

Phylogenetic and Comparative Genomic Analyses of Heterocystous Cyanobacteria

Phylogenetic and Comparative Genomic Analyses of Heterocystous Cyanobacteria

BY

MOHAMMAD HOWARD-AZZEH, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

© Copyright by Mohammad Howard-Azzeh, September 2013

MASTER OF SCIENCE (2013)

McMaster University

(Biology)

Hamilton, Ontario

TITLE: Phylogenetic and Comparative Genomic Analyses of Heterocystous
Cyanobacteria

AUTHOR: Mohammad Howard-Azzeh, B.Sc. (Brock University)

SUPERVISOR: Dr. H.E. Schellhorn

CO-SUPERVISOR: Dr. R.S. Gupta

NUMBER OF PAGES: ix, 235

Abstract

Heterocyst-forming Cyanobacterial species encompass the orders *Nostocales* and *Stigonematales*. These orders are currently differentiated based solely upon morphological characteristics (such as the formation of true or false branches), which may be unrelated to phylogeny. Thus, these bacteria do not form distinct monophyletic groups in the 16S rRNA tree, and as yet, no reliable molecular markers have yet been identified that allow species of these two orders to be distinguished from each other or from other organisms. Using published genome sequences for these species, we have investigated the relationship of species from these two orders. We describe here detailed phylogenetic analyses based on concatenated protein sequences for 45 proteins from 48 Cyanobacterial species/strains whose genomes are now available. In addition, we have performed comprehensive comparative genomic analyses on eight available *Nostocales* and *Stigonematales* genomes to identify conserved signature indels (CSIs) and conserved signature proteins (CSPs) that are specifically present in all *Nostocales/Stigonematales* or any of their subclades. These analyses have yielded >100 CSIs and CSPs that are specific for distinct coherent clades of *Nostocales/Stigonematales*. Seventeen of these CSIs and twelve CSPs are present in all sequenced *Nostocales* and *Stigonematales*, supporting a distinct relationship of the species from these two orders of heterocyst-forming bacteria. Fifty-four CSIs and forty-one CSPs are specific for different subclades of *Nostocales* and many others are diagnostic of individual species/strains. The newly-identified CSIs and CSPs, which are specific for different subclades of *Nostocales/Stigonematales*, may provide novel means for identifying previously unknown members of these orders, for assignment of unsequenced members of heterocystous Cyanobacteria into different identified subclades, and

for detecting individual strains from environmental samples by employing CSIs/CSPs as diagnostic tools.

Acknowledgments

I would like to thank Dr. Schellhorn and Dr. Gupta for giving me the opportunity to work under their supervision. Their leadership and patience allowed me to understand the complicated mechanisms and troubleshooting faced throughout this project. Their intellect and experience inspires me not only in the lab but also on a personal level. Working under Dr. Schellhorn's and Gupta's guidance has been an excellent experience and is truly appreciated

To my fellow lab mates, I would to thank Steven Botts, Mohammad Mohiuddin, Shirely Wong, Lingzi Kong, Pardis Haghighi, Dr. Sarah Chiang, Rosemary Yu, Myra Sohail, Quan Yao, Xiaoli Zhao, Ryan Zheng, Andrew O'Brian, Mobolaji Adeolu, Hafiz Sohail Naushad, Vibhab Bhandari and Aqeel Al-Dahwi for all their support and assistance. Their friendship made the duration of my project a momentous experience.

I thank Dr. Igdoura for taking time away from his schedule to chair my defense.

Finally, I would like to thank my family for their love, patients, support and inspiration through all life's endeavours.

List of Illustrations

Tables

Table 1: General information on toxins produced by cyanobacteria expected in freshwater ecosystems	13
Table 2: Cyanobacterial isolates identified from environmental sample from the Great Lakes. Isolates are recorded from both the Canadian Phycological Culture Centre and published literature	24
Table 3: List of cyanobacterial species that have been identified in lakes within Ontario	25
Table 4: Sequence characteristics of featured Cyanobacterial genomes	35
Table 5: Conserved signature indels specific for all <i>Nostocales</i> and <i>Stigonematales</i>	44
Table 6: Conserved signature indels specific for Subclades of <i>Nostocales</i>	45
Table 7: Conserved signature indels specific for Subclades of <i>Nostocales</i>	46
Table 8: Conserved signature proteins specific for Clades of <i>Nostocales</i> and <i>Stigonematales</i>	46

Figures

Figure 1: General structures of Microcystin, Anatoxin-a, Saxitoxin and Cylindrospermopsin	14
Figure 2: Graphical representation of the Great Lakes indicating (dots) regions subject to cyanobacterial bloom formations	23
Figure 3: A maximum-likelihood consensus tree based on 45 concatenated sequences for 48 species of Cyanobacteria	33
Figure 4: A neighbor-joining distance tree based on 16S rRNA sequences, representing all genera of <i>Nostocales</i> and <i>Stigonematales</i>	34
Figure 5: Partial sequence alignments for the proteins (A) 30S ribosomal protein S3, showing a 5 aa insertion and (B) ATP synthase epsilon subunit, showing a 2 aa insertion	55
Figure 6: Partial sequence alignments for the proteins (A) Recombination protein O, showing a 7 aa insertion and (B) O-succinylbenzoyl-CoA synthetase, showing a 7 aa insertion	57
Figure 7: Partial sequence alignments for the proteins (A) Glucose-1-phosphate thymidyltransferase, showing a 1 aa insertion and (B) Alpha-amylase, showing a 9 aa insertion	59
Figure 8: A summary diagram showing the distribution of identified CSIs and CSPs specific to various clades within <i>Nostocales/Stigonematales</i>	61

Table of Contents

Abstract	iv
Acknowledgments	vi
List of Illustrations	vii
Tables.....	vii
Figures.....	vii
Table of Contents	viii
Chapter 1: Introduction	1
1.1 Project Intent	1
1.2 General Background.....	2
1.2.1 <i>Cyanobacteria</i>	2
1.2.2 <i>Structural Variation and Reproduction</i>	3
1.2.3 <i>Classification</i>	4
1.2.4 <i>Cyanotoxins</i>	7
1.2.5 <i>Extracellular Toxin Removal</i>	9
1.2.6 <i>Geosmin</i>	15
1.2.7 <i>Current Methods of Cyanobacterial Monitoring</i>	16
1.2.8 <i>Cyanobacteria of the Great Lakes</i>	19
1.3 Project Introduction.....	26
Chapter 2: Methods	30
2.1 Phylogenetic Sequence Analysis	30
2.2 Identification of Conserved Signature Indels.....	31
2.3 Identification of Specificity for Previously Reported Conserved Signature Proteins....	32
Chapter 3: Results	36
3.1 Phylogenetic Analyses of <i>Nostocales</i> and <i>Stigonematales</i>	36
3.2 Conserved Signature Indels Specific for <i>Nostocales/Stigonematales</i> and <i>Nostocales</i> ..	39
3.3 Conserved Signature Indels Specific for <i>Nostocales</i> Subclades	40
3.4 Conserved signature Indels Specific to Individual Members of Sequenced <i>Nostocales/Stigonematales</i>	42
3.5 Conserved Signature Proteins (CSPs) Specific for Clades of <i>Nostocales</i> and <i>Stigonematales</i>	43
Chapter 4: Discussion	47
4.1 Phylogenetic and Comparative Genomic Analyses.....	47

4.2	Implications Inferred by Phylogenetic and Comparative Genomic Analyses Pertaining to all <i>Nostocales/Stigonematales</i>	48
4.3	Implications Inferred by Phylogenetic and Comparative Genomic Analyses Pertaining to the First Subclade of <i>Nostocales</i>	51
4.3	Implications Inferred by Phylogenetic and Comparative Genomic Analyses Pertaining to the Second Subclade of <i>Nostocales</i>	52
4.4	Conserved Signature Indels Specific to Individual Members of Sequenced <i>Nostocales/Stigonematales</i>	53
4.5	Conclusion.....	53
4.7	Future work.....	62
	References	65
	Thesis Defence	77
	Supplementary Figures and Tables.....	81

Chapter 1: Introduction

1.1 Project Intent

The initial focus of this project was to understand problematic Cyanobacteria in the Great Lakes and develop a means of monitoring the characteristics of the inhabiting communities. This is of particular interest since human impact on aquatic environments increases rates of water eutrophication, directly increasing the frequency and intensity of algal blooms (excessive cyanobacterial growth). In addition to the reduced aesthetic appearance and overall utility of recreational water, several species of Cyanobacteria responsible for fresh water blooms produce potentially harmful toxins posing a significant public health risk. Other cyanobacterial secondary metabolites produced by several bloom-forming species affect the taste and odour of drinking water, which results in increased filtering processes/run times of water treatment facilities (Westrick, 2003; Rashash et al., 1996). The assay to be developed allows for a rapid means of characterizing fresh water cyanobacteria, which may play a key role in predictive modeling, takes advantage of specific molecular markers (explained later in-depth) that are present only in related groups of organisms. However, these molecular markers allow inferences on the evolution of the organisms being studied. Understanding the evolutionary history that the identified molecular markers describe and the species that they are specific to is necessary to determine their potential detection boundaries. Therefore, this project has focused on identifying an abundance of molecular markers specific to a group of Cyanobacteria as well as their evolutionary implications.

1.2 General Background

1.2.1 *Cyanobacteria*

Cyanobacteria are one of the major phyla of bacteria, defined as the sole prokaryotic oxygenic phototrophs. These organisms require only water, carbon dioxide, nitrogen, phosphorus, light and trace metals for growth, with many species having the capacity to fix nitrogen (Westrick et al., 2010). Cyanobacteria possess photosystems I and II in their photosynthetic apparatus. The photosynthetic chlorophyll is located in thylakoids found in the cell's periphery. In addition to chlorophyll, cyanobacteria produce the accessory pigments phycocyanin, phycoerythrin and allophycocyanin to make use of different wavelengths across the light spectrum (Pandey et al., 2004). These pigments are imbedded in phycobilisomes and found in rows on the outside of thylakoids (Pandey et al., 2004). Members of this phylum are widely distributed over all latitudes, and naturally inhabit a vast array of terrestrial, marine and freshwater environments. They are generally more tolerant of environmental extremes than algae and are often the dominant or sole oxygenic phototrophic organisms in hot springs, hypersaline lakes, and other extreme niches. In terrestrial environments, cyanobacteria are found in soil and rocks, often in extreme conditions such as desert sands by forming extensive crusts over the surface and remain dormant for most of the year (Garcia-Pichel et al., 2001). Some cyanobacteria share symbiotic relationships with plants, such as liverworts, ferns and cycads, fixing nitrogen at their root systems or acting as the phototrophic component for lichens (Westrick et al., 2010).

1.2.2 *Structural Variation and Reproduction*

Many cyanobacteria contain gas vesicles in the cell's cytoplasm. This phenotype is common amid planktonic species living in open aquatic ecosystems. These gas vesicles regulate the buoyancy of the cell to maintain a position within a water column that optimizes photosynthesis. A group of filamentous cyanobacteria can form differentiated specialized cells which fix nitrogen, known as heterocysts. Heterocysts are distributed regularly along or at the ends of filaments (Sandh et al., 2009). Heterocysts have intercellular connections to its neighbouring vegetative cells for material exchange. Most heterocysts are low in accessory pigments and therefore lack photosystem II. Lacking the ability to fix CO₂ means they cannot create the necessary reductant to fix nitrogen. Instead fixed carbon is imported from adjacent vegetative cells to be used as an electron donor for the reduction of N₂ to NH₃. In return, nitrogen is transported back to the vegetative cells as glutamine (Kumar et al., 2010). Unicellular nitrogen-fixing strains also exist, having the capability to store glycogen during the day and fix nitrogen at night (Toepel et al., 2008). Since nitrogenase is sensitive to oxygen (Haystead et al., 1970), these cells use a variety of strategies like temporal regulation (switching between nitrogen fixation/photosynthesis every 24h) to limit the amount of O₂ that enters the cells (Huang et al., 1993).

In water of sufficient temperature and nutrition, some cyanobacterial strains can double their population in 0.3-1.4 days and can reach concentrations of millions of cells/mL (Mur et al., 1999). High cell counts naturally diminish as water temperatures decrease in autumn (Pandey et al., 2004). As conditions become less favourable for cyanobacteria, some vegetative cells differentiate into akinetes, a larger and thicker cell wall type structure in filaments. These akinetes can survive in lowered temperatures where most other cell types die off. They are

capable of resting dormant over winter in water, sediments or surfaces after water has dried. In non-akinete-forming species, as temperatures decline in autumn, photosynthesis becomes more rapid than respiration. The carbohydrate ballast is not consumed and the colonies sink to the bottom of the water column where they can survive winter. Gradually, carbohydrate stores are consumed and cells that ascend to the surface in the spring are unicellular or small cell aggregates (Leao et al., 2012; Watson et al., 2008).

Some filamentous cyanobacteria also differentiate into hormogonia as a mechanism of dispersal and reproduction. Hormogonia are specialized motile cells that detach and migrate from the parent filament to form new colonies (Pandey et al., 2004).

1.2.3 *Classification*

Taxonomic classification of cyanobacteria has been controversial. The morphological classification of cyanobacteria was originally governed by the International Code of Botanical Nomenclature (bacteriological code), prior to their characterization as bacteria (Komarek, 1994; Watson et al., 2008; Euzéby, 1997; Geitler, 1932; Komarek, 2002; Rippka et al., 1979; Stanier et al., 1978; Wilmotte and Herdman, 2001; Watson et al., 2008). Thereafter, the traditional classification of Cyanobacteria was developed and is primarily is still based on morphology as well as ecological and biochemical characteristics. The phylum was classified into five sections/orders: *Chroococcales*, *Pluechlorocales*, *Oscillatoriales*, *Nostocales* and *Stigonematales* (Euzéby, 1997; Rippka et al., 1979; Wilmotte and Herdman, 2001). With more in-depth knowledge of their diversity, further changes have been made to accommodate the modern view of cyanobacteria, most importantly the addition of the order/subsection *Gloeobacterales* (Cavalier-Smith, 2002; Hoffmann et al., 2005).

Though there are several proposed classification systems for Cyanobacteria (Cavalier-Smith, 2002; Euzeby, 1997; Hoffmann et al., 2005; Rippka et al., 1979; Wilmotte and Herdman, 2001), the only groups that are recognized by the International Code of Nomenclature of Bacteria (ICNB) are: 3 classes (Chroobacteria, Hormogoneae and Gloeobacteria), 6 orders (*Chroococcales*, *Pleurocapsales*, *Oscillatoriales*, *Nostocales*, *Stigonematales* and *Gloeobacterales*), and 5 genera (*Prochlorothrix*, *Prochloron*, *Prochlorococcus*, *Planktothricoides* and *Halospirulina*) (Cavalier-Smith, 2002; Euzeby, 1997). Morphological characteristics used to assign Cyanobacteria into their respective orders poorly correlate with Cyanobacterial phylogeny. Some of these classes/orders are known to be polyphyletic (Gugger and Hoffmann, 2004; Gupta and Mathews, 2010; Wilmotte and Herdman, 2001). Therefore, it is unclear which phylogenetic groups can be validly associated with the recognized classes/orders. Further, the polyphyletic genera/species which encompass these classes/orders are yet to be defined. However, the general classification can be outlined as described in the following paragraphs.

Cyanobacterial morphology encompasses unicellular, multicellular filamentous and colonial forms. The class Chroobacteria encompasses 3 orders, *Chroococcales*, *Pleurocapsales* and *Oscillatoriales*. The orders *Chroococcales* and *Pleurocapsales* are spherical and secrete a gel-like biofilm during colony growth which holds the cells in the colony. The traditional defining characteristics of *Chroococcales* includes cyanobacteria which reproduce by budding or through binary division in one, two or three planes (Komarek and Anagnostidis, 1986; Golubic, 1979). However, under the bacteriological code, the description is more ambiguous, describing them as unicellular and colonial (non-filamentous) cyanobacteria (with phycobilisomes and prochlorophytes with chlorophyll b instead (Cavalier-Smith, 2002).

The traditional description of *Pleurocapsales* and that of the bacteriological code are similar (Rippka et al., 1979; Wilmotte and Herdman, 2001). However, unlike the traditional description, under the bacteriological code *Pleurocapsales* include some filamentous members and are defined as colonial or filamentous, reproducing by intramural multiple fission to yield smaller unicellular dispersal stages (Cavalier-Smith, 2002).

The cyanobacterial orders which solely encompass filamentous members are found within two different classes but are divided into three orders, *Oscillatoriales*, *Nostocales* and *Stigonematales*. The main chain of the filament is called the trichome, and any filaments that exist on additional plains are at right angles to the main axis. Trichomes can exist straight or coiled. Under the bacteriological code, *Oscillatoriales* of the Chroobacteria class is described as unbranched linear filaments without heterocysts; cells typically shorter than broad (Cavalier-Smith, 2002). This order also does not differentiate into akinetes or hormogonia and its description is similar to the traditional narrative (Cavalier-Smith, 2002; Rippka et al., 1979; Wilmotte and Herdman, 2001).

Nostocales and *Stigonematales* encompass the class Hormogoneae. This class is monophyletic and is the sole group that can differentiate into heterocysts as well as hormogonia. Both *Nostocales* and *Stigonematales* have members which can also differentiate into Akinetes. *Nostocales* divide in one plane, forming simple filaments and occasionally false branches. Under the bacteriological code they are described as the only members whose filaments are unbranched and can produce heterocysts (Cavalier-Smith, 2002). Traditionally, they are defined the same way as that which is accepted under the bacteriological code (Rippka et al., 1979; Wilmotte and Herdman, 2001).

Stigonematales divide in multiple planes and form true branches. In the bacteriological code they are the only members whose filaments are branched and can produce heterocysts (Cavalier-Smith, 2002). The traditional description is similar to that of the bacteriological code (Rippka et al., 1979; Wilmotte and Herdman, 2001).

Gloeobacterales is the sole order found within the Gloeobacteria class and is a sister group to all other cyanobacteria, representing the deepest branching clade on phylogenetic trees. Traditionally this order is not recognized and their members belong to *Chroococcales* (Rippka et al., 1979; Wilmotte and Herdman, 2001). Members found within this *Gloeobacterales* (namely *Gloeobacter*) are described as not containing thylakoids (Cavalier-Smith, 2002).

1.2.4 Cyanotoxins

There are two major groups of toxins produced by cyanobacteria, referred to as the cyclic peptides and alkaloids. Each cyanotoxin is named after the cyanobacterial genera it was first isolated from (Westrick et al., 2010). **Table 1** presents a general overview of eight cyanotoxins found in freshwater, lethal dose, symptoms and possible treatments.

The general structure of cyclic peptide toxins can be seen in **Figure 1 A**. The cyclic peptides contain two groups of toxins, microcystins and nodularins. Microcystin from the cyanobacterium *Microcystis* is a cyclic heptapeptide. Microcystin has been identified in more than 90 variant conformations. Nodularins have similar toxicity to microcystins, and were first isolated from the cyanobacterium *Nodularia*. The most abundant molecular variation of the cyclic peptides is microcystin-LR, having L-leucine in the "R1" and L-arginine in the "R2" position (**Figure 1 A**) (Yoshida et al., 1997). Microcystin-LR is up to ten times more toxic than other conformations. These cyclic peptide toxins all share a β -linked amino acid component which is

essential for toxicity. Small stereochemical changes in a side chain will greatly reduce its ability to bind to the active site of protein phosphatases in eukaryotic cells. In bloomed cyanobacterial scum, microcystin can be found at concentrations as high as 20 mg/L (Falconer, 1991). However, at a lethal oral dose of 5-10 mg/kg body weight, it is unlikely that humans would die from drinking directly from contaminated water but may show symptoms such as headaches and painful diarrhoea (Yoshida et al., 1997). However, since this molecule may have carcinogenic properties, it is best to minimize contact with microcystin regardless of the concentration (Yoshida et al., 1997; Falconer, 1991).

Similar to higher plants, cyanobacteria synthesize a range of alkaloid toxins. The cyanotoxin anatoxin-a (**Figure 1 B**), first isolated from *Anabaena flos-aquae*, is the smallest of the cyanobacterial toxins. Its chemical structure is similar to cocaine and acetylcholine. Its toxicity stems from its ability to block synaptic transmissions (Haider et al., 2003). Anatoxin-a and homoanatoxin-a, a closely related compound, were identified in several animal poisonings (Chorus et al., 2000). These toxins are produced in both benthic and planktonic colonizing strains but are found only in filamentous species of cyanobacteria (Chorus et al., 2000; Haider et al., 2003).

The most potent congener cyanotoxin by weight are Saxitoxins (**Figure 1 C**) (Pomati et al., 2006). Saxitoxin is a tricyclic alkaloid and is famous for accumulating in filter-feeding shellfish, causing paralytic shellfish poisoning (PSP). This toxin is produced by several freshwater and marine cyanobacteria, and can accumulate to concentrations found high enough to be lethal through oral ingestion. Saxitoxin blocks sodium ion movement through voltage-gated sodium channels in motor nerves, leading to a conduction block that manifests as respiratory failure (Pomati et al., 2006). Severe intoxication results in hypoxia due to partial paralysis of the

diaphragm and accessory respiratory muscles. If left untreated, death by asphyxiation occurs (Pomati et al., 2006).

Cylindrospermopsin, initially isolated from *Cylindrospermopsis raciborskii* was first identified as a human and wildlife risk in Palm Island, Australia. In one outbreak, 149 people were hospitalized with symptoms of hepato-enteritis after consuming drinking water contaminated with cylindrospermopsin (Hawkins et al., 1997). Cylindrospermopsin blocks the *in vitro* activity of uridine monophosphate synthase complex, resulting in protein synthesis inhibition. Compared with other cyanotoxins, however, cylindrospermopsin is considerably less toxic and its concentration, even at its record high, are relatively low. Record high cylindrospermopsin concentrations of 589 µg/L are two orders of magnitude less than record concentrations for microcystin-LR at 25000 µg/L and an order of magnitude smaller than anatoxin-a record concentrations of 3300 µg/L (Saker and Eaglesham, 1999).

Based on animal studies, the tolerable daily intake (TDI) of microcystin-LR has been determined as 0.04 µg/kg/day (Falconer, 2001). Using the TDI value and assuming that the average person drinks approximately 2 L of water per day, The World Health Organization (WHO) determined a provisional guideline of a maximum of 1.0 µg microcystin-LR per litre of drinking water (WHO, 1998; WHO, 2003). Water containing less than 1.0 µg/L, of microcystin-LR, should not cause liver damage or any possible carcinogenesis/tumour growth, even with long-term exposure. Other cyanotoxins are treated as a microcystin-LR equivalent (Falconer, 1991).

1.2.5 Extracellular Toxin Removal

Cyanobacterial cells typically release their toxins under adverse environmental stress or cells lysis. High concentrations of toxins can remain in water columns for many weeks after

cyanobacterial blooms fall below detectable concentrations (Svrcek and Smith, 2004). As toxin concentrations are difficult to monitor and are influenced by several variables, it is difficult for water treatment facilities to predict how to treat water during potentially toxic seasons.

Intracellular toxins are removed with the cell during the sand or membrane filtration stages of water treatment. However, the removal of extracellular toxins requires more complicated and expensive procedures (Codd et al., 2005; Svrcek and Smith, 2004).

Chlorination is the most common treatment in North America used to degrade dissolved cyanobacterial toxins from drinking water. Chlorine is added to water at different concentrations for thirty min at an elevated pH. This process effectively removes microcystin, nodularin, cylindrospermopsin and saxitoxin at different pH ranges, but shows little effect on anatoxin-a. The oxidized by-products of this procedure has yet not been characterised (Svrcek and Smith, 2004).

