und der Toxinproduktion in diesen naheverwandten Arten beigetragen hat. Die kleinen und zweckorientierten Genome dieser beiden Stämme eigen sich optimal als Modelsysteme für weitere Untersuchungen verwandter Prozesse in filamentösen Cyanobakterien.

IV. ABBREVIATIONS

2-OG	2-oxoglutarate
ANI	average nucleotide identity
cDNA	complementary DNA
CDS	coding sequence
cRNA	complementary RNA
CYL	cylindrospermopsin
CYN	cylindrospermopsin
cyr	cylindrospermopsin biosynthetic gene cluster
DNA	deoxyribonucleic acid
dcNEO	decarbamoyl neosaxitoxin
dcSTX	decarbamoyl saxitoxin
doCYN	deoxycylindrospermopsin
GTXs	gonyautoxins
HABs	harmful algal blooms
HGT	horizontal gene transfer
HPLC	high performance liquid chromatography
ITS	internal transcribed spacer
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry
Ma	million years
Mb	mega bases
N_2	dinitrogen
NEO	neosaxitoxin
NRPS	nonribosomal peptide synthetases
ORF	open reading frame
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
PKS	polyketide synthase
PSP	paralytic shellfish poisoning
PST	paralytic shellfish toxins
qPCR	quantitative PCR
rDNA	ribosomal DNA
RNA	ribonucleic acid
STX	saxitoxin
sxt	saxitoxin biosynthetic gene cluster
tsp	transcription start point
TU	transcription unit

1. GENERAL INTRODUCTION

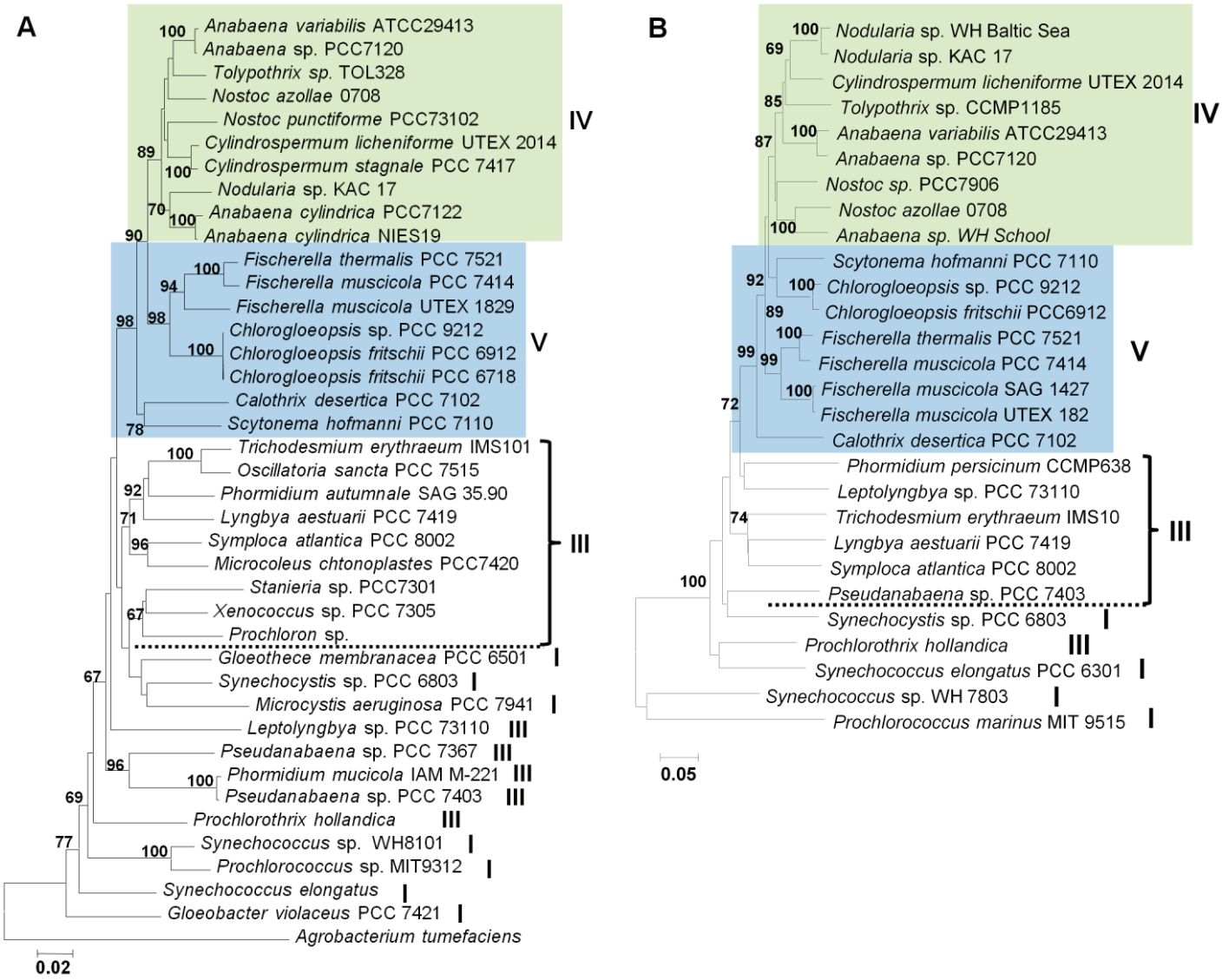
1.1 Cyanobacteria

Cyanobacteria or “blue-green algae” are unique prokaryotes in that they perform oxygenic photosynthesis. This old evolutionary group epitomizes the ancestors of chloroplasts of higher plants and of other photoautotrophic microorganisms. Cyanobacteria have diversified to populate almost all environments on Earth. They are commonly found in oceans and freshwaters and as extremophiles in polar ice and geysers. Cyanobacteria are typically free-living but they can form symbiotic associations with fungi (e.g. as lichens), or higher plants and may occur in mats on various substrates. Marine unicellular cyanobacteria, e.g. represented by *Synechococcus* spp. and *Prochlorococcus* spp., occupy an important position at the base of the marine food web. These cyanobacterial picoplankton are abundant in the world's oceans and as a result are major primary producers on a global scale (Partensky *et al.*, 1999, Chisholm *et al.*, 1988).

Cyanobacterial morphology is diverse, including unicellular (single cells or colonies) and filamentous forms. Filamentous cyanobacteria are often considered as multicellular microorganisms since many are capable of cell differentiation. Under stress or unfavorable environmental conditions, akinetes (spore-like) cells may develop. Alternatively, heterocysts are specialized cells committed for atmospheric nitrogen (N_2) fixation; they are formed by the deposition of a thick polysaccharide layer that isolates the cell from external oxygen allowing the activity of the nitrogenase enzymatic complex, responsible for the reduction of N_2 to ammonium. The capability of N_2 fixation makes cyanobacteria ideal partners for symbiotic relationships. For example, members of the order Nostocales differentiate temporary vegetative filaments into motile filaments called hormogonia that allow the cyanobacteria to colonize its host (Meeks & Elhai, 2002).

The advent of oxygenic photosynthesis, marked by the emergence of cyanobacteria, was a major determinant of the evolution of life on Earth. Geological records have dated the appearance of cyanobacteria before 2,800 million years ago (Ma) (Des Marais, 2000). Although is still under debate, the discovery of Archaean microfossils with cyanobacteria-like morphology in Australia, could extend the records of cyanobacteria to 3,300 – 3,500 Ma (Schopf & Packer, 1987). More integrative analysis, employing phylogenetics and geological data has suggested that

cyanobacteria capable of cell differentiation are monophyletic and that this clade diverged between 2,450 and 2,100 Ma (Tomitani *et al*., 2006). Based on morphological criteria and division patterns, cyanobacteria are classified into five subsections (Rippka *et al*., 1979). Subsections I and II comprise unicellular cyanobacteria and subsections III to V include filamentous non-heterocystous (III) and heterocyst-forming (IV-V) cyanobacteria. Evolutionary inferences of the cyanobacterial origin have been elaborated on a single gene basis or through phylogenomics. Phylogenetic inferences based on sequences of common phylogenetic markers such as the 16S rDNA and the gene for the larger subunit of the Rubisco (*rbcL*), have the advantage that they can cover a large number of organisms while phylogenomics are restricted to the few completed genome sequences, which at present do not include members of subsection V. Either way, there is a consensus supporting a monophyletic origin of heterocyst formers from subsections IV and V (Figure 1.1) (Sánchez-Baracaldo *et al.*, 2005, Tomitani *et al*., 2006, Swingley *et al.*, 2008) –Sánchez-Baracaldo's and Swingley's work did not include members of subsection V- whereas single phylogenetic markers reveal that cyanobacterial members from subsection III group together with unicellular cyanobacteria (Figure 1.1A, B) (Tomitani *et al*., 2006, Honda *et al.*, 1999). It is important to consider that monophyly of cyanobacteria can be compromised by horizontal gene transfer (HGT) events. HGT is a major driver factor in bacterial evolution (Doolittle, 1999, Ochman *et al.*, 2000, Gogarten *et al.*, 2002, Shimodaira & Hasegawa, 1999), and especially in cyanobacteria (Zhaxybayeva *et al.*, 2004, 2006), which have a large number of elements favoring HGT in HGT such as transposases, phage integrases, insertion and repeated sequences, favoring HGT (Frangeul *et al.*, 2008, Kaneko *et al.*, 2007, Elhai *et al.*, 2008). Based upon phylogenetic analysis of multiple data sets, almost 25% of the analyses positioned non-cyanobacteria within the cyanobacterial clade, and of the genes cyanobacterial monophyly, a majority showed evidence of HGT within the cyanobacteria (Zhaxybayeva *et al*., 2006). Similar findings were reported by Shi & Falkowski, (2008).


A
Anabaena variabilis ATCC29413
Anabaena sp. PCC7120
Tolypothrix sp. TOL328
Nostoc azollae 0708
Nostoc punctiforme PCC73102
Cylindrospermum licheniforme UTEX 2014
Cylindrospermum stagnale PCC 7417
Nodularia sp. KAC 17
Anabaena cylindrica PCC7122
Anabaena cylindrica NIES19
IV
Fischerella thermalis PCC 7521
Fischerella muscicola PCC 7414
Fischerella muscicola UTEX 1829
Chlorogloeopsis sp. PCC 9212
Chlorogloeopsis fritschii PCC 6912
Chlorogloeopsis fritschii PCC 6718
Calothrix desertica PCC 7102
Scytonema hofmanni PCC 7110
V
Trichodesmium erythraeum IMS101
Oscillatoria sancta PCC 7515
Phormidium autumnale SAG 35.90
Lyngbya aestuarii PCC 7419
Symploca atlantica PCC 8002
Microcoleus chtonoplastes PCC7420
Stanieria sp. PCC7301
Xenococcus sp. PCC 7305
Prochloron sp.
III
Gloeothece membranacea PCC 6501
Synechocystis sp. PCC 6803
Microcystis aeruginosa PCC 7941
Leptolyngbya sp. PCC 73110
Pseudanabaena sp. PCC 7367
Phormidium mucicola IAM M-221
Pseudanabaena sp. PCC 7403
Prochlorothrix hollandica
Synechococcus sp. WH8101
Prochlorococcus sp. MIT9312
Synechococcus elongatus
Gloeobacter violaceus PCC 7421
Agrobacterium tumefaciens
I
0.02
B
Nodularia sp. WH Baltic Sea
Nodularia sp. KAC 17
Cylindrospermum licheniforme UTEX 2014
Tolypothrix sp. CCMP1185
Anabaena variabilis ATCC29413
Anabaena sp. PCC7120
Nostoc sp. PCC7906
Nostoc azollae 0708
Anabaena sp. WH School
IV
Scytonema hofmanni PCC 7110
Chlorogloeopsis sp. PCC 9212
Chlorogloeopsis fritschii PCC6912
Fischerella thermalis PCC 7521
Fischerella muscicola PCC 7414
Fischerella muscicola SAG 1427
Fischerella muscicola UTEX 182
Calothrix desertica PCC 7102
V
Phormidium persicinum CCMP638
Leptolyngbya sp. PCC 73110
Trichodesmium erythraeum IMS10
Lyngbya aestuarii PCC 7419
Symploca atlantica PCC 8002
Pseudanabaena sp. PCC 7403
III
Synechocystis sp. PCC 6803
Prochlorothrix hollandica
Synechococcus elongatus PCC 6301
Synechococcus sp. WH 7803
Prochlorococcus marinus MIT 9515
I
0.05

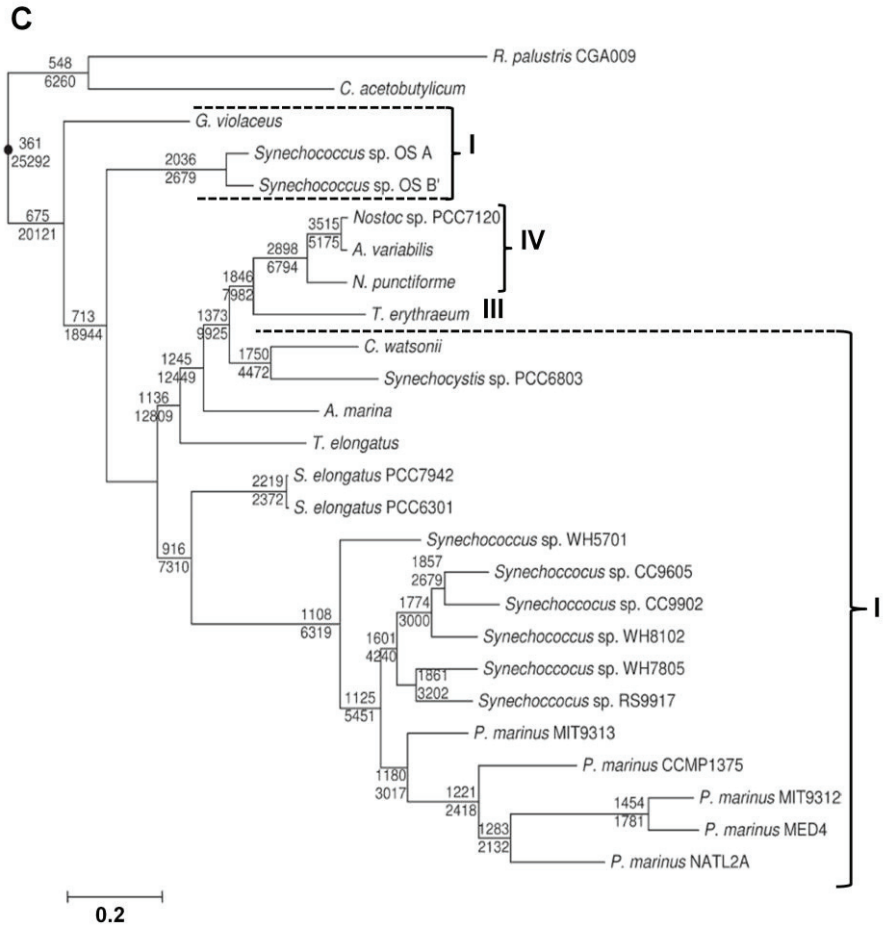


Figure 1.1 Phylogenetic relationships of cyanobacteria. The phylogenetic inference was based on the 16S rDNA **(A)** and *rbcL* **(B)** genes. A and B trees were created from the nucleotide sequences obtained by Tomitani *et al*., (2006), although more taxa were incorporated to the analysis. Nucleotide sequences were aligned using clustalX2 (Higgins & Sharp, 1988, Larkin *et al.*, 2007) and trees were created with MEGA, using the Neighbor Joining algorithm (NJ) (Saitou & Nei, 1987). **(C)** Phylogenomic tree reproduced and modified from Swingley *et al*. (2008). The tree was constructed using the Bayesian maximum likelihood algorithm for a full-concatenated data set, based on 230,415 aligned positions in 26 genomes. Note that this analysis was based in complete genome sequences; therefore there are no representatives from section V and only one representative from section III. Shown at each bifurcation are the predicted core-genome (upper number) and pan-genome (lower number) sizes of an ancestor at that point. The scale bars indicate the number of substitutions per site. The subsections I – V of cyanobacteria are indicated with Roman numerals. Color figures highlight cyanobacteria belonging to group IV (green) and V (light blue). Accession numbers of the taxa included are as follows: 16S rDNA partial sequence: *Tolypothrix sp.* TOL328 (AM230706); *N. azollae* 0708 (ACIR01000045, azoDRAFT_R0016), *C. stagnale* PCC 7417 (AF132789), *A. cylindrica* NIES19 (AF247592), *C. fritschii* PCC 6718 (AF132777), *C. desertica* PCC 7102 (AM230699), *P. autumnale* SAG 35.90 (EF654081), *P. mucicola* IAM M-221 (AB003165), *Prochlorococcus* sp. MIT9312 (AF053398), *A. tumefaciens* (D14500). *rbcL* gene: *Nodularia* sp. WH Baltic Sea (AB075916), *Nostoc* sp. PCC 7906 (AB075918), *N. azollae* 0708 (NZ_ACIR01000080, AazoDRAFT_4615), *Anabaena* sp. WH School (AB075905), *F. muscicola* SAG 1427 (AB075911), *S. elongatus* PCC 6301(NC_006576, syc0130_c), *Synechococcus* sp. WH 7803 (NC_009481, synWH7803_0678), *P. marinus* MIT 9515 (NC_008817, P9515_06141).

1.2 Harmful algal blooms (HABs)

Although cyanobacteria are of ecological importance and are primarily beneficial, excessive proliferation can be detrimental for the environment. The low requirements for cyanobacterial growth (fundamentally light, nitrogen, phosphorous and carbon) in concert with an increase in the eutrophic levels and temperature of water bodies, have promoted the excessive growth of cyanobacteria into what is known as harmful algal blooms (HABs). Cyanobacterial blooms are often accompanied by noxious and/or toxic effects to animals, humans or to ecosystem damage in which case they are referred to as cyanobacterial HABs. A schematic view of all the possible negative effects of a cyanobacterial bloom and its associated consequences to the environment is shown in Figure 1.2. Cyanobacteria can mobilize throughout the water column due to their buoyancy regulating ability depending on the nutrients and light availability, or due to waterbody-mixing processes. In the case of freshwater blooms, especially when the water column remains poorly mixed, cyanobacteria can normally cover large surfaces forming scums shading other phytoplankton and plants inhabiting at lower depths. This is one of the effects caused directly by the presence of a bloom, however, cyanobacterial blooms can negatively affect their environment in several other ways. For example, the elevated photosynthetic rates reduce the levels of CO_2 and increase the water's pH. Bloom decays produce an accumulation of organic matter that is degraded by bacteria, reducing the oxygen available for eukaryotes, causing for example severe fish kills.

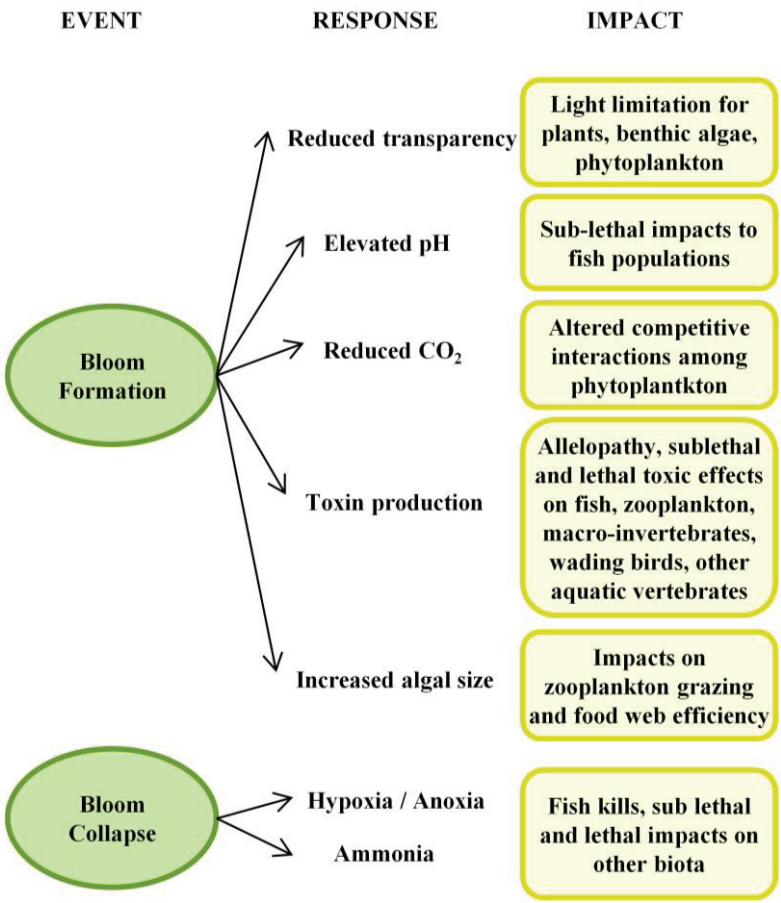


Figure 1.2 Scheme of the detrimental effects of cyanobacterial blooms (adapted from Havens, 2008). Noxious effects can occur directly during the bloom phase, or as a consequence of bloom decay and cell lysis.

The favorable conditions for blooms generally met in tropical, subtropical and temperate water bodies, where there is a seasonal succession of cyanobacteria. The Baltic Sea, the largest brackish water body in the world, experiences a succession of cyanobacterial blooms, with members of *Anabaena* spp. occurring in the summer, whereas late-summer blooms consist mainly of *Nodularia spumigena*, and *Aphanizomenon flos-aquae* (Stal *et al*., 2003). In lakes, ponds and reservoirs the conditions are different; waters are tipically calm and water columns are normally stratified. Under these conditions in tropical waters blooms can persist throughout the year, as is the case for some blooms in Brazil and Australia (McGregor & Fabbro, 2000, Bouvy *et al.*, 1999). Cyanobacterial blooms are not formed in temperate oceans. The reasons for this are still ambiguous, but it was proposed that the hydrodynamic instability of these environments do not allow cyanobacterial to float and settle down in the surface water layers (Stal *et al*., 2003).

1.2.1 Cyanobacterial toxins

Cyanobacterial toxicity, although measurable by biological, biochemical or analytical assays, cannot be predicted by studying the cyanobacterial population in the bloom. This problem arises by the fact that cyanobacterial populations may be dominated by a single species or be composed of a variety of species, some of which may not be toxic. Even within a single-species bloom there may be a mixture of toxic and non-toxic strains. Likewise, the determination of one toxin group cannot be associated with the presence of a specific species since a given toxin can be produced by several cyanobacterial species. Until now, there are no reports of the simultaneous production of more than one toxin class by the same organism. The genera of cyanobacteria known to produce toxins are represented by *Anabaena, Aphanizomenon, Lyngbya, Nodularia, Cylindrospermopsis, Raphidiopsis, Planktothrix, Oscillatoria* and *Microcystis*.

Cyanotoxins are commonly produced by fresh and brackish water cyanobacteria; however, the production of microcystins by a marine *Synechococcus* of the Salton Sea indicates that cyanotoxins may be a more common occurrence in marine environments (Carmichael & Li, 2006).

Cyanobacteria are a rich source of secondary metabolites, including alkaloids, polyketides and non–ribosomal peptides, and therefore they are of potential biotechnological importance. Unfortunately, many of these secondary metabolites are potent toxins (cyanotoxins) affecting mammals as well as other organisms. Cyanotoxins are released to the water bodies when the blooms collapse and cells are lysed. According to the exerted effect, cyanotoxins are classified into dermatotoxins, cytotoxins, hepatotoxins, and neurotoxins (Sivonen & Jones, 1999). Hepatotoxins and neurotoxins are of great concern since at high concentrations they have caused death (Griffiths & Saker, 2003).

The biological role of cyanotoxins is currently unknown, but several hypotheses have been proposed. One of the first conceptions, positions cyanotoxins as allelochemicals, i.e., molecules involved in inter-specific interactions, such as growth inhibition of competitor organisms (Berry *et al*., 2008). The competitive inhibitor hypothesis has been a matter of debate until today. Works that tend to refute this hypothesis indicate that in order to accomplish the allelophatic effects, concentrations higher than those found in nature were used (Babica *et al*., 2006). The demonstrated allelophatic effect of other (non-toxin) metabolites in *Nodularia*

(Suikkanen *et al*., 2006) and the fact that it is not clear if toxins are actively exported to the environment, a likely prerequisite for allelochemical activity, are also problematic for agree with this hypothesis.

The hepatotoxins microcystins are among the most common and well-studied group of cyanotoxins in freshwater reservoirs. This group of toxins comprises cyclic heptapeptides that are synthesized from five non-protein amino acids and two protein amino acids. The two later aminoacids are variable, giving rise to the more than 70 known microcystin variants (Sivonen & Jones, 1999). Microcystins are generally produced by members of the genera *Microcystis*, *Anabaena*, *Planktothrix*, *Anabaenopsis*, and *Nostoc* although they have also been detected in other genera of marine and terrestrial cyanobacteria such as *Radiocystis*, *Synechococcus*, *Phormidium*, *Plectonema* and *Hapalosiphon*. Of similar structure than microcystins, nodularins are another class of hepatotoxins, found in brackish waters. Among the free-living cyanobacterial genera, nodularin production appears to be specific for *Nodularia* species, although a nodularin analogue (motuporin) was reported by cyanobacterial symbionts in a sponge (De Silva *et al*., 1992).

Cylindrospermopsin (CYN) is a guanidine alkaloid that was first described in *Cylindrospermopsis raciborskii* (Ohtani *et al*., 1992). Additional species producing CYN are *Umezakia natans, Aphanizomenon ovalisporum, A. flos-aquae, Raphidiopsis curvata* and *Anabaena bergii.* Unlike mycrocystins and nodularins, only two analogs of cylindrospermopsin have been reported: 7-epicylindrospermopsin (epiCYN) (Banker *et al*., 2000) and a non-toxic analogue lacking a hydroxyl group on carbon 7 of CYN, deoxy-cylindrospermopsin (doCYN) (Norris *et al*., 1999). The toxic effects of CYN could be confined to the presence of this specific functional group. CYN is considered as a hepatotoxin although its mechanism of action does not involve the inhibition of the eukaryotic protein phosphatases Types 1 and 2A, but instead acts by interrupting the biosynthetic steps leading to cell-reduced glutathione (Runnegar *et al*., 1994). Another harmful effect of CYN is suppression of protein synthesis (Froscio *et al*., 2003). CYN has also been considered as a cytotoxin since livers and kidneys of mice injected showed cell death (Falconer *et al*., 1999). Additional toxic properties of CYN are its purine nature and the presence of potentially reactive guanidine and sulfate groups suggesting its interaction with DNA strands probably causing irreversible DNA damage leading to cancer (Shen *et al*., 2002).

Among all cyanotoxin groups, the most potent cyanotoxins are neurotoxins causing rapid death in mammals by respiratory paralysis. Two groups of neurotoxins are defined according to

their structures and mechanisms of actions (targets) - anatoxins and saxitoxins. Anatoxins are a group of alkaloids produced among a number of cyanobacterial, genera including *Anabaena*, *Oscillatoria, Aphanizomenon, Cylindrospermum,* and *Raphidiopsis*. Two analogues of anatoxins have been described: anatoxin-a and homoanatoxin-a both of which are secondary amines. Anatoxin-a and homoanatoxin-a are potent agonists of the nicotinic acetylcholine receptor (Spivak *et al.*, 1980, Lilleheil *et al.*, 1997). A third anatoxin, anatoxin-a(S), is a phosphate ester of a cyclic N-hydroxyguanine, with pharmacological and chemical properties different from anatoxin-a, acting as an acetylcholinesterase inhibitor (Mahmood & Carmichael, 1986a).

Saxitoxin (STX) and its analogues are a family of tricyclic tetrahydropurine alkaloids that act by blocking voltage-gated sodium channels (Narahashi & Moore, 1968, Llewellyn, 2006). STX is considered to be the basic structure of this group, but modifications give rise to more than 30 structural analogues that can be divided in three larger groups: carbamoyl, N-sulfocarbamoyl and decarbamoyl toxins (Llewellyn, 2006). STX is the most toxic analogue among this group due to its highest affinity for the sodium channel; toxicity measured as LD_{50} in mice (intraperitoneal injection) is 10 μg kg^{-1}, 1000-fold more potent than cyanide. On contrary to the other cyanotoxins, saxitoxins are the only cyantoxins also produced by marine phytoplankton, specifically by dinoflagellates from the genus *Alexandrium* (Cembella, 1998). Saxitoxins are accumulated in filter feeding mussels that feed on dinoflagellates. Human consumption of contaminated mussels produces the syndrome of paralytic shellfish poisoning (PSP), giving the name of saxitoxin and its analogues as PSP toxins (or PSTs). In cyanobacteria, PSTs are produced by the genera *Anabaena, Aphanizomenon, Lyngbya, Planktothirx, Cylindrospermopsis* and recently by *Raphidiopsis* (Yunes *et al*., 2009).

1.2.2 Biosynthetic pathways for cyanotoxins

New bioactive compounds produced by cyanobacteria are been discovered and their structures and mechanisms of action elucidated. However, our knowledge of biosynthetic pathways and the molecular mechanisms associated is slowly increasing. The structure of many cyanobacterial bioactive compounds: cyclic or branched-cyclic organization, polyketide or fatty acid side chains, and the presence of unusual and modified amino acids in small peptides suggested the involvement of nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) in cyanotoxin biosynthesis. The involvement of NRPS on microcystin biosynthesis was

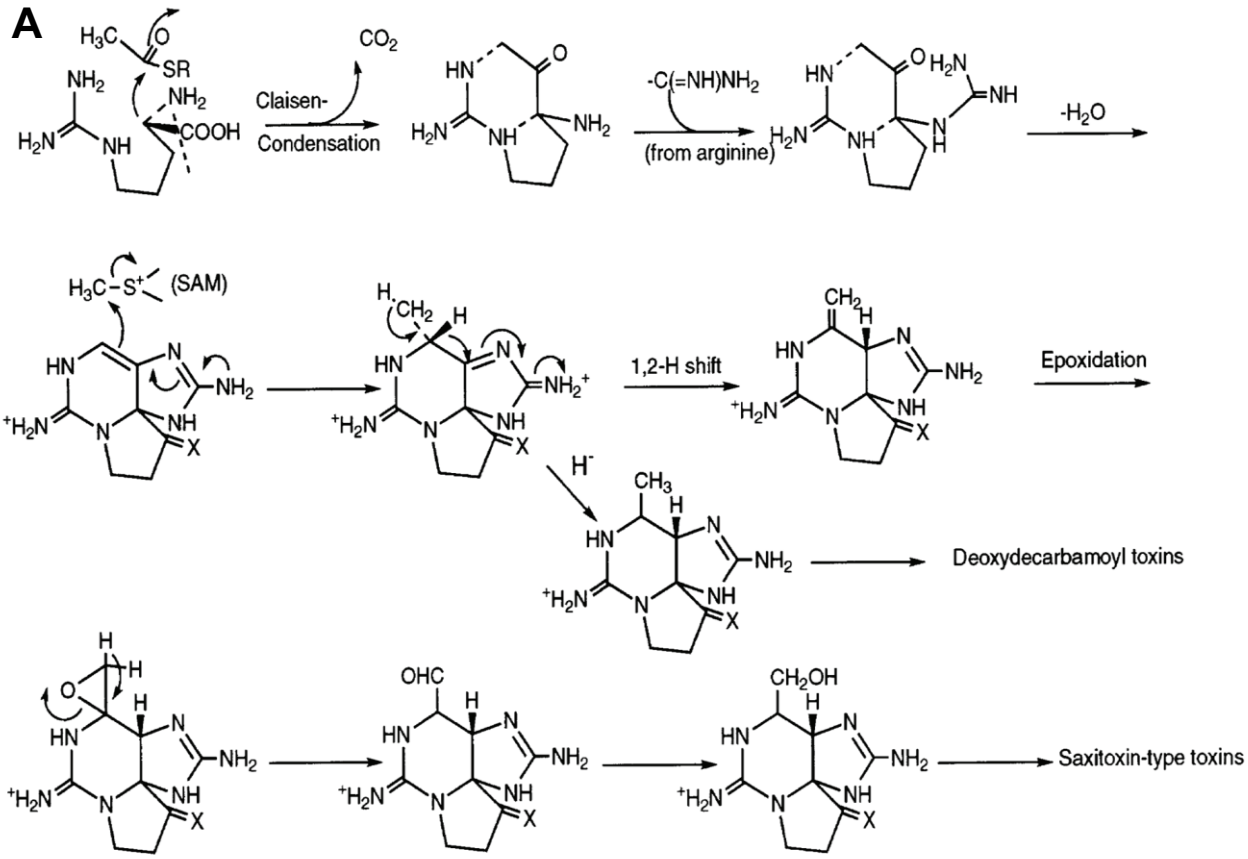
discovered based upon these premises (Dittmann *et al*., 1997); subsequent sequencing of the complete gene cluster (*myc*) in two strains of *Microcystis aeruginosa* revealed the presence of PKS domains in addition to the previously described NRPS, for which the gene cluster for microcystin biosynthesis is formed by a hybrid complex of NRPS/PKS (Nishizawa *et al.*, 2000, Tillett *et al.*, 2000). A similar approach led to the discovery of the nodularin cluster (*nda*) in *Nodularia spumigena.* The *nda* resembles the *myc* cluster in structure and length, comprised of NRPS/PKS; both clusters are flanked by transposases, which may explain the distribution of toxicity (presence of the cluster) among cyanobacterial genera (Moffitt & Neilan, 2004).

The biosynthetic pathway for anatoxin-a was recently described based on the genome sequence of *Oscillatoria* PCC 6506 (Mejean *et al*., 2009). Anatoxin-a is synthesized via a PKS pathway, but so far the genes have not been explored in other anatoxin producing strains.

The gene cluster involved in the biosynthesis pathway of CYN (*cyr*) was described in *Cylindrospermopsis raciborskii* AWT205. The *cyr* gene cluster resembles that of microcystins in that is formed by a hybrid NRPS-PKS complex. New for this gene cluster, are the presence of tailoring enzymes that are not related with the NRPS-PKS complex and a putative toxin transporter (Mihali *et al*., 2008).

The first attempts to determine the biosynthetic pathway of STX were performed in the 1980s. Through culturing studies using stable-isotope labeling, the precursors for STX were proposed as one molecule of acetate, three of arginine and one of methionine in both the dinoflagellate *Alexandrium tamarense* (previously named *Gonyaulax tamarense*) and the cyanobacterium *Aphanizomenon flos-aquae* (Shimizu, 1986). A biosynthesis pathway was proposed involving as the main steps a Claisen condensation between acetate and arginine, an amidino transfer from a second arginine to the formed intermediate and after cyclization, the incorporation of a methyl chain from S-adenosylmethionine (SAM) (Figure 1.3.A) (Shimizu, 1996). Although the structure of STX is highly nitrogenated, this is a relatively small molecule (MW 300) compared with the cyclic peptide toxins (microcystins MW 900-1100; nodularin MW 825). The discovery of the saxitoxin gene cluster in the cyanobacterium *Cylindrospermopsis raciborskii* T3 (Kellmann *et al.*, 2008a) revealed one acyl carrier domain (ACP) -typical for PKS synthases- embedded in a fusion protein containing an aminotransferase domain (SxtA); the remaining genes of the toxin cluster encode for additional non NRPS/PKS proteins. The discovery of this novel PKS modified the old pathway in the order of the methylation reaction. There is no modular biosynthesis of STX, only the first three steps of STX biosynthesis are

catalyzed by the fusion protein SxtA, which can putatively perform a methylation of acetate, and a Claisen condensation reaction between propionate and arginine (Figure 1.3.B). Interestingly, the *sxt* genes have different phylogenetic origins suggesting that several events of horizontal gene transfer (HGT) were involved in the assembly of the *sxt* cluster (Moustafa *et al*., 2009). The *sxt* gene cluster has a mosaic structure that in each toxigenic cyanobacterium studied correlates with its PST profile. The evolutionary time of acquisition of the *sxt* genes has not been estimated. However, if we consider the assembly of the *sxt* gene cluster early in evolution, independent losses could have given rise to non-toxic strains and the differences found among species may have occurred via rearrangements during the course of *sxt*-bearing cyanobacterial evolution. A different scenario involves a recent assembly of the *sxt* cluster followed by rapidly-evolving structural modifications. The endosymbiotic theory for the acquisition of the STX biosynthetic pathway in dinoflagellates is supported only by the first scenario.



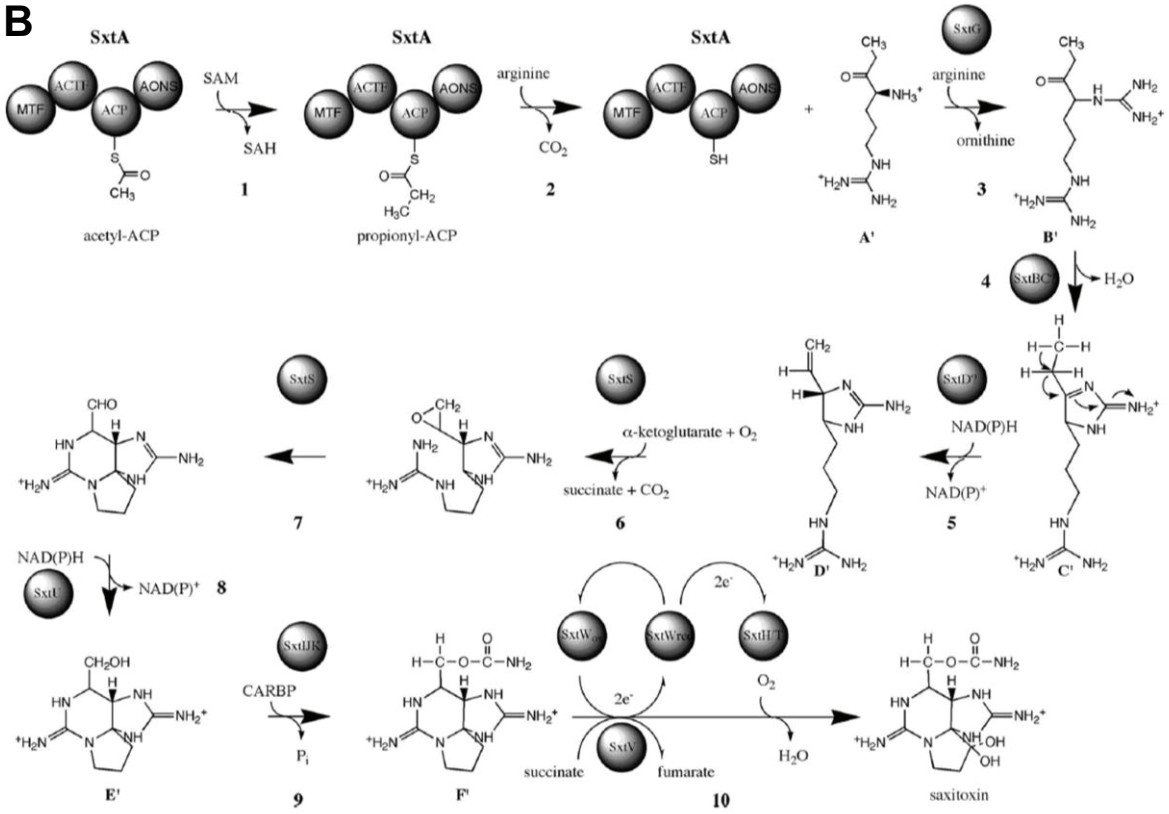


Figure 1.3 Saxitoxin biosynthesis pathways as proposed early by Shimizu (1996) (A) and recently by Kellmann *et al.* (2008a) (B). For more details see text and references.

1.2.3 Regulation of cyanotoxin production

Understanding the factors that affect cyanotoxins production would provide a tool to monitor and prevent the emergence of toxic blooms of cyanobacteria. Production of cyanotoxins in cyanobacteria appears to be constitutive, but evidence suggests that the environmental conditions of a bloom may alter the levels of toxin produced. Various studies have focused on the environmental factors that may influence changes in the production of toxins. Exposing toxic cyanobacteria to different ecological regimes, involving variable temperature, light, nitrogen, iron, phosphate, predators, and other microorganisms, has revealed changes primarily in toxin profile. For example, for the production of microcystins, a correlation between toxin profile and light intensity or after supplementation with aminoacids was observed in the cyanobacterium *Planktothrix agardhii* (Tonk *et al.*, 2005, 2008). The shifts in the toxin profile were accompanied by a modulation of the *mcyA* gene for microcystins biosynthesis. Light was proposed to induce

toxicity by increasing iron uptake in *Microcystis aeruginosa* (Utkilen & Gjolme, 1995). The *mcy* gene cluster in contains a promoter region for a Fur (ferric uptake regulator) regulator, suggesting that the *mcy* genes might be regulated by iron (Martin-Luna *et al*., 2006).

Studies on cylindrospermopsin production have been carried out in two cyanobacterial species, *Aphanizomenon ovalisporum* and *Cylindrospermopsis raciborskii.* Modulation of N sources (ammonium and N deprivation) and light intensities in *C. raciborskii* showed that higher toxin concentrations were produced under N deprivation and at elevated light intensity, where high light cultures grew at similar rates than those grown at moderate light intensities (Saker & Neilan, 2001, Dyble *et al.*, 2006). Toxin measurements were only performed in exponential phase, and gene expression data were not available at that time. In *A. ovalisporum*, starvation for sulfate, phosphorous (P) and N produced a reduction of the cellular CYN content; whereas increasing light intensity resulted in a larger accumulation of CYN in the cells and in the medium (Shalev-Malul *et al.*, 2008, Bácsi *et al.*, 2006). Transcript levels of the CYN biosynthetic genes *aoaA* and *aoaC* (equivalent to *cyrA* and *cyrC* of *C. raciborskii*) were measured in the high light and N starvation experiments. Independent of the amount of toxin recorded, transcript levels decreased in both treatments indicating that toxin production is regulated by additional mechanisms. An ArbB regulator was found to bind the upstream region of the *aoaA* and *aoaC* genes, but no further experiments were carried on to demonstrate the involvement of this transcriptional regulator on CYN biosynthesis (Shalev-Malul *et al*., 2008).

Similar ecological parameters were modulated in *Aphanizomenon flos-aquae* and *C. raciborskii,* both producers of different PST analogues. The amount of produced toxins was not affected in P-limited cultures of *A. flos-aquae*, but the proportion of the different PST analogues changed along with the culture age, towards an increase of the most toxic variant STX. In contrast, N deprivation increased toxin production (cell quota) at the end of the growth phase (Dias *et al*., 2002), but again, in a diazotrophic organism such as *A. flos-aquae*, N limitation is rapidly overcome by N_2 fixation and implications of N limitation at later stages of growth cannot be inferred.

1.3 Cyanobacterial nutrient physiology and nitrogen metabolism

Diazotrophic cyanobacteria, such as the filamentous *Trichodesmium erythraeum* and other unclassified unicellular picocyanobacteria, are of particular importance in oligotrophic oceans

since they are indispensable to maintain nitrogen (N) recycling (Capone *et al.*, 1997, Montoya *et al.*, 2004). Between 160-3,600 μmol N m^{-2} d^{-1} are recycled by *Trichodesmium* spp. N_2 fixation alone in the North Atlantic (Capone *et al.*, 2005, Carpenter & Romans, 1991, Lipschultz & Owens, 1996) and up to 3,955 μmol N m^{-2} d^{-1} was estimated for unicellular picocyanobacteria from the North Pacific (Montoya *et al*., 2004). *Thrichodesmium* blooms occur in tropical and subtropical oceans where denitrification depletes the available N and growth can be limited by phosphorus (P) and Fe (Arrigo, 2005), but they are not found in the more productive waters of estuaries and coastal seas. Most species of filamentous diazotrophic cyanobacteria inhabit fresh and brackish waters where after their decay, they help to maintain the N pool available for non-diazotrophs (Vahtera *et al*., 2007). In freshwaters, as in marine environments, diazotrophs are dominant under P and N limitation, consequently; N limitation in lakes will shift the cyanobacterial population from non-N_2 fixers to N_2 fixers (Schindler *et al*., 2008).

Cyanobacteria can metabolize several forms of nitrogen, inorganic N in the form of ammonium, nitrate or nitrite or organic N as cyanate, urea, and some amino acids, and atmospheric N_2 in the case of N_2 fixers. The strategy followed depends on the environment and reflects the evolutionary adaptation of cyanobacteria to their environment.

Nitrogen fixation is carried on by the nitrogenase enzymatic complex that is irreversibly inactivated by atmospheric oxygen (O_2). Cyanobacteria have thus, evolved various strategies to separate N_2 fixation from O_2-generating photosynthesis. As stated before, cyanobacteria from Subsections IV and V develop heterocysts to protect the nitrogenase and therefore can fix N_2 in parallel with photosynthesis. Unicellular cyanobacteria, incapable of cellular differentiation, separate N_2 fixation and photosynthesis into two different temporal phases in the cell cycle (Mitsui *et al*., 1986). *Trichodesmium erythraeum,* locates the nitrogenase in groups of 3-20 contiguous cells within the filament, called diazocytes (Bergman & Carpenter, 1991, Fredriksson & Bergman, 1997), and reduces photosynthesis when N_2 fixation is at its maximum (Berman-Frank *et al*., 2001).

Ammonium, the most reduced form of N, at high concentrations (mM) enters the cells by passive transport. Oxidized forms of N are enzymatically reduced to ammonium, which is incorporated into carbon skeletons by the glutamine synthetase-glutamate synthase pathway (GS-GOGAT) (Figure 1.4). At low concentration, some N sources are incorporated by active transport into the cells, which in sum make these reactions more energetically costly; thus, ammonium is the preferred source for cyanobacteria. When ammonium is depleted from the medium, several

genes involved in the uptake and reduction of alternative N sources, along with degradation of photosynthetic pigments are activated, and genes for CO_2 fixation are repressed. This cascade of responses, ensuring a balance between N regulation and C:N sensing is known as nitrogen control, where 2-oxoglutarate plays a central role (Herrero *et al*., 2001). Cyanobacteria lack 2-oxoglutarate dehydrogenase, the enzyme responsible for the conversion of 2-oxoglutarate in Succinyl CoA in the tricarboxylic acid cycle (TCA). Hence, the only possible fate of 2-oxoglutarate is its incorporation into ammonium via the GS-GOGAT cycle (Vázquez-Bermúdez *et al*., 2000) (Figure 1.4), which makes 2-oxoglutarate an indicator of the C:N ratios. Nitrogen metabolism in cyanobacteria is thus, regulated internally by the incorporation of ammonium into the carbon skeletons, (i.e. levels of 2-oxoglutarate) instead of the intracellular concentration of ammonium.

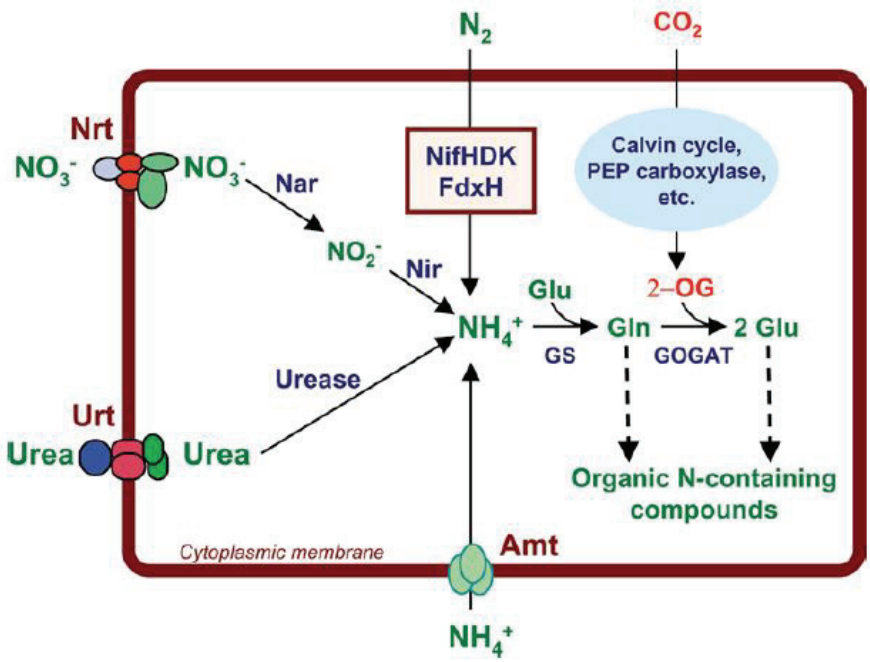


Figure 1.4 Main nitrogen assimilation pathways in cyanobacteria. The scheme was adapted from Flores & Herrero (2005). At low concentration, the different nitrogen sources are incorporated by specific transporters and atmospheric nitrogen (N_2) by diffusion. Combined N sources are metabolized to ammonium, which is incorporated into carbon skeletons through the glutamine synthetase–glutamate synthase pathway. Nitrogen is then distributed from glutamine or glutamate to the other N-containing organic compounds. Nrt, ABC-type nitrate/nitrite transporter; Urt, ABC-type urea transporter; Amt, ammonium permease; Nar, nitrate reductase; Nir, nitrite reductase; NifHDK, nitrogenase complex; FdxH, heterocyst-specific ferredoxin; PEP carboxylase, phosphoenolpyruvate carboxylase; 2-OG, 2-oxoglutarate; GS, glutamine synthetase; GOGAT, glutamate synthase. Nitrogenase and FdxH are boxed to note that in some filamentous cyanobacteria N_2 fixation takes place in heterocysts.

Two central regulators of N control are NtcA and the signaling protein PII. NtcA is a transcriptional regulator for most of the genes involved in N control (Herrero *et al*., 2004). NtcA binds to a conserved box upstream its gene targets; depending on the distance of the binding box to the transcription start point (tsp), gene targets are activated or repressed; although most of the NtcA targets are induced (Luque & Forchhammer, 2008). In the presence of ammonium, *ntcA* is transcribed at basal levels (Luque *et al.*, 1994, Ramasubramanian *et al.*, 1996, Lindell *et al.*, 1998), but upon N-deprivation, the increase in the 2-oxoglutarate levels increases NtcA affinity for DNA through an unknown mechanism triggering the expression (or repression) of NtcA regulated genes (Vázquez-Bermúdez *et al.*, 2002, Tanigawa *et al.*, 2002). The signaling protein PII is the gene product of *glnB*, a target of NtcA. PII proteins are conserved in bacteria, Archaea and plants where they regulate various aspects of N-metabolism. In Proteobacteria it is known that PII reversibly changes from an urydylated to a non-urydylated state in response to glutamine and 2-oxoglutarate (Atkinson *et al.*, 1994). In cyanobacteria, PII is phosphorylated in the presence of 2-oxoglutarate and ATP and coordinates carbon and nitrogen metabolism. Nitrite transport and NtcA dependent gene regulation are known targets of PII in cyanobacteria (Lee *et al.*, 1998, Paz-Yepes *et al.*, 2003).

In heterocyst-forming cyanobacteria, nitrogen control is a more complex process than in non-heterocystous cyanobacteria. More genes are proposed to play crucial roles in the early response to N deprivation, whereby in addition to *ntcA* and *glnB,* the transcriptional regulator *nrrA* and the peptidase *hetR,* known as the master gene of heterocyst differentiation, are also fundamental (Ehira & Ohmori, 2006a, 2006b, Buikema & Haselkorn, 2001). Ammonium-dependent 2-oxoglutarate levels are also important for heterocyst differentiation; however the threshold of activation is different for this process. A first threshold is achieved under growth in an oxidized N-source such as nitrate, in which the levels of 2-oxoglutarate reach a higher level than with ammonium. These levels cause the activation of NtcA and consequently of the genes for nitrate uptake and assimilation; they are, however, not high enough to trigger the expression of *hetR* or of the regulatory cascade leading to heterocyst differentiation that needs a second and higher threshold (high concentrations of 2-oxoglutarate) achieved under N-limitation (Zhang *et al.*, 2006). The expression of *hetR* and *ntcA* is autoregulatory (Ramasubramanian *et al*., 1996) but also mutually dependent and it is this mutual dependence that ensures the *hetR*-dependent transcription of *ntcA* to continue the process of heterocyst differentiation. The newly produced

NtcA will then activate the genes for the stages of the differentiation process (Muro-Pastor *et al.*, 2002).

Heterocyst differentiation occurs in three temporal steps in which the expression of early genes produces the commitment of vegetative cells to form heterocysts and the inhibition of adjacent vegetative cells. The late stages of differentiation include the deposition of the glycolipid and polysaccharide layers and the expression of nitrogenase genes. In *Anabaena* sp. PCC 7120 the commitment to heterocyst formation occurs 8 h after N deprivation (Ehira & Ohmori, 2006b, Buikema & Haselkorn, 2001) and it was shown that this commitment depends on the expression of *hetR* from one of the four transcription start points (tsps) (-271) (Rajagopalan & Callahan, 2010). The presence of several tsps in the promoter region of *hetR* is an additional example of the complexity of this process. The role of *hetR* in non-heterocyst cyanobacteria has so far not been explored.

1.4 Advances in genomics and transcriptomics of cyanobacteria

Advances in cyanobacterial genomic knowledge were first obtained after the release of the genome sequence of *Synechocystis* PCC 6803 (Kaneko & Tabata, 1997). This cyanobacterium became a model organism for the study of photosynthesis and carbon metabolism since is suitable for genetic transformation and grows heterotrophically with glucose as carbon source (Kufryk *et al.*, 2002). However, *Synechocystis* PCC 6803 is not of the same ecological importance as for example, *Prochlorococcus* spp. and *Synechococcus* spp., which are amongst the most abundant phytoplankton in coastal waters and open oceans, respectively. The discovery of different ecotypes of *Prochlorococcus* differing in their light preferences (high- versus low-light adapted) and forming separated phylogenetic clades opened many questions regarding genomic differences underlying their phenotypes (Chisholm *et al.*, 1988, Partensky *et al.*, 1999). Genomic comparison of two high light-adapted (HL) ecotypes that differ by only 0.8% in 16S rRNA sequences revealed the presence of genomic islands, suggesting dynamism in *Prochlorococcus* genomic structure. Completion and further comparison of 8 additional genome sequences has supported the findings based on the previous sequenced genomes and showed that each ecotype phylogroup shares a set of core genes that may be absent in the other, demonstrating the basis of HL and LL adaptations in *Prochlorococcu*s (Kettler *et al*., 2007).

Among filamentous cyanobacteria, the diazotroph *Anabaena* sp. PCC 7120 (heretofore referred as *Anabaena*) develops intercalated heterocysts in a coordinated pattern, in such that a heterocyst is formed at approximately each 10 vegetative cells. Differentiation of a vegetative cell to form a heterocyst occurs, however, only in the absence of combined nitrogen. The possibility of obtaining mutants of *Anabaena* made this cyanobacterium a model for the study of cellular differentiation, pattern formation and N_2 fixation, leading to the sequencing of its genome - the first for a multicellular cyanobacterium.

Another ten genomes of filamentous and almost 50 of unicellular cyanobacteria have since been partially or totally sequenced, allowing phylogenomic comparisons to infer the evolution of this old evolutionary branch. Phylogenomics has allowed deciphering a core cyanobacterial genome of 1,054 protein coding sequences (CDS) and suggested that the cyanobacterial ancestor was a non-N_2 fixing thermophilic microorganism (Mulkidjanian *et al.*, 2006, Shi & Falkowski, 2008). Extraordinarily, comparative genomic analysis has also identified the presence of photosynthetic genes in the genomes of three phages from two viral families distributed among several *Prochlorococcus* clades, suggesting a back and forth transfer between host and phage (Lindell *et al.*, 2004). Transcriptomic studies have shown that many of the differences found at the genomic level are supported by gene expression, whereby phage-encoded photosynthetic genes are transcribed together with phage DNA replication genes in a unit (Lindell *et al*., 2007).

1.5 *Cylindrospermopsis raciborskii* and *Raphidiopsis brookii*

The genus *Cylindrospermopsis* Seenayya et Subba Raju, (1972) belongs to the order Nostocales and is included into the bacteriological code. *Cylindrospermopsis* was described as a tropical organism, with straight and/or curved trichomes and terminal heterocysts. The type species *Cylindrospermopsis raciborskii* was originally described as *Anabaena raciborskii* from a tropical pond in Java, Indonesia (Woloszynska, 1912). *C. raciborskii* is considered as a cosmopolitan highly invasive cyanobacterium because since its first reports in tropical waters identification of this taxon has expanded to include temperate waters (Padisák, 1997), including blooms in temperate regions of northern Europe (Wiedner *et al*., 2007). In Australia, blooms of *C. raciborskii* are commonly represented by CYN-producing strains, although this toxin has also been reported in blooms from northern Europe and south-east Asia (Li *et al.*, 2001a, Wiedner *et*

al., 2007). Brazilian blooms of *C. raciborskii*, in contrast, have only been reported to produce STX and other PSTs (Lagos *et al.*, 1999, Bouvy *et al*., 1999), and non-toxic strains are found scattered worldwide. Unlike most cyanobacteria, *Cylindrospermopsis* does not form scums on the surface of the water; instead is present in highest concentrations below the surface from 1 to 3 m depth (St. Amand, 2002). *C. raciborskii* has been found in highly eutrophic lakes and reservoirs, although it seems to have a high affinity for phosphate uptake since it still dominates the phytoplankton community in low phosphate lakes (Istvánovics *et al.*, 2000). Higher phosphatase activity was also shown in comparison with *Microcystis* and *Aphanizomenon* indicating a greater capacity for metabolizing organic phosphorous (Wu *et al*., 2009).

Blooms of *C. raciborskii* are often associated with *Raphidiopsis* species. The number of works reporting the co-occurrence of these two cyanobacteria has increased in the last few years (Mohamed, 2007, Li *et al.*, 2008); this implies that the nutrient requirements and metabolism are similar between both genera. Studies of *Raphidiopsis* species are yet emerging; blooms have been found in China, Egypt and Brazil (Mohamed, 2007, Li *et al*., 2008) but there is not much knowledge on *Raphidiopsis* physiology. Recently, a new strain of *Raphidiopsis brookii* grown at three nitrate concentrations was shown to constitutively produce PSP toxins (Yunes *et al.*, 2009). Three classes of toxins are produced among different *Raphidiopsis* species, anatoxin-a, CYN and the most recent discovery of the production of PSP toxins by *R. brookii*.

The genera *Cylindrospermopsis* and *Raphidiopsis* are morphologically similar. Both can present coiled or straight trichomes, and develop akinetes in a single unit or adjacent pairs. However, *Raphidiopsis* does not develop heterocysts (Figure 1.5). *Cylindrospermopsis* and *Raphidiopsis* were taxonomically separated based on this morphological criterion (Rippka *et al.,* 1979). However, these two genera cannot be discriminated at the molecular level (Gugger *et al*., 2005). *Raphidiopsis* resembles *C. raciborskii* in the morphological stage of a trichome without heterocysts and it has been suggested that *R. mediterranea* is only a growth stage of *C. raciborskii* (Moustaka-Gouni *et al.*, 2009). This background indicates that there is an unclear taxonomical definition of *C. raciborskii* and *Raphidiopsis* species, thereby posing an interesting model for the study of evolution of toxin phenotype, cellular differentiation and nitrogen metabolism.

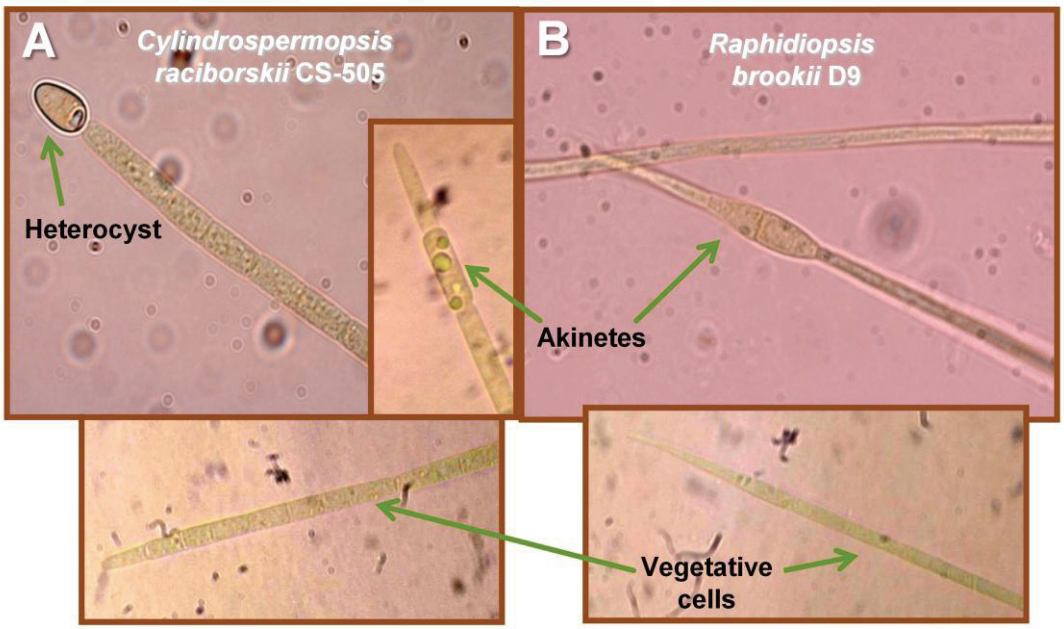


Figure 1.5 Light microscopy of the two filamentous cyanobacteria studied in this thesis. *Cylindrospermopsis raciborskii* CS-505 **(A)** and *Raphidiopsis brookii* D9 (**B**). The different cellular types are indicated: photosynthetic or vegetative cells, heterocyst and akinetes. Note the pointed end of *R. brookii* D9 filament; this structure is also observed in young filaments of *C. raciborskii* before the development of terminal heterocysts.

1.6 Taxonomic considerations

In this thesis much of the focus is directed towards the genomic and transcriptomic comparison between *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.In this context it is therefore important to describe inconsistencies in the nomenclature of *R. brookii* D9 that were deciphered in the course of this thesis work. *R. brookii* D9 was originally isolated from a mixed multiclonal culture dominated by a strain designated as *Cylindrospermopsis raciborskii* SPC338 isolated from the Taquacetuba arm of the Billings freshwater reservoir in Sao Paulo, Brazil. Since it was supposed that the mixed culture contained different strains of *C. raciborskii*, the new clone was initially called *C. raciborskii* D9. During the preparation of **Publication I**, phylogenetic analysis based on 16S rDNA showed that *C. raciborskii* D9 grouped with *Raphidiopsis* strains and during the culturing time heterocysts were not observed (as would be typical for *Raphidiopsis*). These arguments led to the conclusion that the clone belonged to the genus *Raphidiopsis*, although it was unclear to which species. As a consequence, *C. raciborskii* was renamed as *Raphidiopsis* sp. D9. In 2009, a Brazilian report on all cyanobacterial species

isolated during blooms in that country claimed that the mixed culture of *C. raciborskii* received from Brazil, corresponded instead to *Raphidiopsis brookii* SPC339. Based largely upon morphological analysis, this clarifies a mistake in the nomenclature and the isolate was renamed *Raphidiopsis brookii* D9.

1.7 Thesis motivation

The species *Cylindrospermopsis raciborskii* is of particularly high ecological importance due to its worldwide expansion which in many cases positions it as the dominant species. Hence, *C. raciborskii* is known as an invasive, highly competitive species of potential toxicity, comprising strains able to produce either PSTs or cylindrospermopsin, as well as non-toxigenic strains. One of the most interesting aspects of this species is that different morphologies are observed within the growth stages (Singh, 1962). All stages previous to the development of the heterocysts resemble the morphology of the genera *Raphidiopsis*, causing problems to differentiate and isolate both species. Moreover, the differences at the 16 rRNA sequence level among *C. raciborskii* CS-505 and *R. brookii* D9 are close to 0.5%, resembling the differences between different ecotypes of *Prochlorococcus*. If *R. brookii* D9 is seen as an evolutionary stage of *C. raciborskii,* differing only in the diazotrophy and heterocystous phenotypes, a genomic comparison would reveal the extent of genetic similarities among them. Furthermore, both species are ideal models for the study and separation of the whole regulatory cascade involved in cell differentiation and nitrogen-stress response that is triggered only under N deprivation.

Another strong aspect favoring a comparative study of these two cyanobacteria is the evolution of toxin biosynthesis. PSTs are the only toxins produced in both marine dinoflagellates and cyanobacteria; this fact is one of the strongest arguments for implying the acquisition of the PSP biosynthesis pathway by dinoflagellates through endosymbiosis of cyanobacteria. In contrast to the complexity of the dinoflagellate genome (3,000 – 215,000 Mbp), the simplicity of cyanobacterial genomes offers an alternative model for understanding the pathways involved in toxin production and gene regulation.

The strains *C. raciborskii* CS-505 and *R. brookii* D9 were chosen as models for the study of toxin production, cell differentiation and evolution of *C. raciborskii* due to the fact that they differ in their morphology, geographical origin and toxin profile while being phylogenetically

closely related. In addition, preliminary estimations by pulsed field gel electrophoresis (PFGE) indicated a small genome size (about 3 Mb) which made both strains good candidates for genome sequencing. Since genetic diversity was observed in two closely related unicellular *Prochlorococcus* ecotypes, the opportunity to explore the genomic differences of these two cyanobacteria presenting a multicellular phenotype was a major motivation.

1.8 Outline of the Thesis

This thesis is organized into four core chapters corresponding to four publications where the candidate is first author and one separate chapter where the candidate is coauthor.

Toxin production in cyanobacteria is an inconsistent phenomenon in which only certain strains within a species produce toxins. Likewise, several groups of toxins may be produced within a species. In **Publication I**, the correlation between the phylogenetic affiliation of *C. raciborskii* strains and their toxigenic phenotype was explored. In order to achieve an improved discrimination between strains, phylogenetic inference was performed by means of two approaches. The first approach involved a combined analysis of the sequence of a polymorphic region, the internal transcribed spacer (ITS) of the 16S rDNA. The second made use of a general genotyping technique, usually applied to distinguish among pathogenic bacteria, generally clinical isolates that present only small genetic differences. Restriction analysis by pulsed field gel electrophoresis (PFGE) is hence highly discriminatory. With the latter analysis it was possible to estimate genome sizes, number of ribosomal operons and to detect the presence of putative CYN biosynthetic genes in CYN-producers. The candidate laid out the experimental study and performed DNA isolation, amplifications, sequencing analysis, optimized and performed PFGE analysis. Phylogenetic analyses were supported by M.M. and Southern hybridizations were prepared by A.M. The candidate analyzed the data and wrote the paper in collaboration with A.M. and M.V.

The large phenotypic differences between *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9, while being phylogenetically closely related (as inferred by ribosomal and whole genome phylogeny in **Publication I**, recommended both strains as candidates for genome sequencing. In **Publication II**, a comparative analysis of the genomes of *C. raciborskii* CS-505 and *R. brookii* D9 was performed, emphasizing in the unique and common genes of each

strain and the implications of their presence in the phenotypic differentiation of the strains. Additionally, this work described the smallest genomes known for filamentous cyanobacteria allowing a deep analysis into the evolutionary origins of multicellularity, diazotrophy and heterocyst formation. Finally, the genomic structure and plasticity supporting the need for a taxonomic reclassification of the genus *Raphidiopsis* into *Cylindrospermopsis* was discussed. Within this study, the candidate was responsible for the isolation of the DNA employed for sequencing, producing a fosmid library and sequencing the ends of fosmids to aid in the process of genome closure. She led the genome annotation of unique genes, data analysis and wrote the manuscript in collaboration with the co-authors.

The genetic structure of the PSP toxin gene cluster in *R. brookii* D9 (*sxt* cluster) supported the pathway proposed by Kellmann *et al*. (2008a) for the biosynthesis of saxitoxin in *C. raciborskii* T3. The *sxt* gene cluster of D9 is structurally different from that of T3 but in both strains, the genetic structure of the *sxt* cluster correlates with the different toxin profiles produced by each strain. Therefore, a comparative analysis among these two strains and other PST-producing cyanobacterial species would aid in the understanding of PST biosynthesis among cyanobacteria. One of the aims of **Publication III** was to compare the genetic structure with the toxin profile in *C. raciborskii* T3 and *R. brookii* D9, with emphasis on the specific genes encoded by the D9 *sxt* cluster. Within this aim, the candidate analyzed the data and wrote a segment of the manuscript in cooperation with the lead author and co-authors.

Among the phenotypic differences between CS-505 and D9 are the development of heterocysts (and consequently of diazotrophy) only in CS-505, and the production of two different groups of toxins by each strain. Genomic data suggested that both gene clusters for the biosynthesis of the toxins might be N-regulated. Considering the significance of N in the physiology of the strains, in **Publication IV** the responses of CS-505 and D9 in toxin production and gene expression, to alternative N sources were presented. With the availability of the genome sequences of CS-505 and D9, a dual whole-genome microarray was designed in order to perform comparative transcriptomics. In this work, the gene expression data were considered with an integrative view of the genomic differences found for **Publication II**, and the role of toxins in the cyanobacterial metabolism were discussed. The candidate designed the experiments in discussion with U.J and M.V. She performed all laboratory experiments for the molecular analysis,

functional genomics and analytic chemistry. She analyzed the data and wrote the paper in consultation with the co-authors.

Although **Publication IV** gave an overview of the main processes occurring in CS-505 and D9 when grown under a specific N source, the information was restricted to a single point in the growth curve and therefore it was not possible to explore the dynamics of gene expression. With this aim, in **Publication V** gene expression was analyzed in a time series of up to 48 h under growth with selected N sources. The patterns of gene expression leading to the development of a terminal heterocyst in CS-505 were discussed based upon data supported by experimental evidence on the evolution of N_2 fixation in cultures grown under three different N sources. Additionally, the dynamics of gene expression of CS-505 were compared with that of D9 in order to get insights into the gene regulation in these two filamentous and multicellular cyanobacteria that also present a small genome. The candidate designed the experiments in discussion with U.J and M.V. and carried on growth and N_2 fixation measurements, RNA isolations, cDNA synthesis, and microarray hybridizations. She performed all laboratory experiments for the molecular analysis, functional genomics. She analyzed the data with support of R.G. The candidate wrote the paper in consultation with the co-authors.

2. PUBLICATIONS

Publication I

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

Publication II

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. doi:10.1371/journal.pone.0009235.

Publication III

Soto-Liebe K., Murillo AA., Krock B., **Stucken K**., Fuentes JJ., Trefault N., Cembella A., Vásquez M. Reassessment of the toxin profile of *Cylindrospermopsis raciborskii* T3 and function of putative sulfotransferases in synthesis of sulfated and sulfonated PSP toxins. Submitted to Toxicon.

Publication IV

Stucken K., John U., Cembella A., Soto-Liebe K. and Vásquez M. Effects of alternative nitrogen sources on gene expression and toxin production in the closely related cyanobacteria *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9. Submitted to Molecular Microbiology.

Publication V

Stucken K., John U., Gutierrez R., Glöckner G., Vásquez M., and Cembella A. Gene expression dynamics and heterocyst differentiation in *Cylindrospermopsis raciborskii* versus *Raphidiopsis brookii*, a non-diazotroph closely related species. To be submitted to Journal of Bacteriology.

2.1 Publication I

Toxicity phenotype does not correlate with phylogeny of *Cylindrospermopsis raciborskii* strains

Karina Stucken, Alejandro A. Murillo, Katia Soto-Liebe, Juan J. Fuentes-Valdés, Marco A. Méndez, Mónica Vásquez

2.1.1 Abstract

Cylindrospermopsis raciborskii is a species of freshwater, bloom-forming cyanobacterium. *C. raciborskii* produces toxins, including cylindrospermopsin (hepatotoxin) and saxitoxin (neurotoxin), although non toxin-producing strains are also observed. In spite of differences in toxicity, *C. raciborskii* strains comprise a monophyletic group, based upon 16S rRNA gene sequence identities (greater than 99 %). performed comparative phylogenetic analyses of 16S rRNA gene and 16S-23S rRNA gene internally transcribed spacer (ITS-1) sequences, and genomic DNA restriction fragment length polymorphism (RFLP), resolved by pulsed-field gel electrophoresis (PFGE) of strains of *C. raciborskii*, obtained mainly from the Australian phylogeographic cluster. Our results showed no correlation between toxic phenotype and phylogenetic association in the Australian strains. Analyses of the 16S rRNA gene and the respective ITS-1 sequences (long L, and short S) showed an independent evolution of each ribosomal operon. The genes putatively involved in the cylindrospermopsin biosynthetic pathway were present in one locus and only in the hepatotoxic strains, demonstrating a common genomic organization for these genes and the absence of mutated or inactivated biosynthetic genes in the non-toxic strains. In summary, our results support the hypothesis that the genes involved in toxicity may have been transferred as an island by processes of gene lateral transfer, rather than convergent evolution.