Ozone is an effective way to quickly destroy microcystins, nodularins and anatoxin-a. However, alkalinity and dissolved organic carbon reduce the efficacy of ozone treatment (Svrcek and Smith, 2004). Ozone and chlorine both cause cell lysis, which leads to the release intracellular toxins. Therefore, ozone and chlorine treatment is recommended to take place after sand or membrane filtration removes cells and cell debris (Chorus et al., 2000).

Photochemical degradation using ultra violet light to break molecular bonds is advantageous as it does not require chemical additives. It is used to break down organic compounds in water supplies and effectively degrades microcystins, anatoxin-a and cylindrospermopsins. Normal industrial water treatment methods use an ultra violet dose of 40 mJoules/cm² to treat pathogens such as *Cryptosporidium parvum* oocysts (Tsuji et al., 1995). However, the photolytic destruction of cyanotoxins requires an ultra violet dose of 1500 to

20000 mJ/cm² (Tsuji et al., 1995). At this intensity, photochemical treatment of cyanotoxin is not economically feasible under constant treatment scenarios (Ho et al., 2012).

Activated carbon is used to adsorb soluble organic material such as cyanotoxins. Currently, two forms of activated carbon are used: granular activated carbon, which can be used continuously; and powdered activated carbon, which is used for targeted treatments (Shi et al., 2012; Svrcek and Smith, 2004). Activated carbon is effective in the treatment of microcystins and saxitoxins. Cylindrospermopsins are not well adsorbed, and require additional treatment with chlorine or ultra violet light. After treatment, activated carbon will still have toxins adsorbed until they are chemically destroyed or biodegraded (Ho et al., 2012; Svrcek and Smith, 2004).

New drinking water purification technologies such as electrochemical degradation are being developed to be used in water treatment facilities. Experiments show that an applied current of 100 mA results in a half-life of microcystin of only 2.5 minutes. This process, among many others, is a promising approach to cheap and effective cyanotoxin removal (Svrcek and Smith, 2004).

Possibly the best way reduce cyanotoxins in water is to simply prevent their occurrence. The formation of cyanobacterial bloom formations are limited by many factors, such as availability to limiting nutrients such as phosphorus or nitrogen, water turbidity, light intensity and ambient temperature. Even in optimal conditions, cyanobacterial proliferation can still be controlled if identified while cell concentrations are still low. Several measures have been developed to control cyanobacterial populations, such as limiting use of fertilizers, however algal blooms remain and constant treatment of large water bodies is not feasible. Therefore the most effective solution is to treat potentially harmful communities with algaecides as they are identified. Copper sulphate is a well-studied algaecide, capable of eliminating cyanobacteria in

concentrations as low as 9.2 mg/L (Gleason and Case, 1986; Svrcek and Smith, 2004). Although the use of copper sulphates is environmentally controversial, copper sulphate has been regarded as the choice algaecide as it is economical, effective and relatively safe to both humans and the environment (WHO, 1996). On the other hand, cyanotoxins are predominantly intracellular in healthy cells, but will release their toxins upon lysis or environment stress. Algaecide treatment leads to the dispersal of intracellular toxins. If these toxins reach water treatment intake before degradation, it is unlikely to be completely removed by conventional flocculation, sedimentation and filtration procedures. Thus, algaecides must be used with caution and should only be used when cell numbers are low (Gleason and Case, 1986).

Table 1: General information on toxins produced by cyanobacteria expected in freshwater ecosystems

Toxin	LD₅₀ (i.p. in mouse)	Toxin Producing Organism	Effect	Symptoms
Microcystins	45-1000µg/kg (Yoshida et al., 1997)	<i>Microcystis</i> , <i>Anabaena</i> , <i>Planktothrix</i> , <i>Nostoc</i> , <i>Hepalosiphon</i>	Protein phosphatase inhibitor, liver hemorrhage, tumor-promoter, hepatic insufficiency, alterations of actin microfilaments. (Yoshida et al., 1997)	Weakness, G.I. irritation, reluctance to move, liver damage, mental derangement, photosensitization, asphyxiation, headache, death. (Yoshida et al., 1997) (Codd et al., 2005) (Falconer, 1991)
Nodularins	30-50 µg/kg (Falconer, 1991)	<i>Nodularia spumigena</i>	Protein phosphatase inhibitor, tumor promoter, liver hemorrhage. (Backer, 2002)	Skin and eye irritation. (Codd et al., 2005)
Cylindrospermopsin	2100 µg/kg 200 µg/kg/5-6 d (Ohtani et al., 1992)	<i>Cylindrospermopsis</i> , <i>Aphanizomenon</i> , <i>Umezakia</i>	Protein synthesis inhibitor, culminating effects damage liver, kidney, intestines and lung cells. Carcinogenic effects. (Ohtani et al., 1992)	Asphyxiation, diarrhoea, gasping respiration, enlarged liver, malaise, vomiting, headache. (Ohtani et al., 1992)
Saxitoxin, neosaxitoxin	10-30 µg/kg (Falconer, 1991; Haider et al., 2003)	<i>Anabaena</i> , <i>Aphanizomenon flos aquae</i> , <i>Lyngbya</i> , <i>Cylindrospermopsis</i> , <i>Planktothrix</i>	Blocks sodium channels. (Haider et al., 2003)	Loss of coordination, respiratory failure, numbness of lips, face and extremities, weakness, paralysis, death. (Haider et al., 2003)
Anatoxin-a	200-250 µg/kg (Haider et al., 2003)	<i>Anabaena flos aquae</i> , <i>Anabaena lemmermanii</i> , <i>Anabaena</i> , <i>Planktothrix</i> , <i>Aphanizomenon</i> , <i>Cylindrospermopsis</i>	Blocks post-synaptic depolarization. (Chorus et al., 2000)	Muscle fasciculation, Reluctance to move, cyanosis, convulsions, death. (Chorus et al., 2000)
Anatoxin-a(s)	20 µg/kg	<i>Aphanizomenon flos aquae</i>	Acetylcholinesterase inhibitor. (Chorus et al., 2000)	Respiratory paralysis Hypersalivation, mucoid nasal discharge, tremors, diarrhea, paresis, death. (Chorus et al., 2000)

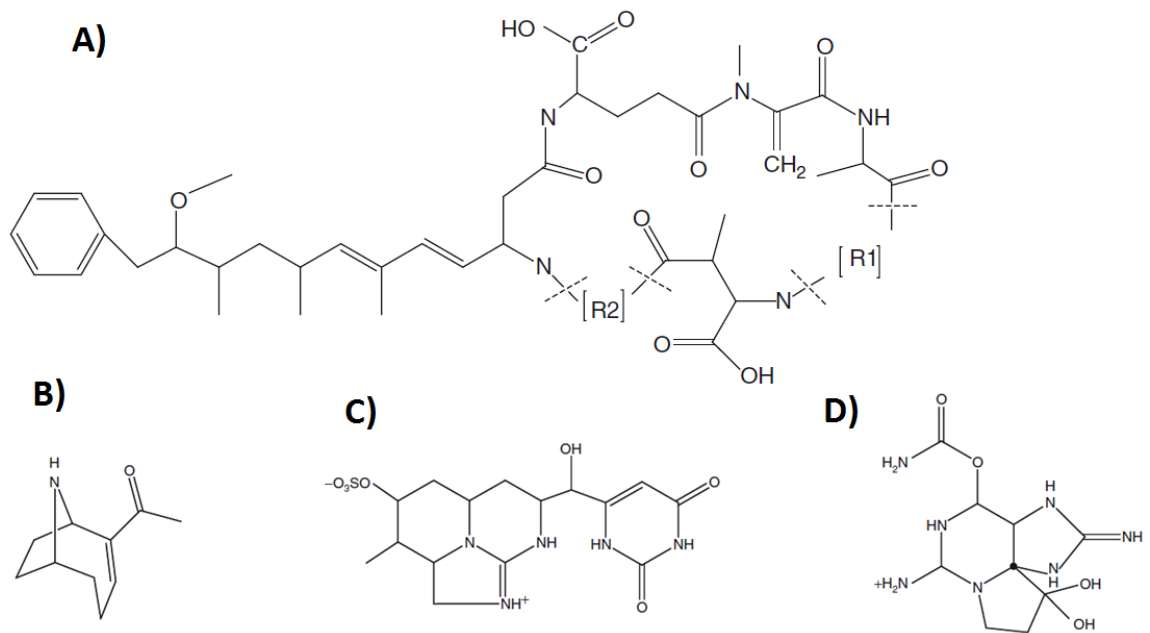


Figure 1 - General structures of A) Microcystin B) Anatoxin-a C) Saxitoxin and D) Cylindrospermopsin (Pelaez et al., 2010).

1.2.6 *Geosmin*

Geosmin (trans-1, 10-dimethyl-trans-9-decalol) is one of the most frequently reported odour-causing compound in the Great Lakes. It is released by several members of cyanobacteria and eubacteria, and gives water an earthy, musty smell and taste (Juttner, 1995). Outbreaks of geosmin are difficult to anticipate and control, as it can be produced in abundance by benthic subsurface populations without the presence of blooms. It is especially difficult to anticipate since it is not species specific. Geosmin is resistant to oxidation and when concentrations are high, elimination by conventional drinking water treatment is often insufficient (Nakamura et al., 2001). Peak geosmin concentrations usually occur in late August and early September. The majority of geosmin is intracellular and will be retained until cells rupture (Watson, 2004). Geosmin has a half-life of approximately 10-25 days, however natural biodegradation of geosmin is accelerated by species such as *Sphingopyxis alaskensis* and *Novosphingobium stygiae*, which breakdown geosmin at an approximate half-life of one day (Hoefel et al., 2006). The cyanobacterium *Anabaena lemmeranii* Richter (*Nostocales*) is suspected to be the predominant producer of geosmin in the Great Lakes (Watson, 2004).

During summer, eastern winds often create a downwelling of the surface water in Lake Ontario. This causes offshore waters to move into the shoreline, transporting offshore planktonic cyanobacteria, toxins and geosmin deep into inshore water, where they can be collected by treatment facility intakes (Rao et al., 2003). Geosmin is poorly removed from these waters, as advanced filtration technology and carbon beds show reduced performance when concentrations of dissolved organic material are high. Sand sedimentation and sand filtration is also very ineffective at removing geosmin and other odour-causing compounds (Watson, 2004).

1.2.7 *Current Methods of Cyanobacterial Monitoring*

Attempts to control both toxin and geosmin episodes have so far been unsuccessful. Monitoring cyanobacteria at both drinking water sources, as well as in problematic areas, for potential cyanobacterial secondary metabolites that may alter water quality is a viable option that has not yet been implemented in the Great Lakes. With further understanding of surrounding cyanobacterial communities, water quality management may prevent blooms, identify the source taxa and prepare treatment facilities accordingly.

Current technologies for the detection of cyanobacterial metabolites rely mainly on ELISA (enzyme-linked immunosorbent assay) and GC-MS (gas chromatography-mass spectrometry). ELISA employs antibodies to detect cyanobacterial metabolites (Chung et al., 1991; Morales-Valle et al., 2010). The lower detection limit of ELISA is for cyanobacterial metabolites is 1 µg/L (Chung et al., 1991). In Ontario suspect algal blooms are reported to the Ministry of Environment most often through the Spills Action Centre public phone line. The local Ministry office will make an assessment and collect samples. Samples are sent to the Ministry lab in Toronto and subject to inverted microscope analysis to determine whether cyanobacteria are present in abundance. Each sample is assessed to identify whether a bloom of algae is present and what genus/species is responsible. The Ministry office is alerted as soon as possible if there are potentially toxin producing cyanobacteria present. Samples are analyzed for toxins (total microcystins using ELISA). Mass spec is then implemented if toxins are detected to determine the specific toxin variant. The local Health Unit then uses these results to determine whether preventative action is necessary. Although this protocol is extremely useful, it relies

heavily on volunteers and not all blooms are reported. Also the process takes time and obviously does not occur in real-time.

PCR (polymerase chain reaction) can be used for strain characterisation or test the presence of toxin-producing genes in water reservoirs. This technique has been used for the characterisation of microcystin-producing species from environmental samples, as only microcystin-producing species encode the three-gene cluster required for microcystin synthesis (Hisbergues et al., 2003). The microcystin synthetase gene cluster spans 55 kb and includes genes for polyketide synthases, peptide synthetases, mixed peptide/polyketide synthases and tailoring enzyme. Amplifying these genes through PCR from environmental samples, one can conclude that microcystin-producing strains are present in the location of sample collection (Rantala et al., 2004). However, this technique does not provide information on the specific species/strain of the organism and cannot relay whether a potential bloom-forming or geosmin-producing strain is present (Hisbergues et al., 2003).

As culture conditions in a laboratory setting differ from the natural environment, cultured cyanobacteria often exhibit different morphological characteristics from their environmental counterparts. As well, many strains of cyanobacteria cannot be cultured in the lab. This leads to many misidentified strains in collected cultures (Nubel et al., 1997). Therefore, genetic markers are necessary for the correct identification of strains. A sequence analysis technique was developed for cyanobacterial strain identification and is now being used to study cyanobacteria. This method utilises genetic markers and PCR in conjunction with a genetic profiling technique called denaturing gradient gel electrophoresis (DGGE). First, primers are designed to specifically amplify the 16S rRNA gene in cyanobacteria. The 16S rRNA gene is a good marker for this procedure as it contains regions which are both highly conserved (for

universal primers) and regions which are very heterogeneous (for high resolution profiling), as well as a large sequence database. DGGE then allows for PCR product separation to profile the diversity of a given cyanobacterial community, or to identify the given strain in the environmental sample. Band separation in DGGE is based on fragment sequence as well as size. To achieve separation, double stranded DNA molecules are partially melted using a linearly increasing gradient of denaturant, which is typically formaldehyde or urea (Muyzer, 1999; Nubel et al., 1997). As the double-stranded DNA passes through the denaturing gradient, nucleotide domains dissociate, which significantly change the dynamic of the strand travelling through the gel matrix. Therefore, slight changes in base pair composition changes melting domain boundaries enough to cause sequences of the same length to migrate to different positions on the gel. As a result, differences in sequence, often as little as one base pair, can be distinguished from its nearly identical counterpart (Nubel et al., 1997). Sequencing is performed after band separation and compared to a 16S rRNA database. *Microcystis*, *Synechococcus*, *Aphanizomenon*, *Planktothrix*, *Prochlorothrix hollandica*, *Oscillatoria* have been successfully identified using this method in Lake Loosdrecht, Netherlands (Zwart et al., 2005).

In general, rRNA genes are more conserved in sequence than protein-coding genes and therefore the physiological diversity of the given organism is not sufficiently reflected by the genetic diversity of the 16S rRNA gene. However, the lack of sequence homogeneity in protein-coding genes would complicate primer design and sequence data interpretation especially from non-axenic samples. Although the 16S rRNA gene contains both conserved and heterogeneous regions, several species/strains contain the same 16S sequence, meaning this gene lacks the resolution to distinguish several species/strains from one another (Nubel et al., 1996; Nubel et

al., 1997). 16S rRNA limited resolution may pose a problem when using it to identify organisms from environmental samples.

The use of fluorometers to relay real-time information of photosynthetic pigment levels in water can be extremely valuable. Fluorometers measure the fluoresced light of chlorophyll-a to monitor the concentration of photosynthetic organisms. The concentrations of chlorophyll-a directly translates into total algal biomass within a given area. Fluorometers also take advantage of pigment diversity (Turner-Designs, 2004). True algae, phytoplankton and cyanobacteria all have unique accessory pigments. These accessory pigments have strong fluorescent signatures that do not interfere with chlorophyll fluorescence, allowing fluorometers to measure algal concentrations and identify them into their respective groups. For cyanobacteria, the fluorometer was most sensitive to the detection of *Cylindrospermopsis raciborskii*, then *Aphanizomenon flos-aquae* and finally *Microcystis aeruginosa*. The lower detection limit for Turner Design's "C3 Submersible Fluorometer" for cyanobacteria is 500 cells/mL (Turner-Designs, 2004). Using fluorometers, cyanobacteria like *Microcystis aeruginosa* can be identified long before they reach peak cell concentrations of approximately 100,000,000 cells/mL. The sensitivity and real-time capabilities of fluorometers make them essential to the management of water quality (Turner-Designs, 2004).

1.2.8 *Cyanobacteria of the Great Lakes*

The Great Lakes are made up of five major freshwater lakes: Lake Superior, Lake Michigan, Lake Huron, Lake Erie and Lake Ontario. These lakes contain approximately one-fifth of the Earth's potable water supply. More than 24 million Canadians and Americans rely on the Great Lakes for drinking and recreational purposes. The past cyanobacterial bloom information

and locations have been collected from published papers, United States Environmental Protection Agency Area of Concerns reports and unpublished drinking water treatment records by Susan Watson and colleagues (Watson, 2004; Watson et al., 2008). Blooms were recorded in two locations in Lake Michigan, three locations in Lake Huron, six locations in Lake Ontario, seven locations in Lake Erie and none for Lake Superior (Watson et al., 2008). These areas of concern can be seen in **Figure 2**.

Lake Huron, Lake Michigan and Lake Superior are the largest of the Great Lakes. These lakes, although affected by cyanobacteria growth, are used by a relatively small population. Therefore they are affected less by human activity and are thus studied less. However, toxins have been detected in these lakes. In particular Lake Huron, which is the second largest Great Lake, has two of the largest freshwater bays; both have reported measurable amounts of microcystin (Watson et al., 2008). Bloom formations are becoming increasingly common in areas of these lakes and it is predicted they may form in other, more northern bays that have so far been unaffected by problematic cyanobacteria (Vanderploeg et al., 2001).

Lake Ontario and Lake Erie are the smallest of the five lakes, however they are used by a much larger population. In these lakes, excessive phosphorus run-off has caused lake-wide eutrophication in the 1950s and 1960s. By the mid-1960s seasonal cyanobacterial blooms in the western basin of Lake Erie were recorded. Those responsible for the blooms include *Aphanizomenon*, *Microcystis* and *Anabaena*. By the 1970s, strict phosphorus discharge limits were enforced to improve the water quality. In the mid 1980s, phosphorus loading decreased significantly, and phytoplankton biomass significantly decreased (Makarewicz et al., 1999). No major blooms occurred in the 1980s, however the cyanobacteria that was most predominant was *Aphanizomenon* (Brittain et al., 2000). By late 1980s, zebra mussels (*Dreissena polymorpha*)

and quagga mussels (*Dreissena bugensis*) colonized the lower Great Lakes and considerably altered the abundance and speciation of phytoplankton through selective filter-feeding and remineralisation by the mussel populations. In the late 1980s, the dominant cyanobacterial genus shifted from *Aphanizomenon* to Microcystin (Vanderploeg et al., 2001; Makarewicz et al., 1999). In 1995 and 1998, Lake Erie experienced large cyanobacterial blooms of *Microcystis* and possibly *Anabaena* and *Planktothrix*, with microcystin concentrations as high as 3.4 µg/L in the western basin. The primary form of microcystin was microcystin-LR followed by microcystin-RR (Dyble et al., 2008). Distinct communities of microcystin-producing *Planktothrix* and *Microcystis* were found in separated locations of Lake Erie, suggesting that populations responsible for the high levels of microcystin are likely to be diverse, and that *Microcystis* is not the only genus synthesising this toxin (Rinta-Kanto and Wilhelm, 2006). Surveys conducted by Environment Canada in 2003 to 2004 in Lake Ontario showed that microcystin concentrations routinely exceeded 1.0 µg/L (Watson et al., 2008). Similar studies in 2003 to 2004 concluded that embayment along New York shore, Hamilton Harbour, waters off of Toronto's shores and near Oswego all contained significant levels of anatoxin-a, though not exceeding 1.0 µg/L (Boyer, 2004).

Gregory Boyer *et al.* (Boyer, 2004) has collected more than 800 samples from the lower Great Lakes between the years of 2000 and 2004, in which 50% tested positive for microcystin-LR and 4% exceeded the WHO's guideline concentration of 1.0 µg/L. Anatoxin-a, cylindrospermopsin and saxitoxin were reported in only 4% of samples and none found in concentrations above 1 µg/L (Boyer, 2004; Boyer, 2008). Little to no toxins were found in the waters in some years, and extremely high toxicity were seen in others. In August of 2003 microcystin concentrations in the samples exceeded 20 µg/L, as a result of a large bloom of

Microcystis aeruginosa formed in the eastern basin of Lake Ontario near Oswego, New York. Several beaches and recreational boating areas were closed due to surface scum washing ashore. Microcystin concentrations from this bloom reached the Onondaga County drinking water intakes in concentrations of 1 µg/L (Boyer, 2008).

Well-defined blooms of *Microcystis aeruginosa* were accountable for blooms occurring in the western basin of Lake Erie, whereas Sandusky Harbor of south western Lake Erie often contained a dense flora of cyanobacteria dominated by *Aphanizomenon flos-aquae*, *Anabaena flos-aquae* or *Cylindrospermopsis raciborskii* (Conroy et al., 2007; Boyer, 2004). A recent analysis shows that the source of some microcystin production is likely from *Planktothrix* species present in the lower, deeper portions of blooms dominated by *Microcystis aeruginosa* (Rintakanto and Wilhelm, 2006). **Table 2** shows a list of cyanobacteria isolates that have been identified from environmental samples taken from the Great Lakes. **Table 3** shows a list of all cyanobacterial species identified within Ontario as well as a list of cyanobacterial species which have been confirmed to form blooms in lakes within Ontario. This information has been provided by Ontario Ministry of the Environment Environmental Monitoring and Reporting Branch.



Figure 2- Graphical representation of the Great Lakes indicating (dots) regions subject to cyanobacterial bloom formations (Watson et al., 2008).

Table 2: Cyanobacterial isolates identified from environmental samples from the Great Lakes. Isolates are recorded from both the Canadian Phycological Culture Centre and published literature

Order	Species	Toxin
Chroococcales	<i>Microcystis aeruginosa</i> (Brittain et al., 2000) (Vanderploeg et al., 2001) (Rinta-Kanto and Wilhelm, 2006) (Watson et al., 2008)	Microcystin
Nostocales	<i>Cylindrospermopsis raciborskii</i> (Conroy et al., 2007)	Cylindrospermopsin Saxitoxin Anatoxin-a
	<i>Aphanizomenon flos-aquae</i> (Sawyer et al., 1968)	Anatoxin-a Saxitox Cylindrospermopsin
	<i>Anabaena lemmermannii</i>	Anatoxin-a(s) Anatoxin-a Microcystins
	<i>Anabaena circinalisi</i>	Anatoxin-a Anatoxin-a(s) Microcystins
	<i>Anabaena</i> sp. (Watson et al., 2008)	Anatoxin-a(s) Anatoxin-a Microcystins
	<i>Anabaena flos-aquae</i>	Anatoxin-a Anatoxin-a(s) Microcystins
Oscillatoriales	<i>Pseudanabaena contorta spec. nova</i> (Kling et al., 2012)	Anatoxin-a
	<i>Planktothrix aghardii</i> (Makarewicz et al., 1999) (Rinta-Kanto and Wilhelm, 2006)	Microcystins Anatoxin-a
	<i>Oscillatoria</i> sp. (Watson et al., 2008)	Anatoxin-a

Table 3: List of cyanobacterial species that have been identified in lakes within Ontario

Cyanobacterial species confirmed in blooms	Cyanobacterial species identified				
<i>Anabaena</i>	<i>Anabaena</i>	<i>Aphanothece</i>	<i>Gloeotrichia</i>	<i>Nostoc</i>	<i>Rhabdoderma lineare</i>
<i>Anabaena lemmermanii</i>	<i>Anabaena affinis</i>	<i>Aphanothece clathrata</i>	<i>Gloeotrichia echinulata</i>	<i>Oscillatoria</i>	<i>Romeria</i>
<i>Anabaena flos-aquae</i>	<i>Anabaena aphanizomenoides</i>	<i>Aphanothece clathrata</i> var. <i>brevis</i>	<i>Gomphosphaeria</i>	<i>Oscillatoria Agardhii</i> var. <i>isothrix</i>	<i>Scytonema</i>
<i>Anabaena planctonica</i>	<i>Anabaena augustunalis</i> ?	<i>Aphanothece nidulans</i>	<i>Gomphosphaeria aponina</i>	<i>Oscillatoria geminata</i>	<i>Spirulina</i>
<i>Anabaena spiroides</i>	<i>Anabaena circinalis</i>	<i>Aphanothece stagnina</i>	<i>Gomphosphaeria lacustris</i>	<i>Oscillatoria limnetica</i>	<i>Spirulina laxa</i>
<i>Aphanizomenon</i>	<i>Anabaena constricta</i>	<i>Arthrospira</i>	<i>Johannesbaptistia</i>	<i>Oscillatoria limnetica</i> var. <i>acicularis</i>	<i>Synechococcus</i>
<i>Aphanizomenon flos-aquae</i>	<i>Anabaena cylindrica</i>	<i>Chroococcus</i>	<i>Limnothrix</i>	<i>Oscillatoria planctonica</i>	<i>Synechocystis</i>
<i>Aphanothece</i>	<i>Anabaena flos-aquae</i>	<i>Chroococcus dispersus</i>	<i>Lyngbya</i>	<i>Oscillatoria redeckeii</i>	<i>Tolypothrix</i>
<i>Gloeotrichia</i>	<i>Anabaena lemmermanii</i>	<i>Chroococcus dispersus</i> var. <i>minor</i>	<i>Lyngbya birgei</i>	<i>Oscillatoria</i> sp. "A"	<i>Trichodesmium</i>
<i>Microcystis</i>	<i>Anabaena limnetica</i>	<i>Chroococcus limneticus</i>	<i>Lyngbya contorta</i>	<i>Oscillatoria</i> sp. "B"	<i>Wollea</i>
<i>Microcystis aeruginosa</i>	<i>Anabaena planctonica</i>	<i>Chroococcus plantonicus</i>	<i>Lyngbya epiphytica</i>	<i>Oscillatoria splendida</i> = <i>Geitlerinema</i>	<i>Woronichinia klingae</i>
<i>Oscillatoria</i>	<i>Anabaena sphaerica</i>	<i>Chroococcus prescottii</i>	<i>Lyngbya limnetica</i>	<i>Oscillatoria utermoehlii</i>	<i>Woronichinia naegeliana</i> (= <i>Coelosphaerium naegelianum</i>)
<i>Lyngbya</i>	<i>Anabaena spiroides</i>	<i>Chroococcus varians</i> ?	<i>Lyngbya vacuolifera</i>	<i>Phormidium</i>	
<i>Planktothrix</i>	<i>Anabaena spiroides</i> var. <i>crassa</i>	<i>Coelomoron</i>	<i>Marsoniella</i>	<i>Phormidium antarcticum</i>	
<i>Planktolyngbya</i>	<i>Anabaena wisconsinense</i>	<i>Coelosphaerium</i>	<i>Merismopedia</i>	<i>Phormidium mucicola</i>	
<i>Scytonema</i>	<i>Anabaenopsis</i>	<i>Coelosphaerium kuetzingianum</i>	<i>Merismopedia elegans</i>	<i>Planktolyngbya</i>	
<i>Woronichinia naegeliana</i> (= <i>Coelosphaerium naegelianum</i>)	<i>Aphanizomenon</i>	<i>Cyanarcus</i>	<i>Merismopedia glauca</i>	<i>Planktothrix</i>	
<i>Nostoc</i>	<i>Aphanizomenon flos-aquae</i>	<i>Cyanococcus</i>	<i>Merismopedia minima</i> ?	<i>Planktothrix agardhii</i> (= <i>Oscillatoria agardhii</i>)	
<i>Gloeocapsa</i>	<i>Aphanizomenon flos-aquae</i> var. <i>gracile</i>	<i>Cyanodictyon</i>	<i>Merismopedia tenuissima</i>	<i>Plectonema</i>	
<i>Chroococcus</i>	<i>Aphanizomenon gracile</i>	Cyanophyceae colonial	<i>Microcystis</i>	<i>Pseudanabaena</i>	
<i>Coelosphaerium</i>	<i>Aphanizomenon issatschenkoi</i>	Cyanophyceae filamentous	<i>Microcystis aeruginosa</i>	<i>Pseudanabaena constricta</i>	
<i>Limnothrix</i>	<i>Aphanizomenon issatschenkoi</i> var. <i>gracilis</i>	Cyanophyceae mixed	<i>Microcystis botrys</i>	<i>Pseudanabaena contorta</i>	
<i>Phormidium</i>	<i>Aphanizomenon schindleri</i>	<i>Cylindrospermum</i>	<i>Microcystis flos-aquae</i>	<i>Pseudanabaena galeata</i>	
	<i>Aphanocapsa</i>	<i>Dactylococcopsis</i>	<i>Microcystis incerta</i>	<i>Radiocystis</i>	
	<i>Aphanocapsa delicatissima</i>	<i>Eucapsis</i>	<i>Microcystis novacekii</i>	<i>Radiocystis geminata</i>	
	<i>Aphanocapsa elachista</i>	<i>Gloeocapsa</i>	<i>Microcystis pulverea</i>	<i>Raphidiopsis</i>	
	<i>Aphanocapsa elachista</i> var. <i>conferta</i>	<i>Gloeothece</i>	<i>Microcystis viridis</i>	<i>Rhabdoderma</i>	
	<i>Aphanocapsa holsatica</i>	<i>Gloeothece linearis</i> var. <i>composita</i>	<i>Microcystis wesenbergii</i>	<i>Rhabdoderma gorskii</i>	

1.3 Project Introduction

Cyanobacteria encompass a large and morphologically heterogeneous group of oxygenic phototrophic bacteria thought to be the original and primary source of atmospheric molecular oxygen (Bekker et al., 2004; Tomitani et al., 2006). Until recently, Cyanobacterial taxonomy was governed by the criteria used in botanical classification. Thus, the Cyanobacteria are largely delineated based upon morphological and biochemical characteristics (Desikachary, 1959; Rippka et al., 1979; Stanier et al., 1978; Geitler, 1925; Geitler, 1932). With improved biochemical understanding, Cyanobacteria have more recently been reclassified many times into different systems comprised of several different and debated orders/subsections (Castenholz, 2001; Cavalier-Smith, 2002; Drouet, 1981; Euzéby, 1997; Florenzano et al., 1986; Anagnostidis and Komarek, 1985; Sayers et al., 2010; Wilmotte and Herdman, 2001). Currently, the use of 16S rRNA sequence data is one of the most widely implemented approaches to understanding microbial taxonomy (Embley and Stackebrandt, 1997; Rossello-Mora and Amann, 2001). While 16S rRNA-based phylogenies are useful in the classification of many bacterial groups, Cyanobacterial phylogeny based on 16S rRNA shows little correlation between the morphological characteristics which are the primary criterion used in the taxonomy of the phylum. Additionally, 16S rRNA gene-based trees group Cyanobacteria into 14 clusters, however the evolutionary and taxonomic relevance is not clear (Wilmotte and Herdman, 2001).

Invasion and proliferation of heterocyst-differentiating Cyanobacteria to new habitats have placed a particular importance on these bacteria, as their spread may significantly alter the function of a given ecosystem and hinder propagation of native communities (Christoffersen, 1996; Padisak, 1997; Wiedner et al., 2007). As a result of the increased eutrophication in fresh water and marine environments, the frequency and intensity of toxic Cyanobacterial blooms has

increased (Anderson et al., 2002). The resulting abundance of cyanotoxins greatly affects water quality and causes significant damage to global aquaculture and fishing industries (Codd, 2000; Steffensen, 2008). As well, contact with or consumption of cyanotoxins in humans and animals can cause skin irritation, liver damage, tumor promotion and muscular paralysis which may result in death (Kaebernick and Neilan, 2001). Thus, heterocystous Cyanobacteria are of increasing importance to both clinical and ecological communities (Pitois et al., 2000).

Cyanobacterial genera that differentiate into heterocysts all belong to and encompass the *Nostocales* and *Stigonematales* orders based on the traditional/accepted classification system or subsections (IV) and (V) according to the proposed classification (Rippka et al., 1979; Stanier et al., 1978; Castenholz, 2001); Cavalier-Smith, 2002). These orders encompass the class Hormogoneae, which is described to be filaments that multiply vegetatively by hormogonia; usually with heterocysts (Cavalier-Smith, 2002). Within this class, the legitimate description which differentiates *Nostocales* from *Stigonematales* details *Nostocales* as unbranched (false branching) and *Stigonematales* as branched (true branching) (Cavalier-Smith, 2002). Most other classification systems propose a relatively similar description of these two orders detailing *Nostocales* as filamentous, heterocystous, uniseriate bacteria that divide by binary fission in one plane at a right angle to the long axis of the trichomes (false branching), while describing *Stigonematales* are classified as filamentous, heterocystous bacteria with longitudinal cell division as well as transverse cell division causing occasional true branching (Castenholz, 2001). *Nostocales* and *Stigonematales* are divided into 35 and 14 genera, respectively, based on characteristics such as motility, trichome length and shape, sheathed and sheath thickness, akinete and heterocyst position and formation, gas vesicles, habitat, false branch formation and 16S rRNA tree branching, however no genera are recognized under the bacteriological code

(Castenholz, 2001; Sayers et al., 2010). 16S rRNA sequence data indicates that all heterocystous members of Cyanobacteria form a coherent monophyletic cluster separate from non-heterocystous differentiating members (Wilmotte and Herdman, 2001; Giovannoni et al., 1988; Honda et al., 1999; Turner et al., 1999). However, within this monophyletic cluster the separation between *Nostocales* and *Stigonematales* is not resolved (Wilmotte and Herdman, 2001). The genetic distance between many *Nostocales* genera is equal to or greater than those that separate them from genera of *Stigonematales* (Wilmotte and Herdman, 2001). Genera of *Nostocales* and *Stigonematales* have been noted to appear in polyphyletic intermixed clusters in the 16S rRNA tree and of a common ancestor; therefore 16S rRNA sequence data does not justify the separation of these two orders based on true or false branching (Wilmotte and Herdman, 2001; Gugger and Hoffmann, 2004). However, evidence solely based on 16S rRNA branching is insufficient to fully understand the relationships shared between these orders. Gugger and Hoffmann suggest *Stigonematales* form one T-branching and one separate Y-branching clade that is divided by *Nostocalean* members. The ever-evolving and debated classification system clearly indicates that the taxonomy of members belonging to the *Nostocales* and *Stigonematales* orders is inadequate. Furthermore, it has been proposed that several species of Cyanobacteria are currently misidentified as a result of morphological characteristics loss upon culturing (Komarek, 1994). Therefore, it is necessary to find new sequence-based characteristics that can offer a reliable means to resolve the taxonomic and evolutionary relationships, as well as provide a useful means to identify species within these orders (Komarek, 2002).

With the increasing availability of genome sequences, we can now examine the phylogeny of heterocystous Cyanobacteria through the use of multiple protein concatenated

trees (Gupta, 2009; Wu et al., 2009), as well as comparative genomic analyses, to identify different types of molecular markers which provide a more reliable means of deciphering the phylogenetic relationship between *Nostocales* and *Stigonematales* (Gao et al., 2009; Gupta, 2000; Gupta, 2009; Wu et al., 2009). This approach requires extensive DNA databases that are only now sufficiently comprehensive to allow analysis. Our recent work has concentrated on concatenated sequences of highly conserved proteins to create phylogenetic trees and the identification of conserved signature inserts or deletions (i.e. Indels) (CSIs) and conserved signature proteins (CSPs) that are specifically present in bacterial groups that share a common ancestor (Gao et al., 2009; Gupta and Mok, 2007; Gupta, 1998; Gupta and Griffiths, 2002; Rokas et al., 2003). CSPs and CSIs that are uniquely present in Cyanobacteria were recently discovered (Gupta, 2009; Gupta and Mathews, 2010). These markers delineate the phylum into distinct taxa in clear molecular terms. Previously identified CSIs and CSPs that are specific for various Cyanobacterial clades strongly support the division of genome sequenced Cyanobacterial species/strains into three main clades at the highest level (Gupta, 2009; Gupta and Mathews, 2010). These clades were also clearly distinct from each other in phylogenetic trees constructed based upon large datasets of concatenated proteins (Gupta, 2009; Gupta and Mathews, 2010). Additionally, these clades also correspond to three main subclasses of Cyanobacteria proposed by Hoffmann et al. (Gupta, 2009; Gupta and Mathews, 2010; Hoffmann, 2005; Hoffmann et al., 2005). However, due to a limited number of genome sequences available, previous studies mainly focused on the higher taxonomic clades of Cyanobacteria. Due to the absence of any sequenced representative from the order *Stigonematales*, the relationship of the order *Nostocales* to *Stigonematales*, both of which are part of the proposed subclass *Nostocophycidae*, was also not examined in the earlier work. Nevertheless, recently genome

sequence information has become available for a number of new *Nostocales* species/strain as well as from a member of the order *Stigonematales*, enabling more detailed investigations on these heterocystous Cyanobacteria. In this study, we have carried out detailed comparative genomic analyses combined with multi-protein phylogenetic analysis to identify molecular markers that are specific for these orders of Cyanobacteria. Our work reported here has identified >100 CSIs and CSPs that are specific to either all *Nostocales/Stigonematales* or a number of distinct subgroups/clades of these bacteria.

Chapter 2: Methods

2.1 Phylogenetic Sequence Analysis

Phylogenetic analysis was performed on a concatenated sequence alignment of 45 highly conserved proteins that are present in most living bacteria (Supplementary Table 1) (Harris et al., 2003). Amino acid sequences were used for all publicly available (as of August 9, 2012) fully sequenced species of Cyanobacteria, as well as *Fischerella sp. JSC-11*, *Raphidopsis brookii D9*, *Cylindrospermopsis raciborskii CS-505* and *Nodularia spumigena CCY9414*. In addition, our dataset also included *Staphylococcus aureus subsp. aureus Mu50* and *Bacillus subtilis str. 168*, which served as outgroups in phylogenetic analysis (Table 4). Multiple sequence alignments were created using Clustal_X 2.1 (Larkin et al., 2007). Poorly aligned regions of the concatenated alignment were removed using Gblocks 0.91b (Castresana, 2000). The resulting concatenated sequence, which contained 15413 aligned positions, was used to construct a neighbor-joining phylogenetic tree and a maximum-likelihood phylogenetic consensus tree with 100 bootstrap replications based on the Jones-Taylor-Thornton model of evolutionary gene change with uniform rates and with three discrete categories of gamma distributed substitution

rates, respectively, using Mega 5.05 (Jones et al., 1992; Nei and Kumar, 2000; Tamura et al., 2011; Tamura et al., 2004).

A phylogenetic analysis was also performed on a 16S rRNA alignment of sequences from at least one good quality (encompassing at least 1200 bp) representative from every genera (except *Riveria*) in the *Nostocales* and *Stigonematales* orders available from The SILVA ribosomal RNA gene database project, as well as *Microcystin aeruginosa NIES-843*, which served as an outgroup (Quast et al., 2013). The 16S rRNA sequences of all 48 species were used to create a neighbor-joining phylogenetic tree of 1000 bootstrap replications based on the maximum composite likelihood model of evolutionary gene change with uniform rates of substitution with Mega 5.05 (Tamura et al., 2011; Tamura et al., 2004).

2.2 Identification of Conserved Signature Indels

To locate CSIs, BLASTp searches were initially performed on proteins in the genome of *Nostoc sp. PCC 7120*. These searches were performed using all of the proteins from accession numbers 17227498 to 177228576 and 17231359 to 17232863. Based on these BLASTp searches, for each protein sequences of 10-15 high scoring homologues were obtained from the available *Nostocales*, as well as several outgroup species. Multiple sequence alignments of these proteins were then constructed using Clustal_X 2.1 (Larkin et al., 2007). The resulting alignments were then examined for conserved indels that are flanked on both sides by at least 5-6 identically conserved amino acids in the neighbouring 30-40 residues. Indels that were not flanked on either side by conserved regions were not further considered, as they do not provide useful molecular markers and could arise from alignment artefacts (Gupta, 2009; Gupta, 2001). Species distribution patterns of all potentially useful indels thus identified were examined further by

performing more detailed BLASTp searches on short sequence regions (approximately 60-80 aa long) containing the indel and its flanking conserved regions. These BLAST searches were carried out using the nr (non-redundant) database and top 250 BLAST hits were examined for the presence or absence of similar indels to determine the specificities of different indels (Gupta, 2009). In this work, we report the results of those CSIs that in most cases are specifically present in species from the orders *Nostocales* and *Stigonematales* and not present in other Cyanobacteria or other bacteria.

2.3 Identification of Specificity for Previously Reported Conserved Signature Proteins

Our earlier work has identified a number of CSPs those were indicated to be specific for the species from the order *Nostocales* (Gupta and Mathews, 2010). As sequence information for many additional Cyanobacterial species/strains (see Table 4) has now become available, BLASTp searches were conducted on all of the proteins from the Table 4 and supplementary table 5 of this earlier publication to re-examine their species specificity. A protein was considered to be specific for a given clade, if all significant BLAST hits (E value $<1 \times 10^{-4}$ or as determined by PSI-BLAST) for this protein are from the indicated group of bacteria.

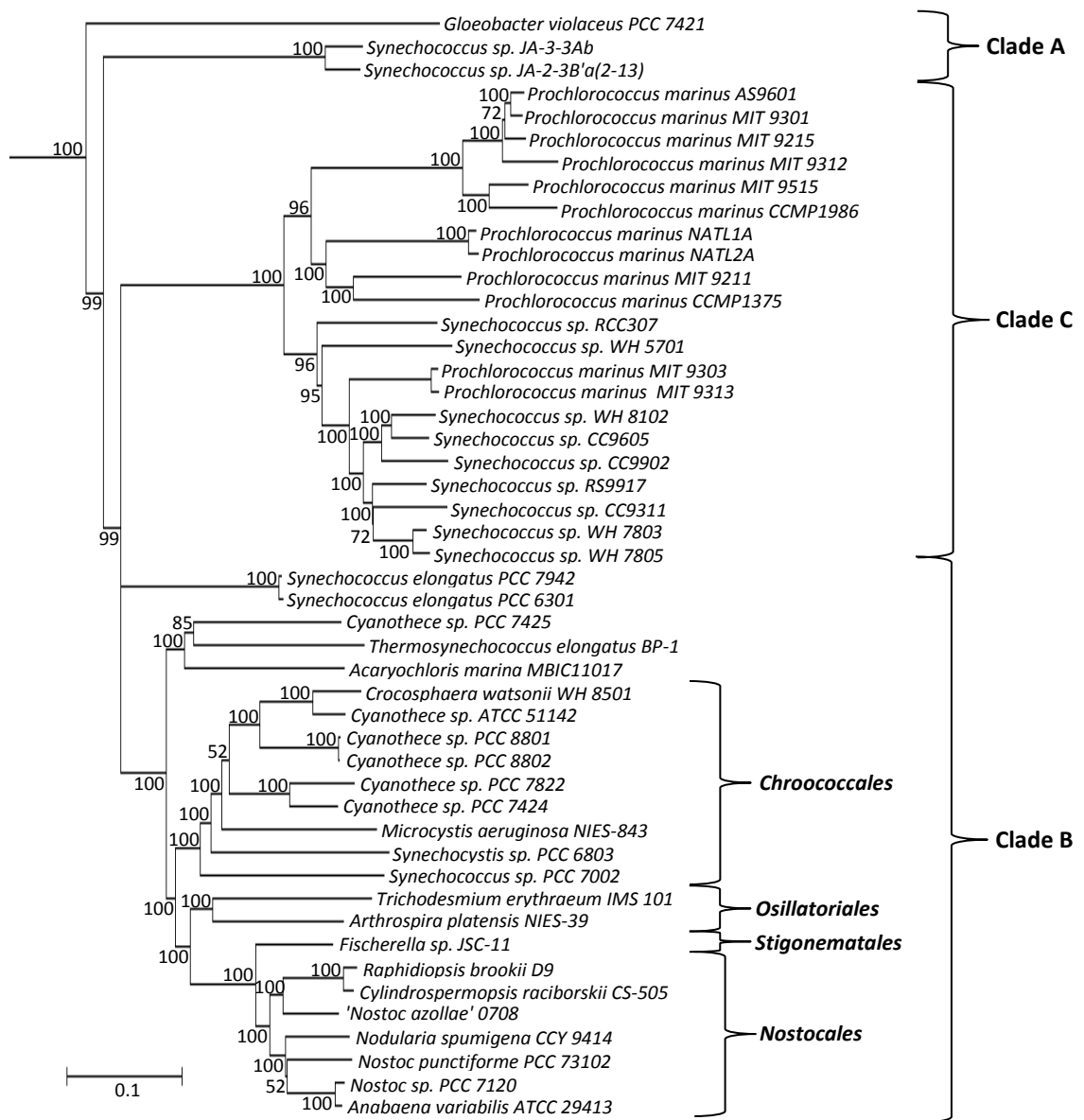


Figure 3: A maximum-likelihood consensus tree based on 45 concatenated sequences for 48 species of Cyanobacteria. The tree was rooted using sequences for *Staphylococcus aureus* subsp. *aureus* Mu50 and *Bacillus subtilis* str. 168. Numbers located at nodes indicate the bootstrap values out of 100. Major clades and subclades of Cyanobacteria resolved in the tree are indicated.



Figure 4: A neighbor-joining distance tree based on 16S rRNA sequences, representing all genera of *Nostocales* and *Stigonematales* (except *Riveria*) and rooted with *Microcystin aeruginosa* NIES-843. Boxed members are those which have sequenced genomes available. ^N and ^S denote members belonging to *Nostocales* and *Stigonematales*, respectively. * indicates members belonging to the *Nostocaceae* family.

Table 4: Sequence characteristics of featured Cyanobacterial genomes.