2.1.2 Introduction

Cylindrospermopsis raciborskii Seenayya et Subba Raju (Woloszynska, 1912) is defined as a planktonic, filamentous, nitrogen-fixing, terminal heterocystous, freshwater, bloom-forming

cyanobacterium. Strains of *C. raciborskii* present two morphologies, with straight or coiled trichomes (Table 2.1.1), and produce either the hepatotoxin cylindrospermopsin (CYL), a potent protein synthesis inhibitor (Griffiths & Saker, 2003), paralytic shellfish poisoning (PSP) toxins (Lagos *et al*., 1999), or do not produce toxins. Since its first report from in the tropical island of Java (Indonesia) (Woloszynska, 1912), blooms of *C. raciborskii* have been reported in diverse countries of tropical, as well as temperate conditions (Padisák, 1997). Besides this temperature dependence, *C. raciborskii* can grow under different nutrient conditions, with urea, nitrate or ammonium, as well as by fixation of atmospheric nitrogen (Fuenzalida, 2005, Saker & Neilan, 2001), and at low phosphate concentrations (Istvánovics *et al*., 2000), making water supplies favorable places for the growth of this cyanobacterium. The occurrence of a bloom of *C. raciborskii* in a water supply in Palm Island, Australia was associated with the first case of CYL human intoxication (Bourke *et al*., 1983), which caused concern among health authorities. Despite the differences in morphology and growth conditions, toxicity and geographical distribution, the 16S rRNA gene sequences of strains of *C. raciborskii* are homogeneous. The genotypic variation between strains isolated from several geographic regions, showed 99.1 % sequence identity in the 16S rRNA genes (Neilan *et al*., 2003). Similarly, studies of strains isolated in Australia exhibited not less than 99.8% sequence identity, demonstrating that *C. raciborskii* is a well-defined species (Neilan *et al*., 2003). A phylogeographic relationship is apparent among *C. raciborskii* strains analyzed so far, although correlations between phylogenetic association (using markers from complete genome sequences) and toxigenic phenotypes have not been thoroughly explored.

Pulsed-field gel electrophoresis (PFGE) is a genomic tool widely used in epidemiological analyses of pathogenic bacteria, due to its high discriminatory power (Bisgard *et al*., 2001). PFGE restriction patterns may allow for inferring phylogenetic relationships among closely related strains. In the case of cyanobacteria, PFGE has been useful for discerning relevant information on genome size and for constructing physical and genetic maps of cyanobacterial chromosomes (Chen & Widger, 1993). Nevertheless, to date, no PFGE studies have been performed on *C. raciborskii*. We investigated genetic variability and genome organization in *C. raciborskii* strains, mainly in Australian phylogeographic groups, and analyzed whether the toxic phenotype is correlated with the phylogenetic relationships.

The goal of this study was to determine whether the presence of toxin genes in *C. raciborskii* is the result of convergent evolution or lateral gene transfer (LGT). The first step corresponded to a phylogenetic analysis based on 16S rDNA plus the 16S-23S rRNA gene

internally transcribed spacer (ITS-1) sequence and restriction fragment length polymorphism (RFLP) of the complete genome resolved by PFGE. The second step was to explore, by Southern blot hybridization, the number and the organization of 16S rDNA genes and to detect in toxic and non-toxic strains the presence of three genes: *aoaA*, *aoaB*, and *aoaC* encoding an amidinotransferase, a hybrid non-ribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) and a PKS, respectively, putatively involved in the CYL biosynthetic pathway (Kellmann *et al*., 2006).

2.1.3 Results

16S rDNA plus 16S-23S rDNA ITS amplification and sequence analysis

Previous studies have described the inclusion of the 16S-23S rRNA gene ITS (ITS-1) region in phylogenetic analyses, effectively increasing the levels of discrimination for strains of a species (Gugger *et al*., 2005). When we analyzed the strains of this study by PCR-amplification primers specific for cyanobacteria (Table 2.1.2), two amplification products were observed (Figure 2.1.S1). All strains of *C. raciborskii* showed ITS-1 of approximately 430 and 630 bp, except for *C. raciborskii* strain D9, which exhibited ITS-1 regions of 400 and 600 bp. The sizes of the PCR-amplified products corresponded to the lengths of the ITS-1 plus 150 bp from 16S rRNA genes (the 30 end) and 50 bp from 23S rRNA genes (the 50 end). Therefore, the sizes of the ITS-1 regions, ITS-1(S) and ITS-1(L), of Australian strains of *C. raciborskii* were calculated to be 238 and 394 bp, respectively. The lengths of ITS-1 regions of the Brazilian strain D9 were 174 and 338 bp, with 20 bp missing at the 30 end of the sequences. Basically, the size differences among the ITS-1 regions are due to the presence of $tRNA^{Ile}$ and $tRNA^{Ala}$ sequences in the longer one (Figure 2.1.1), as was shown in a previous report (Gugger *et al*., 2005). Previous analysis of the ITS-1 regions in *C. raciborskii* indicated a pattern of only the two tRNAs, although numerous short domains that are conserved in cyanobacteria are also present in the strains of this study. As is shown in Figure 2.1.1, the following domains are present: D1, D10, D3, D2, D4 and D5, required for correct folding of rRNA transcripts (Iteman *et al*., 2000), the V3 stem loop and, boxB-boxA anti-terminators, most of them homologous to those observed in *E. coli*. In the case of *A. circinalis* strain ACMB13 (used as an out-group in the phylogenetic analysis), three amplified fragments were observed: approximately 600, 700 and 800 bp (Figure 2.1.S1),

suggesting a minimum of two ribosomal operons in *C. raciborskii* and three in *A. circinalis* strains.

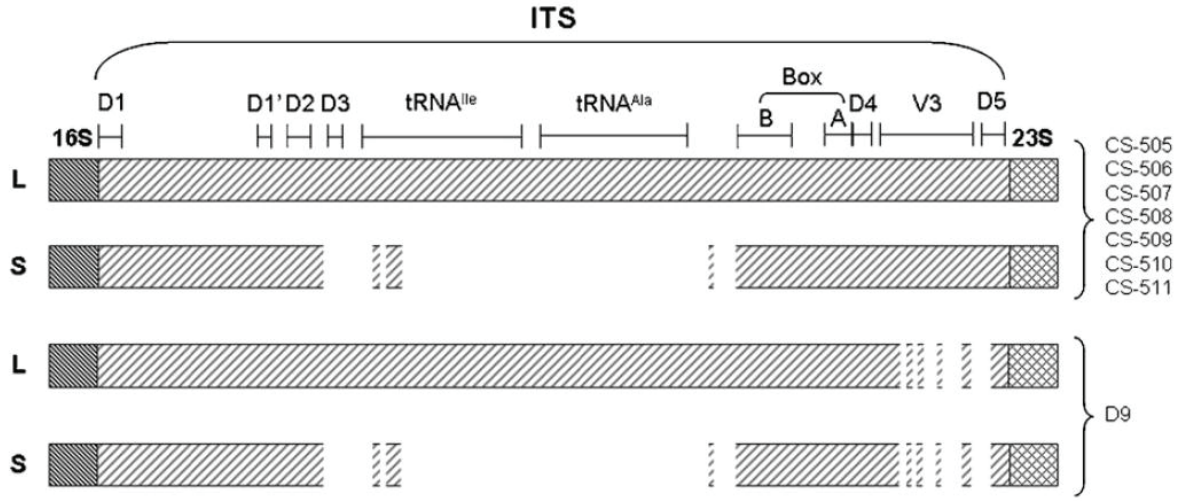


Figure 2.1.1 Schematic view of alignments for ITS-1(S) and (L) sequences from D9, CS-505, CS-506, CS-507, CS-508, CS-509, CS-510 and CS-511 strains of *C. raciborskii*. The conserved domains D1, D10, D2, D3, D4 and D5, the V3 stem, boxA and boxB anti-terminators are indicated. The sequence of strain K-12 of *E. coli* was used to define the positions of conserved domains. The 3' -fragment of the 16S and 5' -fragment of the 23S rRNA gene sequences are flanking the ITS-1 region. The white spaces correspond to sequence gaps in the ITS-1(S) compared with ITS-1(L).

Phylogenetic analysis based on 16S rRNA genes and associated ITS regions

Unlike other studies (Iteman *et al*., 2000, Neilan *et al*., 2003) in which the 16S rRNA gene or ITS-1 region sequences were considered, this study utilized both the 16S rRNA gene sequence and the ITS-1 region sequence for the analyses. The idea was to complement the information contained in both sequences for more robust resolution, through PCR amplification, using suitable primers (Table 2.1.2), cloning and sequencing of the PCR-amplification products of each type of ribosomal operon from all *C. raciborskii* and *A. circinalis* ACMB13 strains. The analyses revealed high sequence identities between Australian strains of *C. raciborskii* (99.9%), although the values decreased to 99.8% when Brazilian strain D9 sequences were included in the analysis (data not shown). When the sequences of 16S rRNA genes, coupled with the associated ITS-1 regions, were analyzed, the intra-strain variation between 16S rRNA gene sequences and the congruent ITS-1(S) or ITS-1(L) was observed to be greater than the variation between sequences from the different *C. raciborskii* strains used in this study. In each analysis, congruence was observed between the NJ and MP trees obtained. Furthermore, these analyses exhibited different

topologies depending on the gene analyzed, ITS-1 (S) or ITS-1 (L). These results suggested that the two operon sequence types have evolved independently (Figure 2.1.S2). The ITS-1 (L) sequences from the strains of this study were compared with the ITS-1 sequences of other cyanobacteria, present in the GenBank data base (see Materials and Methods). As shown in Figure 2.1.2, the strains of *C. raciborskii* of this study clustered with the three Australian strains, but split into two evolutionary branches. No correlation was observed between the toxin phenotypes and the phylogenetic relationships. The D9 strain clustered with other Brazilian strains, supporting the previous results on the phylogeographic clustering of *C. raciborskii* strains (Neilan *et al.,* 2003).

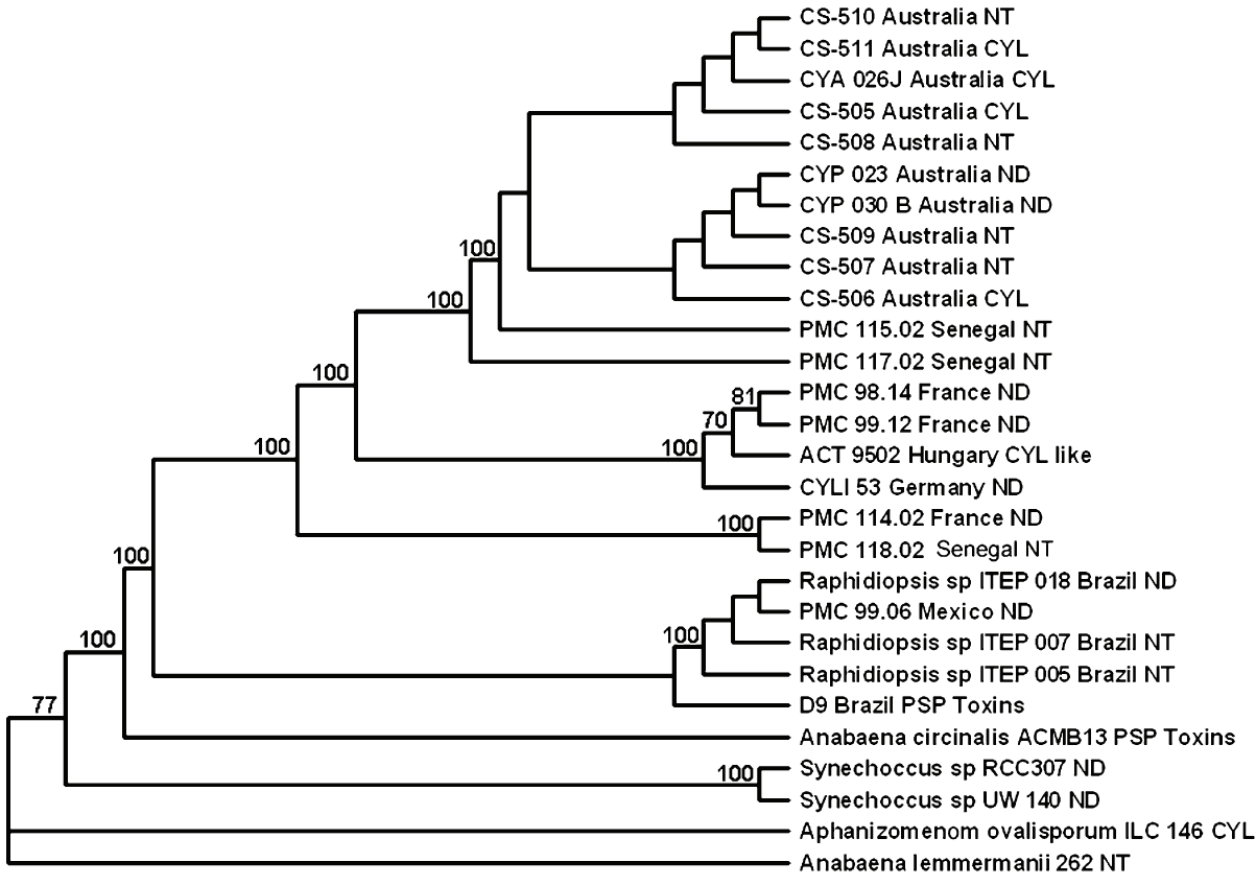


Figure 2.1.2 Phylogenetic inference of *C. raciborskii* strains based on 16S-23S rDNA ITS-1(L) sequence comparisons. The dendrogram was inferred with the NJ method using p distance. The values over the nodes represent the percentage bootstrap values (1000 resamplings). The distribution of the toxin type is shown with the following abbreviations: CYL: cylindrospermopsin; NT: non-toxic; PSP: paralytic shellfish poisoning and ND: not described.

Phylogenetic analysis based on RFLP resolved by PFGE

The two RFLP analyses performed with *Mlu* I and *Xho* I and resolved by PFGE allowed the establishment of a distance matrix based on the presence or absence of individual restriction fragments for each strain. The cluster analysis (Figure 2.1.3) obtained with RFLP–PFGE data showed one branch, including the non-toxin producing strains CS-507 and CS-508, and a separate branch, including the non-toxin-producing strain CS-510 and CS-511, a CYL-producing strain (the analysis was supported with the maximum bootstrap value, 100%). The CS-505 strain, a CYL-producer, grouped with the CS-510 and CS-511 branch, although distant from CS-506, another CYL-producing strain. These results agree with the phylogenetic inference based on the 16S rRNA gene plus ITS-1 sequences, as well as a phylogenetic inference based upon ITS-1(L) sequence data, alone. This demonstrates that even when the complete genome is considered in the analysis, such as by RFLP–PFGE analysis of genomic DNA, there is no correlation between phylogenetic relationship and toxic phenotype.

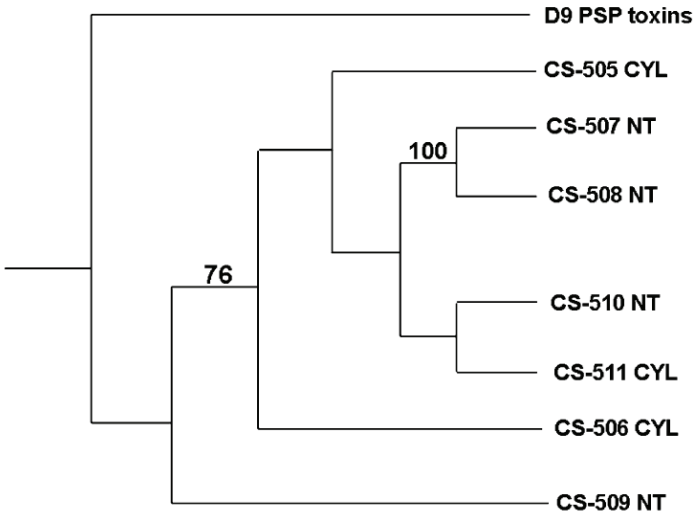


Figure 2.1.3 Phylogenetic inference based on the genomic DNA RFLP profile resolved by PFGE, using *Xho* I and *Mlu* I enzymes. A distance matrix was constructed, based on the presence or absence of restriction fragments in each PFGE gel. The MP algorithm was used (Length: 195 steps, IC: 0.5744; IR: 0.4503). The values over the nodes represent the percentage bootstrap values (1000 pseudoreplicates), only values over 70% are shown. The distribution of toxicity type is shown using the following abbreviations: CYL: cylindrospermopsin; NT: non-toxic; PSP: paralytic shellfish poisoning.

Genomic distribution of the 16S rRNA gene

Although the 16S rRNA gene sequence is generally uninformative at the strain level, the genomic distribution of the rRNA gene sequence can provide information about the phylogenetic association among strains of species within the same genus (Iteman *et al*., 2000). RFLP analysis with *Xho* I digestion were selected for Southern blot hybridizations of 16S rRNA sequences (amplified with the cyanobacterial-specific primers CYA106F/CYA781R (Table 2.1.2)) since this enzyme did not cut within the 16S rRNA gene sequences of the strains assayed. The results shown in Figure 2.1.4 demonstrated that all strains have at least three 16S rRNA gene sequence types and, therefore, probably three ribosomal loci. The distribution of this sequence type is not exactly the same but in fragments with very similar sizes among all the Australian strains. The differences were slight suggesting a common genomic organization, although genome sequencing is the only definitive test. The D9 strain showed a different pattern, with one of the hybridizing loci in a different restriction fragment of approximately 80 Kb (Figure 2.1.4). These results correlated with all previous phylogenetic analyses in which strain D9 has been found to be distantly related to the Australian strains.

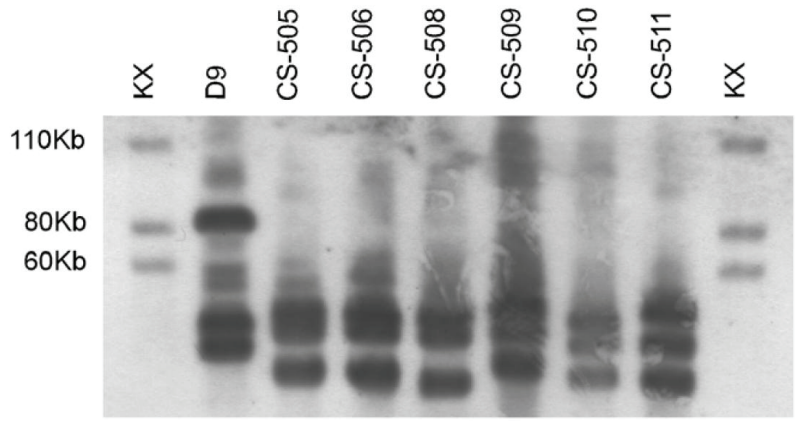


Figure 2.1.4 16S rRNA Southern blot hybridizations. The hybridization probe corresponds to a PCR product of 663 bp of the 16S rRNA gene 5' region obtained with primers CYA106F/ CYA781R. *V. parahaemolyticus* strain KX digested with *Not* I restriction enzyme served as a molecular weight marker.

Genes putatively involved in CYL biosynthesis

Given the non-correlation of toxic phenotype and phylogeny, the genes putatively involved in the CYL biosynthetic pathway (Kellmann *et al*., 2006) were analyzed to determine whether they are present in the non-toxic *C. raciborskii* strains. First, the *aoaA*, *aoaB* and *aoaC* genes, encoding an amidinotransferase, a hybrid NRPS/PKS and a PKS, respectively, were amplified by PCR with the primers CYLATF/CYLATR, CPSF/CPSR and A205PKF/A205PKR (Table 2.1.2; Figure 2.1.S3). A product of the expected size for the *aoaC* gene was obtained from the three CYL-producing strains (CS-505, CS-506 and CS-511) and only non-specific amplification was observed in other strains. In the case of the *aoaB* gene sequence, the CYL producers showed the expected product size (strong band) and a weak amplification of the estimated product size observed from the non CYL producers. The presence of the CYL biosynthetic genes in one or more than one locus in the genome of a strain was confirmed, and also checked as to whether the genes could also be present in non-toxic strains, with the amplified product of each gene as a hybridization probe in a Southern blot analysis (Figure 2.1.S4). The three probes hybridized specifically to the CYL-producing strains and only in one unique position, in the same size range (110 Kb), thus demonstrating that the three genes are conserved within CYL-producing strains. The three genes were observed to be located close to each other in the genomes, suggesting a cluster organization of the toxic genes in the genomes of CYL-producing strains. The cluster, with all the genes required for CYL biosynthesis, regulation, and export of the toxin, has been estimated to be approximately 43 Kb in size (Mihali *et al*., 2008) and can be completely contained in the 110 Kb hybridization fragment. The complete CYL biosynthesis cluster may be hypothesized to be present only in CYL-producing strains and absent in non-producing ones, suggesting a loss or acquisition of large genomic fragments, rather than gene mutation or an inactivation process. This hypothesis is supported by the lack of correlation between phylogenetic relationships and toxic phenotype in *C. raciborskii* species.

2.1.4 Discussion

The genetic variation in 26 *C. raciborskii* strains isolated from freshwater rivers and reservoirs from Europe, North and South America, based on 16S rRNA gene sequence comparisons, has already allowed for delineation of phylogeographic clustering for this species (Neilan *et al.*, 2003). As we show in the current study, estimation of the phylogenetic relationships based only on 16S rRNA gene sequence analysis is just a first approach. Neilan *et*

al. (2003) did not analyze the correlation between toxin phenotype and phylogeny with consideration of the whole genome and additional genetic markers. We improved the resolution of the phylogenetic analysis for *Cylindrospermopsis* species (Table 2.1.1), by employing the 16S rRNA gene sequences associated with the contiguous 16S-23S rRNA gene ITS-1 region sequences, although relatively little information about ITS regions in cyanobacteria has yet been accumulated. The data of this study suggest that the number of ITS-1 regions is variable among different genera of cyanobacteria (Boyer *et al.*, 2001, Iteman *et al*., 2000).

In several cyanobacterial genera, such as *Anabaena, Aphanizomenon, Anabaenopsis, Nostoc, Cylindrospermopsis* and *Nodularia*, the number of ITS-1 regions varies between 1 and 4 (Iteman *et al*., 2000, Neilan *et al*., 1997). Among cyanobacteria whose genomes are sequenced, either four or five rRNA operons have been reported (http://rrndb.cme.msu.edu). Previous reports and the results of this study including different strains have shown that *C. raciborskii* contains two ITS sequence types, ITS-1(L) and ITS-1(S) that differ mainly in the inclusion of $tRNA^{Ile}$ and $tRNA^{Ala}$ sequences, a common feature among cyanobacteria. This variability of the occurrence of tRNA genes, together with the structural domains, D1, D10, D2, D3, D4, D5, the V3 stem loop and boxB-boxA anti-terminators, are clearly more phylogenetically informative than the 16S rRNA gene sequence alone and are useful for discriminating higher taxonomic levels (Wang *et al*., 2003). Accordingly, the use of the 16S rRNA gene sequence plus 16S-23S rRNA gene ITS-1(L) and 16S rRNA, plus 16S-23S rRNA gene ITS-1(S), enriches the phylogenetic analysis for our *C. raciborskii* strains. The phylogenetic trees generated for each group of sequence types showed different topologies (Figure 2.1.S2), suggesting a divergent evolution for the different ribosomal operons in the same strain. This pattern has been observed in prokaryotes with two or more rRNA operons when these operons are separated by as much as 7Kb, reducing recombination rates by two orders of magnitude (Lovett *et al*., 1994) and allowing the operons to escape homogenization via homologous recombination and more likely to diverge (Hashimoto *et al*., 2003). This phenomenon could be operating in the *C. raciborskii* strains analyzed in the present study.

The ITS-1 (L) sequences of *C. raciborskii* strains in this study were compared against 20 other ITS-1(L) cyanobacterial sequences available in GenBank (Figure 2.1.2). The high bootstrap value separating Australian from Brazilian strains indicates that the phylogenetic relationships among the strains are principally defined by geographical origin, as previously suggested (Gugger *et al.,* 2005). If we observe the distribution of CYL producing strains in the branch of Australian strains, it is clear that a correlation between phylogenetic association and toxin

phenotype for *C. raciborskii* strains does not exist. However, consistency in the number of ribosomal operons in the *C. raciborskii* strains was observed in analyses by Southern blot hybridization. Partial 16S rRNA gene sequences, PCR-amplified and applied as a probe, served to establish the presence of at least three ribosomal operons for each strain, whereas the ITS-1 amplification by PCR detected only two types. It seems that, besides the phylogeographic relationship of the strains, the genome organization is similar among the analyzed strains, i.e. a similar distribution of ribosomal operon in the Australian strains. The multi-faceted evidence from this study indicates that the toxic phenotype does not correlate phylogeny with the toxic phenotype in *C. raciborskii* strains, leading to questions regarding the mechanisms to explain this non-correlation. The analysis of genes presumed to be involved in the CYL-biosynthetic pathway showed that the genes were specific and conserved for the three CYL-producing strains and were located in the same region in the genome, most probably forming a cluster as has been described for other non-ribosomal synthesized hepatotoxins (Tillett *et al*., 2000). The absence of the complete CYL biosynthetic gene cluster in the non-toxic strains may be explained as the primordial absence of these genes or via loss of large genomic fragments, rather than by partial deletions or point mutations. Our data strongly support the LGT process as being responsible for acquisition or loss of toxin gene clusters in *C. raciborskii*. We infer, if LGT is a plausible explanation for generating the toxic phenotype, that the presence of several features related with this process are expected in the *C. raciborskii* genome, such as: 1) the presence of tRNA genes associated with the cluster, which are common integration sites for mobile elements; 2) the presence of transposases associated with the flanking regions of the cluster, as evidence for a mobile genomic island conformation; 3) a complete or cryptic phage sequence, capable of transferring large genomic sequences between bacteria and representing a major driving force for evolutionary processes (Frost *et al*., 2005); and 4) a high number of repeated regions, which are important for reorganization and plasticity of bacterial genomes. These presumptions should be further elucidated by complete genome sequencing.

Regarding the Brazilian phylogeographical cluster, our results suggest that there is incongruence in the taxonomic classification of the genera *Cylindrospermopsis* and *Raphidiopsis*. Careful and further analysis must to be done to clarify the taxonomic characterization between strains of *Raphidiopsis* and *Cylindrospermopsis* species. The ecological consequences of horizontal transfer of such gene clusters responsible for toxin phenotypes between cyanobacteria are significant because of the spreading and increasing occurrence of cyanobacterial blooms in freshwater supplies all over the world (Saker, 2000). Furthermore, the co-existence of toxic and

non-toxic cyanobacteria in the same bloom could generate new toxic strains or, worse yet, new biosynthetic pathways that give rise to new toxins. Such events have important implications for future water supply reserve problems with dangerous consequences to human and animal health.

2.1.5 Materials and methods

Strains and culture conditions

Non-axenic cultures of *C. raciborskii* and *Anabaena circinalis* were obtained from the culture collection of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) (Table 2.1.1). *C. raciborskii* strain D9 corresponds to a clonal culture isolated from the mixed culture SPC-338, obtained from the Billings water reservoir, Sao Paulo, Brazil. Strain D9 is genetically indistinguishable of strain C10, previously described by Castro *et al*., (2004). More information, including strain source, morphology and toxicity of strains are detailed in Table 2.1.1. Cyanobacteria were cultured in 250 ml flasks containing 100 ml of MLA medium (Castro *et al*., 2004), without aeration, at 25–28 °C, under fluorescent light with a photon flux density of 40 μmol m^{-2} s^{-1} and on a 12:12-h light/dark photocycle.

Table 2.1.1 Cyanobacterial strains used in this study

Species	Strain	Toxin[b]	Morphology	Source	Isolated by	Reference
	D9	STX, GTX2, GTX3 dcSTX, dcGTX2, dcGTX3	Straight	Billing reservoir, Sao Paulo, Brazil	L. Fuenzalida	(Fuenzalida, 2005)
	CS-505 (CR1/SDS)[a]	CYL, doCYL	Straight	Solomon Dam, North Queensland, Australia	M. Saker	(Saker *et al*., 1999)
	CS-506 (CR2/SDC)[a]	CYL, doCYL	Coiled	Solomon Dam, North Queensland, Australia	M. Saker	(Saker *et al*., 1999)
	CS-507 (CR3/AQS)[a]	Non-toxic	Straight	Townsville, North Queensland, Australia	M. Saker	http://www.cmar.csiro.au/microalgae/
	CS-508 (CR6/AQC)[a]	Non-toxic	Straight	Townsville, North Queensland, Australia	M. Saker	http://www.cmar.csiro.au/microalgae/
	CS-509 (CR7/LJ)[a]	Non-toxic	Straight	Lake Julius, Townsville, Queensland, Australia	M. Saker	http://www.cmar.csiro.au/microalgae/
C. raciborskii	CS-510 (CR5/ANN)[a]	Non-toxic	Straight	McKinlay farm dam, North Queensland, Australia	M. Saker	Unpublished
	CS-511 (CR5/MCK)[a]	CYL, doCYL	Straight	McKinlay farm dam, North Queensland, Australia	M. Saker	Unpublished
	CYP-026J	CYL	c	Bourke, NSW, Australia	M. F. Gugger	(Gugger *et al*., 2005)
	CYP-023	c	c	Bourke, NSW, Australia	M. F. Gugger	(Gugger *et al*., 2005)
	CYP-030B	c	c	Bourke, NSW, Australia	M. F. Gugger	(Gugger *et al*., 2005)
	PMC115.02	Non-toxic	Straight	Guiers, Senegal	C. Berger	(Gugger *et al*., 2005)
	PMC117.02	Non-toxic	Flexuous	Guiers, Senegal	C. Berger	(Gugger *et al*., 2005)
	PMC118.02	Non-toxic	Straight	Guiers, Senegal	C. Berger	(Gugger *et al*., 2005)
	PMC98.14	c	c	Viry-Châtillon, France	M. F. Gugger	(Gugger *et al*., 2005)
	PMC99.12	c	c	Chanteraines, France	M. F. Gugger	(Gugger *et al*., 2005)
	PMC114.02	c	c	Courneuve, France	M. F. Gugger	(Gugger *et al*., 2005)
	ACT-9502	Anatoxin-a-like CYL-like	Straight	Balaton, Hungary	M. Présing	(Kiss *et al*., 2002)
	CYLI 53	c	c	Lake Melangsee, Germany	M. F. Gugger	(Gugger *et al*., 2005)
	PMC99.06	c	c	Epazote, Mexico	M. F. Gugger	(Gugger *et al*., 2005)

	ITEP-018	c	c	Tabocas reservoir, Brazil	M. F. Gugger	(Gugger *et al*., 2005)
Raphidiopsis sp.	ITEP-005	Non-toxic	c	Tapacura', Brazil	M. F. Gugger	(Gugger *et al*., 2005)
	ITEP-007	Non-toxic	c	Ingazeira, Brazil	M. F. Gugger	(Gugger *et al*., 2005)
Synechococcus sp.	RCC307	c	c	Sargasso Sea, depth: 170 m	N.A. Ahlgren	(Ahlgren & Rocap, 2006)
	UW 140	c	c	Sargasso Sea, depth: 170 m	N.A. Ahlgren	(Ahlgren & Rocap, 2006)
Anabaena circinalis	CS-537/13 (ACMB13)[a]	STX, GTX2, GTX3, dcSTX, dcGTX2,dcGTX3	Coiled	Mt Bold Reservoir, South Australia, Australia	W. Van Dok	Unpublished
Anabaena lemmermanii	262	Non-toxic	Straight	Lake Vesijärvi, Finland	M. F. Gugger	(Gugger *et al*., 2005)
Aphanizomenon ovalisporum	ILC 14	CYL	Straight	Israel: Lake Kinneret	E. Masseret	Unpublished

[a] Nomenclature used by CSIRO and in references.

[b] Toxicity confirmed by LC/MS-MS.

[c] Unknown.

DNA extraction and PCR amplification

Genomic DNA was extracted from 20 ml of fresh cyanobacterial culture, using the CTAB method described by Ausubel *et al*. (1992). PCR-amplifications of 16S rRNA genes and 16S-23S rRNA gene ITS-1 regions were performed with primers 27F and 340R (Table 2.1.2). The PCR reagents for each amplification were: 0.25 U Taq DNA polymerase (Invitrogen®); 3 ml 10X PCR buffer (Invitrogen®); 2.5 mM MgCl (Invitrogen®); 0.4 mM primers; and 0.93 mM of each deoxynucleotide. Thermal cycling was performed in an Eppendorf Mastercycler, under the following conditions: initial DNA denaturation for 3 min at 95 °C; primer annealing for 6 min at 58 °C; primer-extension for 1.5 min at 72 °C; a second cycle of 1.5 min at 95 °C; 3 min at 58 °C; 1.5 min at 72 °C; followed by 29 cycles of 1.5 min at 95 °C; 1.5 min at 58 °C; and 1.5 min at 72 °C; with a final elongation of 10 min at 72 °C. PCR products were checked by electrophoresis in 8 % polyacrylamide gels stained with silver nitrate (Bassam & Gresshoff, 2007). The PCR reaction for the analyses of putative CYL genes was performed with primers previously described by Kellmann *et al*. (2006), in 30 ml reaction volumes containing: 1X Taq polymerase buffer; 2.5 mM $MgCl_2$; 0.46 mM of each dNTP; 1 pmol forward and reverse primer; 50–100 ng of genomic DNA; and 0.3 U of Taq polymerase (Invitrogen®). Thermal cycling was performed in an Eppendorf Mastercycler. The PCR-amplifications were performed as follows: an initial step of 94 °C for 3 min; 30 cycles of 94 °C, 10 s; 55 °C, 20 s; 72 °C, 1 min; with a final elongation at 72 °C for 7 min. The PCR products were electrophoresed in a polyacrylamide/bisacrylamide gel (15 %/0.4 %) in a Mini-PROTEAN Electrophoresis system (BIO-RAD®) and stained with silver nitrate.

Table 2.1.2 PCR and sequencing primers and target sites

Primer	Sequence (5' - 3')	Target site	Reference
M13F	GACGTTGTAAAACGACGGCCAG	2956–2972[d]	pGEM®-T Easy vector, Promega
27F	AGAGTTTGATCCTGGCTCAG	8–27 (16S)[b]	(Lane *et al*., 1985)
CYA106F[a]	CGGACGGGTGAGTAACGCGTGA	106–127 (16S)[c]	(Nübel *et al*., 1997)
322F	TGTACACACCGCCCGTC	1332–1348 (16S)[c]	(Iteman *et al*., 2000)
M13R	CACAGGAAACAGCTATGACCATG	176–192[d]	pGEM®-T Easy vector, Promega
CYA781R(A/B)	GACTAC(T/A)GGGGTATCTAATCCCATT	781–805[c]	(Nübel *et al*., 1997)
1492R	GGTTACCTTGTTACGACTT	1492–1513 (16S)[b]	(Lane *et al*., 1985)
1221R[a]	ATTGTAG(T/C)ACGTGTGTAGCC	1221–1240 (16S)[b]	(Neilan *et al*., 1997)
340R	CTCTGTGTGCCTAGGTATCC	1999–1979 (23S)[c]	(Iteman *et al*., 2000)
CPSF	AGTATATGTTGCGGGACTCG	338–357 (*aoaB*)	(Kellmann *et al*., 2006)
CPSR	CCCGCCAAGACAGAGGGTAG	797–816 (*aoaB*)	(Kellmann *et al*., 2006)
A205PKF	AATGACAGAGACTTGTGCGGGG	966–987 (*aoaC*)	(Kellmann *et al*., 2006)
A205PKR	TTATCGGTATTGGTGGTAGCAACT	429–452 (*aoaC*)	(Kellmann *et al*., 2006)
CYLATF	ATTGTAAATAGCTGGAATGAGTGG	13–36 (*aoaA*)	(Kellmann *et al*., 2006)
CYLATR	TTAGGGAAGTAATCTTCACAG	1098–1118 (*aoaA*)	(Kellmann *et al*., 2006)

[a]Primers used only for sequencing.
[b]Primer position based on the *E. coli* numbering of 16S rRNA nucleotides
[c]Primer position based on the *Synechocystis* PCC 6803 ribosomal operon.
[d]Primer position according to pGEM®-T Easy vector.

Cloning and sequencing of PCR-amplified 16S rRNA gene and ITS-1 regions

The16S rRNA gene plus ITS-1 region PCR-amplification products were separated by electrophoresis in 0.8 % (w/v) low melting point (LMP) agarose (US Biological), stained with ethidium bromide (1 %). Bands were excised from the gel and purified using the Wizard® PCR Purification System (Promega®). Cloning of the purified fragments was done with the pGEM®-T Easy Vector System (Promega®) and transformed into *E. coli* DH5α cells, following the manufacturer's instructions. Two representative clones of the shorter and longer PCR fragments of each strain were selected for sequencing. DNA was sequenced directly from plasmids, with primers M13F, CYA106F, CYA781R(A/B), 1221R, 340R and M13R (Table 2.1.2) and Big Dye Terminator v3.1. Sequencing reactions were performed in Macrogen LTD (www.macrogen.com) with an ABI 3730XL sequencer (Applied Biosystems) and POP7 polymer. All sequences were checked manually.

RFLP and PFGE

The following protocol was based on those developed for the unicellular cyanobacterium *Prochlorococcus marinus* (Hess *et al*., 1999) and for the bacterium *Streptomyces coelicolor* (Alduina *et al*., 2003). Cyanobacterial cultures (100 ml) were harvested at mid-exponential phase of growth and washed with sterile water through a 3.0 mm filter, to remove heterotrophic bacteria in the cultures. The removal of accompanying bacteria from each preparation was checked by acridine orange staining with epifluorescence microscopy (Hobbie *et al*., 1977). The washed cells were resuspended in 500 ml EDTA (50 mM), pH 8.0, and vortex-mixed with 500 ml of preheated (42 °C) 1.5 % LMP agarose in EDTA (125 mM), pH 8.0. The LMP agarose-cell suspension was allowed to solidify and plugs were prepared, using the BioRad® system. Agarose plugs were incubated with 5 ml of ESP buffer [EDTA (0.5M), pH 8.0; N-lauroylsarcosine 1 % (w/v); proteinase K 1 % (w/v)] at 55 °C for 24 h, and for another 48 h with a lower concentration of proteinase K (0.05 % w/v) in ESP buffer, changing fresh buffer every 24 h. Agarose plugs were washed for 1 h in 5 ml of TE25Suc solution (EDTA (25 mM), pH 8.0; Tris (25 mM), pH 8.0; Sucrose (0.3 M)) containing 1 µl · ml^{-1} of 0.1M phenylmethylsulfonylfluoride (PMSF), and finally washed three times in TE buffer (Tris-HCl (10 mM); EDTA (1 mM), pH 8.0). The agarose plugs were kept in 0.5 M EDTA (pH 8.0) at 4 °C, until used. For restriction analysis, in situ digestion was performed with *Mlu* I and *Xho* I restriction enzymes (Fermentas) as follows: half of a plug was washed three times (1 h each time) with TE buffer at room temperature; then each plug was equilibrated in 250 ml of restriction endonuclease buffer at room temperature for 1 h; the buffer was then replaced by a 250 ml reaction mixture containing 20 U of restriction endonuclease; and DNA was digested overnight at 37 °C. Electrophoresis was performed at 14 °C with a CHEF DRII system (BioRad®), in 1.2 % agarose (BioRad®) in 0.5 X TBE buffer supplemented with 50 mM thiourea. The field strength was maintained constant at 6 V · cm^{-1}. The pulse times were set according to the sizes of the fragments, with pulses between 0.1–25 s for duration of 21 h, for fragments obtained after *Xho* I digestion, and two blocks of pulses between 1–18 and 3–80 s for duration of 10 h duration each, for fragments obtained after *Mlu* I digestion. Finally, the gels were stained with ethidium bromide and visualized in a UV transilluminator.

Phylogenetic analysis based on 16S rRNA gene and associated ITS-1 sequences

For the 16S rRNA genes and associated ITS-1 regions of *C. raciborskii* strains, sequences of cloned PCR amplification fragments between 1,735 and 1,920 bp were edited using the BioEdit program package (Hall, 1999), and aligned with options of ClustalW (with default parameters) and checked by the SILVA rRNA database (http://www.arb-silva.de). Amplified fragments of 1,683 and 1,820 bp, corresponding to the 16S rRNA gene plus ITS-1(S) (shorter) or ITS-1(L) (longer), respectively, were analyzed individually. The sequences of the same region from *A. circinalis* ACMB13 were utilized as out-groups. Edited sequences of uniform size were used in all analyses. Primer sites and ambiguous regions observed in the alignments were excluded. The longer ITS-1 sequences (504 bp) of all *C. raciborskii* strains utilized in this study were obtained from GenBank. The accession numbers are: AJ582284 (ITEP-A3); AJ582283 (PMC00.01); AJ582282 (PMC99.06); AJ582281 (CYLI 53); AJ582280 (ACT-9502); AJ582279 (PMC114.02); AJ582278 (CYP-023); AJ582277 (CYP-030B); AJ582276 (ITEP-018); AJ582273 (PMC99.12); AJ582271 (PMC117.02); AJ582270 (PMC118.02); AJ582269 (CYP-026J); AJ582268 (PMC98.14). The ITS-1(L) sequences of other filamentous and nitrogen-fixing cyanobacteria belonging to the order Nostocales were used as outgroups: *A. circinalis* ACMB13 (EU636199, determined in this study); *Aphanizomenon ovalisporum* ILC-146 (AY335547); *Anabaena lemermanni* 262 (AJ293101); and the unicellular cyanobacteria of the order Chroococcales: *Synechococcus* sp. RCC307 (DQ351316) and *Synechococcus* sp. UW 140 (DQ351315). Phylogenetic analysis by maximum parsimony (MP), neighbor joining (NJ), and maximum likelihood (ML) were performed with the PAUP* program package (Swofford, 2003). MP included the following options: exhaustive research; tree bisection-reconnection (TBR); and ACCTRAN. NJ included p-distance options. For the ML analysis, the best evolutionary model was selected according to the Akaike information criterion (AIC), using the program ModelTest 3.06 (Posada & Crandall, 1998). The statistical confidence of the nodes was evaluated by 1000 bootstrap pseudoreplicates (Felsenstein, 1985) for all analyses.

GenBank accession numbers

The 16S rRNA gene and ITS-1 nucleotide sequences described in this study have been deposited in the GenBank database under accession numbers: *Raphidiopsis* sp. D9 16S+ITS-1(S) (EU552070); *Raphidiopsis* sp. D9 16S+ITS-1(L) (EU552069); CS-505 16S+ITS-1(S) (EU552062); CS-506 16S+ITS-1(S) (EU552063); CS-507 16S+ITS-1(S) (EU552064); CS-508 16S+ITS-1(S) (EU552065); CS-509 16S+ITS-1(S) (EU552066); CS-510 16S+ITS-1(S) (EU552067); CS-511 16S+ITS-1(S) (EU552068); CS-505 16S+ITS-1(L) (EU552055); CS-506 16S+ITS-1(L) (EU552056); CS-507 16S+ITS-1(L) (EU552057); CS-508 16S+ITS-1(L) (EU552058); CS-509 16S+ITS-1(L) (EU552059); CS-510 16S+ITS-1(L) (EU552060); CS-511 16S+ITS-1(L) (EU552061).

Phylogenetic analysis-based RFLP resolved by PFGE

A binary matrix was constructed based on the presence or absence of 59 and 58 restriction fragments obtained with the *Mlu* I and *Xho* I enzymes, respectively. MP analysis was performed with PAUP* v4.0b10 (Swofford, 2003), conducted using equal weight parsimony (heuristic search, 10 random addition replicates, TBR and branch and swapping options). Nodal support was estimated with 1000 parsimony bootstrap pseudo-replicates (Felsenstein, 1985).

Southern blot analysis

DNA from seven strains of *C. raciborskii*, digested with *Xho* I restriction enzyme (Fermentas®) and resolved using PFGE, were transferred to nylon membranes (GeneScreen Plus) and Southern blot hybridizations were carried out with different probes: a PCR product (663 bp) from the 16S rRNA gene 5' region obtained with primers CYA106F/CYA781R; a PCR product (478 bp) from a NRPS-like sequence; a PCR product (514 bp) from a PKS-like sequence; and a PCR product (1,105 bp) from a putative amidinotransferase enzyme. The probes were labeled with biotin in a random primer reaction (KPL, DNA Biotinylation Kit) and detected via a chemiluminescent reaction (KPL, Detector HRP Chemiluminescent Blotting Kit). The molecular weights of the digested DNA fragments were estimated with reference to *Vibrio parahaemolyticus* strain KX DNA digested with *Not* I restriction enzyme as a molecular marker (data not shown).

2.1.6 Supplementary material

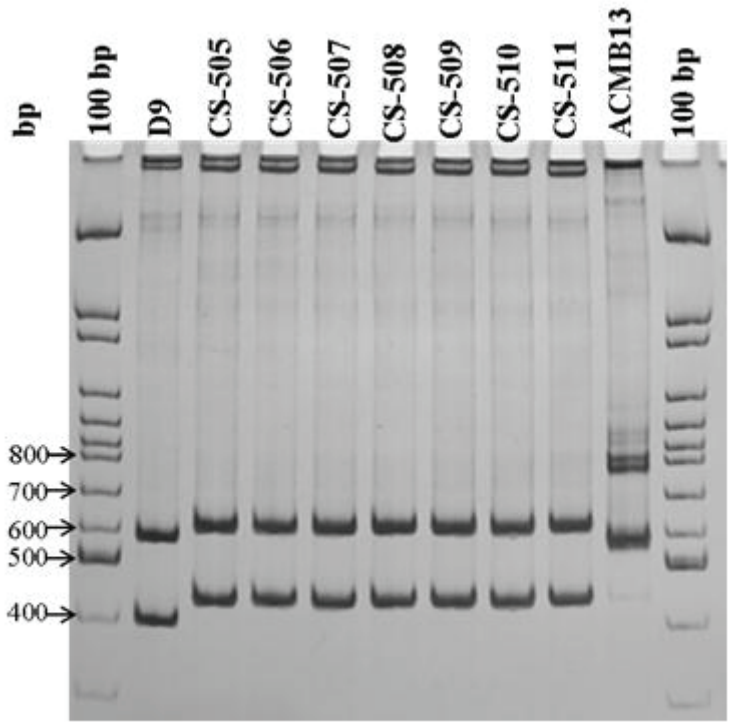


Figure 2.1.S1 PCR products of the ITS region from the *C. raciborskii* strains CS-505-CS-510, *Raphidiopsis* sp. D9, and *A. circinalis* ACMB13. The presence of two amplification products of different sizes in *C. raciborskii* and *Raphidiopsis* sp. D9 suggests that at least two ITS regions (and therefore ribosomal operons) are present in these strains but the size differences show the divergence of strain D9 with respect to *C. raciborskii.* *A. circinalis* as outgroup, has at least three ITS regions of larger size than *Raphidiopsis* and *C. raciborskii*. Electrophoresis was carried on an 8% polyacrylamide gel, and DNA was visualized by silver nitrate staining. 100 bp plus (Fermentas ®) was used as molecular weight marker.

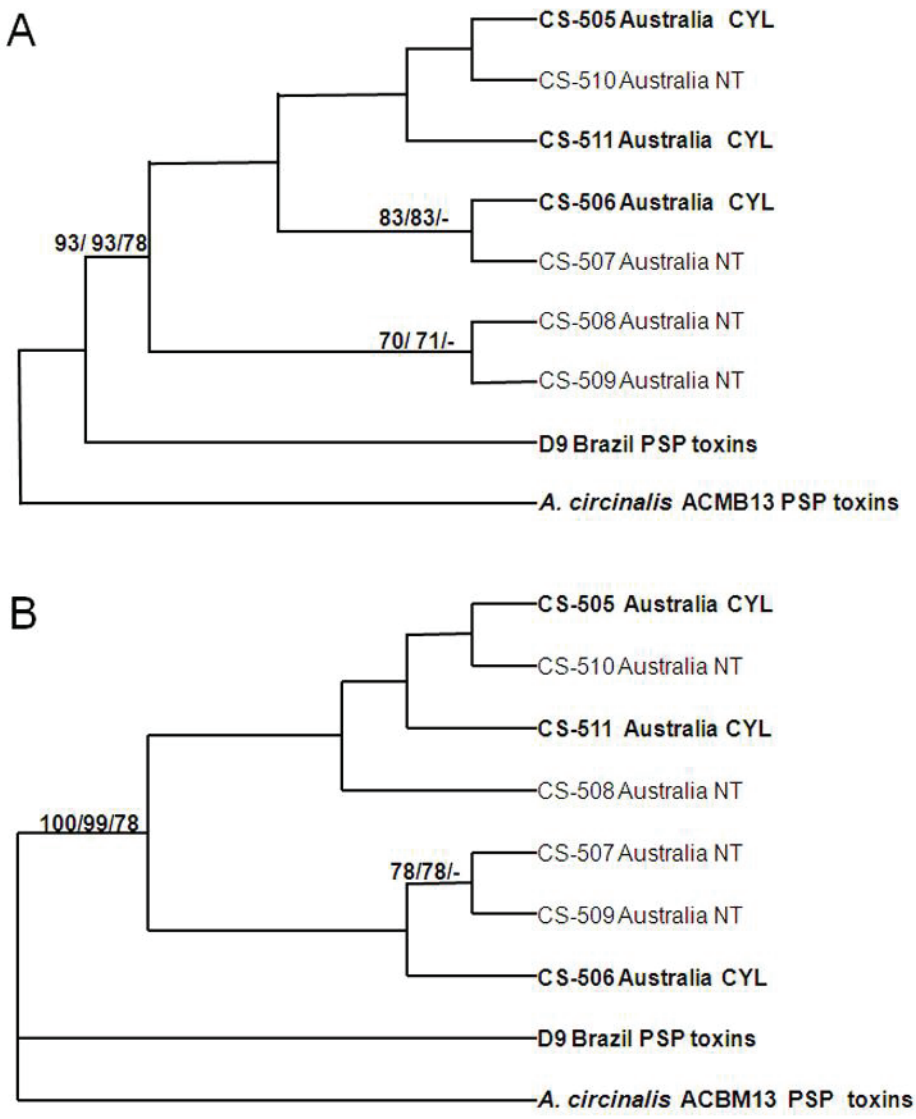


Figure 2.1.S2 Phylogenetic inference of *C. raciborskii* strains based on sequences of (A) 16S rDNA plus 16S-23S rDNA ITS (S) and (B) 16S rDNA plus 16S-23S rDNA ITS (L). Trees were inferred from Neighbor Joining (NJ), Maximum Parsimony (MP), and Maximum Likelihood (ML) methods. Values over nodes represent the percentage of 1000 bootstrap pseudoreplicates (NJ/MP/ML).

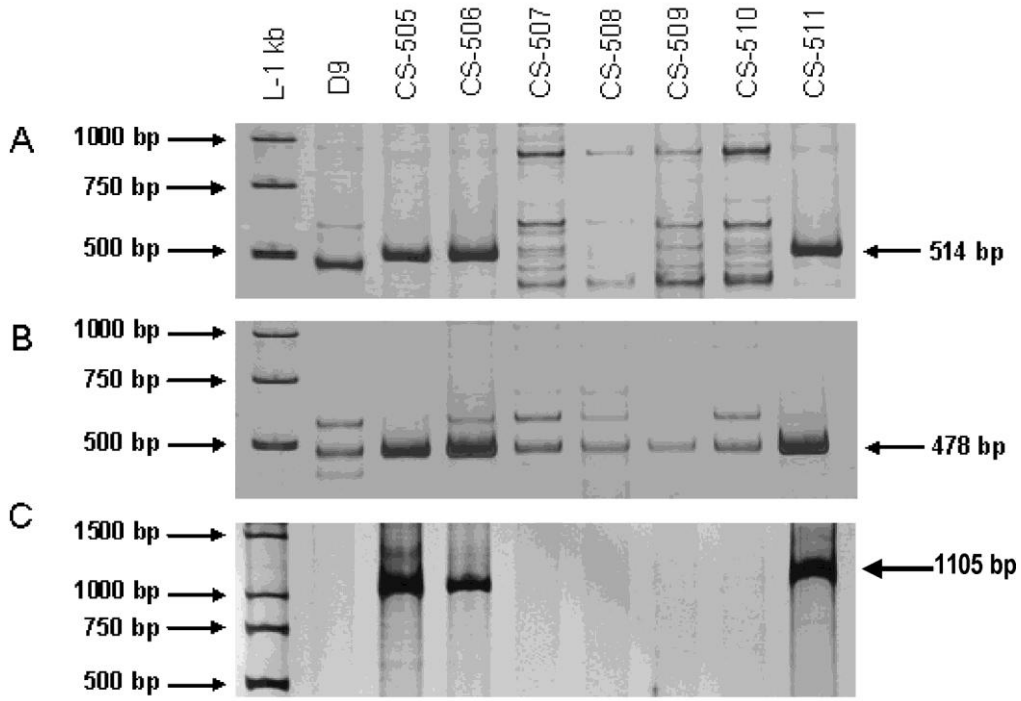


Figure 2.1.S3 PCR amplification with primers A205PKF/R (514bp), CPSF/R (478bp) and CYLATF/R (1,105bp). Primers are specific for sequences putatively associated with the production of CYL. Strains CS-505, 506 and 511 showed the predicted amplification product for the tree sequences, while in the case of PKS-like sequence (*aoaC*) the other strains showed non-specific amplifications **(A)**. For the hybrid nonribosomal peptide synthetase/polyketide synthase sequence (*aoaB*) we observed a weak amplification of the predicted product in the other strains **(B)** and amidinotransferase PCR product (*aoaA*) was specific for the tree three CYL-producing strains **(C)**.

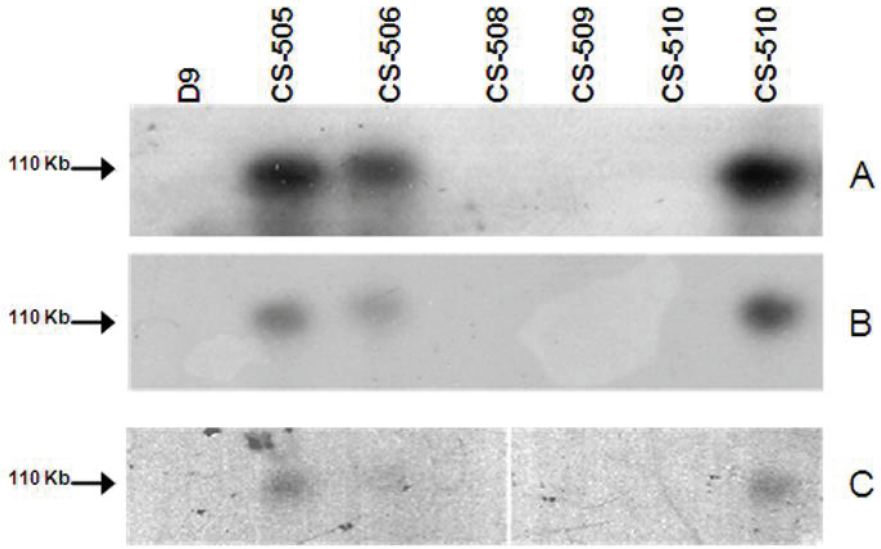


Figure 2.1.S4 Southern blot analysis of the putative CYL biosynthesis genes. The following PCR products were used consecutively as probes: **(A)** 514 bp fragment of the polyketide synthetase b-ketosynthase domain (PKS-like sequence); **(B)** 478 bp fragment of peptide synthetase A domain (hybrid nonribosomal peptide synthetase/polyketide synthase sequence); **(C)** 1,105 bp fragment of putative amidinotransferase enzyme. The hybridization was specific for the three CYL-producing strains for each probe. Target genes (indicated by the arrow) are located in the same position in the genome.

2.2 Publication II

The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications

Karina Stucken, Uwe John, Allan Cembella, Alejandro A. Murillo, Katia Soto-Liebe, Juan J. Fuentes-Valdés, Maik Friedel, Alvaro M. Plominsky, Mónica Vásquez and Gernot Glöckner

2.2.1 Abstract

Cyanobacterial morphology is diverse, ranging from unicellular spheres or rods to multicellular structures such as colonies and filaments. Multicellular species represent an evolutionary strategy to differentiate and compartmentalize certain metabolic functions for reproduction and nitrogen (N_2) fixation into specialized cell types (e.g. akinetes, heterocysts and diazocytes). Only a few filamentous, differentiated cyanobacterial species, with genome sizes over 5 Mb, have been sequenced. We sequenced the genomes of two strains of closely related filamentous cyanobacterial species to yield further insights into the molecular basis of the traits of N_2 fixation, filament formation and cell differentiation. *Cylindrospermopsis raciborskii* CS-505 is a cylindrospermopsin-producing strain from Australia, whereas *Raphidiopsis brookii* D9 from Brazil synthesizes neurotoxins associated with paralytic shellfish poisoning (PSP). Despite their different morphology, toxin composition and disjunct geographical distribution, these strains form a monophyletic group. With genome sizes of approximately 3.9 (CS-505) and 3.2 (D9) Mb, these are the smallest genomes described for free-living filamentous cyanobacteria. We observed remarkable gene order conservation (synteny) between these genomes despite the difference in repetitive element content, which accounts for most of the genome size difference between them. We show here that the strains share a specific set of 2539 genes with >90% average nucleotide identity. The fact that the CS-505 and D9 genomes are small and streamlined compared to those of other filamentous cyanobacterial species and the lack of the ability for heterocyst formation in strain D9 allowed us to define a core set of genes responsible for each trait in filamentous species.

We presume that in strain D9 the ability to form proper heterocysts was secondarily lost together with N_2 fixation capacity. Further comparisons to all available cyanobacterial genomes covering almost the entire evolutionary branch revealed a common minimal gene set for each of these cyanobacterial traits.

2.2.2 Introduction

Cyanobacteria are among the most successful primary producing aquatic organisms, having populated the Earth for approximately 2.8 billion years (Des Marais, 2000). Extant species are major (occasionally dominant) components of marine, brackish and freshwater environments, where they play crucial roles in global biological solar energy conversion and nitrogen (N_2) fixation, but are also found in terrestrial ecosystems (in mats), and as extreme thermophiles in hot springs and polar ice. In high biomass concentration, cyanobacteria are responsible for noxious or harmful algal blooms (HABs), and this phenomenon is compounded by the fact that some cyanobacteria also produce potent cyanotoxins (microcystins, nodularins, saxitoxins, anatoxins, cylindrospermopsins, etc.), which have been classified according to their mode of action and effects on mammals (Sivonen & Jones, 1999).

Cyanobacteria have evolved alternative morphologies, including unicellular and diverse multicellular forms ranging from simple colonies to branched filaments. Phylogenetic analysis has suggested that cyanobacteria capable of cell differentiation are monophyletic (Tomitani *et al.,* 2006). Within this monophyletic group some cyanobacteria further evolved from filaments in which a small number of vegetative cells differentiated into either heterocysts or akinetes (resting stages). Nitrogen (N_2) fixation, or diazotrophy, also appears to be monophyletic among cyanobacteria although a polyphyletic origin has also been proposed (Shi & Falkowski, 2008, Swingley *et al.*, 2008). When mineral and organic nitrogen sources, such as nitrate or ammonium, are depleted from the growth medium, some filamentous cyanobacteria maintain photosynthetic activity (including O_2 generation) in vegetative cells and differentiate heterocysts to provide an anoxic environment suitable for N_2 fixation (Herrero *et al.,* 2001).

The proposed evolutionary sequence of heterocyst-forming filamentous cyanobacteria is still under debate. However, a likely scenario is that diazotrophy was first established in filamentous cyanobacteria (who acquired it either by horizontal gene transfer (HGT) or by vertical evolution of a not necessarily filamentous ancestor), and only after the establishment of

diazotrophy, the capacity for heterocyst formation in filamentous diazotrophs developed (Swingley *et al.,* 2008).

Among filamentous cyanobacteria, the toxigenic species *Cylindrospermopsis raciborskii* is highly successful in freshwater environments. This species has been reported to be rapidly expanding worldwide, from tropical to temperate freshwater bodies (Padisák, 1997). *C. raciborskii* can also co-exist with morphotypes assigned to the closely related genus *Raphidiopsis* (also with toxic members), which unlike *Cylindrospermopsis* does not develop heterocysts or fix N_2 (Mohamed, 2007).

One remarkable characteristic of some cyanobacteria is their ability to form toxic blooms. Nevertheless, toxigenicity is not a ubiquitous feature at the generic level or even within a species; for example, both non-toxic and toxic strains of *Cylindrospermopsis* and *Raphidiopsis* have been isolated from natural populations. The genes responsible for toxin production are organized into clusters that might be subject to frequent horizontal gene transfer (HGT), a possible explanation for the evolution and biogeography of both toxigenic and non-toxigenic strains within species or genera (Kellmann *et al.*, 2008b) Among *C. raciborskii* strains, two totally different types of toxins may be produced: the hepatotoxin cylindrospermopsin (CYN), a tricyclic alkaloid inhibitor of protein synthesis (Falconer & Humpage, 2006), or neurotoxins associated with paralytic shellfish poisoning (PSP), specifically the tetrahydropurine saxitoxin (STX) and analogues (Lagos *et al.,* 1999). Cylindrospermopsin is biosynthesized via combined polyketide synthase/nonribosomal peptide synthetase (PKS/NRPS) pathways (Mihali *et al.,* 2008), whereas cyanobacterial STX and its analogues are likely generated by a unique gene cluster recently described in *C. raciborskii* strain T3 (Kellmann *et al.*, 2008a) and also in a few other toxic species (Mihali *et al.,* 2009). Toxic strains of *Raphidiopsis* are reported to produce CYN and/or deoxycylindrospermopsin (doCYN) (Li *et al.,* 2001b), the bicyclic amine alkaloid anatoxin-a (Namikoshi *et al.,* 2003), which affects mammalian nicotinic acetylcholine receptors, or PSP toxins (Yunes *et al.,* 2009). These cyanotoxin classes exhibit completely different mechanisms of action in mammalian systems (Falconer & Humpage, 2006, Llewellyn, 2006) and are also structurally dissimilar.

Cyanobacteria are of high ecological importance, and given their relatively small genome, they are an ideal target for genome sequencing and analysis with current genomic tools. Knowledge gained from such projects has yielded important insights into the evolution of photosynthesis (Mulkidjanian *et al.,* 2006), and the adaptation of these microorganisms to the

environment (Rocap *et al.,* 2003). Nevertheless, to date, only 9 filamentous cyanobacteria have been either completely or partially sequenced.

Comparative genomics has also revealed high genetic variability even between closely related cyanobacterial strains (Coleman *et al.,* 2006). Our objective was to conduct a genomic comparison of phylogenetically closely related filamentous cyanobacteria with a particular focus on the elucidation of the genetic background of their morphological and metabolic differences. Accordingly, we chose two cyanobacterial strains, *C. raciborskii* CS-505 and *R. brookii* D9 isolated from geographically disjunct regions in Australia (CS-505) and Brazil (D9) (Figure 2.2.1). The strains under study have been morphologically classified into different genera, since D9 produces no functional heterocysts and is therefore unable to fix N_2. Nevertheless, based on 16S rDNA analysis they share 99.5% identity and are thus part of the same monophyletic cluster (**Publication I**) The two chosen strains also express a radically different toxin profile: while CS-505 produces exclusively CYN and doCYN, strain D9 produces the PSP-toxin analogues STX, gonyaulaxtoxins 2/3 (GTX2/3) and the respective decarbamoylated analogues (**Publication I**).

We sequenced and analyzed the complete genomes of both strains CS-505 and D9, and thereby found the two smallest genomes thus far described for filamentous cyanobacteria. A comparative genomic analysis of these strains in relation to other members of filamentous cyanobacteria allowed us to propose minimal sets of core genes that provide insight into the evolution of diazotrophy and multicellularity, and heterocyst development in these minimal genomes.

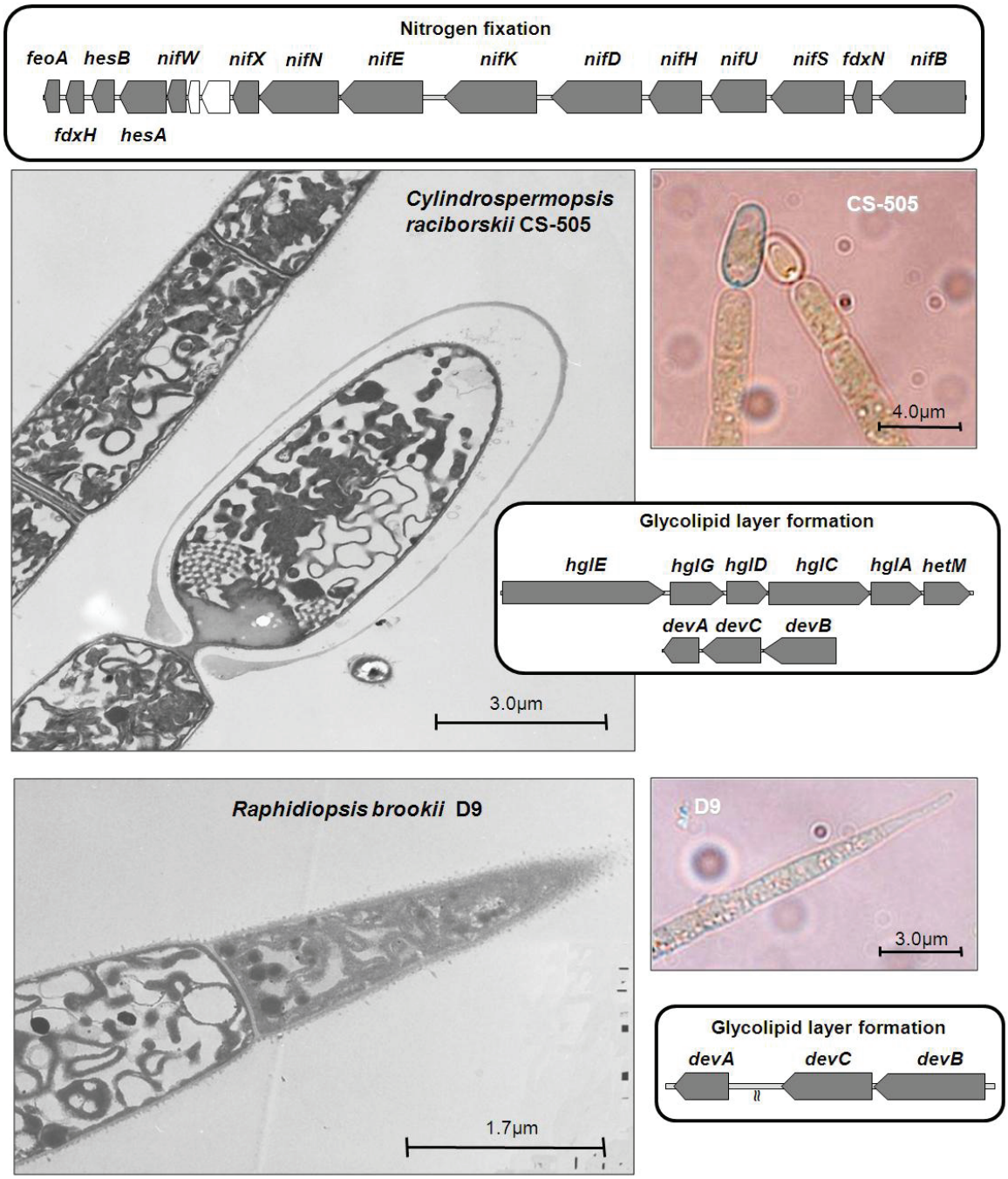


Figure 2.2.1 Overview of the main gene clusters involved in nitrogen metabolism and heterocyst development in strains CS-505 and D9. Transmission electron micrographs in the left panels show the heterocyst of CS-505 and the apically differentiated cell of D9. Optical micrographs on the right panels exhibit the Alcian blue staining characteristic of polysaccharides in the heterocyst.

2.2.3 Results and Discussion

Genome structure comparison

We sequenced the genomes to >20-fold depth with 454/Roche pyrosequencing technology (Table 2.2.1), thereby rendering >99.9% complete genomes (Lander & Waterman, 1988). Although a number of small gaps caused mainly by repetitive elements remain in both sequences, it is thus unlikely that we missed a significant portion of the genomes. The additional sequences of the long and short insert libraries from the Sanger sequencing (Table 2.2.1) also served to mitigate this deficiency, and the extra Sanger sequences derived from the short insert libraries were used to correct for 454/Roche technology intrinsic errors.

Table 2.2.1 Sequencing and assembly statistics for the two strains

	D9	**CS-505**
454 GS sequence coverage	27	34
small insert library	188	3909
fosmid library	491	--
finishing reads	253	161
Total sequencing depth	27	35
contigs	157	268
assembled (Mb)	3.20	3.89
contigs >3.5 kb	33	94
largest contig (kb)	543	259
repeats (regions)	53	406
repeats (bases)	53,870	244,280
repeats (% total)	1.7	6.3

An initial assembly of all sequencing data for each strain yielded 182 contigs (larger than 3 kb) for CS-505 and 105 for D9. Due to limitations in the assembly of next generation sequencing (NGS)-derived, repeated sequences are commonly represented only once in such an assembly. Thus, only plasmid shotgun and fosmid clone end-sequencing and clone walking enabled us to close further gaps, such that the current assembly consists of 94 contigs for CS-505 and 33 for D9 (Table 2.2.1). The highly repetitive nature of the remaining gaps prevented us from reconstructing gap-free genomes.

Contigs lengths circumscribe an overall genome size of 3.89 Mb for strain CS-505 versus a smaller genome size of 3.20 Mb for D9 (Table 2.2.1); sequences in gaps accounted for an

additional estimated 100 to 150 kb in both strains. The size of the genomes was further supported in the assessment by restriction fragment length polymorphism with Pulsed Field gel electrophoresis (PFGE). The later method rendered an estimated genome size of 3.49 Mb for CS-505 and 3.09 Mb for D9 using the restriction enzyme *Mlu* I. The smaller estimated sizes by PFGE can be attributed to the low resolution of the some bands in the electrophoresis, which may lead to an underestimation of the genome sizes (Figure 2.2.S1). Indicative of the presence of plasmids, we observed a faint band in CS-505 (Figure 2.2.S2) plus a second band (data not shown) of approximately 30 kb. Plasmids may be integrated into the genome and thereby the plasmid sequences can be present in the assemblies surrounded by two different sequence environments (plasmid only sequences or adjacent genome parts), making the integration site a low coverage region. Thus, if present at all, any plasmids are very likely represented by a single contig in our assembly.

The genomes we sequenced are almost a factor of two smaller than that of the most closely related fully sequenced cyanobacterium, *Anabaena* sp. PCC 7120 (hereafter referred as *Anabaena*) (6.41 Mb) (Table 2.2.2). The genome sizes of filamentous cyanobacterial species are previously reported to range between 5.0 and 8.7 Mb (NCBI database). Curiously, the genomes of our filamentous cyanobacteria are comparable to the genome size of those of unicellular cyanobacteria such as *Synechocystis* sp. PCC 6803 (3.57 Mb). Moreover, the number of ribosomal operons (3) and regulatory systems in both CS-505 and D9 (81 and 75 sensor-regulator components, respectively), is more similar to that of *Synechocystis* sp. PCC 6803 (3 ribosomal operons and 89 sensor-regulator systems), than to that of the filamentous *Anabaena* (4 ribosomal operons and 175 sensor-regulator systems).