Organisms	GenBank Accession No.	Clade	Size (Mpb)	No. of proteins	References
<i>Acaryochloris marina</i> MBIC11017	CP000828.1	B	8.36	8383	(Swingley et al., 2008b)
<i>Anabaena variabilis</i> ATCC 29413	CP000117.1	B	7.11	5710	DOE JGI ^b
<i>Arthrospira platensis</i> NIES-39 ^f	AP011615.1	B	6.79	6630	(Fujiisawa et al., 2010)
<i>Crocospaera watsonii</i> WH 8501 ^a	AADV000000000.2	B	6.24	5958	DOE JGI ^b
<i>Cyanothece</i> sp. ATCC 51142	CP000806.1	B	5.46	5304	(Welsh et al., 2008)
<i>Cyanothece</i> sp. PCC 7424 ^f	CP001291.1	B	6.55	5710	DOE JGI ^b
<i>Cyanothece</i> sp. PCC 7425 ^f	CP001344.1	B	5.79	5327	DOE JGI ^b
<i>Cyanothece</i> sp. PCC 7822 ^f	CP002198.1	B	7.84	6642	DOE JGI ^b
<i>Cyanothece</i> sp. PCC 8801	CP001287.1	B	4.79	4367	DOE JGI ^b
<i>Cyanothece</i> sp. PCC 8802 ^f	CP001701.1	B	4.8	4444	DOE JGI ^b
<i>Cylindrospermopsis raciborskii</i> CS-505 ^{a,f}	ACYA000000000.1	B	3.88	3449	FLI Jena ^c
<i>Fischerella</i> sp. JSC-11 ^{a,f}	AGIZ000000000.1	B	5.38	4627	DOE JGI ^b
<i>Gloeobacter violaceus</i> PCC 7421	BA000045.2	A	4.66	4430	(Nakamura et al., 2003)
<i>Microcystis aeruginosa</i> NIES-843	AP009552.1	B	5.84	6312	(Kaneko et al., 2007)
<i>Nodularia spumigena</i> CCY 9414 ^a	AAVW000000000.1	B	5.32	4860	GBM Foundation ^d
' <i>Nostoc azollae</i> ' 0708 ^f	CP002059.1	B	5.49	3,651	(Ran et al., 2010)
<i>Nostoc punctiforme</i> PCC 73102	CP001037.1	B	9.06	6689	DOE JGI ^b
<i>Nostoc</i> sp. PCC 7120	BA000019.2	B	7.21	6129	(Kaneko et al., 2001)
<i>Prochlorococcus marinus</i> AS 9601	CP000551.1	C	1.67	1920	GBM Foundation ^d
<i>Prochlorococcus marinus</i> CCMP1375	AE017126.1	C	1.75	1883	(Dufresne et al., 2003)
<i>Prochlorococcus marinus</i> CCMP1986	BX548174.1	C	1.66	1717	(Rocap et al., 2003)
<i>Prochlorococcus marinus</i> MIT 9211	CP000878.1	C	1.69	1854	GBM Foundation ^d
<i>Prochlorococcus marinus</i> MIT 9215	CP000825.1	C	1.74	1982	DOE JGI ^b
<i>Prochlorococcus marinus</i> MIT 9301	CP000576.1	C	1.64	1906	GBM Foundation ^d
<i>Prochlorococcus marinus</i> MIT 9303	CP000554.1	C	2.68	2997	GBM Foundation ^d
<i>Prochlorococcus marinus</i> MIT 9312	CP000111.1	C	1.71	1810	DOE JGI ^b
<i>Prochlorococcus marinus</i> MIT 9313	BX548175.1	C	2.41	2269	DOE JGI ^b
<i>Prochlorococcus marinus</i> MIT 9515	CP000552.1	C	1.7	1905	GBM Foundation ^d
<i>Prochlorococcus marinus</i> NATL1A	CP000553.1	C	1.86	2193	GBM Foundation ^d
<i>Prochlorococcus marinus</i> NATL2A	CP000095.2	C	1.84	2162	DOE JGI ^b
<i>Raphidiopsis brookii</i> D9 ^f	CP000095.2	B	3.19	3007	DOE JGI ^b
<i>Synechococcus elongatus</i> PCC 6301	AP008231.1	B	2.7	2523	(Sugita et al., 2007)
<i>Synechococcus elongatus</i> PCC 7942	CP000100.1	B	2.74	2662	DOE JGI ^b
<i>Synechococcus</i> sp. CC9311	CP000435.1	C	2.61	2892	(Palenik et al., 2006)
<i>Synechococcus</i> sp. CC9605	CP000110.1	C	2.51	2645	DOE JGI ^b
<i>Synechococcus</i> sp. CC9902	CP000097.1	C	2.23	2306	DOE JGI ^b
<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	CP000240.1	A	3.05	2862	TIGR ^e
<i>Synechococcus</i> sp. JA-3-3Ab	CP000239.1	A	2.93	2760	TIGR ^e
<i>Synechococcus</i> sp. PCC 7002	CP000951.1	B	3.41	3187	Penn. State University
<i>Synechococcus</i> sp. RCC307	CT978603.1	C	2.22	2534	Institut Pasteur
<i>Synechococcus</i> sp. RS9917 ^a	AANP000000000	C	2.58	2770	GBM Foundation ^d
<i>Synechococcus</i> sp. WH 5701 ^a	AANO000000000	C	3.04	3346	GBM Foundation ^d
<i>Synechococcus</i> sp. WH 7803	CT971583.1	C	2.37	2533	Institut Pasteur
<i>Synechococcus</i> sp. WH 7805 ^a	AAOK000000000	C	2.62	2883	GBM Foundation ^d
<i>Synechococcus</i> sp. WH 8102	BX548020.1	C	2.43	2519	(Palenik et al., 2003)
<i>Synechocystis</i> sp. PCC 6803	BA000022.2	B	3.95	3575	(Kaneko et al., 1996)
<i>Thermosynechococcus elongatus</i> BP-1	BA000039.2	B	2.59	2476	(Nakamura et al., 2002)
<i>Trichodesmium erythraeum</i> IMS 101	CP000393.1	B	7.75	4451	DOE JGI ^b

^a Associated genomes currently under scaffold/contig status^b Department of Energy Joint Genome Institute

^c Fritz-Lipmann-Institute

^d The Gordon and Betty Moore Foundation

^e The Institute of Genome Research

^f Species/Strains are not represented in previous phylogenetic studies of our lab

Chapter 3: Results

3.1 Phylogenetic Analyses of *Nostocales* and *Stigonematales*

Current understanding of evolutionary relationships amongst microorganisms relies primarily on phylogenetic trees based on the 16S rRNA gene and individual protein sequences. However, phylogenetic analyses based on a combined dataset of sequences for many proteins can more reliably portray phylogenetic relationships than those based on any single gene or protein (Wu et al., 2009; Rokas et al., 2003; Ciccarelli et al., 2006). Phylogenetic trees for 44 Cyanobacterial species/strains based on concatenated sequences for 44 conserved proteins were described in our recent work (Gupta and Mathews, 2010). The trees based upon concatenated sequences for a limited number of Cyanobacteria (between 14-24) have also been reported by other investigators (Shi and Falkowski, 2008; Swingley et al., 2008a; Zhaxybayeva et al., 2006; Sanchez-Baracaldo et al., 2005). While these earlier studies provide important information regarding the evolutionary relationships among Cyanobacteria, as they contain only a limited number of *Nostocales* species/strains and no member of the order *Stigonematales*, they did not provide any information regarding the relationship among the species/strains belonging to the order *Nostocales* and *Stigonematales*. Hence, in the present work, we have constructed an updated phylogenetic tree for 48 Cyanobacterial species/strains, which include information for seven members of the order *Nostocales* and also *Fischerella* sp. JS-11 belonging to the order *Stigonematales*.

Phylogenetic trees for the 48 Cyanobacteria species/strains for which sequences are currently available (Table 4) were constructed based on concatenated amino acid sequences of 45 conserved proteins and rooted with *Staphylococcus aureus* subsp. *Aureus* Mu50 and *Bacillus*

subtilis str. 168. The trees were constructed using both maximum-likelihood (ML) and neighbor-joining (NJ) algorithms and the tree obtained using ML method is shown in Figure 3, whereas the information for NJ tree is provided in supplementary Figure 1. The branching pattern of different Cyanobacteria in both the ML and NJ trees were very similar, most of the nodes were supported by bootstrap values between 70-100%. The relationships among different Cyanobacterial lineages observed in these trees are very similar to those observed in our earlier (Gupta and Mathews, 2010; Shi and Falkowski, 2008; Swingley et al., 2008a; Sanchez-Baracaldo et al., 2005). Several genera such as *Synechococcus*, *Prochlorococcus*, *Cyanothece* and *Nostoc* were observed to be polyphyletic and species/strains from them branched into a number of different clusters. The three distinct clades of Cyanobacteria observed in earlier studies that are also supported by several previously identified CSIs and CSPs, were also clearly distinguished in the tree shown in Figure 3 (Wu et al., 2009; Gupta and Mathews, 2010). The first clade, referred to as clade A, forms the deepest branching lineage on this tree and is comprised of the thylakoid lacking *Gloeobacter violaceus* and two diazotrophic *Synechococcus* sp. *JA-3-3AB* and *JA2-3-B'a*. Most other strains used in this study can be grouped into two main clades, Clade B and Clade C. Clade C presently contains only marine unicellular *Prochlorococcus* and *Synechococcus* strains. Clade B, constitutes the second main clade which encompasses an array of diverse Cyanobacteria grouped into subclades such as *Chroococcales* (*Synechococcus*, *Cyanothece*, *Crocospaera*, *Microcystis*, *Synechocystis*), *Oscillatoriales* (*Trichodesmium* and *Arthrospira*) and *Nostocales* (*Nodularia*, *Nostoc*, *Anabaena*, *Cylindrospermopsis*, *Raphidiopsis*). The Cyanobacterial species/strains comprising of the freshwater thermophilic unicellular *Thermosynechococcus elongatus* BP-1, marine unicellular chlorophyll d producing *Acaryochloris marina* and freshwater diazotrophic unicellular *Cyanothece* sp. PCC 7425 formed a separate deep branching clade that

formed the deepest branch of the Clade B Cyanobacteria. As observed in earlier studies, the two strains of the freshwater unicellular Cyanobacteria *Synechococcus elongatus* branched distinctly from the other three main clades and it may comprise a distinct deep branching lineage of Cyanobacteria.

In terms of the relationship among the species/strains from the order *Nostocales* and *Stigonematales*, members of these two orders formed a strongly supported clade that was separated from all other Clade B Cyanobacteria by a long branch. *Fischerella* sp. *JSC-11*, which is sole representative of the order *Stigonematales*, formed an outgroup of the *Nostocales* clade. The available species/strains of *Nostocales* were also found to form a number of distinct subclades. Within one of these subclades, a close affinity was observed between *Cylindrospermopsis raciborskii* *CS-505* and *Raphidiopsis brookii* *D9* and a clade consisting of these two species grouped together with '*Nostoc azollae*' *0708*. In another subclade, a close relationship was observed between the *Nostoc* sp. *PCC7120* and *Anabaena variabilis* *ATCC2914*, and these species/strains branched with the *Nostoc punctiforme* *PCC73102* and *Nodularia spumigena* *CCY9414*. However, the low bootstrap scores for this larger subclade suggest that the branching position of *Nostoc punctiforme* *PCC73102* and *Nodularia spumigena* *CCY9414* in relation to other *Nostocales* is not resolved.

We have also constructed a NJ tree based on 16S rRNA gene sequences of 47 members, representing all genera of *Nostocales* and *Stigonematales*. This tree, which also included sequence information for different species/strains from these orders that are part of the protein tree, was rooted with *Microcystin aeruginosa* *NIES-843* (Figure 4). The branching topology of different species/strain in the rRNA tree is similar to those observed in earlier studies for

members of these orders (Gugger and Hoffmann, 2004). In this tree, neither species/strains of the *Nostocales* or *Stigonematales* orders formed a monophyletic lineage. The members of both orders were extensively intermixed in the constructed tree. In addition, both T and Y-branching members of *Stigonematales* do not form monophyletic clades, with members of both branch type as well as *Nostoclean* species/strains branching intermixed. Importantly, the sequenced species used in the concatenated protein tree and for comparative genomic studies, were found to be parts of different clades, indicating that they represent this large assemblage of heterocystous Cyanobacteria. The close affinity of *Nostoc sp. PCC 7120* to *Anabaena variabilis ATCC 29413* and of *Cylindrospermopsis raciborskii CS-505* to *Raphidiopsis brookii D9* observed in the protein tree was also supported by the rRNA gene tree.

3.2 Conserved Signature Indels Specific for *Nostocales/Stigonematales* and *Nostocales*

The main thrust of this study was to perform comparative genomic analyses to identify conserved signature indels (CSIs) that are specific for some or all members of the order *Nostocales* commonly shared with the *Fischerella sp. JSC-11* representing the order *Stigonematales*. The identification and evaluation of these CSIs was carried out independently of the phylogenetic studies. Hence, the species distribution patterns of the identified CSIs and the relationships revealed by them provided independent evidence for the shared derived characteristics of the species/strains from these orders of Cyanobacteria. Our work has identified 70 CSIs in different proteins that are specific for either some of the sequenced members of the order *Nostocales/Stigonematales*. A brief description of the species-specificity of these CSIs is provided below.

Of the discovered CSIs, 17 were specific for all of the sequenced species/strains from the

orders *Nostocales* and *Stigonematales* and are not found in the homologous proteins from any other organisms. One example of a CSI that is specific for the *Nostocales/Stigonematales* group of species is shown in Figure 5A. In this case, a five amino acid insert in a highly conserved region of the ribosomal protein S3 is uniquely present in all available sequences from *Nostocales/Stigonematales*, but it is not found in any other Cyanobacteria or other bacteria (Figure 5A). Another example of a CSI that is specific for all sequenced *Nostocales/Stigonematales* is a two amino acid insertion in a conserved region of ATP synthase epsilon subunit (Figure 5B), a protein that plays an essential role in ATP synthesis. Similar to these two CSIs, 15 other CSIs were identified that are uniquely found in the homologs of all available *Nostocales* and *Stigonematales* species/strains and not in any other organism. Some characteristics of these CSIs are summarized in Table 5 and the sequence information for them, showing the presence of the indicated CSIs, are provided in Supplementary Figures 2-16. In addition to these CSIs, our analyses have also identified one CSI consisting of a two aa insert in the hypothetical protein Npun_F3904 that is uniquely present in all sequenced *Nostocales* species, but it is not found in the *Fischerella sp. JSC-11* as well as other Cyanobacteria. The characteristics of this CSI are also summarized in Table 5 and its sequence alignment can be found in Supplementary Figure 17.

3.3 Conserved Signature Indels Specific for *Nostocales* Subclades

In addition to the CSIs that are commonly shared by all available *Nostocales/Stigonematales*, our analyses have also identified 55 CSIs which are specific for smaller subclades of the *Nostocales* providing information about the relationships/affinities among these species/strains. Based upon their branching in the concatenated protein trees, the

sequenced *Nostocales* species can be grouped into two main clades. The first of these two subclades, which contains *Cylindrospermopsis raciborskii*, *Raphidopsis brookii* D9 and '*Nostoc azollae*', is independently supported by five of the discovered CSIs that are uniquely shared by all three species/strains that are part of this clade (Table 6A). One example of a CSI specific to this subclade of species comprising of a seven amino acid insert in the recombination protein O is show in Figure 6A. Within this subclade, *Raphidiopsis brookii* D9 and *Cylindrospermopsis raciborskii* CS-505 show a close relationship in both protein tree and the rRNA tree. Importantly, this close and specific relationship between these two species is also strongly and independently supported by 40 discovered CSIs that are uniquely shared by these two species/strains, but not found in any other Cyanobacteria including '*Nostoc azollae*'. One example is a CSI consisting of a seven amino acid insert in a highly-conserved region of the protein O-succinylbenzoyl-CoA synthetase, which is specifically present in these two *Nostocales* species and is shown in Figure 6B. Information for other CSIs that are also uniquely shared by these two *Nostocales* species is summarized in Table 6B and sequence alignments for these groups are shown in Supplementary Figure 18-60.

Another subclade of *Nostocales* species/strains that is supported by the discovered CSIs corresponds to the species *Nodularia spumigena* CCY9414, *Nostoc punctiforme* PCC73102, *Nostoc sp.* PCC 7120 and *Anabaena variabilis* ATCC2914. Our analyses have identified three CSIs that are uniquely present in all species/strains from this subclade (Table 7A). One example of a CSI specific to this subclade consisting of a one amino acid insert in a highly conserved region of the protein glucose-1-phosphate thymidyltransferase (Figure 7A), an enzyme responsible for the catalysis of an essential step in the production of l-rhamnose. Within this clade, comprised a closer relationship among the species/strains *Nostoc punctiforme* PCC 73102, *Nostoc sp.* PCC

7120 and *Anabaena variabilis* ATCC 29413 is supported by four of the identified CSIs that are exclusively present in the protein homologs from these bacteria (Table 7B). One of the CSIs is specific to this subclade consisting of a nine amino acid insertion in a highly conserved region of the protein Alpha-amylase is shown in Figure 7B. *Nostoc sp. PCC 7120* and *Anabaena variabilis* ATCC 29413 form a strongly supported clade in the phylogenetic trees. A close relationship between these species/strains is also independently supported by five of the identified CSIs in different proteins that are uniquely present in the homologs of these two species (Table 7C). Some characteristics of different CSIs that are specific for smaller subclades of *Nostocales* are listed in Table 7C and the sequence alignments of the protein showing these CSIs are provided in Supplementary Figures 61-70.

3.4 Conserved signature Indels Specific to Individual Members of Sequenced *Nostocales*/*Stigonematales*

In addition to the CSIs that are commonly shared by at least two or more sequenced *Nostocales* species, our analyses have also identified 83 CSIs that are only present in a single sequenced species/strain of the *Nostocales* or *Stigonematales* (Supplementary Table 2). Although the species/strain specificities of these CSIs need to be confirmed, they provide potential useful molecular markers for the detection of individual closely related members of this group of bacteria. Many of these CSIs are likely present only in a narrow spectrum of closely related Cyanobacterial members. Information on CSIs that are uniquely shared by a single member of *Nostocales*/*Stigonematales* is summarized in Supplementary Table 2 and sequence alignments for these CSIs are shown in Supplementary Figure 71-150.

3.5 Conserved Signature Proteins (CSPs) Specific for Clades of *Nostocales* and *Stigonematales*

In our earlier work, BLASTp searches were performed on each open reading frame of the *Nostoc sp. PCC 7120* genome to identify proteins specific to the *Nostocales* order (Gupta and Mathews, 2010). This study identified several CSPs, which at that time, were specific to all sequenced *Nostocales* (*Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413*, *Nostoc punctiforme PCC73102* and *Nodularia spumigena CCY9414*). Since sequence information for a number of other *Nostocales* species/strain as well as other Cyanobacteria has become available in the past few years, BLASTp searches were performed on each of these proteins to determine their current cladal or species specificities (Table 8). These studies have revealed that of the CSPs identified in our earlier work, significant BLAST hits for 12 proteins are observed only for all sequenced *Nostocales* members as well as *Fischerella sp. JSC-11*, whereas six other proteins are specific only for the *Nostocales* species/strains. In addition, 10 CSPs are specific for a subclade comprising of the species/strains *Nodularia spumigena CCY9414*, *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413* and *Nostoc punctiforme PCC 73102*; 13 other proteins are only found in *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413* and *Nostoc punctiforme PCC 73102*; and 12 CSPs are specifically present in the species *Nostoc sp. PCC 7120* and *Anabaena variabilis ATCC 29413*. All of these CSPs are of unknown function and information regarding their accession numbers and cladal specificity is provided in Table 8.

Table 5: Conserved signature indels specific for all *Nostocales* and *Stigonematales*

Protein Name	GeneBank Identifier	Figure #	Indel Size	Indel Position
30S ribosomal protein S3	354565800 ^F	Fig # 5A	5 aa ins	29-68
ATP synthase epsilon subunit	17232530 ^N	Fig # 5B	2 aa ins	36-80
Hypothetical protein Npun_R4929	17231417 ^N	S. Fig # 2	5 aa ins	484-531
Outer membrane adhesin (OpcA)	17231510 ^N	S. Fig # 3	1 aa ins	370-438
Hypothetical protein DUF111	282899760 ^C	S. Fig # 4	7 aa ins	123-186
Hypothetical protein Npun_R5589	17227740 ^N	S. Fig # 5	3 aa ins	12-58
XRE family transcriptional regulator	17231636 ^N	S. Fig # 6	1 aa ins	124-166
Pentapeptide repeat protein	282897379 ^R	S. Fig # 7	7 aa ins	203-250
Hypothetical protein Npun_R4490	17232408 ^N	S. Fig # 8	1 aa ins	150-191
2-hydroxy-6-oxohepta-2,4-dienoate hydrolase	17227812 ^N	S. Fig # 9	1 aa ins	203-264
Hypothetical protein Aazo_3898	282896270 ^R	S. Fig # 10	2 aa ins	15-49
PilT domain-containing protein	282900401 ^C	S. Fig # 11	1 aa del	126-184
Putative sensor protein	282899023 ^C	S. Fig # 12	3 aa ins	342-400
N-acetylglucosaminyltransferase	282896617 ^R	S. Fig # 13	1 aa ins	127-205
Ribonuclease II	282901218 ^C	S. Fig # 14	14 aa ins	177-237
Arsenite efflux ATP-binding protein	75909548 ^A	S. Fig # 15	2 aa ins	150-188
TPR repeat protein	282901375 ^C	S. Fig # 16	5 aa ins	268-308
hypothetical protein Npun_F390*	17231801 ^N	S. Fig # 17	2 aa ins	17-78

*Indicated CSI is present only in *Nostocales* species/strains

^F GeneBank Identifier accession numbers are associated with *Fischerella* sp. *JSC-11*

^N GeneBank Identifier accession numbers are associated with *Nostoc* sp. *PCC 7120*

^A GeneBank Identifier accession numbers are associated with *Anabaena variabilis* ATCC 29413

^C GeneBank Identifier accession numbers are associated with *Cylindrospermopsis raciborskii*

^R GeneBank Identifier accession numbers are associated with *Raphidopsis brookii* D9

Table 6: Conserved signature indels specific for Subclades of *Nostocales*

Protein Name	GeneBank Identifier	Figure #	Indel Size	Indel Position
(A) CSIs specific for <i>R. Brookii D9</i>, <i>C. raciborskii CS-505</i> and '<i>N. azollae</i>' 0708				
Recombination protein O	282899614 ^C	Fig # 6A	4 aa ins	46-90
Hypothetical protein CRC_01609	282900218 ^C	S. Fig # 18	7 aa ins	148-204
Rod shape-determining protein	298490229 ^N	S. Fig # 19	3 aa ins	131-193
Family 4 glycosyl transferase	298492981 ^N	S. Fig # 20	1 aa ins	262-310
Beta-ketoacyl synthase	282896136 ^R	S. Fig # 21	3 aa ins	195-245
(B) CSIs specific for <i>R. Brookii D9</i> and <i>C. raciborskii CS-505</i>				
O-succinylbenzoyl-CoA synthetase	282901079 ^C	Fig # 6B	7 aa ins	278-321
Prolipoprotein diacylglycerol transferase	282898069 ^R	S. Fig # 22	2 aa ins	16-43
Succinyl-CoA synthetase beta chain	282898786 ^C	S. Fig # 23	2 aa del	97-145
DNA mismatch repair protein	282900091 ^C	S. Fig # 24	4 aa ins	273-325
Methyltransferase type 11	282899194 ^C	S. Fig # 25	1 aa ins	177-212
Prolipoprotein diacylglycerol transferase	282896706 ^R	S. Fig # 26	4 aa ins	170-201
ABC transporter, transmembrane region protein	282899923 ^C	S. Fig # 27	20 aa ins	424-493
Membrane bound O-acyl transferase	282897199 ^R	S. Fig # 28	3 aa ins	46-107
SNF2-related helicase	282898570 ^R	S. Fig # 29	22 aa ins	1007-1072
Hypothetical protein Aazo_0426	282896286 ^R	S. Fig # 30	4 aa ins	21-66
Hypothetical protein Aazo_2731	282898291 ^R	S. Fig # 31	2 aa ins	79-140
Zn-dependent hydrolase	282897472 ^C	S. Fig # 32	7 aa ins	82-140
ABC transporter-like protein	282898549 ^C	S. Fig # 33	1 aa ins	117-155
Small GTP-binding protein domain protein	282900497 ^C	S. Fig # 34	2 aa ins	564-601
Phosphoglucomutase	282899881 ^C	S. Fig # 35	3 aa ins	228-266
Mg chelatase-related protein	282898311 ^R	S. Fig # 36	1 aa ins	12-49
Hypothetical protein Aazo_1936	282901672 ^C	S. Fig # 37	2 aa ins	198-242
Alpha/beta hydrolase fold protein	282900791 ^C	S. Fig # 38	3 aa ins	48-102
Aldo/keto reductase	282896230 ^R	S. Fig # 39	4 aa ins	54-101
Menaquinone biosynthesis protein	282897995 ^R	S. Fig # 40	1 aa ins	434-476
DNA sulfur modification protein	282898767 ^C	S. Fig # 41	4 aa ins	298-350
STHK ^a	282898911 ^C	S. Fig # 42	3 aa ins	348-389
Ribosomal protein L4/L1e	282898356 ^R	S. Fig # 43	1 aa ins	142-194
Metallophosphoesterase	282896194 ^R	S. Fig # 44	4 aa ins	371-416
STHK ^a	282898911 ^C	S. Fig # 45	2 aa ins	490-546
Hypothetical protein Npun_R4078	282895994 ^R	S. Fig # 46	1 aa ins	44-98
ATP-binding protein of ABC transporter	282897025 ^R	S. Fig # 47	4 aa ins	167-218
Leucine dehydrogenase	282897902 ^R	S. Fig # 48	2 aa ins	27-46
Hypothetical protein Npun_F3878	282898976 ^C	S. Fig # 49	3 aa ins	365-422
Zn-dependent hydrolase (beta- lactamase fold protein)	282897472 ^R	S. Fig # 50	9 aa ins	86-140
Phosphoglucomutase	282899881 ^C	S. Fig # 51	3 aa ins	214-283
Hypothetical protein N9414_10373	282897719 ^R	S. Fig # 52	1 aa ins	323-364
1-aminocyclopropane-1-carboxylate deaminase	282901339 ^C	S. Fig # 53	4 aa ins	107-162
Hypothetical protein Aazo_0549	282901667 ^C	S. Fig # 54	7 aa ins	209-257
Deoxyxylulose-5-phosphate synthase	282900458 ^C	S. Fig # 55	4 aa del	199-243
FAD linked oxidase-like protein	282895973 ^R	S. Fig # 56	3 aa ins	271-320
Hydrogenase maturation protein	282899854 ^C	S. Fig # 57	3 aa ins	63-112
Hydrogenase formation protein	282898405 ^R	S. Fig # 58	2 aa ins	1-48
STHK ^a	282898911 ^C	S. Fig # 59	3 aa ins	691-742
Neurotrophic tyrosine kinase receptor type 1	282895647 ^R	S. Fig # 60	1 aa del	217-263

^a Serine/threonine protein kinase and signal transduction histidine kinase^C GeneBank Identifier accession numbers are associated with *Cylindrospermopsis raciborskii*^N GeneBank Identifier accession numbers are associated with '*Nostoc azollae*'^R GeneBank Identifier accession numbers are associated with *Raphidopsis brookii D9*

Table 7: Conserved signature indels specific for Subclades of *Nostocales*

Protein Name	GeneBank Identifier	Figure #	Indel Size	Indel Position
(A) CSIs specific for <i>N. sp. PCC 7120</i>, <i>N. spumigena</i>, <i>N. punctiforme PCC 73102</i> and <i>A. variabilis ATCC 29413</i>				
Glucose-1-phosphate thymidyltransferase	17231983 ^N	Fig # 7A	1 aa ins	100-140
Phospholipid/glycerol acyltransferase	17227730 ^N	S. Fig # 61	1 aa ins	57-115
Hypothetical protein N9414_16197	17227750 ^N	S. Fig # 62	3 aa ins	76-122
(B) CSIs specific for <i>N. sp. PCC 7120</i>, <i>N. punctiforme PCC 73102</i> and <i>A. variabilis ATCC 29413</i>				
Alpha-amylase	17227664 ^N	Fig # 7B	9 aa ins	152-194
Alpha-2-macroglobulin-like protein	17232592 ^N	S. Fig # 63	6 aa del	413-467
Hypothetical protein Npun_R1017	17228462 ^N	S. Fig # 64	2 aa ins	419-476
Ammonium transporter	75909166 ^A	S. Fig # 65	9 aa ins	249-312
(C) CSIs specific for <i>N. sp. PCC 7120</i> and <i>A. variabilis ATCC 29413</i>				
Short-chain dehydrogenase/reductase	17231389 ^N	S. Fig # 66	2 aa ins	135-189
Hypothetical protein Npun_F4300	17227628 ^N	S. Fig # 67	2 aa ins	388-424
Hypothetical protein Npun_R6172	17231457 ^N	S. Fig # 68	2 aa ins	127-171
Mercuric reductase	17232677 ^N	S. Fig # 69	6 aa del	383-421
SNF2-like helicase	17231890 ^N	S. Fig # 70	8 aa ins	884-929

^N GeneBank Identifier accession numbers are associated with *Nostoc sp. PCC 7120*

^A GeneBank Identifier accession numbers are associated with *Anabaena variabilis ATCC 29413*

Table 8: Conserved signature proteins specific for Clades of *Nostocales* and *Stigonematales*

Protein	Function (Amino Acids)	Protein	Function (Amino Acids)
(A) CSPs Specific to <i>Nostocales/Stigonematales</i>			
NP_484526	Hypothetical (64)	NP_488605	Hypothetical (92)
NP_485180	Hypothetical (107)	NP_485189 ^b	Hypothetical (847)
NP_486414	Hypothetical (129)	NP_487215 ^b	Hypothetical (264)
NP_488828	Hypothetical (173)	NP_485191 ^b	Hypothetical (364)
NP_488852	Heterocyst differentiation (215)	NP_489045 ^b	Hypothetical (153)
NP_488867	GDP-D-mannose dehydratase (344)	NP_485774 ^b	Hypothetical (93)
(B) CSPs Specific to <i>Nostocales</i>			
NP_485976	Hypothetical (81)	NP_485974	Hypothetical (77)
NP_485977	Hypothetical (83)	NP_485975	Hypothetical (77)
NP_485973	Hypothetical (77)	NP_485332 ^b	Hypothetical (70)
(C) CSPs Specific to <i>N. PCC 7120</i>, <i>A. variabilis ATCC 2913</i>, <i>N. punctiforme PCC 73102</i> & <i>N. spumigenaCCY9414</i>			
NP_484046	Hypothetical (245)	NP_486339	Hypothetical(74)
NP_487290	Hypothetical (69)	NP_486399	Hypothetical (135)
NP_487429	Hypothetical (75)	NP_489027	Hypothetical (74)
NP_487185	Hypothetical (122)	NP_488822 ^b	Hypothetical (124)
NP_484357	Hypothetical (110)	NP_486953 ^b	Hypothetical (261)
(D) CSPs Specific to <i>N. PCC 7120</i>, <i>A. variabilis ATCC 2913</i> & <i>N. punctiforme PCC 73102</i>			
NP_484767	Hypothetical (69)	NP_487377	Hypothetical (115)
NP_485238	Hypothetical (90)	NP_487521	Hypothetical (185)
NP_485423	Hypothetical (192)	NP_487405 ^b	Hypothetical (332)
NP_488602	Hypothetical (70)	NP_485098 ^b	Hypothetical (181)
NP_489020	Hypothetical (131)	NP_485099 ^b	Hypothetical (151)
NP_489353	Hypothetical (71)	NP_485311 ^b	Hypothetical (131)
NP_489409	Hypothetical (160)		
(E) CSPs Specific to <i>N. PCC 7120</i> & <i>A. variabilis ATCC 2913</i>			
NP_489183	Hypothetical (116)	NP_485855	Hypothetical (212)
NP_484250	Hypothetical (60)	NP_485978	Hypothetical (85)
NP_484860	Hypothetical (66)	NP_487792	Hypothetical (183)
NP_484933	Hypothetical (75)	NP_489180	Hypothetical (103)
NP_485344	Hypothetical (219)	NP_489181	Hypothetical (110)
NP_485765	Hypothetical (136)	NP_489182	Hypothetical (80)

CSPs were identified by BLASTP searches for proteins of the *Nostoc sp. PCC 7120* genome

^b Indicated proteins are also found in 1 to 3 homologous proteins from other bacterial species

Chapter 4: Discussion

4.1 Phylogenetic and Comparative Genomic Analyses

Cyanobacteria are currently classified into several comparable classifications based upon their morphological characteristics (Rippka et al., 1979; Stanier et al., 1978; Castenholz, 2001; Cavalier-Smith, 2002; Euzéby, 1997; Sayers et al., 2010). In phylogenetic trees based upon 16S rRNA, Cyanobacterial members form 14 clusters. However, several of these clusters contain members of a number of different genera and their taxonomic relevance is uncertain (Wilmotte and Herdman, 2001). The availability of genome sequences have made it possible to determine Cyanobacterial phylogeny based upon large datasets of protein sequences. Recent studies have consistently revealed that the sequenced Cyanobacterial species/strains can be divided into at least three main clades at the highest level (Gupta, 2009; Gupta and Mathews, 2010). The emerging relationship closely resembles the four subclass classification system of Cyanobacteria recently proposed by Hoffman et al. (2005). Clade A corresponds to the subclass *Gloeobacteriophycidae*, which contains *Gloeobacter* species and possibly misclassified strains *Synechococcus JA-3-3A* and *S. JA-2-3B*. Clade C, corresponding to the subclass *Synechococcophycidae*, is comprised of the marine unicellular *Synechococcus* and *Prochlorococcus* species/strains. Clade B contains species from four orders, in which *Nostocales* and *Stigonematales* form a monophyletic clade which corresponds to the subclass *Nostocophycidae*. The orders *Oscillatoriales* and *Chroococcales*, which are also part of Clade B have been proposed to form another subclass, *Oscillatoriophycidae*. The distinctness of all three main Clades (A, B and C) is supported by large numbers of CSIs and CSPs discovered in earlier work that are unique to all sequenced species/strains from these clades.

Several molecular signatures have previously been reported which support the three clades found in our trees (Gupta, 2009; Gupta and Mathews, 2010). CSIs found in highly-conserved regions of widely distributed proteins specific to a related group of organisms are extremely rare and specific molecular markers (Gupta, 1998; Rokas and Holland, 2000). CSIs have been found to be associated to cell survival, hence, it is plausible that these CSIs specific to certain Cyanobacterial clades also have an essential role (Gupta and Bhandari, 2011; Singh and Gupta, 2009). Specificity of the identified CSIs in species from *Nostocales/Stigonematales* suggests that the genetic changes responsible originated once in a common ancestor and were subsequently passed on to its descendants, as indicated in Figure 8 (Gupta, 1998; Rokas and Holland, 2000; Gogarten et al., 2002). Previously we have identified CSPs to be, at the time, specific to all or most sequenced *Nostocales* members (*Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413*, *Nostoc punctiforme PCC73102* and *Nodularia spumigena CCY9414*) (Gupta and Mathews, 2010). In this study, BLASTp searches were performed to assess protein specificity, and it was observed that 53 CSPs remain specific to various taxonomic groupings of sequenced *Nostocales/Stigonematales* members (Table 8). Moreover, we extended this approach to identify a large group of >70 novel CSIs in widely-distributed proteins that are uniquely found in all heterocystous Cyanobacteria (*Nostocales/Stigonematales*) or one of their many subclades.

4.2 Implications Inferred by Phylogenetic and Comparative Genomic Analyses Pertaining to all *Nostocales/Stigonematales*

In contradiction with the phylogeny based on morphological characteristics, our phylogenetic studies do not resolve *Stigonematales* and *Nostocales* into two distinct orders. We report 17 CSIs that are specific for all sequenced members of the *Nostocales/Stigonematales*

subclade and one CSI specific for only *Nostocales* members. Interestingly, updated BLASTp searches of previously identified CSPs, we found 12 CSPs that are specific to all *Nostocales/Stigonematales* members and six CSPs that are specific for all *Nostocales* members. Independent from phylogenetic analyses, these ancestrally inherited CSIs and CSPs indicate that the orders are monophyletically derived, which is distinct from other order level taxonomic groupings. The relative abundance of molecular markers specific for both *Nostocales* and *Stigonematales* compared to the few that are specific only to *Nostocales*, suggests an evolutionarily close relationship shared between them and *Fischerella sp. JSC-11*. Molecular markers identified are consistent with concatenated protein trees (Figure 3 & Supplementary Figure 1), which are indicative of a very close monophyletically derived relationship. Protein trees show members from both orders forming a strongly supported clade that is separated from all other Clade B Cyanobacteria by a long branch. *Fischerella sp. JSC-11*, the sole representative of *Stigonematales*, formed an outgroup of the *Nostocales* clade. In the 16S rRNA tree (Figure 4), *Nostocales* members are entrenched within *Stigonematales* members. Within *Nostocales* and *Stigonematales*, members are not monophyletic, as they have representatives from both orders extensively intermixed within each group. Thus, the rRNA tree supports these orders as not different, derived from a common ancestor rather than distinct evolutionary groups divided on the basis of true/false branching. As suggested by both molecular markers and protein trees, the rRNA tree also supports a close relationship shared between sequenced *Nostocales* members relative to *Fischerella sp. JSC-11*. Moreover, both T and Y-branching *Stigonematalean* members do not form monophyletic clades, having members of *Nostocales* as well as both branch types intermixed. Although sequenced members used in this study all belong to a single family, *Nostocaceae* (except *Fischerella*), these species span the majority of

the rRNA tree and encompass parts of several different clades, thus indicating that CSIs and CSPs represent a large portion of *Nostocales/Stigonematales* members as opposed to a single closely related group. Members of the *Nostocaceae* family are not monophyletic, spanning almost the entirety of the tree with members intermixed between several other *Nostocales* families as well as many members of *Stigonematales*. Although contradictory to the rRNA tree created in this study, it is possible that the sequenced *Nostocales/Stigonematales* members all belong to a closely related clade, and the above molecular markers identified are not specific to heterocystous Cyanobacteria but a small group of related species. Similarly, in the event that *Fischerella sp. JSC-11* was misidentified, it is possible that these molecular markers are specific only to a small group of related Heterocystous Cyanobacteria. In this event, *Fischerella sp. JSC-11* may also not represent the *Stigonematales* order and our analyses may not be studying the relationship shared between *Nostocales* and *Stigonematales* but rather a clade within *Nostocales*. Nevertheless, the evidence provided by phylogenetic and comparative genomic studies suggests that organisms belonging to both orders, *Nostocales* and *Stigonematales*, share a close, likely intermixed relationship, forming a monophyletic group that represents the sole heterocyst differentiating Cyanobacteria. However, a more comprehensive database of sequenced members is needed to address the limited availability of sequenced genomes within this group, fully understand the relationship shared between *Nostocales* and *Stigonematales*, as well as the cladal specificity of identified molecular markers.

4.3 Implications Inferred by Phylogenetic and Comparative Genomic Analyses Pertaining to the First Subclade of *Nostocales*

The species/strains of *Nostocales* with genome sequences available, form two distinct subclades and provide information about the relationships/affinities shared among these members. Branching patterns observed in concatenated protein trees suggest a subclade consisting of *Cylindrospermopsis raciborskii* CS-505, *Raphidiopsis brookii* D9 and '*Nostoc azollae*' 0708. This clade is independently supported by five identified CSIs that are uniquely shared by all three species/strains. Additionally, the rRNA tree further indicates these members branch together and are part of a clade which also encompasses *Sphaerospermopsis reniformis* ACC 105. Within this subclade, *Raphidiopsis brookii* D9 and *Cylindrospermopsis raciborskii* CS-505 show a close affinity in both protein tree and the 16s rRNA tree. This close relationship is also independently supported by 40 CSIs that are uniquely shared by these two species/strains. The relative abundance of CSIs identified as specific for this group, further reflects the close affinity shared by these two members. However, the close relationship shared within the rRNA tree by species/strains within this clade suggests that these members may only form a small related collection that is a part of a much bigger clade. CSPs specific for this clade were not identified as only the genome of *Nostoc sp.* 7120 was used for analyses. Thus, only clades which include *Nostoc sp.* 7120 could be studied. Although there is strong support for the suggested relationships shared between sequenced members of this group, these connotations cannot be concluded until a more inclusive database of genomes is composed.

4.3 Implications Inferred by Phylogenetic and Comparative Genomic Analyses Pertaining to the Second Subclade of *Nostocales*

A separate subclade of *Nostocales* members is supported by a number of identified molecular markers and consists of the species; *Nodularia spumigena* CCY9414, *Nostoc punctiforme* PCC73102, *Nostoc sp.* PCC 7120 and *Anabaena variabilis* ATCC2914. Our analyses have identified three CSIs and 10 CSPs that are exclusively present in all members of this group. Within this subclade, there is a closer relationship among the members *Nostoc punctiforme* PCC 73102, *Nostoc sp.* PCC 7120 and *Anabaena variabilis* ATCC 29413 that is supported by four CSIs and 13 CSPs specifically present in these bacteria. Furthermore, a close affinity is observed between *Nostoc sp.* PCC 7120 and *Anabaena variabilis* ATCC 29413 in the rRNA tree and is supported by five CSIs and 12 CSPs uniquely found in these two species/strains. Additionally, these three clades/affinities are all independently supported based on the branching of sequenced species/strains in the concatenated protein trees. Polyphyletic branching of the *Nostoc* genus in both the rRNA and protein trees, infers that these members may have been misidentified. Aforementioned CSIs found only in subclades of *Nostocales* may be useful for diagnostics. Due to the specificity of CSIs located in conserved regions of proteins, degenerate PCR primer based analysis can resolve the presence or absence of clade specific CSIs in unsequenced species, which may assist in the circumscription of these species into their respective bacterial groups (Gao and Gupta, 2005; Griffiths and Gupta, 2002; Griffiths and Gupta, 2006; Gupta, 2005). These CSIs may also be used to identify and characterize unknown species whose genomes are present in metagenomic analyses containing data derived from *Nostocales* and *Stigonematales* members (Gao and Gupta, 2005).

4.4 Conserved Signature Indels Specific to Individual Members of Sequenced *Nostocales*/*Stigonematales*

Our analyses have also identified CSIs that are only present in a single sequenced species/strain of the *Nostocales* or *Stigonematales*. Although, many of these CSIs are likely to be present only in a small spectrum of closely related members, they provide potentially useful molecular markers for the detection of individual or closely related members of this group. The specificities of these CSIs however, need to be confirmed as they may be attributed to artifacts. Sequence based detection of Cyanobacteria is particularly important since morphological based identification is unreliable, as many morphological characteristics are lost upon culturing (Komarek, 1994). While PCR characterization based on degenerate primers specific Cyanobacterial 16S rRNA genes, have proven invaluable, 16S rDNA has limited resolution and its characterization requires the 16s rDNA to be sequenced (Nubel et al., 1997) (Nubel et al., 1997). A similar concept can be used to provide a very precise means for the explicit identification of individual *Nostocales*/*Stigonematales* members by using PCR primers to amplify CSIs that are specific to one species (Ahmod et al., 2011; Naushad and Gupta, 2012).

4.5 Conclusion

As the absence of a consensus phylogeny clearly indicates that the taxonomy of *Nostocales* and *Stigonematales* members is inadequate. Recognizing new sequence based characteristics that offer a reliable means to resolving the evolutionary relationships, as well as provide a useful means to identifying species within these orders, is critical. The CSIs identified in this study will help elucidate the phylogeny of the *Nostocales* and *Stigonematales* orders. In addition, further studies of novel biochemical or physiological characteristics unique to specific

groups of *Nostocales* and *Stigonematales* may provide better insight of the deep branching relationships exhibited. CSIs and CSPs specific to clades found within Cyanobacterial orders have not yet been studied for any other Cyanobacterial groups. The development and implementation of protocols to specifically characterize *Nostocales/Stigonematales* through the amplification of identified molecular markers is of particular interest for future work.

		29	68
(A) Nostocales/ Stigonematales	Fischerella sp. JSC-11	354565800	YPELLQEDHKLRFIEEKL GRYAQ NNAGISEVRIERKADQ
	Nostoc sp. PCC 7120	17231701	-----QY--Q-- --L--
	'Nostoc azollae' 0708	298491404	-----KY-DQ--
	Anabaena variabilis ATCC 29413	75906920	-----QY--Q-- --L--
	Nostoc punctiforme PCC 73102	186684504	-----Y--QY--Q-- --L--
	Nodularia spumigena CCY9414	119511173	-----Y--QY--K-- -KL-
	Raphidiopsis brookii D9	282898351	-----QY--Q-- --N--
	Cylindrospermopsis raciborskii	282899958	-----QY--Q-- --N--
	Moorea producta 3L	332712163	--L-----YQI-QYVDKN-
	Trichodesmium erythraeum 101	113476573	--DT-----I-QYVKAT-
	Arthrospira platensis Paraca	284053309	--Q-----YTI-QY-QKN-
	Microcoleus chthonoplastes 7420	254412256	--V-----YII-QYV-KN-
	Oscillatoria sp. PCC 6506	300865408	-----S-I-QYVQKN-
	Synechococcus sp. JA-2-3B'a(2-13)	86610034	--A-----DQI-TYLTQK-
	Thermosynechococcus elongatus	22297631	-----RI-T--NQQ-
	Acaryochloris marina MBIC11017	158337816	-----F-V--YVKN-
	Synechocystis sp. PCC 6803	16329935	-----I-QY--KT-
	Cyanothece sp. PCC 7822	307154836	-----R-I-SY-DAN-
	Other Cyanobacteria	Synechococcus elongatus PCC 6301	56751880
Synechococcus sp. WH 7805		88807890	--T-----DRI-K--HK-Y
Synechococcus sp. RS9917		87124968	--T-----DRI-K--HK-Y
Synechococcus sp. PCC 7002		170077679	--A-V---Y-I-QH-DAN-
Synechococcus sp. CC9605		78211922	--A-----DRI-K--HK-Y
Synechococcus sp. WH 8102		33866605	--A-----DRI-K--HK-Y
Synechococcus sp. RCC307		148243221	--T-----ERI-K-VNK-Y
Crocospaera watsonii WH 8501		67923589	-----LVV-EY--KN-
Cyanothece sp. ATCC 51142		172038933	-----LAI-QHV-KN-
Cyanothece sp. PCC 8801		218245134	--T-----RQI-QY-TKN-
Microcystis aeruginosa NIES-843		166368478	-----RRI-QYV-KN-
Prochlorococcus marinus MIT 9202		254526486	--I-----F-I-T--K-Y
Prochlorococcus marinus CCMP1986		33862108	--T-----D-I-T--QK-Y
Prochlorococcus marinus AS9601		123969291	--I-----F-I-T--QK-Y
Prochlorococcus marinus NATL1A		124026702	--L-----DRI-T--KK-Y
Gloeobacter violaceus PCC 7421		37523490	--A--A-----I-AY-VK--
Chlorokybus atmophyticus		124112119	--RQ-----NSI--LRT--
Trebouxia aggregata		162134335	--Q-VF--KF-RHYLS-RF
Other Bacteria		Clostridium cellulovorans 743B	302876528
	Sporolactobacillus inulinus CASD	374709486	--A-Y-H--I-I--Y-----
	Bacillus subtilis BEST7613	407957827	-----I-QY--T-

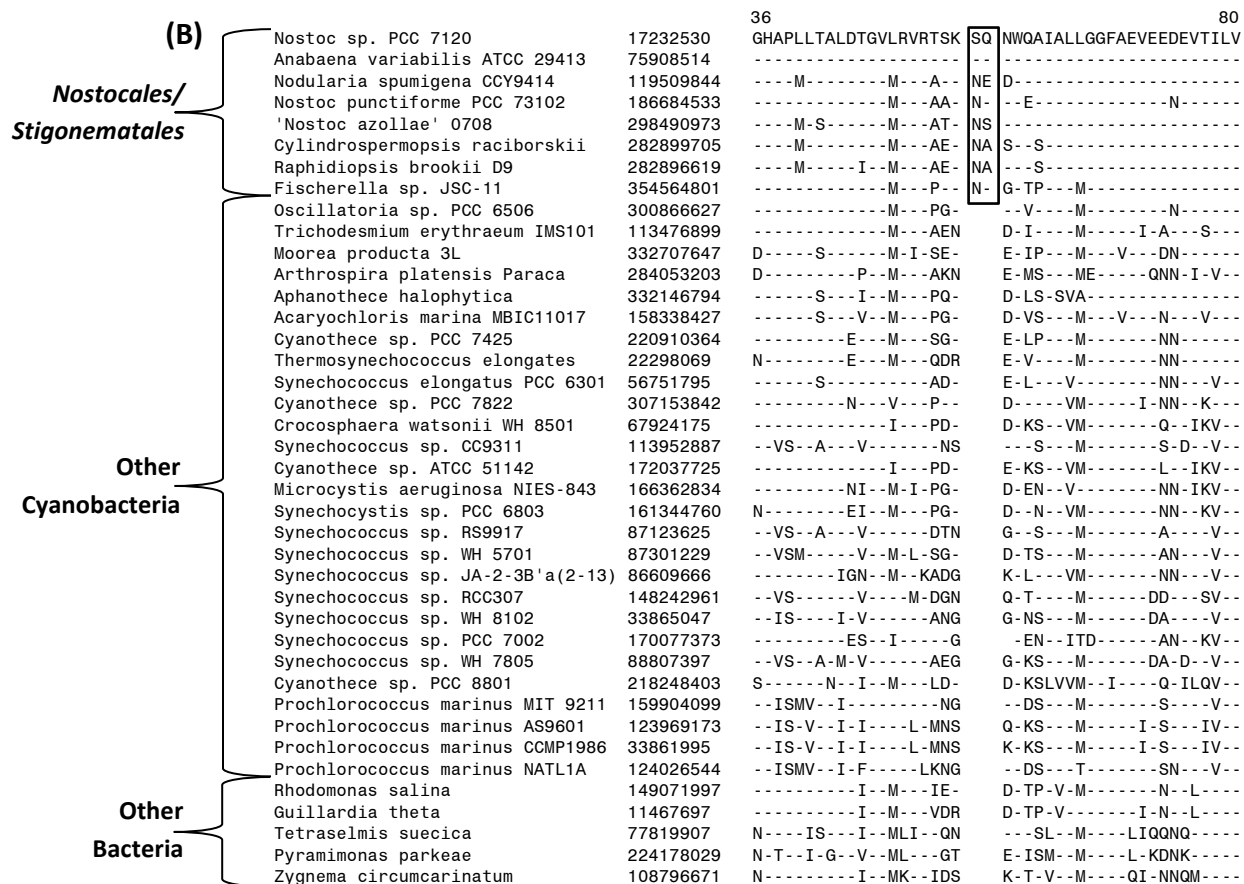


Figure 5: Partial sequence alignments for the proteins (A) 30S ribosomal protein S3, showing a 5 aa insertion and (B) ATP synthase epsilon subunit, showing a 2 aa insertion, both flanked by conserved regions that are specifically present in all sequenced *Nostocales/Stigonematales* members. Dashes in the sequence alignments indicate amino acids which align identically with the first row. Not all sequence information available for species from other Cyanobacterial groups are represented here.