Genome reduction is a well-known evolutionary strategy to streamline genomes and get rid of superfluous functions. This strategy is followed by most obligate pathogens because their metabolic processes are strongly dependent upon the host. However, free-living cyanobacteria undergo genome reduction as well (Rocap *et al.,* 2003). The reasons for this genome reduction phenomenon are unknown, but are likely related to genomic efficiency and relatively lax selective pressure on certain aspects of metabolism.

The G+C content is similar between both genomes (approximately 40%) and also similar to that of other fully sequenced genomes of filamentous cyanobacteria (Table 2.2.2). The genomes share 2,539 clearly orthologous protein coding sequences (CDS) (referred to here as shared CDS), representing 73.6% and 84.4% of all predicted CDS from CS-505 and D9,

respectively. We found 112 additional CDS in the CS-505 genome with similarities to D9 counterparts. Further analysis indicated that this surplus of CDS is mainly due to coding parts of transposable elements. Of the shared genes, the average nucleotide identity is >90% and the rate of synonymous substitutions is 0.29. These values are similar to those found for conspecific bacterial strains that have evolved in different ecological habitats (Konstantinidis & Tiedje, 2005) and are consistent with the level of similarity between 16S rRNA sequences from CS-505 and D9.

Table 2.2.2 General features of the genomes of strains CS-505 and D9 in comparison with four other fully sequenced genomes of filamentous cyanobacteria.

	D9	**CS-505**	**Avar**	**Anab**	**Tery**	**Npun**
Genome size (Mb)	3.20	3.89	6.34	6.41	7.75	8.23
G+C content (%)	40	40.2	41	41	40.8	41
Genes	3,088	3,968	5,134	5,432	5,542	6,501
CDS	3,010	3,452	5,039	5,368	4,452	6,087
Function assigned (COGs)	1,979	1,922	3,799	3,892	2,729	0
Unclassified	1,031	1,530	1,244	1,474	2,347	6,087
rRNA genes	9	9	12	12	5	12
tRNA genes	42	42	47	48	38	98
Transposases	9	77	57	145	260	112
Phage integrases	-	2	10	-	3	22
Repeated regions	53	406				
Plasmids	?	?	3	6	-	5
Unique CDS	394	794				
Function assigned (COGs)	157	291				
Unclassified	237	503				

Abbreviations: Avar: *Anabaena variabilis* ATCC 29413; Anab: *Anabaena* sp. PCC 7120; Tery: *Trichodesmium erythraeum* IMS101; Npun: *Nostoc punctiforme* PCC 73102. * Function assigned according to COGs

Unique CDS in the genomes

Comparative analysis via Best Bidirectional Hits (BBH) revealed large differences in the number of unique CDS between the two strains. CS-505 has 794 (23%) unique CDS whereas D9 contains 394 (13%). The presence and number of unique CDS among two closely related strains may represent the different potential for ecological adaptation and physiological plasticity. This relationship has been proven, particularly for pathogenic bacterial isolates that acquire

pathogenicity islands conferring the toxic phenotype (Juhas *et al.,* 2009). Even the acquisition of single genes can yield adaptations to a specific strain. For example, the two ecotypes of *Prochlorococcus marinus* MIT9313 and MED4 differ in the presence of certain Photosystem II and nitrite transport and reduction genes, among others. These differences correlate with the distribution of the ecotypes in the water column (Rocap *et al.,* 2003).

The classification of the unique CDS into Clusters of Orthologous Groups (COGs) showed that for two thirds of the unique CDS no function could be assigned (503 of 794 for CS-505 and 237 of 394 for D9). Yet of the remainder, there was a homogeneous distribution within most of the COG categories, indicating common functions between CS-505 and D9 (Figure 2.2.S3, Table 2.2.S1). Only a minor fraction of the unique CDS of both strains showed evident differences in their distribution into the different categories. Those differences were restricted to seven COGs (Figure 2.2.2) from which only two categories were better represented by D9-unique CDS: coenzyme and amino acid transport and metabolism. The COG distribution clearly showed the greater metabolic capabilities of CS-505 than D9 in relation to: 1) secondary metabolite biosynthesis, transport and catabolism; 2) replication, recombination and repair (this category is overrepresented partially due to transposons), 3) energy production and conversion, 4) cell cycle control and 5) cell wall and membrane biogenesis. On closer inspection, most of the identifiable unique CDS of CS-505 were organized into gene clusters and could be attributed to toxin production and heterocyst differentiation coupled with diazotrophy. Thus, the lack of these genes and the scarcity of unique genes in D9 points to the fact that this genome was shaped by gene and function losses rather than gains.

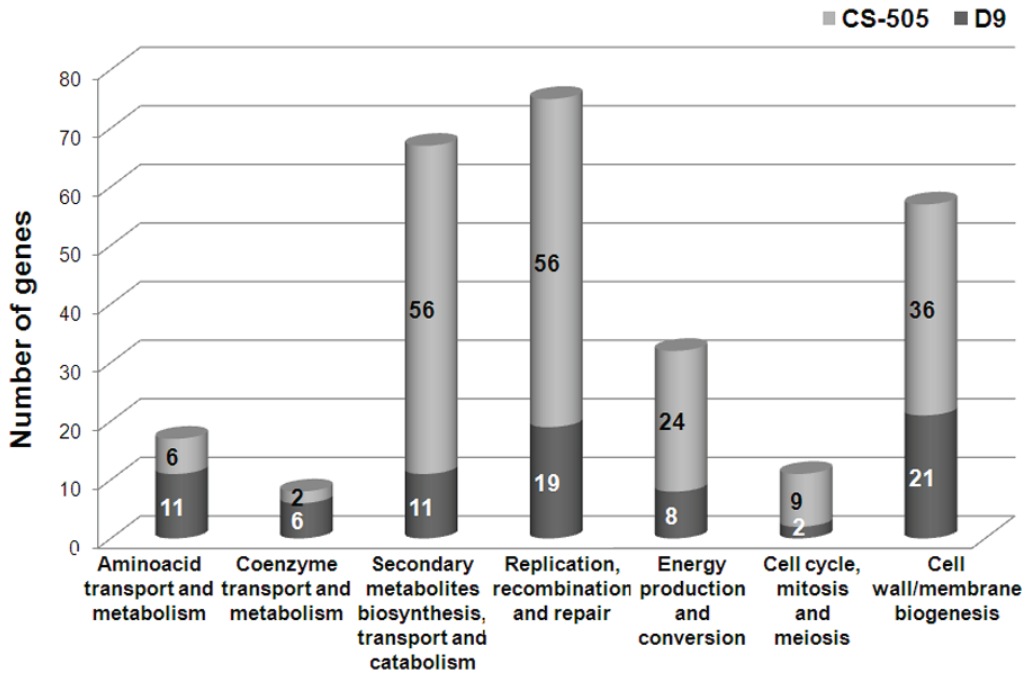


Figure 2.2.2 Distribution of the unique CDS of CS-505 and D9 into Cluster of Orthologous Groups (COGs). Only COG categories over-represented by CDS of CS-505 or D9 are shown (see text for more details). Unique CDS were obtained by a Best-Bidirectional Hits (BBHs) search between both genomes using a 30% cutoff.

Repetitive elements and synteny

The most prominent difference between the genomes of CS-505 and D9 is the overwhelming number of repeated insertion elements or transposon-derived sequences in the CS-505 genome, which accounts for a considerable part of the genome size difference (nearly 0.6 Mb or 20% of the D9 genome) between the two strains (Table 2.2.1). Repetitive elements are not rare in cyanobacteria. On the contrary, a high percentage of repeated sequences was found for the genomes of *Crocosphaera watsonii* WH8501 (19.8%) and of the only two sequenced *Microcystis aeruginosa* strains (11.7% each) (Kaneko *et al.*, 2007, Frangeul *et al.*, 2008). Nevertheless, our study represents the first time that large differences in repeat numbers have been observed between closely related strains. A low number (~100) of surplus CDS in the CS-505 genome reflects apparent gene duplications or functional redundancies. However, since we produced significantly more large- and small-insert library derived reads for the CS-505 strain from Sanger-based sequencing, a small portion of the observed genome size difference could be due to the better resolution of repeats in this strain. Such redundancy of long stretches of nearly identical sequences also contributes to our difficulties in closing the gaps in the genome sequences,

particularly for CS-505. The total number of nearly identical repeated sequences with coding potential in the CS-505 genome accounts for 6.3% of its genome length. In addition, we identified two phage integrase genes and 77 transposases among them, from which only 28 were full sequences reflecting possible functionality.

The low number of mobile elements in the D9 genome compared to other cyanobacteria is indeed remarkable. The presence of only 9 transposases (just one is a full sequence), 53 repeated regions and no phage integrase genes points to a high plasticity of the CS-505 genome relative to the transposon-poor genome of D9 (Table 2.2.2). Pursuit of repeated sequence elements not necessarily coding for proteins, employing a strategy described in Abouelhoda *et al.,* (2004) (see methods), allowed us to define 20 clusters that are present more than once in the CS-505 genome. Interestingly, one of those clusters is internally repetitive, i.e. a short sequence stretch is repeated identically within this cluster several times (Figure 2.2.S4) and occurs 39 times in the genome.

Repetitive elements can be a source for genome rearrangements. This genomic plasticity could be partly responsible for niche adaptation of organisms to their environment. We counted the number of syntenic regions between the two species to estimate the number of rearrangements that occurred after their evolutionary separation. Interestingly, all 2,539 orthologous gene pairs are located in syntenic regions, meaning that at least one neighboring CDS is common between the two species. This excludes the possibility that single genes were relocated to other genomic regions during evolution. In total we found 280 synteny groups with a mean of 9 members in a group. The largest group comprised the ribosomal cluster and an adjacent CDS with 55 orthologous pairs. If we compare the *A. variabilis* genome to that of D9 in the same way we observe 464 synteny groups with only 1,651 members. Thus, not unexpectedly, the mean of 3.6 CDS per synteny group is much lower than in the CS-505/D9 comparison. The high sequence similarity between CS-505 and D9 emphasizes the close relationship between the two strains, whereas the synteny analysis shows that rearrangements occurred relatively frequently during evolution. This high plasticity may be partly due to the high number of repeated elements.

Genomic islands for N_2 fixation and toxin production

We did not find any region matching the criteria for definition of a genomic island, i.e. differing G+C content, presence of direct repeats, transposition elements or tRNA sequences (Juhas *et al.,* 2009), within the CS-505 and D9 genomes. Nevertheless, we found gene clusters present in one or the other strain; thus we cannot discard the possibility that islands containing those gene clusters were transferred from genomes with a similar G+C content. Filamentous cyanobacteria are known to have a homogenous G+C content (Table 2.2.2). The most prominent examples of such identifiable gene clusters in our strains are those for N_2 fixation and toxin production in CS-505, and the toxin production gene cluster in D9. In strain CS-505 the *nif* gene cluster encoding for the Fe-Mo cofactor-dependent nitrogenase and 13 other genes related to N_2 fixation are all together within a tight 15 kb cluster. The gene content is therefore similar to the *nif* cluster of heterocystous cyanobacteria (Herrero *et al.,* 2001). The gene organization, however, is comparable to that of the second *nif* cluster expressed in vegetative cells of *Anabaena variabilis* (Thiel *et al.,* 1997), and of the *nif* cluster of the symbiotic *Nostoc azollae* 078 (see Figure 2.2.3 for the comparison with *A. variabilis*). The distinguishing feature of this gene organization is that it does not exhibit excision elements interrupting the *nifD* sequence, a characteristic of many other heterocyst-forming cyanobacteria. A second nitrogenase operon *nifVZT* (also commonly present in diazotrophic cyanobacteria) is located at a different locus in CS-505. The D9 strain is not able to fix N_2 and is therefore dependent on the uptake of N-containing compounds from the environment. This dependency is nicely reflected by the absence of the N_2-fixation gene clusters (*nif*) and the prevalence of several unique CDS for coenzyme- and amino acid-transport in the D9 genome (Figure 2.2.2). We note as significant that there is shared synteny in the regions surrounding the *nif* clusters in the compared strains (Figure 2.2.3). The *nif* clusters in *R. brookii* D9 might thus have been selectively lost along with the corresponding function. Nevertheless, the D9 genome encodes and expresses (see Materials and Methods) *hetR*, an important regulator of heterocyst differentiation and pattern formation in N-fixing cyanobacteria (Khudyakov & Golden, 2004), under normal culture conditions (with nitrate as N-source). As reported by Zhang *et al.,* (2009), the presence of *hetR* and its expression have been detected in non-heterocyst producing cyanobacteria that also do not fix nitrogen, pointing to a more global role of HetR.

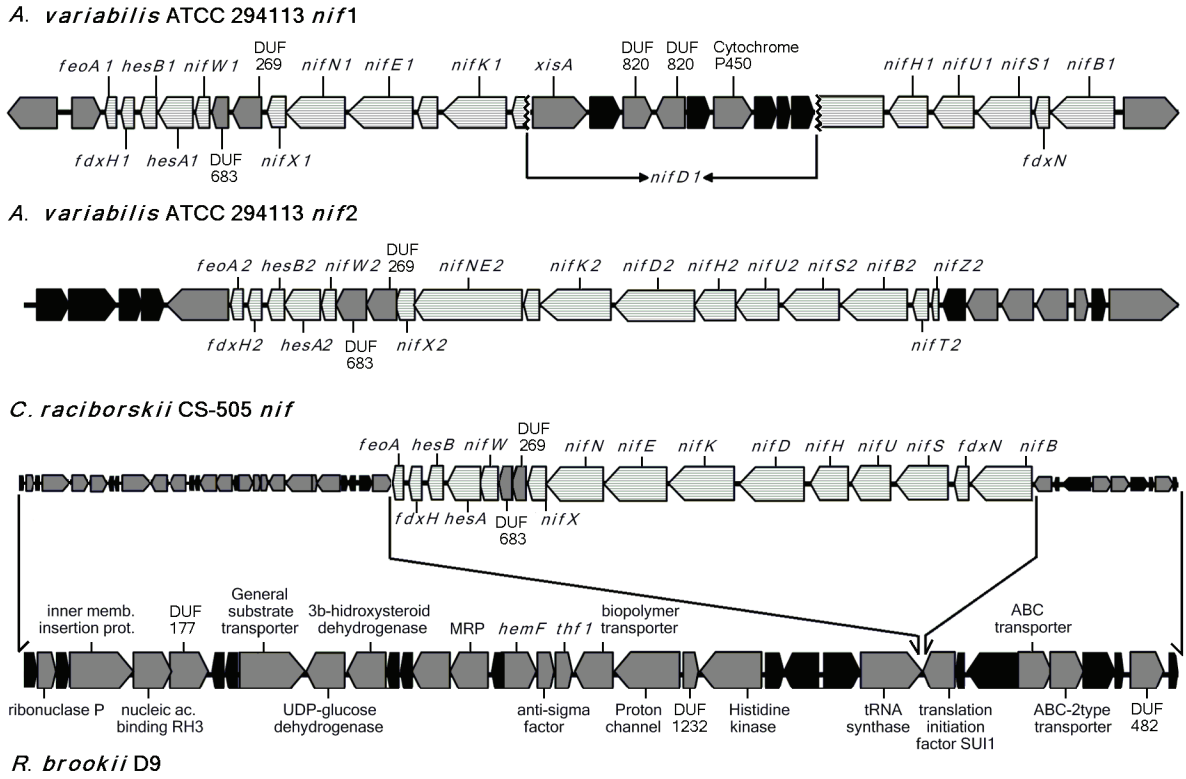


Figure 2.2.3 Schematic representation of the synteny within the vicinity of the *nif* gene clusters. The scheme represents the 15kb gene cluster containing the *nifHDK* and the other 13 nitrogen fixation related genes in CS-505 compared with the *nif1* and *nif2* gene clusters of *Anabaena variabilis* ATCC 29413 and the synteny regions between CS-505 and D9. The synteny regions between CS-505 and D9 are delimited by the arrows. *nif* genes are represented by light grey. Genes in black correspond to hypothetical proteins and dark grey genes to proteins with assigned function.

A similar example of common and conservative elements is observed in the toxin gene clusters of CS-505 and D9. The different cyanotoxins produced by strains CS-505 and D9 are the most prominent known secondary metabolites in these cyanobacteria. The tricyclic alkaloids CYN/doCYN and the tetrahydropurine STX and analogues are N-rich molecules, but these toxin groups are synthesized by two independent and apparently unrelated biosynthetic pathways in cyanobacteria. The *C. raciborskii* CS-505 genome encodes for only one hybrid NRPS-PKS pathway, corresponding to the CYN/doCYN biosynthesis cluster. The cluster spans 41.6 Kb, encodes for 16 Open Reading Frames (ORFs) and has complete synteny with the CYN cluster of *C. raciborskii* AWT205 (Mihali *et al.,* 2008) (Figure 2.2.S5A), flanked at both ends by genes from the hydrogenase gene cluster (*hypABCDEF*). In addition to the two transposase ORFs, *cyrL* and *cyrM*, CRC_01709 is only present in CS-505. This latter gene fragment of 219 bp is located

between *cyrC* and *cyrE,* and matches with part of a transposase from *Synechococcus* BO 8402. The *cyrM* and CRC_01709 components are only vestiges of transposases, indicating that rearrangements have occurred in this section of the genome. The same genetic structure neighboring the CYN biosynthetic gene cluster in CS-505 is present in the strain D9, as another example of synteny (Figure 2.2.4A). As this conservation has been shown in other non-CYN producing Australian strains of *C. raciborskii* that contain the uninterrupted hydrogenase cluster (Figure 2.2.4B), we find it plausible that each cluster could be inserted or deleted at common genetic loci.

Likewise, the genes adjacent to the STX gene cluster in the D9 genome form a syntenic region within the CS-505 genome (Figure 2.2.4C). The STX gene cluster in D9 covers 25.7 Kb, and encodes for 24 ORFs, in comparison with 35 Kb and 31 ORFs described in the published STX gene cluster of *C. raciborskii* T3 (Kellmann *et al.,* 2008a) (Figure 2.2.S5B). Only 20 ORFs are shared between these clusters (19 ORFs share 100% similarity); among these ORFs are all of the proposed genes necessary to synthesize STX. Thus, according to the genome size of D9, this is the minimum gene cluster thus far described for STX production.

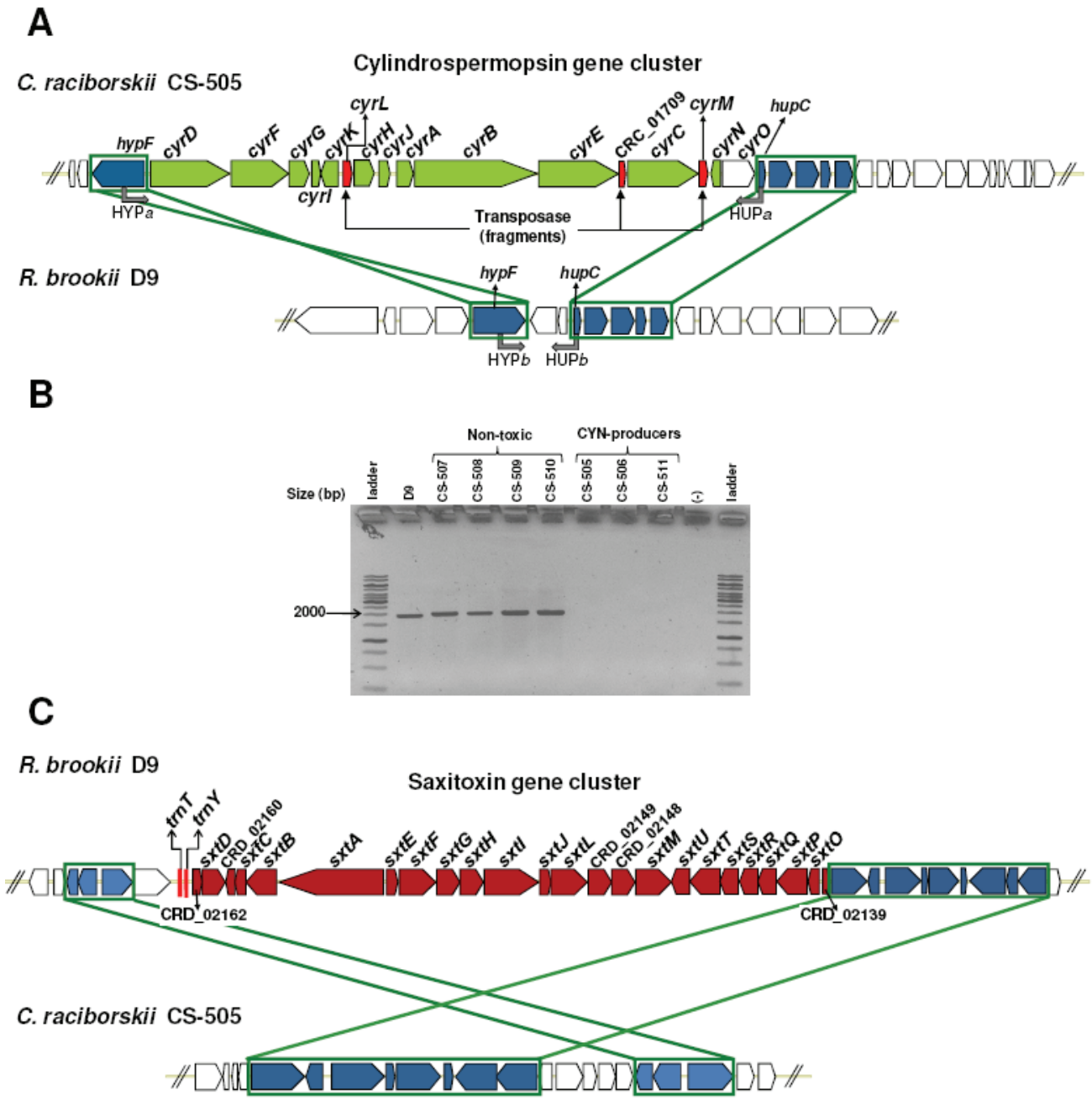


Figure 2.2.4 Schematic representation of the syntenic regions within the toxin gene clusters in CS-505 and D9. **(A)** Location of the CYN gene cluster of CS-505 compared with the syntenic genomic region in D9. **(B)** Gel electrophoresis of the PCR products from the *hypF/hupC* amplification in *R. brookii* D9 and in the non-toxic strains of *C. raciborskii*: CS-507, CS-508, CS-509 and CS-510. Producers of CYN: CS-505, CS-506 and CS-511 do not present amplification of the *hypF/hupC* region. **(C)** Location of the STX gene cluster of D9 compared with the syntenic genomic region in CS-505. Genes participating in syntenic regions are depicted in blue and highlighted in the green boxes within the arrows; genes outside the syntenic regions are depicted in white. tRNAs and transposases are shown in red. The grey arrows show the position of the primer pairs HYPa/HUPa and HYPb/HUPb used to amplify the region between *hypF* and *hupC* genes in different strains of *C. raciborskii* and in *R. brookii* D9, respectively. Ladder: GeneRuler 1kb DNA ladder (Fermentas, Ontario, Canada).

Tracing the evolution of traits in cyanobacteria

Access to the smallest known genomes of filamentous and heterocyst-forming cyanobacteria provided insights into the molecular basis and evolution of traits such as diazotrophy, filamentous growth, and the capacity for cellular differentiation. We assumed that protein sequences had to be drastically changed or newly developed to achieve new functions. Of course it is possible that only slight neofunctionalizations could be responsible for the observed phenotypic changes without major restructuring. In the latter case, the observable gene repertoire of all cyanobacteria would remain relatively stable with only the addition of paralogous genes with acquired new functions. Genes with new functions would turn up at specific evolutionary time points and then remain stable as long as the respective trait is expressed and positively selected. Our analysis by definition excluded genes that might have become indispensible over the time course of evolution in one or another species, but not in all species analyzed. We thus aimed at only the description of key innovations for the establishment of major evolutionary branches.

To this end, we collected available genomes of cyanobacteria from the databases and compared their gene repertoire. Unfortunately, due to difficulties in culturing, no genome of a branching filamentous species (e.g. from Stigonematales) is available for comparison (Gugger & Hoffmann, 2004), but all other major groups are well represented (see Materials and Methods). The availability of streamlined genomes of *C. raciborskii* and *R. brookii* further enhanced the resolution of our analysis. We made use of the whole genome sequences and subtracted step-wise common sets from the sets represented only in species with specific traits. We cannot exclude that some genes in these sets are not related with the specific trait, but the broader the species sampling the better is the resolution of the genes of the trait in question. Since our analysis is based on BLAST hits, the orthologous relationships between the genes in the respective species may not be clear. But as it turned out most genes in the gene sets had only one counterpart in each genome, thus representing most likely orthologous gene groups. Table 2.2.3 shows a summary of the number of core genes found for each of the three traits under study.

Table 2.2.3 Common genes for the different traits

	Hits between species* (see methods)	CS-505	D9	Core set present in wider spectrum of species
Filament formation	32	23	20	10
Diazotrophy	49	38	6	10
Heterocyst development	149	58	54	41

*Paralogs are not removed

Filament formation

Filaments are formed in groups III, IV and V of cyanobacteria (Tomitani *et al.,* 2006). Filament formation is also observed in unicellular cyanobacteria as well as in bacteria when several genes involved in cell division are interrupted by transposon mutagenesis (Miyagishima *et al.,* 2005). If filament formation is generally a loss-of-function mutation then filamentous species should lack some cell division genes. However, orthologs of all genes examined for their effect on this artificial filament formation are present in the filament-forming *Anabaena* (Table 2.2.S2). Filament formation is thus more likely a gain-of-function in the evolutionary context. When we compared all the available genomes of filamentous species, we found 32 genes present in all (Table 2.2.3, Table 2.2.4). Comparison of this set with the more streamlined genomes of CS-505 and D9 showed that only 23 and 20 genes are present, respectively. Since the D9 strain is able to form proper filaments, the additional three genes found in the CS-505 genome are unlikely to be directly involved in filament formation. The absence of these three genes in D9 points to the probability that some of the remaining 20 genes are also not associated with ability to form filaments. This is further underlined by the fact that the additional screen of the unfinished genome sequences of *Nostoc azollae* 078 and *Microcoleus chtonoplastes* PCC 7420 yielded a common set of only 10 genes. We conclude that filament formation in cyanobacteria needs at most 10 different gene products. Interestingly, besides the three genes previously thought to be associated with heterocyst formation (*hetR* and *patU3* and *hetZ*) all other seven genes correspond to only hypothetical proteins. Although mutations in these three genes do not produce a unicellular phenotype, it has been shown that they affect heterocyst development, and

hetZ and *patU3* also affect pattern formation (Buikema & Haselkorn, 1991, Zhang *et al.*, 2007). Their presence in cyanobacteria that does not present these phenotypes is suggestive of a different and more general function, which could be filament formation.

Insights into evolution of the 10 core genes were obtained by phylogenetic analysis (Figure 2.2.S6). The trees clearly show the high phylogenetic affiliation between CS-505 and D9, supported by bootstrap values of 100%, and the closest association to the CS-505/D9 branch to *Nostoc azollae*, supported for 9 of the core genes. All the core genes support the monophyly of heterocystous cyanobacteria (belonging to Subsection IV), consistent with previous reports based on 16S rRNA, *hetR* (Tomitani *et al.,* 2006) or phylogenomics (Shi & Falkowski, 2008, Swingley *et al.,* 2008). It is remarkable that seven of the core genes have an ortholog in *Synechococcus* sp. PCC 7335. The closest relationship of this organism with filamentous cyanobacteria has been reported by 16S rRNA phylogeny (Honda *et al.,* 1999), and an ortholog of HetR was also described (Zhang *et al.,* 2009). Our results strongly indicate that this organism could be the closest ancestor of filamentous cyanobacteria.

Although our BLAST analysis selected the gene pair CRC_00038/CRD_02583 as part of the core genes, only the branch formed by heterocystous cyanobacteria is resolved on the phylogenetic tree. Non-heterocystous cyanobacteria cluster with unicellular taxa, suggesting that this gene is part of a different family and therefore was removed from the core.

Table 2.2.4 Genes present only in filamentous species

Npun		gene product description	Anab	D9	CS-505
186680616		hypothetical protein	all1770	CRD_00231	CRC_00822
186680621	**core set**	hypothetical protein	all1765	CRD_00230	CRC_00821
186681198	**core set**	hypothetical protein	alr0202	CRD_00387	CRC_01215
186681299		hypothetical protein	all1340	no	no
186681300		hypothetical protein	all1339	no	no
186681350		PpiC-type peptidyl-prolyl cis-trans isomerase	alr1613	no	CRC_02567
186681409		HEAT repeat-containing PBS lyase	alr2986	CRD_00077	CRC_02169
186681476		peptidoglycan binding domain-containing protein	alr4984	CRD_02468	CRC_02058
186681631		nuclease	all2918	CRD_01392	CRC_01535
186681697	**core set**	hypothetical protein	alr2393	CRD_02002	CRC_01280
186681814		hypothetical protein	all3643	CRD_01982	CRC_01258
186681958	**core set**	hypothetical protein	all1729	CRD_02583	CRC_00038
186682138	**core set**	peptidase S48, HetR	alr2339	CRD_01519	CRC_03184
186682240	**core set**	PatU3	alr0101	CRD_02293	CRC_02800
186682241	**core set**	HetZ	alr0099	CRD_02292	CRC_02801
186682787		hypothetical protein	alr1555	no	no
186682808		peptidoglycan binding domain-containing protein	all1861	no	no
186683172		hypothetical protein	all5122	CRD_01021	CRC_02539
186683174		hypothetical protein	all2077	no	no
186683213		hypothetical protein	all1154	CRD_00512	CRC_00964
186683474		GDSL family lipase	all0976	no	no
186683904		hypothetical protein	all0215	CRD_00210	CRC_03261
186683953		GDSL family lipase	all1288	no	no
186684054		hypothetical protein	asr1049	no	no
186684093	**core set**	hypothetical protein	all2344	CRD_00085	CRC_00676
186684579	**core set**	hypothetical protein	all2320	CRD_01527	CRC_01389
186684586		hypothetical protein	all5091	CRD_02655	CRC_00188
186685511	**core set**	hypothetical protein	alr4863	CRD_02120	CRC_01594
186685539		NUDIX hydrolase	alr2015	CRD_01916	CRC_01834
186685973		hypothetical protein	all1007	no	CRC_00879
186685974		hypothetical protein	all1006	no	CRC_00878
186686413		hypothetical protein	all4622	no	no

Nitrogen fixation

Diazotrophy is an ancient character marking the lineage from which filamentous cyanobacteria seem to have evolved (Tomitani *et al.,* 2006, Swingley *et al.,* 2008). In comparing the gene repertoire of all available diazotrophic species with that of non-diazotrophic, we ended up with 49 genes that were present in at least eight of the nine genomes we chose for the first

analysis. These 49 genes comprise the upper limit of true inventions at this evolutionary juncture. As the functional classification confirms, most of the gene products are indeed involved in N_2 fixation (Table 2.2.5). The data set can be dissected into three distinctive categories: 1) the *nif* cluster and related genes, 2) the uptake hydrogenase gene cluster (*hupSL*) and endopeptidase specific for the uptake hydrogenase *hupW,* and 3) finally, a set of genes involved in general metabolism and hypothetical proteins.

Three genes coding for hypothetical proteins normally located between *hupSL* and maturation hydrogenase gene clusters (*hypABCDEF*) (Tamagnini *et al.,* 2007) belong to the group of 49 genes, suggesting their key role in N-metabolism. Part of the set also comprises genes found to be up-regulated under N-depletion in *Anabaena* (Ehira *et al.,* 2003) (Table 2.2.5). Interestingly, the CS-505 strain does not have the full set of 49 genes. Genome comparison with this strain thus further narrows the set of gene products needed for diazotrophy down to only 38. This indicates that a streamlined genome like that of *C. raciborskii* may be able to dispense with some otherwise needed genes. Analysis of several further genomes to account for species variability allowed us to define an indispensable core gene set for all species. It turned out that in some *Cyanothece* and extremophile *Synechococcus* genomes many of the previously found common genes were not present, e.g. the uptake hydrogenase and related genes and genes that show changes in expression in heterocysts are missing (Table 2.2.5). *Microcoleus chthonoplastes* has not been classified as a N_2 fixing cyanobacteria, however, its genome contains the *nifHDK* and *nifEN* gene clusters with similarity to δ-proteobacteria rather than cyanobacteria suggesting that this cluster was transferred by HGT (Bolhuis *et al.,* 2009). When we considered the *Cyanothece*, *Synechococcus* and *M. chthonoplastes* PCC 7420 genomes, our core set was highly reduced to 10 genes: the *nif* gene cluster and related genes and *patB*. PatB has an N-terminal-ferredoxin and a C-terminal helix-turn-helix domain suggesting its function as a redox-sensitive transcription factor (Liang *et al.,* 1992). Furthermore, in *Anabaena,* a *patB* deletion mutant was completely defective for diazotrophic growth, but in the wildtype, its expression was restricted to heterocysts (Jones *et al.,* 2003). The presence of *patB* as part of the core gene for diazotrophic cyanobacteria suggests that this gene is also essential in unicellular and non-heterocystous diazotrophic cyanobacteria.

Table 2.2.5 Genes present in N_2- fixing species

Npun	Gene		Gene product	Anab	D9	CS-505	Absent in
186680715			glycerophosphoryl diester phosphodiesterase	all1051	CRD_01538	CRC_01381	SynJA3, SynJA2, Cya7425
186680864	*patB*	**core set**	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein	all2512	no	CRC_01763	
186680869			hypothetical protein	alr2517*	no	CRC_03082	SynJA3, SynJA2, Cya8801
186680870			cupin 2 domain-containing protein	alr2518*	no	CRC_03081	SynJA3, SynJA2, Mcht
186680871			nitrogenase-associated protein	alr2520*	no	CRC_03080	Mcht
186680875			hypothetical protein	asr2523*	no	CRC_02152	SynJA3, SynJA2, Cya7425
186680876			hypothetical protein	alr2524*	no	CRC_02151	SynJA3, SynJA2, Cya7425, Mcht
186680892			NHL repeat-containing protein	alr0693	no	no	SynJA3, SynJA2, Cya7425, Mcht
186680893			Rieske (2Fe-2S) domain-containing protein	alr0692	no	no	
186680895			hypothetical protein	asr0689	no	CRC_01692	SynJA3, SynJA2, Cya7425, Mcht
186680897	*hupS*		Ni Fe-hydrogenase small subunit, HupS	all0688	no	CRC_02736	SynJA3, SynJA2, Cya7425, Mcht
186680898	*hupL*		Ni Fe-hydrogenase large subunit, HupL	all0687	no	CRC_02737	SynJA3, SynJA2
186680903	*hupW*		hydrogenase maturation protease	alr1423	no	CRC_01049	SynJA3, SynJA2, Cya7425
186680908	*feoA*		FeoA family protein	asl1429	no	CRC_02875	Tery, Lyng, Nspu
186680909	*fdxH*		ferredoxin (2Fe-2S)	all1430	no	CRC_02876	Mcht
186680910	*hesB*		iron-sulfur cluster assembly accessory protein	all1431	no	CRC_02877	SynJA3, SynJA2
186680911	*hesA*		UBA/THIF-type NAD/FAD binding protein	all1432	no	CRC_02878	SynJA3, SynJA2
186680912	*nifW*		nitrogen fixation protein	all1433	no	CRC_02879	Mcht
186680913			protein of unknown function DUF683	asl1434	no	CRC_02880	Tery
186680914			protein of unknown function DUF269	all1435	no	CRC_02881	Mcht
186680915	*nifX*		nitrogen fixation protein	all1436	no	CRC_02882	Mcht
186680916	*nifN*	**core set**	nitrogenase molybdenum-iron cofactor biosynthesis protein NifN	all1437	no	CRC_02883	
186680917	*nifE*	**core set**	nitrogenase MoFe cofactor biosynthesis protein	all1438	no	CRC_02884	
186680918			Mo-dependent nitrogenase family protein	all1439	no	no	Tery
186680919	*nifK*	**core set**	nitrogenase molybdenum-iron protein beta	all1440	no	CRC_02885	

			chain				
186680550	*nifD*	**core set**	nitrogenase molybdenum-iron protein alpha chain	all1454	no	CRC_02886	
186680941	*nifH*	**core set**	nitrogenase iron protein NifH	all1455	no	CRC_02887	
186680943	*nifU*	**core set**	Fe-S cluster assembly protein NifU	all1456	no	CRC_02888	
186680944	*nifS*	**core set**	Nitrogenase metalloclusters biosynthesis protein NifS	all1457	no	CRC_02889	
186680946	*nifB*	**core set**	nitrogenase cofactor biosynthesis protein	all1517	no	CRC_02891	
186680953	*cysE*		serine acetyltransferase	alr1404	no	no	SynJA3, SynJA2
186680954			hypothetical protein	asr1405*	no	no	
186680955		**core set**	hypothetical protein	asr1406*	no	CRC_02071	
186680956	*nifV*		homocitrate synthase	alr1407	no	CRC_02070	Tery, Mcht
186680957	*nifZ*		NifZ family protein	asr1408	no	CRC_02069	Mcht
186680958	*nifT*		NifT/FixU family protein	asr1409	no	CRC_02068	Mcht
186680959			hypothetical protein	alr1410	no	CRC_02067	Tery
186682206			hypothetical protein	all0969	no	no	SynJA3, SynJA2
186682693			ribokinase-like domain-containing protein	alr4681	CRD_01205	CRC_01938	SynJA3, SynJA2, Cya7425
186683057			hypothetical protein	alr0857	no	no	SynJA3, SynJA2
186683906			pathogenesis related protein-like protein	all0217	no	CRC_03259	SynJA3, SynJA2, Cya7425, Mcht
186684105			glycosyl transferase, group 1	all1345	CRD_02459	no	SynJA3, SynJA2, Cya7424
186684241			hypothetical protein	all4434	CRD_01931	CRC_02458	SynJA3, SynJA2, Cya7425
186685158			phosphoglycerate mutase	alr2972	CRD_00352	CRC_03094	Tery, SynJA3, SynJA2, Cya7425
186685476			hypothetical protein	asl0163	no	no	SynJA3, SynJA2
186685625			hypothetical protein	all3713*	no	no	SynJA3, SynJA2, Cya7425, Cya8801
186685845			hypothetical protein	asl0597	no	no	SynJA3, SynJA2, Cya7425
186686227			Arginyl tRNA synthetase anticodon binding	all3951	CRD_01597	CRC_03274	
186686347			cytochrome P450	all1361	no	no	SynJA3, SynJA2, Cya7425

* Genes that show regulation in *Anabaena* under N_2- depletion (Ehira *et al.,* 2003)

Further evidence for the correct logic of our approach is provided from genomic data of the D9 strain. This cyanobacterium has lost the ability to fix N_2, and as we show in our analysis, it lost genes involved in this process. Indeed, only 6 of the 49 genes were detected in this strain. The gene products of these are related to phosphoglycerol metabolism and therefore are possibly involved in membrane degradation/synthesis. If they were once involved in N_2 fixation they likely now fulfill indispensable functions, such that a loss would lead to decreased fitness or lethality.

Heterocyst development

The process of heterocyst differentiation has been described in detail only in *Anabaena* and *Nostoc punctiforme* ATCC 29113, which develop intercalated heterocysts in a specific pattern (Zhang *et al.*, 2006). There are no published studies on heterocyst differentiation in cyanobacterial genera that develop terminal heterocysts, such as *Cylindrospermopsis* or *Cylindrospermum*. Our comparative genomic screen for genes restricted to heterocyst-forming species delivered an overwhelming number of 149 genes (Table 2.2.3, Table 2.2.S3). This high number can be explained by the fact that only a few genomes of heterocyst-forming species are currently known, but they are also rather closely related. Some of the 149 genes may be involved in the formation of intercalating heterocysts. If we include the genomes of our strains in this analysis only 58 unique genes remain as common to all heterocyst-forming species. A further slight reduction in gene numbers to 41 was achieved by including the symbiont *Nostoc azollae* in the analysis. Of these genes, only one, *patN*, is currently described as involved in pattern formation.

In *Anabaena*, 77 genes are described as having a function in heterocyst differentiation, but only 55 of these have a homolog in the *C. raciborskii* CS-505 genome (Table 2.2.S4). Surprisingly, one of the genes thus far seen as essential for heterocyst differentiation, *hetC*, is absent. This gene represses the expression of *ftsZ* in early stages of heterocyst differentiation in *Anabaena*, and the Δ*hetC* mutant generates multiple clusters of uncompromised pro-heterocysts along the filament that are capable of cell division and elongation (Xu & Wolk, 2001). Other genes related with key steps in heterocyst differentiation that are absent in CS-505 include: *ccbP*, whose product has been shown to regulate the calcium availability for heterocyst formation and negatively regulate the heterocyst differentiation (Zhao *et al.,* 2005); *hetL*, a positive regulator of the differentiation process that interferes with *patS* inhibition process (Liu & Golden, 2002); and

hetN, a suppressor of heterocyst differentiation involved in the maintenance of the delayed heterocyst spacing pattern (Callahan & Buikema, 2001). Lack of these genes again points to streamlining in the *C. raciborskii* genome and could possibly be attributed to a terminal rather than an intercalating heterocyst formation.

PatS, or a pentapeptide PatS-5, have been proposed to be diffusible molecules acting as an inhibitor of heterocyst differentiation (Yoon & Golden, 1998). In the current model, HetR activates the expression of *patS*, and PatS or a derivative diffuses laterally to inhibit the differentiation of the neighboring cells by acting negatively on the DNA-binding activity of HetR (Zhang *et al.,* 2006). No CDS for a protein with the typical characteristics described for PatS, i.e. a diffusible penta-peptide with distinctive last C-terminal amino-acid, was found in the CS-505 and D9 genomes. Zhang *et al.,* (2009) showed that a *patS*-like CDS of the non-diazotrophic *Arthrospira platensis* could complement a *patS* mutant of *Anabaena*, despite the fact that its conserved penta-peptide (RGSGR) was not in the last amino acids of the predicted protein. Thus, other penta-peptide-containing proteins could have taken over the function of PatS in *C. raciborskii*. A deeper analysis for proteins containing the penta-peptide revealed a CDS that had the penta-peptide in the C-terminal region in the genomes of CS-505 and D9. We propose therefore that if *patS* exists in these cyanobacteria, the most likely candidates are CRC_02157 and CRD_02133 in CS-505 and D9, respectively.

Our analysis also revealed that 39 of the 55 heterocyst differentiation-related genes in CS-505 are also present in the non-heterocystous *R. brookii* D9 genome (Table 2.2.S4). In *Nostoc punctiforme,* the cellular differentiation pathways for hormogonia, akinetes and heterocysts are reported to have genes with common expression profiles (Campbell *et al.,* 2007). Since we observed akinetes in strain D9 by optical microscopy, it is most likely that these genes with described functions in heterocyst differentiation have additional functions and/or are involved in other cellular differentiation processes such as akinete formation.

Terminal cell differentiation was evident from electron micrographs (Figure 2.2.1) of both strains. Nevertheless, terminal cell differentiation in D9 resembles the morphology of an immature heterocyst of CS-505 (not shown), suggesting incomplete heterocyst development. The final steps in heterocyst development involve the synthesis and deposition of an inner glycolipid layer and then covering with a polysaccharide envelope (Zhang *et al.*, 2006). These layers isolate the newly differentiated cell from external oxygen. The inner glycolipid layer is synthesized by a

cluster of genes involving a polyketide synthase (PKS) pathway and glycosyltransferases in *Anabaena* (Awai & Wolk, 2007, Fan *et al.*, 2005). *Cylindrospermopsis raciborskii* CS-505 contains most of these genes (*hglEGDCA* and *hetM*), with the exception of the aforementioned *hetN* (Figure 2.2.1, Table 2.2.S4). In *Anabaena*, *hetN* is located adjacent to the phosphopanteinyltransferase *hetI*. In CS-505, however, a sequence similar to *hetI* is located at a different locus, implying structural differences in the glycolipid layer between the CS-505 and *Anabaena* heterocysts. In any case, further analysis must be done to understand the implications on the glycolipid structure of *C. raciborskii.*

The N_2-fixation and heterocyst glycolipid clusters are not present in the D9 genome (Figure 2.2.1). Only *hetI* is present in the D9 genome, which suggests a different role of this gene in this strain. Strain CS-505 and D9 genomes do, however, contain an identical gene arrangement of the genes necessary for the synthesis of the polysaccharide envelope of the heterocysts (Huang *et al.*, 2005) (Table 2.2.S4, Figure 2.2.S7). Since neither heterocyst formation nor N_2-fixation occur in D9 it seems unlikely that the polysaccharide layer is properly deposited in the terminal D9 cells. Indeed, when we stained for polysaccharide with Alcian blue, D9 filaments were homogeneously but only slightly stained (Figure 2.2.1), indicating that polysaccharide was being synthesized, but not as a heterocyst-protective layer.

We note that because the D9 strain shows features of heterocyst formation, we expect that most gene products responsible for this trait are also encoded for in this strain. Indeed, only five genes of the smallest common set of 41 genes are not present in the D9 strain. The lack of these five genes could be entirely responsible for the incomplete heterocyst formation in D9. Unfortunately, no function is as yet assigned to these genes. Most of the other genes have also no assigned function for their gene products.

Although all five genes are annotated as conserved hypothetical proteins, the intensive studies of N-metabolism and heterocyst development in *Anabaena* allowed us to search for possible functions of these five genes absent in D9. Indeed, we found possible functions for three genes. Both alr2522 and all0721 were shown to be up- regulated in a mutant expressing HE0277, a homolog of the sigma factor *sigJ* of alr0277 that confers resistance to desiccation by up-regulating genes involved in polysaccharide synthesis (Yoshimura *et al.,* 2007). Furthermore, all1814 was up-regulated after 8h of N-depletion and showed no significant regulation in an *nrrA* mutant, an N-regulator that facilitates heterocyst development (Ehira & Ohmori, 2006). Together

this evidence suggests that alr2522 and all0721 are involved in polysaccharide biosynthesis and that all1814 is related to a stage of heterocyst development. Their absence in D9 makes them ideal targets for further functional studies on heterocyst development.

2.2.4 Conclusions

The innovations of diazotrophy, filamentous growth, photosynthesis and the capacity for cellular differentiation are major defining events in the evolution of cyanobacteria. Given that the free-living cyanobacteria *C. raciborskii* CS-505 and *R. brookii* D9 have the smallest known genomes among filamentous cyanobacteria, they are ideal subjects for exploration of the development and modifications of these characteristics among cyanobacteria. In spite of their relatively small genomes, these strains are nevertheless capable of cell differentiation. We are likely observing evidence of genetic streamlining, pointing towards the minimum set of genes required for these traits. Remarkably, strain CS-505 is able to develop a functional heterocyst without supposedly "essential" genes, such as, *hetC, hetN, hetL* and *ccbp*.

The *C. raciborskii* CS-505 and *R. brookii* D9 strains have geographically disjunct origins within tropical freshwater ecosystems. We expect that they have been genetically isolated and hence have evolved independently. Nevertheless, on the basis of 16S rRNA they are virtually identical and form a monophyletic cluster, with a close phylogenetic affiliation among filamentous cyanobacteria. The morphological criteria originally used to discriminate between these strains and to assign them to different genera obviously reflect differential genetic editing primarily associated with cell differentiation and functional heterocyst formation rather than their phylogenetic relationships. With respect to this high similarity in 16S rRNA, as well as that revealed in our phylogenomic analysis (>90% identity their 2,539 shared genes), we propose that these strains are congeneric. This evidence suggests that the strain differences may represent an example of allopatric speciation.

Our genomic analysis provides support for the idea that cyanobacteria are capable of evolving according to highly diverse strategies for genomic organization and adaptive mechanisms. Whereas CS-505 shows evidence of phenotypic plasticity and has a more elaborate genome, perhaps via gene acquisition and rearrangement, D9 has apparently adapted by avoiding lateral gene transfer and by losing genes. These alternative strategies have important implications

for the adaptive radiation of filamentous cyanobacteria and at least partially account for their evolutionary success in a multitude of environments over enormously long time-scales.

2.2.5 Materials and Methods

Isolation and culture of cyanobacterial strains

Cylindrospermopsis raciborskii strain CS-505 was clonally isolated in 1996 from a water supply at the Solomon Dam, Australia (Saker *et al.,* 1999) and obtained from the culture collection of the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. *Raphidiopsis brookii* D9 (originally classified as *C. raciborskii*) was isolated from a mixed plankton sample collected in 1996 from the Billings freshwater reservoir near Sao Paulo, Brazil and subsequently recloned from a single filament. Strain CS-505 produces cylindrospermopsin (CYN) and deoxy-cylindrospermopsin (doCYN) (Saker *et al.,* 1999), but no PSP toxins. Strain D9 constitutively produces the following PSP toxins, as confirmed by LC-MS/MS: saxitoxin (STX), C-11 O-sulfated gonyautoxins (GTX2/3), and their respective decarbamoyl derivatives (dcSTX and dcGTX2/3) as minor components (Castro *et al.,* 2004).

The non-axenic cultures were grown in 250 ml flasks containing 100 ml of MLA growth medium (Castro *et al.,* 2004) without aeration at 23 ºC under fluorescent light at a photon flux density of 35 μmol m^{-2} s^{-1} on a 12:12h light/dark photocycle. To minimize bacterial contamination several wash steps were performed after harvesting and the absence of eubacterial DNA was checked by PCR as previously described (**Publication I**).

Preparation and sequencing of genomic DNA

Long strands of genomic DNA were obtained by purifying DNA embedded on low melting point (LMP) agarose plugs. Intact chromosomal DNA embedded on agarose plugs was obtained from 100 ml of healthy cultures in mid-exponential growth phase as previously described (**Publication I**).

Sequencing was conducted with the BigDye kit from ABI (Foster City, USA) using standard forward and reverse primers; pre-assembly trimming was performed with a modified version of Phred (Ewing *et al.*, 1998, Ewing & Green, 1998).

Genomic libraries for the 454/gs20 system were prepared according to the manufacturer's protocols (454 Life Sciences Corporation, Branford, CT, USA). Three runs each were performed on the 454/gs20 sequencing system. All 454/gs20 sequence data were assembled according to species-specific criteria with the newbler assembler software (http://www.454.com). The Sanger-based sequencing reads were assembled onto this backbone. Clone gaps then were filled by a primer walking strategy with custom primers. The genome sequences of CS-505 and D9 were deposited in the NCBI genome database under the main accession numbers: ACYA00000000 (CS-505) and ACYB00000000 (D9).

RFLP by PFGE and genome size estimation

Intact chromosomal DNA from strains CS-505 and D9 embedded into agarose plugs was digested with the high frequency cut restriction enzyme *Mlu* I as described previously in **Publication I**. The estimation of the genomic sizes was achieved using as standard, genomic DNA from *Vibrio parahaemolyticus* RIMD 2210633 (Vpkx) digested with *Not* I (Hara-Kudo *et al.,* 2003). The genome sequence of Vpkx is available and comprises two chromosomes of 3.3 and 1.8 Mb. A calibration curve (r^2 = 0.99012) was generated assuming a logarithmic migration of the DNA fragments during the electrophoresis. Bands showing higher intensity in CS-505 and D9 were considered as double.

RNA extraction, cDNA synthesis and qPCR detection of genes involved in N-metabolism and heterocyst pattern formation

Total RNA was extracted from mid-exponential growth cultures using Rneasy mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Briefly, frozen samples were thawed on ice, and approximately two small spatulas full of 0.1 mm diameter glass beads were added to the sample. The cells were sheared for 2 X 120 s using a Qiagen Bead Beater (Hilden, Germany). The supernatant was separated from the glass beads and cell debris by

centrifugation (10 min, 13,200 rpm, 4°C). A DNA digestion was performed after the RNA isolation for 1 h at RT and a final clean up with a second DNA on tube digestion was performed to remove any possible DNA remaining on the RNA samples. RNA was quantified using the NanoDrop ND- 1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA) and the RNA purity was assessed using the ratio A_{260}/A_{230}. If required, samples were further cleaned using the Microcon Elute System (Millipore, Massachusetts, USA). RNA integrity was checked on RNA nano chips using the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA).

cDNA was synthesized from 500 ng total RNA using random hexamers with the Omniscript RT kit from Qiagen. All primers were designed using the Primer Express 3.0 (Applied Biosystems, Darmstadt, Germany) software and synthesised from MWG Biotechnologies (Eurofins MWG Operon, Ebersberg, Germany). Primer sequences are as follows:

q505nifHF 5' GCTGCTGAGAAAGGTGCTGTAG 3`	q505nifHR 5' GGAATCCGGCCAGCATTACT 3'
q505fdxHF 5' CAGCAGCAGAAGCAGATATTGAA 3'	q505fdxHR 5'CCACACAACTAGAGCAAGAACCA3'
q505nifBF 5' TCAACCAGGTCCGACTAATAAAGAA 3'	q505nifBR 5' CGCATCTGCTCGACACTGA 3'
q505nifVF 5' TTGGAACCGAGCAGTTATTTCA 3'	q505nifVR 5' GCTAGCGATTTGGATTCCAGAA 3'
qhetRF 5' AGTTGCCACAGCAGCATCAA 3'	qhetRR 5' GGAGAGTCAATCCGGGTAACC 3'
qntcAF 5' TTTTACTGCGGTGGAATTGCT 3'	qntcAR 5' TTCCTTGAGGGCCTGCTCTAC 3'
q505patAF 5' AACGTCAATTCCTTGGTGCAA 3'	q505patAR 5' AGACCAGATTCTTCTCGGGAAAC 3'
q505hetMF 5' TAGGCGCGGAAGCAGTTCT 3'	q505hetMR 5' CCAGTTGCTCCGGTTATCAAG 3'
qhepAF 5' TGGTGCATCAGGTGGTGGTA 3'	qhepAR 5' CCCGTTAGTTGGTTCATGGAA 3'

The expression of nitrogen metabolism genes was tested in CS-505 and D9 (Table 2.2.6). A 20 µl qPCR reaction was composed of 1 µl of a 10-fold diluted cDNA, forward and reverse primers at a concentration of 100 nM, and 10 µl 2x SYBR Green PCR Master Mix (Applied Biosystems, Darmstadt, Germany). Pure RNA was used to test for the presence of genomic DNA in the RNA samples. Cycle parameters were as follows: initial denaturation at 95 °C/10 min, followed by 40 cycles of 95 °C/15 sec and 59 °C/1 min. Finally a product-primer dissociation step was utilized to verify formation of a single unique product/primer dimerization.

Table 2.2.6 Expression of N_2 fixation and heterocyst differentiation genes

Gene	Efficiency (%)	Slope	r^2	Ct CS-505	Ct D9
nifH	93.8	-3.11	0.9934	22.30	ND
fdxH	NT	NT	NT	26.23	ND
nifB	NT	NT	NT	25.22	ND
nifV	NT	NT	NT	25.29	ND
hetR	104	-3.46	0.9996	21.45	14.11
ntcA	102.7	-3.41	0.9986	21.38	15.01
patA	89.8	-2.99	0.9918	20.77	ND
hepA	NT	NT	NT	26.04	14.40
hetM	NT	NT	NT	21.38	ND

NT: Not tested. ND: Not detected

Genomic DNA isolation and PCR amplification of the hydrogenase genes *hupC* and *hypF*

DNA was extracted using the CTAB method (Ausubel *et al.,* 1992). The following PCR primers targeting the region upstream and downstream the CYN gene cluster were used: HYPa 5-GGGGTGGACAGTGGTCATAC-3, HUPa 5-TGGGTGTTCCTCATCAACAA-3 forward and reverse repectively, and targeting the region between the maturation hydrogenase genes *hypF* and *hupC* of D9: HYPb 5-CCTCCAAACGATGGAGGAAT-3, HUPb 5-GGGTGTTCCTCATCCACAAT-3 forward and reverse, respectively (Figure 2.2.4). The PCR reaction contained 50–100 ng of genomic DNA, the reagents for each 30 µl amplification reaction were: 0.25 U Taq DNA polymerase (Invitrogen, California, USA); 3 µl 10X PCR buffer; 2.5 mM MgCl; 0.4 mM primers; and 0.93 mM of each deoxynucleoside triphosphate (Promega, Wisconsin, USA). Thermal cycling was performed in an Eppendorf Mastercycler, under the following conditions: initial DNA denaturation at 99 ºC/1 min, 30 cycles 94 ºC/15 sec, 53 ºC/1 min, 72 ºC/1.5 min and a final extension at 72 ºC/7 min. All PCR products were verified by gel electrophoresis (1 % agarose) and visualized under UV transilumination after staining with ethidium bromide.

Repeat analysis

For the CS-505 genome we calculated all supermaximal repeats. A supermaximal repeat is defined as follows:
A pair of substrings $R = ((i_1, j_1), (i_2, j_2))$ is a *repeated pair* if and only if $(i_1, j_1) \neq (i_2, j_2)$ and $S[i_1..j_1] = S[i_2..j_2]$. The length of R is $j_1 - i_1 + 1$. A repeated pair $((i_1, j_1), (i_2, j_2))$ is called *left maximal* if $S[i_1 - 1] \neq S[i_2 - 1]$ and *right maximal* if $S[j_1 + 1] \neq S[j_2 + 1]$. A repeated pair is called *maximal* if it is left and right maximal.
A substring ω of S is a (*maximal*) *repeat* if there is a (maximal) repeated pair $((i_1, j_1), (i_2, j_2))$ such that $\omega = S[i_1..j_1]$. A *supermaximal repeat* is a maximal repeat that never occurs as a substring of any other maximal repeat.
For the given contigs of *C. raciborskii* CS-505 we found 258,229 different supermaximal repeats covering 98.52 % of the whole sequence.

In a second step we clustered all supermaximal repeats close to each other and with similar distances between their positions in the genome. This helps to find larger degenerated repeats because they contain several exact super maximal repeats.

For the clustering, each supermaximal repeat containing more than two copies was decomposed into all possible copy pairs. Those pairs were then clustered according to similar first positions of the first copy respectively and according to similar distances between the copies. We selected 500 nt as the maximal allowed difference between the two first positions. The maximal allowed difference between the distances was 100 nt.

With the given parameter setting we got 5,390 different clusters. For the 20 clusters with the best score (= total copy length · copy pair amount), we performed a motif search with at least 60 % sequence identity on both strands and in both directions. All hits of the 20 clusters cover 3.94 % of the whole sequence.

Using *Nostoc punctiforme* as member of the group with all traits analyzed as “template”, we performed blastp analyses against all other genomes. We applied a score threshold of 150 to get rid of spurious hits. Remaining hits were analyzed with respect to their occurrence in four different groups: non-N_2-fixing, N_2-fixing, filamentous N_2-fixing and filamentous heterocyst-forming N_2-fixing.

A further analysis was then done with these sets including genomes from a wider range of species to get the true core sets for the traits: *Nostoc azollae* strain 0708 is a symbiotic cyanobacterium with duckweed; *Microcoleus chtonoplastes* PCC 7420 possesses multiple filaments in one mucous sheath, and *Arthrospira maxima* CS-328 belongs to Section III of the cyanobacteria. Different *Cyanothece* and *Synechococcus* strains were used to account for species variability.

Bioinformatics analysis

The cyanobacterial taxa used for comparative genomic analyses are listed in Table 2.2.7.

Table 2.2.7 Characteristics of the cyanobacterial taxa used for comparative genomic analyses

Species	Morphology	Diazotrophy	Accession number	Genome sequence status
Nostoc punctiforme PCC 73102 (Npun)	Filamentous, heterocystous	N_2-fixing	NC_010628	finished
Nodularia spumigena CCY9414 (Nspu)	Filamentous, heterocystous	N_2-fixing	NZ_AAVW00000000	unfinished
Anabaena sp. PCC 7120 (Anab)	Filamentous, heterocystous	N_2-fixing	NC_003272	finished
Anabaena variabilis ATCC 29413 (Avar)	Filamentous, heterocystous	N_2-fixing	NC_007413	finished
Nostoc azollae strain 0708* (Nazo)	Filamentous, heterocystous	N_2-fixing	NZ_ACIR00000000	unfinished
Trichodesmium erythraeum IMS101 (Tery)	Filamentous	N_2-fixing	NC_008312	finished
Lyngbya sp. PCC 8106 (Lyng)	Filamentous	N_2-fixing	NZ_AAVU00000000	unfinished
Microcoleus chthonoplastes PCC 7420* (Mcth)	Filamentous	N_2-fixing	NZ_ABRS00000000	unfinished
Arthrospira maxima CS-328* (Amax)	Filamentous	non-N_2-fixing	NZ_ABYK00000000	unfinished
Crocosphaera watsonii WH8501 (Cwat)	Unicellular	N_2-fixing	NZ_AADV00000000	unfinished
Cyanothece sp. ATCC 51142 (Cya51142)	Unicellular	N_2-fixing	NC_010546	finished
Cyanothece sp. PCC 8801* (Cya8801)	Unicellular	N_2-fixing	NC_011726	finished
Cyanothece sp. PCC 7424* (Cya7424)	Unicellular	N_2-fixing	NC_011729	finished
Cyanothece sp. PCC 7425* (Cya7425)	Unicellular	N_2-fixing	NC_011884	finished
Synechococcus sp. JA-3-3Ab* (SynJA3)	Unicellular	N_2-fixing	NC_007775	finished
Synechococcus sp. JA-2-3B'a(2-13)* (SynJA2)	Unicellular	N_2-fixing	NC_007776	finished
Synechococcus sp. CC9311	Unicellular	non-N_2-fixing	NC_008319	finished
Synechocystis sp. PCC 6803	Unicellular	non-N_2-fixing	NC_000911	finished
Acaryochloris marina MBIC11017	Unicellular	non-N_2-fixing	NC_009925	finished
Gloeobacter violaceus PCC 7421	Unicellular	non-N_2-fixing	NC_005125	finished
Microcystis aeruginosa NIES-843	Unicellular	non-N_2-fixing	NC_010296	finished
Prochlorococcus marinus MIT 9301	Unicellular	non-N_2-fixing	NC_009091	finished
Synechococcus elongatus PCC 7942	Unicellular	non-N_2-fixing	NC_007604	finished
Thermosynechococcus elongatus BP-1	Unicellular	non-N_2-fixing	NC_004113	finished

*Species included in the second part of the analysis.

2.2.6 Supplementary material

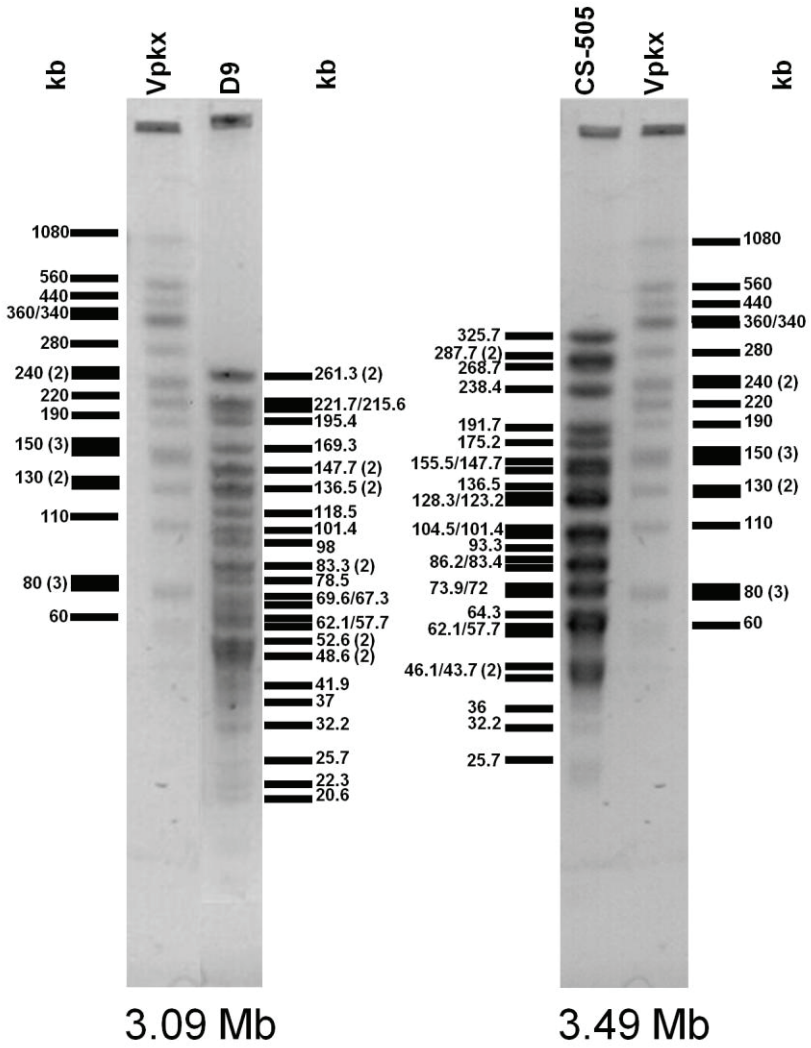


Figure 2.2.S1 Size estimation of D9 and CS-505 genomes by PFGE restriction analysis. Restriction profiles were obtained by *Mlu* I digestion. SC: Chromosomic DNA from *Saccharomyces cerevisiae.* Vpkx: Genomic DNA from *Vibrio parahaemolyticus* RIMD 2210633 digested with *Not* I. PFGE electrophoresis conditions are described in **Publication I**

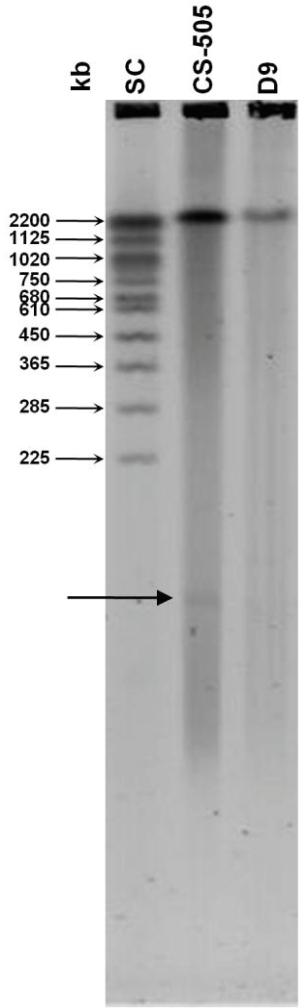


Figure 2.2.S2 Possible extrachromosomal element in the CS-505 genome. PFGE of chromosomic DNA from strains D9 and CS-505, the possible plasmid is indicated by the arrow. SC: Chromosomic DNA from *Saccharomyces cerevisiae.*

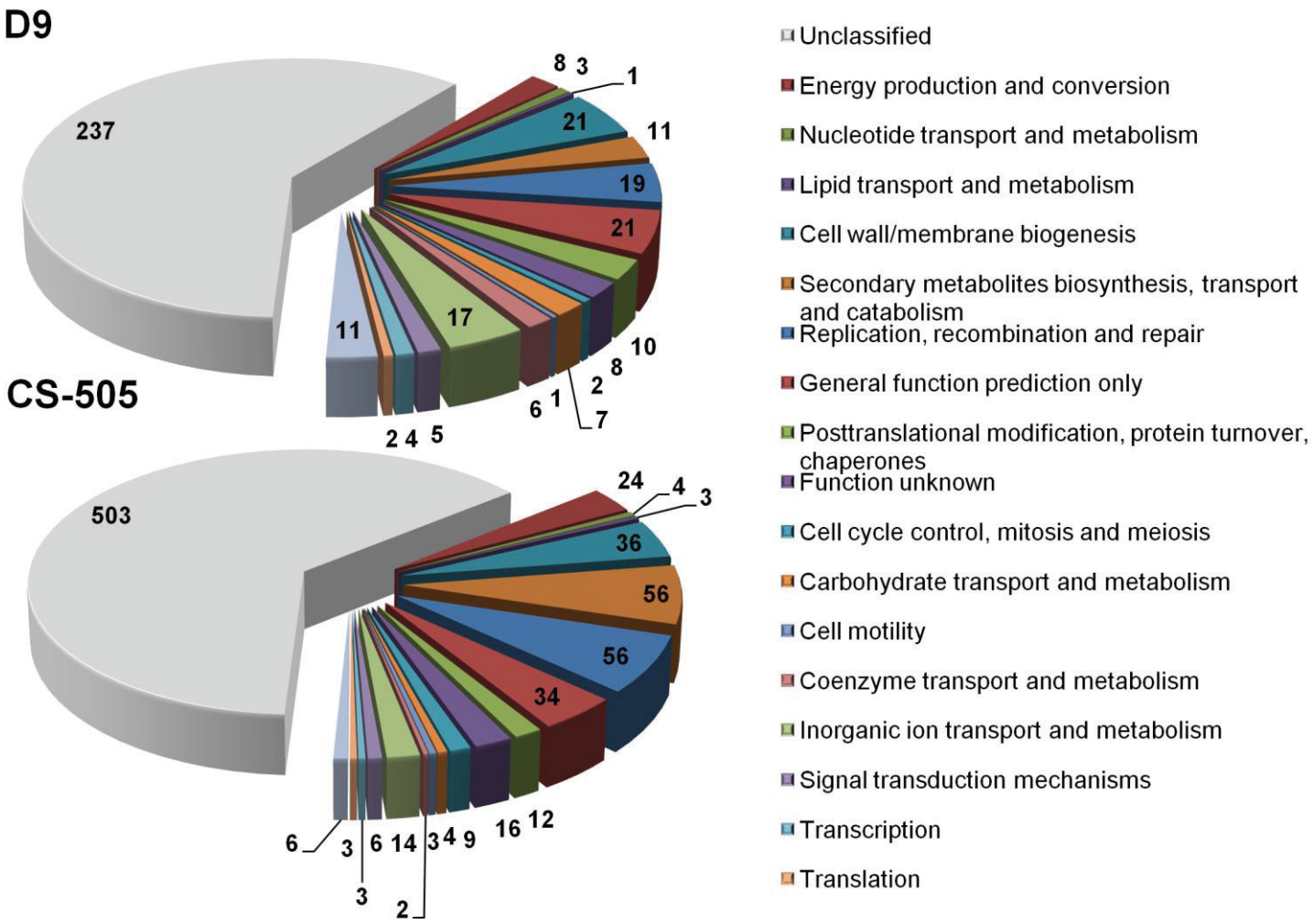


Figure 2.2.S3 Distribution of the total unique CDS of CS-505 and D9 into Cluster of Orthologous Groups (COGs). Unique CDS were obtained by a Best-Bidirectional Hits (BBHs) search between both genomes using a 30% cutoff.