(A)

		46	90	
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i>	282899614	SSLNGRSGIFTVNELLISLGKS NPTQTSN LYKINQAQTLKTYPL	
	<i>Raphidiopsis brookii</i> D9	282898320	-N-----M-----P--- -----G --R-----HA--	
	' <i>Nostoc azollae</i> ' 0708	298491381	---G-----V-Q---Q-R- SQPA-PT -D--T--E-I----S-	
	<i>Anabaena variabilis</i> ATCC 29413	75906952	---G---M-V-----AK-R- -D--T-----I-----	
	<i>Nostoc</i> sp. PCC 7120	17231667	---G---AM-V-----AK-R- -D--T-----I-----	
	<i>Nostoc punctiforme</i> PCC 73102	186684466	---G---M-V-----AK-R- -D--T-----I-----	
	<i>Nodularia spumigena</i> CCY9414	119513475	---G---M-V-----AK-R- -DR-T-----V----D-	
	<i>Fischerella</i> sp. JSC-11	354565593	---G---M-V-----AK-R- -D--T-----V-S----	
	<i>Microcoleus chthonoplastes</i> 7420	254411982	-K-A--M-L-V-Q---YK-R- -D--T--E--ES----	
	<i>Moorea producta</i> 3L	332706215	-K-G---L-V-Q--LAK-R- -D--I--E--TES----	
	<i>Oscillatoria</i> sp. PCC 6506	300866448	-H-G---EL-V-----LAK-R- -D--T--E--IES-S----	
	<i>Arthrospira platensis</i> NIES-39	291565919	-R-G-TT-L-V--QW--AK-R- -D--T--E-IINSHS--	
	<i>Trichodesmium erythraeum</i> IMS101	113477991	-K-G--GEL-V-Q---K-R- -DR----E-ITS-T--	
	<i>Arthrospira platensis</i> Paraca	284052146	-R-G-TT-L-V--QW--AK-R- -D--T--E-IINSHS--	
	<i>Lyngbya</i> sp. PCC 8106	119485167	-K-G--EL-I-Q--LYK-R- -DR-T--DLI-S-G--	
	<i>Cyanothece</i> sp. PCC 7425	220906414	---G---L-V-----AP-R- -H-LS--E--TS-AR-	
	<i>Acaryochloris marina</i> MBIC11017	158337750	-K-G--AL-V-D-MLVQ-R- -D--A--E--ES----	
	<i>Cyanothece</i> sp. CCY0110	126658910	-K-R--CEL-VI----AK-R- -D--T--E--ES----	
	Other Cyanobacteria	<i>Crocospaera watsonii</i> WH 0003	357265200	-R-R--EL-VI----AK-R- -D--T--N--ES----
		<i>Crocospaera watsonii</i> WH 8501	67925437	-R-R--EL-VI----AK-R- -D--T--N--ES----
<i>Cyanothece</i> sp. PCC 7424		218439778	-R-R--EL-V-Q---VK-R- -D-LI--E--ES----	
<i>Cyanothece</i> sp. PCC 7822		307152071	-R-R--AEL-V-Q---VK-R- -E--I--E--ES----	
<i>Microcystis aeruginosa</i> NIES-843		166365593	---R--EL-V-N---VA--- -D-VV--E--IES--K-	
<i>Cyanothece</i> sp. PCC 8802		257060399	-R-R--EL-VI-D---AK-R- -D--T--D--ES--K-	
<i>Cyanothece</i> sp. PCC 8801		218248277	-R-R--EL-VI-D---AK-R- -D--T--D--ES--K-	
<i>Cyanothece</i> sp. ATCC 51142		172039459	-K-R--CEL-VI----AK-R- -D--T--E--ES----	
<i>cyano bacterium</i> UCYN-A		284928635	-P-S--EL-M-QF--IK-H- -D--A--E--QF----	
<i>Synechococcus</i> sp. PCC 7335		254421751	-K-A---L-V---FVK--- -D-LV--E--VRSF---	
<i>Synechococcus</i> sp. RCC307		148242899	---AAA-AL-LIKAQVAR-R- -DRLR--VIRS-SR-	
<i>Synechocystis</i> sp. PCC 6803		161344755	---R--TEL-V-QF-LFA--- -DR----E-QYS----	
<i>Synechococcus</i> sp. PCC 7002		170077380	-A-R--TELLV---FVLTR-R- -DR-Q--DM-TS-S--	
<i>Thermosynechococcus elongatus</i>		22299777	-R-G--VDL-V-D-F--P-RN -DR-L--E-VA-Q--	
<i>Synechococcus elongatus</i> PCC 6301		56750569	-K-G--TAL-V-GDC--AC--- -DRLQ-VE--TSHTA-	
<i>Gloeobacter violaceus</i> PCC 7421		37519653	ARIG--TEQ-V--D-QLYR-R- -DQLT--ES-R-F---	
<i>Prochlorococcus marinus</i> MIT 9211		159902937	---SATTPL-FL--Q-GGKTG -RRV--IKI--SFSK-	
<i>Prochlorococcus marinus</i> NATL1A		124025163	-R--ATSP-NLLD-H-VGK-D -K-VT- IKII-S-GN-	
Other Bacteria		<i>Bacillus subtilis</i> BEST7613	407961348	---R--TEL-V-QF-LFA--- -DR----ETQYS----
		<i>Bifidobacterium dentium</i> 27678	171742448	-RFG--LEP-MRVDV-VAE-R- -DV-S--ESISA-A-P
	<i>Subdoligranulum variabile</i> 15176	261366883	NK-FSAC-L-CYS-FSLTS-R- H-FVES--VK-VFH-V	

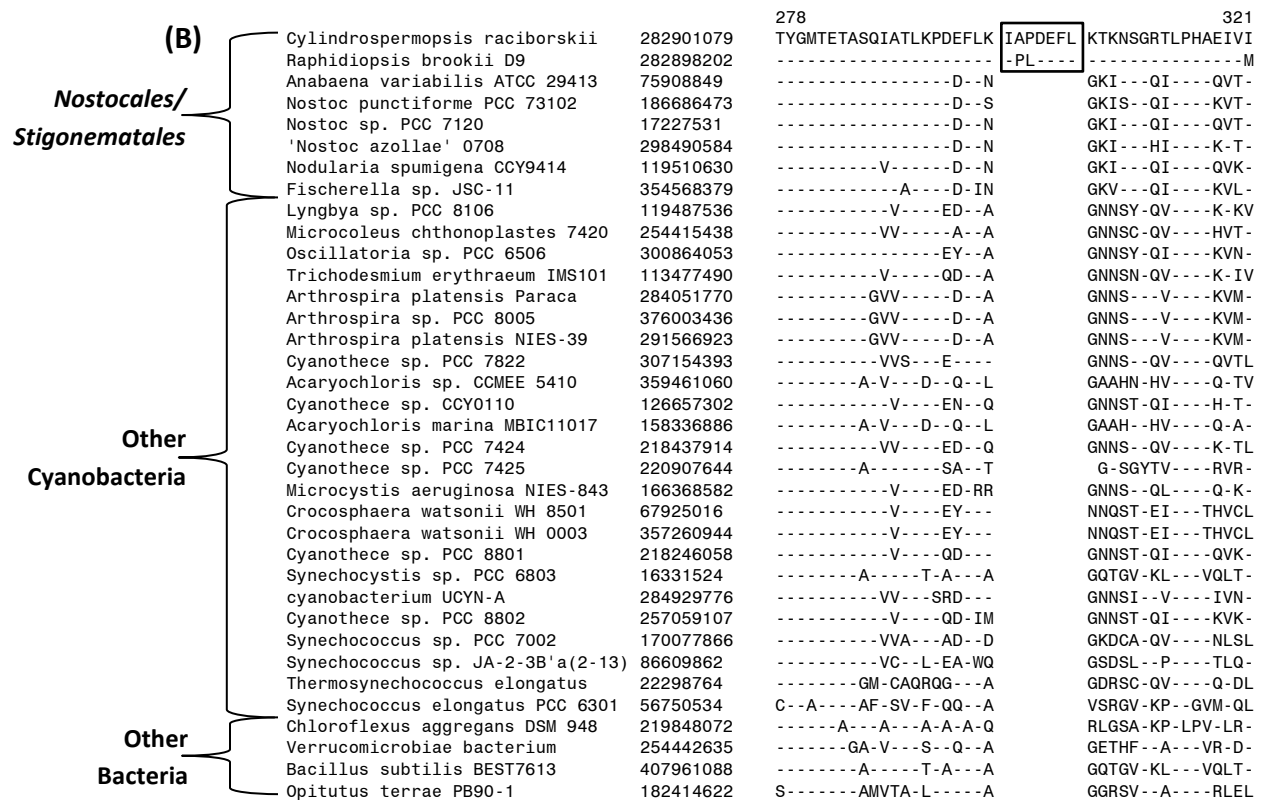


Figure 6: Partial sequence alignments for the proteins (A) Recombination protein O, showing a 7 aa insertion and (B) O-succinylbenzoyl-CoA synthetase, showing a 7 aa insertion, both flanked by conserved regions that are specifically present in all members which represent a subclade/close affinity within *Nostocales*. Dashes in the sequence alignments indicate amino acids which align identically with the first row. Not all sequence information available for species from other Cyanobacterial groups are represented here.

(A)

			100		140
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17231983	SPFVMYLGDNLIQQ	G	DLSNFLQKFIQEQPDALILLREVVNP
	Anabaena variabilis ATCC 29413	75909562	-----	-	-----Q-----I--
	Nodularia spumigena CCY9414	119513384	---I-----	---	---Y---E---Q---E-----
	Nostoc punctiforme PCC 73102	186684648	-----L---	-	E-RY---Q-S-Q-----S-A-
	Microcystis aeruginosa NIES-843	166362761	---I-----	-ED	---P--GS-QKQSL-----K-S--
	Cyanothece sp. PCC 7425	220909660	---I-----	-ES	E-GE-VDQ-K--RL-----KA-P--
	Microcystis sp. T1-4	390441716	---I-----	-ED	---P--DS-QKQSL-----K-S--
	Synechococcus sp. PCC 7335	254424288	---L-----	-D	S-MG-V-E-L-R-L--M---KH-P--
	Cyanothece sp. PCC 8801	218245361	A--I-----	-D	--KP--S--ESQDL-G----S-S--
	Cyanothece sp. PCC 8802	257058396	A--I-----	-D	--KP--S--ESQDL-G----S-S--
Other Cyanobacteria	Cyanothece sp. CCY0110	126654757	D--I-----	-D	--NL--D--KNKHL-G----S-P--
	Crocospaera watsonii WH 0003	357260613	---I-----	-D	-INV--D--KNQHL-E----S-S--
	Cyanothece sp. PCC 7424	218438468	---I-----	-D	-VNP-VET-KTQHL-----T-S--
	Cyanothece sp. ATCC 51142	172037462	D--I-----	-D	--NL--D--KNKHL-G----S-P--
	Synechocystis sp. PCC 6803	16331420	-----D		H-EQ--AH-QAKSL-S----R-S--
	Synechococcus sp. PCC 7335	254424288	---L-----	-D	S-MG-V-E-L-R-L--M---KH-P--
	Cyanothece sp. PCC 7822	307153970	---I-----	-D	EVIP-VET-SKQHL-----T-S--
	Lyngbya sp. PCC 8106	119485670	---I-----	-N	Q--P-VEQ-KSK-L-----S-E--
	Oscillatoria sp. PCC 6506	300866612	---I-----	-S	Q-DP---T-KNNNL-----P-T--
	Trichodesmium erythraeum IMS101	113475230	---I-----	-I-ES	Q-G--DT-KKK-L-----K-P--
Other Bacteria	Arthrospira sp. PCC 8005	376004467	---I-----	-N	P--P--EQ-T-K-L-G----P-E--
	Arthrospira platensis Paraca	284051014	---I-----	-N	P--P--EQ-K-K-L-S----P-E--
	Arthrospira platensis NIES-39	291568778	---I-----	-N	P--P--EQ-K-K-L-S----P-E--
	Arthrospira maxima CS-328	209526053	---I-----	-N	P--P--EQ-K-K-L-G----P-E--
	Microcoleus vaginatus FGP-2	334117921	---I-----	-EN	Q-DP--EI-NKQKL-----P-P--
	Moorea producta 3L	332705311	-----V-S		E-NL-VEN-QNKNL---T---D--
	Microcoleus chthonoplastes 7420	254416323	---L-----	-V-S	---L-I-R-QERNL---T---CT-P--
	Gloeobacter violaceus PCC 7421	37522806	-S-A-----	-S	G-AAPIER-EASR-E----KR-S--
	Streptomyces peucetius ATCC 27952	238560596	DD-----	-F-VG	GI-GLVDE-RA-R--Q---TK-P--
	Desulfotomaculum ruminis DSM 2154	334339704	Q--L-F-----	-V-G	GVVEVVRE-QR-K-----K--PD-
Candidatus Magnetobacterium	297374643	DD-----	-LS-	GVREAIER-NKSR-----F-K--ED-	
Conexibacter woesei DSM 14684	284047223	A-----	-L-G	GIQELVDA-RSSE-----TP-ED-	

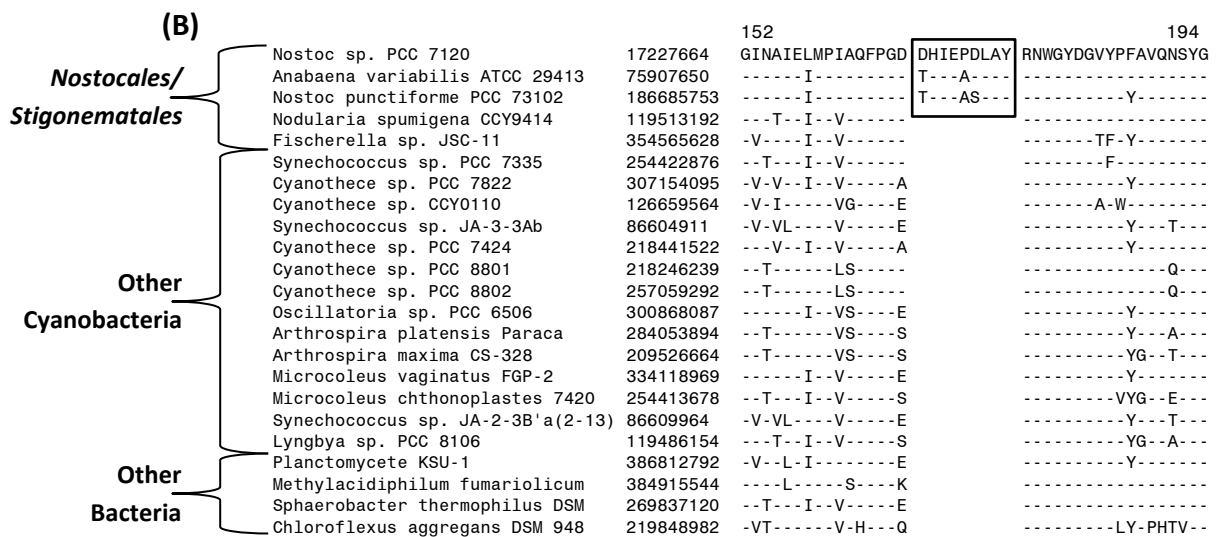


Figure 7: Partial sequence alignments for the proteins **(A)** Glucose-1-phosphate thymidyltransferase, showing a 1 aa insertion and **(B)** Alpha-amylase, showing a 9 aa insertion, both flanked by conserved regions that are specifically present in all members belonging to two consecutive subclades within *Nostocales*. Dashes in the sequence alignments indicate amino acids which align identically with the first row. Not all sequence information available for species from other Cyanobacterial groups are represented here.

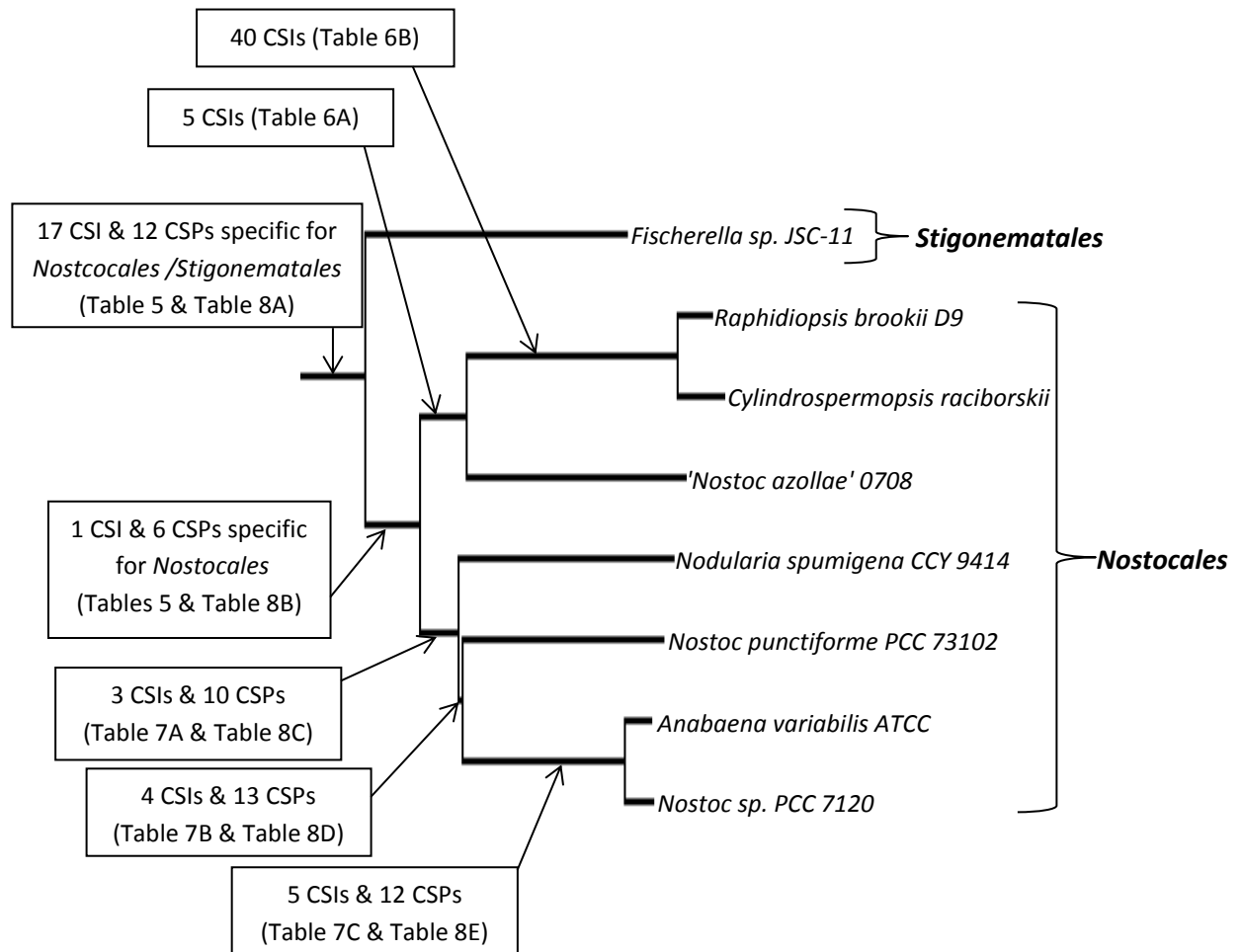


Figure 8: A summary diagram showing the distribution of identified CSIs and CSPs specific to various clades within *Nostocales*/*Stigonematales*. The identified clades supported by both phylogenetic studies as well as various molecular markers are indicated. CSIs specific for only one sequenced species of *Nostocales* are not presented.

4.7 Future work

As this project focuses on sequenced genomes, the number of available genomes dictates the amount of information that can be analysed. This assignment has focused primarily on the orders *Nostocales* and *Stigonematales*, therefore the most important genomes are those which pertain to these groups. When this project began, there were approximately 48 cyanobacterial genomes available, 7 of which belonging to *Nostocales* and 1 belonging to *Stigonematales*. Throughout this study, extensive work focused on sequencing members of this phylum. However none of the genomes sequenced had a large effect on our research as the members sequenced (other than 1) did not pertain to the group our work focused on. As the project reached its final stage before submission, Shih, et al. (2013) published 54 cyanobacteria genomes, 16 of which belonging to *Nostocales/Stigonematales*, in a single manuscript and made the data publicly available through the Department of Energy-Joint Genome Institute (DOE-JGI). Currently there are more than 135 cyanobacterial genomes available, far more than the 48 this project began studying. As detailed earlier, this study portrays an incomplete picture limited by the small number of genomes that were available at the time of analysis. The relationship observed between *Nostocales/Stigonematales* cannot be conclusive with only one *Stigonematales* member available to represent the order. Therefore, there is not only far too much information to ignore, the available data gives us the opportunity to overcome the limitations this project has faced and we now have the possibility to draw more conclusive observations. Thus, we have decided not to publish this manuscript and instead use the data we have collected for downstream work.