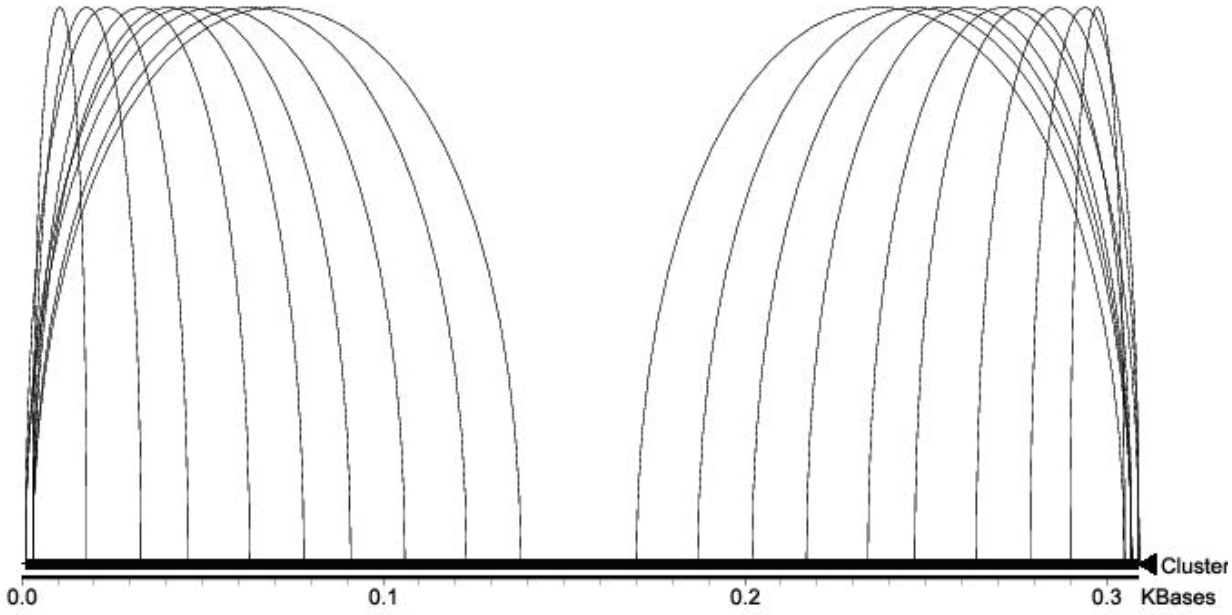


Figure 2.2.S4 Repeated sequences in a repeat unit as revealed by an analysis using miropeats. The analysis was performed according to Parsons (1995) with a threshold score of 100.

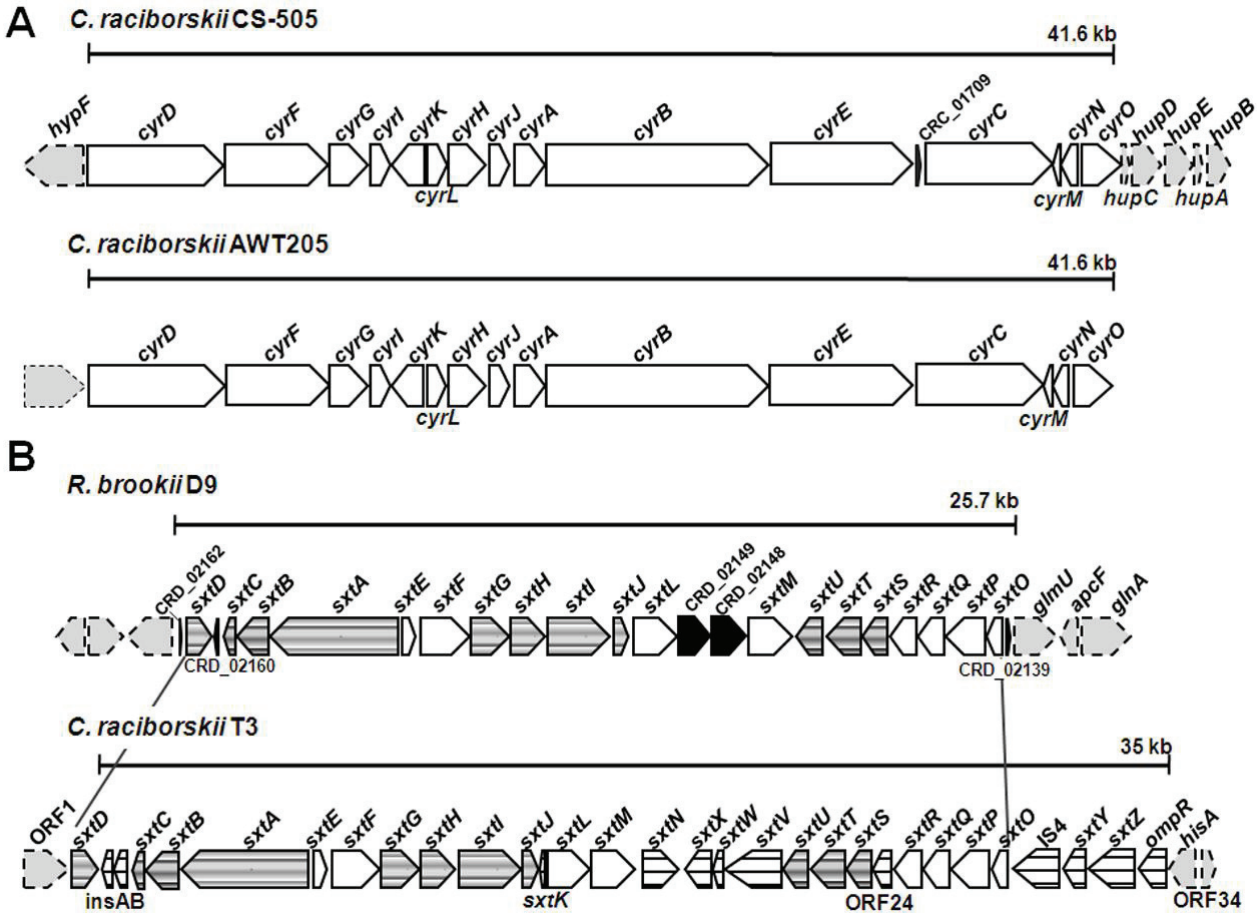
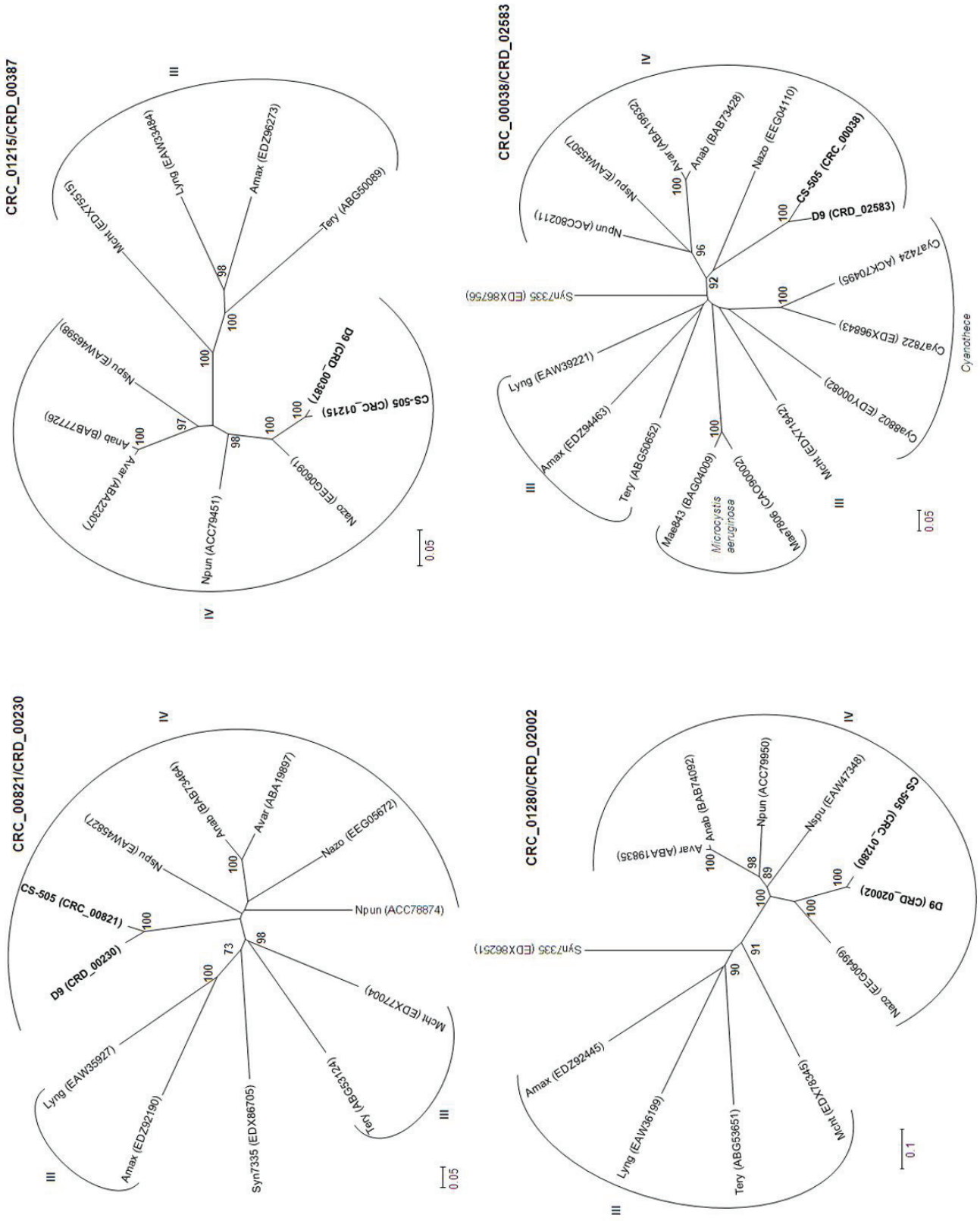
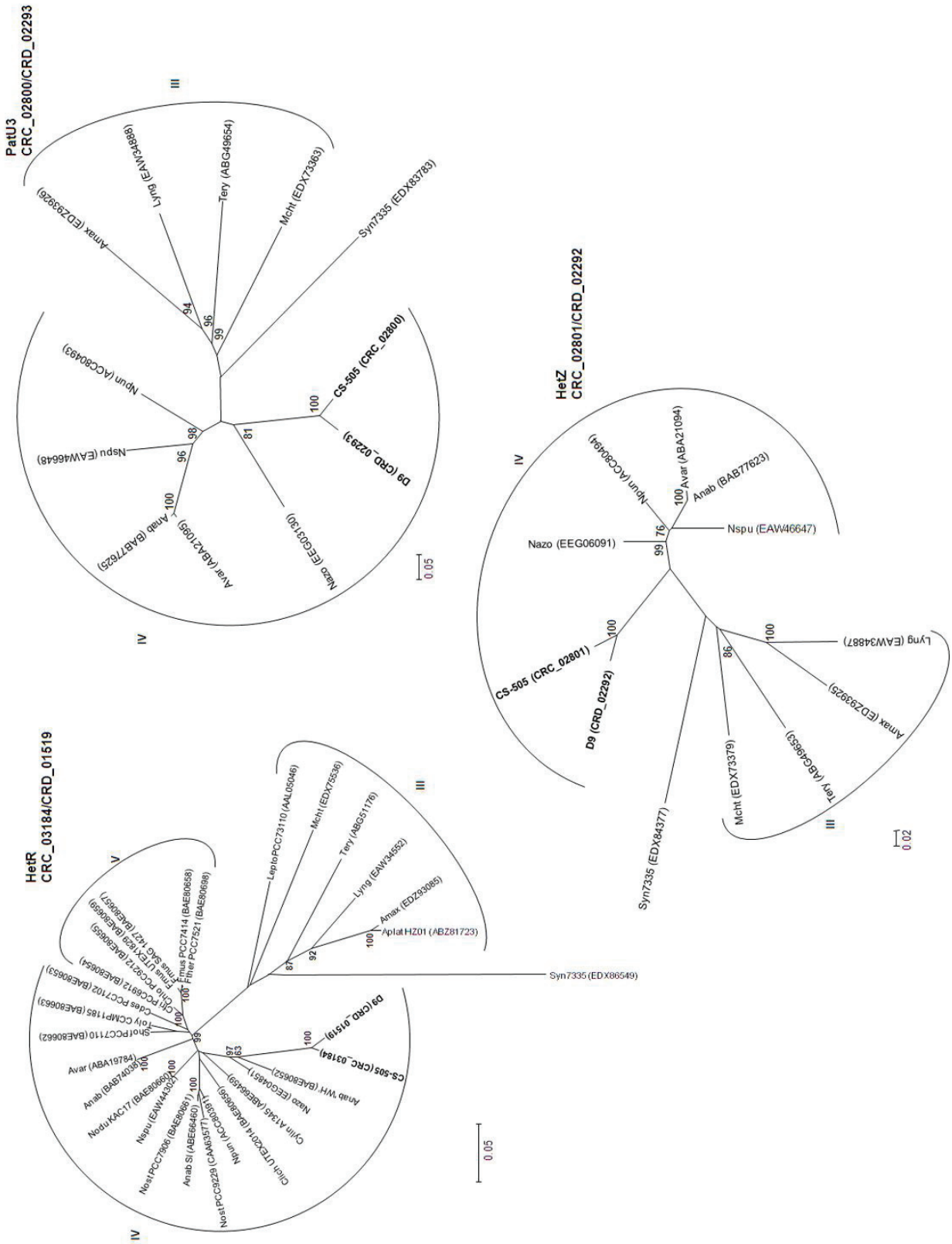


Figure 2.2.S5 Structure and comparison of the toxin gene clusters in CS-505 and D9 with those previously described. **(A)** Comparison of the CYN gene cluster of strain CS-505 with the *cyr* gene cluster described in *C. raciborskii* AWT205 (Mihali *et al.,* 2008); **(B)** Comparison of the STX gene cluster of strain D9 with the *sxt* gene cluster described in *C. raciborskii* T3 (Kellmann *et al.,* 2008a). Identical ORFs between D9/T3 and CS-505/AWT205 are depicted in white; genes involved in the biosynthesis of STX are highlighted with horizontal gray lines and shading. The ORFs unique to D9 and CS-505, with respect to T3 and AWT205, are indicated in black. Unique ORFs in T3 are represented by black horizontal stripes. ORFs outside the clusters are represented by marginal dashed lines and gray fill.


CRC_01215/CRD_00387
III
Mch (EDX75515)
Lyng (EAW34844)
Amax (EDZ96273)
Tery (ABG50089)
D9 (CRD_00387)
CS-505 (CRC_01215)
Nazo (EEG06091)
Npun (ACC79451)
Avar (ABA22307)
Anab (BAB77726)
Nspu (EAW46598)
IV
CRC_00038/CRD_02583
IV
Avar (ABA19852)
Anab (BAB73428)
Nazo (EEG04110)
CS-505 (CRC_00038)
D9 (CRD_02583)
Npun (ACC80211)
Nspu (EAW45507)
Cya7424 (ACK70495)
Syn7335 (EDX86756)
Cya7822 (EDX96843)
Cya8802 (EDY00082)
Lyng (EAW39221)
Mch (EDX71842)
Amax (EDZ94463)
Tery (ABG50652)
Mae843 (BAG04009)
Mae7806 (CAO90002)
Microcystis aeruginosa
III
Cyanothece
CRC_00821/CRD_00230
IV
Avar (ABA18997)
Anab (BAB73464)
Nazo (EEG05672)
Nspu (EAW45827)
CS-505 (CRC_00821)
Npun (ACC78874)
D9 (CRD_00230)
Mch (EDX77004)
Lyng (EAW35927)
Tery (ABG53124)
Amax (EDZ92190)
Syn7335 (EDX86705)
III
CRC_01280/CRD_02002
IV
Anab (BAB74092)
Npun (ACC79950)
Nspu (EAW47348)
Avar (ABA19835)
CS-505 (CRC_01280)
D9 (CRD_02002)
Nazo (EEG06489)
Syn7335 (EDX86251)
Amax (EDZ92445)
Lyng (EAW36199)
Tery (ABG53651)
Mch (EDX78345)
III


PatU3
CRC_02800/CRD_02293
III
Anax (EDZ39926)
Lyng (EAW34888)
Tery (ABG49654)
Mchl (EDX73363)
Syn7335 (EDX83783)
CS-505 (CRC_02800)
D9 (CRD_02293)
Npun (ACC80493)
Nspu (EAW46648)
Anab (BAB77623)
Avar (ABA21095)
Nazo (EEG03130)
IV
0.05
HetZ
CRC_02801/CRD_02292
IV
Npun (ACC80494)
Avar (ABA21094)
Anab (BAB77623)
Nspu (EAW46647)
Nazo (EEG06091)
CS-505 (CRC_02801)
D9 (CRD_02292)
Lyng (EAW34887)
Anax (EDZ39925)
Tery (ABG49653)
Mchl (EDX73379)
Syn7335 (EDX84377)
III
0.02
HetR
CRC_03184/CRD_01519
V
Fnus PCC7414 (BAE80658)
Fthe PCC7521 (BAE80698)
Cth PCC8972 (BAE80654)
Fnus UTEX1829 (BAE80655)
Fnus SAG 14.27 (BAE80657)
Cdes PCC7102 (BAE80653)
Holy CCMP1185 (BAE80663)
Shot PCC7110 (BAE80662)
III
LeptoPCC73110 (AAL06046)
Mchl (EDX75536)
Tery (ABG51176)
Lyng (EAW34552)
Anax (EDZ93085)
Aplat HZ01 (ABZ81723)
Syn7335 (EDX86549)
IV
CS-505 (CRC_03184)
D9 (CRD_01519)
Anab WH (BAE80652)
Nazo (EEG04851)
Cylin A1345 (ABE66459)
Clich UTEX2014 (BAE80656)
Npun (ACC09091)
Nost PCC9229 (CAA63577)
Anab SI (ABE66460)
Nost PCC7906 (BAE80661)
Nspu (EAW44302)
Nodu KAC17 (BAE80660)
Anab (BAB74035)
Avar (ABA19784)
0.05

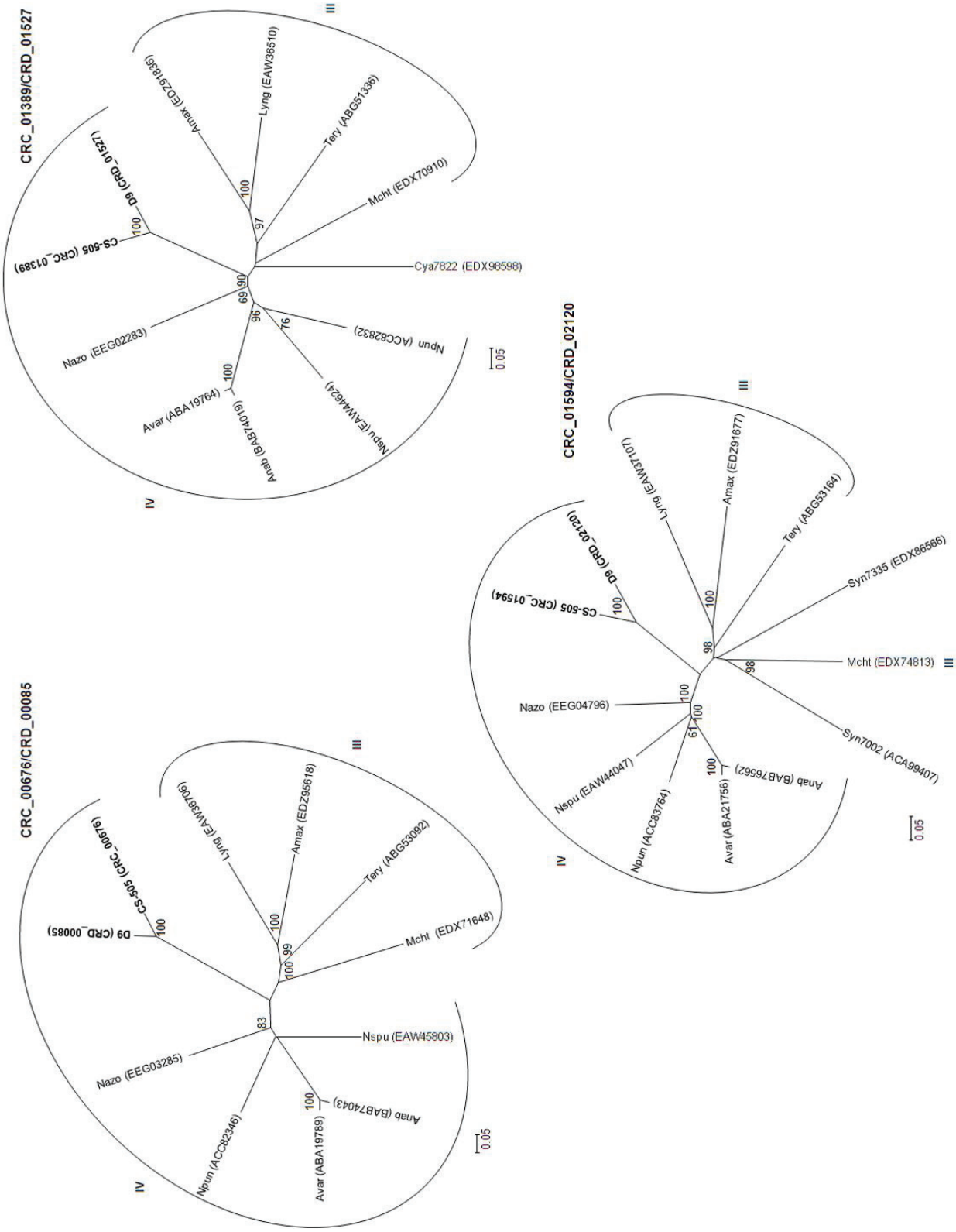

CRC_01389/CRD_01527
III
Lyng (EAW36510)
Tery (ABG51336)
Mcht (EDX70910)
Cya7822 (EDX98598)
Npun (ACC82832)
Nspu (EAM44624)
Anab (BAB74019)
Avar (ABA19764)
Nazo (EEG02283)
CS-505 (CRC_01389)
D9 (CRD_01527)
Amax (EDZ91836)
IV
100
97
90
69
96
76
100
100
0.05
CRC_01594/CRD_02120
III
Lyng (EAM37107)
Amax (EDZ91677)
Tery (ABG53164)
Syn7335 (EDX86566)
Mcht (EDX74813)
Syn7002 (ACA99407)
Anab (BAB76562)
Avar (ABA21756)
Npun (ACC83764)
Nspu (EAW44047)
Nazo (EEG04796)
CS-505 (CRC_01594)
D9 (CRD_02120)
IV
100
100
98
98
100
61
100
0.05
CRC_00676/CRD_00085
III
Lyng (EAM36706)
Amax (EDZ95618)
Tery (ABG53092)
Mcht (EDX71648)
Nspu (EAW45803)
Anab (BAB74043)
Avar (ABA19789)
Npun (ACC82346)
Nazo (EEG03285)
CS-505 (CRC_00676)
D9 (CRD_00085)
IV
100
100
99
100
83
100
0.05

Figure 2.2.S6 Phylogenetic relationships of the 10 CDS found as core in 9 filamentous cyanobacteria. Affiliations to the cyanobacterial subsections are shown in brackets. The trees were constructed with clustalX, using the Neighbor-Joining algorithm with bootstrap of 1000; only bootstrap values higher than 60% are shown over the nodes. When available, unicellular strains were used as outgroup taxa. Trees are organized according to the appearance of each CDS pair in Table 2.2.4. GenBank accession numbers are indicated after species designation (names in bold-face correspond to sequences belonging to CS-505 and D9). Species name abbreviations were used as in materials and methods with the exception of the new sequences used in phylogenetic analyses: Anab WH: *Anabaena* sp. WH School st. isolate; Cylin A1345: *Cylindrospermum* sp. A1345; Clich UTEX2014: *Cylindrospermum licheniforme* UTEX 2014: Nost PCC9229: *Nostoc* sp. PCC 9229; Anab SI: *Anabaena* sp. South India 2006; Nost PCC7906: *Nostoc* sp. PCC 7906; Nodu KAC17: *Nodularia* sp. KAC 17; Shof PCC7110: *Scytonema hofmanni* PCC 7110; Toly CCMP1185: *Tolypothrix* sp. CCMP1185; Cdes PCC7102: *Calothrix desertica* PCC 7102; Cfri PCC6912: *Chlorogloeopsis fritschii* PCC 6912; Chlo PCC9212: *Chlorogloeopsis* sp. PCC 9212; Fmus UTEX1829: *Fischerella muscicola* UTEX 1829; Fmus SAG 1427: *Fischerella muscicola* SAG 1427-1; Fmus PCC7414: *Fischerella muscicola* PCC 7414; Fther PCC7521: *Fischerella thermalis* PCC 7521; LeptoPCC73110: *Leptolyngbya* sp. PCC 73110; Aplat HZ01: *Arthrospira platensis* HZ01; Mae843: *Microcystis aeruginosa* NIES-843; Mae7806: *Microcystis aeruginosa* PCC 7806; Cya7822: *Cyanothece* sp. PCC 7822; Syn7002: *Synechococcus* sp. PCC 7002; Syn7335: *Synechococcus* sp. PCC 7335.

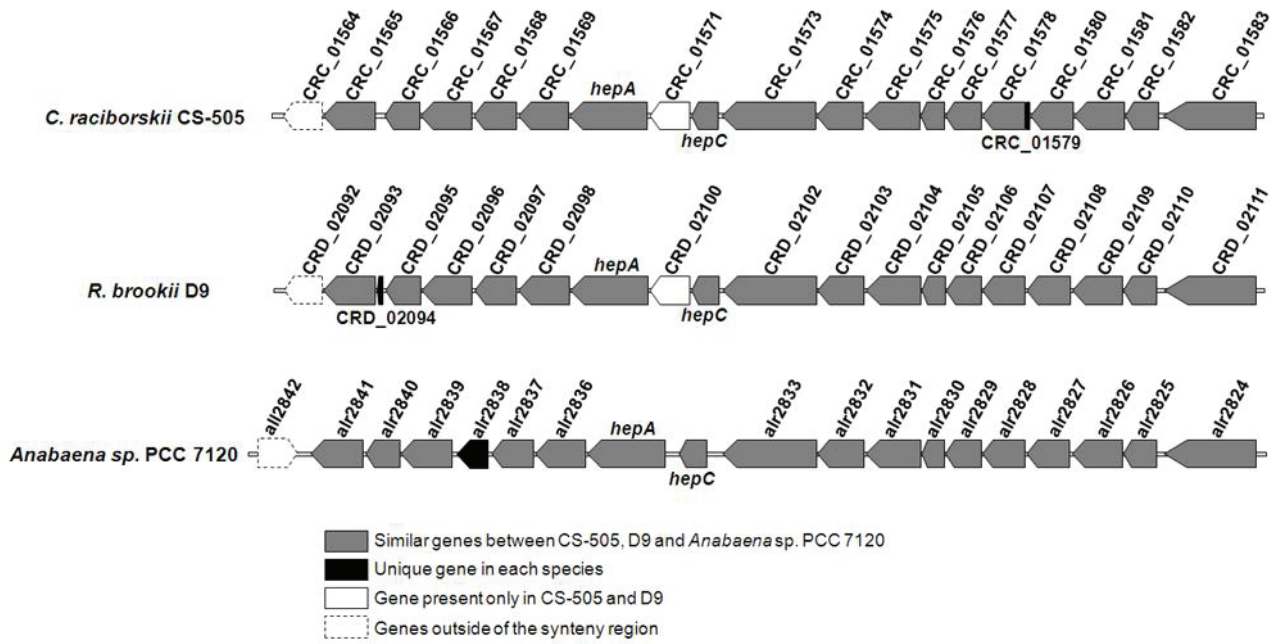


Figure 2.2.S7 Comparison of the gene clusters for heterocyst polysaccharide biosynthesis. The comparison was based in the gene cluster described for *Anabaena* sp. PCC 7120 (Huang *et al.*, 2005).

Table 2.2.S1 List of the unique CDS of CS-505 and D9 and their classification into the different COG categories, Table 2.2.S1 is available at the PLoS ONE website http://www.plosone.org

Table 2.2.S2 Cell division genes in cyanobacteria

Name	CS-505 Locus tag	% identity AA	D9 Locus tag	% identity AA	Product / Description
ftsW°	CRC_02478	77.14	CRD_02755	76.88	Cell cycle protein
ftsZ°	CRC_03281	87.56	CRD_01589	86.42	Cell division protein FtsZ
*ftn2**	CRC_01266	59.25	CRD_01988	56.73	Heat shock protein DnaJ-like
*cikA**	CRC_01431	67.40	CRD_00661	69.02	Multi-sensor Signal Transduction Histidine Kinase
ftsI°	CRC_03074	64.90	CRD_01146	64.13	Penicillin-binding protein, transpeptidase
*cdv1**	CRC_02542	69.70	CRD_01024	69.70	Peptidyl-prolyl cis-trans isomerase, cyclophilin type
*cdv2(ylmF)**	CRC_03484	81.35	CRD_01858	82.38	Protein of unknown function DUF552
ylmE+	CRC_03485	70.27	CRD_01859	70.72	Protein of unknown function UPF0001
ylmG+	CRC_00226	70.65	CRD_01524	70.65	Protein of unknown function YGGT
ylmH+	CRC_00425	85.71	CRD_02533	85.71	RNA-binding S4
minC°	CRC_03004	60.70	CRD_00104	60.40	Septum formation inhibitor MinC
minE°	CRC_03006	72.53	CRD_00106	70.10	Septum formation topological specificity factor MinE
minD°	CRC_03005	84.53	CRD_00105	86.04	Septum site-determining protein MinD
ftsQ°	CRC_03280	54.24	CRD_01590	54.24	cell division protein FtsQ
*cdv3(divIVA)**	CRC_00206	72.55	CRD_02623	72.22	conserved hypothetical protein
yfhF$^{2°}$	CRC_01074	59.44	CRD_02634	59.18	conserved hypothetical protein
*ftn6**	CRC_02893	46.46	CRD_02832	45.10	conserved hypothetical protein

Similarity percentages are based on comparison with the orthologs in *Anabaena* sp. PCC 7120
* Identified cell division genes in *Synechochoccus elongatus* PCC 7942
+ Candidate cell division genes in *S. elongatus* PCC 7942
° Genes described in *E. coli* (1) and *B. subtillis* (2) responsible for cell division
See Miyagishima *et al.,* (2005) and listed references for a detailed explanation on the functional characterization of cell division genes in cyanobacteria and eubacteria.

Table 2.2.S3 Core genes of heterocystous cyanobacteria, Table 2.2.S3 is available at the PLoS ONE website http://www.plosone.org

Table 2.2.S4 83 previously described regulatory genes present in the genomes of the terminal heterocyst forming cyanobacteria *C. raciborskii* strain CS-505 and the non-heterocystous *R. brookii* D9.

Anabaena **PCC 7120**	**Gene**	***C. raciborskii*** **CS-505**	**Identity (%)**	***R. brookii*** **D9**	**Identity (%)**	**Described function/role**
all1940	*abp1*	CRC_02963	76.30	CRD_02427	77.77	Heterocyst glycolipid-layer formation (Koksharova & Wolk, 2002)
all1939	*abp2*	CRC_02964	72.69	CRD_02426	72.89	
alr3608	*abp3*	CRC_00830	56.18	-	-	
alr4240	*abp4 (hlyD)*	CRC_02581	65.90	CRD_01244	64.24	
alr1010	*ccbP*	-	-	-	-	Heterocyst differentiation (Zhao *et al.,* 2005)
all0601	*cnaT*	CRC_01748	65.60	CRD_01497	64.45	*Nir* operon transcriptional regulator (Frias *et al.,* 2003)
all0187	*conR*	CRC_00948	62.62	CRD_02925	62.46	Heterocyst functionality (Fan *et al.,* 2006)
alr3712	*devA*	CRC_00436	80.70	-	-	Heterocyst polysaccharide formation (Cai & Wolk, 1997, Fiedler *et al.*, 1998, Ramirez *et al.*, 2005, Campbell *et al.*, 1996)
alr3710	*devB*	CRC_00438	65.32	-	-	
alr3711	*devC*	CRC_00437	77.74	-	-	
alr3952	*devH*	CRC_03275	98.32	CRD_01596	98.32	
alr0442	*devRa*	CRC_02562	83.72	CRD_01268	79.84	
all1430	*fdxH*†	CRC_02876	70.40	-	-	Heterocyst ferredoxin (Schrautemeier & Böhme, 1985)
alr3858	*ftsZ*	CRC_03281	86.34	CRD_01589	86.41	Septum formation during cell division (Sakr *et al.,* 2006)
all1691	*furA*	CRC_01433	88.07	CRD_00658	88.74	Iron metabolism regulation (Lopez-Gomollon *et al.,* 2007)
all2319	*glnB*	CRC_02206	92.85	CRD_00313	93.75	N_2 metabolism regulation (Laurent *et al.,* 2004)
asr3935	*hanA*	CRC_01160	92.47	CRD_00595	90.32	Heterocyst differentiation (Khudyakov & Wolk, 1996)
alr0093	*hcwA*	-	-	-	-	Heterocyst polysaccharide and envelope formation (Fan *et al.,* 2006)
alr1086	*henR*	CRC_00384	69.04	CRD_03033	68.51	
alr2835	*hepA*†	CRC_01570	57.19	CRD_02099	56.09	
alr3698	*hepB*	CRC_03468	67.94	CRD_01842	68.46	
alr2834	*hepC*	CRC_01572	41.50	CRD_02101	43.19	
all4496	*hepK*	-	-	-	-	
alr0117	*hepN*	CRC_00406	48.58	CRD_01826	47.33	
all2760	*hepS*	-	-	-	-	
alr2817	*hetC*	-	-	-	-	Required for pro-heterocyst cell division cessation(Xu & Wolk, 2001)

alr3546	*hetF*	CRC_02207	52.46	CRD_00315	58.04	Heterocyst differentiation (Wong & Meeks, 2001, Liu & Golden, 2002)
all3740	*hetL*	-	-	-	-	
alr5357	*hetM*† *(hglB)*	CRC_02052	65.60	-	-	Heterocyst glycolipid-layer formation (Maldener *et al.,* 2003)
alr5358	*hetN*	-	-	-	-	Late stage heterocyst pattern maintenance (Callahan & Buikema, 2001)
alr2818	*hetP*	CRC_01587	58.11	-	-	Heterocyst differentiation and maturation(Fernandez-Pinas *et al.*, 1994, Buikema & Haselkorn, 1991, Yoon *et al.*, 2003, Zhang *et al.,* 2007)
alr3234	*hetP-like*	CRC_00980	69.33	CRD_00495	69.33	
alr3234	*hetP-like*	CRC_02972	41.26	-	-	
alr2902	*hetP-like*	CRC_01674	61.19	CRD_02196	62.68	
alr2339	*hetR*†	CRC_03184	83.94	CRD_01519	82.94	
alr2300	*hetY*	CRC_02119	65.94	CRD_02372	64.74	
alr0099	*hetZ*	CRC_02801	79.07	CRD_02292	79.69	
all5359	*hetI*	CRC_01556	51.96	CRD_01379	53.27	Heterocyst glycolipid-layer formation(Fan *et al.,* 2005, Black *et al.*, 1995, Bauer *et al.*, 1997, Campbell *et al.*, 1997)
all5345	*hgdA*	CRC_02043	69.23	-	-	
all5347	*hgdB*	CRC_02045	63.93	CRD_00126	48	
all5346	*hgdC*	CRC_02044	82.39	CRD_00127	61.47	
alr5356	*hglA*	CRC_02051	81.40	CRD_00121	75.62	
alr5355	*hglC*	CRC_02050	61.19	-	-	
alr5354	*hglD*	CRC_02049	71.49	-	-	
alr5351	$hglE_A$	CRC_02047	65.96	-	-	
all1646	$hglE_2$	-	-	-	-	
alr5352	*hgl*F	-	-	-	-	
alr5353	*hgl*G	CRC_02048	64	-	-	
all0813	*hgl*K	CRC_00692	54.93	CRD_01636	54.24	
asr4517	*nbl*A	CRC_02186	82.25	CRD_00293	80.64	Phycobiliprotein degradation (Baier *et al.,* 2004)
all4312	*nrr*A	CRC_00297	86.43	CRD_01945	85.27	Transcriptional regulator of *hetR* (Ehira & Ohmori, 2006)
alr4392	*ntcA*†	CRC_00858	96.80	CRD_00550	91.94	N_2 metabolism and heterocyst differentiation regulators (Wei *et al.*, 1994, Frias *et al.*, 2000)
all0602	*ntcB*	CRC_01749	77.59	CRD_01496	77.27	
all0521	*patA*†	CRC_01956	40.15	-	-	Heterocyst differentiation and initial pattern formation (Liang *et al.,* 1992, Jones *et al.,* 2003, Meeks & Elhai, 2002, Yoon & Golden, 1998, Zhang *et al.,* 2007)
all2512	*patB*	CRC_01763	73.48	-	-	
alr4812	*patN*	CRC_03501	60.64	CRD_01874	61.11	
asl2301	*patS*	CRC_02157	29*	CRD_02133	23*	

asr0101	*patU3*	CRC_02800	51.76	CRD_02293	56.20	
alr2502	*pkn22*	-	-	-	-	NtcA mediated iron metabolism regulation (Cheng *et al.*, 2006, Xu *et al.*, 2003)
alr0709	*pkn41*	-	-	-	-	
alr0710	*pkn42*	-	-	-	-	
alr4366	*pknA*	-	-	-	-	Heterocyst differentiation (Zhang, 1993)
alr4368	*pknD*	CRC_02958	50.49	CRD_00759	47	N_2 regulation (Zhang & Libs, 1998)
alr3732	*pknE*	-	-	-	-	Heterocyst maturation (Zhang & Libs, 1998, Jang *et al.*, 2007)
alr3731	*prpA*	-	-	-	-	
all1731	*prpJ*	CRC_01339	55.29	CRD_02900	53.52	
alr4516	*prpS*	CRC_02187	68.46	CRD_00294	61.95	Modification of PII (GlnB) (Laurent *et al.,* 2004)
all1692	*sigC*	CRC_01434	79.21	CRD_00657	78.60	Transcriptional Regulators (Aldea *et al.,* 2007)
alr3800	*sigE*	CRC_02165	83.48	CRD_00081	84.40	
alr3280	*sigG*	CRC_00607	87.09	CRD_02710	86.23	
asr1734	NA	CRC_01342	87.09	CRD_02898	84.94	Heterocyst development inhibitor (Wu *et al.,* 2007)
all5343	NA	CRC_02041	55.84	-	-	Heterocyst glycolipid-layer formation (Fan *et al.,* 2005)
alr5348	NA	-	-	-	-	
asr5349	NA	-	-	-	-	
alr2825	NA	CRC_01582	89.10	CRD_02110	89.10	Heterocyst polysaccharide envelope formation (Huang *et al.,* 2005)
alr2827	NA	CRC_01580	87.64	CRD_02108	86.09	
alr2831	NA	CRC_01575	79.44	CRD_02104	79.21	
alr2833	NA	CRC_01573	70.88	CRD_02102	70.08	
alr2837	NA	CRC_01568	54.76	CRD_02097	54.76	
alr2839	NA	CRC_01567	62.46	CRD_02096	62.72	
alr2841	NA	CRC_01565	49.77	CRD_02093	49.51	

* To assign the corresponding CDS as a *patS* homologue other features than amino acid sequence identity were considered (see text). NA: not assigned.

† Expression confirmed by qPCR.

2.3 Publication III

Reassessment of the toxin profile of *Cylindrospermopsis raciborskii* T3 and function of putative sulfotransferases in synthesis of sulfated and sulfonated PSP toxins

Katia Soto-Liebe, Alejandro A. Murillo, Bernd Krock, Karina Stucken, Juan J. Fuentes-Valdés, Nicole Trefault, Allan Cembella, Mónica Vásquez.

2.3.1 Abstract

Cylindrospermopsis raciborskii T3, a toxigenic freshwater cyanobacterium, has been used as a model to study and elucidate the biosynthetic pathway of tetrahydropurine neurotoxins associated with paralytic shellfish poisoning (PSP). There are nevertheless several inconsistencies and contradictions in the published toxin profiles of this strain. An initial report shortly after isolation and culture claimed that this strain produced the carbamoyl analogues saxitoxin (STX) and gonyautoxins 2 and 3 (GTX2/3), whereas a subsequent study by pre-column derivatization and HPLC analysis with fluorescence detection gave the profile of strain T3 strain as STX and the N-sulfocarbamoyl derivatives C1/2. When *C. raciborskii* T3 was utilized to identify the sequence of genes putatively involved in PSP toxin biosynthesis (*sxt* gene cluster), the toxin profile was considered to comprise STX, neoSTX (NEO), decarbamoyl STX (dcSTX) and B1. Our analysis of the complete genome of another PSP toxin-producing cyanobacterium, *Raphidiopsis brookii* D9, which is closely related to *C. raciborskii* T3, as well as the toxin profile of PSP toxins in both strains by liquid chromatography coupled to mass spectrometry (LC-MS/MS) has resolved many issues regarding the correlation between biosynthetic pathways, putative genes and the T3 toxin profile. The putative *sxt* gene cluster in *R. brookii* D9 has a high synteny with the T3 *sxt* cluster with 100% nucleotide identity among the shared genes. However, the published toxin profiles for T3 do not correlate with the toxin profile of D9 and the putative functions assigned to each open reading frame (ORF). Our reassessment of the PSP toxin profile of T3 by LC-MS/MS confirmed production of only STX, NEO and dcNEO production. Significant insights were gained via correlation between specific *sxt* genes and their role in PSP-toxin production in both D9 and T3 strains. In particular, analysis of sulfotransferase functions

for SxtN (N-sulfotransferase) and SxtSUL (O-sulfotransferase) enzymes yielded a revised view of the genetic mechanisms underlying the synthesis of sulfated and sulfonated analogues of STX in toxigenic cyanobacteria.

2.3.2 Introduction

Cylindrospermopsis raciborskii was described as a planktonic cosmopolitan filamentous cyanobacterium (Order Nostocales) from freshwater, able to fix nitrogen in terminal heterocysts (Woloszynska, 1912). *C. raciborskii* has become one of the most notorious blue-green algal species (Padisák, 1997) because of its toxicity and tendency to form dense blooms that interfere in multiple ways with water use. The species comprises strains that can produce either the hepatotoxin cylindrospermopsin (CYN), a potent protein synthesis inhibitor, or the neurotoxins saxitoxin and its analogues. The latter toxins are associated with paralytic shellfish poisoning (PSP), which can cause illness and even death of humans after consumption of seafood contaminated with these toxins, and are also responsible for deterious effects on organisms in aquatic ecosystems (Zingone & Enevoldsen, 2000, Ressom *et al.*, 1994). Nevertheless, some strains of *C. raciborskii* do not produce any known toxins.

The first report of PSP toxin production in *C. raciborskii* (Lagos *et al*. 1999), characterized the toxin profile of three strains T1, T2 and T3. Strains T2 and T3 were isolated in 1996 from the same location in Brazil, a branch of Billings water reservoir called Taquacetuba.The toxin profiles of cultured isolates were determined by liquid chromatography with fluorescence detection (LC-FD) after post-column derivatization. Confirmation of toxin identity was provided by liquid chromatography with detection by electro-spray ionization mass spectrometry (LC-ESI-MS). Lagos *et al*. (1999) mentioned that the strains T2 and T3 had identical toxin profiles, comprising saxitoxin (STX), and the epimers gonyautoxin 2 and gonyautoxin 3 (GTX2/3) in a 1:9 ratio, but showed only the LC-chromatograms of strains T1 and T2. Curiously, despite the lack of definitive chromatographic evidence of the toxin profile of T3, this strain was subsequently selected by different researchers to investigate the biosynthetic pathway for PSP toxins. The first chromatogram for *C. raciborskii* strain T3 was published by Pomati *et al.* (2003), based upon a pre-column oxidation method followed by liquid chromatographic separation of the oxidation products (Lawrence *et al.*, 1996). Furthermore, although Pomati *et al*. (2003) used a mixture of STX and GTX2/3 as a calibration standard and claimed to find only STX in T3, they noted that the strain also produces the low potency N-

sulfocarbamoyl derivatives C1/C2, referencing the work of Lagos *et al*. (1999). This was a misquotation of the T3 toxin profile as described by Lagos *et al.* (1999), who reported only the presence of STX and GTX2/3. This confusion was later compounded by Pomati *et al.* (2004), who considered the "STX and C1/2" profile of T3 as the basis for experimentation and hypothesis of a possible new pathway for synthesis of C1/2 analogues, thereby modifying the biosynthetic pathways proposed by Shimizu (1996) and Sako *et al.* (2001).

The sequence of the PSP toxin biosynthesis gene cluster, known as the *sxt* gene cluster, in strain T3 was recently published (Kellmann *et al.*, 2008a). Based upon their LC-MS analysis, Kellmann *et al.* (2008a) described yet another toxin profile for T3: STX, neosaxitoxin (NEO), decarbamoyl saxitoxin (dcSTX) and the N-sulfocarbamoyl toxin B1 (GTX5). These authors did not refer to synthesis of C1/2 analogues, but confusingly in the discussion they mentioned that both *Anabaena circinalis* AWQC131C and *C. raciborskii* T3 produce N-sulfonated and O-sulfated analogues of STX (B1, C2/C3, dcGTX3/dcGTX4). Kellmann *et al*. (2008a) analyzed the structure of the gene cluster and the possible function of each open reading frame (ORF) based on the toxin profile they reported.

Obviously the interpretation of existing reports on the toxin profile and molecular genetic evidence for the biosynthetic pathway to STX and analogues in strain T3 has been complicated if not confounded by these discrepancies. The inconsistencies could conceivably have arisen via a combination of cross-contamination, miss-identification, inappropriate application or interpretation of analytical methodologies, and/or errors in citation of the literature. Furthermore, the Brazilian group responsible for the distribution of the original strain T3 did not specify the toxin profile of this strain, and only expressed the sum of all peak areas of toxins as concentration equivalents of STX (Ferrão Filho *et al.*, 2008).

We recently sequenced the genome of *Raphidiopsis brookii* D9 (**Publication II**), a PSP-toxin producing cyanobacterium closely related to and formerly assigned to *C. raciborskii* (**Publication I**) and later to *Raphidiopsis* (Plominsky *et al.*, 2009). The toxin profile for strain D9, based upon detailed LC-MS/MS analysis, comprises STX, dcSTX, GTX2/3 and dcGTX2/3. We identified a gene cluster for PSP toxin synthesis that is highly similar to that published for T3, but in D9 the cluster is not flanked at one end by the proposed regulatory genes *sxtY*, *sxtZ* and *ompR* and lacks four open reading frames (ORFs) related to PSP-toxin production in T3 (*sxtN*, *sxtX*, *sxtW* and *sxtV*) (**Publication II**). Considering the putative differences in toxin profiles

between T3 and D9, these four ORFs are of interest to evaluate and correlate the toxin profile in *C. raciborskii* T3.

The primary aim of this study is clarify the toxin profile of *C. raciborskii* T3, a reference strain in the study of PSP toxin biosynthesis, and to assess the coherence between the PSP toxin profile and the gene content, with reference to other cyanobacterial strains. We provide further evidence on the role of sulfotransferase-like genes in the synthesis of sulfated and sulfonated analogues in cyanobacteria.

2.3.3 Results and Discussion

Morphological and genetic analysis of *C. raciborskii* T3

Morphological analysis of *C. raciborskii* T3 by light microscopy was consistent with previous observations. As distinctive for *C. raciborskii*, filaments showed typical vegetative cells (Figure 2.3.1A) and specialized cells, specifically cone-shaped heterocysts located at the end of the filament (Figure 2.3.1B). Characteristic ovoid-shaped akinetes (Figure 2.3.1C) were also present, sometimes in tandem of four akinetes cells (inset in Figure 2.3.1C).

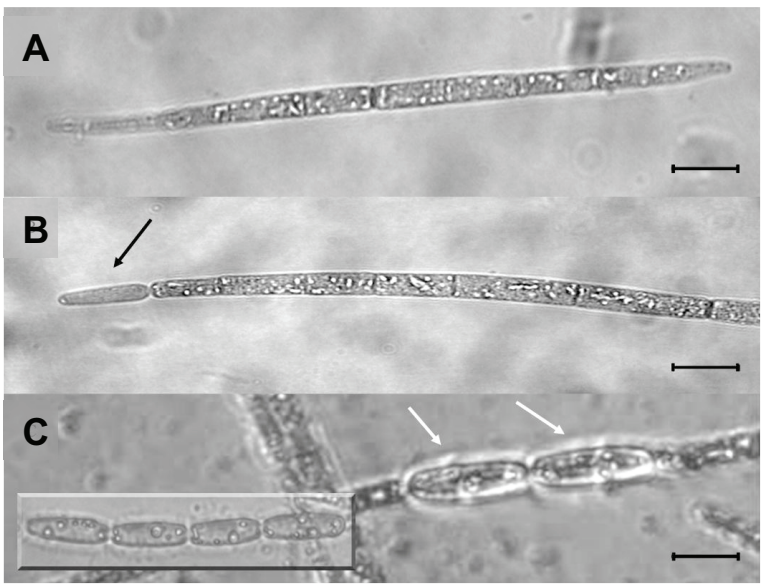


Figure 2.3.1 Filaments of *C. raciborskii* T3. (A) Vegetative cell. Scale bar, 10 μm. **(B)** Solitary trichome with heterocyst (black arrow) in terminal position of the filament. **(C)** Trichome with two adjacent oval-shaped akinetes (white arrows).

Genetic confirmation of the identity of *C. raciborskii* T3 was performed by amplification and sequencing of the 16S rDNA and five genes belonging to the *sxt* gene cluster (*sxtA*, *sxtN*, *sxtX*, *sxtW, sxtV*). In the other four cyanobacterial species from which the cluster is known, the genetic context of the *sxt* gene cluster is different from that of T3 strain allowing for a proper identification of the T3 *sxt* cluster. In this strain, an unknown orf1 is located upstream de *sxtD* gene (*sxtD*-context) and an insertion sequence, from IS4 family, is located upstream of *sxtO* gene (*sxtO*-context) (Figure 2.3.2A). PCR amplicons of the expected size were obtained for all genes studied (16S rDNA: 1367 bp; *sxtA*: 3600; *sxtN*: 1045 bp; *sxtX*: 921 bp; *sxtW*: 430 pb; *sxtV*: 1653 pb; *sxtD*-context: 853 bp; and, *sxtO*-context: 1240 bp) (Figure 2.3.2B).

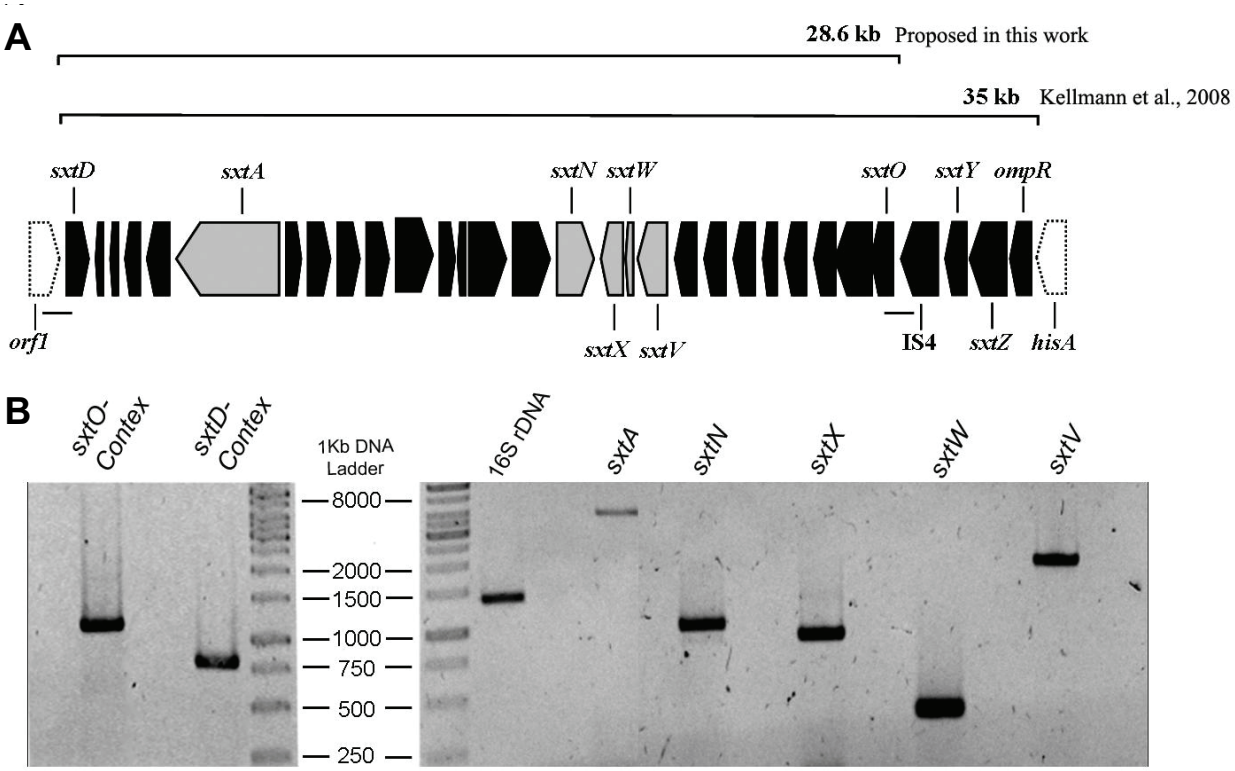


Figure 2.3.2 Schematic organization of the *sxt* gene cluster in *C. raciborskii* T3 (A) and PCR amplicons of *sxt* genes (B). Genes amplified to confirm the cluster structure in T3 are shown in grey. *sxtD* and *sxtO* genes (both at the flanking regions of the cluster) and their context are indicated by a black bar. The scale of the gene cluster lengths is indicated as the number of base pairs. Electrophoresis in 1.0% agarose gel stained with ethidium bromide, showing amplification fragment of 16S rDNA, *sxt* genes from *C. raciborski* T3 and genetic context of the *sxt* gene cluster ends.

Comparison of the sequences of the 16S rDNA of *C. raciborskii* T3 and the corresponding genes present in the putative *sxt* gene cluster with those available in the NCBI database described by Kellmann *et al*. (2008a) yielded very high similarities. Alignment of the 16S rDNA sequences gave identity matches of 99.8%; the difference of 0.2% corresponded to one gap and a mismatch of five nucleotides (Figure 2.3.S1). Further comparison of the 16S rDNA sequence of strain T3 with *C. raciborskii* strains from Australia (CS-505, 506 and 511), showed that the differences among the strains are restricted mostly to specific polymorphic sites (Figure 2.3.S1). We found a nucleotide identity of 100% in all genes of the *sxt* gene cluster tested (Table 2.3.1).

Table 2.3.1 PCR primers used for amplification and sequencing. The length of PCR products and the 5'-binding sites are given in parentheses.

Target gene	%ID with T3	Primer pair	Primer sequence
16S rDNA	100	CYA106Fa (5' -45)	5'-CGGACGGGTGAGTAACGCGTGA-3'
(1400 bp)		1492R (5' -1411)	5'-GGTTACCTTGTTACGACTT-3'
sxtA	100	SXTAF (5' -5146)	5'-CTCCTCTTCGGTATTGGCGG-3'
(3600 bp)		SXTAR (5' -8745)	5'-GCGGTTCCCGTTATTCTTGC-3'
sxtK	100	SXTKF (5' -15630)	5'-ATCACACCAATAGGATTCA-3'
(324 bp)		SXTKR (5' -15934)	5'-CCCTTGAAGTTACTCATAAT-3'
sxtN	100	SXTNF (5' -18931)	5'- CTGCTAGTTTGCGGCTGGTG-3'
(1045 bp)		SXTNR (5' -19975)	5'- CCCCCTCTGAACGGTTACGA-3'
sxtX (921	100	SXTXF (5' -20212)	5'-AAAAAGTGCAAGTTAAGAGG-3'
bp)		SXTXR (5' -21132)	5'-TAAAACACGGGGACTACATC-3'
sxtW	100	SXTWF (5' -21141)	5'-ACTTAGGTAAAGCGGCTTTG-3'
(430 bp)		SXTWR (5' -21512)	5'-GCGCTACACCACCAGTATTA-3'
sxtV	100	SXTVF (5' -21511)	5'-CCTGCTACAACTTTAATACT-3'
(1829 bp)		SXTVR (5' -23338)	5'-AGCTTAGTAAAGAACAACT-3'
sxtD'	100	SXTDcontF (5' -868)	5'-ATACCTCCAGAGCGAGGACA-3'
(853 bp)		SXTDR (5' -1720)	5'-ATGGAAGGGAGAGCGAATTT-3'
sxtO'	100	SXTOF (5' -29976)	5'-TCCTGACTCCGCAAAGAGTT-3'
(1240 bp)		SXTOcontR (5' -31215)	5'-CGGGGTTGGATATGTTTTTG-3'

Re-assessment of the *C. raciborskii* T3 toxin profile

Clarification of the *C. raciborskii* T3 toxin profile is an important consideration for defining correct gene functions because the gene cluster responsible for the biosynthesis of PSP toxins was first described in T3, with putative functions assigned based upon the reported toxin profile (STX, NEO, dcSTX and B1) (Kellmann *et al*., 2008a). Nevertheless, the previously described toxin profile of *C. raciborskii* (Lagos *et al.*, 1999, Kellmann *et al.,* 2008a) is inconsistent among studies, with the following PSP-toxin analogues: STX, GTX2/3, C1/2, B1, NEO, dcNEO and dcSTX, reported in different combinations depending on the publication. Our analysis of the toxin profile of T3 by LC-FD revealed three fluorescent peaks that coeluted with the toxin standards for NEO-dcNEO, dcSTX and STX, respectively (Figure 2.3.3A,B). However, the analyte peak at the approximate retention time of dcSTX of the T3 chromatogram was eluted slightly earlier than the standard dcSTX, which makes it difficult to confirm the presence of dcSTX in this strain. We also observed several peaks with retention times that corresponded with those of standards for various GTXs. These peaks were found to represent toxin artefacts ("false positives") because they continued to appear as fluorescent components, even when post-column oxidation (derivatization) was suspended (Figure 2.3.3C). In contrast, the three peaks corresponding to NEO-dcNEO, dcSTX and STX were missing in chromatograms without post-column oxidation, supporting their legitimacy as putative PSP toxin peaks.

In any case, without further confirmatory analysis (e.g. by LC-MS), reports of PSP toxin profiles with definitive peak identities for cyanobacteria must be viewed rather sceptically. For example, the pre-column oxidation LC method (Lawrence *et al*., 1996) generates multiple fluorescent derivatives for certain toxin analogues and was originally designed to estimate overall PSP toxicity in shellfish rather than to directly infer the exact toxin profile.

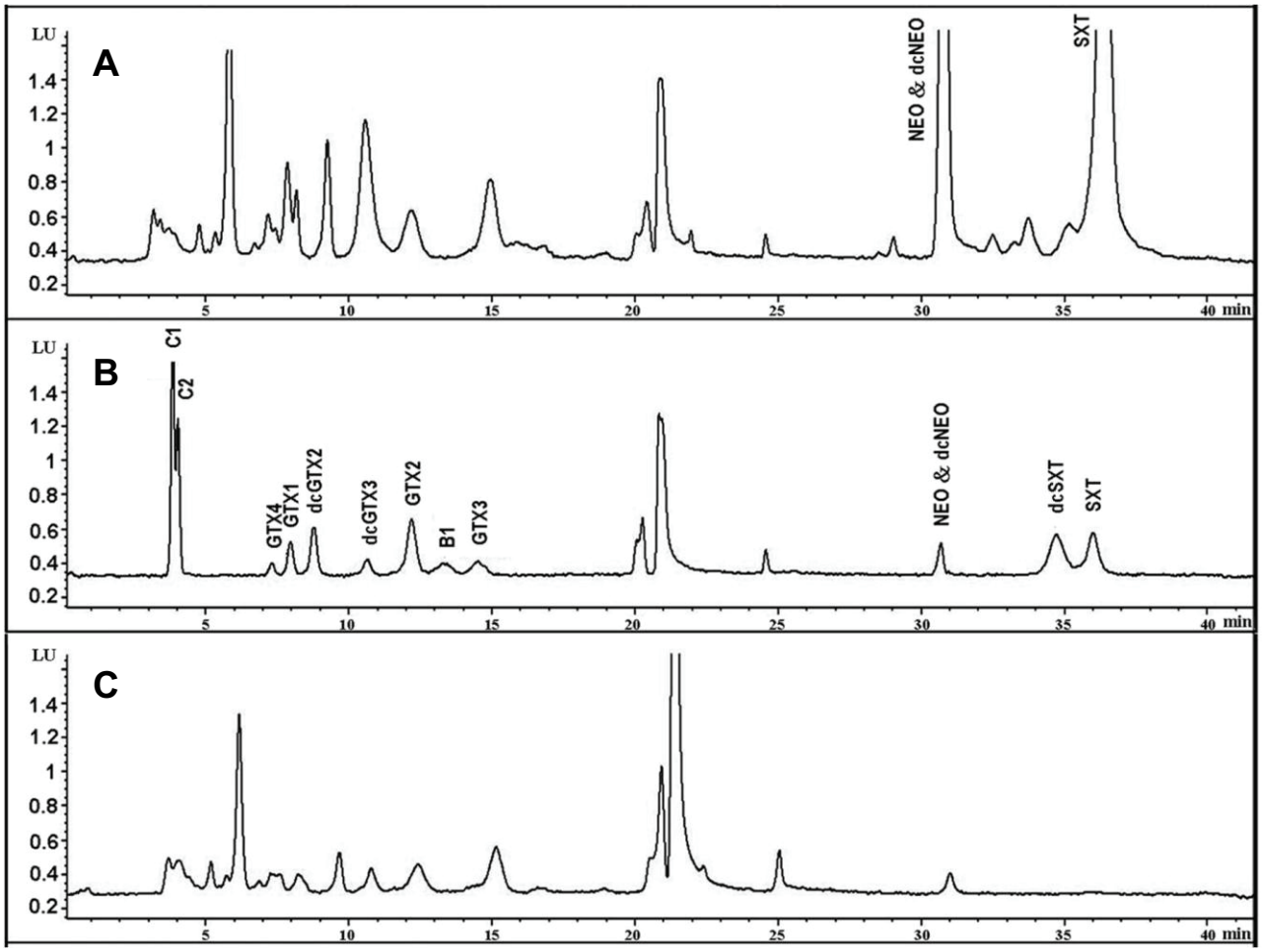


Figure 2.3.3 (A) LC-FD chromatogram of a *C. raciborskii* T3 extract with post-column derivatization. **(B)** LC-FD chromatogram of the PSP toxins (C1/2, GTX1/4, dcGTX2/3, GTX2/3, B1, NEO, dcNEO, dcSTX, STX) standard solution, analyzed with post-column derivatization. **(C)** LC-FD chromatogram of a *C. raciborskii* T3 extract without post-column derivatization.

We confirmed the T3 toxin profile by liquid chromatography with tandem mass spectrometry (LC-MS/MS) (Figure 2.3.4A), consisting of NEO, dcNEO and STX in decreasing order of relative abundance as indicated in Figure 2.3.4B. The LC-MS/MS results were essentially consistent with the LC-FD analysis, but also provided mass spectral resolution of the co-eluting peaks of NEO and dcNEO. There was a faint peak in the T3 LC-MS/MS ion-trace for dcSTX (*m/z* 257>156, data not shown), however, this was below the limit of detection (signal to noise ratio <3). This means that also by LC-MS/MS we cannot unambiguously confirm the presence of dcSTX in T3, but if present, its abundance is lower than 0.3 fmol mg^{-1} fresh weight (FW) of cyanobacteria or less than 0.5 % of the total PSP toxin content. Whereas Kellmann *et al.* (2008a) employed LC-MS to identify key intermediates in the STX biosynthetic pathway, they

only mention that the toxin profile of T3 comprises STX, NEO, dcSTX and B1, but did not provide chromatographic support for this profile. We found no evidence for the presence of the latter two toxin analogues in this strain.

According to our previous LC-MS/MS analysis, with respect to T3, *R. brookii* D9 produces a more complex toxin spectrum, characterised by the presence of STX and O-sulfated toxins GTX2/3, in addition to decarbamoyl analogues, dcSTX, and dcGTX2/3 (Soto-Liebe *et al.*, 2008).

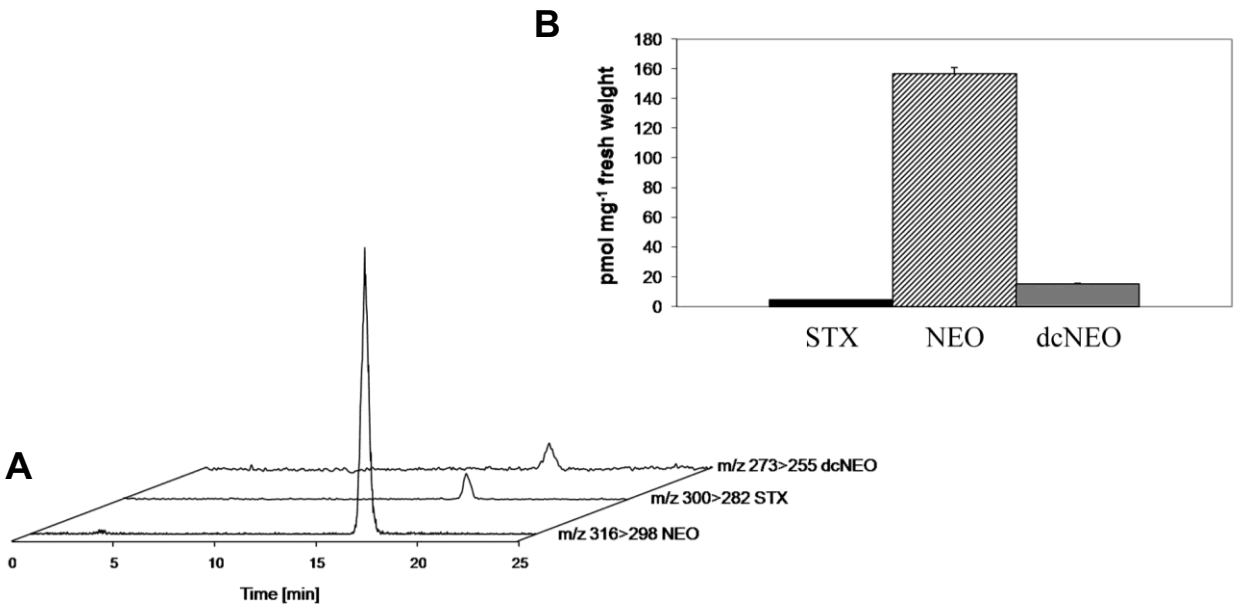


Figure 2.3.4 (A) LC-MS spectra of a *C. raciborskii* T3 extract. **(B)** Toxin content (pmol mg^{-1} of fresh weight) for *C. raciborskii* T3.

Comparative analysis of the *sxt* gene cluster in *C. raciborskii* T3 with other cyanobacteria

Comparison of the gene cluster for biosynthesis of PSP toxins in *C. raciborskii* T3 with the *sxt* gene cluster of other cyanobacteria provides information on gene functions and also reflects the differences in toxin profile. For example, the published *sxt* cluster of *C. raciborskii* T3 comprises 35 Kb and 31 open reading frames (ORFs), whereas the smaller gene cluster of D9 covers only 25.7 Kb, and encodes for 24 ORFs, (Figure 2.3.5). Only 20 ORFs are shared between these clusters, but 19 ORFs share 100% similarity). Four genes in *C. raciborskii* T3: *sxtX*, *sxtN*, *sxtW* and *sxtV* are absent from the *sxt* gene cluster of *R. brookii* D9. The product of *sxtX* in T3 is putatively involved in the synthesis of NEO (Kellmann *et al*., 2008a), and consistent with the lack of this gene in D9, this latter strain does not produce NEO. The *sxtV* and *sxtW* genes were

described as a putative electron transport system to reposition electrons for the correct functioning of oxygenases. The *sxtN* gene was proposed to encode for a sulfotransferase that would produce the sulfated analogues of PSP toxins (Kellmann *et al*., 2008a). Finally, in contrast to what was described for T3, genes which codified as regulatory proteins are not present in the D9 *sxt* gene cluster. The D9 cluster is flanked at the 3' end by ORFs with translation products similar to the bifunctional protein GlmU, allophycocyanin β-subunit (ApcF), and glutamine synthase (GS). Since expression of *apcF* and *glnA* genes is regulated by nitrogen (Herrero *et al.*, 2004, Johnson *et al.*, 1988), STX production may also be regulated by nitrogen availability in D9.

Since the description of the *sxt* gene cluster in *C. raciborskii* T3, corresponding clusters have also been described in three different genera of filamentous cyanobacteria. These clusters differ in size and gene content (Figure 2.3.5). *A. circinalis* AWQC131C encodes a cluster of 30.9 Kb corresponding to 28 *sxt*-like ORFs, three transposase sequences, three disrupted and repeated fragments of *sxtN* and two disrupted fragments corresponding to *sxtW* and *sxtV* (Mihali *et al*., 2009). *Aphanizomenon* sp. NH-5 encodes a 29 Kb cluster comprising 24 *sxt*-like ORFs, where the most remarkable difference is the absence of *sxtO*, putatively involved in PAPS production, a fundamental precursor for the sulfation or sulfonation of STX to O-sulfated or N-sulfonated analogues (Mihali *et al*., 2009). Finally, *Lyngbya wollei* encodes a 36 Kb cluster formed by 26 *sxt*-like ORFs, including within these genes three copies of *sxtM* (MATE family efflux pump) and two copies of *sxtN* (Neilan and Mihali, submitted to the NCBI-database; accession number: EU603711).

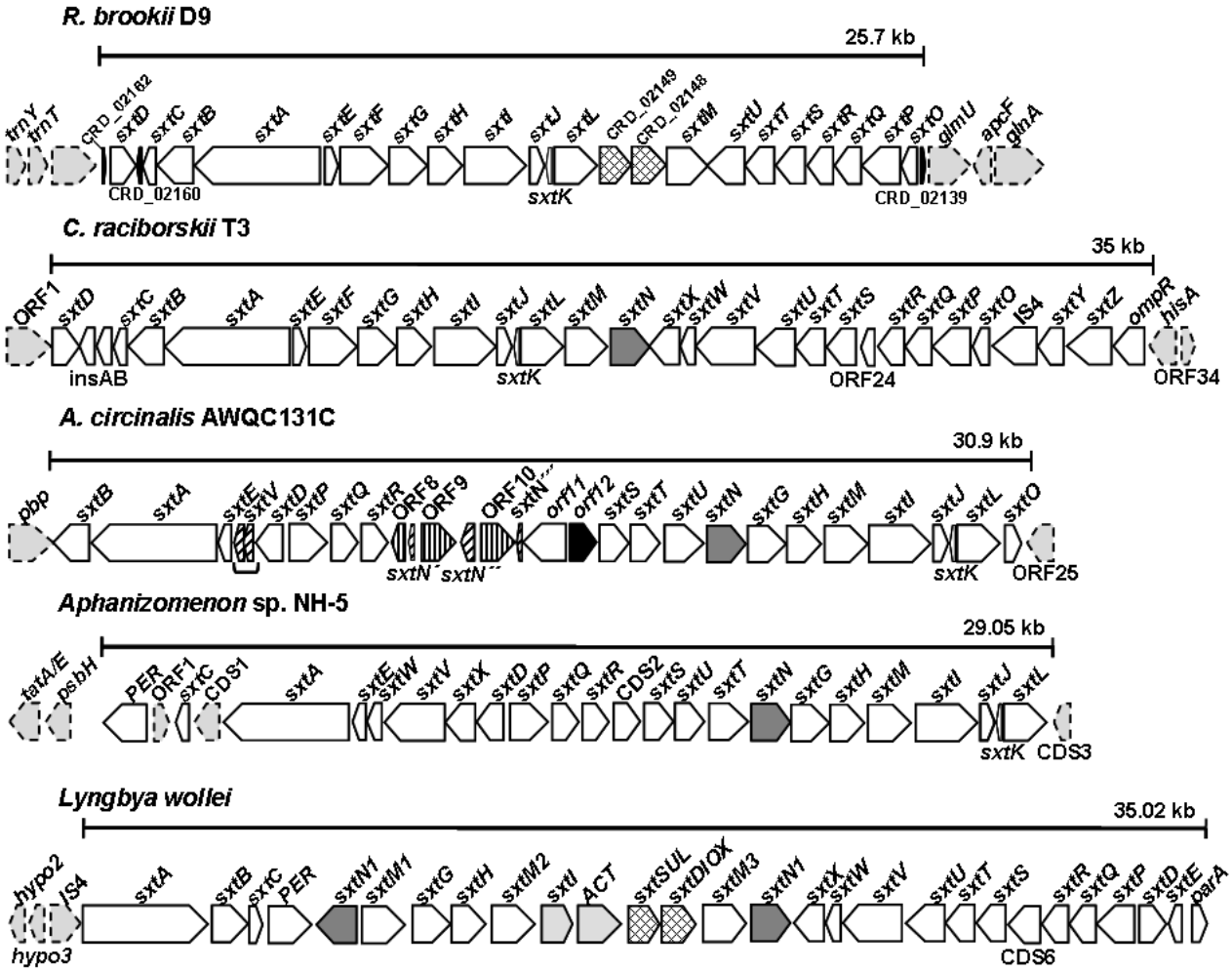


Figure 2.3.5 Comparative analysis of *sxt* gene clusters in five species of filamentous cyanobacteria. The ORFs shared among the five clusters are shown in dark grey, ORFs with unknown function are given in black, the ORFs filled with straight lines codify for transposases and those filled with slanted lines are pseudogenes. *sxtSUL* and *sxtDIOX* are presented with grid lines and ORFs flanking the clusters or without function in the clusters are shown in light grey and dashed lines.