With the emergence of several diverse cyanobacterial genomes, classification of this phylum will likely soon change. The broadened coverage has brought some clarity to what is still a poorly classified group of bacteria. The published ML tree by Shih et al. (2013), comprised of 31 concatenated proteins and encompassing 126 cyanobacteria, shows several clades consistent to our tree and several clades which have not yet been seen until now. Clade A and C from our work remains almost identical to the published tree. Clade B also remains distinct from clade A and C, and still clearly forms at least 3 distinct groups. We also see what looks to be a clade(s) branching between our deepest branching group (clade A) and clade C. However, as there are too few species sequenced which belong to this group, more data is needed to clarify their branching resolution. Members belonging to both *Chroococcales* and *Oscillatoriales* encompass the entirety of the tree except in within the heterocystous group (*Nostocales/Stigonematales*). Therefore, it is unclear which phylogenetic group Cavalier-Smith intended to represent *Chroococcales* and *Oscillatoriales* orders. As *Chroococcales*, *Oscillatoriales* and *Pleurocapsales* all belong to the same class, it is likely they should represent a monophyletic group of bacteria. Therefore the only option is that these orders encompass the *Chroococcales*, *Oscillatoriales* and *Pleurocapsales* species of Clade B. Thus, Clade C, the clearest and most distinct clade within this phylum remains completely unclassified. In addition, we see that *Nostocales* and *Stigonematales* are monophyletic, sharing a common ancestor but distinct (other than *Calothrix* PCC 7103 and 6303) from one another, so far contradicting the 16S rRNA tree we created. Although not supported by strong bootstrap, we see two *Calothrix* species (*Nostocales*) branching more closely to *Stigonematales* than *Nostocales*. Although *Nostocales* and *Stigonematales* members on this tree are not intermixed, with still a small number of sequenced *Stigonematales* it is not conclusive that these orders are distinct. Our Future work will initially re-examine the specificity

of CSIs identified in this project to see if they remain specific to all heterocystous cyanobacteria or a group within. Moreover, we intend to re-examine the specificity of all previously identified CSIs which were observed to be specific for major clades of cyanobacteria in hopes of giving a clearer picture of groups within this phylum.

References

- Ahmod, N.Z., Gupta, R.S., and Shah, H.N. (2011). Identification of a *Bacillus anthracis* specific indel in the *yeaC* gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *Journal of microbiological methods* 87, 278-285.
- Anagnostidis, K. and Komarek, J. (1985). Modern approach to the classification system of cyanophytes. 1-Introduction. *Algological Studies/Archiv fur Hydrobiologie* 38, 291-302.
- Backer, L.C. (2002). Cyanobacterial harmful algal blooms (CyanoHABs): Developing a public health response. *Lake and Reservoir Management* 18, 20-31.
- Bekker, A., Holland, H.D., Wang, P.L., Rumble, D.I.I.I., Stein, H.J., Hannah, J.L., Coetzee, L.L., and Beukes, N.J. (2004). Dating the rise of atmospheric oxygen. *Nature* 427, 117-120.
- Boyer, G. (2004). Monitoring strategies for harmful algal blooms in the lower great lakes; Lakes Erie, Ontario and Champlain, USA. 11th Internat. Conference on Harmful Algae. Capetown, South Africa.
- Boyer, G.L. (2008). Cyanobacterial toxins in New York and the lower Great Lakes ecosystems. In *Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs*, Springer, pp. 153-165.
- Brittain, S., Mohamed, Z.A., Wang, J., Lehmann, V.K.B., Carmichael, W.W., and Rinehart, K.L. (2000). Isolation and characterization of microcystins from a River Nile strain of *Oscillatoria tenuis* Agardh ex Gomont. *Toxicon* 38, 1759-1771.
- Castenholz, R.W. (2001). Phylum BX. Cyanobacteria Oxygenic photosynthetic bacteria. In *Bergey's Manual of Systematic Bacteriology*. (New York: Springer), pp. 474-487.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* 17, 540-552.
- Cavalier-Smith, T. (2002). The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *International Journal of Systematic and Evolutionary Microbiology* 52, 7-76.
- Chorus, I., Falconer, I.R., Salas, H.J., and Bartram, J. (2000). Health risks caused by freshwater cyanobacteria in recreational waters. *Journal of Toxicology and Environmental Health Part B: Critical Reviews* 3, 323-347.
- Christoffersen, K. (1996). Ecological implications of cyanobacterial toxins in aquatic food webs. *Phycologia* 35, 42-50.

- Chung,S.Y., Vercellotti,J.R., Johnsen,P.B., and Klesius,P.H. (1991). Development of an enzyme-linked immunosorbent assay for geosmin. *Journal of Agricultural and Food Chemistry* 39, 764-769.
- Ciccarelli,F.D., Doerks,T., Von Mering,C., Creevey,C.J., Snel,B., and Bork,P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283-1287.
- Codd,G.A. (2000). Cyanobacterial toxins, the perception of water quality, and the prioritisation of eutrophication control. *Ecological engineering* 16, 51-60.
- Codd,G.A., Morrison,L.F., and Metcalf,J.S. (2005). Cyanobacterial toxins: risk management for health protection. *Toxicology and applied pharmacology* 203, 264-272.
- Conroy,J.D., Quinlan,E.L., Kane,D.D., and Culver,D.A. (2007). *Cylindrospermopsis* in Lake Erie: testing its association with other cyanobacterial genera and major limnological parameters. *Journal of Great Lakes Research* 33, 519-535.
- Desikachary,T.V. (1959). *Cyanophyta*, Indian Council of Agricultural Research, monographs on Algae. New Delhi, India.
- Drouet,F. (1981). Summary of the classification of blue-green algae. *Nova Hedwigia. Beiheft* 66, 135-209.
- Dufresne,A., Salanoubat,M., Partensky,F., Artiguenave,F., Axmann,I.M., Barbe,V., Duprat,S., Galperin,M.Y., Koonin,E.V., and Le Gall,F. (2003). Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences* 100, 10020-10025.
- Dyble,J., Fahnenstiel,G.L., Litaker,R.W., Millie,D.F., and Tester,P.A. (2008). Microcystin concentrations and genetic diversity of *Microcystis* in the lower Great Lakes. *Environmental toxicology* 23, 507-516.
- Embley,T.M. and Stackebrandt,E. (1997). Species in practice: exploring uncultured prokaryote diversity in natural samples. *Systematics Association* 54, 61-82.
- Euzeby,J.P. (1997). List of bacterial names with standing in nomenclature: A folder available on the Internet. *International Journal of Systematic Bacteriology* 47, 590-592.
- Falconer,I.R. (1991). Tumor promotion and liver injury caused by oral consumption of cyanobacteria. *Environmental toxicology and water quality* 6, 177-184.
- Falconer,I.R. (2001). Toxic cyanobacterial bloom problems in Australian waters: risks and impacts on human health. *Phycologia* 40, 228-233.
- Florenzano,G., Balloni,W., and Materassi,R. (1986). Nomenclature of *Prochloron didemni* (Lewin 1977) sp. nov., nom. rev., *Prochloron* (Lewin 1976) gen. nov., nom. rev., *Prochloraceae* fam.

nov., Prochlorales ord. nov., nom. rev. in the class Photobacteria Gibbons and Murray 1978. *International Journal of Systematic Bacteriology* 36, 351-353.

Fujisawa,T., Narikawa,R., Okamoto,S., Ehira,S., Yoshimura,H., Suzuki,I., Masuda,T., Mochimaru,M., Takaichi,S., and Awai,K. (2010). Genomic structure of an economically important cyanobacterium, *Arthrospira (Spirulina) platensis* NIES-39. *DNA research* 17, 85-103.

Gao,B. and Gupta,R.S. (2005). Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *International Journal of Systematic and Evolutionary Microbiology* 55, 2401-2412.

Gao,B., Mohan,R., and Gupta,R.S. (2009). Phylogenomics and protein signatures elucidating the evolutionary relationships among the Gammaproteobacteria. *International Journal of Systematic and Evolutionary Microbiology* 59, 234-247.

Garcia-Pichel,F., Lopez-Cortes,A., and Nubel,U. (2001). Phylogenetic and morphological diversity of cyanobacteria in soil desert crusts from the Colorado Plateau. *Applied and Environmental Microbiology* 67, 1902-1910.

Geitler,L. (1925). Synoptische Darstellung der Cyanophyceen in morphologischer und systematischer Hinsicht. *Beih. Bot. Centralbl* 2, 163-324.

Geitler,L. (1932). Cyanophyceae. Rabenhorst's Kryptogamen-Flora von Deutschland, Österreich und der Schweiz, Reprint 1985 14.

Giovannoni,S.J., Turner,S., Olsen,G.J., Barns,S., Lane,D.J., and Pace,N.R. (1988). Evolutionary relationships among cyanobacteria and green chloroplasts. *Journal of bacteriology* 170, 3584-3592.

Gleason,F.K. and Case,D.E. (1986). Activity of the natural algicide, cyanobacterin, on angiosperms. *Plant physiology* 80, 834-837.

Gogarten,J.P., Doolittle,W.F., and Lawrence,J.G. (2002). Prokaryotic evolution in light of gene transfer. *Molecular biology and evolution* 19, 2226-2238.

Golubic,S. (1979). Cyanobacteria (blue-green algae) under the bacteriological code? An ecological objection. *Taxon* 28, 387-389.

Griffiths,E. and Gupta,R.S. (2002). Protein signatures distinctive of chlamydial species: horizontal transfers of cell wall biosynthesis genes *glmU* from archaea to chlamydiae and *murA* between chlamydiae and *Streptomyces*. *Microbiology* 148, 2541-2549.

Griffiths,E. and Gupta,R.S. (2006). Molecular signatures in protein sequences that are characteristics of the phylum Aquificae. *International Journal of Systematic and Evolutionary Microbiology* 56, 99-107.

- Gugger, M.F. and Hoffmann, L. (2004). Polyphyly of true branching cyanobacteria (Stigonematales). *International Journal of Systematic and Evolutionary Microbiology* 54, 349-357.
- Gupta, R.S. (1998). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiology and Molecular Biology Reviews* 62, 1435-1491.
- Gupta, R.S. (2000). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS microbiology reviews* 24, 367-402.
- Gupta, R.S. (2001). The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. *International Microbiology* 4, 187-202.
- Gupta, R.S. (2005). Protein signatures distinctive of alpha proteobacteria and its subgroups and a model for alpha-proteobacterial evolution. *Critical reviews in microbiology* 31, 101-135.
- Gupta, R.S. (2009). Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. *International Journal of Systematic and Evolutionary Microbiology* 59, 2510-2526.
- Gupta, R.S. and Bhandari, V. (2011). Phylogeny and molecular signatures for the phylum Thermotogae and its subgroups. *Antonie Van Leeuwenhoek* 100, 1-34.
- Gupta, R.S. and Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theoretical population biology* 61, 423-434.
- Gupta, R.S. and Mathews, D.W. (2010). Signature proteins for the major clades of Cyanobacteria. *Bmc Evolutionary Biology* 10, 24.
- Gupta, R.S. and Mok, A. (2007). Phylogenomics and signature proteins for the alpha Proteobacteria and its main groups. *BMC microbiology* 7, 106.
- Haider, S., Naithani, V., Viswanathan, P.N., and Kakkar, P. (2003). Cyanobacterial toxins: a growing environmental concern. *Chemosphere* 52, 1-21.
- Harris, J.K., Kelley, S.T., Spiegelman, G.B., and Pace, N.R. (2003). The genetic core of the universal ancestor. *Genome research* 13, 407-412.
- Hawkins, P.R., Chandrasena, N.R., Jones, G.J., Humpage, A.R., and Falconer, I.R. (1997). Isolation and toxicity of *Cylindrospermopsis raciborskii* from an ornamental lake. *Toxicon* 35, 341-346.
- Haystead, A., Robinson, R., and Stewart, W.D.P. (1970). Nitrogenase activity in extracts of heterocystous and non-heterocystous blue-green algae. *Archiv fur Mikrobiologie* 74, 235-243.

- Hisbergues, M., Christiansen, G., Rouhiainen, L., Sivonen, K., and Borner, T. (2003). PCR-based identification of microcystin-producing genotypes of different cyanobacterial genera. *Archives of Microbiology* *180*, 402-410.
- Ho, L., Sawade, E., and Newcombe, G. (2012). Biological treatment options for cyanobacteria metabolite removal-A review. *Water research* *46*, 1536-1548.
- Hoefel, D., Ho, L., Aunkofer, W., Monis, P.T., Keegan, A., Newcombe, G., and Saint, C.P. (2006). Cooperative biodegradation of geosmin by a consortium comprising three gram-negative bacteria isolated from the biofilm of a sand filter column. *Letters in applied microbiology* *43*, 417-423.
- Hoffmann, L. (2005). Nomenclature of Cyanophyta/Cyanobacteria: roundtable on the unification of the nomenclature under the Botanical and Bacteriological Codes. *Algological Studies* *117*, 13-29.
- Hoffmann, L., Komarek, J., and Kastovsky, J. (2005). System of cyanoprokaryotes (cyanobacteria) state in 2004. *Algological Studies* *117*, 95-115.
- Honda, D., Yokota, A., and Sugiyama, J. (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. *Journal of Molecular Evolution* *48*, 723-739.
- Huang, Y.F., Chen, C.T., and Kao, C.H. (1993). Salicylic acid inhibits the biosynthesis of ethylene in detached rice leaves. *Plant growth regulation* *12*, 79-82.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS* *8*, 275-282.
- Juttner, F. (1995). Physiology and biochemistry of odorous compounds from freshwater cyanobacteria and algae. *Water Science and Technology* *31*, 69-78.
- Kaebernick, M. and Neilan, B.A. (2001). Ecological and molecular investigations of cyanotoxin production. *FEMS microbiology ecology* *35*, 1-9.
- Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., Nakamura, Y., Kasai, F., Watanabe, A., and Kawashima, K. (2007). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA research* *14*, 247-256.
- Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., and Kimura, T. (2001). Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA research* *8*, 205-213.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirose, M., Sugiura, M., and Sasamoto, S. (1996). Sequence analysis of the genome of the unicellular

cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA research* 3, 109-136.

Kling,H.J., Smarda,J., Komarek,J., Acerman,J., BRUUN,K., Watson ,S.B., and Chen ,F. (2012). A new red colonial *Pseudanabaena* (Cyanoprokaryota, Oscillatoriales) from North American large lakes. *Fottea, Olomouc: Czech Phycological Society* 12, 327-339.

Komarek,J. (1994). Current trends and species delimitation in the cyanoprokaryote taxonomy.- Introductory lecture of the 12th IAC Symposium, Sharp, August 1992. *Algological Studies/Archiv fur Hydrobiologie, Supplement Volumes* 75, 11-29.

Komarek,J. (2002). Problems in cyanobacterial taxonomy: implication for most common toxin producing species. *Rapporti Istisa* 6-43.

Komarek,J. and Anagnostidis,K. (1986). Modern approach to the classification system of cyanophytes. 2. Chroococcales. *Arch. Hydrobiol. Suppl* 73, 157-226.

Kumar,K., Mella-Herrera,R.A., and Golden,J.W. (2010). Cyanobacterial heterocysts. *Cold Spring Harbor perspectives in biology* 2.

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

Leao,P.N., Engene,N., Antunes,A., Gerwick,W.H., and Vasconcelos,V. (2012). The chemical ecology of cyanobacteria. *Natural Product Reports* 29, 372-391.

Makarewicz,J.C., Lewis,T.W., and Bertram,P. (1999). Phytoplankton Composition and Biomass in the Offshore Waters of Lake Erie: Pre-and Post *Dreissena* Introduction (1983ΓÇô1993). *Journal of Great Lakes Research* 25, 135-148.

Morales-Valle,H., Silva,L.C., Paterson,R.R.M., Oliveira,J.M., Venóncio,A., and Lima,N. (2010). Microextraction and Gas Chromatography/Mass Spectrometry for improved analysis of geosmin and other fungal "off" volatiles in grape juice. *Journal of microbiological methods* 83, 48-52.

Mur,R., Skulberg,O.M., and Utkilen,H. (1999). Cyanobacteria in the Environment. In *Toxic Cyanobacteria in Water*. (London, U.K: Spon Press), pp. 15-40.

Muyzer,G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystems. *Current opinion in microbiology* 2, 317-322.

Nakamura,S., Nakamura,N., and Ito,S. (2001). Determination of 2-methylisoborneol and geosmin in water by gas chromatography/mass spectrometry using stir bar sorptive extraction. *Journal of separation science* 24, 674-677.

- Nakamura,Y., Kaneko,T., Sato,S., Ikeuchi,M., Katoh,H., Sasamoto,S., Watanabe,A., Iriguchi,M., Kawashima,K., and Kimura,T. (2002). Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA research* *9*, 123-130.
- Nakamura,Y., Kaneko,T., Sato,S., Mimuro,M., Miyashita,H., Tsuchiya,T., Sasamoto,S., Watanabe,A., Kawashima,K., and Kishida,Y. (2003). Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA research* *10*, 137-145.
- Naushad,H.S. and Gupta,R.S. (2012). Molecular signatures (conserved indels) in protein sequences that are specific for the order Pasteurellales and distinguish two of its main clades. *Antonie Van Leeuwenhoek* *101*, 105-124.
- Nei,M. and Kumar,S. (2000). *Molecular evolution and phylogenetics*. Oxford University Press, USA).
- Nubel,U., Engelen,B., Felske,A., Snaidr,J., Wieshuber,A., Amann,R.I., Ludwig,W., and Backhaus,H. (1996). Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *Journal of bacteriology* *178*, 5636-5643.
- Nubel,U., Garcia-Pichel,F., and Muyzer,G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology* *63*, 3327-3332.
- Ohtani,I., Moore,R.E., and Runnegar,M.T. (1992). Cyliospermopsin: a potent hepatotoxin from the blue-green alga *Cyliospermopsis raciborskii*. *Journal of the American Chemical Society* *114*, 7941-7942.
- Padisak,J. (1997). *Cyliospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. *Archiv für Hydrobiologie Supplementband Monographische Beiträge* *107*, 563-593.
- Palenik,B., Brahmsha,B., Larimer,F.W., Land,M., Hauser,L., Chain,P., Lamerdin,J., Regala,W., Allen,E.E., and McCarren,J. (2003). The genome of a motile marine *Synechococcus*. *Nature* *424*, 1037-1042.
- Palenik,B., Ren,Q., Dupont,C.L., Myers,G.S., Heidelberg,J.F., Badger,J.H., Madupu,R., Nelson,W.C., Brinkac,L.M., and Dodson,R.J. (2006). Genome sequence of *Synechococcus* CC9311: insights into adaptation to a coastal environment. *Proceedings of the National Academy of Sciences* *103*, 13555-13559.
- Pandey,K.D., Shukla,S.P., Shukla,P.N., Giri,D.D., Singh,J.S., Singh,P., and Kashyap,A.K. (2004). Cyanobacteria in Antarctica: Ecology, physiology and cold adaptation. *Cellular and Molecular Biology-Paris-Wegman* *50*, 575-584.
- Pelaez,M., Antoniou,M.G., He,X., Dionysiou,D.D., Armah,A., Tsimeli,K., Triantis,T., Hiskia,A., Kaloudis,T., and Williams,C. (2010). Sources and occurrence of cyanotoxins worldwide. In *Xenobiotics in the Urban Water Cycle*, Springer), pp. 101-127.

- Pitois, S., Jackson, M.H., and Wood, B.J.B. (2000). Problems associated with the presence of cyanobacteria in recreational and drinking waters. *International Journal of Environmental Health Research* 10, 203-218.
- Pomati, F., Kellmann, R., Cavalieri, R., Burns, B.P., and Neilan, B.A. (2006). Comparative gene expression of PSP-toxin producing and non-toxic *Anabaena circinalis* strains. *Environment international* 32, 743-748.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* 41, D590-D596.
- Ran, L., Larsson, J., Vigil-Stenman, T., Nylander, J.A., Ininbergs, K., Zheng, W.W., Lapidus, A., Lowry, S., Haselkorn, R., and Bergman, B. (2010). Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *Plos One* 5, e11486.
- Rantala, A., Fewer, D.P., Hisbergues, M., Rouhiainen, L., Vaitomaa, J., Borner, T., and Sivonen, K. (2004). Phylogenetic evidence for the early evolution of microcystin synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 101, 568-573.
- Rao, Y.R., Skafel, M.G., Howell, T., and Murthy, R.C. (2003). Physical processes controlling taste and odor episodes in Lake Ontario drinking water. *Journal of Great Lakes Research* 29, 70-78.
- Rashash, D.M., Hoehn, R.C., Dietrich, A.M., Grizzard, T.J., and Parker, B.C. (1996). Identification and control of odorous algal metabolites. (AWWA Research Foundation and American Water Works Association).
- Rinta-Kanto, J.M. and Wilhelm, S.W. (2006). Diversity of microcystin-producing cyanobacteria in spatially isolated regions of Lake Erie. *Applied and Environmental Microbiology* 72, 5083-5085.
- Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Journal of General Microbiology* 111, 1-61.
- Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., and Hess, W.R. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424, 1042-1047.
- Rokas, A. and Holland, P.W. (2000). Rare genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution* 15, 454-459.
- Rokas, A., Williams, B.L., King, N., and Carroll, S.B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798-804.
- Rossello-Mora, R. and Amann, R. (2001). The species concept for prokaryotes. *FEMS microbiology reviews* 25, 39-67.

- Saker, M.L. and Eaglesham, G.K. (1999). The accumulation of cylindrospermopsin from the cyanobacterium *Cylindrospermopsis raciborskii* in tissues of the Redclaw crayfish *Cherax quadricarinatus*. *Toxicon* 37, 1065-1077.
- Sanchez-Baracaldo, P., Hayes, P.K., and Blank, C.E. (2005). Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology* 3, 145-165.
- Sandh, G., El-Shehawy, R., Diez, B., and Bergman, B. (2009). Temporal separation of cell division and diazotrophy in the marine diazotrophic cyanobacterium *Trichodesmium erythraeum* IMS101. *FEMS microbiology letters* 295, 281-288.
- Sawyer, P.J., Gentile, J.H., and Sasner Jr, J.J. (1968). Demonstration of a toxin from *Aphanizomenon flos-aquae* (L.) Ralfs. *Canadian journal of microbiology* 14, 1199-1204.
- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., and Federhen, S. (2010). Database resources of the national center for biotechnology information. *Nucleic acids research* 38, D5-D16.
- Shi, H., Cheng, X., Wu, Q., Mu, R., and Ma, Y. (2012). Assessment and Removal of Emerging Water Contaminants. *J Environ Anal Toxicol* 32, 2161-0525.
- Shi, T. and Falkowski, P.G. (2008). Genome evolution in cyanobacteria: the stable core and the variable shell. *Proceedings of the National Academy of Sciences* 105, 2510-2515.
- Singh, B. and Gupta, R.S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Molecular Genetics and Genomics* 281, 361-373.
- Stanier, R.Y., Siström, W.R., Hansen, T.A., Whitton, B.A., Castenholz, R.W., Pfennig, N., Gorlenko, V.N., Kondratieva, E.N., Eimhjellen, K.E., and Whittenbury, R. (1978). Proposal to place the nomenclature of the cyanobacteria (blue-green algae) under the rules of the International Code of Nomenclature of Bacteria. *International Journal of Systematic Bacteriology* 28, 335-336.
- Steffensen, D.A. (2008). Economic cost of cyanobacterial blooms. In *Cyanobacterial harmful algal blooms: state of the science and research needs*, Springer, pp. 855-865.
- Sugita, C., Ogata, K., Shikata, M., Jikuya, H., Takano, J., Furumichi, M., Kanehisa, M., Omata, T., Sugiura, M., and Sugita, M. (2007). Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: gene content and organization. *Photosynthesis Research* 93, 55-67.
- Svrcek, C. and Smith, D.W. (2004). Cyanobacteria toxins and the current state of knowledge on water treatment options: a review. *Journal of Environmental Engineering and Science* 3, 155-185.