Sulfotransferase-like genes and their role in PSP toxin synthesis

Although details of the complete synthesis of STX remain to be determined, recent studies showed 11,12,12-trideshydroxy-decarbamoyl STX (E') as the last detectable intermediary to STX (Kellmann *et al.,* 2008a). Whereas our data essentially confirm the STX pathway in cyanobacteria as previously described by Kellmann *et al.* (2008a), we propose an extension of the pathway from E' to yield GTX2/3. Only three steps involving double hydroxylation of C12, and the addition of the carbamoyl group to 11, 12, 12-trideshydroxy-decarbamoyl STX, would render STX (Figure 2.3.6A, B).

Inferences based on the structure of sulfated and sulfonated PSP toxins in cyanobacteria indicate that two tailoring reactions, carried on by two different enzymes, are needed to form the different sulfated or sulfonated analogues. With STX as precursor, the first reaction would require an O-sulfotransferase for sulfation of the hydroxyl group at the C-11 position of STX, thereby rendering GTX2/3. The second reaction would involve an N-sulfotransferase that incorporates a sulfonyl group at the N-1 position of the STX carbamoyl group, yielding B1. The joint action of O- and N-sufotransferases would produce C1/C2 toxins (Figure 2.3.6D, E). In the *sxt* cluster, two ORFs have been related with sulfotransferase activity, *sxtN* (Kellmann *et al.,* 2008a) and *sxtSUL,* from *Lyngbya wollei* (submitted by Neilan and Mihali, NCBI-database); neither of them, however, has been experimentally linked to the specific sulfotransferase functions.

The D9 *sxt* gene cluster contains two ORFs (CRD_02149 and CRD_02148) that are absent from the T3 cluster. Our bioinformatics analysis of the gene product of CRD_02149 shows clear similarities to a sulfotransferase, and has an 89% amino acid identity with SxtSUL of *L. wollei*, whereas the gene product of CRD_02148 shows 85% amino acid identity with SxtDIOX of *L. wollei*, corresponding with a hydroxylase. Both SxtSUL and SxtDIOX probably modify STX to form the C11-sulfated derivatives GTX2/3 (Figure 2.3.6C, D).

Despite the lack of experimental or bioinformatic support, Mihali *et al*. (2009) proposed a dual function for SxtN in *A. circinalis* AWQC131C, a producer of STX, GTX2/3 and C1/2. The O-sulfotransferense function was attributed to the synthesis of GTX2/3, whereas the N-sulfotransferase function would yield C1/2.

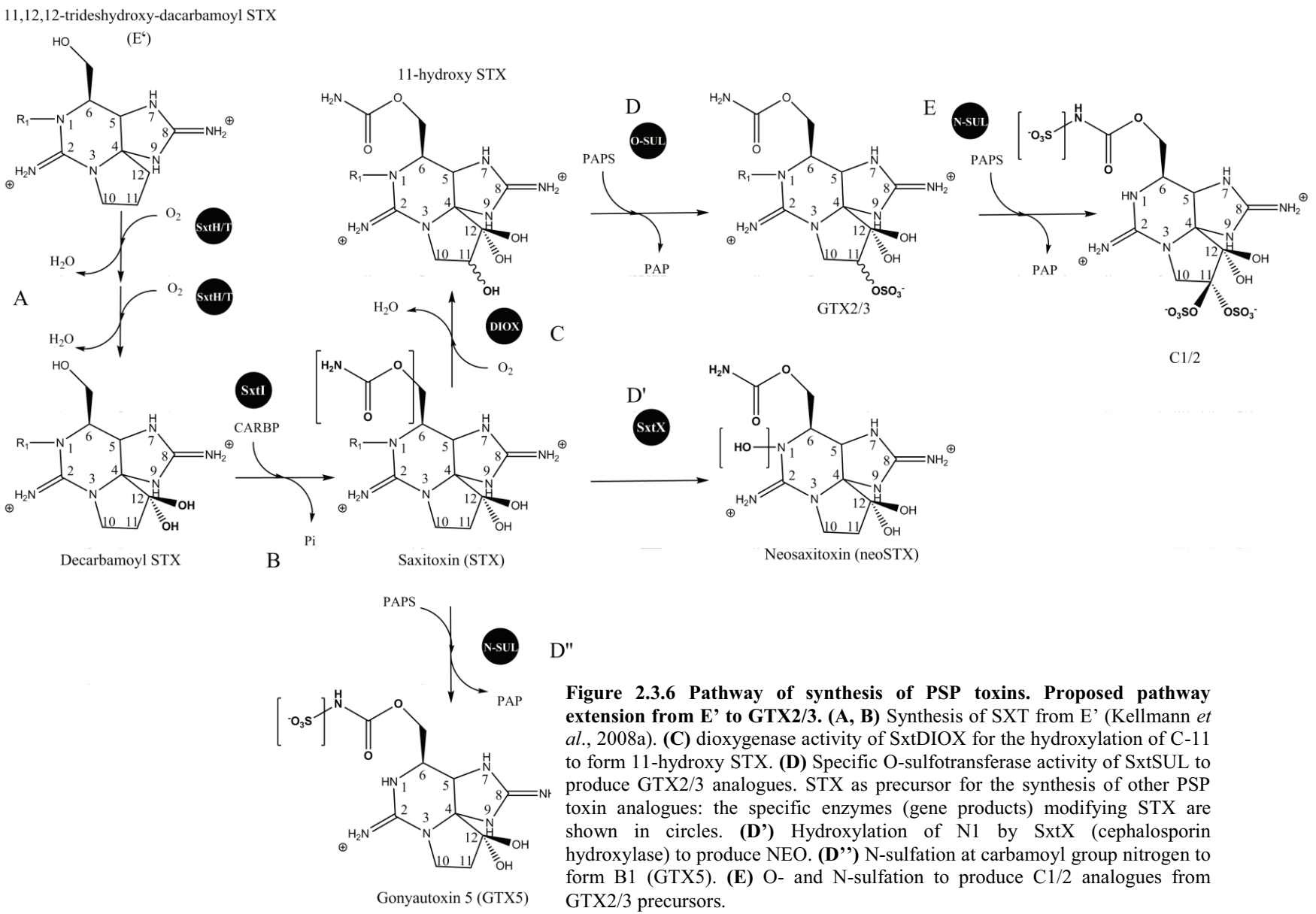


Figure 2.3.6 Pathway of synthesis of PSP toxins. Proposed pathway extension from E' to GTX2/3. (A, B) Synthesis of SXT from E' (Kellmann *et al*., 2008a). **(C)** dioxygenase activity of SxtDIOX for the hydroxylation of C-11 to form 11-hydroxy STX. **(D)** Specific O-sulfotransferase activity of SxtSUL to produce GTX2/3 analogues. STX as precursor for the synthesis of other PSP toxin analogues: the specific enzymes (gene products) modifying STX are shown in circles. **(D')** Hydroxylation of N1 by SxtX (cephalosporin hydroxylase) to produce NEO. **(D'')** N-sulfation at carbamoyl group nitrogen to form B1 (GTX5). **(E)** O- and N-sulfation to produce C1/2 analogues from GTX2/3 precursors.

In the *sxt* gene cluster of T3, *sxtN* is the only ORF described with a putative sulfotransferase function, assigned by a profile library search from which the two best hits in the database were with proteins with sulfotransferase functions and the next best hit was with a putative branched-chain amino acid aminotransferase (Kellmann *et al.,* 2008a). The function assigned to SxtN was supported by structural relatedness to an estrogen sulfotranferase 1AQU, and by similarity of the conserved N-terminal sequence to a PAPS binding domain (Kakuta *et al*., 1997). Our re-analysis of the sequence, however, found that there are several features in the 1AQU sequence related to PAPS binding that are not conserved in the SxtN sequence and may not allow SxtN to bind the sulfate donor (PAPS) (Figure 2.3.S2). Moreover, our analysis of SxtN showed no significant hit with any sulfotransferase after PSI-blastp. This is consistent with the absence of sulfated and sulfonated toxins in the T3 toxin profile (STX, NEO and dcNEO) as we confirmed by LC-MS/MS (Figure 2.3.4). An alternative explanation would be that SxtN is not functional as a sulfotransferase in T3 or that the proper precursor for the enzyme function is not available. We argue that the classification of the *sxtN* gene was forced to explain the toxin profile of T3.

The confirmed PSP toxin profile of *R. brookii* D9 includes STX and dcSTX, as well as the sulfated analogues GTX2/3 and dcGTX2/3. We postulate that differences in the sequence of *sxtN* and *sxtSUL* can explain the synthesis of GTX2/3 in D9 and the lack of function of the *sxtN* gene in T3, which is not acting as a sulfotransferase in T3, as proposed by Kellmann *et al.* (2008a).

With exception of *Lyngbya wollei,* all cyanobacterial strains which contain *sxtX* are able to synthesize the carbamoylated NEO analog (Table 2.3.2). In the case of *L. wollei* the absence is probably due to loss of function of *sxtI* preventing R4 carbamoylation and subsequent hydroxylation at the N-1 position. Furthermore, the two NEO-producing cyanobacteria are incapable of synthesizing the O- or N-sulfated analogues, whereas *L. wollei* can produce O-sulfated dcGTX2/3. This correlates with the presence of *sxtSUL* in the *sxt* gene cluster of this cyanobacterium.

The presence of the *sxtN* gene may correlate with synthesis of N-sulfonated analogues, but only *A. circinalis* AWQC131C that contains *sxtN* is able to produce the N-sulfonated analogues C1/2. However, in order to synthesize C1/2, it is necessary to synthesize GTX2/3 as precursors for the C1/2 N-sulfonation (Figure 2.3.6E); therefore an O-sulfotransferase function is needed. As *A. circinalis* AWQC131C produces GTX2/3, we postulate that the *sxtSUL* gene is located in another region of the genome.

Table 2.3.2 PSP toxin profiles and *sxt* gene content in cyanobacterial species.

Strain	PSP toxin profile (percentage)										*sxt* genes			Procedence	Ref.
	STX	**dcSTX**	**GTX2/3**	**dcGTX2/3**	**NEO**	**dcNEO**	**C1/2**	**B1 (GTX5)**	**B2 (GTX6)**	**LW1-6**	***sxtX***	***sxtN***	***sxtSUL***		
R. brookii D9	4.5	2.9	**94.6**	0.7							x	x	√	Brazil	This work
A. circinalis ACMB13	0.8	0.1	6.9				**86.6**	5.7			x	√	√	Australia	This work
A. circinalis AWQC131C	2.0	11	16	7			64				x	√	**	Australia	(Llewellyn *et al*., 2001)
C. raciborskii T3	2.8				**87**	10.1					√	√	x	Brazil	This work
Aphanizomenon sp. NH5	+				+						√	√	x	USA	(Mahmood & Carmichael, 1986b)
L. wollei		1.4		5.4						**(1) 24.5** (2) 13.9 **(3) 32.7** (4) 0.7 (5) 17.8 (6) 3.4	√	√	√	USA	(Onodera *et al*., 1997) (Carmichael, 1997)
Aph. gracile AB2008/31	10.5	6.3			2.9			**80.2**			?	?	?	Germany	(Ballot *et al*., 2010)
Aph. gracile AB2008/65	12.2	1.6			**69.4**			16.9			?	?	?	Germany	(Ballot *et al*., 2010)
Aph. issatschenkoii LMEYA 31	5.4	6.1			23			**64.5**	1.1		?	?	?	Portugal	(Pereira *et al*., 2000)
A. circinalis AWQC323B	8.0							**93**			?	?	?	Australia	(Llewellyn *et al*., 2001)

+ The analog is present but the accurate PSP toxin profile has not been determined by LC/MS.

** The presence could be checked.

? The sequences are not available.

We could not access ACWQ131C for direct comparison but we sequenced and analyzed the *sxt* gene cluster of the closely related *A. circinalis* ACMB13. This strain was also isolated from Australia and produces a toxin profile (STX, dcSTX, GTX2/3, dcGTX2/3, C1/2 and B1) similar to that of ACWQ131C. Most importantly, the *sxt* clusters of these strains are identical (100%). The genome of ACMB13 contains *sxtSUL*, as detected by positive PCR amplification with specific primers against the *sxtSUL* sequence, with 91.6% and 90.1% identity to SxtSUL of *L. wollei* and *R. brookii* D9, respectively.

Analysis of the *sxt* gene clusters and toxin profile of *A. circinalis* AWQC131C, *Aphanizomenon* sp. NH-5, *L. wollei* and the other four strains (whose sequences have been not described) allowed us to conclude that the toxin profile reflects an equilibrium between STX and production of terminal analogues, typically present in high molar percentage (bold numbers in Table 2.3.2). Terminal analogues could be sulfated and/or sulfonated (GTX2/3, B1, C1/2, LW1-3) or N-1-hydroxylated (NEO) derivatives (Table 2.3.2, Figure 2.3.6), indicating very efficient and specific enzymatic activity to utilize the available substrate (STX). Conversion of STX to NEO is completely favored in T3, even though the presence of SxtN could also allow conversion to B1, as in *Aphanizomenon issatschenkoi* LMEYA 31 (Table 2.3.2). This is more clearly illustrated by comparing the profiles of *Aph. gracile* AB2008/31 and AB2008/65, conspecific isolates from lakes in Germany, but indicating that in the former strain B1 synthesis is strongly favored versus NEO synthesis in the latter (Table 2.3.2).

In *R. brookii* D9, which does not possesses either SxtX or SxtN, conversion of STX to GTX2/3 is completely favored, supporting the proposed function of SxtSUL (CRD_02149) as a specific O-sulfotransferase. The case of *A. circinalis* AWQC131C is quite remarkable, because this cyanobacterium produces C1/2 analogues in higher proportion than all other analogues (as does strain ACMB13) (Table 2.3.2), but only encodes for SxtN inside the *sxt* gene cluster. As we propose, an O-sulfotransferase function is needed in the first place to form GTX2/3 as the precursors of C1/2. The presence of an *sxtSUL* sequence (91.6% identity to SxtSUL of *L. wollei*) in the genome of ACMB13, however, suggests the presence of an O-sulfotransferase function encoded outside of the *sxt* gene cluster in *A. circinalis*. Our hypothesis of the requirement for a specific precursor for a specific enzymatic function to produce the sulfated and sulfonated analogues from STX has no experimental support, but is highly correlated with the toxin profile of *L. wollei* and consistent with its *sxt* gene cluster composition.

L. wollei provides another remarkable example of the plasticity of the *sxt* gene cluster in cyanobacteria. The *L. wollei* cluster encodes the *sxtX* gene, two copies of *sxtN* gene and also the *sxtSUL* gene, but only produces dcSTX, dcGTX2/3 and LWTX 1-6. The latter toxins are result of *sxtI* gene disruption by *sxtACT* (predicted acyltransferases) that would transfer an acyl group to R4 and render the base of the LWTX structure. *L. wollei* is therefore unable to carbamoylate dcSTX (Figure 2.3.6B); such that the proper precursor for SxtX and SxtN is not present and only SxtSUL can act to O-sulfate dcSTX to dcGTX2/3.

2.3.4 Conclusions

The *sxt* gene cluster sequences and gene expression levels could give us a good approach to understand the genetic and enzymatic toxin basis of these toxin profiles differences.

We propose that STX is the necessary substrate for GTX2/3, B1 and NEO biosynthesis, corresponding to O-sulfotransferase *(sxtSUL*), N-sulfotransferase (*sxtN*) and N1-hydroxylation (*sxtX*) activities, respectively. Furthermore, GTX2/3 are the precursors needed for C1/2 biosynthesis by N-sulfonation of the carbamoyl group. Therefore, the absence of sulfated and sulfonated PSP toxin in *C. raciborskii* T3 and *A. flos-aquae* NH5, despite the presence of *sxtN* in each *sxt* cluster, could be explained by the substrate absence (no GTX2/3 production), and a complete synthesis favored to produce NEO, leaving no free substrate to produce B1 (Figure 2.3.6D``).

The accurate identification of the toxin profiles in this group of cyanobacteria was a primary task, in order to identify the function of the genes present in the cluster. At present, it has not been possible to perform mutagenesis experiments or heterologous expression of this kind of cluster. A strong bioinformatic basis could help to assign a correct hypothetical function to a given gene for the synthesis of STX and analogues, until a proper experimental condition is developed to test this function. Application of this “gene profile analysis” to the prediction of toxin profiles in new toxic isolates could also assist in the assessment of toxicity risk due to cyanobacterial blooms in natural water systems.

2.3.5 Materials and methods

Cyanobacterial cultures and growth conditions

The non-axenic *C. raciborskii* T3 strain, was kindly provided by Sandra Azevedo (Universidade Federal do Rio de Janeiro, Brazil). *Cylindrospermopsis raciborskii* CS-505, 506 and 511 were isolated from Australia and obtained from the CSIRO collection, Hobart, Tasmania. *Raphidiopsis brookii* D9 was isolated by re-cloning from the multiclonal isolate SPC338 collected in 1996 from the Billings freshwater reservoir near Sao Paulo (Brazil) (Castro *et al.*, 2004). Cyanobacteria were cultured in MLA medium according to Castro *et al.* (2004), at 25–28 ºC under fluorescent light at a photon flux density of 40 μmol m^{-2} s^{-1} on a light/dark photocycle of 12:12 h without aeration.

In order to demonstrate the morphological identity of *C. raciborskii* T3, as directly received from S. Acevedo, who is responsible for maintenance and distribution of the original stain, the isolate was examined and compared with our previous results. The morphological identification was assessed by light microscopy with oil immersion in a Nikon ECLIPSE TS100 inverted microscope, at 1000X amplification.

Genomic DNA isolation, amplification and sequencing

The DNA was extracted with the CTAB method described by Wilson, (1990). For PCR amplification of 16S rDNA, the general primers CYA106a forward and 1492 reverse were used, as described by Nübel *et al.* (1997) and Lane *et al.* (1985), respectively. Amplification of the putative *sxt* gene cluster, was carried out with *sxtA, sxtN, sxtX, sxtW, sxtV, sxtD*-context, and *sxtO*-context primers (forward and reverse). Primer sequences and PCR amplicon sizes are shown in Table 2.3.1. The PCR reaction contained 50–100 ng of genomic DNA. Reagents for each amplification were: 0.25U Taq DNA polymerase (Invitrogen®); 3 μl 10X PCR buffer (Invitrogen®); 2.5 mM MgCl (Invitrogen®); 0.4 mM primers; and 0.93 mM of each deoxynucleoside triphosphate (Promega®). Thermal cycling for 16S rDNA was performed in an Eppendorf Mastercycler, under the following conditions: initial DNA denaturation at 92 ºC for 2 min; 30 cycles at 94 ºC for 1 min, 56 ºC for 1 min, 72 ºC for 2 min and a final elongation at 72 ºC for 5 min.

Thermal cycling amplification for the *sxt* gene cluster was performed in an Eppendorf Mastercycler, under the following conditions: initial DNA denaturation at 99 ºC for 1 min, 30

cycles at 94 ºC for 15 s, 53 ºC for 1 min, 72 ºC for 1.5 min and a final extension at 72 ºC for 7 min.

All PCR products were checked by electrophoresis in 1.0% agarose electrophoresis gel and visualized under UV transillumination after staining with ethidium bromide.

PCR-generated fragments of the 16S rDNA and *sxt* gene cluster were purified, with a Wizard®*Plus* SV Miniprep, DNA Purification System (Promega®). PCR primers were used for sequencing of both DNA strands (Macrogen, Korea). All sequences were checked manually by Vector NTI software (Invitrogen®).

GenBank accession numbers

Published sequences were obtained from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) under the accession numbers: *C. raciborskii* T3 16S rDNA gene, partial sequence (EU439566); *C. raciborskii* T3 *sxt* gene cluster, complete sequence (DQ787200); *R. brookii* D9 *sxt* gene cluster, complete genome sequence (ACYB00000000); *Lyngbya wolleii sxt* gene cluster, complete sequence (EU603711); *C. raciborskii* CSIRO (CS) collection 16S+ITS-1(S): CS-505 (EU552062), CS-506 (EU552063), CS-511 (EU552068); *R. brookii* D9 16S+ITS-1(S) (EU552070). *A. circinalis* ACMB13 *sxtSUL* gene, partial sequence was submitted in this work under the accession number HM163164.

Toxin extraction of *C. raciborskii* T3

Cyanobacteria were harvested in exponential growth phase by centrifugation at 16,000 x g for 15 min to yield a cell pellet. The cell pellets were dried (DNA Speed Vac, SAVANT mod. DNA 110-230) and extracted in 300 µl of 0.05 M acetic acid. Samples were disrupted three times (20 s each) with an ultrasonic cell disruptor (Microson XL, Misonix, Farmingdale, USA). 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.

Liquid Chromatography with Fluorescence Detection (LC-FD)

The LC-FD analysis was carried out on a LC1100 series liquid chromatograph (Agilent Technologies, Waldbronn, Germany) coupled to a PCX 2500 post-column derivatization system (Pickering Laboratories, Mountain View, CA, USA). The LC-system consisted of a degasser

(G1379A), a quaternary pump (G1311A), an autosampler (G1229A), an autosampler thermostat (G1330B), a column thermostat (G1316A) and a fluorescence detector (G1321A).

Chromatographic conditions were as follows: mobile phase A: 6 mM 1-octanesulphonic acid and 6 mM 1-heptanesulfonic acid in 40 mM ammonium phosphate, adjusted to pH 7.0 with diluted phosphoric acid and 0.75% THF for the gonyautoxin group; mobile phase B: 13 mM 1-octanesulphonic acid in 50 mM phosphoric acid adjusted to pH 6.9 with ammonium hydroxide and 15% (v/v) of acetonitrile and 1.5% of THF for the saxitoxin group. The flow rate was 1 ml min^{-1} with the following gradient: 0 min, 100% A isocratic to 15 min, switch to 100% B until 16 min, isocratic B until 35 min, switch to 100% A until 36 min, isocratic 100% A until 45 min (= total run time). The autosampler was cooled to 4 °C and the injection volume was 20 µl. The separation of analytes was performed on a 250 × 4.6 mm i.d., 5 µm, Luna C18 reversed-phase column (Phenomenex, Aschaffenburg, Germany) equipped with a Phenomenex SecuriGuard pre-column. The eluate from the column was continuously oxidized with 10 mM of periodic acid in 550 mM ammonium hydroxide at a flow rate of 0.4 ml min^{-1} in a reaction coil set at 50 °C. Subsequently, the eluate was continuously acidified with 0.75 N nitric acid at a flow rate of 0.4 ml min^{-1} and the toxins were detected by a dual monochromator fluorescence detector (λ_{ex} 333 nm; λ_{em} 395 nm). Data acquisition and processing was performed with the HP ChemStation software. PSP toxin concentrations were determined by external calibration.

Standard solutions of PSP toxins (STX, NEO, dcSTX, dcNEO, gonyautoxins 1/4 (GTX1/4), GTX2/3, decarbamoyl gonyautoxins 2/3 (dcGTX2/3), B1 and C1/2) were purchased from the Certified Reference Materials Programme of the Institute for Marine Biosciences (National Research Council, Halifax, NS, Canada).

Liquid Chromatography Coupled with Tandem Mass Spectrometry (LC-MS/MS)

Mass spectrometric analyses for PSP toxins were performed according to a hydrophilic interaction liquid ion-chromatography (HILIC) method (Diener *et al.*, 2007) with slight modifications. Mass spectral experiments were performed on an ABI-SCIEX-4000 Q Trap, triple quadrupole mass spectrometer equipped with a TurboSpray® interface coupled to an Agilent model 1100 LC. The LC equipment included a solvent reservoir, in-line degasser (G1379A), binary pump (G1311A), refrigerated autosampler (G1329A/G1330B), and temperature-controlled column oven (G1316A).

The analytical column (150 × 4.6 mm) was packed with 5 µm ZIC-HILIC (SeQuant, Lund Sweden) and maintained at 35°C. The flow rate was 0.7 ml min^{-1} and gradient elution was performed with two eluants, where eluant A was 2 mM formic acid and 5 mM ammonium formate in acetonitrile/water (80:20 v/v) and B was 10 mM formic acid and 10 mM ammonium formate in water. The gradient was as follows: 20 min column equilibration with 80% A, linear gradient until 5 min to 65% A, then until 10 min to 60% A, then until 20 min 55% A, subsequent isocratic elution with 55% A until 24 min and finally return to initial 80% A until 25 min. Total run time: 45 min. 5 µl of sample were injected in triplicates.

Multiple reaction monitoring (MRM) experiments were carried out in positive ion mode by selecting the following transitions (precursor ion > fragment ion), period 1 (B-, C- and gonyautoxins): *m/z* 412>332 and *m/z* 412>314 (for GTX1/4 and C3/4), *m/z* 396>316 and *m/z* 396>298 (for GTX2/3, C1/2 and B2), *m/z* 380>300 and *m/z* 380>282 (for B1 = GTX5), *m/z* 353>273 (for dcGTX2/3), *m/z* 369>289 (for dcGTX1/4); period 2 (STX, NEO and their decarbamoyl derivatives): *m/z* 300>282 and *m/z* 300>204 (for STX), *m/z* 316>298 and *m/z* 316>196 (for NEO), *m/z* 257>196 and *m/z* 257>156 (for dcSTX) and *m/z* 273>255 (for dcNEO). Dwell times of 100–200 ms were used for each transition. For these studies the source parameters were as follows: curtain gas: 30 psi, temperature: 650 °C, ion-spray voltage: 5000 V, gas 1 and 2: 70 psi, interface heater: on, collision gas: high, declustering potential: 66 V, entrance potential 10 V, collision energy: 30 V and collision cell exit potential: 12 V.

2.3.6 Supplementary material

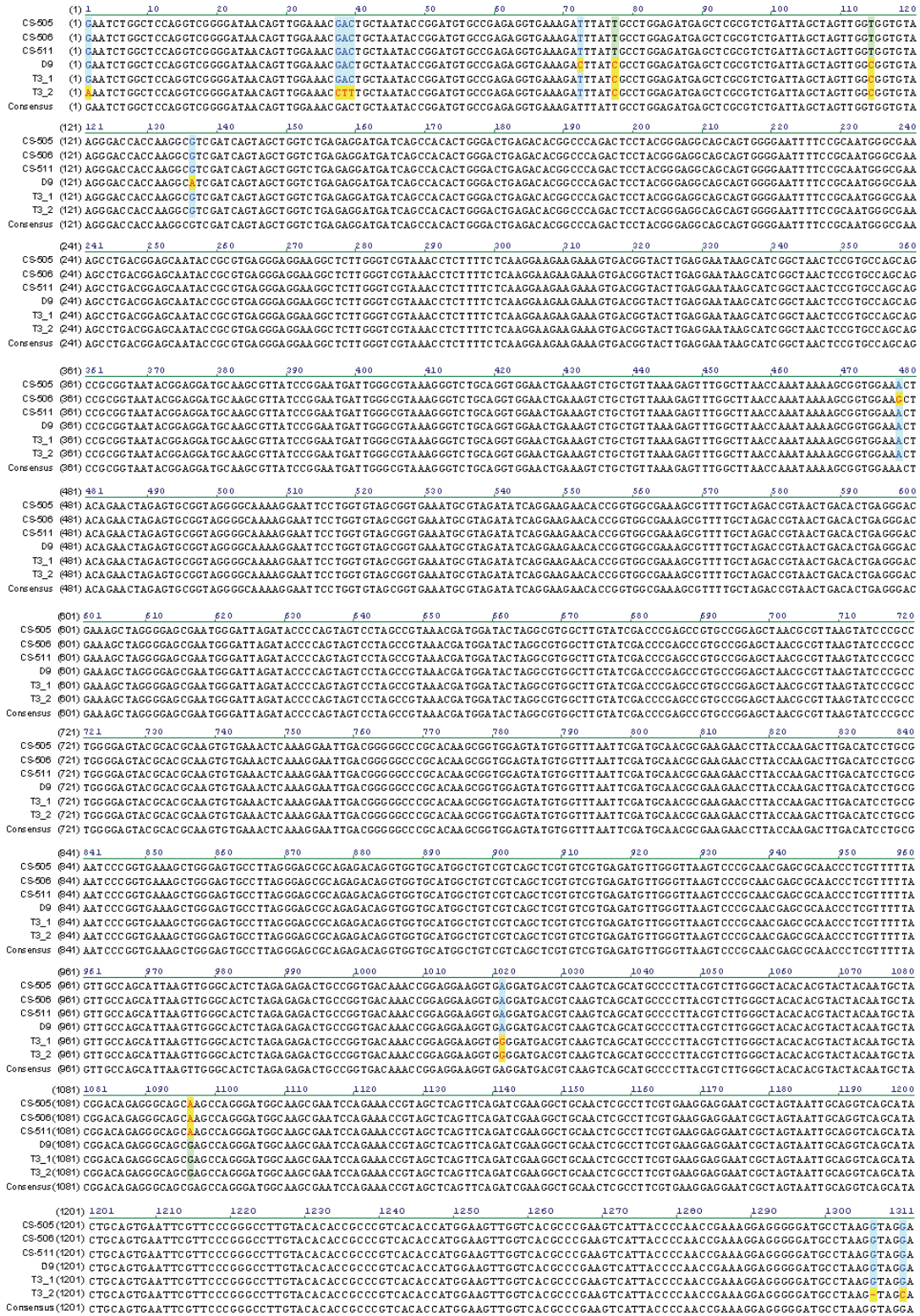


Figure 2.3.S1 Alignment of partial 16S rRNA nucleotide sequences of *C. raciborskii* T3 with other *C. raciborskii* strains and *Raphidiopsis brookii* D9. Alignments were performed with clustalX. The degree of nucleotide conservation is indicated by colors as follows: yellow blocks (non-similar), light blue blocks (conserved), light green blocks (similar), and white (identical). Accession numbers from the sequences analyzed are: T3_1: *C. raciborskii* T3 from this work; T3_2: *C. raciborskii* T3 (EU439566); D9: *R. brookii* D9 (EU552070); CS-505, CS-506 and CS-511: *C. raciborskii* (EU552062, EU552063 and EU552068, respectively).

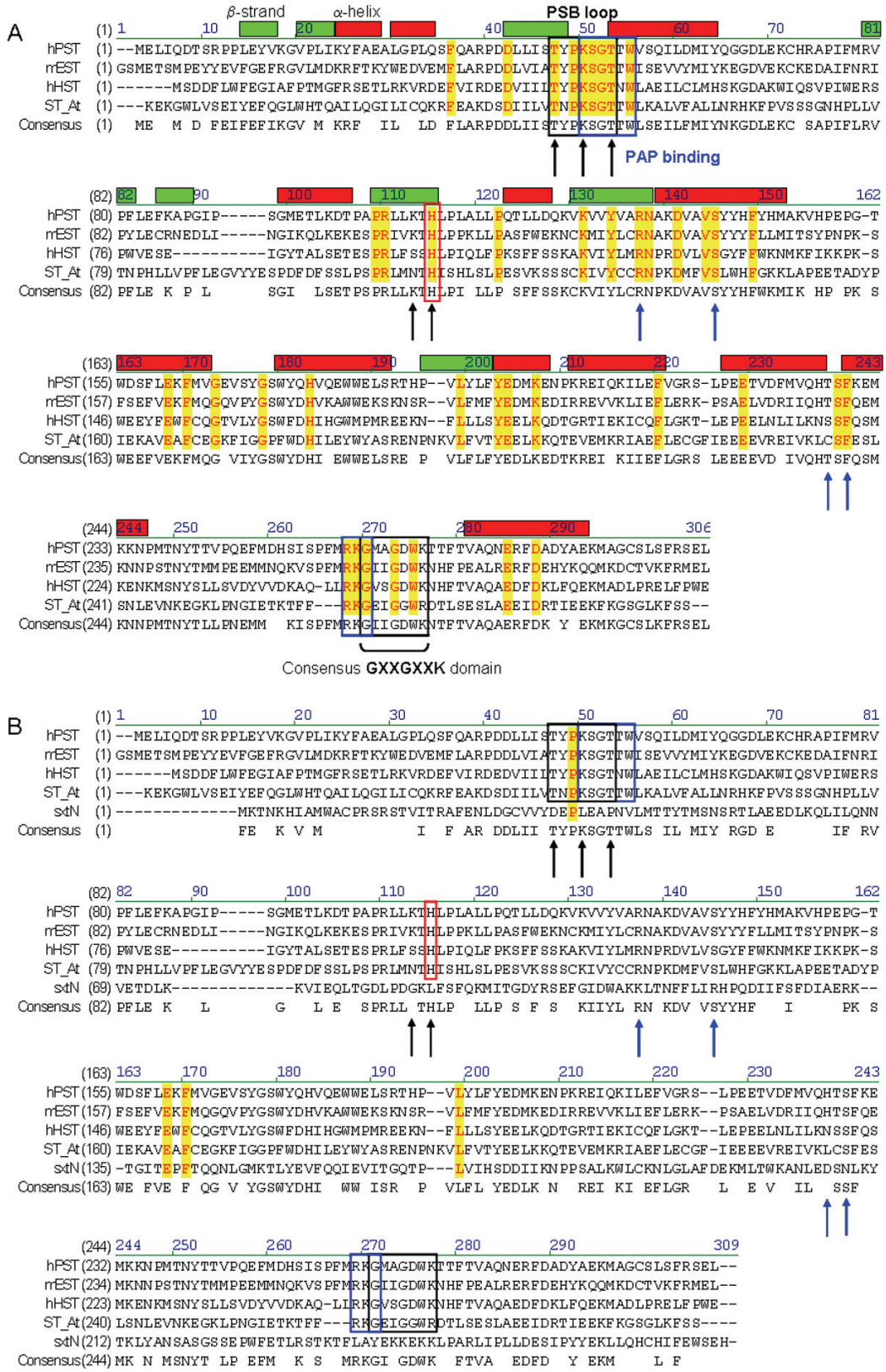

A
β-strand
α-helix
PSB loop
PAP binding
hPST
mEST
hHST
ST_At
Consensus
Consensus GXXGXXK domain
B
hPST
mEST
hHST
ST_At
sxtN
Consensus

Figure 2.3.S2 Amino acid alignment of sulfotransferase sequences described by Kakuta *et al.* (1997), for crystallographic prediction of this type of enzyme structure. (A) Mouse-estrogen sulfotranferase (EST), human phenol sulfotransferase (PST), human hydroxysteroid sulfotransferase (HST), and flavonol 3-sulfotransferase from *Arabidopsis thaliana*. The amino acids in red with yellow background are conserved among the protein families (31 sequences analyzed). The secondary structure elements are shown in green boxes (β-strain) and red boxes (α-helix). Residues involved in PAPS binding are marked in blue boxes and blue arrows. The residues suggested to be involved in catalytic mechanism are marked with black arrows and the histidine believed to be the proton acceptor in the reaction is labeled in red box. The consensus GXXGXXK and PSB loop are boxed. **(B)** The same alignment including SxtN sequence.

2.4 Publication IV

Effects of alternative nitrogen sources on gene expression and toxin production in the closely related cyanobacteria *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9

Karina Stucken, Uwe John, Allan Cembella, Katia Soto-Liebe and Mónica Vásquez.

2.4.1 Abstract

The closely related cyanobacteria *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9 share 2,539 orthologous genes, but morphologically differ by the presence of heterocysts (diazotrophy) in the former. CS-505 and D9 produce cylindrospermopsin (CYN) and paralytic shellfish toxins (PSTs), respectively, constitutively biosynthesized by independent gene clusters. It is not known how or if toxin biosynthetic genes are regulated, particularly by N-source dependency. Herein we show the N-dependent regulatory pathways and production of CYN and PSTs. In early exponential phase, growth inhibition by N-deprivation or high ammonium ion concentration, arrest PST- and slow CYN-production. In CS-505, up to 23% of repeated elements and several transposases were differentially expressed under all N-regimes, apparently playing critical regulatory roles in metabolism. Comparative transcriptomics revealed a differential regulation in N control and heterocyst differentiation genes between CS-505 and D9, and late regulation of *ntcA*. Quantification of the main toxin biosynthesis genes by qPCR showed a tendency for repression of the PST and CYN genes in both control cells and under different N-regimes. The late regulatory responses and the pattern of biosynthetic gene regulation suggest that additional mechanisms of N control and posttranscriptional mechanisms are involved in toxin production.

2.4.2 Introduction

Characteristic seasonal blooming of planktonic populations of unicellular and filamentous cyanobacteria, which often dominate the photosynthetic plankton in eutrophic freshwater and coastal brackish waters, has been documented for many years (Francis, 1978). Among cyanobacteria numerous different forms of nitrogen (N), primarily nitrate/nitrite, ammonium, urea and cyanate can be assimilated, as well as dinitrogen (N_2) through fixation in some cases (Herrero *et al*., 2001). Many members of both unicellular and filamentous cyanobacterial groups are diazothrophs, i.e. able to fix N_2, and therefore can survive and even thrive in N-poor environments.

A wide array of neurotoxins and hepatotoxic components are produced among the cyanobacteria (Carmichael, 1997, Sivonen & Jones, 1999). These toxins include the paralytic shellfish toxins (PSTs), comprising the tetrahydropurine saxitoxin (STX) and more than 20 naturally occurring analogues, potent neurotoxins that act by blocking voltage-gated sodium channels in excitable cells (Llewellyn, 2006). The production of PSTs is long known in cyanobacterial genera, such as *Anabaena*, *Aphanizomenon, Lyngbya, Cylindrospermopsis* (Humpage *et al.*, 1994, Carmichael *et al.*, 1997, Lagos *et al.*, 1999), and more recently these toxins have been confirmed in *Raphidiopsis* (Yunes *et al*., 2009). Cylindrospermopsins (CYNs) are polycyclic uracil derivatives with broad cytotoxic properties found among several cyanobacterial genera, including in species of *Anabaena*, *Aphanizomenon*, *Umezakia, and Raphidiopsis.* The distribution of biosynthetic capacity for either PST or CYN production indicates that although these toxin groups may occur among closely related taxa and among conspecific isolates from a given biogeographical region, they do not usually co-occur within a given strain.

The cosmopolitan filamentous diazotroph *Cylindrospermopsis raciborskii* is a frequent major component of cyanobacterial blooms in tropical and temperate waters (Padisák, 1997). Within this species, certain strains can produce either CYN or PSTs (Falconer & Humpage, 2006, Lagos *et al*., 1999). The planktonic filamentous genus *Raphidiopsis* is less frequently found within cyanobacterial blooms than *C. raciborskii*. The association of *C. raciborskii* with blooms of *Raphidiopsis* spp. is, however, increasingly reported in freshwater reservoirs and lakes in Australia, Brazil, Argentina and Egypt, where they coexist under the same environmental

conditions (McGregor & Fabbro, 2000, Chellappa & Costa, 2003, Zalocar de Domitrovic *et al.*, 2007, Mohamed, 2007).

Cylindrospermopsis raciborskii and *Raphidiopsis* spp. differ morphologically in the development of functional heterocysts (specialized cells for N_2-fixation) only in the former species. In fact, this morphological difference in heterocyst formation was the primary reason for the taxonomic separation of *Cylindrospermopsis* and *Raphidiopsis* into different genera (Komárek & Anagnostidis, 1989). Nevertheless, phylogenetic analysis has shown that taxa of both genera form part of the same species-complex (Gugger *et al.*, 2005, **Publication I**). Our recent report of the genome sequences of *C. raciborskii* CS-505 and *R. brookii* D9, showed that these taxa share a core genome of 2,539 genes (>90% amino acid identity). Furthermore, all core genes are present in syntenic regions, suggesting that these close relatives should be reclassified as species within the same genus (**Publication II**).

The genomes of both strains *C. raciborskii* CS-505 and *R. brookii* D9 contain the genes involved in the classical pathways for N uptake and assimilation, and accordingly both strains grow well on ammonium, nitrate, or urea as N sources. Among the more prominent differences between CS-505 and D9 at the genomic level are the absence of nitrogenase and heterocyst development gene clusters, and the presence of genes involved in amino acid transport and metabolism in D9. This latter group of genes could presumably provide a compensating advantage in N-limited environments.

Despite genomic differences related to N-metabolic capacities, both strains produce highly nitrogenated toxins. Strain CS-505 produces cylindrospermopsin (CYN) (Saker *et al*., 1999), whereas D9 synthesizes several PSTs, specifically STX, the gonyautoxins GTX2/3 and the respective decarbamoyl analogues dcSTX and dcGTX2/3 (Soto-Liebe *et al*., 2008). Cylindrospermopsin is biosynthesized via combined polyketide synthase/nonribosomal peptide synthetase (PKS/NRPS) pathways (Mihali *et al*., 2008), whereas STX and analogues are derived via a novel gene cluster whose mosaic structure determines the synthesis of the different PST analogues (Kellmann *et al.*, 2008a, Mihali *et al.*, 2009). Interestingly, the gene clusters involved in the biosynthesis of CYN and PSTs in CS-505 and D9, respectively, are located adjacent to genes regulated by *ntcA*, the master regulator for N availability (Vega-Palas *et al*., 1990). This raises the hypothesis that toxin synthesis could be regulated by nitrogen in these two cyanobacteria.

Given the critical importance of N supply and metabolism in growth, bloom development and life history transitions of cyanobacteria, we compared the respective effects of nitrate, ammonium and urea as N-sources with N-deprivation in CS-505 and D9. These strains were selected as optimal candidates for investigating the cellular growth responses and toxin production, based upon their close taxonomic association, the availability of comprehensive genome sequences, and proximity of the toxin clusters to NtcA-regulated genes in both CS-505 and D9. The availability of the sequences of the CYN and STX clusters allowed us to perform a bioinformatic search for *ntcA* regulatory boxes. We assayed the expression of the main genes involved in toxin biosynthesis and transport by qPCR. Microarray hybridization of the transcriptome 24 h after the change in the N-source yielded an integrated view of the regulatory pathways under these conditions. Our work is the first comparative study on gene expression related to growth, N metabolism and toxin production in these cyanobacteria to determine the different regulatory mechanisms that could underlie their phenotypic plasticity.

2.4.3 Results

Growth and toxin production under different N-regimes

Cylindrospermopsis raciborskii CS-505 and *Raphidiopsis brookii* D9 were adapted to grow on nitrate, but they were also able to use either urea or ammonium as substrate, measured as growth normalized per unit chl *a* or dry weight (Figure 2.4.1). These cyanobacteria were, however, not able to grow on cyanate (data not shown), which can be spontaneously derived from urea. With the exception of lower biomass yield in D9 grown on ammonium (Figure 2.4.1D), growth was not highly dependent on the type of N-substrate. Although D9 does not grow diazotrophically, the cultures showed an increase in biomass up to 48 h in -N medium (Figure 2.4.1D), indicating usage of intracellular N. After 72 h, the biomass decreased and the filaments appeared fragmented, as an indication of the start of cell death. Ammonium affected D9 by producing a growth-lag that was less apparent for CS-505 (Figure 2.4.1C,D).

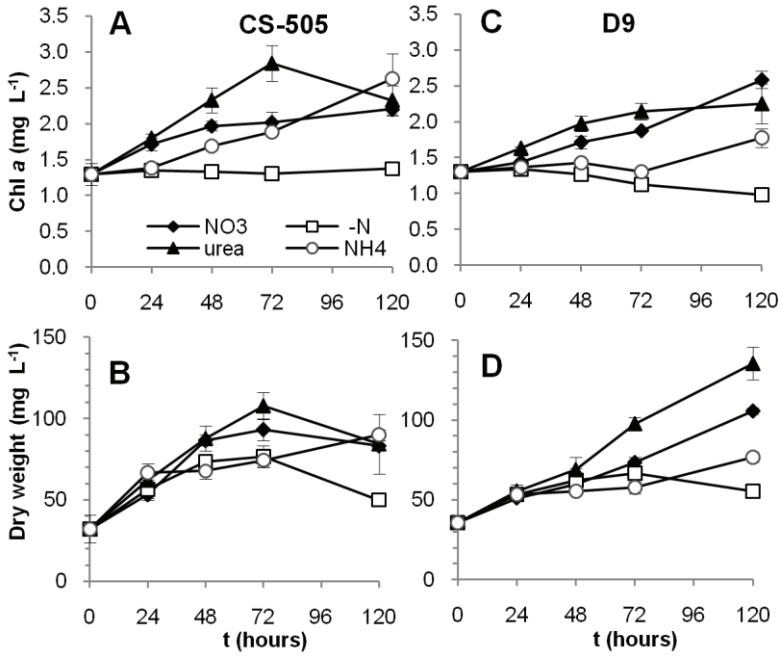


Figure 2.4.1 Growth curves of *C. raciborskii* CS-505 (A, B) and *R. brookii* D9 (C, D). Curves were plotted by chlorophyll *a* and dry weight per unit culture volume. Values are shown as the average of three biological replicates; error bars indicate ±SD from the mean (n = 3).

The initially high C:N ratios further increased in both strains after 24 h and remained high in the absence of added N (Figure 2.4.2). In CS-505, however, high C:N ratios were maintained until 72 h, but after 120 h, the ratio significantly decreased (Figure 2.4.2A), temporally correlating with the formation of heterocysts and consequent N_2 fixation (**Publication V**). In contrast, lacking the ability for N_2 fixation, D9 cultures remained N-starved as indicated by the high C:N ratio during the entire experiment. The C:N ratios also revealed that CS-505 was more efficient at assimilation of ammonium than D9; in the latter strain grown on ammonium, the cellular C:N ratio did not decrease until 72 h, indicating that D9 cells were not effectively retaining externally supplied N in the form of ammonium (Figure 2.4.2B). Both strains showed a rapid assimilation of urea as N-substrate until 48 h, but later the C:N ratio decreased below the initial values in CS-505, whereas in D9 this ratio increased even above the values shown with nitrate.

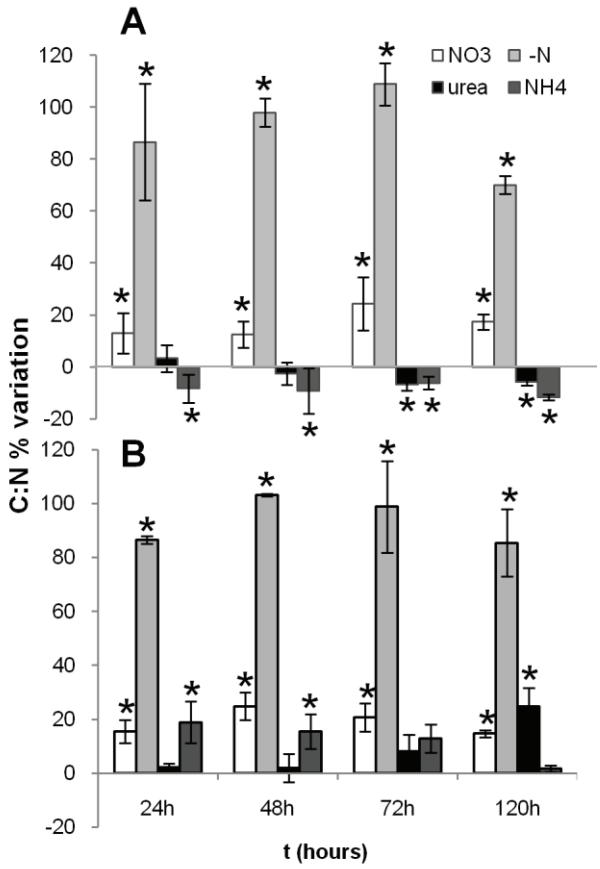


Figure 2.4.2 C:N ratios of CS-505 (A) and D9 (B) grown under alternative N-regimes. Values are shown as a percentage of variation with respect to t = 0; error bars indicate ±SD from the mean (n = 3). Values significantly different (t-test p-value< 0.05) from t = 0 are marked with a * symbol.

Total intra- and extracellular CYN and deoxycylindrospermopsin (doCYN) were increasingly produced in CS-505 under all N-treatments (Figure 2.4.3A). This response was however, much lower in the treatment without fixed N (Figure 2.4.3A). When normalized to biomass, the effect of N-deprivation on toxin production correlated with the available N pool in the cell; the toxin content per unit biomass that remained low until 72 h, increased after 120 h (Figure 2.4.3B), along with a drop by almost 50% of the C:N ratio (Figure 2.4.2A). These changes reflect an increase in cellular N, probably derived from N_2 fixation. Similarly, only N-deprivation significantly affected the intra- and extracellular CYN:doCYN, ratio (Student's t-test, p<0.05) which increased as reflected by a dramatic decrease in the precursor doCYN (Figure 2.4.3C,D). The CYN:doCYN ratios progressively decreased in the cellular and extracellular fraction under all fixed N-sources (Figure 2.4.3C,D), although this trend was not as marked by 120 h in the extracellular fraction (Figure 2.4.3D). Small differences in this tendency are likely attributable to cell lysis and/or leakage of toxins rather than to active toxin transport.

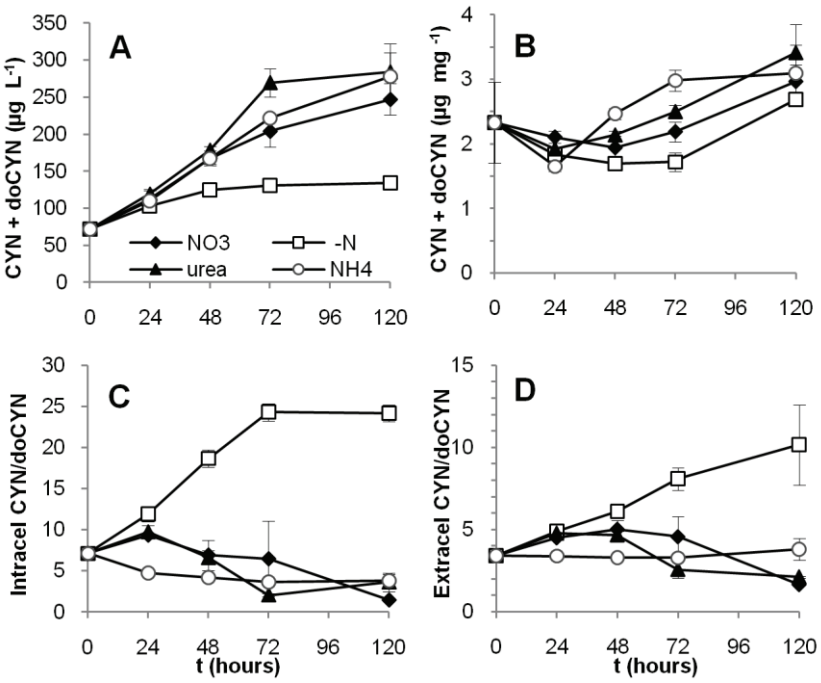


Figure 2.4.3 Toxin production by CS-505 grown under four alternative N-regimes. **(A)** Total CYN + doCYN content over the time in the intra- and extracellular component per unit culture volume. **(B)** Total toxin content normalized to biomass (dry weight). **(C)** and **(D)** Intra- and extracellular ratios of CYN:doCYN, respectively.

Rates of toxin production (μ_{tox}) and growth rate (μ_c) showed a significant correlation in cultures grown on nitrate and urea over the entire growth curve, with a slope not significantly different from 1 (ANCOVA: p = 0.204550 for nitrate and p = 0.566347 for urea), suggesting a balanced equilibrium between net growth and toxin production. In contrast, cultures grown under N-deprivation showed a correlation between μ_{tox} and μ_c whose slope (m = 0.447) differed significantly from a slope = 1 (ANCOVA: p = 0.000930). No significant correlation was observed for the ammonium treatment, for which toxin production rate appeared to be independent of growth rates (Figure 2.4.S1A).

Raphidiopsis brookii D9 produces six PST analogues in the following order of descending relative abundance: GTX2/3>STX>dcSTX and dcGTX2/3. The total PST content (intra- and extracellular) of the D9 cultures grown on nitrate or urea increased along the time course of growth, but remained roughly constant in cultures grown with ammonium or under N–deprivation over the 120 h experiment (Figure 2.4.4A).

The total toxin per unit biomass (µg PSTs mg^{-1}) went down during the first 24 h in all N-treatments, and continued to decrease in the ammonium and urea treatments, where it reached a minimum of 0.86 µg mg^{-1} (*ca.* 50% of the initial amount) at 120 h (Figure 2.4.4B). In contrast, under –N conditions and with nitrate, the total PST content per unit biomass increased towards the end of the experiment but did not attain the levels measured at $t = 0$. The increase of total PST in the –N treatment may be explained by the low biomass at this time point while the total PST remained constant (relatively refractory to degradation) in the medium. The trend towards decreasing intracellular toxins, except for the nitrate treatment, was only reflected in the increase in the extracellular STX pools in cultures grown on urea; under ammonium and N-deprivation, STX did not substantially increase (Figure 2.4.4C, D).

In the intracellular fraction, the toxin composition, expressed as the STX:GTX2/3 ratio, rapidly shifted down in the –N and ammonium treatments, while decreasing only later in the urea-grown cultures (Figure 2.4.4E). In the ammonium treatment, the intracellular content of STX went down to the limit of detection (Figure 2.4.4C), thereby producing lower STX:GTX2/3 ratios; the same effect was observed in the –N treatment (Figure 2.4.4E,F). In these cases, the cultures were in two different physiological stages; N-depleted cultures were destined to die from starvation, whereas cultures grown in ammonium only showed a lag phase until the fourth day, after which they increased in biomass and chlorophyll *a* content (Figure 2.4.1). With nitrate as growth substrate, the STX:GTX2/3 ratio (approximately 1:5) did not significantly change (Student's t-test $p<0.05$) over the first 48 h, but increased slightly in the latter half of the experiment (Figure 2.4.4E,F). In general, the ratios in the intracellular and extracellular fractions followed similar patterns, with the exception of the urea treatment, which showed a downshift only in the intracellular component, presumably due to the release of STX to the extracellular medium.

The rate of PST production was apparently independent of the growth rate in the urea, nitrate and ammonium treatments; only the –N treatment showed a significant positive correlation between these two parameters. Nevertheless, the slope of the regression line ($m = 0.1647$) was significantly different from a slope = 1 (ANCOVA $p = 0.000116$), showing that under N-deprivation, a doubling in the biomass did not yield a corresponding doubling in toxin content (Figure 2.4.S1B).

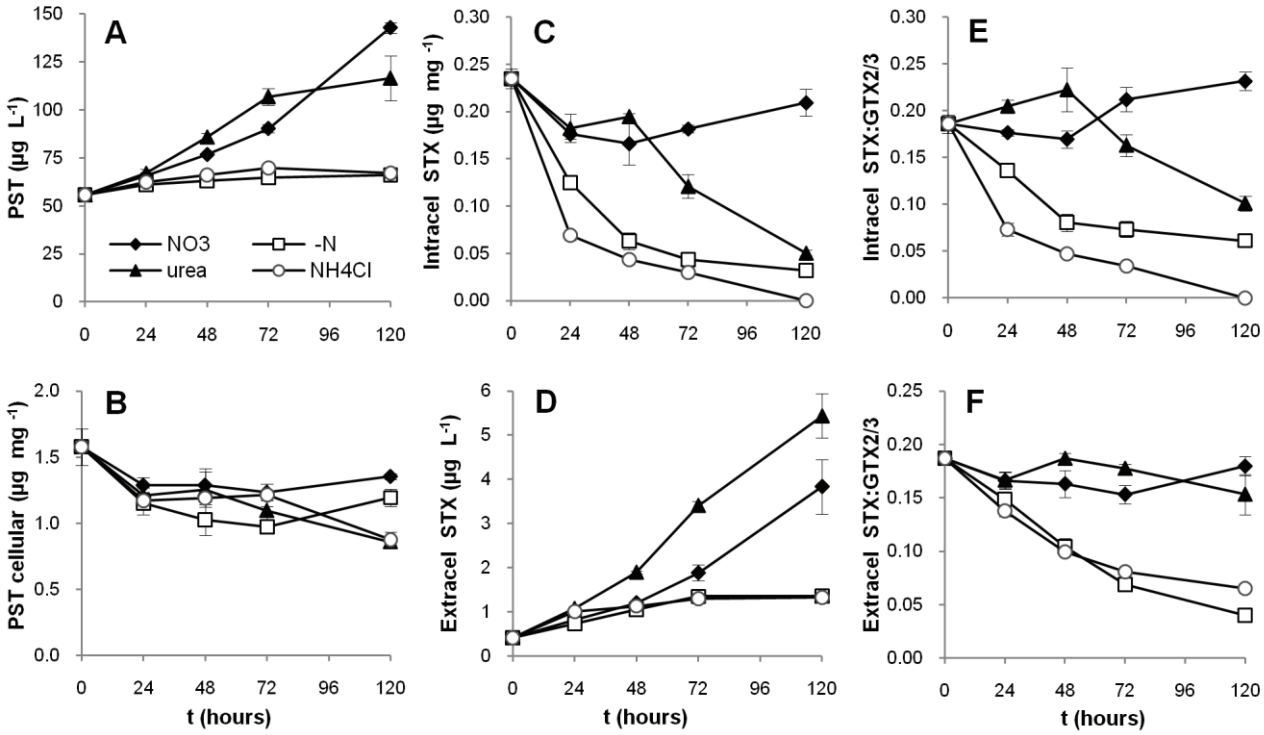


Figure 2.4.4 Toxin production by D9 grown under four alternative N-regimes. **(A)** Total PST content over the time in the intra- and extracellular component per unit culture volume. **(B)** Total toxin content normalized to biomass (dry weight). **(C)** and **(D)** Intra- and extracellular STX content. **(E)** and **(F)** Intra- and extracellular STX:GTX2/3 ratios.

Core genes and transcriptional responses to alternative N sources

Analysis of the coding sequences (CoDing Sequence, CDS) of the cyanobacterial strains revealed that of the 3,492 CDS identified in the CS-505 genome, 2,539 (73%) have an ortholog in the D9 genome, comprising 3,010 CDS (of which 84% are shared with CS-505). We refer to this set of 2,539 orthologs as the "core genome" shared between CS-505 and D9. Following growth in all four N-regimes, there was an overlap between strains in the expression of 12% of the genes belonging to the core (Figure 2.4.5). This fragment of the core represents the main processes known to be affected in cyanobacteria under N-limitation of growth and hence includes genes highly conserved in cyanobacteria and eubacteria (Table 2.4.S1). Transferring CS-505 and D9 cultures to nitrate-replete medium provoked a mild induction/repression in a large number of genes, of which a few were regulated with fold changes -3>FC>3. The common response

between the strains was reflected by the overlap in the expression of 44 core genes, a low number considering that 257 genes belonging to the core were expressed in D9 alone. As in the nitrate treatment, the overlap between strains in the core gene expression was low (98 core genes) when compared to the 471 genes from the core expressed in D9. From the overall pattern of expressed genes in CS-505 and D9 under N-deprivation, 218 belonged to the core and were expressed in both strains. Among all N treatments, in both CS-505 and D9 we observed the lowest number of regulated genes with ammonium, but with the highest magnitude in gene expression (repression) levels. From the 29 core genes regulated in both strains, almost all (27) were down-regulated.

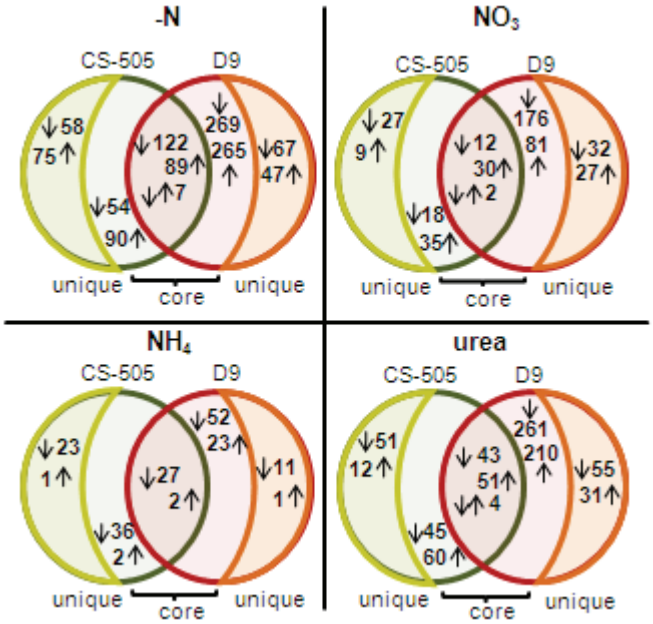


Figure 2.4.5 Venn diagrams representing the overlap in expression of the core genes between CS-505 and D9. Arrows indicate the direction of regulation (up/down).

Although for both strains only a fraction of the expressed genes belonged to the core genome, not all genes were regulated in the same direction (Figure 2.4.5, Table 2.4.S1). For example, in the –N treatment, seven genes were regulated in an opposite manner; *patU3*, typically associated with heterocyst development, was up-regulated in CS-505 but down-regulated in D9. Likewise, under nitrate-dependent growth, two genes were differentially regulated, but they are only annotated as hypothetical proteins. Genes differentially regulated with urea clearly pointed towards different N-uptake and utilization responses between CS-505 and D9. Whereas in CS-505 the urease transport gene *urtA* and a GCN5-related N-acetyltransferase (CRD_00043/CRC_02981) were up-regulated, in D9 an ABC transporter

(CRD_00713/CRC_00490) and a monooxygenase (CRD_00742/CRC_02427) were induced. No genes with opposite expression patterns were observed in the ammonium treatment.

The significant shifts detected at 24 h in both toxin and C:N ratios were early enough to exclude the possibility that they were due to an acclimation response of CS-505 to N-deprivation (trigger of N_2 fixation and heterocyst development genes), or to cell death in D9 caused by extreme N-stress. More genes were regulated in D9 than in CS-505 under the nitrate-, urea- and –N-treatments (Table 2.4.1). The transcriptional responses to ammonium, on the other hand, exhibited that lowest total of regulated genes among all N-treatments, but the number of down-regulated genes was similar between the strains (Table 2.4.1).

Table 2.4.1 Number of genes and repeated sequences regulated in CS-505 and D9 at 24 h under alternative N-regimes. FC = fold change

q = 0, FC > 1.5 LOESS	**CS-505**		**D9**	
Nitrogen regime	**Up-regulated**	**Down-regulated**	**Up-regulated**	**Down-regulated**
nitrate	75 + 2 repeats	57 +41 repeats	139	221
ammonium	5	86 + 7 repeats	26	90
urea	126 + 1 repeat	140 + 52 repeats	294	361
-N added	258 +10 repeats	237 + 86 repeats	404	461

Particularly remarkable was the differential expression in CS-505 of several transposases and repeat sequences, almost all of which were repressed in all N-treatments. In CS-505 grown on nitrate, the number of regulated repeats reached 70% of expressed genes. Large numbers of repeats were also regulated in the urea and –N treatments. In contrast, none of the few repeats of the D9 genome changed their expression levels in any treatment; under ammonium, however, two transposases were up-regulated.

Although expression patterns of the core genes provided an overview of the main processes occurring in parallel in CS-505 and D9, repressed processes were poorly represented. Among the repressed genes in CS-505 under nitrate, the condition that showed the largest number of repeats and transposases regulated, three quarters comprised hypothetical proteins, and genes similar to those of ABC domains. Half of these genes are encoded in CS-505 without orthologs in D9 (unique CS-505 genes). The nature of these processes is therefore difficult to predict, but if we add the number of repressed repeated sequences, the processes most inhibited in CS-505 are recombination and genome rearrangements. The large number of genes repressed in D9 alone (209) mostly represented processes of DNA replication and exogenous transfer, transcriptional regulation, carbohydrate biosynthesis and cell wall degradation

Nitrogen control genes

Several genes and gene clusters were regulated under most or all N-regimes in D9 or CS-505 or in both strains. The regulation of some of these genes followed the typical pattern of N regulation for alternative N sources, i.e. induction under N-deprivation and repression when ammonium is present in the culture medium. Genes that followed the pattern of gene regulation of N control in D9 were the ammonium transporters *amt1, amt2*, the *urtA* and *nirA* genes, the gene for the N-regulatory protein P-II *glnB,* the inactivating factor for glutamine synthase (GS) *gifA,* and a gene cluster (CRD_00336-9) with similarity to nitrile hydratases (*nhlABE*) with an associated cobalamin synthesis protein necessary for the cobalt cofactor center of the enzyme. On the contrary, in CS-505 only the *nrt* genes, *gifA* and the glutamine synthase gene *glnA,* followed the N control expression pattern (Table 2.4.2).

These differences in regulation patterns were reflected mainly in the expression of genes in the urea and N-deprivation treatments. For example, under N-deprivation the *urt, nrt,* and *glnA* genes were induced in CS-505 alone. The pattern of expression of the *glnB* and of *nrrA* a transcriptional regulator, whose function has been linked to heterocyst development in *Anabaena* sp. PCC 7120 (Ehira & Ohmori, 2006b), was of particular interest. Both genes were induced under N-deprivation in CS-505 and D9; however, when ammonium or urea was supplied as N source, their expression remained unchanged in CS-505, whereas in D9 they were strongly repressed in both ammonium and urea.

Table 2.4.2 Regulation of N control, heterocyst differentiation and carbon metabolism genes. Expression values are given as fold changes.

				CS-505				D9			
gene	**CS-505**	**D9**	**Product / Description**	**-N**	**NO_3**	**NH_4**	**urea**	**-N**	**NO_3**	**NH_4**	**urea**
ntcA	CRC_00858	CRD_00550	Nitrogen-responsive regulator	--	--	--	--	-1.7	--	--	--
hetR	CRC_03184	CRD_01519	Heterocyst differentiation protein	--	--	--	--	1.9	--	--	--
nrrA	CRC_00297	CRD_01945	Transcriptional regulator	5.4	2.3	--	--	3.8	2.1	-6.8	-4.5
glnB	CRC_02206	CRD_00313	Nitrogen regulatory protein P-II	3.9	--	--	--	2.5	1.7	-3.8	-2.4
glnA	CRC_02160	CRD_02136	Glutamine synthetase type I	2.0	2.4	-2.2	--	--	--	-3.0	-2.2
gifA	CRC_02159	CRD_02135	GS inactivating factor, IF7	-6.9	-4.0	5.5	2.2	-6.6	-9.1	2.6	--
nirA	CRC_00047	CRD_02595	Nitrite reductase	3.5	--	-43.7	-6.0	1.8	1.7	-47.1	-37.0
nrtA	CRC_00046	CRD_02594	Nitrate transporter	4.2	2.4	-23.9	-4.3	--	--	-34.9	-30.4
nrtB	CRC_00045	CRD_02593	Nitrate transporter	4.1	1.9	-14.8	-4.0	--	--	-25.5	-35.0
nrtB		CRD_02592	Nitrate transporter	-----	-----	-----	-----	--	--	-24.1	-29.6
nrtC	CRC_00044	CRD_02591	Nitrate transporter	4.1	1.9	-10.7	-4.9	--	--	-13.1	-22.5
nrtC		CRD_02590	Nitrate transporter	--	--	--	--	--	--	-16.2	-21.5
nrtD	CRC_00043	CRD_02589	Nitrate transporter	3.8	--	-19.2	--	--	--	-27.3	-18.9
narB	CRC_00042	CRD_02588	Nitrate reductase	--	--	-3.8	-2.9	--	--	-11.9	-12.1
narB		CRD_02587	Nitrate reductase	-----	-----	-----	-----	--	--	-10.2	-9.9
urtA		CRD_01584	Urea transporter	-----	-----	-----	-----	3.1	2.1	-6.5	-3.8
urtA	CRC_00451	CRD_02952	Urea transporter	4.2	3.0	--	2.4	1.7	--	--	--
urtB	CRC_00450	CRD_02953	Urea transporter	1.8	--	--	--	--	--	--	--
urtC	CRC_00449	CRD_02954	Urea transporter	2.1	--	--	--	--	--	--	--
urtD	CRC_00448	CRD_02955	Urea transporter	2.4	2.3	--	2.4	--	--	--	--
urtD		CRD_02956	Urea transporter	-----	-----	-----	-----	--	--	--	--
urtE	CRC_00447	CRD_02957	Urea transporter	1.7	2.4	--	2.1	--	--	--	1.5
amt1	CRC_00109	CRD_01628	Ammonium transporter	5.6	3.9	-5.1	--	2.3	2.2	-9.8	-2.5
amt2	----------	CRD_00091	Ammonium transporter	-----	-----	-----	-----	2.5	2.3	-5.0	-2.1
abp3	CRC_00830	----------	Heterocyst glycolipid formation	--	--	-2.7	--	-----	-----	-----	-----

devH	CRC_03275	CRD_01596	Heterocyst polysaccharide formation	1.8	2.2	--	2.4	2.0	1.9	--	--
hanA	CRC_01160	CRD_00595	Heterocyst differentiation	1.9	--	--	--	--	--	--	--
hglA	CRC_02051	CRD_00121	Heterocyst glycolipid formation	1.9	--	-1.9	--	--	--	--	--
hetM	CRC_02052	----------	Heterocyst glycolipid formation	1.5	--	--	--	-----	-----	-----	-----
hetP	CRC_01587	----------	Heterocyst differentiation	2.2	--	--	--	-----	-----	-----	-----
patU3	CRC_02800	CRD_02293	Heterocyst differentiation	2.0	--	--	--	-1.6	--	--	--
hetZ	CRC_02801	CRD_02292	Heterocyst differentiation	--	--	--	--	-1.8	--	--	--
sigG	CRC_00607	CRD_02710	Transcriptional regulator	--	--	-2.3	-2.0	-1.7	-2.2	--	-2.2
sigE	CRC_02165	CRD_00081	Transcriptional regulator	--	--	--	--	--	--	--	-1.6
patN	CRC_03501	CRD_01874	Heterocyst differentiation	--	--	--	--	1.7	--	--	--
asr1734	CRC_01342	CRD_02898	Heterocyst development inhibitor	--	--	--	--	--	--	-2.1	-2.2
NA	CRC_01573	CRD_02102	Heterocyst polysaccharide formation	--	--	--	--	1.8	--	--	--
NA	CRC_01567	CRD_02096	Heterocyst polysaccharide formation	--	--	--	--	1.7	--	--	--
nblA	CRC_02186	CRD_00293	phycobilisome degradation	8.8	--	--	--	8.1	2.0	--	-2.7
glgA	CRC_02999	CRD_00097	Glycogen/starch synthases, ADP-glucose type	--	--	--	1.9	--	--	--	2.0
agp	CRC_00827	CRD_00238	Glucose-1-phosphate adenylyltransferase	-1.8	--	--	--	-1.7	1.6	--	1.7
gap1	CRC_00851	CRD_00555	Glyceraldehyde-3-phosphate dehydrogenase	--	--	--	--	1.7	1.6	--	-1.5
tal	CRC_00849	CRD_00556	Transaldolase AB	1.9	--	--	--	1.7	1.7	--	-1.6
rbcS	CRC_02521	CRD_00636	Ribulose bisphosphate carboxylase	-2.0	--	--	--	-2.0	--	--	2.2
NA	CRC_02522	CRD_00637	Chaperonin-like RbcX	-2.6	--	--	--	-1.7	--	--	1.8
rbcL	CRC_02523	CRD_00638	Ribulose bisphosphate carboxylase	-2.7	--	--	1.7	-2.3	--	--	1.6
fbp	CRC_01685	CRD_02186	Fructose 1,6-bisphosphatase	--	--	--	--	1.9	--	--	-1.5
zwf	CRC_01684	CRD_02187	Glucose-6-phosphate dehydrogenase	2.5	--	--	--	3.0	2.0	--	--
pfkA	CRC_01677	CRD_02193	Phosphofructokinase	1.7	2.3	--	2.6	--	--	--	--
icd	CRC_01085	CRD_02610	Isocitrate dehydrogenase	--	--	--	--	1.6	--	--	1.6

-- Gene not differentially regulated
----- Absent gene

Likewise, the largest differences in the transcriptional response between CS-505 and D9 were detected under the urea treatment in which the *nir* operon, repressed in both strains, was down-regulated approximately five-fold in CS-505, whereas in D9 the transcript levels decreased by 37-fold. Furthermore, other N-control genes were down-regulated in D9 but did not show any changes in CS-505: *amt1, amt2, nhlABE, nrrA, glnB, nblA, urtA* and *nirA,* two hypothetical proteins, and genes from the sugar catabolic pathways: *gap1* and *tal* (Table 2.4.2). The level of inhibition of all these genes resembled the inhibition caused by ammonium. We did not see, however, any changes in the expression levels of the urease genes.

In contrast to all other N-regimes tested, N-deprivation caused the expression of the urease genes in D9, but not of the *nir* operon as in CS-505 (Table 2.4.2). The induction of the urease genes only under N-deprivation may reflect an increase in the intracellular levels of urea. Urea is a secondary metabolite can be produced by degradation of arginine. Three enzymes from the arginase pathway were induced under N-deprivation; the carbamoyl phosphate synthase, the gene *argG*, for the argininosuccinate synthase and *argD*, encoding for the ornithine transaminase enzyme were all up-regulated under N-deprivation, although this occurred in both strains and the transcript levels of arginase, the direct enzyme that releases urea from the cleavage of arginine, did not change.

Clear differences reflecting a different environmental acclimation pattern of CS-505 and D9 to N-availability were also observed under nitrate growth. In CS-505, nitrate and urea uptake were induced by activating (FC~2-3) the transporter genes *nrtABC* and *urtADE*, and the glutamine synthetase gene *glnA,* whereas in D9 the protein PII gene, *glnB*, *amt2*, and the nitrite reductase *nirA* were induced (Table 2.4.2).

Heterocyst and N_2 fixation

Heterocysts were developed in the initial culture of CS-505 growing on nitrate; after 24 h of replacement of the N-source heterocysts were still observed in all N-regimes. However, from all combined N sources, only ammonium triggered the repression of the N_2 fixation genes (*nif*), as well as of some heterocyst development-related genes, such as the sigma factor *sigG,* and *hglA and abp3,* both involved in the biosynthesis of the glycolipid layer (Table 2.4.2). The expression of the *nif* genes did not change after 24 h of N-deprivation, but the small subunit (*hupS*) of the uptake hydrogenase (specific for diazotrophic cyanobacteria) was induced together with a few heterocyst development genes. One of them *hetP* is a heterocyst early response gene, two, *hglA*

and *hglB* (or *hetM*), are involved in the maturation of the heterocyst, specifically in the heterocyst glycolipid biosynthesis and the last is the histone-like protein *hanA,* whose time of expression has not been determined.

Interestingly, we found several genes that are related with the process of heterocyst differentiation among the group of core genes regulated under N-deprivation in D9, a non-diazotrophic, non-heterocystous cyanobacterium. These included *hetR,* the master gene for heterocyst development, and *patN,* involved in heterocyst pattern formation, together with two genes from heterocyst polysaccharide biosynthesis (CRD_02102, CRD_02096), which were all induced. The *hetZ* and *patU3* genes both related to heterocyst differentiation, and *ntcA,* which is critically important to N control in cyanobacteria were found to be repressed (Table 2.4.2). The repression of *ntcA* presumably reflects that D9 cells were under extreme stress conditions, as inferred by the high number of genes regulated in this treatment.

A gene similar to *asr1734* from *Anabaena* sp. PCC 7120 (CRD_02898), proposed as inhibitor of heterocyst development, was found to be down-regulated in D9 under ammonium and urea. The expression of *devH*, related to polysaccharide biosynthesis, was induced in both strains under growth on nitrate and under N-deprivation, and in D9 also under urea. On the contrary, the sigma factor *sigG* was repressed in CS-505 under nitrate, N-deprivation, and urea, and in D9 under ammonium and urea growth (Table 2.4.2).

Energy and growth

Diazotrophic growth physiologically separates CS-505 from D9, and therefore these strains should respond differently to N-deprivation. In fact, 865 genes (28% of total) were regulated in D9 at 24 h after N-deprivation, in comparison to only 495 (12%) in the CS-505 genome (Table 2.4.1). Inhibition of growth, photosynthetic and general energy metabolism was part of the common response of both strains as determined by the repression of ribosomal genes, the cell division-related gene *minE*, phycobilisome synthesis (*acpDE*) and assembly genes (*cpcABCDEF*, *cpcG1G2G4*) along with the strong induction of the *nblA* gene for phycobilisome degradation, inorganic carbon acquisition via CO_2 (*ndh3*) or HCO_3^- uptake (*cmp*) and carbon storage genes. Genes associated with the use of organic carbon reserves were induced (*zwf, tal*) (Table 2.4.2). Nitrate and urea positively induced the growth of CS-505 and D9, which was reflected in the induction of the energy metabolism genes in the strains, such as carbon acquisition mechanisms, photosynthetic electron transport and ATP synthesis genes (Table

2.4.S1). A more inductive effect of urea was, however, observed in D9 energy metabolism, in which in addition to photosynthetic ATP synthesis, the Photosystem II (*psbU, psbK, psbN*), antennae complex (*cpcCEF, cpcG4, acpACE*) and Rubisco genes were induced.

Although part of the core, 534 genes were expressed in D9 under N-deprivation that were not expressed in CS-505. The expression pattern of these genes together with that of the 129 unique genes in D9 should indicate general stress response to N-deprivation, but most fold change (FC) values were within the range of 1.5 to 2 indicating only weak induction/repression responses. As indication of growth arrestment in D9, we found a repression of additional ribosomal proteins and other genes involved in energy metabolism, metabolite transport, protein synthesis and stability, DNA repair, polysaccharide biosynthesis and proteins of unknown function.

Ammonium triggered in D9 the expression a large number of genes compared to CS-505 (26 vs 5, respectively). Among these were genes known to be induced as a response to oxidative damage, such as *sufS* and *sufD* (a conserved hypothetical protein downstream *sufD*), involved in synthesis of Fe-S centers, a ferredoxin, and the bidirectional hydrogenase genes *hoxE* and *hoxF*. The response also included the genes encoding for the chaperonin GroES, a peptidase from the FtsH family, and an ATPase, with chaperone or post-transcriptional activity.

A gene cluster comprised of six CDS was found to be repressed under all four N-regimes in CS-505 and D9 and might have an important role in N-metabolism in the strains. Each CDS shares high similarity with its homologue in the other strain, but low similarity when compared to other cyanobacteria. Within the cluster there are a putative arginase and two proteins containing CobW/HypB/UreG nucleotide-binding domains. This domain is found in enzymes that bind nickel, such as the accessory hydrogenase protein HypB and ureaseG. The other two CDS encode for hypothetical proteins with similarity to ABC transporters. The gene cluster was down-regulated with similar intensities in both strains, although the highest repression was observed under urea, sequentially decreasing with ammonium and nitrate and N-deprivation in CS-505. In contrast, in D9 the cluster was equally repressed in nitrate and N-deprivation.

Putative NtcA binding sites within the *cyr* and *sxt* clusters

An arrest in PST production was observed under both ammonium-replete and N-deprivation conditions. Based on the assumption of gene proximity (i.e. intergenic region <50 nt) and the lack of regulatory boxes inside the intergenic regions, we could allocate 5 TU, formed by more than one contiguous gene within the *sxt* gene cluster: *sxtQRST*; *sxtLsxtSUL*; *sxtIJ*, *sxtFGH,* and *sxtABC* (Figure 2.4.S2A). The *sxt* gene cluster is flanked by a hypothetical protein (CRD_02162), tRNAThr and tRNATyr and by *glmU*, coding for a bifunctional protein, *acpF,* an allophycocyanin gene, and the glutamine synthetase gene *glnA*, known to be a target for NtcA regulation (Luque *et al*., 1994). The putative NtcA binding region for *glnA* in *R. brooki* D9 is located 128 nt upstream from the start of the open reading frame (ORF) and 92 nt downstream from the -10 box (Figure 2.4.S2B). Two additional putative NtcA binding regions were found inside the *sxt* gene cluster, upstream the *sxtM* gene (Figure 2.4.S2C) and *sxtABC* (Figure 2.4.S2D); both regions had one mismatch with the consensus motif ($GTAN_8TAC$), but they were located upstream from the -10 box with the conserved motif TAN_3T, important for NtcA binding (Herrero *et al*., 2004). Interestingly, we found four other putative NtcA binding regions that completely matched the canonical ntcA motif; however, these regions were located inside the sequence of the *sxtA*, *sxtP*, *sxtQ*, and *sxtS* genes, the two latter forming part of one transcriptional unit (Figure 2.4.S2E).

The cylindrospermopsin gene cluster in CS-505 is located within the maturation hydrogenase *hypABDCEF* gene cluster (**Publication II**). The *hyp* genes are known to be N regulated in *Nostoc* sp. strain PCC 73102 (Hansel *et al*., 2001). We could allocate only two transcriptional units (TUs) comprising more than one gene: *cyrDFGI* and *cyrABE* (Figure 2.4.S3A). We found no NtcA binding sites for the first TU. For *cyrABE*, one NtcA binding site completely matching the canonical motif is located 552 nt upstream of the ORF start of *cyrA*. Although far away from the ORF, two possible -10 boxes with the conserved motif were located downstream of this NtcA site (Figure 2.4.S3A). We found four NtcA binding boxes in the intergenic regions of the *cyr* gene cluster and flanking regions. As in *Nostoc* sp. PCC 73102, two of the NtcA sites have one mismatch with the canonical motif and are located 290 and 247 nt upstream of *hypF*. A -10 box with the conserved motif TAN_3T is present although further apart from the NtcA sites (208 and 251 nt) (Figure 2.4.S3B). Similarly, the third NtcA binding site has one mismatch with the canonical motif, and is located upstream *cyrJ,* 464 nt apart from the predicted -10 box (Figure 2.4.S3C). The fourth NtcA binding motif has a perfect match with the

consensus motif and is located 74 nt upstream of the *cyrK* gene, coding for the putative transporter of CYN. For this NtcA binding site, the predicted -10 box is located only 40 nt downstream (Figure 2.4.S3D). The last NtcA binding box is located within the coding region of *cyrJ* and far away from the ORF start of the *cyrABE* TU (Figure 2.4.3E).

qPCR quantification of genes of the *cyr* and *sxt* clusters

Only a few genes from the *cyr* and *sxt* gene clusters showed modified expression in our microarray data. Five genes from the *sxt* gene cluster were all down-regulated in the –N treatment (*sxtS, sxtJ, sxtE, sxtC* and *sxtD*) and two in the nitrate treatment (*sxtS, sxtJ*). Urea was the only treatment where *sxt* genes were induced (*sxtP, sxtT, sxtU* and *sxtE*). The *cyrB* and *cyrE* genes but not *cyrA*, which is part of the same TU (Figure 2.4.S3) were down-regulated only under N-deprivation.

As qPCR targets, the selected candidate genes for the biosynthesis of the toxin precursor molecules (STX or CYN), tailoring reactions, toxin transport, and predicted functions are shown in Table 2.4.3. The expression of four *cyr* and eight *sxt* genes in cultures grown in nitrate, urea, or ammonium, were compared with N-deprivation, considering nitrate-dependent growth as a control under which CYN and PST are constitutively synthesized. N-deprivation results for qPCR in D9 were not obtained because the filaments did not grow under this condition and produced only poor quality RNA after 24 h.

Table 2.4.3 Genes used in qPCR analysis and their predicted functions in toxin biosynthesis.

Gene	Predicted function	Reference
ntcA	Transcriptional regulator in nitrogen metabolism	(Vega-Palas *et al*., 1990)
cyrB	NRPS/PKS, second step in CYN biosynthesis	(Mihali *et al*., 2008)
cyrI	Hydroxylation of C-7 in doCYN to form CYN	(Mihali *et al*., 2008)
cyrJ	Sulfotransferase	(Mihali *et al*., 2008)
cyrK	Transport of CYN/doCYN	(Mihali *et al*., 2008)
sxtSUL	Sulfotransferase rendering GTX2/3	**Publication III**
sxtDIOX	Hydroxylation of STX prior to GTX2/3 formation	**Publication III**
sxtI	Carbamoyltransferase	(Kellmann *et al*., 2008a)
sxtO	Donor of sulfate group	(Kellmann *et al*., 2008a)
sxtA	ACP/aminotransferase, first step in STX biosynthesis	(Kellmann *et al*., 2008a)
sxtU	Reduction of C-1, eighth step in STX biosynthesis	(Kellmann *et al*., 2008a)
sxtM	Transport of PST	(Kellmann *et al*., 2008a)
sxtF	Transport of PST	(Kellmann *et al*., 2008a)

The expression of the N-regulator gene *ntcA* was determined as a proxy control for N-regulation. With ammonium, the levels of *ntcA* showed a significant decrease after 72 h in CS-505 (Figure 2.4.6A). In D9, the mRNA levels of *ntcA* continuously decreased until 72 h in cells grown on nitrate and ammonium, as we expected for the negative regulation of *ntcA* under N-replete conditions (Figure 2.4.6B). In the urea treatment, *ntcA* showed a two-fold regulation at 72 h in CS-505, whereas in D9 expression remained unchanged. Surprisingly, the *ntcA* transcript levels increased 3-fold only after 48 h of N-deprivation, yet the C:N ratios were high after the first 24 h.

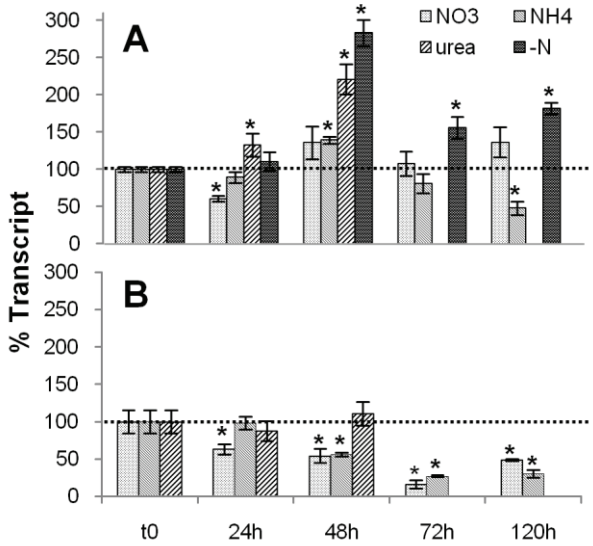


Figure 2.4.6 Comparison of the *ntcA* gene expression of CS-505 (A) and D9 (B). Cultures were grown under nitrate (white bars with black dots), ammonium (left thin stripes), urea (right thick stripes), and N-deprivation (black bars with white dots). Error bars indicate ±SD of the mean (n = 3). Values significantly different (t-test p-value <0.05) from t = 0 are marked with a symbol (*).

Cylindrospermopsin and saxitoxin biosynthetic genes were constitutively expressed under all treatments tested (Fig 2.4.S4). Toxin production was slower during the 120 h of the experiment in the –N treatment. In this treatment, the *cyrB* transcript level dropped 65% after the first 24 h; after 48 h the levels increased but remained below the initial values until the end of the experiment. The *cyrI* and *cyrJ* transcripts decreased at a lesser although significant extent (35%) at 24 h (t-test $p < 0.05$), but they returned to the initial levels and significantly decreased at 120h whereas *cyrK* transcripts did not change. In N-replete treatments, the gene expression response was rather surprising. After 120 h, the transcript levels did not rise above the control, with the exception of *cyrB* at 48h in ammonium, and at 72 h in nitrate, and *cyrJ* at 48 h in urea. Gene expression in the ammonium and nitrate treatments showed a similar response, namely a

significant down-regulation between 72-120 h. With urea, *cyrI* and *cyrK* were repressed at 24 and 48 h. The eight STX biosynthetic genes assayed responded similarly and comparably to the *cyr* genes; a significant down-regulation was observed near the end of the time series (Figure 2.4.S5).

2.4.4 Discussion

This work represents the first integrated assessment of transcriptional and toxin physiological responses of filamentous toxigenic cyanobacteria grown in alternative N regimes. Our comparison, also supported by a previous phylogenetic analysis (**Publication I**, **II**) suggests that the closely related strains *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9 evolved from a common ancestor possessing a diazotrophic and heterocyst-forming phenotype. In this scenario, D9 would have evolved by losing genes necessary for diazotrophic growth and formation of a functional heterocyst, but in compensation gained or kept genes to assist in more efficient metabolism of alternative N sources (**Publication II**).

Growth, transcriptomic and metabolic responses to alternative nitrogen regimes

Both *C. raciborskii* CS-505 and *R. brookii* D9 demonstrated the capacity to grow on a variety of fixed N-sources, as is typical for cyanobacteria. Independent of the N regime, cells committed to cell division upon transfer from nitrate stock medium completed division, but under N-deprivation further growth was arrested in both CS-505 and D9. The effect of N-deprivation in CS-505 is in contrast with that observed in *Anabaena* sp. PCC 7120, which acclimates more quickly and grows equally well in nitrate-replete and in N-deprived medium (Ehira & Ohmori, 2006b). Although ammonium is considered as the preferred N source for cyanobacteria (Herrero *et al*., 2001), CS-505 grew less rapidly on ammonium than on urea or nitrate and growth of D9 was also negatively affected by this compound. Toxicity triggered by ammonia is common among plants and oxygenic photosynthetic microorganisms, since NH_3 produced by alkaline conversion from NH_4^+ inhibits photosynthesis probably by uncoupling photophosphorylation (Krogmann *et al.*, 1959, Azov & Goldman, 1982). The genes induced in D9 under ammonium were involved in responses to oxidative damage (*suf* genes). In the case of the *hox* genes, the dissipation of photosynthetic electron transport is one of the possible roles assigned to the bidirectional hydrogenase (Appel *et al*., 2000). These genes could be acting in concert to protect D9 from photodamage. Alternatively, we found a peptidase induced in D9, with similarity to the

peptidase FtsH2 involved in photodamage protection in *Synechocystis* sp. PCC 6803 (Drath *et al*., 2008). There is more than one copy of these peptidases in the D9 genome and we cannot therefore assign functionality without further studies.

Repeat sequences are stretches of DNA with coding potential that are either present as single sequences or arranged in tandem. When compared to the CS-505 genome, the D9 genome has a low number of repeated elements and transposases (53 repeats and 9 transposases versus 406 and 77, respectively) (**Publication II**). The differential regulation (mostly repression) of repeats in CS-505 grown under all N-regimes indicates that these repeats were expressed during exponential growth and therefore they may be critical during this period of maximum fitness. Since 505 and D9 are phylogenetically closely related, the expression of repeat sequences might be an important factor contributing to phenotypic differentiation, although we cannot exclude the participation of other post-transcriptional regulation processes acting jointly.

Repeats are also found among other cyanobacteria (Frangeul *et al*., 2008), but no functions have been assigned. In bacteria, repeated sequences are involved in genomic rearrangements promoting genome plasticity (Aras *et al*., 2003). In *Mycobacterium leprae,* whose genome includes 2% repeated DNA sequences (Cole *et al*., 2001); microarray studies recently determined that repeats are actively transcribed during host infection. Since *M. leprae* has a low percentage of protein-coding regions, the function of the repeats, together with pseudogenes and non-coding regions, is proposed to be an efficient way to regulate the small pool of coding sequences (Akama *et al*., 2009). In our cyanobacterial comparison, a higher number of repeated sequences (also expressed) was found in CS-505 - the strain with the larger genome – than in D9. The expression of repeats is thus unlikely to be related to tight regulation of small genomes. Non-coding RNAs are responsible for gene regulation of *Prochlorococcus* MED4, whose genome is only 1.66 Mb (Steglich *et al*., 2008). Unfortunately, our arrays did not include the sequences of small RNAs, although the genomes of both CS-505 and D9 contain some, such as Yfr1. Additional studies to detect and describe the expression of non-coding RNAs in CS-505 and D9 will aid in understanding gene regulation of these strains.

The large number of genes expressed after N-deprivation in D9 probably reflects the fact that for those filaments that completed cell division, after exhaustion of the internal N, death is forthcoming. What we observe is therefore a gradual shutdown in general metabolism. On the other hand, as is typical for *C. raciborskii* (Shafik *et al*., 2003) CS-505 should compensate by N_2

fixation when N becomes limiting. In that case, in CS-505 the transcriptional response should not be as pronounced as in D9. Indeed, the number of genes regulated in CS-505 under N-deprivation (12% of which 6% are up-regulated) is similar to the responses of *Anabaena*, where 10% of its genes are up-regulated (Ehira *et al.,* 2003). The fact that there is always a low but constant N_2 fixation in filaments grown on nitrate (**Publication V)** further supports the proposition of a background N pool in CS-505, even under external N-limiting conditions. Consequently, when CS-505 was shifted to ammonium, it down-regulated the nitrogenase gene cluster and some of the genes necessary for the biosynthesis of heterocyst polysaccharides.

A comparative model of the processes affected in CS-505 and D9 by different N regimes provides a detailed scheme for differential pathway analysis and regulation (Figure 2.4.7). Ammonium, the most reduced form of nitrogen, is directly ready to be incorporated into carbon skeletons via the GS/GOGAT pathway (Luque *et al*., 1994, Herrero *et al*., 2001). As expected for both strains, replacing nitrate with ammonium triggered the repression of genes involved in acquisition and assimilation of alternative N-sources and of general N-metabolism genes. Of the two known glutamine synthase inhibitor factors, IF7 (*gifA*) and IF17 (*gifB*), only *gifA* is encoded in the CS-505 and D9 genomes and up-regulated in both. Although the ammonium-inactivation of glutamine synthetase via IFs has been demonstrated only for *Synechocystis* PCC 6803 (García-Domínguez *et al*., 1999), the induction of *gifA* in CS-505 and D9 suggests that GS is inactivated by the same mechanism.

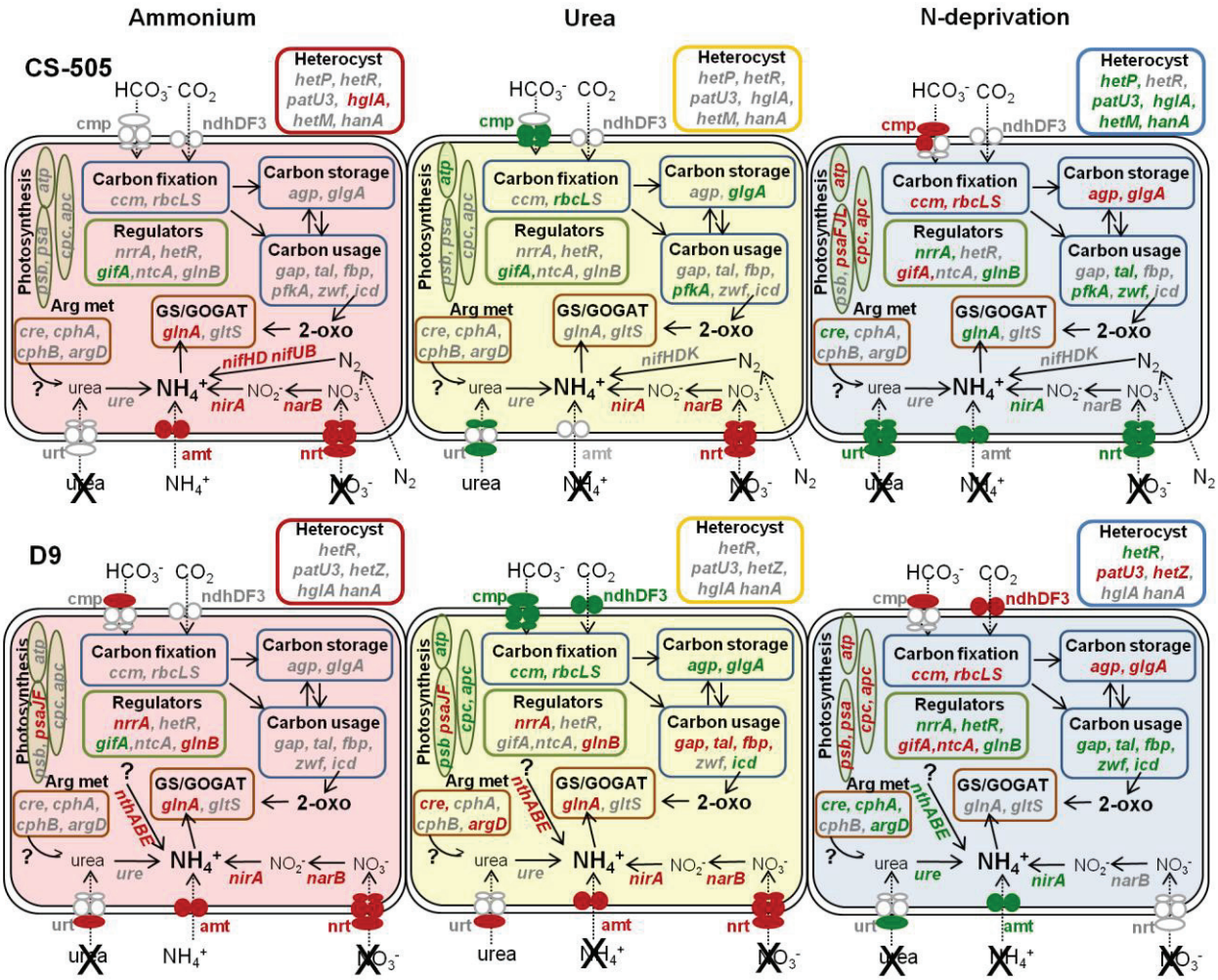


Figure 2.4.7 Schematic model of the main processes affected in CS-505 and D9 by growth under alternative N regimes. A hypothetical cyanobacterial filament is illustrated, in which a single color cell represents the transcriptional profile of CS-505 or D9 under one alternative N-source. Although N_2 fixation occurs in heterocysts in CS-505, the model simplifies the transcriptional profile to a single cell and heterocyst-specific genes are presented in colored boxes on top of each cell. Red cells represent growth under ammonium, yellow under urea and blue under N-deprivation. Different processes are presented by squares and metabolic reactions by full arrows. Patterned lines represent uptake of metabolites, external N sources are marked with a cross to indicate that they were not supplied in the medium. Up-regulated genes are depicted in green and down-regulated in red. Genes that did not show changes in expression are shown in gray. For structural proteins, structures are colored according to the same gene color coding.

Taken as a whole, our results indicate that urea is easily assimilated by D9 because after 24 h the C:N ratios did not change relative to initial conditions, whereas with nitrate, ammonium and N-deprivation the ratios significantly increased (Figure 2.4.2B). Among the most remarkable findings on gene expression in D9 were the transcriptional changes of the genes subject to N-

control when grown on urea. The extent of regulation of some of these genes was comparable to that observed under ammonium-dependent growth, e.g. the repression of the *nir* operon in D9 (Table 2.4.2), although growth was induced with urea but delayed with ammonium. Additional data supporting the preference of D9 for urea comes from the induction of general metabolism, within which photosynthesis, carbon fixation and storage-related genes were induced in D9, but only a few genes were induced in CS-505 (Figure 2.4.7). The expression of the urea transporter (*urtABCDE*) and urease genes (*ure*) showed a different regulation pattern between strains. The *urt* genes are subject to ammonium-repression in cyanobacteria (Valladares *et al*., 2002). In high concentration, urea can diffuse through membranes, but at low concentrations active transport by *urt* transporters is necessary. At the high concentration in our experiments (1mM), uptake of urea was presumably via diffusion. On this N source, the *urtA* gene was repressed in D9, but in contrast, the *urtADE* genes were up-regulated in CS-505. In the later strain, the urea transporters seem to play an active role in N- metabolism, because they were strongly induced under N-deprivation and when cells were grown on nitrate.

High urea uptake also depends on metabolism by urease genes (Valladares *et al*., 2002). These genes appear to be constitutively expressed in CS-505, as reported for other cyanobacteria (Ge *et al.*, 1990, Palinska *et al.*, 2000, Valladares *et al*., 2002). On the other hand, urease genes were induced in D9 under N-deprivation but not under urea-dependent growth. The induction of the *ure* genes might reflect an increase in internal urea levels, but we found no evidence of differences in D9 and CS-505 that could explain an increase in intracellular urea.

One intriguing aspect of the D9 genome is the number of encoded genes whose functions can be related to heterocyst differentiation in *Anabena* sp. PCC 7120 (e.g., *nrrA*, *hetR, hetZ, patU3*). Our study showed that there are striking differences between CS-505 and D9 in N-regulation of these genes, along with genes involved in N control in cyanobacteria (Table 2.4.2, Figure 2.4.7). For example, under N-deprivation, the *glnB* gene encoding for the N-regulatory protein P-II was up-regulated in both strains (as is common in cyanobacteria); for cultures grown on ammonium and urea, however, this gene was down-regulated in D9, whereas in CS-505 there was no change in expression. Protein PII is ubiquitous among cyanobacteria; its function is to coordinate the C:N balance in cells and depends on cellular 2-oxoglutarate levels (Forchhammer & Tandeau de Marsac, 1995). Protein PII is required for ammonium-promoted inhibition of active uptake of nitrate and nitrite (Lee *et al*., 1998). The role of protein PII in heterocyst-forming cyanobacteria is not clearly understood, but in heterocyst-forming *Anabaena*, *glnB* seems to have

an essential role because transcript levels are high when cells are grown on nitrate or ammonium (Paz-Yepes *et al*., 2009). The differences between CS-505 and D9 in regulation of *glnB* are probably related to heterocyst differentiation, whereby the dispensability of PII by D9 under N-replete conditions is contrasted with the PII requirement by CS-505, which maintains both heterocyst differentiation and N_2 fixation. These differences were observed for *nrrA*, encoding for a transcriptional regulator that facilitates heterocyst differentiation. In *Anabaena*, upon N-deprivation, NtcA up-regulates the expression of *nrrA* and then NrrA up-regulates the expression of *hetR,* which leads to heterocyst differentiation (Ehira & Ohmori, 2006a, Ehira & Ohmori, 2006b). The *nrrA* gene is not expressed under growth on nitrate in *Anabaena* (Ehira and Ohmori, 2006a), which is consistent with the lack of repression of this gene in CS-505 grown in ammonium and urea, but contrasts with the repression of *nrrA* in D9 under the same N-regime. The expression of *nrrA* and its involvement in N-metabolism of non-diazotrophic cyanobacteria warrants further investigation to clarify this point.

Other critical examples that reflect the clear differences between CS-505 and D9 and between them and *Anabaena* were observed under N-deprivation. The cascade of gene expression leading to heterocyst differentiation in *Anabaena* starts with the induction of *ntcA, nrrA* and *hetR*. Although N-deprivation induced *nrrA* in both strains, *hetR* was only up-regulated in D9. In *Anabaena*, the expression of *hetR* initiated from one of its four transcription start points (tsp); tsp -271 is the signal that indicates the point of no return in heterocyst differentiation, occurring between 9 and 14 h after induction of differentiation (Rajagopalan & Callahan, 2010). Additionally, after the discovery of *hetR* in non-heterocystous diazotrophic cyanobacteria, such as *Symploca* PCC 8002 and *Trichodesmium erythraeum* (Janson *et al*., 1998), which fix N_2 in contiguous groups of cells called diazocytes (Fredriksson & Bergman, 1997, Berman-Frank *et al.*, 2001), the role of *hetR* as unique for heterocyst differentiation has been reconsidered in favor of a more general role in cell differentiation. In *T. erythraeum* the expression of *hetR* correlates with that of the cell division gene *ftsZ*, both preceding the development of N_2-fixing diazocytes (Sandh *et al*., 2009). The lack of expression changes in CS-505 points to a late regulation of heterocyst differentiation in this strain. In D9, however, the induction of *hetR* cannot be related to the differentiation of N_2-fixing cell types, but it might be an indication of differentiation of other cell types. Akinetes, resistant cells produced under stress conditions in cyanobacteria (Adams & Duggan, 1999), are the only other differentiated cell type besides vegetative cells in D9. In our experiments, D9 cultures were N-starved at 24 h after N-deprivation; therefore the increased levels of *hetR* might be related to akinete formation. Alternatively, *hetR* levels were shown to

gradually increase after N-deprivation in *Arthrospira platensis*, a non-diazotroph unicellular cyanobacterium, and the presence of *hetR* has been described for other filamentous non-diazotrophic cyanobacteria (Zhang *et al*., 2009). We recently showed (**Publication II**) that *hetR*, together with *patU3* and *hetZ*, two genes involved in heterocyst differentiation in *Anabaena* (Zhang *et al*., 2007), are part of the core genes of filamentous cyanobacteria and suggested that these genes are involved in filament formation. Since cell death is manifest in both D9 and CS-505 by filament fragmentation and *patU3* and *hetZ* are down-regulated in D9, regulation of these three genes might be an indication of ongoing cell death in D9.

Our results showed that gene expression responses are rather slow in CS-505 and D9 in comparison to what is known for the closely related genus *Anabaena*, in which *ntcA* levels were high 8 h following N withdrawal (Ehira *et al.*, 2003). Although expression of the gene targets of NtcA regulation in cyanobacteria, specifically of the *nir* and *amt* transporter genes, changed in both our strains within 24 h after replacing the N-source, in CS-505, *ntcA* was not significantly expressed in the -N treatment. Induction of *ntcA* was observed by qPCR only after 48 h, after which the transcript levels remained high. More unusually, *ntcA* was down-regulated in D9 at 24 h. The repression of *ntcA* has been observed in *Anabaena* when exposed to cold stress; however, repression was transitory until cells were acclimated to the cold (Mori *et al.*, 2002). In D9, cells were maintained under constant N-stress for 120 h, which led to cell death. The repression of *ntcA* after 24 h reflects that the cells did not acclimate to N stress; yet the expression of NtcA regulated genes was still maintained after 24 h. This may indicate either that the turnover of NtcA is low in CS-505 and D9, or that there are additional posttranscriptional modifications involved in the regulation of NtcA gene targets.

Nitrogen regulation of STX and CYN biosynthetic genes

Following the above scenario described for NtcA-regulated genes, other gene targets of NtcA should be induced or repressed after 24 h by NtcA activity beyond the basal expression levels. As our search for NtcA binding boxes has shown, some genes of the *sxt* and *cyr* clusters are potential NtcA targets. Based on the distance of the binding box to the start of the open reading frame (ORF), NtcA should act as an activator of the toxin genes (Herrero *et al*., 2004). We showed here, however, that for most of the eight *sxt* and four *cyr* gene transcripts quantified by qPCR, there was no transcriptional regulation for the period studied. In fact, the tendency of

transcripts was towards repression. Microarray data showed that only two CYN biosynthetic genes (*cyrB* and *cyrE*) were slightly down-regulated under N-deprivation, and some of the STX biosynthetic genes were inhibited upon N-deprivation and with nitrate, but induced with urea. Interestingly, these genes encode for hypothetical proteins (*sxtE, sxtC, sxtJ*) or for proteins with putative functions in oxidoreduction reactions (*sxtD, sxtU, sxtS* and *sxtT*) (Kellmann *et al.,* 2008a). The pattern of gene regulation of these *sxt* genes is consistent with the fact that both N-deprivation and urea affect the energetic status of D9 cells by repressing or inducing ATP generation, respectively. These genes are therefore probably involved in general electron transport in addition to PST production.

Neither *sxt* nor *cyr* gene clusters form part of a single transcriptional unit, as revealed by the search of promoters within the gene clusters. In fact, absolute quantification showed that genes belonging to the same cluster did not respond equally, indicating that they were not being co-transcribed. Evidence of horizontal gene transfer (HGT) within the *sxt* gene cluster has been shown by phylogenetic affiliation of *sxt* genes with members of Proteobacteria and Cyanobacteria (Moustafa *et al*., 2009). Although the evolution of the *cyr* genes has not been explored in detail, the transfer of the CYN cluster as a unit has been proposed for *C. raciborskii* CS-505 and AWT205 (Mihali *et al*., 2008, **Publication II**). Both gene clusters are almost identical and contain fragments of transposases that could help in genomic rearrangements or reallocation of gene units. The *cyrG, cyrH* and *cyrN* sequences blast with members of Proteobacteria and Firmicutes.

In *C. raciborskii,* the highest concentrations of CYN have been recorded at late stationary phase in cultures grown at a moderate photon flux density of 140 μmol m^{-2} m^{-1} (Dyble *et al*., 2006) and in the absence of a fixed-N source; the lowest concentrations were found in cultures supplied with ammonium (Saker & Neilan, 2001). Our observations at early exponential phase showed constitutive biosynthesis of cylindrospermopsins in all treatments were N was supplied as a fixed source; however, both CYN and doCYN production was affected by N-deprivation. This highly contrasts with the findings of Saker and Neilan (2001), but we argue that in late exponential growth phase, N-deprived cultures of *C. raciborskii* have restored N_2 fixation and therefore the cells are no longer N-depleted. In addition Saker and Neilan's experiments were carried on batch cultures for which even after ammonium supplementation, nutrients are progressively depleted.

In our cultures, N-deprivation reduced the doCYN:CYN ratio. In CYN biosynthesis, doCYN is proposed to be the precursor that is hydroxylated by the enzyme CyrI (Mihali *et al*., 2008) to form CYN. One possible explanation for the increased CYN:doCYN ratio would be an accelerated rate of conversion from doCYN to CYN, but in this case we would have observed a higher toxin production rate. Since CYN and doCYN production was low, the most probable explanation is that the rate of hydroxylation from doCYN to CYN was maintained, while the input of the precursor was reduced. Although our qPCR and microarray data showed a slight down-regulation of some CYN biosynthesis genes, *cyrB* and *cyrE* were affected and both are involved in the synthesis of the precursor doCYN (Mihali *et al*., 2008).

It is not known whether STX is synthesized by protein complexes with low turnover rates. Two groups (Pomati *et al.*, 2004, Soto-Liebe *et al*., 2008) independently studied the effect of the protein synthesis inhibitor chloramphenicol (CAM) on PST biosynthesis in *Cylindrospermopsis raciborskii* T3 and *Raphidiopsis brookii* D9 (formerly classified as *C. raciborskii* D9), respectively, but their results were contradictory. Whereas Pomati *et al*. (2004) suggested that biosynthetic PST enzymes have a high turnover rate (based on 76% inhibition of STX with CAM after 24 h), Soto-Liebe *et al*. (2008) found an early accumulation of STX (but not of GTX2/3), but they could not measure intracellular toxins after 24 h because CAM produced total cell lysis. The decrease of STX in D9 cultures grown on ammonium (down to the analytical limit of detection) and under N-deprivation points towards a high turnover rate of STX biosynthetic enzymes; we showed, however, that STX biosynthetic genes were constitutively transcribed. This suggests that STX biosynthesis is not regulated at the transcriptomic level at early stages of growth. The lack of negative feedback tends to indicate a central role of PSTs, functioning more akin to primary metabolites in cyanobacterial physiology, than as classic inducible/repressible secondary metabolites as has often been assumed. In D9 grown with ammonium or under N-deprivation, a similar effect was observed in which at low growth rates, the stable ratio of STX:GTX2/3 found in nitrate-rich medium, was shifted lower. The correlation for the –N treatment in D9 demonstrates that in this regime the cells divided without duplicating the toxin content, hence there was a slowdown of toxin production. This was not the case in the ammonium treatment, where there was no correlation between these two parameters. The general implication is that the effect of the N-source is not directly related with toxin production in D9, but rather conditions that negatively affect growth also affect toxin production.

Notwithstanding several efforts to identify the ecological and evolutionary roles of cyanobacterial toxins, this aspect remains elusive. Intriguingly, both cylindrospermopsins and PSTs in CS-505 and D9, respectively, are constitutively produced along the growth curve and, to our knowledge, there are no conditions that could arrest toxin production whilst growth conditions are positive. Here we showed that N-deprivation and cytotoxicity caused by ammonia (only in D9), negatively affect STX and CYN biosynthesis in *C. raciborskii* CS-505 and *R. brookii* D9, respectively, but in neither case is there evidence that these toxins are transcriptionally regulated as classic inducible/repressible secondary metabolites in response to environmental stress factors. In fact their production and dynamics and regulatory features are more typical of components of primary or intermediary metabolism, albeit of unknown function.

In summary, the general mechanisms of N control and heterocyst differentiation in cyanobacteria are now widely understood, at least partially thanks to the model cyanobacterium *Anabaena* sp. PCC 7120. Our new findings on the time-frame and regulation of heterocyst development and N_2 fixation in *C. raciborskii* CS-505 significantly adds to interpretation of these mechanisms. This cyanobacterium develops a terminal heterocyst and also has a small genome thereby rendering an alternative model for heterocyst differentiation in a much simpler organism. *Raphidiopsis brookii* D9 represents an important counterpart evolutionary stage for comparison, in that the remnant of heterocyst differentiation genes have been apparently recycled to other cell differentiation stages. It is remarkable that despite the close relationship between CS-505 and D9, genes that are shared in some cases with 100% identity are regulated in opposite directions under the same N-dependent stimuli. Differential regulation between CS-505 and D9 indicates that other cryptic factors are involved in N-dependent metabolism and control, particularly in the early response to N-deprivation in CS-505. Finally, late regulatory responses and the pattern of biosynthetic gene regulation strongly indicate that post-transcriptional mechanisms are involved in toxin biosynthesis and regulation.

2.4.5 Materials and Methods

Experimental culture and harvest

Cylindrospermopsis raciborskii strain CS-505 was isolated in Australia and obtained from the CSIRO culture collection, Hobart, Australia. *Raphidiopsis brookii* D9 (originally classified as

C. raciborskii D9) is a clonal isolate from the Billings freshwater reservoir near Sao Paulo, Brazil. Batch cultures of non-axenic CS-505 and D9 were grown in MLA medium on 2 mM nitrate (Castro *et al*., 2004) in a controlled environment chamber on a 12/12 h light/dark photocycle at a photon flux density of 35 µmol m^{-2} s^{-1} at 25 °C. Stock cultures were harvested for experiments in early exponential growth phase (OD_{750} = 0.2). To change the experimental N regime, 800 ml of culture were filtered through an 8 µm pore-size cellulose ester membrane (Millipore, Eschborn, Germany) and cells were resuspended in twice the volume of new MLA medium containing 2 mM $NaNO_3$ (control medium, nitrate treatment), 2 mM ammonium chloride (ammonium treatment), 1 mM urea (urea treatment), or without a fixed external N-source but with 2 mM NaCl to restore osmotic balance (-N treatment). Experiments were performed for 120 h under a continuous light regime at a photon flux density described above.

Three biological replicates were harvested at time points of 0, 24, 48, 72 and 120 h for determination of biomass, chlorophyll a, C:N ratios, toxins and for RNA isolation. Estimation of biomass was based on dry weight of harvested cells. Chlorophyll *a* was extracted from 1 ml samples in technical replicates and measured according to the ISO method (Lawton *et al*., 1999). For C:N ratios, technical duplicates (4 ml) from each biological replicate were filtered through a prebaked (500 °C, 5 h) GF/F filter (Whatman, Dassel, Germany) and analysed with an Elemental Analyzer (Euro EA, HEKAtech, Wegberg, Germany).

Cellular CYN and PST toxins were harvested from 30 ml of experimental cultures after centrifugation (20 min at 3,220 x g) at room temperature and pellets were frozen at -20 °C until extracted. For the extracellular toxin fraction, 10 ml of supernatant after centrifugation were filtered through a 0.2 µm pore-size nitrocellulose membrane (Carl Roth, Karlsruhe, Germany). To avoid degradation of PSTs, immediately after filtration, the pH of the supernatant was adjusted to 2.5 with HCl and the extract was stored at -20 °C.

Frozen cell pellets and supernatants were lyophilized (Beta I Freeze Dryer, Martin Christ, Osterode am Harz, Germany) at -20 °C and ca. 0.004 mbar vacuum. Intra- and extracellular toxins were extracted in 500 µl and 300 µl of 0.05 M acetic acid, respectively. Samples were disrupted with an ultrasonic cell disruptor (Sonoplus LS6, Bandelin Electronics, Berlin, Germany) for 1 minute; extracellular toxins were transferred to vials and stored at -20 °C until analyzed. Intracellular toxin extracts were centrifuged at 5,000 x g for 10 min at 4 °C and filtered through a 0.45 µm pore-size membrane filter (Millipore, Eschborn, Germany) and stored at -20 °C until analysis.

Analysis of Toxins

PSTs were determined by high-performance liquid chromatography with fluorescence detection (LC-FD) following post-column oxidation as previously described (**Publication III**). The LC-FD analysis was carried out on a LC1100 series liquid chromatograph (Agilent Technologies, Waldbronn, Germany) coupled to a PCX 2500 post-column derivatization system (Pickering Laboratories, Mountain View, CA, USA). The separation of analytes was performed on a 250 × 4.6 mm i.d., 5 μm, Luna C18 reversed-phase column (Phenomenex, Aschaffenburg, Germany) equipped with a Phenomenex SecuriGuard pre-column.

Standard solutions of PSP toxins: saxitoxin (STX), neosaxitoxin (NEO), decarbamoyl saxitoxin (dcSTX), gonyautoxins 1 and 4 (GTX1/GTX4), gonyautoxins 2 and 3 (GTX2/GTX3), decarbamoyl gonyautoxins 2 and 3 (dcGTX2/dcGTX3), and B1, were purchased from the Certified Reference Material Programme of the Institute for Marine Biosciences, National Research Council, Halifax, NS, Canada. Epimeric pairs of compounds (e.g. GTX2/GTX3, dcGTX2/dcGTX3) are reported as single components due to facile epimerization and thus unstable ratios.

Cylindrospermopsin (CYN) and deoxycylindrospermopsin (doCYN) were identified and quantified by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Mass spectral experiments were performed on an ABI-SCIEX-2000 triple quadrupole mass spectrometer (Applied Biosystems, Darmstadt, Germany) coupled to a Model 1100 liquid chromatographic system (Agilent, Waldbronn, Germany).

Separation of CYN and doCYN was performed by reverse- phase chromatography on an analytical column (150 × 3 mm) packed with 3 μm Luna C18 (2) 120 Å (Phenomenex, Aschaffenburg, Germany) and maintained at 20 °C. The flow rate was 0.3 ml min^{-1} and gradient elution was performed with two eluents, where eluent A was water and eluent B was methanol/water (95:5 v/v), both containing 2.0 mM ammonium formate and 50 mM formic acid. Initial conditions were 10 min column equilibration with 10% B, followed by a linear gradient to 90% B in 15 min and isocratic elution until 19 min with 90% B. The system was then returned to initial conditions until 20 min (total run time: 30 min).

Multiple reaction monitoring (MRM) experiments were carried out in positive ion mode by selecting the following transitions (precursor ion > fragment ion): *m/z* 400>194 and *m/z* 400>176 for doCYN and *m/z* 416>194 and *m/z* 416>176 for CYN. Dwell times of 100 ms were

used for each transition. Sample concentrations were calibrated against external standards of CYN and doCYN.

Specific growth and toxin production rates (μ_c and μ_{tox}, respectively) for both PSTs and CYN/doCYN were calculated for each interval in the growth and toxin production curves according to Anderson *et al.* (1990).

RNA isolation

Cells for RNA isolation were collected by filtration of 50 ml of culture through an 8 µm pore-size cellulose ester membrane (Millipore, Eschborn, Germany). Filtered cells were resuspended in 800 µl of RLT lysis buffer (Qiagen, Hilden, Germany) containing β-mercaptoethanol, then transferred to a microcentrifuge tube, flash frozen in liquid nitrogen and stored at -70 °C. Total RNA was extracted with an RNeasy mini-kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Briefly, frozen samples were thawed on ice, and cells were twice disrupted for 2 min with 0.1 mm diameter acid-washed glass beads in a tissue lyser (Qiagen, Hilden, Germany). The supernatant was separated from the glass beads and cell debris by centrifugation (10 min, 16,000 x g, 4 °C). On-column DNA digestion was performed after RNA isolation for 1 h at RT, followed by a final clean-up with a second DNA on-column digestion to remove any DNA. RNA was quantified by spectrophotometry (ND-1000, NanoDrop Technologies, Wilmington, DE, USA) and purity was assessed based upon the ratios A_{260}/A_{230} and A_{260}/A_{280}. If required, samples were further cleaned with a Microcon Elute System (Millipore, Eschborn, Germany). RNA integrity was checked on RNA nano-chips using an Agilent Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany).

Microarray design and data analysis

DNA microarrays were designed for use with the eArray system (Agilent Technologies, Waldbronn, Germany). Two 60-mer oligonucleotides were created for the 3,968 genes of CS-505 and 3,085 D9 genes, such that the genomes of both strains were represented in each 15,000-oligo expression array. Intergenic regions were absent in these arrays, but repeats and transposases were included.

Microarray hybridizations were performed with a two-color microarray-based gene expression analysis kit (Agilent Technologies, Waldbronn, Germany). The procedure was carried out following the manufacturer's instructions with slight modifications. Specifically, to cope with the absence of poly-A tailed mRNA, 200 ng of random nonamers coupled with the sequence of the promoter of the RNA polymerase T7 were added in each reaction (Moreno-Paz & Parro, 2006), instead of the T7 primer supplied for eukaryotic mRNA. 300 ng of total RNA were used as input RNA for cDNA synthesis and cRNA amplification. cRNA from CS-505 was labeled with Cy5 (control samples t = 0 h) and Cy3 (treatment samples t = 24 h). cRNA from D9 was labeled with Cy3 (control samples t = 0 h) and Cy5 (treatment samples t = 24 h). Samples from t = 0 h were hybridized against the samples at t = 24 h.

Microarrays were scanned on an Agilent G2565BA scanner, and raw data were extracted with the Agilent Feature Extraction (FE) Software version 9.1.3.1. Array quality was monitored using the Agilent QC Tool (v1.0) with the metric set GE2_105_Dec08.

Microarray data normalization and statistical analysis

Raw data obtained with the FE software were normalized by LOESS normalization without background subtraction using the limma package (Smyth, 2005) in the R software (http://www.r-project.org/). Normalization with limma was chosen over FE normalization due to less dispersion of the data points in the MA plot obtained with limma (data not shown).

Significance analyses of the normalized data were performed with the Significance Analysis of Microarrays (SAM) tool (Tusher *et al*., 2001) from the TMEV software (Saeed *et al*., 2003). Significantly expressed genes were selected with a q-value criteria of <0.01. The q-value is the fraction of false positives of differentially expressed genes, expected when the threshold has been set to a specific value. In addition, a cutoff of >1.5- fold change was applied to the preselected genes. First class SAM was applied to obtain the overall number of genes whose expression was significant in each treatment at 24 h after the change in N-regime. Gene expression data are supplied as electronic material (CS-505 gene expression data.xls and D9 gene expression data.xls)

qPCR conditions and motif search

The cDNA for qPCR was synthesized from 500 ng total RNA using random hexamers with the Omniscript RT kit (Qiagen, Hilden, Germany). qPCR reactions were performed in 20 µl reaction mixtures composed of 1 µl 10-fold diluted cDNA, forward and reverse primers at a concentration of 100 nM, and 10 µl 2x SYBR Green PCR Master Mix (Applied Biosystems, Darmstadt, Germany). Pure RNA was used to test for the presence of genomic DNA in the RNA samples. Cycle parameters were as follows: initial denaturation at 95 °C (10 min), followed by 40 cycles of 95 °C (15 s) and 59 °C (1 min). Finally, a product-primer dissociation step was utilized to verify formation of a single unique product/primer dimerization.

All qPCR primers were designed with Primer Express 3.0 software (Applied Biosystems, Darmstadt, Germany) and synthesised by Eurofins MWG Operon (Ebersberg, Germany). Samples were run in biological triplicates, and mean values were taken among these replicates, including standard deviation. For each primer pair, a standard curve was established by 10-fold dilutions of a PCR template, spanning concentration differences from 100 pg to 1 fg. Linear regression analysis between PCR product concentration and the cycle number (Ct-value) was used to determine primer efficiency. qPCR amplicon sizes, primer sequences and efficiencies are shown in Table 2.4.S2.

The Motif Search tool from the Nano+Bio-Center at the University of Kaiserslautern (http://nbc11.biologie.uni-kl.de) was employed to search for putative NtcA-binding domains in the intergenic regions outside and within the *cyr* (48.5 Kb) and *sxt* (34 Kb) gene clusters. The NtcA-binding domains have a canonical sequence defined as $GTAN_8TAC$, and they are separated by approximately 22 nucleotides from a -10 box with the consensus sequence TAN_3T (Herrero *et al*., 2004). However, differences in the canonical structure and in the length of separation between the promoter and binding site have been described and probed for genes such as *petH*, *nifH* and *cphA1* (Herrero *et al*., 2004). Following this criteria, we set the mismatches to 1, to allow variation in the 6 conserved positions of the NtcA-binding site. The results were further filtered according to the presence of a -10 box with the conserved motif TAN_3T for NtcA-biding regions.

2.4.6 Supplementary material

Table 2.4.S1 Microarray expression values of the core genes shared between *C. raciborskii* CS-505 and *R. brookii* D9 (supplied electronically).

Table 2.4.S2 qPCR primer sequences, efficiencies and amplicon sizes.

Gene	Amplicon size (nt)	q PCR primer sequence 5'-3'	Slope	Efficiency (%)
ntcA	60	qD9ntcAF TCGCTGAAGCAATAGGATCCA qD9ntcAR TCCCGCAAGTCTCCTAGCAA	-3.33	100.3
	60	q505ntcAF TTTTACTGCGGTGGAATTGCT q505ntcAR TTCCTTGAGGGCCTGCTCTAC	-3.41	102.7
sxtSUL	79	qsxt14F CTGAAGTAGCGGCTCAACAGATAA qsxt14R CAGGAGGATTGCGAAGCATAA	-3.67	110.5
sxtDIOX	99	qsxt15F CGATGGGAGAAATTGCGAAT qsxt15R ATGTCAGGGTGAGCTGGGATA	-3.76	113.1
sxtO	69	qsxtOF TTGGGTGAGGTTGCCAAACT qsxtOR CGGTCATTCCTGTAGGGTGAGA	-3.53	106.3
sxtU	69	qsxtUF GCGATCGCCGCAAGAC qsxtUR CTTGACCACCACTGGCTTCA	-3.31	99.7
sxtA	59	qACPR GGGAGAGCGAGCCTTGAAT qACPR ATGCGGGACAACATAGGAGTGT	-3.30	99.0
sxtI	69	qsxtIF TTGTTGCTGCAGCTCAGGAA qsxtIR ATCGCTCCAGTCGGAAACC	-3.29	99.1
sxtM	100	qsxtMF GGCTAGCAACGGCCTTGTC qsxtMR TGCCAGTAACACATTGCTTTGTT	-3.16	95.2
sxtF	60	qsxtFF GCCCATGATATTGGCTTCCA qsxtFR TGCCGACTCCGAGTGGTATAAC	-3.47	104.5
cyrB	99	qcyrBF ACTGCACAGACTGCGATTTCA qcyrBR TGCGCCCCATGTTAATGTATT	-3.48	104.8
cyrI	94	qcyrIF TCCCGGTCATCCATCAGTAAG qcyrIR ATGTGGGTGCGTCTTCTTGATA	-3.65	109.9
cyrJ	60	qcyrJF TGAACGATCCGCGGAGAA qcyrJR TCGGTAAACTCAACCCCTACAAC	-3.32	100.0
cyrK	99	qcyrKF AGATGACCGAGAGGGCGTACT qcyrKR CACCCCTAACGGGTACTGTAACA	-3.97	119.5

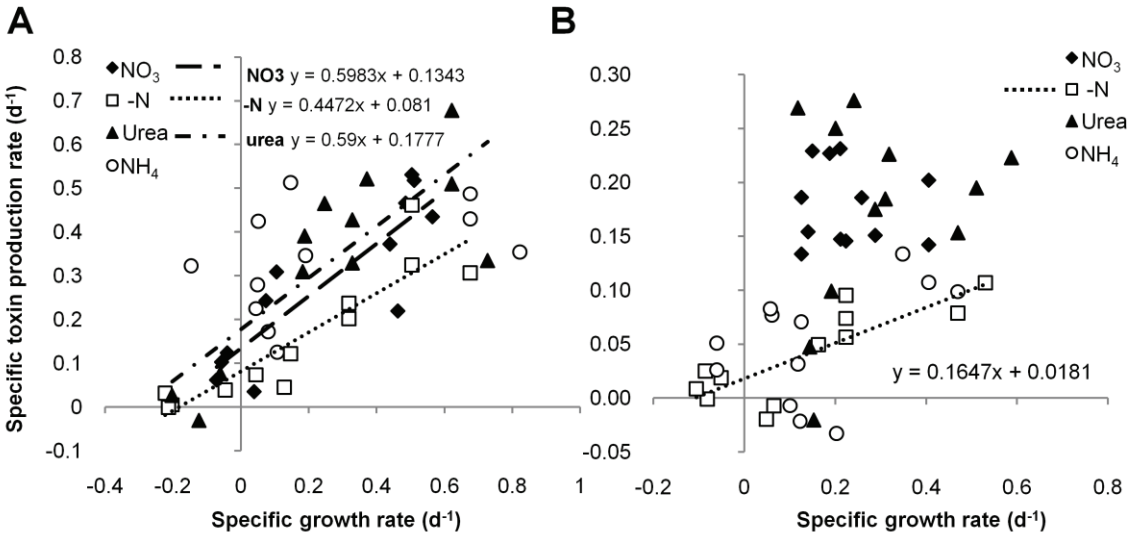


Figure 2.4.S1 Specific toxin production rate as function of specific growth rate in the four N- regimes for CS-505 (A) and D9 (B).

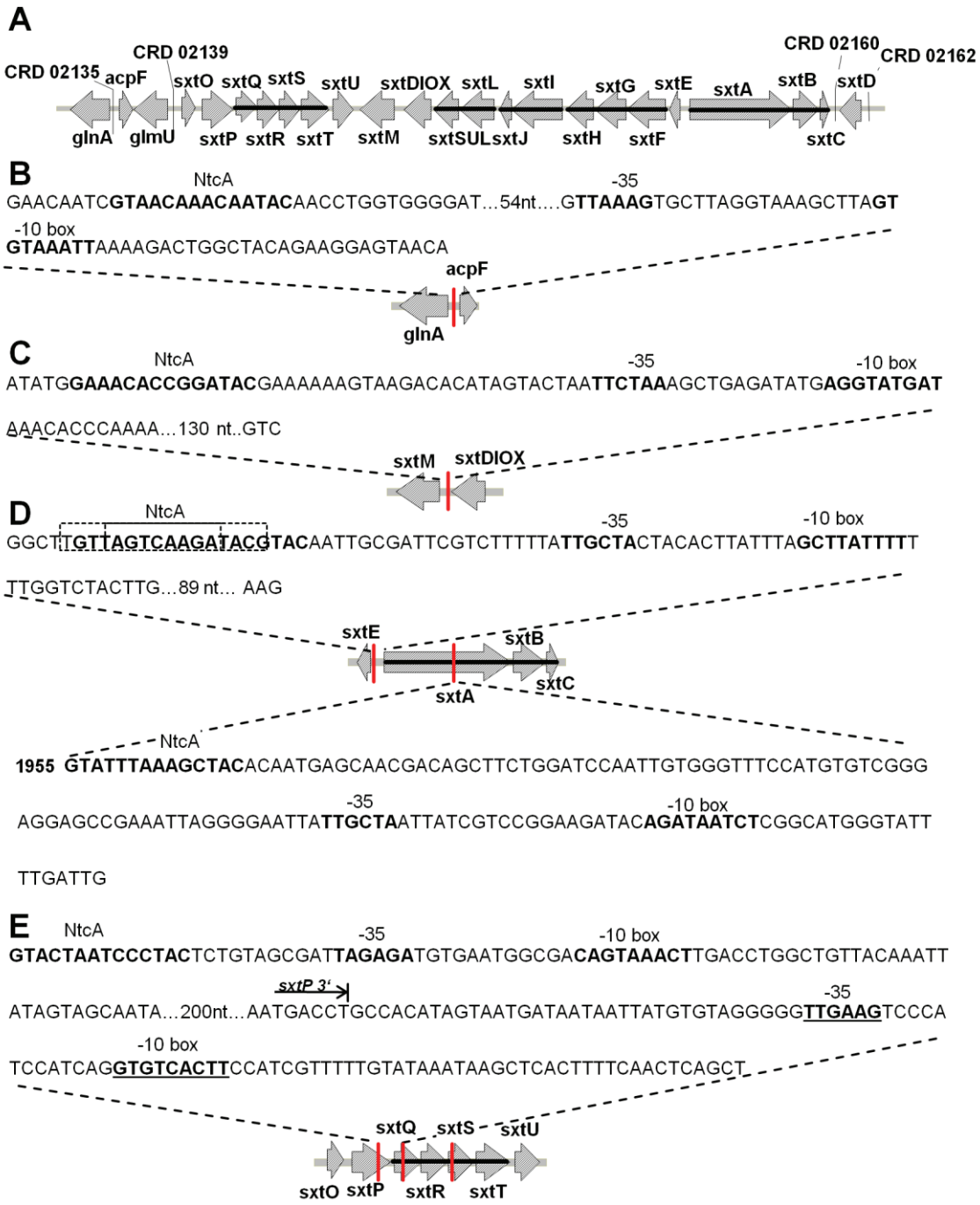


Figure 2.4.S2 In silico analysis of transcriptional units (TU), NtcA binding boxes and promoters ouside and within the *sxt* gene cluster. In silico analysis of transcriptional units (TU), NtcA binding boxes and promoters ouside and within the *sxt* gene cluster. Genes forming part of TUs are marked by a black line them. NtcA binding boxes are shown as a thick red line within the cluster and enclosed in rectangles in the sequences. -10 and -35 regulatory boxes are shown only for the -10 box with the conserved motif TAN3T. In the presence of two possible -10 and -35 boxes within the same intergenic region, the second sequence appear underlined. The reverse complement strand is shown for the NtcA binding region of *glnA* and *sxtM.* The direct strand is shown for NtcA binding region of *sxtABC.*

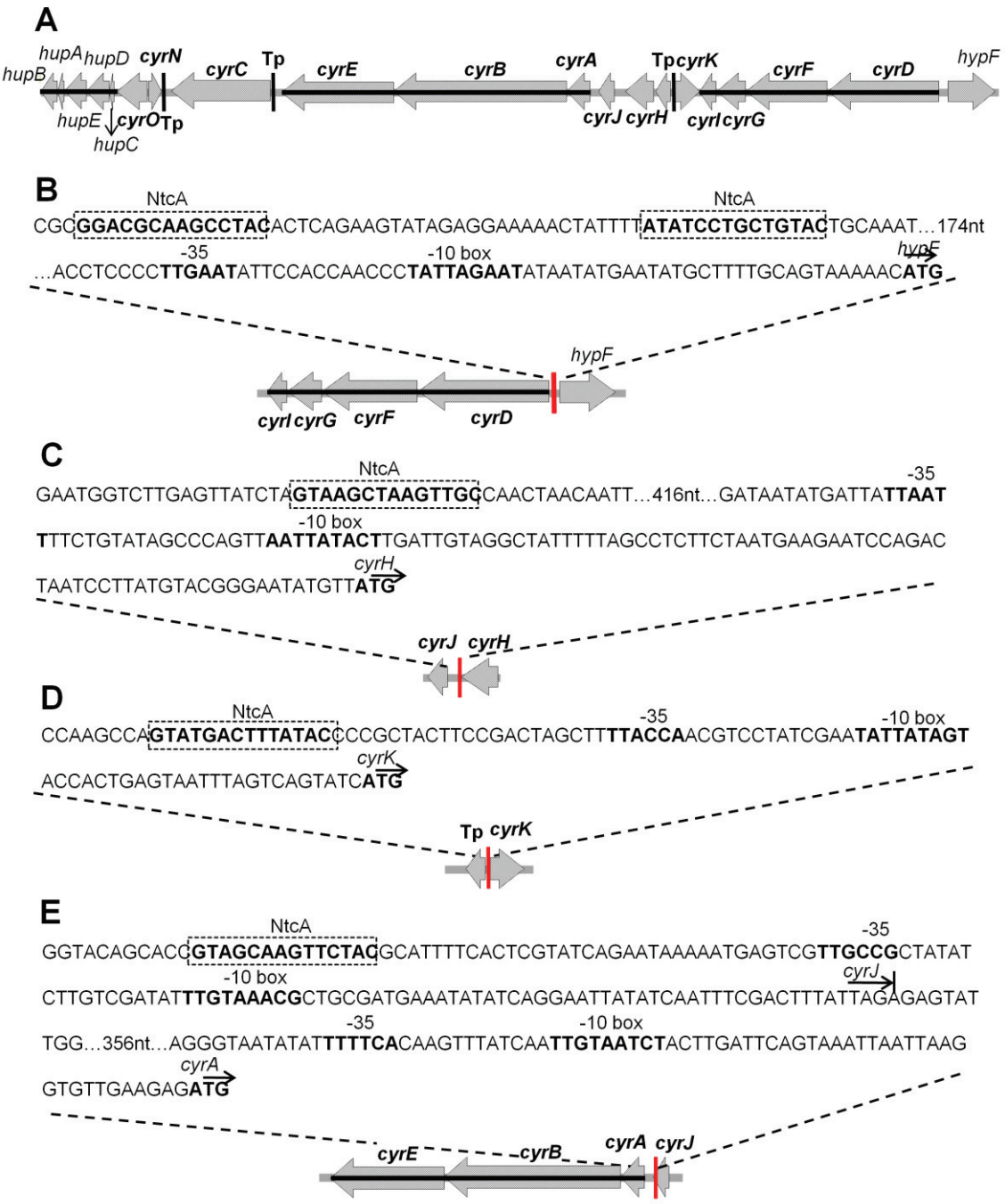


Figure 2.4.S3 In silico analysis of transcriptional units (TU), NtcA binding boxes and promoters ouside and within the *cyr* gene cluster. In silico analysis of transcriptional units (TU), NtcA binding boxes and promoters ouside and within the *cyr* gene cluster. Genes forming part of TUs are marked by a black line across them. NtcA binding boxes are shown as a thick red line within the cluster and enclosed in rectangles in the sequences. -10 and -35 regulatory boxes are shown only for the -10 box with the conserved motif TAN3T.

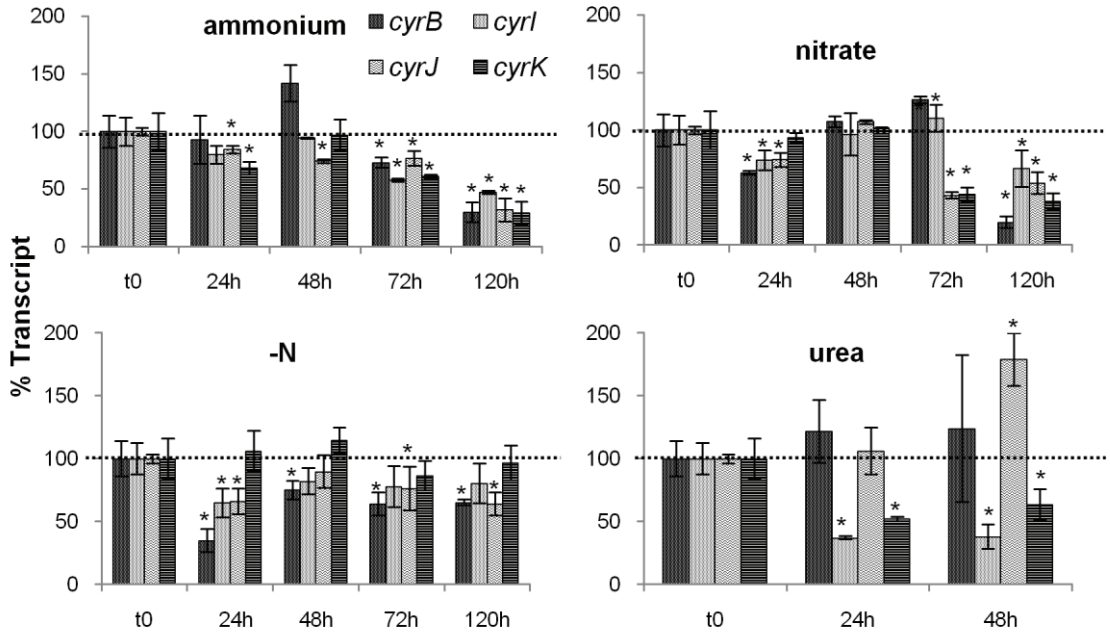


Figure 2.4.S4 Comparison of the expression of four *cyr* genes from *C. raciborskii* CS-505 grown under ammonium, nitrate, N-deprivation and urea. Error bars indicate ±SD of the mean (n = 3). Values significantly different (t-test p-value < 0.05) from t = 0 are marked with a * symbol.

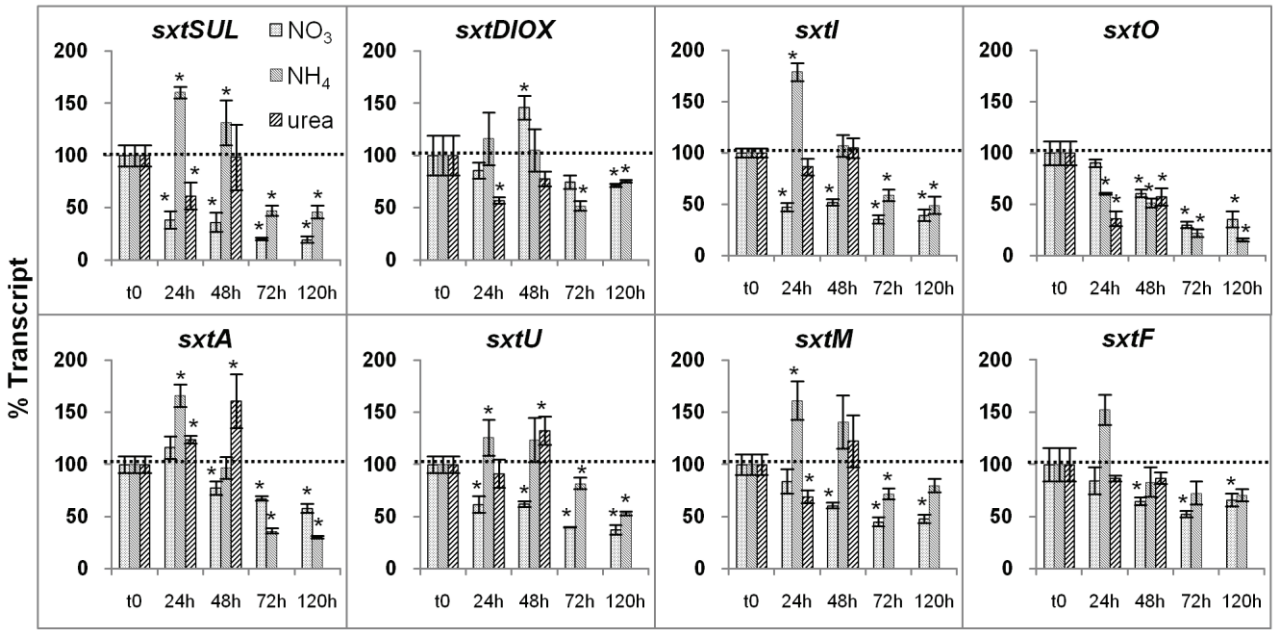


Figure 2.4.S5 Comparison of the expression of eight *sxt* genes from *R. brookii* D9 grown under alternative N regimes. Nitrate (white bars with black dots), ammonium (left thin stripes) and urea (right thick stripes). Error bars indicate ±SD of the mean (n = 3). Values significantly different (t-test p-value < 0.05) from t = 0 are marked with a * symbol.

2.5 Publication V

Gene expression dynamics and heterocyst differentiation in *Cylindrospermopsis raciborskii* versus *Raphidiopsis brookii*, a non-diazotroph closely related species

Karina Stucken, Uwe John, Rodrigo Gutierrez, Gernot Glöckner, Mónica Vásquez and Allan Cembella.

2.5.1 Abstract

The filamentous cyanobacterium *Cylindrospermopsis raciborskii* CS-505 differentiates the apical cells to heterocysts, specialized cells for dinitrogen (N_2) fixation under fixed nitrogen (N) i.e. nitrate. The terminal heterocyst phenotype in *C. raciborskii* CS-505 is likely a consequence of the absence of heterocyst pattern formation genes. The process of heterocyst differentiation in *C. raciborskii* CS-505, evolution of N_2 fixation, heterocyst frequency and the time series of gene expression were studied in cultures grown under three different N regimes, with ammonium, nitrate or under N-deprivation. Upon N deprivation, nitrogenase activity increased two-fold after 24 h, whereas heterocyst numbers doubled only after 48 h. Gene expression dynamics revealed the induction of NtcA gene targets shortly after N step-down, although *ntcA* was induced between 24-48 h together with heterocyst early-differentiation and maturation genes. In addition, comparison of ammonium-regulated genes between CS-505 and the closely related *Raphidiopsis brookii* D9 showed similar expression patterns, stimulating translation machinery. Energy metabolism was repressed in D9, however, as a possible reflection of toxicity following growth on high ammonium levels. We conclude that the response of CS-505 to alternative nitrogen regimes is dramatically different from that of *Anabaena* sp. PCC 7120, the type strain for studies of cell differentiation in cyanobacteria. In CS-505 the responses are slower, but nevertheless very efficient, and probably involve unknown factors of transcriptional and translational regulation.

2.5.2 Introduction

Cyanobacteria can assimilate a variety of nitrogen (N) forms: nitrate, nitrite, ammonium, urea, cyanate and in some cases dinitrogen (N_2) through fixation (diazotrophy). Under N limitation, some filamentous cyanobacteria may differentiate heterocysts, specialized cells for N_2 fixation in order to protect the nitrogenase form oxygen-inhibition. The process of N_2 fixation and differentiation of heterocysts has been well studied in *Anabaena* sp. PCC 7120, *A. variabilis* ATCC 29413 and *Nostoc punctiforme* ATCC 29113, all of which develop heterocysts in a intercalated pattern of 10-15 cells each, occurring within the first 24 h of N-deprivation (Zhang *et al.*, 2006, Meeks & Elhai, 2002, Xu *et al.*, 2008). Upon N-starvation, a cascade of gene reactions is triggered leading sequentially to the differentiation of cells and later to the expression of N_2 fixation genes. These include the induction of the main transcriptional regulator in N metabolism *ntcA,* and the master gene for heterocyst differentiation *hetR* via *nrrA* (Herrero *et al.*, 2004, Muro-Pastor *et al.*, 2002, Ehira & Ohmori, 2006a). As part of the early response, genes involved in pattern formation (*patS, patA, hetN*) and inhibition of cell division (*hetC*) are expressed. Finally, genes involved in the maturation of the heterocyst, i.e., deposition of the glycolipids (*hgl*, *hgd*, and *dev* genes), polysaccharide (*hep* genes and cluster *alr2825-alr2841* of *Anabaena* sp. PCC 7120) and N_2 fixation genes (*nif*) are induced (Ehira *et al.*, 2003, Xu *et al.*, 2008).

In *Nodularia spumigena,* there is an apparent decoupling between N_2 fixation and heterocyst differentiation, since ammonium ions inhibited N_2 fixation while not affecting heterocyst development (Vintila & El-Shehawy, 2007). The species *Cylindrospermum licheniforme* and *Cylindrospermopsis raciborskii* develop only terminal heterocysts. Studies on *C. licheniforme* showed a similar timing of N_2 fixation and heterocyst differentiation on this species with *Anabaena* and *Nostoc* (Van De Water & Simon, 1982). In *C. raciborskii* however, there was no inhibition of N_2 fixation when concentrations of nitrate or ammonium up to 3000 μg N l^{-1} were used, suggesting a more relaxed mechanism of regulation in this species (Sprober *et al*., 2003).

The genome of *Cylindrospermopsis raciborskii* CS-505 is the smallest described for a heterocyst-forming cyanobacterium. Comparative analysis revealed that some of the genes described above as essential for pattern formation are absent (*patS, hetN*), as well as other related genes (e.g., *hetC, hetL, ccbp*, among others), implying that this cyanobacterium contains the minimum set of genes required for differentiation of a functional heterocyst (**Publication II**). In addition, our recent investigation into the main transcriptional regulated processes under N-

deprivation in CS-505 showed that *ntcA* was expressed only after 48 h, while there were no changes in the expression of the *nif* genes or *hetR.* This strongly suggested a late response to N-deprivation when compared to other heterocystous cyanobacteria such as *Anabaena* sp. PCC 7120 (heretofore referred as *Anabaena*) (**Publication IV**).

On the other hand, the genus *Raphidiopsis* comprises filamentous non-heterocystous cyanobacteria unable to fix N_2. Species of this genus have not been deeply investigated but recent reports show that they are also widely distributed, and often thrive in coexistence with *C. raciborskii* (Mohamed, 2007). Although *C. raciborskii* and *Raphidiopsis* have been assigned to different genera, species of both genera form part of a monophyletic cluster (Alster *et al.*, 2009, Gugger *et al.*, 2005, **Publication I**). In addition, we have shown that the genomes of *C. raciborskii* CS-505 and *R. brookii* D9 share 2,539 CDS with an average nucleotide identity (ANI) >90%. These values correspond to strains from the same species that have evolved independently (allopatric speciation) (**Publication II**).

The ability to fix N_2 is a major difference between the strains CS-505 and D9. The loss of the *nif* operon and several other genes related with N_2 fixation but not of all genes necessary for the development of a mature heterocyst in D9 is one example of how gene losses have shaped the D9 genomic structure. The presence in D9 of several genes involved in N metabolism that are absent in the CS-505 genome seem to compensate for the inability of fix N_2 (**Publication II**). Some of those genes include an amino acid transporter, four asparagine synthases and a fragment of the nitrile hydratase (*nhlEAB*) operon.

We questioned how CS-505 and D9 with such small but related genomes respond to the variability in an environmental condition as ecologically important as N availability. Using various N regimes to model transcription behaviour, we compared gene expression of CS-505 and D9 in a time-series of 1-24 h under "normal" culture conditions, i.e., cultures adapted to grow on nitrate versus ammonium-replete cultures. In addition, to unravel the timing of gene expression involved with heterocyst formation in CS-505, we measured nitrogenase activity, heterocyst frequency and gene expression in N-deprived cultures.

2.5.3 Results

Global time-dependent transcriptomic responses to N availability

The 24 h transcriptomic responses of both strains to four N-regimes were previously reported (**Publication IV**). In order to understand the timing of responses after shifts in N availability, time series experiments were conducted in cultures of *C. raciborskii* CS-505 and *R. brookii* D9 subject to growth on nitrate, ammonium and N-deprivation (the last one only to CS-505). To evaluate the effect of ammonium and N-deprivation on gene expression in each one of the microarray impressed genes, we performed regression analysis to retrieve significantly expressed genes with respect to nitrate.

Differentially expressed genes as determined by regression showed that after the first hour, the response elicited by ammonium in CS-505 and D9 was similar (140 and 156 genes regulated, respectively). However, there was a larger increase over time of regulated genes under ammonium in CS-505 (from 140 to 510 genes) compared to D9 (from 156 to 239 genes) (Table 2.5.1). We have previously reported that 12% (495 genes) of the CS-505 gene content responds after 24 h of N-deprivation. Similarly, the number of genes regulated in CS-505 under N-deprivation, reached 506 after 24 h. This response appeared to have already been achieved after 8 h, when the number of regulated genes was 488. After 48 h, the number of regulated genes doubled (1,018), representing 26% of the CS-505 genome. The proportion of up- and down-regulated genes in all treatments was 1:1 for CS-505 and almost 1:1 for D9 (Table 2.5.1). As we previously reported (**Publication IV**), transposases and repeat sequences formed part of the regulated genes of CS-505. Repeats were regulated irrespective of the N regime, and after the first hour, expression was detected in the ammonium treatment.

Table 2.5.1 Number of genes regulated over time by the different N regimes.

		CS-505				D9	
		-N		Ammonium		Ammonium	
Time (h)		up	down	up	down	up	down
1	genes	64	60	81	59	67	89
	repeats	--	1	10	5	--	--
3	genes	172	171	104	133	49	75
	repeats	11	5	14	10	--	--
8	genes	244	244	145	149	73	110
	repeats	9	3	13	12	--	--
24	genes	248	258	260	250	104	131
	repeats	3	13	24	22	--	--
48	genes	487	531	--	--	--	--
	repeats	60	21	--	--	--	--

Comparative transcriptional responses of *C. raciborskii* CS-505 and *R. brookii* D9

Differentially expressed genes that responded in a similar pattern over the time course for each N-regime and may thus be functionally related, were explored by K-means clustering. We distinguished from the expression patterns an early (1-8 h) and late (24-48 h) response that would give a maximum of seven gene clusters if genes do not highly differ in their magnitude of regulation. Following this biological consideration and by means of hierarchical clustering, genes regulated by N-deprivation in CS-505 were divided into eight clusters of expression and nine versus six clusters were determined for ammonium-regulated genes in CS-505 and D9, respectively (Figure 2.5.1).

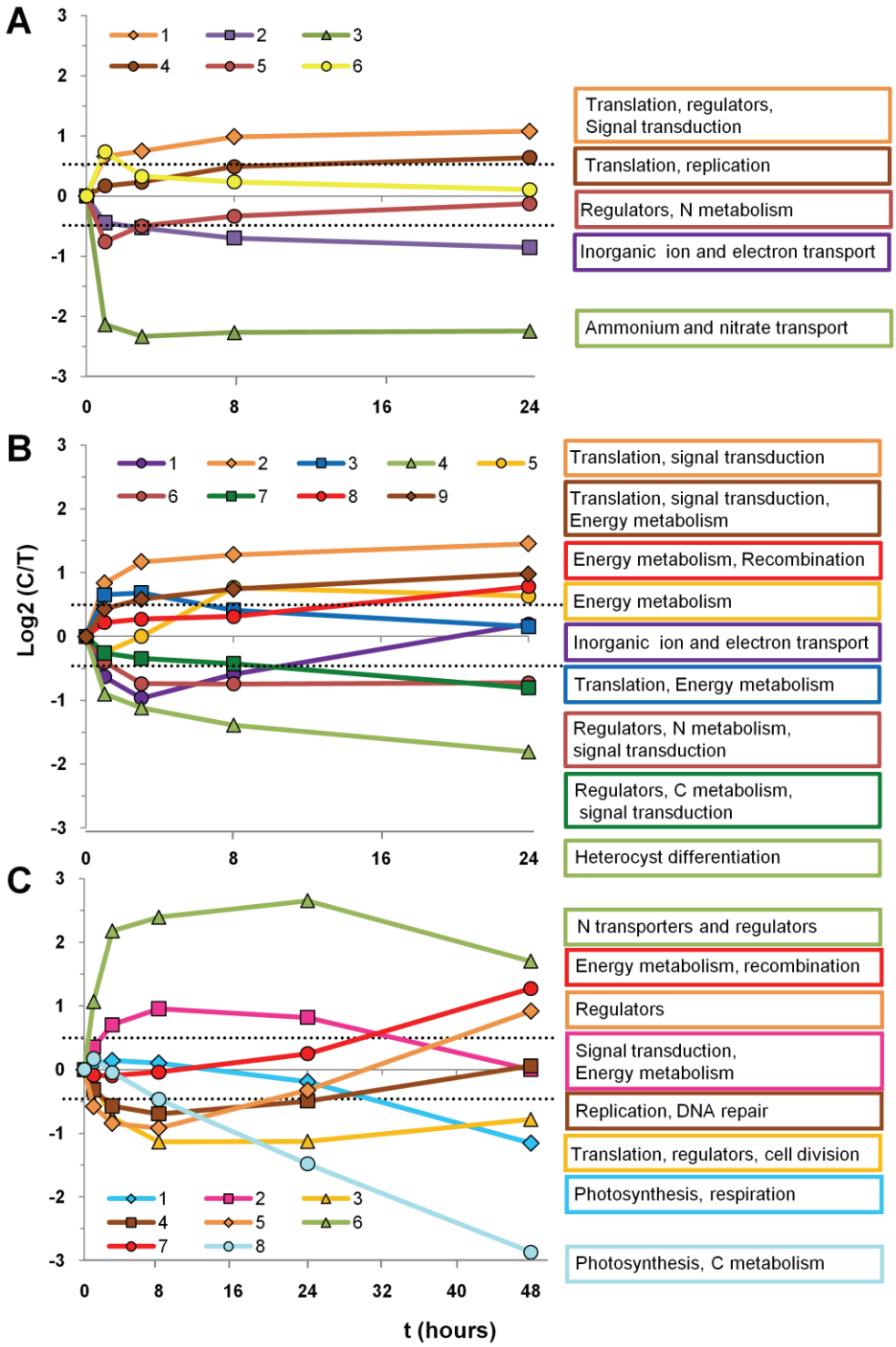


Figure 2.5.1 Temporal expression of differentially regulated genes of D9 (A) and CS-505 (B) grown on ammonium and CS-505 under N-deprivation (C). Expression patterns were determined by hierarchical and KMC clustering; dotted lines delimit the region of low intensity (log_2 ratios < [±0.58]) values, where genes were considered as not regulated. The processes better represented in each cluster are indicated into the color boxes, color boxes: denoting N metabolism, ion and electron transport, general energy metabolism, photosynthesis and carbon metabolism and replication, regulation, and signal transduction. Genes belonging to signal transduction were mainly cell wall or polysaccharide biosynthesis that may be involved in this process. Detailed lists with intensity values and standard deviations for each cluster are provided as supplementary material.

Comparative transcriptomic responses to ammonium availability

Although more regulation patterns were found for CS-505, in general the response of D9 to ammonium was similar to that of CS-505 with the exception of D9 cluster 3, which was rapidly and strongly repressed after the first hour. Down-regulated genes belonged mostly to N (and other inorganic ion) transport and metabolism, and were represented in all D9-repressed clusters and in three of four in CS-505. In these experiments, the genes that showed the highest regulatory response (intensity) were related to N control, but interestingly, this response was not comparable to CS-505, where the genes for N transport and metabolism *amt* and *nrt* were not significantly down-regulated with respect to nitrate. Other differences in N regulation were conveyed by the expression of the nitrilase (*nhl*) gene cluster in D9 cluster 3, and of the N-regulator genes *nrrA* and *glnB,* which were highly repressed in D9 after 1 h (*nrrA*) and 24 h (*glnB*), whereas in CS-505, *nrrA* was slightly repressed after 8 h (cluster 7) and changes in the expression of *glnB* were not detected. The highest down-regulated (progressively) genes from CS-505 (cluster 4) were *nirA* and some heterocyst differentiation genes *hetF, fraG, abp3* and a *hetP*-like gene, including the sigma factor *sigG.*

The early up-regulated genes of both strains with ammonium (D9 cluster 1, CS-505 cluster 2), corresponded mostly to the process of translation, but interestingly, a set of genes clustered in the genome and unique for each strain respect to the other, was part of the highly expressed genes. Although unique for each strain, the clustered genes included several sugar- and methyl transferases that may have relation to cell to cell communication and therefore were classified as signal transduction.

There was a common response in both CS-505 and D9 to an early and transient regulation (1-3 h); these patterns although similar, did not represent the same functional categories. Early repressed genes from CS-505 cluster 1 were part of a set of clustered genes in the genome encoding for a NADH-ubiquinone oxidoreductase complex, whereas D9 cluster 5 included the *nrtCD* genes, posttranscriptional regulators and chaperones (*groES, groEL*). Interestingly, within the transient upregulated genes of CS-505 cluster 3, were *nifHK*, although *nifH* was further downregulated after 24 h; other genes following this pattern belonged to the translation category. There were no clear functional categories representing D9 cluster 6 genes, indicating that diverse processes might have been affected shortly after the incorporation of ammonium to the medium.

The additional patterns of expression of CS-505 (clusters 5, 8 and 7) indicated that besides an increased protein synthesis in the cells, metabolism and recombination were also activated. Four from five up-regulated clusters of CS 505 (clusters 3, 5, 8 and 9) contained genes related with energy metabolism, among others. For example, CS-505 clusters 3, 8 and 9 included the urease, fatty acid biosynthesis, Photosystem II, ATP synthesis and cytochrome oxidase genes. This same category was repressed in D9, where Photosynthesis, ATP synthesis, inorganic carbon transport and respiration were part of progressively repressed genes of cluster 2. Other better represented genes in CS-505 in comparison with D9 were from the signal transduction category, present in up- and down- regulated clusters (2, 6, 7 and 8). Transposases were found to be repressed in CS-505 clusters 8 and 9, repeated sequences however, were regulated without a specific pattern and they were scattered among all clusters but 1 and 5.

Temporal expression of genes under N-deprivation in CS-505

The dynamics of gene regulation of CS-505 under N-deprivation showed that after the first hour of removal of the N source, there was a high induction of genes known to be controlled by N-deprivation (cluster 6), including *amt1, nrrA, urtA* and *glnA*. Accompanying the induction of *glnA, gifA*, encoding for the inhibitor of glutamine synthase, was repressed. In addition, the heterocyst differentiation gene *hetP,* and a gene cluster of unknown proteins were expressed as well. The expression of all these genes remained induced until 48 h, although they showed tendency to decrease in intensity towards 48 h. Genes belonging to cluster 2 were progressively induced until 8-24 h, and returned to the initial levels after 48 h. None of these genes were known to be involved with N metabolism, but rather with oxidoreduction reactions and several glycosyltransferases may be related with signal transduction. Within this cluster, a gene cluster unique for CS-505 and D9 (compared to other cyanobacteria) that encodes for a putative arginase, ABC transporters and putative nickel binding proteins was also expressed, as suggestive of new N transport and metabolism capacity in these strains.

As part of the early response, the protein synthesis machinery was inhibited within the first hour as represented in cluster 3. These inhibited elements included ribosomal proteins, transcriptional regulators, chaperones, cell division proteins (*minC, minD, ftsZ, ftsY*), the RNA polymerase genes (*rpoB, rpoC*), the DNA gyrase (*gyrB*) and additional transcription factors. Among the latter group, was a transcriptional regulator AbrB (CRC_00661), similar to sll0359 of

Synechocystis PCC 6803, which was proposed to be essential to growth; Sll0822, a second AbrB regulator was demonstrated to be necessary for the regulation of several N regulated genes (*nirA, glnA*, *amt1, glnB,* among others) (Ishii & Hihara, 2008). The homolog to sll0822 in CS-505 is CRC_00151 and although its expression was modified by N-deprivation, induction was observed only after 48 h. From 24 h on, the growth-arrestment strategy followed in the first hours incorporated energy saving or N recycling by the inhibition of photosynthesis and ATP synthesis (*psa, psb, apc, cpc, atp*), bicarbonate (*cmp*), and sulfate transport (*cys*), CO_2 fixation (*rbcSL, ccm*) and glucose metabolism (clusters 1 and 8). At 48 h most of the repressed genes from cluster 5 were induced and from cluster 4 de-repressed. Genes from cluster 5 were barely coherent in functional categories; however, transcriptional regulators and three genes related to heterocyst differentiation followed this pattern *hetF*, *sigE* and a *hetP*-like gene. Cluster 4 comprised ribosomal proteins, DNA repair genes, transcriptional and posttranscriptional regulators, which indicated that at this time point, the cells may start actively metabolizing again. Following this hypothesis, the largest number of genes (411) appeared to be up-regulated at 48 h (cluster 7). Despite the large number of genes, two main functional categories were evident (Table 2.5.2). First, one functional category comprised genes related to biosynthesis of cofactors or molecules related to modification of the energetic status of the cell, which included genes for the biosynthesis of cobalamin, thiamine, terpenes and pyridoxal-phosphate biosynthesis, among others. Second, the additional category included genes for DNA replication, recombination and repair, as well as for RNA modification; here were included transposases, repeat sequences (56), DNA and RNA methyltransferases, signal transduction-related proteins, transcriptional regulators and competence proteins. These two categories indicated that although the cells started to metabolize fundamental cofactors, genetic rearrangements and possible transformation are active parts of CS-505 metabolism.

Table 2.5.2 Genes representing the two main processes induced in CS-505 after 48 h of N-deprivation.

Locus tag	Product / Description	Log_2 (T/C)*				
Energy metabolism (cofactors)		**1h**	**3h**	**8h**	**24h**	**48h**
CRC_03493	Thiamine-monophosphate kinase	0.07	0.12	0.31	0.3	1.8
CRC_01276	Phosphoribosylglycinamide formyltransferase	-0.22	-0.4	-0.36	-0.1	1.73
CRC_00155	Phosphoribosyltransferase	-0.17	-0.12	-0.04	0.25	1.73
CRC_00287	Phospholipid/glycerol acyltransferase	-0.17	-0.3	0.01	0.43	1.68
CRC_02941	Hydantoinase B/oxoprolinase	-0.06	-0.24	-0.14	-0.04	1.59
CRC_03453	Terpene synthase	-0.04	-0.16	-0.3	0.08	1.48
CRC_01892	Shikimate kinase	-0.18	-0.18	-0.08	0.49	1.46
CRC_01162	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	-0.1	-0.22	0.09	0.42	1.44
CRC_02942	Hydantoinase/oxoprolinase	-0.16	-0.12	-0.09	0.01	1.43
CRC_00797	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (HisA)	0.18	0.47	0.3	0.3	1.2
CRC_01332	Phosphoribosylformylglycinamidine synthase II	-0.43	-0.29	-0.47	-0.07	1.16
CRC_00180	Cobyrinic acid a,c-diamide synthase	-0.14	-0.03	-0.06	0.13	1.06
CRC_01679	Phosphoenolpyruvate synthase	-0.54	-0.76	-0.68	-0.35	1.04
CRC_00118	Cobalbumin biosynthesis enzyme	-0.15	-0.25	-0.01	0.35	1.03
CRC_03258	Pyridoxal phosphate biosynthetic protein (PdxJ)	-0.19	-0.44	-0.51	0.02	0.9
CRC_02524	Ketopantoate hydroxymethyltransferase	-0.09	-0.37	-0.35	0.21	0.88
CRC_01907	Dihydrodipicolinate synthase subfamily	-0.26	-0.4	-0.3	0.31	0.85
CRC_01954	Cobalamin (vitamin B12) biosynthesis CbiX protein	-0.17	-0.19	-0.38	0	0.81
CRC_00589	Glucose-1-phosphate thymidylyltransferase, short form	-0.13	-0.21	-0.08	0.19	0.75
CRC_00386	Molybdopterin cofactor biosynthesis MoaC region	-0.4	-0.76	-0.85	-0.3	0.72
CRC_00542	Glyoxalase/bleomycin resistance protein/dioxygenase	-0.14	-0.26	-0.08	0.14	0.7
CRC_00439	Pyridoxal phosphate biosynthetic protein PdxA	0.12	0.08	0.08	-0.05	0.68
CRC_00161	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit	-0.18	-0.18	-0.26	-0.16	0.64
DNA modification, repair, genomic rearrangements						
CRC_00006	Cyclic nucleotide-regulated ABC bacteriocin	-0.09	-0.03	-0.13	0.05	1.91
CRC_00310	Radical SAM	-0.34	-0.65	-0.52	-0.23	0.74
CRC_00330	PilQ (type II and III secretion system protein)	-0.21	-0.21	-0.17	0.02	0.76
CRC_00410	PilN-like (Fimbrial assembly)	-0.14	-0.23	-0.18	0.04	0.80
CRC_00420	Protein splicing (intein) site	-0.21	-0.25	-0.22	0.15	1.80
CRC_00453	Heat shock protein DnaJ-like	-0.09	0.02	0.10	0.13	1.88
CRC_00888	Chromosome segregation protein SMC	-0.09	0.04	-0.01	0.38	0.78
CRC_00898	Threonine dehydratase I	-0.01	-0.59	-0.52	-0.35	1.14
CRC_00960	RNA methyltransferase TrmH, group 3	-0.12	-0.75	-0.82	-0.32	1.24
CRC_00961	Ribonuclease III	-0.19	-0.69	-0.82	-0.27	1.42
CRC_00990	Uroporphyrin-III C/tetrapyrrole methyltransferase	0.06	0.13	0.25	0.35	0.96
CRC_01013	Peptidase U62, modulator of DNA gyrase	-0.07	0.24	0.46	0.98	1.33
CRC_01055	Endonuclease III/Nth	-0.34	-0.34	-0.24	0.56	1.39
CRC_01062	Recombination protein O, RecO	-0.32	-0.67	-0.57	-0.31	0.76
CRC_01078	Helix-turn-helix protein, CopG	0.08	0.06	0.42	0.95	1.66

CRC_01266	Heat shock protein DnaJ-like	0.00	0.02	0.19	0.54	1.28
CRC_01477	Exonuclease (RecJ)	-0.16	-0.25	-0.29	0.08	1.17
CRC_01855	Thermonuclease	0.01	-0.05	-0.01	-0.01	1.79
CRC_01859	Essential recombination function protein	0.04	-0.12	-0.04	-0.17	1.91
CRC_01898	Exonuclease SbcC	-0.20	-0.11	-0.03	0.38	0.91
CRC_01955	Uroporphyrin-III C-methyltransferase-like	0.05	-0.24	-0.46	0.06	0.95
CRC_02095	RNA methyltransferase TrmH, group 1	0.03	0.11	0.09	0.46	1.73
CRC_02120	Modification methylase (HemK)	-0.44	-0.38	-0.36	0.17	0.95
CRC_02230	Putative restriction enzyme (Fragment)	-0.24	-0.05	-0.12	0.63	1.34
CRC_02781	Endonuclease/exonuclease/phosphatase	-0.05	-0.33	-0.34	-0.31	1.10
CRC_02987	tRNA modification GTPase TrmE	-0.05	-0.07	-0.10	0.05	1.46
CRC_03036	DNA repair protein RadC	-0.30	-0.37	-0.33	0.02	1.14
CRC_03079	DNA repair protein RecN	-0.04	-0.16	-0.14	0.15	1.40
CRC_03259	Pathogenesis related protein-like	-0.36	-0.8	-0.91	-0.15	1.34
CRC_03302	Restriction modification system DNA specificity domain	-0.11	-0.07	0.11	-0.01	0.84
CRC_01707	Transposase (Fragment)	-0.01	-0.12	-0.11	-0.14	0.80
CRC_01067	Transposase (Fragment)	0.01	-0.05	0.32	0.65	0.99
CRC_03348	Transposase family IS5 (Fragment)	-0.07	0.08	0.06	0.02	0.97
CRC_02221	Transposase IS607 family	0.12	0.09	-0.16	-0.07	2.01
CRC_02347	Transposase IS607 family	0.15	0.08	-0.17	-0.12	2.02
CRC_03464	Transposase IS607 family	0.06	0.01	-0.22	0.05	2.13
CRC_01211	Transposase IS607 family	0.17	0.06	-0.17	-0.02	2.16
CRC_01884	Transposase IS607 family	0.10	0.03	-0.22	-0.01	2.21
CRC_01065	Transposase IS607 family	0.10	0.02	-0.20	0.04	2.23
CRC_03382	Transposase ISL2 family (Fragment)	-0.01	0.15	0.07	0.14	0.75
CRC_00337	Transposase OrfB	0.03	0.15	0.36	0.59	0.63
CRC_02011	Transposase, IS4 (Fragment)	0.05	0.13	-0.09	-0.63	1.34
CRC_01443	Transposase, IS891/IS1136/IS1341	0.09	0.16	0.38	0.61	0.59
CRC_01210	Transposase, IS891/IS1136/IS1341	0.00	0.09	0.31	0.58	0.63
CRC_01616	Transposase, IS891/IS1136/IS1341	0.02	0.12	0.33	0.63	0.64
CRC_03268	MutS 1 protein	0.10	0.09	-0.01	-0.03	1.65

* Control represent intensity values at time zero; treatment at every time point (1, 3, 8, 24 and 48 h)

Nitrogen fixation and heterocyst development in CS-505

The experiments carried out in this work employed adapted cultures to nitrate which were washed and transferred to ammonium or to a medium depleted of N. The presence of heterocysts in *C. raciborskii* in nitrate-containing medium accounts for a basal activity of the nitrogenase enzymatic complex. Measurement of nitrogenase activity, and the frequency of heterocysts developed under nitrate, ammonium and N-deprivation conditions, indicated that a constant rate of N_2 fixation was maintained in cultures of CS-505 grown in nitrate-containing medium, whereas N_2 fixation increased to 230% within the first 24 h of N-deprivation (Figure 2.5.2A). At 24 h, however, there were no observable changes in heterocyst frequency; the heterocyst numbers doubled only after 48 h of N-deprivation (Figure 2.5.2B), while the rate of N_2 fixation remained at 230%. After 120 h, N_2 fixation and heterocyst counts had increased to 860% of the initial levels in N-deprived cells. Heterocysts did not disappear in the ammonium-grown cultures; counts decreased to 80% only after 120 h, but nitrogenase activity decreased earlier (at 72 h) to 6% and to the detection limit after 120 h.

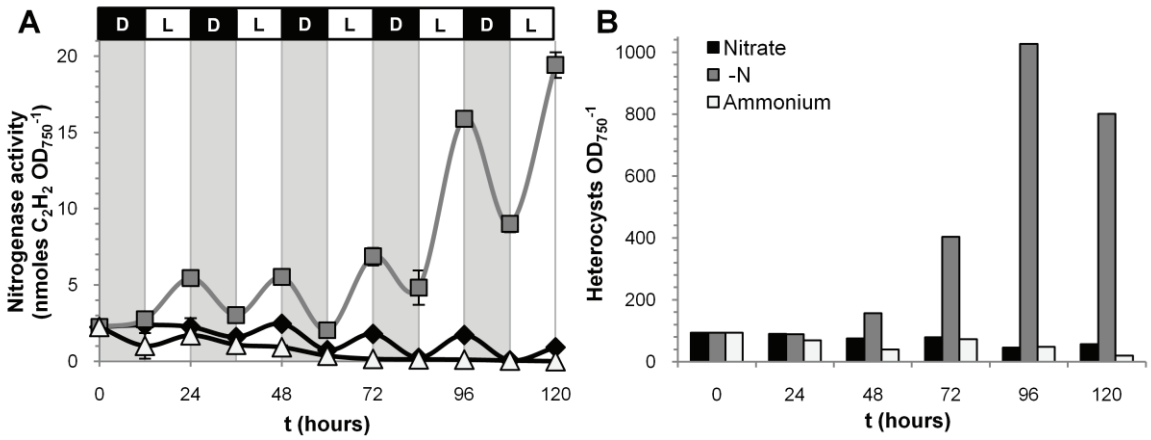


Figure 2.5.2 Diel activity of the nitrogenase (A) and heterocyst frequency (B) of CS-505 under alternative N-regimes; squares: N-deprivation, diamonds: nitrate and triangles: ammonium. Activity was measured as nmoles of ethylene formed by unit of biomass (OD_{750}) at the end of the dark (black boxes) and light (white boxes) phase. Error bars indicate ±SD of the mean (n = 3).

Late responses of heterocyst-specific genes

In order to correlate the expression of heterocyst differentiation genes with the physiological response of CS-505 to N-deprivation, we selected a compendium of 119 genes from which 55 have been experimentally linked to heterocyst development, 41 are part of the core genes of heterocystous cyanobacteria and 23 are involved in N_2 fixation (**Publication II**). 50% (55) of these selected genes were regulated at one or more time points during N-deprivation in CS-505 (Figure 2.5.3).

The *glnB* (encoding for the regulatory protein PII), *nrrA, hanA*, *hetP, patU3* genes and two codifying for hypothetical proteins (CRC_00582 and CRC_03071) were all induced within 3 h but showed a maximum of expression between 8-24 h, decaying at 48 h. After 24 h, but with maximum expression at 48 h were induced genes that in *Anabaena* respond close to 8-24 h, and represent the initial stages of heterocyst differentiation and maturation (*hetM, hgdABC, devC, abp1, conR*). Within this group was found *ntcA*, however, *hetR* whose expression is fundamental for the regulation of heterocyst specific genes and of *ntcA*, showed no expression changes within the complete time series. A different pattern of expression shared the sigma factors *sigE* and *sigG*, together with *hetF*, a *hetP*-like gene, CRC_00688, *hetZ* and CRC_02624, all repressed from the first to 8 or 24 h, and then finally induced at 48 h (*hetZ* and CRC_02624 were not induced, rather de-repressed at 48 h). Nitrogen fixation genes are expected to be expressed only in heterocysts and after the development process is completed. Contrarily, in CS-505, the expression of *nifHDK* genes was upregulated during the first 1-8 h, decayed after 24 h and increased again after 48 h. Other N_2 fixation genes (*nifZ, fdxH, hesA* and *hesB*), *patB* and the hydrogenase uptake genes (*hupSL*) were all upregulated after 48 h. The nitrogenase accessory gene *nifB* and *prpJ,* a phosphatase genes were repressed after 48 h.

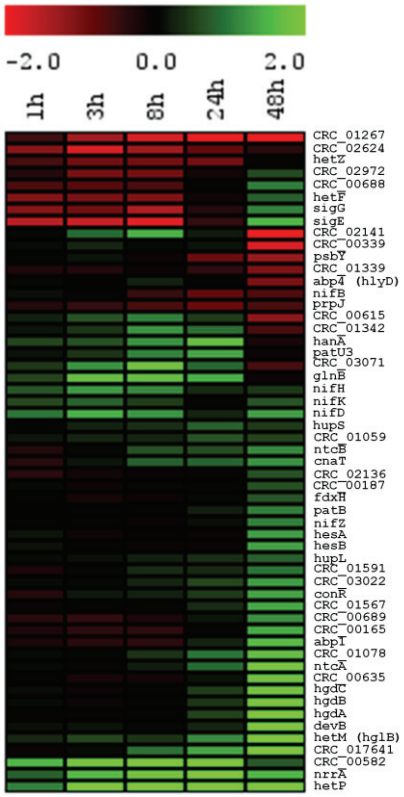


Figure 2.5.3 Differential expression of 55 genes related to heterocyst differentiation and N_2 fixation in *C. raciborskii* CS-505 under N-deprivation. The heat map shows up-regulated genes in green and down-regulated in red boxes. Genes are organized according to their expression patterns.

2.5.4 Discussion

This work provides insights into the gene expression dynamics, and processes of heterocyst differentiation and N_2 fixation in *Cylindrospermopsis raciborskii* CS-505 and represents a transcriptomic comparative study of the closely related cyanobacteria *C. raciborskii* CS-505 and *Raphidiopsis brookii* D9 to ammonium repression, in order to understand the common pathways controlled by N.

The repression of N assimilation pathways when cells are transferred from other N sources to ammonium (or to other N source that is easily assimilated) is a process referred as N control, and is well known in bacteria and cyanobacteria (Herrero *et al*., 2001). This response was clearly observed in D9, where after the first hour, N metabolism genes were strongly repressed. The responses of CS-505 although similar were not as pronounced as in D9, where the highest repressed genes were progressively down-regulated and included only the nitrite reductase gene (*nirA*). The common response of both strains to ammonium was related to the increased transcription of the translation machinery and of signal transduction related proteins. The induction of these genes is an indication of protein synthesis and signaling that may lead to

growth. However ammonium (2 mM) affects differentially the growth of CS-505 and D9, where growth of D9 is delayed (**Publication IV**). Despite the induction of the translation machinery, energy metabolism was negatively affected in D9. The repression of photosynthesis (phycobilisomes, ATP synthesis, and both Photosystem I and II) besides CO_2 fixation and respiration only in D9, support the putatively toxic effect of ammonium in D9 on the photosynthetic machinery, as it is known in photosynthetic microorganisms (Krogmann *et al.*, 1959, Azov & Goldman, 1982) and as we have shown for D9 in **Publication IV**. On the contrary, some photosynthetic genes were induced in CS-505, and CO_2 fixation and respiration were not affected.

Among the highly repressed genes of CS-505 with ammonium, were heterocyst-specific genes (*fraG, hetF, hetP*-like, *abp3*, and *sigF*), suggesting a degree of inhibition of the differentiation process despite of the constant frequency of heterocysts within the first 24 h of ammonium growth. N_2 fixation in the other hand, decreased by 50% with respect time zero during the first 24 h, and continued to decrease until five days (Figure 2.5.2). The repressive effect of ammonium on diazotrophic growth was observed by the repression of the *nifHK* genes after 24 h. However, at early stages (1-8 h), the nitrogenase genes were induced irrespective of the N-regime (ammonium or N-deprivation). Providing that N_2 fixation in CS-505 follows a diel pattern, and nitrogenase activity behaves as detected in the first 48 h in CS-505 growing under nitrate, ammonium or N-deprivation, the early expression of the *nif* genes could exhibit endogenous rhythms controlled by circadian regulators, as in the filamentous cyanobacterium *Trichodesmium* sp. IMS101 (Chen *et al.*, 1998, Berman-Frank *et al.*, 2001). This same pattern of expression was observed for several other genes regulated under N-deprivation in CS-505 (cluster 2) including two nitrate assimilation enhancers (*cnaT* and *ntcB*).

The relationship between site-specific recombination and cellular differentiation has been described for several organisms, and occurs in combination with developmentally regulated DNA rearrangements. The development of a functional heterocyst in cyanobacteria involves heterocyst-localized genetic rearrangements within the coding regions of the *nifD, fdxN* and *hupL* genes (Carrasco *et al.*, 1995, Carrasco & Golden, 1995, Golden *et al.*, 1987, Haselkorn, 1992). In CS-505, none of these genes are interrupted by excision elements, however, high rearrangement activity was observed after 48 h of N-deprivation (Table 2.5.2) at the time where heterocyst differentiation genes were induced. Repeat regions, which are also potential recombinatory elements, were shown to be regulated (mostly repressed) in CS-505 after 24 h of growth in

alternative N sources (**Publication IV**). Similarly, our time series experiments revealed that repeats were up- and down regulated in an equal proportion irrespective of the N-source and a without a specific expression pattern. This random pattern reflects a N-independent regulation of repeats, however, after 48 h of N-deprivation, when there was also an increase on the energy metabolism in the cells, most repeats were induced. The function of transposases and repeats thus, seems to be different with repeats induced along the cell cycle and transposases induced later. Whether this recombination activity is related to general metabolism in vegetative cells in the period of better fitness or if it is related to regulation of heterocyst development in CS-505 should be clarified by gene expression of heterocyst enriched RNA.

The strategies followed by CS-505 after the removal of combined N were sequentially organized as first, a stress response reflected in the decrease of protein and DNA synthesis; therefore cells stop growing (not shown but see **Publication IV**). Decrease in the protein content has been related to the process of heterocyst differentiation in *Anabaena variabilis,* which upon N-deprivation and before heterocyst development degrade of 20-40% of their total protein (Thiel, 1990). The energetic metabolism of CS-505 was not affected at early stages; degradation of photosynthetic pigments (*cpc, apc* genes), inorganic carbon (Ci), CO_2 fixation and gluconeogenesis, as determined by the repression of the related genes, occurred after 24 h and was maintained until 48 h. This response is not comparable to that of *Anabaena variabilis* ATCC 29413 under N-deprivation, in which phycobilisome proteins are inhibited as early as after the first hour of N-deprivation (Wealand *et al.*, 1989). CS-505 cells however, sensed the absence of N during the first hour of N-deprivation, where the C:N ratios increased with respect to time zero after the first hour (Figure 2.5.4) and N metabolism genes (*amt, urtA, glnA, glnB*) were induced, a common response in diazotrophic and non-diazotrophic cyanobacteria. These early responses are determined by the activated form of NtcA, generally transcribed at a basal level (Herrero *et al*., 2004). N-controlled genes were downregulated until 48 h, although at that time, their regulation was attenuated, showing the start of the end of the need to obtain nitrogenated compounds. The C:N values however, decreased after 48 h, implying that CS-505 was still deficient on N (Figure 2.5.4).

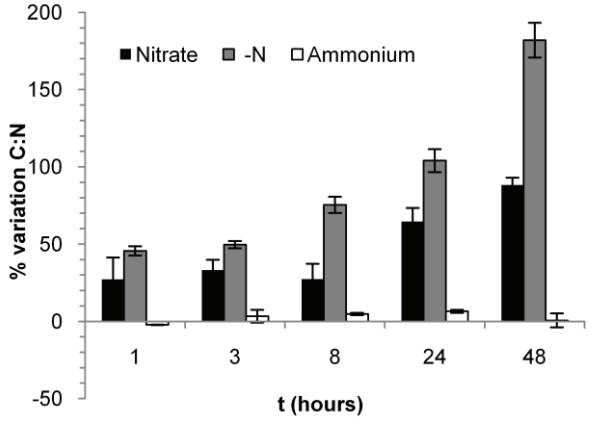


Figure 2.5.4 Cellular C:N ratios of CS-505 grown under nitrate, ammonium and deprived from fixed N.

If N-deprivation continues, in heterocystous cyanobacteria the development of heterocysts is triggered. In *Anabaena*, approximately after 8 h of N step-down, NtcA directly induces its own transcription and that of other heterocyst-specific early response genes (*hetC* and *hetP*, *devH*, *patS*, *patA*, *patB,* glycolipid and polysaccharide genes), and indirectly via NrrA the transcription of *hetR* (Ehira & Ohmori, 2006a). The timing of gene expression in CS-505, as with photosynthetic genes, showed a delay in 24 h with respect to *Anabaena* and *Nostoc punctiforme* (Campbell *et al.*, 2007, Ehira *et al.,* 2003). Based on the physiological and transcriptomic data obtained in this work, the transcriptional responses leading to new heterocysts in CS-505 and *Anabaena* were compared (Figure 2.5.5). In CS-505, the *hetP* gene was the only one among early heterocyst-specific genes induced at early stages that remained induced until 48 h. The timing of response of *hetP* resembled that of *Anabaena* but in the latter strain, *hetP* expression occurs 4-6 h after the induction of *hetR* (Ehira *et al*., 2003). Since there was no increment in the transcription of *hetR*, *hetP* might be a direct target of NtcA regulation in CS-505. In fact, there are three possible NtcA binding boxes in the upstream region of *hetP* (not shown). Other three early heterocyst genes, *nrrA patU3* (of unknown function) and *hanA*, encoding for a histone HU like, peaked at 8 h to decrease in expression at 48 h. The timing of expression of *hanA* has not been determined, but the presence of an NtcA box suggested its early transcription (Khudyakov & Wolk, 1996), our results support that proposition. The expression of *nrrA* on the other hand, depends on *hetR*; although the levels of *hetR* did not change with respect to the initial values, transcript levels are high in nitrate containing media (**Publication II**). These basal levels of *hetR* may therefore be sufficient to regulate the expression of *nrrA* and of the heterocyst specific regulatory cascade. The expression of *ntcA* reached a maximum at 48 h, which correlated with the expression of its gene targets in heterocyst differentiation and with the appearance of new

heterocysts (Figure 2.5.2). This timing of gene expression and of cell differentiation indicates that the transcriptional changes must occur within the interval between 24 and 48 h (Figure 2.5.5).

In cyanobacteria, 2-Oxoglutarate (2-OG) is the carbon skeleton required for ammonium incorporation, and is considered as a signaling metabolite of the C:N status. For heterocyst differentiation to be triggered, the cells must sense a high level of 2-OG, accumulated when the levels of available intracellular ammonium are low (Muro-Pastor *et al*., 2001). High 2-OG levels but under this threshold, would only produce the early response common in non-diazotrophic and diazotrophic cyanobacteria (Zhang *et al.,* 2006). Although we did not measure the intracellular 2-OG, the C:N ratios at 48 h of N-deprived cultures were almost twice as high as the values at 24 h and of those of nitrate-grown cultures at 48 h (Figure 2.5.4). Therefore, if the levels of 2-OG need to reach a threshold for heterocyst differentiation in CS-505, it is likely that the threshold is reached within the interval between 24-48 h, where *ntcA* and the cascade of heterocyst differentiation is induced.

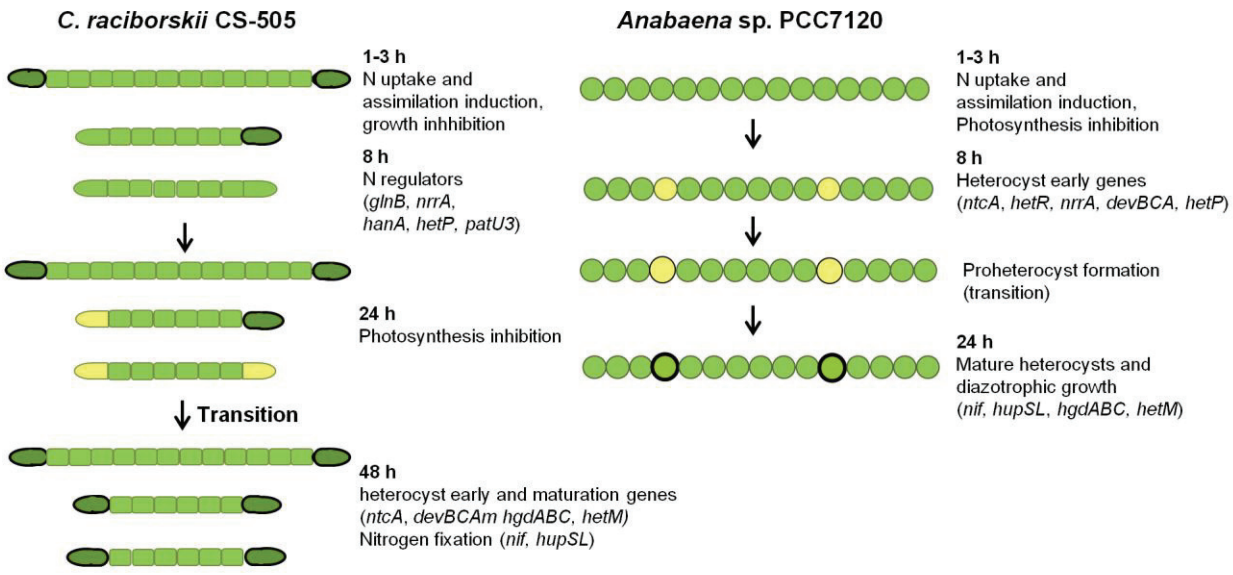


Figure 2.5.5 Diagram of morphological, transcriptional changes and processes affected during heterocyst differentiation in CS-505 compared to *Anabaena sp.* PCC7120. Three main stages of growth are observed in the starting CS-505 nitrate-grown cultures; filaments with terminal heterocysts (dark green cells) at both ends, at only one apical cell, and undifferentiated filaments. Proheterocysts (light yellow cells) and heterocysts are developed from apical cells in CS-505 and intercalated in *Anabaena*. Transcriptional changes in *Anabaena* were obtained from Ehira *et al.,* 2003).

The data gathered in this work shows that in CS-505 the induction of N uptake and assimilatory pathways occurs rapidly after N-deprivation. However, the time necessary to double the number of heterocysts is much longer (48 h). This delay, however, is not indicative of a less efficient management of intracellular N, since with a terminal heterocyst CS-505 is able to maintain vegetative cells along the filament with the sufficient energy to differentiate a new heterocyst when cells have ceased growth.

2.5.5 Materials and methods

Cultures and experimental growth conditions

Batch cultures of *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9 were grown in nitrate containing MLA medium (Castro *et al*., 2004) under a of 12h/12h light/dark cycle at 35 μE m^{-2} s^{-1} and 25°C. Time course experiments were performed under a continuous light regime. 800 ml of nitrate-grown culture (OD_{750} = 0.2) were filtered to remove the old medium and resuspended in twice the volume of new MLA containing 2mM $NaNO_3$ (control medium), 2 mM NH_4Cl or without any combined nitrogen source but with the addition of 2 mM NaCl to restore the osmotic balance. D9 does not grow diazotrophically and therefore the last treatment was not applied for this strain.

To explore the time dependent expression of heterocyst differentiation and N_2 fixation genes, samples from CS-505 were taken from three biological replicates at time points of 0, 1, 3, 8, 24 and 48 h for determination of C:N ratios and RNA isolation. Samples from D9 and CS-505 under ammonium were taken only until 24 h. Samples for biomass measurements (OD_{750}) were taken at 0, 24 and 48 h. For C:N ratios, from each biological replicate, technical duplicates of 4 ml were filtered with low vacuum through a prebaked (500°C, 5h) GF/F filter (Whatman Ltd, United Kingdom). Filters were stored at -20°C and samples analysed with a Carlo Erba NA-1500 Analyzer. Biomass was determined spectrophotometrically as OD_{750} values, which does not take into account the pigments present in the sample but considers only sample turbidity.

Acetylene reduction assay

Nitrogen fixation was assessed for cultures grown during 5 days in media containing 2 mM of $NaNO_3$ or NH_4Cl as N source or without the addition of combined N. The experimental setup was carried out as described above for gene expression analysis. In order to determine if the nitrogenase is active during the light and/or dark phase, cultures adapted to a 12h/12h light/dark cycle were employed and N_2 fixation was measured at the end of the light and of the dark phase.

Nitrogenase activity was assayed by the acetylene reduction assay based on the method of Capone (1993). Shortly, 7 ml of sample culture were placed in 12 ml glass vials. The vials were crimp sealed and 1 ml (20% head volume) of acetylene (100 ppm) was injected into the glass vials. The samples were incubated for 1.5 h under the same experimental conditions in an orbital agitator to achieve gas mixture and allow the reduction of acetylene by the nitrogenase. Ethylene measurements were performed in a Thermo Finnigan trace GC (Waltham, USA) gas chromatograph equipped with a flame detector. The components in 1 ml of sample volume were separated on a column of 80/100 bft, 2-mm diameter, 1/8 in O.D. and 1.8 m length (Restek, Bad Homburg, Germany) and at a temperature of 100° C with a run time of 2 min. A calibration curve for ethylene ($R^2 = 0.9957$) was generated before each measurement and acetylene reduction rates (nmoles ethylene biomass $^{-1}$) were normalized by biomass determined as OD_{750}.

RNA isolation and microarray hybridizations

Cell pellets for RNA isolation were collected by filtration of 50 ml of culture through an 8 µm polycarbonate filter. Pellet samples were resuspended in 800 µl of Qiagen RLT lysis buffer containing β-mercaptoethanol, transferred to a microcentrifuge tube, flash frozen in liquid N_2 and stored at -70°C. Total RNA extraction, quantification and quality control were performed as described in **Publication II**. Microarray hybridizations (from cRNA synthesis to scanning) were performed using the two-color microarray-based gene expression analysis kit from Agilent technologies (Agilent, Santa Clara, CA, USA). 300 ng of total RNA were used as input RNA for cDNA synthesis and cRNA amplification. Labeled RNA was hybridized to microarrays containing probes against the annotated genome of both CS-505 and D9 strains. Samples from t0 were hybridized against the different time points. The procedure was carried out following the manufacturer's instructions with slight modifications as described in **Publication IV**.

Data analysis

Microarrays were scanned on an Agilent G2565AA scanner, and raw data was extracted with the Agilent Feature Extraction Software version 9.1.3.1 (FE). Array quality was monitored using the Agilent QC Tool (v1.0) with the metric set GE2_105_Dec08. The raw intensity signals obtained from the microarray analysis were normalized using the lowess procedure implemented in the normalizeWithinArrays() function of the limma package (Smyth, 2005) in the R software (http://www.r-project.org/). Normalization with limma was chosen over FE normalization due to less dispersion of the data points in the MA plot obtained with limma (data not shown). Each probe in the microarray was represented at least twice. We analyzed the correlation of the normalized intensity values for each probe set and discarded probe sets with correlation values equal or smaller than 0.5. This quality checks identified 3,527 (out of 3,968) and 1,854 (from 3,452) reproducible probe sets from CS-505 and D9 respectively, and were selected for further analysis. As expected, the rejected probes from the D9 strain reside in the low intensity region of the MA plot (not shown) and are likely explained by low signal to background ratio.

The selected reproducible probes were used to determine differential gene expression by the treatments. We used average linkage hierarchical clustering analysis to determine global patterns of gene response to the treatments and the qualitative models determined by the clustering analysis to fit the expression of each gene using linear and non-linear regression analysis. The simplest best model for each gene was selected using Akaike's information criterion. Of the selected data set, 2,651 genes (66% of CS-505 genome) were found to be controlled by at least one treatment (ammonium, N-deprivation) in CS-505 and 1,048 genes (30% of D9 genome) were controlled by ammonium independent on the time. Statistically significant genes were further filtered by a threshold of intensity set to a fold change > 1.5 (log2 ratios higher than [0.58]). Gene expression data of the selected data set are supplied as electronic material (CS-505 –N data.xls; CS-505 NH4 data.xls and D9 NH4 data.xls)

Hierarchical clustering and K means analysis were performed with the Cluster 3.0 software. The number of clusters for K means was estimated first by unsupervised hierarchical clustering of significant genes as determined regression, and clusters were separated by a distance threshold in a range of 2-2.5. The K means function was used with Euclidean distance (the default) and with 1000 replicates.

2.5.6 Supplementary material

Table 2.5.S1 Gene lists, intensity values and standard deviations for all expression clusters are provided as excel sheets under the name of: Table 2.5.S1 KMC clusters.

3. SYNTHESIS

3.1 Evolutionary divergence of *C. raciborskii* CS-505 and *R. brookii* D9

Phylogenomic studies agree on the evolutionary position of the unicellular *Gloeobacter violaceous* as the living ancestor of all contemporary cyanobacteria (Swingley *et al.*, 2008, Blank & Sánchez-Baracaldo, 2010). Due to their polyphyletic origin, the ancestor of filamentous cyanobacteria remains unclear; however, phylogenetic associations based on 16S rDNA sequences have shown that the unicellular *Synechococcus* sp. PCC 7335 groups within the branch of filamentous cyanobacteria from subsection III (Honda *et al.*, 1999). A core set of genes for filamentous cyanobacteria was defined based upon a comparative genomics approach (**Publication II)**. Interestingly, most of these genes were found to have a homolog in the genome of *Synechococcus* sp. PCC 7335, leading to the proposition of this organism as the more feasible ancestor of filamentous cyanobacteria. A clear monophyletic origin is found for heterocyst-forming cyanobacteria, within which *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9 form a clear branch, associated to the symbiont *Nostoc azollae* 078 (**Publication II**).

C. raciborskii is a homogeneous species, as revealed in this thesis (**Publication I**) and in the literature by phylogenetic inferences based on the 16S rDNA and/or ITS markers (Neilan *et al.*, 2003, Gugger *et al.*, 2005, Alster *et al.*, 2009). The monophyletic branch containing *C. raciborskii* species also includes species of the genus *Raphidiopsis*, thereby comprising a species complex. Within this complex, branches are distinguishable only with respect to the geographical origin of the isolates, independently of their taxonomic classification. Accordingly, *C. raciborskii* CS-505 groups with other Australian isolates, whereas *R. brookii* D9 groups with Brazilian isolates of *C. raciborskii* (**Publication I**). Furthermore, pairwise distances of the 16S rDNA sequences, calculated within the Australian *C. raciborskii* clade and including *Raphidiopsis brookii* D9 average 0.0023.

Notwithstanding the homogeneity found among the *C. raciborskii*/*Raphidiopsis* species complex, whole genome analysis by PFGE showed genetic heterogeneity within *C. raciborskii*, whereby strains could be discriminated with 96% certainty. Phylogenetic inference of the PFGE restriction patterns located *R. brookii* D9 as outgroup of the Australian *C. raciborskii* clade. These antecedents indicated a high homogeneity in conserved gene markers, whereas there was genomic heterogeneity. Genome sequencing of CS-505 and D9 demonstrated the high similarity

between the strains, which share 2,539 CDS with an Average Nucleotide Identity (ANI) higher than 90% (93.5%) - a value slightly below the 94-96% threshold accepted to describe two organisms as within a species (Konstantinidis & Tiedje, 2005, Richter & Rosselló-Mora, 2009). However, large differences in the number of plasticity-related elements were detected between both genomes (**Publication II**). Plasticity, viewed as a consequence of genomic rearrangements and acquisition of new functional genes (although not necessarily functional), i.e. by horizontal gene transfer (HGT), was highly represented in the CS-505 genome, which has large number of such elements (transposases, vestiges of transposases and repeated sequences). If these elements are common in the species, they might cause the high genetic heterogeneity found among *C. raciborskii* strains by PFGE profiling, while conserved genes, such as 16S rDNA, would remain highly similar. In turn, the D9 genome, with high similarity in the 2,539 shared genes, shows few plasticity-related elements. This low plasticity may not allow D9 to evolve at the same rate as CS-505, and would also reduce the frequency of HGT events. Although some apparent HGT events are observed in the D9 genome, as for example the occurrence of the *sxt* gene cluster, the synteny observed between the genomes supports the hypothesis that the structure of the D9 genome was shaped by gene losses. The syntenic regions of CS-505 and D9 are occasionally interrupted by transposition elements or gene losses/acquisitions. Although it was not possible to close the genomes, synteny analysis within contigs shows that all genes shared between CS-505 and D9 are located in syntenic regions and that the CS-505/D9 synteny was greater than that of CS-505 (or D9)/*Anabaena variabilis* ATCC 29413.

The ANI between CS-505 and D9 suggests that these two cyanobacteria have been separated for a relatively short evolutionary time. Heterocyst-forming cyanobacteria appear to have evolved before 2,450-2,100 million years ago (Ma) (Tomitani *et al.*, 2006), and recent relaxed molecular clocks based on geochemical and fossil records date the appearance of the Nostocales to between 2,100-2,130 Ma (Blank & Sánchez-Baracaldo, 2010). Based on these estimated dates, a simple molecular clock based on 16S rDNA distances among cyanobacteria estimated the appearance of *C. raciborskii* and *R. brookii* as between 121-135 Ma, and before the differentiation of *A. variabilis* ATCC 27413 and *Anabaena* between 80-90 Ma (Figure 3.1). The close relationship between CS-505 and D9 reflects their recent evolutionary separation, and thereby did not allow calculation of their evolutionary diversification time.

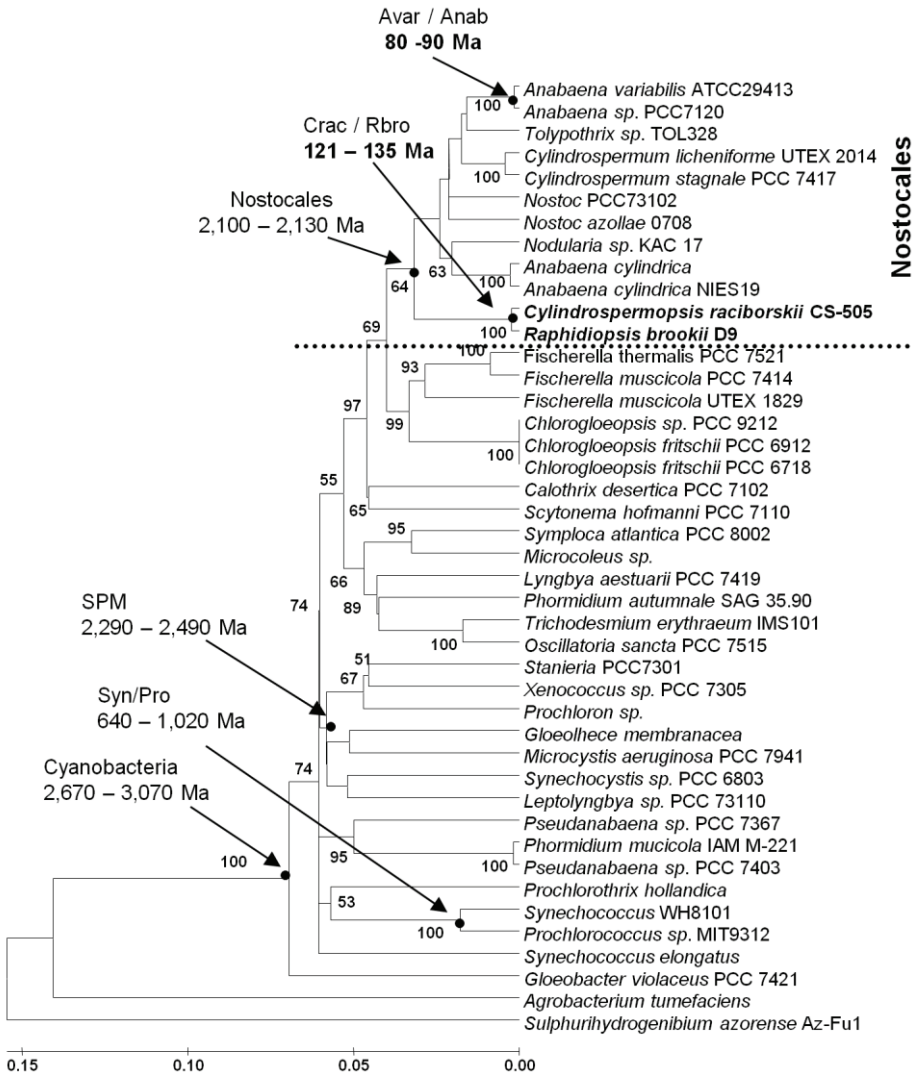


Figure 3.1 Evolutionary divergences of cyanobacteria. A linearized tree based on 16S rDNA sequences is shown. Ages from the nodes indicated by dots were retrieved from Blank and Sánchez-Baracaldo (2010). A calibration curve was determined for the minimum and maximum ages and evolutionary diversification of the *C. raciborskii* CS-505 and *R. brookii* D9 (Crac/Rbro) node was calculated. SPM: cyanobacterial clade containing *Synechocystis*, *Microcystis,* and species from the genus Pleurocapsa. Nucleotide sequences were aligned using clustalX2 (Higgins & Sharp, 1988, Larkin *et al.*, 2007) and trees were created with MEGA, using the Neighbor Joining algorithm (NJ) (Saitou & Nei, 1987).

Regarding the evolution of *C. raciborskii* CS-505 and *R. brookii* D9, this thesis provided the following findings:

i. High plasticity in the CS-505 genome may explain the heterogeneity found at the genome level among *C. raciborskii* strains. The same differences in genomic plasticity and synteny would be a signature of genome reduction in D9.
ii. Although the evolutionary time of divergence of CS-505 and D9 cannot be resolved, it is possible to infer that the species complex evolved before *Anabaena*.
iii. Similarity of the genomes, given by a high conservation of the shared genes among the two strains, strongly supports the necessity of a taxonomic reclassification of *Raphidiopsis* within the genus *Cylindrospermopsis.*

3.1.1 Minimal genomes for filamentous cyanobacteria

A genomic structure shaped by gene losses is highly consistent with the small size (3.2 Mb) of the D9 genome, which is much smaller than that known for other sequenced free-living filamentous cyanobacteria (5.3 – 8.2 Mb), and even when compared with that of the symbiont *N. azollae* (5.0 Mb) The genome size of *C. raciborskii* CS-505 (3.9 Mb) is somewhat larger than that of D9, but together these strains represent the smallest known genomes among filamentous cyanobacteria. The reasons of the small genome size are unknown, but in addition to the genome reduction indicated by the D9 genomic structure, the high number of plasticity elements in the CS-505 genome may also explain some genome reduction.

The genomes of *R. brookii* D9 and *C. raciborskii* CS-505 may be considered as minimal genomes and thus further studies will aid in the field of comparative genomics. By comparative genomic analysis a core set of genes was described for N_2 fixation and the multicellular traits, filament formation and heterocyst differentiation (**Publication II**). A remarkable example of the benefits of considering these minimal genomes for comparative genomics is inherent in the core gene set for heterocystous cyanobacteria. From 149 genes common to all analyzed members of Nostocales, the inclusion of *C. raciborskii* CS-505 reduced the number of core genes for this trait to only 41, a more feasible number for further studies on determining their involvement in this phenotype.

3.2 Nitrogen as a trigger of cell differentiation

The mechanisms whereby cyanobacteria respond to nitrogen (N) availability have been extensively studied in both non-diazotrophic and diazotrophic (including heterocyst-forming) cyanobacteria. In heterocystous cyanobacteria, the effect of shifts in N availability may be dual, in that in addition to the rapid response observed in all cyanobacteria, a second effect, triggered by depletion of N, involves a complex network of regulatory cascades that interconnect N control and heterocyst-specific genes. The physiological and transcriptomic responses of CS-505 and D9 to the combined N sources (urea, nitrate and ammonium) showed that both strains respond as expected with respect to the first effect. In general, although the strains did not differ in their growth under alternative N sources, an equivalent concentration of ammonium produced an inhibitory effect on the growth of D9. Urea, apparently taken up rapidly by both strains, enhanced the energetic metabolism of D9, and elicited repression of N control genes to the same degree as ammonium. The responses to N deprivation, on the other hand, triggered the second effect of cell differentiation in CS-505 as expected, but in D9 initiated a large transcriptomic response (involving 28% of its gene content) from which some genes could be related to cell differentiation.

3.2.2 Remnants of a past diazotrophy

Raphidiopsis brookii D9 cannot grow diazotrophically. Morphological taxonomy, however, classifies the genus *Raphidiopsis* within the order Nostocales, as an exception of the rule for heterocyst formation. Single marker phylogeny and phylogenomics as reported in this thesis (**Publication I, II**) support the taxonomic classification of *R. brookii* D9 within the Nostocales.

The morphological changes leading to the development of a terminal heterocyst in *Cylindrospermopsis* start from the sharpening and later thickening of the apical cell (Hindák, 1988, Horecká & Komárek, 1979, Singh, 1962). Based on the structure of the apical cell in D9, it seems that this structure may represent the first stage of heterocyst differentiation that did not develop further.

Among the genes involved in the development and maturation of a heterocyst, three groups are organized as cluster units in the genome of *C. raciborskii* CS-505: the genes of the nitrogenase enzymatic complex (*nif*) and related genes, and the glycolipid and polysaccharide gene clusters. While the heterocyst polysaccharide gene cluster is present in D9 and shows a structural organization (synteny) identical to its homologues in CS-505 and *Anabaena*, both *nif* and glycolipid gene clusters are absent in the D9 genome. The flanking regions of these two clusters in CS-505 show, however, synteny with the D9 genomic structure. These morphological and genetic features provide evidence that the ancestor of D9 had a diazotrophic and heterocystous phenotype and that gene losses (probably independent) were responsible for the loss of both phenotypes, although is not possible to infer the chronology of those events. In addition to the heterocyst polysaccharide gene cluster, several other genes that have been experimentally demonstrated as involved in heterocyst differentiation are encoded by the D9 genome. Among them, D9 has 48 (60%) but CS-505 has 64 (80%). With the 80% of the heterocyst-specific genes, CS-505 can already develop a functional heterocyst; therefore if we consider this number as the total, 75% of heterocyst specific genes are present in the D9 genome. Given the general functions of, for example, all transcriptional and posttranscriptional regulators, including all sigma factors (*nrrA, conR, devH, devR, henR, hepN, hetR, hanA, pknD, prpJ, prpS*), and the polysaccharide biosynthesis genes (*hepA, hepB, hepC, hetI, hgdABC* and *hglK* plus an unnamed gene cluster described in *Anabaena* sp. PCC 7120), these genes could be easily recycled to general metabolism. Genes from the unnamed gene cluster were expressed in D9 under growth with nitrate and, furthermore, when D9 filaments were stained with alcian blue, a faint staining was observed in the outside layer along the complete filaments, indicating that the genes fulfill functions in D9 cells (**Publication II**). Heterocyst-specific genes whose encoded proteins do not present any similarities to conserved domains or whose predicted functions are described as general are difficult to associate with metabolic processes in D9, other than heterocyst differentiation (*hetF, hetZ, patU3, abp1, abp2, hetP-like, hetY, patN*). Remarkably, when D9 cells were depleted of combined N over 24 h, the expression of some of these "remnant genes" was modified, in some cases, even when under the same conditions, in CS-505 there were no expression changes or the regulation was contrary (**Publication IV**). Of particular interest was *hetR*, induced in the absence of combined N. HetR, a peptidase with an essential role in heterocyst development (Buikema & Haselkorn, 2001), has also been related to akinete differentiation since mutants of this gene in *Nostoc ellipsosporum* block heterocyst and akinete

differentiation (Leganés *et al.*, 1994). A possibility is therefore that *hetR* was associated with both processes in the ancestor of *R. brookii* D9 and was maintained for akinete differentiation. The same conclusion can be drawn for the transcriptional regulator *nrrA*, which activates the expression of *hetR*, and was induced under N-deprivation. Additional information on possible recycled functions for heterocyst-specific genes comes from the genomic comparison and expression pattern of *hetZ* and *patU3*. These genes were recently assigned as coordinators of heterocyst pattern and differentiation in *Anabaena*, both located in clusters but fulfilling a different function (Zhang *et al.*, 2009). Comparative genomics located both genes as part of the core gene set of filamentous cyanobacteria (**Publication II**), and, in contrast to CS-505, for which *patU3* was induced and *hetZ* remained unchanged under N-deprivation, both genes were repressed in D9, strongly suggesting their role in filament formation.

The secondary loss of a functional heterocyst and of N_2 fixation in *R. brookii* D9 was evidenced in this thesis by the following arguments:

i. The close evolutionary relationship between *R. brookii* D9 and *C. raciborskii* CS-505 as members of the order Nostocales
ii. The genomic structure of D9, which was shaped by gene losses, in some cases of complete gene clusters, such as those for nitrogenase and heterocyst
iii. Most genes experimentally linked to heterocyst differentiation are present in the D9 genome, and some of the fundamental genes (*hetR, nrrA*) for this process are expressed under N-deprivation

3.2.1 Towards differentiation of a terminal heterocyst

This thesis provided insights into understanding the development of a terminal heterocyst in *C. raciborskii* CS-505. The genomic information revealed that 17 genes whose function has been related with heterocyst differentiation in *Anabaena* are absent in CS-505 (**Publication II**). Half of these absent genes are involved in glycolipid and polysaccharide synthesis and deposition, therefore implying differences in the composition and structure of the heterocyst between species. Four absent genes (*hetC*, *hetN*, *hetL* and *ccbP*) have a clear function in heterocyst pattern formation, in which mutants of all except *ccbP* present a phenotype of multiple contiguous heterocysts (Mch) (Xu & Wolk, 2001, Callahan & Buikema, 2001, Liu & Golden,

2002, Zhao *et al.*, 2005). Another six absent genes are part of signaling pathways acting as serine/threonine kinases or phosphatases *pkn22*, *pkn41* and *pkn42, pknA, pknE* and *prpA*. Mutants of these last three genes still develop heterocysts (Zhang, 1993, Zhang & Libs, 1998, Jang *et al.*, 2007, Cheng *et al.*, 2006). Due to the general character of their functions, they can easily be substituted by additional kinases or phosphatases encoded by the CS-505 genome.

The role of NtcA is fundamental for growth under N-deficiency, and the regulatory interconnections between this gene and *hetR* are the first steps to heterocyst differentiation. One of the most intriguing findings in this thesis work was the regulation timing of *ntcA* in CS-505. It is well known in *Anabaena* and other cyanobacteria that the expression of *ntcA* occurs early (within the first 12 h) after N step-down (Lindell & Post, 2001, Ehira *et al.*, 2003, Tolonen *et al.*, 2006). Absolute quantification of *ntcA* transcripts taken from daily samples over five days showed that transcription was induced 48 to 72 h after the removal of combined N (**Publication IV**). The late expression of *ntcA* was not the result of a second induction, since dynamics of short term gene expression (1-48 h) corroborated that *ntcA* transcripts were not accumulated before 48 h (**Publication V**). These experiments showed, however, that increased transcription of NtcA gene targets was achieved following the first hour after N step-down (i.e. *amt1, glnA, glnB, nrrA* genes), occurring as a consequence of the basal amounts of NtcA produced in the vegetative cells (quantified by qPCR). As high levels of NtcA (higher 2-oxo levels) are also necessary to trigger the expression of a second response of heterocyst differentiation genes, the induction of most early genes correlated well with the dynamics of *ntcA* gene expression that occured between 24-48 h. However, *hetR* did not increase in expression during the 48 h of study, although this gene typically responds between 30 min and 8 h after retrieval of N in *Anabaena* (Buikema & Haselkorn, 2001, Ehira *et al*., 2003). In accordance with the gene expression data, the frequency of heterocysts duplicated after 48 h in N-depleted CS-505 cultures, whereas the nitrogenase activity (estimated as acetylene reduction) already doubled after 24 h. The rise of N_2 fixation correlated with the increased expression of the *nifHDK* genes during the first 8 h after N step-down.

Heterocyst differentiation was delayed, but degradation of photosynthetic pigments and ATP synthesis was triggered after 24 h, which corresponded to a C:N ratio twice as high as at time zero. This response remained high after 48 h. However, 48 h seems to be a transition time at which cells started to return to normality (defined by active metabolism, i.e. photosynthesis, CO_2

fixation and protein synthesis) because *glnA, amt1, nrrA, urtA* decreased in induction intensity and *glnB* and ribosomal proteins returned to initial values.

The delayed differentiation of new heterocysts is probably related to a higher storage capacity of nitrogenated compounds in CS-505, rather than to the constant production of ammonium by nitrogenase activity of heterocysts in nitrate-containing medium. This is supported by the fact that the period of differentiation of new heterocysts is similar when cells are transferred from nitrate- or ammonium-rich to N-depleted medium (A. Muñoz, personal communication).

Regarding the responses of CS-505 to N-deprivation, the following consequences were uncovered in this thesis work:

i. The formation of a terminal heterocyst in CS-505 is determined by the lost of genes responsible for pattern formation in *Anabaena*
ii. As a rapid response to N-deprivation, genes responsible for transport and metabolism are induced and N_2 fixation activated, however, heterocyst differentiation is not induced until 48 h
iii. The delay of heterocyst differentiation is evidenced by a late transcription of *ntcA* that is followed by its target genes in heterocyst differentiation

3.3 Toxin biosynthesis

3.3.1 Comparative analysis

With current knowledge on diversification of toxin production in cyanobacteria, it is not possible to correlate the production of a certain group of toxins, or even a particular toxin analogue, with a given cyanobacterial species. For example, despite the close phylogenetic relationship between *R. brookii* D9 and *C. raciborskii* T3 (99.8% identity in 16S rDNA sequences), their toxin profiles comprise in common the biosynthesis of STX, but also different PST analogues (dcSTX, GTX2/3 and dcGTX2/3 in D9 versus NEO in T3). The genetic structure of the saxitoxin gene clusters (*sxt*) differs by an additional 11 genes in T3; one of these genes *sxtX*, was specifically found in NEO-producing cyanobacteria (Kellmann *et al.*, 2008a). The D9

sxt cluster contains the *sxtSUL* and *sxtDIOX* genes as unique elements, encoding for a sulfotransferase and a dioxygenase, respectively. Their gene products were proposed to be responsible for biosynthesis of the analogues GTX2/3, and an elongation of the STX biosynthesis pathway was hypothesized for *R. brookii* D9 (**Publication III**). The correct functional classification of both genes was strongly supported by their presence in other GTX2/3-producing cyanobacteria. This correlation between gene products and PST profile was observed in D9, as well as in three other species of toxigenic cyanobacteria. It is therefore clear that the toxigenic phenotype is derived from the genetic structure of the *sxt* cluster. Furthermore, the flanking regions of *sxt* clusters are completely different in all but two strains of *Anabaena circinalis* (AWQC131C and ACMB13), indicating that the cluster was inserted in each species in a different region of the genome, or alternatively, that the cluster was moved within the genome by genomic rearrangements. Among all reported *sxt* clusters, that of D9 is the smallest known that remains functional, providing another example of genetic reduction events in the D9 genome.

The mosaic structure of the *sxt* gene cluster, specifically the presence of unique profile-related genes, will allow for improvements in the prediction of toxigenicity and hence in monitoring of putatively toxic cyanobacterial blooms.

3.3.2 Regulation

The genetic structure of the putative toxin gene clusters was elucidated by comparing the genomes of *C. raciborskii* CS-505 and *R. brookii* D9 with the described gene clusters for cylindrospermopsin (*cyr*) and saxitoxin (*sxt*) (Kellmann *et al*., 2008a, Mihali *et al.*, 2008). This success nevertheless raises the questions of if and how these genes are regulated. A common feature among both gene clusters is the presence of genes known to be targets for regulation of NtcA, located in the cluster flanking regions, and of putative binding boxes for NtcA within the *cyr* and *sxt* clusters. This implies a possible regulation of toxin production by N availability (**Publication III, IV**). This finding, in addition to the significance of N_2 fixation as a trait separating CS-505 from D9, led in this thesis work to the investigation of an N-dependent regulatory mechanism of toxin biosynthesis. Knowledge of toxin regulation in cyanobacteria is poor and often contradictory. Nevetheless, one aspect that has been shown to be fairly consistent in the literature is the negative effect of nutrient stress on toxin biosynthesis, as found for N (and light) and P (as phosphate) on CYN production in *Aphanizomenon ovalisporum* (Shalev-Malul *et*

al., 2008) and *C. raciborskii* (Dyble *et al.*, 2006), and for N and P on PST production in *Lyngbya wollei* (Yin *et al.*, 1997). In this thesis study, stress conditions rather than N stimuli were shown to negatively affect toxin production in both CS-505 and D9, contributing knowledge on stress factors affecting toxin production independent of the toxin analogue and/or organism (**Publication IV**). Furthermore, among all synthesized analogues, the proposed precursors doCYN and STX were drastically reduced in intracellular concentration, with STX even decreasing down to the limit of detection when D9 cultures were grown with ammonium as N source. This effect, however, could not be related to transcriptional regulation of the biosynthetic genes because absolute quantification did not show repression values that may explain the reduction in toxin precursors. We therefore hypothesize that post-transcriptional regulation plays a fundamental role in the early response to stress-arrested toxin production in *C. raciborskii* CS-505 and *R. brookii* D9.

.

This thesis provided insights into biosynthesis and regulation of toxin production in *C. raciborskii* CS-505 and *R. brookii* D9:

i. Two genes, *sxtSUL* and *sxtDIOX* were proposed to determine biosynthesis of GTX2/3 analogues in *R. brookii* D9 and in other cyanobacteria. These genes may be exploited as monitoring tools to discriminate among PST profiles in cyanobacteria.
ii. The production of CYN and PST in CS-505 and D9, respectively, are negatively affected by N stress; at growth rates at or close to zero, production of the proposed toxin precursors is rapidly arrested
iii. Independent of the N source or growth rate, at early stages of the growth cycle in culture, there is no transcriptional regulation of either *cyr* or *sxt* genes.

3.4 Perspectives of future research

The work performed in this thesis represents the foundation for understanding phenotypic and genotypic differentiation between the closely related species *Cylindrospermopsis raciborskii* and *Raphidiopsis brookii*. At present, the morphological methods employed for discrimination of the genera *Cylindrospermopsis* and *Raphidiopsis* require isolation and following growth stages of single filaments - a long and complicated procedure. Furthermore, the morphological classification is not supported by molecular genotypic analysis; on the contrary, molecular data

situates both genera within a species complex. Strong supporting data collected in the course of this PhD thesis work also supports transfer of *Cylindrospermopsis raciborskii* and *Raphidiopsis brookii* from two different genera into two species within the same genus. The period during which species were defined on the basis of 16S rRNA sequences and DNA-DNA renaturation kinetics is being progressively replaced by genome comparisons. The Average Nucleotide Identity (ANI) percentage between the shared genes of a genome pair is easily computed. And indicates that the ANI between *C. raciborskii* CS-505 and *R. brookii* D9 (93.5%) is slightly below the threshold value for species differentiation. It is critical to compile the findings of this thesis together with the molecular data gathered for other species of *Raphidiopsis* in order to define a new taxon that includes these species within a redefined genus *Cylindrospermopsis*.

The description of the two smallest genomes of free-living and multicellular cyanobacteria brings unanswered questions regarding to the evolutionary pathway followed by *C. raciborskii* CS-505 and *R. brookii* D9. It is clear that both strains evolved from a common ancestor, but did their ancestor possess a large genome? If so, in the course of evolution did both CS-505 and D9 undergo genomic reduction, at higher rates for D9 than CS-505? If this is the case, what are the selective pressures leading to gene losses? A similar example of genome reduction was recently described in the fern-symbiotic cyanobacterium *Nostoc azollae* (genome size of 5.1 Mb), which forms filaments and develops heterocysts. The successful relationship of this symbiont pair is due to the fact that *N. azollae* provides fixed nitrogen in the heterocysts to the fern plant, while the fern provides carbon compounds. The presence of heterocysts in CS-505 also gives it potential as a symbiotic cyanobacterium, but what benefits would *R. brookii* D9 provide to a symbiotic partner? Evidence from phylogenomic analysis suggests an evolutionary developmental stage of D9 derived from CS-505 in which D9 has lost the ability for N_2 fixation. In order to understand genome reduction as the consequence of a putative symbiotic partnership it is necessary to study natural populations of *C. raciborskii* and its possible hosts.

From another perspective, the evolutionary path separating *C. raciborskii* and *Raphidiopsis* spp. can deciphered by a comparative study on the genomic structure of isolates from both genera differing for example in geographic location and/or toxin phenotype. DNA from these isolates could be hybridized against that of CS-505 as a base genome for the design of genomic arrays. This comparison would clearly describe a core genome for the *C. raciborskii/Raphidiopsis* species complex that may be followed over different geographical locations. Based on the rates of evolution estimated for the different genes, a relationship between

gene conservation and environmental parameters would reveal gene groups that may be related to the biogeography and global success of these species.

As a result of genome reduction, gene redundancies are commonly eliminated and complete pathways are lost. In order to maintain an undersized genome an efficient regulation of the small genetic pool is necessary. Non-coding small RNAs, located in intergenic regions, have been proven to participate in the efficient regulation of small genomes, e.g. in the *Prochlorococcus* genome - the smallest for unicellular cyanobacteria. The expression of repeat sequences, whose functions remain unknown, may have a regulatory role in CS-505. In order to further explore the regulation of these two genomes it is necessary to develop new microarrays targeting intergenic regions. In addition, errors in the sequence probably derived from inefficient transcription, as represented in D9 by the presence of stop codons within conserved genes such as *nirC* and *narB* needs to be further elucidated.

The coding potential of D9 for heterocyst development genes and for developing a heterocyst polysaccharide layer while not being able to synthesize the glycolipid or fix N_2, opened several questions regarding the role of heterocyst development genes in D9. A complicated mechanism of transcriptional regulation characterizes heterocyst development in *Anabaena*. The presence of several transcription start points (tsp) ensures the basal or enhanced transcription depending on the availability of N. Determining the start of transcription would allow on the one hand a better understanding of the delayed responses of CS-505 to heterocyst differentiation in comparison with *Anabaena*. On the other hand, it would also illuminate the function of these genes in N-metabolism in D9. Furthermore, cellular localization of important proteins such as those encoded by *hetR, patU3, hetZ* and *nrrA* in D9, is necessary to understand whether or not they are part of recycled products of the heterocyst remnant or if they fulfill roles in other cell differentiation processes such as akinete development, cell division, or cell death.

4. REFERENCES

Abouelhoda, M. I., S. Kurtz & E. Ohlebusch, (2004) Replacing suffix trees with enhanced suffix arrays. *J. of Discrete Algorithms* **2**: 53-86.

Adams, D. G. & P. S. Duggan, (1999) Tansley Review No. 107. Heterocyst and akinete differentiation in cyanobacteria. *New Phytol* **144**: 3-33.

Ahlgren, N. A. & G. Rocap, (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* **72**: 7193-7204.

Akama, T., K. Suzuki, K. Tanigawa, A. Kawashima, H. Wu, N. Nakata, Y. Osana, Y. Sakakibara & N. Ishii, (2009) Whole-genome tiling array analysis of *Mycobacterium leprae* RNA reveals high expression of pseudogenes and noncoding regions. *J Bacteriol* **191**: 3321-3327.

Aldea, M. R., R. A. Mella-Herrera & J. W. Golden, (2007) Sigma factor genes *sigC*, *sigE*, and *sigG* are upregulated in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **189**: 8392-8396.

Alduina, R., S. De Grazia, L. Dolce, P. Salerno, M. Sosio, S. Donadio & A. M. Puglia, (2003) Artificial chromosome libraries of *Streptomyces coelicolor* A3(2) and *Planobispora rosea*. *FEMS Microbiol Lett* **218**: 181-186.

Alster, A., R. N. Kaplan-Levy, A. Sukenik & T. Zohary, (2009) Morphology and phylogeny of a non-toxic invasive *Cylindrospermopsis raciborskii* from a Mediterranean Lake. *Hydrobiologia* **639**: 115-128.

Anderson, D. M., D. M. Kulis, J. J. Sullivan, S. Hall & C. Lee, (1990) Dynamics and physiology of saxitoxin production by the dinoflagellates Alexandrium spp. *Mar. Biol.* **104**: 511-524.

Appel, J., S. Phunpruch, K. Steinmuller & R. Schulz, (2000) The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis. *Arch Microbiol* **173**: 333-338.

Aras, R. A., J. Kang, A. I. Tschumi, Y. Harasaki & M. J. Blaser, (2003) Extensive repetitive DNA facilitates prokaryotic genome plasticity. *Proc Natl Acad Sci U S A* **100**: 13579-13584.

Arrigo, K. R., (2005) Marine microorganisms and global nutrient cycles. *Nature* **437**: 349-355.

Atkinson, M. R., E. S. Kamberov, R. L. Weiss & A. J. Ninfa, (1994) Reversible uridylylation of the *Escherichia coli* PII signal transduction protein regulates its ability to stimulate the dephosphorylation of the transcription factor nitrogen regulator I (NRI or NtrC). *J Biol Chem* **269**: 28288-28293.

Ausubel, F. M., R. Brent, R. Kingston, D. Moore, J. G. Seidman, J. A. Smith & K. Struhl, (1992) Current protocols in molecular biology. In. New York: Greene Publishing Associates and Wiley-Inter-science, pp. 2.1.1–2.4.5.

Awai, K. & C. P. Wolk, (2007) Identification of the glycosyl transferase required for synthesis of the principal glycolipid characteristic of heterocysts of *Anabaena* sp. strain PCC 7120. *FEMS Microbiol Lett* **266**: 98-102.

Azov, Y. & J. C. Goldman, (1982) Free ammonia inhibition of algal photosynthesis in intensive cultures. *Appl Environ Microbiol* **43**: 735-739.

Babica, P., L. Blaha & B. Marsalek, (2006) Exploring the natural role of microcystins - A review of effects on photoautotrophic organisms. *J Phycol* **42**: 9-20.

Bácsi, I., G. Vasas, G. Surányi, M. M. Hamvas, C. Máthé, E. Tóth, I. Grigorszky, A. Gáspár, S. Tóth & G. Borbely, (2006) Alteration of cylindrospermopsin production in sulfate- or phosphate-starved cyanobacterium *Aphanizomenon ovalisporum*. *FEMS Microbiol Lett* **259**: 303-310.

Baier, K., H. Lehmann, D. P. Stephan & W. Lockau, (2004) NblA is essential for phycobilisome degradation in *Anabaena* sp. strain PCC 7120 but not for development of functional heterocysts. *Microbiology* **150**: 2739-2749.

Ballot, A., J. Fastner & C. Wiedner, (2010) Paralytic shellfish poisoning toxin-producing cyanobacterium *Aphanizomenon gracile* in northeast Germany. *Appl Environ Microbiol* **76**: 1173-1180.

Banker, R., B. Teltsch, A. Sukenik & S. Carmeli, (2000) 7-Epicylindrospermopsin, a toxic minor metabolite of the cyanobacterium *Aphanizomenon ovalisporum* from lake Kinneret, Israel. *J Nat Prod* **63**: 387-389.

Bassam, B. J. & P. M. Gresshoff, (2007) Silver staining DNA in polyacrylamide gels. *Nat Protoc* **2**: 2649-2654.

Bauer, C. C., K. S. Ramaswamy, S. Endley, L. A. Scappino, J. W. Golden & R. Haselkorn, (1997) Suppression of heterocyst differentiation in *Anabaena* PCC 7120 by a cosmid carrying wild-type genes encoding enzymes for fatty acid synthesis. *FEMS Microbiol Lett* **151**: 23-30.

Bergman, B. & E. J. Carpenter, (1991) Nitrogenase confined to randomly distributed trichomes in the marine cyanobacterium *Trichodesmium thiebautii*. *J Phycol* **27**: 158-165.

Berman-Frank, I., P. Lundgren, Y. B. Chen, H. Kupper, Z. Kolber, B. Bergman & P. Falkowski, (2001) Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* **294**: 1534-1537.

Berry, J. P., M. Gantar, M. H. Perez, G. Berry & F. G. Noriega, (2008) Cyanobacterial toxins as allelochemicals with potential applications as algaecides, herbicides and insecticides. *Mar Drugs* **6**: 117-146.

Bisgard, K. M., C. D. C. Christie, S. F. Reising, G. N. Sanden, P. K. Cassiday, C. Gomersall, W. A. Wattigney, N. E. Roberts & P. M. Strebel, (2001) Molecular epidemiology of *Bordetella pertussis* by pulsed-field gel electrophoresis profile: Cincinnati, 1989-1996. *J Infect Dis* **183**: 1360-1367.

Black, K., W. J. Buikema & R. Haselkorn, (1995) The *hglK* gene is required for localization of heterocyst-specific glycolipids in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **177**: 6440-6448.

Blank, C. E. & P. Sánchez-Baracaldo, (2010) Timing of morphological and ecological innovations in the cyanobacteria: a key to understanding the rise in atmospheric oxygen. *Geobiology* **8**: 1-23.

Bolhuis, H., I. Severin, V. Confurius-Guns, U. I. A. Wollenzien & L. J. Stal, (2009) Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium *Microcoleus chthonoplastes*. *ISME J* **4**: 121-130.

Bourke, A. T. C., R. B. Hawes, A. Neilson & N. D. Stallman, (1983) An outbreak of hepato-enteritis (the Palm Island mystery disease) possibly caused by algal intoxication. *Toxicon* **21**: 45-48.

Bouvy, M., R. Molica, S. De Oliveira, M. Marinho & B. Beker, (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**: 285-297.

Boyer, S. L., V. R. Flechtner & J. R. Johansen, (2001) Is the 16S-23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria. *Mol Biol Evol* **18**: 1057-1069.

Buikema, W. J. & R. Haselkorn, (1991) Characterization of a gene controlling heterocyst differentiation in the cyanobacterium *Anabaena* 7120. *Genes Dev* **5**: 321-330.

Buikema, W. J. & R. Haselkorn, (2001) Expression of the *Anabaena hetR* gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. *Proc Natl Acad Sci U S A* **98**: 2729-2734.

Cai, Y. & C. P. Wolk, (1997) *Anabaena* sp. strain PCC 7120 responds to nitrogen deprivation with a cascade-like sequence of transcriptional activations. *J Bacteriol* **179**: 267-271.

Callahan, S. M. & W. J. Buikema, (2001) The role of HetN in maintenance of the heterocyst pattern in *Anabaena* sp. PCC 7120. *Mol Microbiol* **40**: 941-950.

Campbell, E. L., M. F. Cohen & J. C. Meeks, (1997) A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. *Arch Microbiol* **167**: 251-258.

Campbell, E. L., K. D. Hagen, M. F. Cohen, M. L. Summers & J. C. Meeks, (1996) The *devR* gene product is characteristic of receivers of two-component regulatory systems and is essential for heterocyst development in the filamentous cyanobacterium *Nostoc* sp. strain ATCC 29133. *J Bacteriol* **178**: 2037-2043.

Campbell, E. L., M. L. Summers, H. Christman, M. E. Martin & J. C. Meeks, (2007) Global gene expression patterns of *Nostoc punctiforme* in steady-state dinitrogen-grown heterocyst-containing cultures and at single time points during the differentiation of akinetes and hormogonia. *J Bacteriol* **189**: 5247-5256.

Capone, D. G., (1993) Determination of nitrogenase activity in aquatic samples using the acetylene reduction procedure. In: Handbook of Methods in Aquatic Ecology. P. F. Kemp, B. F. Sherr, E. B. Sherr & J. J. Cole (eds). Lewis Publishers, pp. 621–631.

Capone, D. G., J. A. Burns, J. P. Montoya, A. Subramaniam, C. Mahaffey, T. Gunderson, A. F. Michaels & E. J. Carpenter, (2005) Nitrogen fixation by *Trichodesmium* spp.: an important source of new nitrogen to the tropical and subtropical north Atlantic ocean. *Glob Biogeochem Cycle* **19**: 17.

Capone, D. G., J. P. Zehr, H. Paerl, B. Bergman & E. J. Carpenter, (1997) *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**: 1221-1229.

Carmichael, W. & R. Li, (2006) Cyanobacteria toxins in the Salton Sea. *Saline Systems* **2**: 5.

Carmichael, W. W., (1997) The cyanotoxins. *Adv Bot Res* **27**: 211 - 256.

Carmichael, W. W., W. R. Evans, Q. Q. Yin, P. Bell & E. Moczydlowski, (1997) Evidence for paralytic shellfish poisons in the freshwater cyanobacterium *Lyngbya wollei* (Farlow ex Gomont) comb. nov. *Appl Environ Microbiol* **63**: 3104-3110.

Carpenter, E. J. & K. Romans, (1991) Mayor role of the cyanobacterium *Trichodesmium* in nutrient cycling in the north-Atlantic ocean. *Science* **254**: 1356-1358.

Carrasco, C. D., J. A. Buettner & J. W. Golden, (1995) Programmed DNA rearrangement of a cyanobacterial *hupL* gene in heterocysts. *Proc Natl Acad Sci U S A* **92**: 791-795.

Carrasco, C. D. & J. W. Golden, (1995) Two heterocyst-specific DNA rearrangements of *nif* operons in *Anabaena cylindrica* and *Nostoc* sp. strain Mac. *Microbiology* **141**: 2479-2487.

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

Cembella, A., (1998) Ecophysiology and metabolism of paralytic shellfish toxins in marine microalgae. In: Physiological ecology of harmful algal blooms. D. M. Anderson, A. D. Cembella & G. M. Hallegraeff (eds). Heidelberg: Springer-Verlag, pp. 381–403.

Chellappa, N. T. & M. A. M. Costa, (2003) Dominant and co-existing species of cyanobacteria from a eutrophicated reservoir of Rio Grande do norte state, Brazil. *Acta Oecologica* **24**: S3-S10.

Chen, X. G. & W. R. Widger, (1993) Physical genome map of the unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. *J Bacteriol* **175**: 5106-5116.

Chen, Y.-B., B. Dominic, M. T. Mellon & J. P. Zehr, (1998) Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. strain IMS 101. *J Bacteriol* **180**: 3598-3605.

Cheng, Y., J. H. Li, L. Shi, L. Wang, A. Latifi & C. C. Zhang, (2006) A pair of iron-responsive genes encoding protein kinases with a Ser/Thr kinase domain and a His kinase domain are regulated by NtcA in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **188**: 4822-4829.

Chisholm, S. W., R. J. Olson, E. R. Zettler, R. Goericke, J. B. Waterbury & N. A. Welschmeyer, (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**: 340-343.

Cole, S. T., P. Supply & N. Honore, (2001) Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity. *Lepr Rev* **72**: 449-461.

Coleman, M. L., M. B. Sullivan, A. C. Martiny, C. Steglich, K. Barry, E. F. DeLong & S. W. Chisholm, (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768-1770.

De Silva, E. D., D. E. Williams, R. J. Andersen, H. Klix, C. F. B. Holmes & T. M. Allen, (1992) Motuporin, a potent potein phosphatase inhibitor isolated from the Papua New Guinea sponge *Theonella swinhoei* gray. *Tetrahedron Lett* **33**: 1561-1564.

Des Marais, D. J., (2000) EVOLUTION: when did photosynthesis emerge on Earth? *Science* **289**: 1703-1705.

Dias, E., P. Pereira & S. Franca, (2002) Production of paralytic shellfish toxins by *Aphanizonmenon* sp. LMECYA 31 (Cyanobacteria). *J Phycol* **38**: 705-712.

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

Dittmann, E., B. A. Neilan, M. Erhard, H. vonDohren & T. Borner, (1997) Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. *Mol Microbiol* **26**: 779-787.

Doolittle, W. F., (1999) Lateral genomics. *Trends Cell Biol* **9**: M5-M8.

Drath, M., N. Kloft, A. Batschauer, K. Marin, J. Novak & K. Forchhammer, (2008) Ammonia triggers photodamage of photosystem II in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Plant Physiol* **147**: 206-215.

Dyble, J., P. A. Tester & R. W. Litaker, (2006) Effects of light intensity on cylindrospermopsin production in the cyanobacterial HAB species *Cylindrospermopsis raciborskii*. *Afr J Mar Sci* **28**: 309-312.

Ehira, S. & M. Ohmori, (2006a) NrrA directly regulates expression of *hetR* during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **188**: 8520-8525.

Ehira, S. & M. Ohmori, (2006b) NrrA, a nitrogen-responsive response regulator facilitates heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol Microbiol* **59**: 1692-1703.

Ehira, S., M. Ohmori & N. Sato, (2003) Genome-wide expression analysis of the responses to nitrogen deprivation in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res* **10**: 97-113.

Elhai, J., M. Kato, S. Cousins, P. Lindblad & J. L. Costa, (2008) Very small mobile repeated elements in cyanobacterial genomes. *Genome Res* **18**: 1484-1499.

Ewing, B. & P. Green, (1998) Base-calling of automated sequencer traces using Phred.II. Error Probabilities. *Genome Res* **8**: 186-194.

Ewing, B., L. Hillier, M. C. Wendl & P. Green, (1998) Base-calling of automated sequencer traces using Phred.I. Accuracy Assessment. *Genome Res* **8**: 175-185.

Falconer, I. R., S. J. Hardy, A. R. Humpage, S. M. Froscio, G. J. Tozer & P. R. Hawkins, (1999) Hepatic and renal toxicity of the blue-green alga (cyanobacterium) *Cylindrospermopsis raciborskii* in male Swiss albino mice. *Environ Toxicol* **14**: 143-150.

Falconer, I. R. & A. R. Humpage, (2006) Cyanobacterial (blue-green algal) toxins in water supplies: cylindrospermopsins. *Environ Toxicol* **21**: 299-304.

Fan, Q., G. Huang, S. Lechno-Yossef, C. P. Wolk, T. Kaneko & S. Tabata, (2005) Clustered genes required for synthesis and deposition of envelope glycolipids in *Anabaena* sp. strain PCC 7120. *Mol Microbiol* **58**: 227-243.

Fan, Q., S. Lechno-Yossef, S. Ehira, T. Kaneko, M. Ohmori, N. Sato, S. Tabata & C. P. Wolk, (2006) Signal transduction genes required for heterocyst maturation in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **188**: 6688-6693.

Felsenstein, J., (1985) Confidence-limits on phylogenies- an approach using the bootstrap. *Evolution* **39**: 783-791.

Fernandez-Pinas, F., F. Leganes & C. P. Wolk, (1994) A third genetic locus required for the formation of heterocysts in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **176**: 5277-5283.

Ferrão Filho, A. d. S., S. M. da Costa, M. G. L. Ribeiro & M. F. O. Azevedo, (2008) Effects of a saxitoxin-producer strain of *Cylindrospermopsis raciborskii* (cyanobacteria) on the swimming movements of cladocerans. *Environ Toxicol* **23**: 161-168.

Fiedler, G., M. Arnold, S. Hannus & I. Maldener, (1998) The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol Microbiol* **27**: 1193-1202.

Flores, E. & A. Herrero, (2005) Nitrogen assimilation and nitrogen control in cyanobacteria. *Biochem Soc Trans* **33**: 164-167.

Forchhammer, K. & N. Tandeau de Marsac, (1995) Phosphorylation of the PII protein (*glnB* gene product) in the cyanobacterium *Synechococcus* sp. strain PCC 7942: analysis of in vitro kinase activity. *J Bacteriol* **177**: 5812-5817.

Francis, G., (1978) Poisonous Australian lake *Nature* **18**: 11-12.

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

Fredriksson, C. & B. Bergman, (1997) Ultrastructural characterisation of cells specialised for nitrogen fixation in a non-heterocystous cyanobacterium, *Trichodesmium* spp. *Protoplasma* **197**: 76-85.

Frias, J. E., E. Flores & A. Herrero, (2000) Activation of the *Anabaena nir* operon promoter requires both NtcA (CAP family) and NtcB (LysR family) transcription factors. *Mol Microbiol* **38**: 613-625.

Frias, J. E., A. Herrero & E. Flores, (2003) Open reading frame all0601 from *Anabaena* sp. strain PCC 7120 represents a novel gene, *cnaT*, required for expression of the nitrate assimilation *nir* operon. *J Bacteriol* **185**: 5037-5044.

Froscio, S. M., A. R. Humpage, P. C. Burcham & I. R. Falconer, (2003) Cylindrospermopsin-induced protein synthesis inhibition and its dissociation from acute toxicity in mouse hepatocytes. *Environ Toxicol* **18**: 243-251.

Frost, L. S., R. Leplae, A. O. Summers & A. Toussaint, (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**: 722-732.

Fuenzalida, L., (2005) Genetic and physiologic studies in the cyanobacterium *Cylindrospermopsis raciborskii*, a PSP-toxin producer. In: Facultad de ciencias. Santiago: Universidad de Chile.

García-Domínguez, M., J. C. Reyes & F. J. Florencio, (1999) Glutamine synthetase inactivation by protein-protein interaction. *Proc Natl Acad Sci U S A* **96**: 7161-7166.

Ge, X., K. Cain & R. Hirschberg, (1990) Urea metabolism and urease regulation in the cyanobacterium *Anabaena variabilis*. *Can J Microbiol* **36**: 218-222.

Gogarten, J. P., W. F. Doolittle & J. G. Lawrence, (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**: 2226-2238.

Golden, J. W., M. E. Mulligan & R. Haselkorn, (1987) Different recombination site specificity of two developmentally regulated genome rearrangements. *Nature* **327**: 526-529.

Griffiths, D. J. & M. L. Saker, (2003) The Palm Island mystery disease 20 years on: a review of research on the cyanotoxin cylindrospermopsin. *Environ Toxicol* **18**: 78-93.

Gugger, M., R. Molica, B. Le Berre, P. Dufour, C. Bernard & J. F. Humbert, (2005) Genetic diversity of *Cylindrospermopsis* strains (cyanobacteria) isolated from four continents. *Appl Environ Microbiol* **71**: 1097-1100.

Gugger, M. F. & L. Hoffmann, (2004) Polyphyly of true branching cyanobacteria (Stigonematales). *Int J Syst Evol Microbiol* **54**: 349-357.

Hall, T. A., (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95-98.

Hansel, A., R. Axelsson, P. Lindberg, O. Y. Troshina, R. Wunschiers & P. Lindblad, (2001) Cloning and characterisation of a hyp gene cluster in the filamentous cyanobacterium *Nostoc* sp. strain PCC 73102. *FEMS Microbiol Lett* **201**: 59-64.

Hara-Kudo, Y., K. Sugiyama, M. Nishibuchi, A. Chowdhury, J. Yatsuyanagi, Y. Ohtomo, A. Saito, H. Nagano, T. Nishina, H. Nakagawa, H. Konuma, M. Miyahara & S. Kumagai, (2003) Prevalence of pandemic thermostable direct hemolysin-producing *Vibrio parahaemolyticus* O3:K6 in seafood and the coastal environment in Japan. *Appl Environ Microbiol* **69**: 3883-3891.

Haselkorn, R., (1992) Developmentally regulated gene rearrangements in Prokaryotes. *Annu Rev Genet* **26**: 113-130.

Hashimoto, J. Q., B. S. Stevenson & T. M. Schmidt, (2003) Rates and consequences of recombination between rRNA operons. *J Bacteriol* **185**: 966-972.

Havens, K. E., (2008) Cyanobacteria blooms: effects on aquatic ecosystems. In: Cyanobacterial harmful algal blooms: State of the Science and Research Needs. H. K. Hudnell (ed). New York: Springer, pp. 733-747.

Herrero, A., A. M. Muro-Pastor & E. Flores, (2001) Nitrogen control in cyanobacteria. *J Bacteriol* **183**: 411-425.

Herrero, A., A. M. Muro-Pastor, A. Valladares & E. Flores, (2004) Cellular differentiation and the NtcA transcription factor in filamentous cyanobacteria. *FEMS Microbiol Rev* **28**: 469-487.

Hess, W. R., C. Steglich, C. Lichtle & F. Partensky, (1999) Phycoerythrins of the oxyphotobacterium *Prochlorococcus marinus* are associated to the thylakoid membrane and are encoded by a single large gene cluster. *Plant Mol Biol* **40**: 507-521.

Higgins, D. G. & P. M. Sharp, (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244.

Hindák, F., (1988) Planktic species of two related genera *Cylindrospermopsis* and *Anabaenopsis* from western Slovakia. *Arch. Hydrobiol. Suppl. 80(1–4), Algological Studies* **50–53**: 283–302.

Hobbie, J. E., R. J. Daley & S. Jasper, (1977) Use of nucleopore filters for counting bacteria by fluorescence microscopy. *Appl Environ Microbiol* **33**: 1225-1228.

Honda, D., A. Yokota & J. Sugiyama, (1999) Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *J Mol Evol* **48**: 723-739.

Horecká, M. & J. Komárek, (1979) Taxonomic position of three planktonic blue-green algae from the genera *Aphanizomenon* and *Cylindrospermopsis*. *Preslia* **51**: 289–312.

Huang, G., Q. Fan, S. Lechno-Yossef, E. Wojciuch, C. P. Wolk, T. Kaneko & S. Tabata, (2005) Clustered genes required for the synthesis of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **187**: 1114-1123.

Humpage, A. R., J. Rositano, A. H. Bretag, R. Brown, P. D. Baker, B. C. Nicholson & D. A. Steffensen, (1994) Paralytic shellfish poisons from Australian cyanobacterial blooms. *Aust J Mar Freshwat Res* **45**: 761-771.

Ishii, A. & Y. Hihara, (2008) An AbrB-like transcriptional regulator, Sll0822, is essential for the activation of nitrogen-regulated genes in *Synechocystis* sp. PCC 6803. *Plant Physiol* **148**: 660-670.

Istvánovics, V., H. M. Shafik, M. Presing & S. Juhos, (2000) Growth and phosphate uptake kinetics of the cyanobacterium, *Cylindrospermopsis raciborskii* (Cyanophyceae) in throughflow cultures. *Freshw Biol* **43**: 257-275.

Iteman, I., R. Rippka, N. T. de Marsac & M. Herdman, (2000) Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria. *Microbiology* **146**: 1275-1286.

Jang, J. C., L. Wang, R. Jeanjean & C. C. Zhang, (2007) PrpJ, a PP2C-type protein phosphatase located on the plasma membrane, is involved in heterocyst maturation in the cyanobacterium *Anabaena* sp. PCC 7120. *Mol Microbiol* **64**: 347-358.

Janson, S., A. Matveyev & B. Bergman, (1998) The presence and expression of *hetR* in the non-heterocystous cyanobacterium *Symploca* PCC 8002. *FEMS Microbiol Lett* **168**: 173-179.

Johnson, T. R., J. I. Haynes, 2nd, J. L. Wealand, L. R. Yarbrough & R. Hirschberg, (1988) Structure and regulation of genes encoding phycocyanin and allophycocyanin from *Anabaena variabilis* ATCC 29413. *J Bacteriol* **170**: 1858-1865.

Jones, K. M., W. J. Buikema & R. Haselkorn, (2003) Heterocyst-specific expression of *patB*, a gene required for nitrogen fixation in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **185**: 2306-2314.

Juhas, M., M. Jan Roelof van der, G. Muriel, M. H. Rosalind, W. H. Derek & W. C. Derrick, (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33**: 376-393.

Kakuta, Y., L. G. Pedersen, C. W. Carter, M. Negishi & L. C. Pedersen, (1997) Crystal structure of estrogen sulphotransferase. *Nat Struct Biol* **4**: 904-908.

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

Kaneko, T. & S. Tabata, (1997) Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol* **38**: 1171-1176.

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

Kellmann, R., T. K. Mihali & B. A. Neilan, (2008b) Identification of a saxitoxin biosynthesis gene with a history of frequent horizontal gene transfers. *J Mol Evol* **67**: 526 - 538.

Kellmann, R., T. Mills & B. A. Neilan, (2006) Functional modeling and phylogenetic distribution of putative cylindrospermopsin biosynthesis enzymes. *J Mol Evol* **62**: 267-280.

Kettler, G. C., A. C. Martiny, K. Huang, J. Zucker, M. L. Coleman, S. Rodrigue, F. Chen, A. Lapidus, S. Ferriera, J. Johnson, C. Steglich, G. M. Church, P. Richardson & S. W. Chisholm, (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* **3**: e231.

Khudyakov, I. & C. P. Wolk, (1996) Evidence that the *hanA* gene coding for HU protein is essential for heterocyst differentiation in, and cyanophage A-4(L) sensitivity of, *Anabaena* sp. strain PCC 7120. *J Bacteriol* **178**: 3572-3577.

Khudyakov, I. Y. & J. W. Golden, (2004) Different functions of HetR, a master regulator of heterocyst differentiation in *Anabaena* sp. PCC 7120, can be separated by mutation. *Proc Natl Acad Sci U S A* **101**: 16040-16045.

Kiss, T., A. Vehovszky, L. Hiripi, A. Kovacs & L. Voros, (2002) Membrane effects of toxins isolated from a cyanobacterium, *Cylindrospermopsis raciborskii*, on identified molluscan neurones. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology* **131**: 167-176.

Koksharova, O. A. & C. P. Wolk, (2002) Novel DNA-binding proteins in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **184**: 3931-3940.

Komárek, J. & K. Anagnostidis, (1989) Modern approach to the classification system of Cyanophytes. 4 - Nostocales. *Arch für Hydrobiol. Supp volumes* **82**: 247-345.

Konstantinidis, K. T. & J. M. Tiedje, (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**: 2567-2572.

Krogmann, D. W., A. T. Jagendorf & M. Avron, (1959) Uncouplers of spinach chloroplast photosynthetic phosphorylation. *Plant Physiol* **34**: 272-277.

Kufryk, G., I., M. Sachet, G. Schmetterer & W. Vermaas, F. J., (2002) Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency. *FEMS Microbiol Lett* **206**: 215-219.

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

Lander, E. S. & M. S. Waterman, (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**: 231-239.

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin & N. R. Pace, (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* **82**: 6955-6959.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & D. G. Higgins, (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Laurent, S., K. Forchhammer, L. Gonzalez, T. Heulin, C. C. Zhang & S. Bedu, (2004) Cell-type specific modification of PII is involved in the regulation of nitrogen metabolism in the cyanobacterium *Anabaena* PCC 7120. *FEBS Lett* **576**: 261-265.

Lawrence, J. F., B. Wong & C. Menard, (1996) Determination of decarbamoyl saxitoxin and its analogues in shellfish by prechromatographic oxidation and liquid chromatography with fluorescence detection. *J AOAC Int* **79**: 1111-1115.

Lawton, L., B. Marsalek, J. Padisák & I. Chorus, (1999) Determination of cyanobacteria in the laboratory. In: Toxic Cyanobacteria in Water: A guide to their public health consequences, monitoring and management. I. Chorus & J. Bartram (eds). London: E&FN Spon.

Lee, H.-M., E. Flores, A. Herrero, J. Houmard & N. Tandeau de Marsac, (1998) A role for the signal transduction protein PII in the control of nitrate/nitrite uptake in a cyanobacterium. *FEBS Lett* **427**: 291-295.

Leganés, F., F. Fernández-Piñas & P. C. Wolk, (1994) Two mutations that block heterocyst differentiation have different effects on akinete differentiation in *Nostoc ellipsosporum*. *Mol Microbiol* **12**: 679-684.

Li, R., W. W. Carmichael, S. Brittain, G. K. Eaglesham, G. R. Shaw, A. Mahakhant, N. Noparatnaraporn, W. Yongmanitchai, K. Kaya & M. M. Watanabe, (2001a) Isolation and identification of the cyanotoxin cylindrospermopsin and deoxy-cylindrospermopsin from a Thailand strain of *Cylindrospermopsis raciborskii* (cyanobacteria). *Toxicon* **39**: 973-980.

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

Li, R., S. W. Wilhelm, W. W. Carmichael & M. M. Watanabe, (2008) Polyphasic characterization of water bloom forming *Raphidiopsis* species (cyanobacteria) from central China. *Harmful Algae* **7**: 146-153.

Liang, J., L. Scappino & R. Haselkorn, (1992) The *patA* gene-product, which contains a region similar to CheY of *Escherichia coli*, controls heterocyst pattern-formation in the cyanobacterium *Anabaena* 7120. *Proc Natl Acad Sci U S A* **89**: 5655-5659.

Lilleheil, G., R. A. Andersen, O. M. Skulberg & J. Alexander, (1997) Effects of a homoanatoxin-a-containing extract from *Oscillatoria formosa* (cyanophyceae/cyanobacteria) on neuromuscular transmission. *Toxicon* **35**: 1275-1289.

Lindell, D., J. D. Jaffe, M. L. Coleman, M. E. Futschik, I. M. Axmann, T. Rector, G. Kettler, M. B. Sullivan, R. Steen, W. R. Hess, G. M. Church & S. W. Chisholm, (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83-86.

Lindell, D., E. Padan & A. F. Post, (1998) Regulation of *ntcA* expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J Bacteriol* **180**: 1878-1886.

Lindell, D., M. B. Sullivan, Z. I. Johnson, A. C. Tolonen, F. Rohwer & S. W. Chisholm, (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* **101**: 11013-11018.

Lipschultz, F. & N. J. P. Owens, (1996) An assessment of nitrogen fixation as a source of nitrogen to the North Atlantic Ocean. *Biogeochemistry* **35**: 261-274.

Liu, D. A. & J. W. Golden, (2002) *hetL* overexpression stimulates heterocyst formation in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **184**: 6873-6881.

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

Llewellyn, L. E., A. P. Negri, J. Doyle, P. D. Baker, E. C. Beltran & B. A. Neilan, (2001) Radioreceptor assays for sensitive detection and quantitation of saxitoxin and its analogues from strains of the freshwater cyanobacterium, *Anabaena circinalis*. *Environ Sci Technol* **35**: 1445-1451.

Lopez-Gomollon, S., J. A. Hernandez, C. P. Wolk, M. L. Peleato & M. F. Fillat, (2007) Expression of *furA* is modulated by NtcA and strongly enhanced in heterocysts of *Anabaena* sp. PCC 7120. *Microbiology* **153**: 42-50.

Lovett, S. T., T. J. Gluckman, P. J. Simon, V. A. Sutera & P. T. Drapkin, (1994) Recombination between repeats in *Escherichia coli* by a *recA*-independent, proximity-sensitive mechanism *Mol Gen Genet* **245**: 294-300.

Luque, I., E. Flores & A. Herrero, (1994) Molecular mechanism for the operation of nitrogen control in cyanobaceria. *EMBO J* **13**: 5794-5794.

Luque, I. & K. Forchhammer, (2008) Nitrogen assimilation and C/N balance sensing. In: The cyanobacteria: molecular biology, genomics and evolution. A. Herrero & E. Flores (eds). Norfolk: Caister Academic Press, pp. 335-382.

Mahmood, N. A. & W. W. Carmichael, (1986a) The pharmacology of anatoxin-a(s), a neurotoxin produced by the freshwater cyanobacterium *Anabaena flos-aquae* NRC 525-17. *Toxicon* **24**: 425-434.

Mahmood, N. A. & W. W. Carmichael, (1986b) Paralytic shellfish poisons produced by the freshwater cyanobacterium *Aphanizomenon flos-aquae* NH-5. *Toxicon* **24**: 175-186.

Maldener, I., S. Hannus & M. Kammerer, (2003) Description of five mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120 affected in heterocyst differentiation and identification of the transposon-tagged genes. *FEMS Microbiol Lett* **224**: 205-213.

Martin-Luna, B., E. Sevilla, J. A. Hernandez, M. T. Bes, M. F. Fillat & M. L. Peleato, (2006) Fur from *Microcystis aeruginosa* binds in vitro promoter regions of the microcystin biosynthesis gene cluster. *Phytochemistry* **67**: 876-881.

McGregor, G. B. & L. D. Fabbro, (2000) Dominance of *Cylindrospermopsis raciborskii* (Nostocales, cyanoprokaryota) in Queensland tropical and subtropical reservoirs: implications for monitoring and management. *Lakes & Reservoirs: Research and Management* **5**: 195-205.

Meeks, J. C. & J. Elhai, (2002) Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiol Mol Biol Rev* **66**: 94-121.

Mejean, A., S. Mann, T. Maldiney, G. Vassiliadis, O. Lequin & O. Ploux, (2009) Evidence that biosynthesis of the neurotoxic alkaloids anatoxin-a and homoanatoxin-a in the cyanobacterium *Oscillatoria* PCC 6506 occurs on a modular polyketide synthase initiated by L-proline. *J Am Chem Soc* **131**: 7512.

Mihali, T., R. Kellmann & B. Neilan, (2009) Characterization of the paralytic shellfish toxin biosynthesis gene clusters in *Anabaena circinalis* AWQC131C and *Aphanizomenon* sp. NH-5. *BMC Biochem* **10**: 8.

Mihali, T. K., R. Kellmann, J. Muenchhoff, K. D. Barrow & B. A. Neilan, (2008) Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. *Appl Environ Microbiol* **74**: 716-722.

Mitsui, A., S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao & T. Arai, (1986) Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature* **323**: 720-722.

Miyagishima, S.-y., C. P. Wolk & K. Osteryoung, W., (2005) Identification of cyanobacterial cell division genes by comparative and mutational analyses. *Mol Microbiol* **56**: 126-143.

Moffitt, M. C. & B. A. Neilan, (2004) Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Appl Environ Microbiol* **70**: 6353-6362.

Mohamed, Z. A., (2007) First report of toxic *Cylindrospermopsis raciborskii* and *Raphidiopsis mediterranea* (Cyanoprokaryota) in Egyptian fresh waters. *FEMS Microbiol Ecol* **59**: 749-761.

Montoya, J. P., C. M. Holl, J. P. Zehr, A. Hansen, T. A. Villareal & D. G. Capone, (2004) High rates of N_2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430**: 1027-1032.

Moreno-Paz, M. & V. Parro, (2006) Amplification of low quantity bacterial RNA for microarray studies: time-course analysis of *Leptospirillum ferrooxidans* under nitrogen-fixing conditions. *Environ Microbiol* **8**: 1064-1073.

Mori, S., A. Castoreno & P. J. Lammers, (2002) Transcript levels of *rbcR1*, *ntcA*, and *rbcL/S* genes in cyanobacterium *Anabaena* sp. PCC 7120 are downregulated in response to cold and osmotic stress. *FEMS Microbiol Lett* **213**: 167-173.

Moustafa, A., J. E. Loram, J. D. Hackett, D. M. Anderson, F. G. Plumley & D. Bhattacharya, (2009) Origin of saxitoxin biosynthetic genes in cyanobacteria. *PLoS One* **4**: e5758.

Moustaka-Gouni, M., K. A. Kormas, E. Vardaka, M. Katsiapi & S. Gkelis, (2009) *Raphidiopsis mediterranea* Skuja represents non-heterocytous life-cycle stages of *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju in lake Kastoria (Greece), its type locality: evidence by morphological and phylogenetic analysis. *Harmful Algae* **8**: 864-872.

Mulkidjanian, A. Y., E. V. Koonin, K. S. Makarova, S. L. Mekhedov, A. Sorokin, Y. I. Wolf, A. Dufresne, F. Partensky, H. Burd, D. Kaznadzey, R. Haselkorn & M. Y. Galperin, (2006) The cyanobacterial genome core and the origin of photosynthesis. *Proc Natl Acad Sci U S A* **103**: 13126-13131.

Muro-Pastor, A. M., A. Valladares, E. Flores & A. Herrero, (2002) Mutual dependence of the expression of the cell differentiation regulatory protein HetR and the global nitrogen regulator NtcA during heterocyst development. *Mol Microbiol* **44**: 1377-1385.

Muro-Pastor, M. I., J. C. Reyes & F. J. Florencio, (2001) Cyanobacteria perceive nitrogen status by sensing Intracellular 2-oxoglutarate levels. *J Biol Chem* **276**: 38320-38328.

Namikoshi, M., T. Murakami, M. F. Watanabe, T. Oda, J. Yamada, S. Tsujimura, H. Nagai & S. Oishi, (2003) Simultaneous production of homoanatoxin-a, anatoxin-a, and a new non-toxic 4-hydroxyhomoanatoxin-a by the cyanobacterium *Raphidiopsis mediterranea* Skuja. *Toxicon* **42**: 533-538.

Narahashi, T. & J. W. Moore, (1968) Neuroactive agents and nerve membrane conductances. *J Gen Physiol* **51**: 93-101.

Neilan, B. A., D. Jacobs, T. DelDot, L. L. Blackall, P. R. Hawkins, P. T. Cox & A. E. Goodman, (1997) rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus *Microcystis*. *Int J Syst Bacteriol* **47**: 693-697.

Neilan, B. A., M. L. Saker, J. Fastner, A. Torokne & B. P. Burns, (2003) Phylogeography of the invasive cyanobacterium *Cylindrospermopsis raciborskii*. *Mol Ecol* **12**: 133-140.

Nishizawa, T., A. Ueda, M. Asayama, K. Fujii, K. Harada, K. Ochi & M. Shirai, (2000) Polyketide synthase gene coupled to the peptide synthetase module involved in the biosynthesis of the cyclic heptapeptide microcystin. *J Biochem* **127**: 779-789.

Norris, R. L., G. K. Eaglesham, G. Pierens, G. R. Shaw, M. J. Smith, R. K. Chiswell, A. A. Seawright & M. R. Moore, (1999) Deoxycylindrospermopsin, an analog of cylindrospermopsin from *Cylindrospermopsis raciborskii*. *Environ Toxicol* **14**: 163-165.

Nübel, U., F. GarciaPichel & G. Muyzer, (1997) PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* **63**: 3327-3332.

Ochman, H., J. G. Lawrence & E. A. Groisman, (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.

Ohtani, I., R. E. Moore & M. T. C. Runnegar, (1992) Cylindrospermopsin: a potent hepatotoxin from the blue-green alga *Cylindrospermopsis raciborskii*. *J Am Chem Soc* **114**: 7941-7942.

Onodera, H., M. Satake, Y. Oshima, T. Yasumoto & W. W. Carmichael, (1997) New saxitoxin analogues from the freshwater filamentous cyanobacterium *Lyngbya wollei*. *Nat Toxins* **5**: 146-151.

Padisák, J., (1997) *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. *Archiv für Hydrobiologie. Supplement volumes, Monographic studies* **107**: 563-593.

Palinska, K. A., T. Jahns, R. Rippka & N. Tandeau de Marsac, (2000) *Prochlorococcus marinus* strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease. *Microbiology* **146**: 3099-3107.

Parsons, J. D., (1995) Miropeats: Graphical DNA sequence comparisons. *Comput Appl Biosci* **11**: 615-619.

Partensky, F., W. R. Hess & D. Vaulot, (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106-127.

Paz-Yepes, J., E. Flores & A. Herrero, (2003) Transcriptional effects of the signal transduction protein PII (*glnB* gene product) on NtcA-dependent genes in *Synechococcus* sp. PCC 7942. *FEBS Lett* **543**: 42-46.

Paz-Yepes, J., E. Flores & A. Herrero, (2009) Expression and mutational analysis of the *glnB* genomic region in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **191**: 2353-2361.

Pereira, P., H. Onodera, D. Andrinolo, S. Franca, F. Araujo, N. Lagos & Y. Oshima, (2000) Paralytic shellfish toxins in the freshwater cyanobacterium *Aphanizomenon flos-aquae*, isolated from Montargil reservoir, Portugal. *Toxicon* **38**: 1689-1702.

Plominsky, A. M., K. Soto-Liebe & M. Vasquez, (2009) Optimization of 2D-PAGE protocols for proteomic analysis of two nonaxenic toxin-producing freshwater cyanobacteria: *Cylindrospermopsis raciborskii* and *Raphidiopsis* sp. *Lett Appl Microbiol* **49**: 332-337.

Pomati, F., M. C. Moffitt, R. Cavaliere & B. A. Neilan, (2004) Evidence for differences in the metabolism of saxitoxin and C1+2 toxins in the freshwater cyanobacterium *Cylindrospermopsis raciborskii* T3. *Biochim Biophys Acta-Gen Subj* **1674**: 60-67.

Pomati, F., B. A. Neilan, T. Suzuki, G. Manarolla & C. Rossetti, (2003) Enhancement of intracellular saxitoxin accumulation by lidocaine hydrochloride in the cyanobacterium *Cylindrospermopsis raciborskii* T3 (Nostocales). *J Phycol* **39**: 535–542.

Posada, D. & K. A. Crandall, (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Rajagopalan, R. & S. M. Callahan, (2010) Temporal and spatial regulation of the four transcription start sites of *hetR* from *Anabaena* sp. strain PCC 7120. *J Bacteriol* **192**: 1088-1096.

Ramasubramanian, T. S., T. F. Wei, A. K. Oldham & J. W. Golden, (1996) Transcription of the *Anabaena* sp. strain PCC 7120 *ntcA* gene: multiple transcripts and NtcA binding. *J Bacteriol* **178**: 922-926.

Ramirez, M. E., P. B. Hebbar, R. B. Zhou, C. P. Wolk & S. E. Curtis, (2005) *Anabaena* sp. strain PCC 7120 gene *devH* is required for synthesis of the heterocyst glycolipid layer. *J Bacteriol* **187**: 2326-2331.

Ressom, R., T. Maynard, D. Roder, O. El Saadi, L. Turczynowicz, J. Fitzgerald, F. San Soong & I. R. Falconer, (1994) Health effects of toxic cyanobacteria (blue-green algae). In.: Australian Government Publishing Service.

Richter, M. & R. Rosselló-Mora, (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106**: 19126-19131.

Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman & R. Y. Stanier, (1979) Generic assingnments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**: 1-61.

Rocap, G., F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser & S. W. Chisholm, (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042-1047.

Runnegar, M. T., S. M. Kong, Y. Z. Zhong, J. L. Ge & S. C. Lu, (1994) The role of glutathione in the toxicity of a novel cyanobacterial alkaloid cylindrospermopsin in cultured rat hepatocytes. *Biochem Biophys Res Commun* **201**: 235-241.

Saeed, A. I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush & J. Quackenbush, (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* **34**: 374-378.

Saitou, N. & M. Nei, (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.

Saker, M., (2000) Cyanobacterial blooms in tropical north Queensland water bodies. In. Townsville, Australia: James Cook University.

Saker, M. L. & B. A. Neilan, (2001) Varied diazotrophies, morphologies, and toxicities of genetically similar isolates of *Cylindrospermopsis raciborskii* (Nostocales, Cyanophyceae) from northern Australia. *Appl Environ Microbiol* **67**: 1839-1845.

Saker, M. L., B. A. Neilan & D. J. Griffiths, (1999) Two morphological forms of *Cylindrospermopsis raciborskii* (cyanobacteria) isolated from Solomon Dam, Palm Island, Queensland. *J Phycol* **35**: 599-606.

Sako, Y., T. Yoshida, A. Uchida, O. Arakawa, T. Noguchi & Y. Ishida, (2001) Purification and characterization of a sulfotransferase specific to N-21 of saxitoxin and gonyautoxin 2+3 from the toxic dinoflagellate *Gymnodinium catenatum* (Dinophyceae). *J Phycol* **37**: 1044-1051.

Sakr, S., R. Jeanjean, C. C. Zhang & T. Arcondeguy, (2006) Inhibition of cell division suppresses heterocyst development in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **188**: 1396-1404.

Sánchez-Baracaldo, P., P. K. Hayes & C. E. Blank, (2005) Morphological and habitat evolution in the cyanobacteria using a compartmentalization approach. *Geobiology* **3**: 145-165.

Sandh, G., R. El-Shehawy, B. Diez & B. Bergman, (2009) Temporal separation of cell division and diazotrophy in the marine diazotrophic cyanobacterium *Trichodesmium erythraeum* IMS101. *FEMS Microbiol Lett* **295**: 281-288.

Schindler, D. W., R. E. Hecky, D. L. Findlay, M. P. Stainton, B. R. Parker, M. J. Paterson, K. G. Beaty, M. Lyng & S. E. M. Kasian, (2008) Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci U S A* **105**: 11254-11258.

Schopf, J. W. & B. M. Packer, (1987) Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona group, Australia *Science* **237**: 70-73.

Schrautemeier, B. & H. Böhme, (1985) A distinct ferredoxin for nitrogen fixation isolated from heterocysts of the cyanobacterium *Anabaena variabilis*. *FEBS Lett* **184**: 304-308.

Shafik, H. M., L. Voros, P. Sprober, M. Presing & A. W. Kovacs, (2003) Some special morphological features of *Cylindrospermopsis raciborskii* in batch and continuous cultures. *Hydrobiologia* **506**: 163-167.

Shalev-Malul, G., J. Lieman-Hurwitz, Y. Viner-Mozzini, A. Sukenik, A. Gaathon, M. Lebendiker & A. Kaplan, (2008) An AbrB-like protein might be involved in the regulation of cylindrospermopsin production by *Aphanizomenon ovalisporum*. *Environ Microbiol* **10**: 988-999.

Shen, X., P. K. S. Lam, G. R. Shaw & W. Wickramasinghe, (2002) Genotoxicity investigation of a cyanobacterial toxin, cylindrospermopsin. *Toxicon* **40**: 1499-1501.

Shi, T. & P. G. Falkowski, (2008) Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc Natl Acad Sci U S A* **105**: 2510-2515.

Shimizu, Y., (1986) Toxigenesis and biosynthesis of saxitoxin analogs. *Pure Appl Chem* **58**: 257-262.

Shimizu, Y., (1996) Microalgal metabolites: a new perspective. *Annu Rev Microbiol* **50**: 431-465.

Shimodaira, H. & M. Hasegawa, (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**: 1114-1116.

Singh, R. N., (1962) Seasonal variants of *Anabaenopsis raciborskii* Wolosz. *Hydrobiologia* **20**: 87–91.

Sivonen, K. & G. Jones, (1999) Cyanobacterial toxins. In: Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management. I. Chorus & J. Bartram (eds). London: E & FN Spon, pp. 41-111.

Smyth, G., (2005) Limma: linear models for microarray data. In: Bioinformatics and computational biology solutions using R and Bioconductor. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry & W. Huber (eds). New York: Springer, pp. 397–420.

Soto-Liebe, K., K. Stucken, M. A. Méndez, N. Lagos, A. Cembella, B. Krock & M. Vásquez, (2008) The effects of chloramphenicol, arginine and temperature on PST-production by *Cylindrospermopsis raciborskii* strain D9. In: Proceedings of the 12th International Conference on Harmful Algae Ø. Moestrup, G. Doucette, H. Enevoldsen, A. Godhe, G. Hallegraeff, B. Lukas, N. Lundholm, J. Lewis, K. Rengefors, K. Sellner, K. Steidinger, P. Tester & A. Zingone (eds). Copenhagen: International Society for the Study of Harmful Algae (ISSHA) and the Intergovernmental Oceanographic Commission (IOC of UNESCO), pp. 330-333.

Spivak, C. E., B. Witkop & E. X. Albuquerque, (1980) Anatoxin-a: a novel, potent agonist at the nicotinic receptor. *Mol Pharmacol* **18**: 384-394.

Spróber, P., H. M. Shafik, M. Présing, A. W. Kovács & S. Herodek, (2003) Nitrogen uptake and fixation in the cyanobacterium *Cylindrospermopsis raciborskii* under different nitrogen conditions. *Hydrobiologia* **506**: 169-174.

St. Amand, A., (2002) *Cylindrospermopsis*: an invasive toxic alga. *LakeLine* **22**: 36-38.

Stal, L. J., P. Albertano, B. Bergman, K. v. Bröckel, J. R. Gallon, P. K. Hayes, K. Sivonen & A. E. Walsby, (2003) BASIC: Baltic Sea cyanobacteria. An investigation of the structure and dynamics of water blooms of cyanobacteria in the Baltic Sea-responses to a changing environment. *Cont Shelf Res* **23**: 1695-1714.

Steglich, C., M. E. Futschik, D. Lindell, B. Voss, S. W. Chisholm & W. R. Hess, (2008) The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genet* **4**: 14.

Suikkanen, S., J. Engstrom-Ost, J. Jokela, K. Sivonen & M. Viitasalo, (2006) Allelopathy of Baltic Sea cyanobacteria: no evidence for the role of nodularin. *J Plankton Res* **28**: 543-550.

Swingley, W. D., R. E. Blankenship & J. Raymond, (2008) Integrating markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Mol Biol Evol* **25**: 643-654.

Swofford, D. L., (2003) PAUP*: phylogenetic analysis using parsimony, version 4.0b10.

Tamagnini, P., L. Elsa, O. Paulo, F. Daniela, P. Filipe, H. David James, H. Thorsten & L. Peter, (2007) Cyanobacterial hydrogenases: diversity, regulation and applications. *FEMS Microbiol Rev* **31**: 692-720.

Tanigawa, R., M. Shirokane, S.-i. Maeda, T. Omata, K. Tanaka & H. Takahashi, (2002) Transcriptional activation of NtcA-dependent promoters of *Synechococcus* sp. PCC 7942 by 2-oxoglutarate invitro. *Proc Natl Acad Sci U S A* **99**: 4251-4255.

Thiel, T., (1990) Protein turnover and heterocyst differentiation in the cyanobacterium *Anabaena variabilis J Phycol* **26**: 50-54.

Thiel, T., E. M. Lyons & J. C. Erker, (1997) Characterization of genes for a second Mo-dependent nitrogenase in the cyanobacterium *Anabaena variabilis*. *J Bacteriol* **179**: 5222-5225.

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

Tolonen, A. C., J. Aach, D. Lindell, Z. I. Johnson, T. Rector, R. Steen, G. M. Church & S. W. Chisholm, (2006) Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol Syst Biol* **2**: 53.

Tomitani, A., A. H. Knoll, C. M. Cavanaugh & T. Ohno, (2006) The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A* **103**: 5442-5447.

Tonk, L., D. B. van de Waal, P. Slot, J. Huisman, H. C. P. Matthijs & P. M. Visser, (2008) Amino acid availability determines the ratio of microcystin variants in the cyanobacterium *Planktothrix agardhii*. *FEMS Microbiol Ecol* **65**: 383-390.

Tonk, L., P. M. Visser, G. Christiansen, E. Dittmann, E. Snelder, C. Wiedner, L. R. Mur & J. Huisman, (2005) The microcystin composition of the cyanobacterium *Planktothrix agardhii* changes toward a more toxic variant with increasing light intensity. *Appl Environ Microbiol* **71**: 5177-5181.

Tusher, V. G., R. Tibshirani & G. Chu, (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**: 5116-5121.

Utkilen, H. & N. Gjolme, (1995) Iron-stimulated toxin production in *Microcystis aeruginosa*. *Appl Environ Microbiol* **61**: 797-800.

Vahtera, E., D. J. Conley, B. G. Gustafsson, H. Kuosa, H. Pitkanen, O. P. Savchuk, T. Tamminen, M. Viitasalo, M. Voss, N. Wasmund & F. Wulff, (2007) Internal ecosystem feedbacks enhance nitrogen-fixing cyanobacteria blooms and complicate management in the Baltic Sea. *Ambio* **36**: 186-194.

Valladares, A., M. L. Montesinos, A. Herrero & E. Flores, (2002) An ABC-type, high-affinity urea permease identified in cyanobacteria. *Mol Microbiol* **43**: 703-715.

Van De Water, S. D. & R. D. Simon, (1982) Induction and differentiation of heterocysts in the filamentous cyanobacterium *Cylindrospermum licheniforme*. *J Gen Microbiol* **128**: 917-925.

Vázquez-Bermúdez, M. F., A. Herrero & E. Flores, (2000) Uptake of 2-oxoglutarate in *Synechococcus* strains transformed with the *Escherichia coli kgtP* gene. *J Bacteriol* **182**: 211-215.

Vázquez-Bermúdez, M. F., A. Herrero & E. Flores, (2002) 2-oxoglutarate increases the binding affinity of the NtcA (nitrogen control) transcription factor for the *Synechococcus glnA* promoter. *FEBS Lett* **512**: 71-74.

Vega-Palas, M. A., F. Madueno, A. Herrero & E. Flores, (1990) Identification and cloning of a regulatory gene for nitrogen assimilation in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *J Bacteriol* **172**: 643-647.

Vintila, S. & R. El-Shehawy, (2007) Ammonium ions inhibit nitrogen fixation but do not affect heterocyst frequency in the bloom-forming cyanobacterium *Nodularia spumigena* strain AV1. *Microbiology* **153**: 3704-3712.

Wang, K., C. Hiruki & F. Yeh, (2003) Molecular evolution of phytoplasmas based on polymorphisms in the 16S rRNA genes and the 16/23S spacer regions. *Proc Jpn Acad Ser B Phys Biol Sci* **79**: 155-162.

Wealand, J. L., J. A. Myers & R. Hirschberg, (1989) Changes in gene expression during nitrogen starvation in *Anabaena variabilis* ATCC 29413. *J Bacteriol* **171**: 1309-1313.

Wei, T. F., T. S. Ramasubramanian & J. W. Golden, (1994) *Anabaena* sp. strain PCC 7120 *ntcA* gene required for growth on nitrate and heterocyst development. *J Bacteriol* **176**: 4473-4482.

Wiedner, C., J. Ruecker, R. Brueggemann & B. Nixdorf, (2007) Climate change affects timing and size of populations of an invasive cyanobacterium in temperate regions. *Oecologia* **152**: 473-484.

Wilson, K., (1990) Preparation of Genomic DNA from Bacteria. In: Current protocols in molecular biology. F. Ausubel, R. Brent, R. Kingston, J. Seidman, J. Smith & K. Struhl (eds). New York: Wiley Interscience, pp. 2.4.1–2.4.5.

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

Wong, F. C. Y. & J. C. Meeks, (2001) The *hetF* gene product is essential to heterocyst differentiation and affects HetR function in the cyanobacterium *Nostoc punctiforme*. *J Bacteriol* **183**: 2654-2661.

Wu, X. Q., D. W. Lee, R. A. Mella & J. W. Golden, (2007) The *Anabaena* sp. strain PCC 7120 asr1734 gene encodes a negative regulator of heterocyst development. *Mol Microbiol* **64**: 782-794.

Wu, Z. X., J. Q. Shi & R. H. Li, (2009) Comparative studies on photosynthesis and phosphate metabolism of *Cylindrospermopsis raciborskii* with *Microcystis aeruginosa* and *Aphanizomenon flos-aquae*. *Harmful Algae* **8**: 910-915.

Xu, W. L., R. Jeanjean, Y. D. Liu & C. C. Zhang, (2003) *pkn22* (alr2502) encoding a putative Ser/Thr kinase in the cyanobacterium *Anabaena* sp. PCC 7120 is induced by both iron starvation and oxidative stress and regulates the expression of *isiA*. *FEBS Lett* **553**: 179-182.

Xu, X., J. Elhai & C. P. Wolk, (2008) Transcriptional and developmental responses by *Anabaena* to deprivation of fixed nitrogen. In: The cyanobacteria: molecular biology, genomics and evolution. A. Herrero & E. Flores (eds). Norfolk: Caister Academic Press, pp. 383-422.

Xu, X. & C. P. Wolk, (2001) Role for *hetC* in the transition to a nondividing state during heterocyst differentiation in *Anabaena* sp. *J Bacteriol* **183**: 393-396.

Yin, Q. Q., W. W. Carmichael & W. R. Evans, (1997) Factors influencing growth and toxin production by cultures of the freshwater cyanobacterium *Lyngbya wollei* Farlow ex Gomont. *J Appl Phycol* **9**: 55-63.

Yoon, H. S. & J. W. Golden, (1998) Heterocyst pattern formation controlled by a diffusible peptide. *Science* **282**: 935-938.

Yoon, H. S., M. H. Lee, J. Xiong & J. W. Golden, (2003) *Anabaena* sp. strain PCC 7120 *hetY* gene influences heterocyst development. *J Bacteriol* **185**: 6995-7000.

Yoshimura, H., S. Okamoto, Y. Tsumuraya & M. Ohmori, (2007) Group 3 sigma factor gene, *sigJ*, a key regulator of desiccation tolerance, regulates the synthesis of extracellular polysaccharide in cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res* **14**: 13-24.

Yunes, J. S., S. De La Rocha, D. Giroldo, S. B. da Silveira, R. Comin, M. D. Bicho, S. S. Melcher, C. L. Sant'anna & A. A. H. Vieira, (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.

Zalocar de Domitrovic, Y., A. S. G. Poi de Neiff & S. L. Casco, (2007) Abundance and diversity of phytoplankton in the Paraná River (Argentina) 220 km downstream of the Yacyretá reservoir. *Braz J Biol* **67**: 53-63.

Zhang, C.-C., S. Laurent, S. Sakr, L. Peng & S. Bédu, (2006) Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Mol Microbiol* **59**: 367-375.

Zhang, C. C., (1993) A gene encoding a protein related to eukaryotic protein kinases from the filamentous heterocystous cyanobacterium *Anabaena* PCC 7120. *Proc Natl Acad Sci U S A* **90**: 11840-11844.

Zhang, C. C. & L. Libs, (1998) Cloning and characterisation of the *pknD* gene encoding an eukaryotic-type protein kinase in the cyanobacterium *Anabaena* sp. PCC7120. *Mol Gen Genet* **258**: 26-33.

Zhang, J. Y., W. L. Chen & C. C. Zhang, (2009) *hetR* and *patS*, two genes necessary for heterocyst pattern formation, are widespread in filamentous non heterocyst-forming cyanobacteria. *Microbiology* **155**: 1418-1426.

Zhang, W., Y. Du, I. Khudyakov, Q. Fan, H. Gao, D. G. Ning, C. P. Wolk & X. D. Xu, (2007) A gene cluster that regulates both heterocyst differentiation and pattern formation in *Anabaena* sp. strain PCC 7120. *Mol Microbiol* **66**: 1429-1443.

Zhao, Y. H., Y. M. Shi, W. X. Zhao, X. Huang, D. H. Wang, N. Brown, J. Brand & J. D. Zhao, (2005) CcbP, a calcium-binding protein from *Anabaena* sp. PCC 7120 provides evidence that calcium ions regulate heterocyst differentiation. *Proc Natl Acad Sci U S A* **102**: 5744-5748.

Zhaxybayeva, O., J. P. Gogarten, R. L. Charlebois, W. F. Doolittle & R. T. Papke, (2006) Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Res* **16**: 1099-1108.

Zhaxybayeva, O., P. Lapierre & J. P. Gogarten, (2004) Genome mosaicism and organismal lineages. *Trends Genet* **20**: 254-260.

Zingone, A. & H. O. Enevoldsen, (2000) The diversity of harmful algal blooms: a challenge for science and management. *Ocean Coast Manag* **43**: 725-748.