Swingley,W.D., Blankenship,R.E., and Raymond,J. (2008a). Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Molecular biology and evolution* 25, 643-654.

Swingley,W.D., Chen,M., Cheung,P.C., Conrad,A.L., Dejesa,L.C., Hao,J., Honchak,B.M., Karbach,L.E., Kurdoglu,A., and Lahiri,S. (2008b). Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium *Acaryochloris marina*. *Proceedings of the National Academy of Sciences* 105, 2005-2010.

Tamura,K., Nei,M., and Kumar,S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America* 101, 11030-11035.

Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M., and Kumar,S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 28, 2731-2739.

Toepel,J., Welsh,E., Summerfield,T.C., Pakrasi,H.B., and Sherman,L.A. (2008). Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp. strain ATCC 51142 during light-dark and continuous-light growth. *Journal of bacteriology* 190, 3904-3913.

Tomitani,A., Knoll,A.H., Cavanaugh,C.M., and Ohno,T. (2006). The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proceedings of the National Academy of Sciences* 103, 5442-5447.

Tsuji,K., Watanuki,T., Kondo,F., Watanabe,M.F., Suzuki,S., Nakazawa,H., Suzuki,M., Uchida,H., and Harada,K.I. (1995). Stability of microcystins from cyanobacteria—II. Effect of UV light on decomposition and isomerization. *Toxicon* 33, 1619-1631.

Turner,S.E., PRYER,K.M., Miao,V.P.W., and Palmer,J.D. (1999). Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis1. *Journal of Eukaryotic Microbiology* 46, 327-338.

Turner-Designs (2004). *Cyanobacteria detection in water using in vivo fluorometry*. Turner-Designs.

Vanderploeg,H.A., Liebig,J.R., Carmichael,W.W., Agy,M.A., Johengen,T.H., Fahnenstiel,G.L., and Nalepa,T.F. (2001). Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic *Microcystis* blooms in Saginaw Bay (Lake Huron) and Lake Erie. *Canadian Journal of Fisheries and Aquatic Sciences* 58, 1208-1221.

Watson,S.B. (2004). Aquatic taste and odor: a primary signal of drinking-water integrity. *Journal of Toxicology and Environmental Health, Part A* 67, 1779-1795.

Watson,S.B., Ridal,J., and Boyer,G.L. (2008). Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 65, 1779-1796.

Welsh,E.A., Liberton,M., Stockel,J., Loh,T., Elvitigala,T., Wang,C., Wollam,A., Fulton,R.S., Clifton,S.W., and Jacobs,J.M. (2008). The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proceedings of the National Academy of Sciences* 105, 15094-15099.

Westrick, Judy A. Everything a manager should know about algal toxins but was afraid to ask. 2003. AMER WATER WORKS ASSOC 6666 W QUINCY AVE, DENVER, CO 80235 USA.
Ref Type: Generic

Westrick,J.A., Szlag,D.C., Southwell,B.J., and Sinclair,J. (2010). A review of cyanobacteria and cyanotoxins removal/inactivation in drinking water treatment. *Analytical and bioanalytical chemistry* 397, 1705-1714.

WHO (1996). Guidelines for drinking water quality. world health organization 2, 219-226.

WHO (1998). Cyanobacterial toxins: microcystin-LR. In: Guidelines for drinking water quality. World Health Organization 2, 95-110.

WHO (2003). Algae and Cyanobacteria in fresh water. In: Guidelines for safe recreational water environments. Coastal and fresh waters. World Health Organization 1, 136-158.

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

Wilmotte,A. and Herdman,M. (2001). Phylogenetic relationships among the cyanobacteria based on 16S rRNA sequences. In *Bergey's Manual of Systematic Bacteriology*, D.R.Boone, R.W.Castenholz, and G.M.Garrity, eds. (New York: Springer), pp. 487-493.

Wu,D., Hugenholtz,P., Mavromatis,K., Pukall,R., Dalin,E., Ivanova,N.N., Kunin,V., Goodwin,L., Wu,M., and Tindall,B.J. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462, 1056-1060.

Yoshida,T., Makita,Y., Nagata,S., Tsutsumi,T., Yoshida,F., Sekijima,M., Tamura,S., and Ueno,Y. (1997). Acute oral toxicity of microcystin-LR, a cyanobacterial hepatotoxin, in mice. *Natural toxins* 5, 91-95.

Zhaxybayeva,O., Gogarten,J.P., Charlebois,R.L., Doolittle,W.F., and Papke,R.T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome research* 16, 1099-1108.

Zwart,G., Kamst–van Agterveld,M.P., Der Werff–Staverman,V., Hagen,F., Hoogveld,H.L., and Gons,H.J. (2005). Molecular characterization of cyanobacterial diversity in a shallow eutrophic lake. *Environmental Microbiology* 7, 365-377.

Thesis Defence

Phylogenetic and Comparative Genomic Analyses of Heterocystous Cyanobacteria

Mohammad Howard-Azzeh

Background

- Cyanobacteria encompass a large and morphologically heterogeneous cluster of oxygenic phototrophic bacteria
 - Theorized to be the primary and original source of atmospheric molecular oxygen (Bekker et al., 2004)
- Taxonomy originally governed by the botanical code
 - Largely delineated based upon morphological and biochemical characteristics (Desinkachary, 1959; Stanier et al., 1978; Rappka et al., 1979)
- With biochemical advancements, cyanobacteria have been reclassified many times into different systems
 - Comprising of several different and debated orders/subsections (Cavalier-Smith, 2002; Euzéby, 1997; Wilmotte, 2004)
- Cyanobacteria are most notorious for their formation of blooms and production of toxins.

16S rRNA

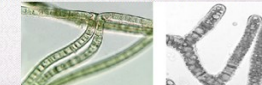
- Currently, the use of 16S rRNA sequence data is one of the most widely implemented approaches to understanding microbial taxonomy (Embley and Stackebrandt, 1997; Roselle-Mora and Amann, 2001)
- Cyanobacterial phylogeny based on 16S rRNA show little correlation with the morphological and biochemical characteristics used to delineate the taxonomy of the phylum (Wilmotte, 2004)
- 16S rRNA gene-based studies group cyanobacteria into 14 clusters
 - The evolutionary and taxonomic relevance of these clusters are ambiguous (Wilmotte, 2004)

Classification

- The only groups that are accepted by the International Code of Nomenclature of Bacteria (ICNB) are: (Cavalier-Smith, 2002; Euzéby, 1997)
 - 3 classes - Chroococcales, Homoeococcales and Gloeobacteriales
 - 6 orders - Chroococcales, Pleurocapsales, Oscillatoriales, Nostocales, Stigonematales and Gloeobacteriales
 - 5 genera - Prochlorothrix, Prochloron, Prochlorococcus, Planktothrixoidis and Halospirulina
- Cyanobacteria taxonomy has several contradictions with phylogeny
 - Extensive polyphyletic branching of the orders Chroococcales and Oscillatoriales across the majority of phylogenetic trees
 - Polyphyletic branching of most genera
 - Unclassified phylogenetic groups

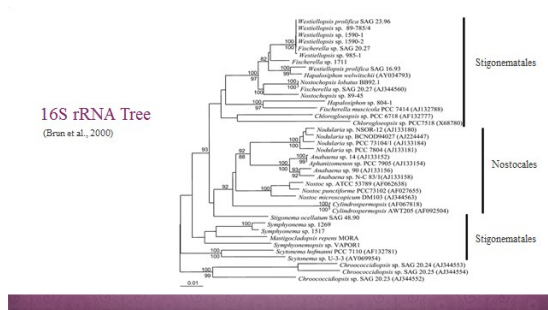
Nostocales & Stigonematales

- Cyanobacterial genera that differentiate into heterocysts all belong to one of two groups, Nostocales and Stigonematales
- Nostocales (35 genera) - Filamentous, heterocystous, uniseriate and divide by binary fission in one plane at a right angle to the long axis of the trichomes (false branching) (Left) (Casterholz, 2001)
 - Form homonomia and are unbranched (Cavalier-Smith, 2002)
- Stigonematales (14 genera) - Filamentous, heterocystous, with longitudinal cell division as well as transverse cell division causing occasional true branching (Right) (Casterholz, 2001)
 - Form homonomia and are branched (Cavalier-Smith, 2002)



16S rRNA

- 16S rRNA sequence data indicates that all heterocystous members of cyanobacteria:
 - Delineated from a common ancestor
 - Form a monophyletic coherent cluster separate from non-heterocystous members
- The same sequence data reveals that species from both orders are extensively intermixed
 - Division of Nostocales and Stigonematales based on the presence or absence of true branching is not reasonable



Rationale

- Taxonomic classification of members belonging to Nostocales and Stigonematales orders are not clear
 - Contradictions between phylogenetic analyses and taxonomy
 - Absence of a consensus phylogeny
 - Ever-evolving classification system
- It is necessary to recognize new sequence-based characteristics
 - Define heterocystous Cyanobacteria in molecular terms
 - Resolve taxonomic and evolutionary relationships
 - Provide a useful means to identifying species within these orders

Cyanobacterial Maximum-Likelihood Distance Tree (Gupta & Mathers 2010)



Methods

Phylogenetic Analysis

- Phylogenetic analysis was performed on:
 - Concatenated sequence of 45 highly conserved proteins (Harris et al., 2003)
 - Present in most living bacteria
- Amino acid sequences were retrieved from NCBI for 48 species of Cyanobacteria
- Multiple sequence alignments were created using Clustal_X 2.1
- Poorly aligned regions of the were removed using Gblocks 0.91b

Concatenated Protein Trees

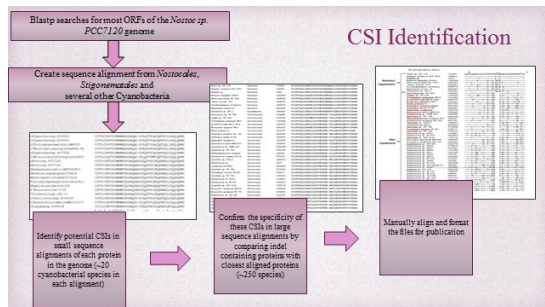
- Resulting concatenated amino acid sequences were used to construct:
 - Neighbour-joining phylogenetic tree
 - 100 bootstrap replications
 - Based on the Jones-Taylor-Thornton model of evolutionary gene change with uniform rates using Mega 5.05
- Concatenated amino acid sequences were also used to construct:
 - Maximum-likelihood phylogenetic consensus tree
 - 100 bootstrap replications
 - Based on the Jones-Taylor-Thornton model of evolutionary gene change with 3 discrete categories gamma distributed substitution rates using Mega 5.05

16S rRNA Trees

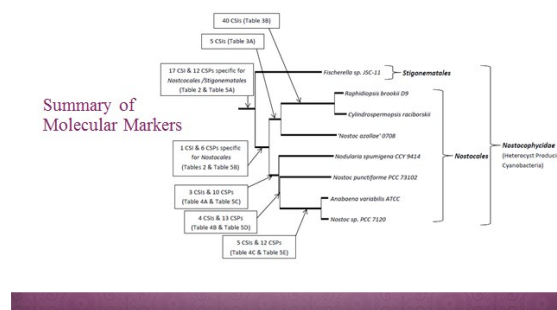
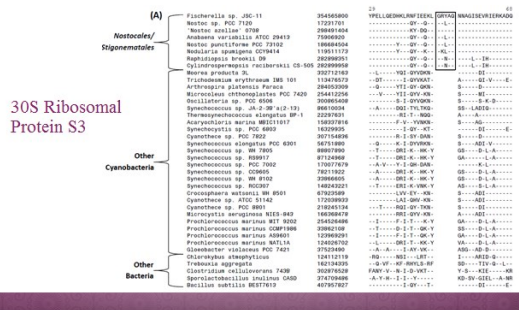
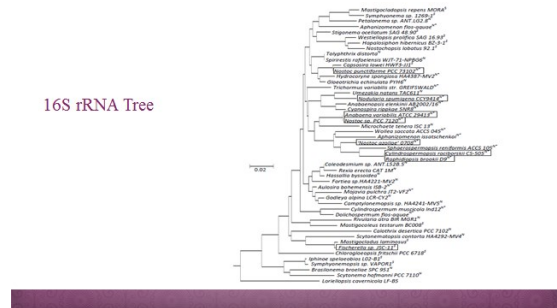
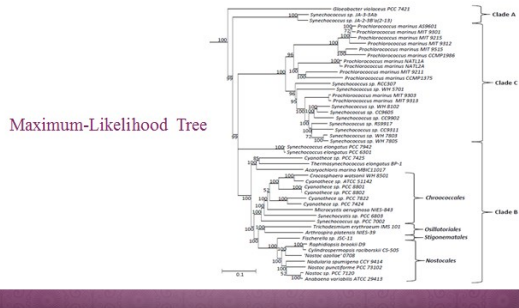
- Phylogenetic analysis was also performed on a 16S rRNA alignment of sequences:
 - From at least one good quality representative from every genera (except *Rhizaria*) in the *Nostocales* and *Stigonematales*
 - Available from the SILVA ribosomal RNA gene database project
- 16S rRNA sequences of all 48 species were used to create a neighbor-joining phylogenetic tree:
 - 1000 bootstrap replications
 - Based on the maximum composite likelihood model of evolutionary gene change with uniform rates of substitution with Mega 5.05

CSP Identification

- Dr. Gupta's earlier work has identified a number of CSPs which were shown to be specific for the species of *Nostocales*
- New sequence information for additional *Nostocales/Stigonematales* has now become available
 - Blastp searches were conducted on all identified proteins from his earlier publication to re-examine their species specificity



Results



Discussion

Nostocales & Stigonematales

- Independent from phylogenetic analyses
 - Ancestrally inherited CSIs/CSPs indicate that the orders are monophyletically derived
 - Distinct from other order level taxonomic groupings
- The relative abundance of molecular markers specific for both *Nostocales* and *Stigonematales* compared to the few that are specific only to *Nostocales*
 - An evolutionarily close relationship shared between them and *Fischerella* sp. JSC-11
- Molecular markers identified are consistent with concatenated protein trees
 - which indicate a very close monophyletically derived relationship
 - Also showing members from both orders forming a strongly supported clade that was separated from all other Clade B Cyanobacteria by a long branch
- In the 16S rRNA tree, *Nostocales* members are entrenched within *Stigonematales* members
 - Both orders are not monophyletic, having representatives from both orders extensively intermingled within each group
 - Thus, the rRNA tree suggests these orders deriving from a common ancestor rather than distinct evolutionary groups divided on the basis of true/false branching

Nostocales Subclade 1

- Species/strains of *Nostocales* with genome sequences available, form two distinct subclades
 - Providing information about the relationships/affinities shared among these members
- Branching patterns observed in concatenated protein trees suggest a subclade consisting of:
 - Cylindrocapsa raciborskii* CS-505
 - Raphidopsis brookii* D9
 - Nostoc azollae* 0708
- This clade is independently supported by several identified CSIs/CSPs that are uniquely shared by all three species/strains
- The rRNA tree further indicates these members branch together and are part of a clade which also encompasses *Spizizenocapsa reifenformis* ACC 105
- Within this subclade, *Raphidopsis brookii* D9 and *Cylindrocapsa raciborskii* CS-505 show a close affinity in both protein tree and the 16S rRNA tree

Nostocales Subclade 2

- A separate subclade of *Nostocales* members is supported by a number of identified molecular markers and consists of the species:
 - Nodularia spumigena* CCY 3414
 - Nostoc punctiforme* PCC 73102
 - Nostoc* sp. PCC 7120
 - Anabaena variabilis* ATCC 29414
- Within this subclade, a more recent divergence among the members is observed:
 - Nostoc punctiforme* PCC 73102
 - Nostoc* sp. PCC 7120
 - Anabaena variabilis* ATCC 29414
- A close affinity is observed between *Nostoc* sp. PCC 7120 and *Anabaena variabilis* ATCC 29414 in the rRNA tree and is supported by several CSIs/CSPs uniquely found in these two species/strains
- Additionally, these three clades/affinities are all independently supported based on the branching of sequenced species/strains in the concatenated protein trees

New Data

Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing (2013)

Patrick M. Shih, Dongying Wu, Amel Latif, Seth D. Axen, David P. Fewer, Emmanuel Talla, Alexandra Calteau, Fei Cai, Nicole Tardieu de Marne, Rosmarie Ruppke, Michael Herdman, Katarina Svoboda, Theresa Courain, Thierry Laurent, Lynne Goodwin, Matt Nolan, Karen W. Davenport, Cliff S. Han, Edward M. Rubin, Jonathan A. Eisen, Tanja Woyke, Mariel Gagger, and Cheryl A. Kerfeld

Maximum Likelihood Trees

(Shih et al., 2013)

0.3

Comparison Between Accepted Classification & Data

Summary of Molecular Markers

Summary

- The evidence provided by phylogenetic and comparative genomic studies suggests:
 - Organisms belonging to *Nostocales* and *Stigonematales*
 - Share a close, likely internixed relationship
 - Forming a monophyletic group that represents the sole heterocyst differentiating Cyanobacteria

References

- Beikar, A., Hollan, H.D., Wang, P.L., Rumba, D.L.H., Stein, H.J., Hansen, L.L., and Beikar, N.J. (2004) Dating the rise of atmospheric oxygen. *Nature* 47, 117-120.
- Dasikachary, T.V. (1959). Cyanophyta, Indian Council of Agricultural Research, monographs on Algae. New Delhi, India.
- Rippka, R., Derr, G.J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979). Genetic assignments, strain histories and properties of pure cultures of cyanobacteria. *Journal of General Microbiology* 111, 1-61.
- Stanier, R.Y., Strain, W.F., Hansen, T.A., Whitton, B.A., Castenholz, R.W., Pflügel, N., Gorlenko, V.N., Kondratieva, E.N., Eimhjellen, K.E., and Waterbury, G. (1973). Proposal to place the nomenclature of the cyanobacteria (blue-green algae) under the rules of the International Code of Nomenclature of Bacteria. *International Journal of Systematic Bacteriology* 23, 333-336.
- Wilmore, A. (2004). Molecular evolution and taxonomy of the cyanobacteria. Springer, pp. 1-25.
- Cavalier-Smith, T. (2002). The neoeukaryotic origin of archaeobacteria, the neobacterial root of the universal tree and bacterial magclassification. *International Journal of Systematic and Evolutionary Microbiology* 22, 7-6.
- Emlay, T.M., and Stackebrandt, E. (1997). Species in practice: exploring uncultured prokaryotic diversity in natural samples. *SYSTEMATIC ASSOCIATION SPECIAL VOLUME* 34, 41-62.
- Harris, J.K., Kelley, J.T., Spigelman, G.B., and Pace, N.R. (2003). The genetic core of the universal ancestor. *Genome research* 13, 407-412.

References

- Rossello-Mora, R., and Amann, R.L. (2001). The species concept for prokaryotes. *FEMS microbiology reviews* 25, 39-67.
- Castenholz, R.W. (2001). *Phyllum BX. Cyanobacteria*. Springer.
- Fluxus J.P. (1997). List of bacterial names with standing in nomenclature. A folder available on the Internet. *International Journal of Systematic Bacteriology* 47, 590-592.
- Bran, Y.V., & Shinkov, L.I. (2000). Prokaryotic development. Washington, DC: ASM press. 11-31
- Gupta, R. S., & Mathews, D. W. (2010). Signature proteins for the major clades of Cyanobacteria. *BMC evolutionary biology*, 10(1), 24.
- Gagger, M.F., and Holmbeck, L. (2004). Polyphyly of true branching cyanobacteria (Stigonematales). *International Journal of Systematic and Evolutionary Microbiology* 54, 349-357.
- Patrick M. Shih, Dongying Wu, Amel Latif, Seth D. Axen, David P. Fewer, Emmanuel Talla, Alexandra Calteau, Fei Cai, Nicole Tardieu de Marne, Rosmarie Ruppke, Michael Herdman, Katarina Svoboda, Theresa Courain, Thierry Laurent, Lynne Goodwin, Matt Nolan, Karen W. Davenport, Cliff S. Han, Edward M. Rubin, Jonathan A. Eisen, Tanja Woyke, Mariel Gagger, and Cheryl A. Kerfeld (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences* 110(3), 1053-1058.

Supplementary Figures and Tables

Supplementary Table 1: Highly conserved proteins used for the construction of Maximum-Likelihood and Neighbor-Joining distance trees.

Protein	Gene Name	Length (Amino Acids)
Alanyl-tRNA synthetase	AlaRS	880
Phosphatidate cytidyltransferase	CdsA	294
Elongation factor G	EF-G	692
Elongation factor P	EF-P	185
Signal recognition particle protein	Ffh	490
Glycine/serine hydroxymethyltransferase	GlyA	427
Cell division transporter substrate-binding protein	FtsY	546
DNA gyrase subunit A	GyrA	872
DNA gyrase subunit B	GyrB	645
DNA polymerase III gamma/tau subunits	HolB	707
Translation initiation factor 2	IF-2	1039
Isoleucyl-tRNA synthetase	IleRS	960
Dimethyladenosine transferase	KsgA	271
Leucyl-tRNA synthetase	LeuRS	872
Methionyl-tRNA synthetase	MetRS	530
Phenylalanyl-tRNA synthetase alpha chain	PheRS	330
DNA polymerase I	PolA	977
Prolyl-tRNA synthetase	ProRS	604
DNA recombination protein A	RecA	357
50S ribosomal protein L2	RibProtL2	287
50S ribosomal protein L4	RibProtL4	210
50S ribosomal protein L5	RibProtL5	182
50S ribosomal protein L6	RibProtL6	182
50S ribosomal protein L15	RibProtL15	148
30S ribosomal protein S2	RibS2	265
30S ribosomal protein S3	RibS3	260
30S ribosomal protein S5	RibS5	174
30S ribosomal protein S8	RibS8	133
30S ribosomal protein S11	RibS11	131
30S ribosomal protein S12	RibS12	127
30S ribosomal protein S13	RibS13	126
30S ribosomal protein S15	RibS15	89
30S ribosomal protein S17	RibS17	176
30S ribosomal protein S19	RibS19	92
RNA polymerase alpha subunit	RpoA	315
RNA polymerase beta subunit	RpoB	1131
RNA polymerase beta prime subunit	RpoC1	1350
Preprotein translocase SecA subunit	SecA	930
Preprotein translocase SecY subunit	SecY	437
Seryl-tRNA synthetase	SerRS	426
Tryptophanyl-tRNA synthetase	TrpS	335
Tyrosyl-tRNA synthetase	TyrRS	398
ATP-dependent DNA helicase	UvrD	772
Valine-tRNA ligase	ValRS	1014
GTP-binding protein	YchF	363

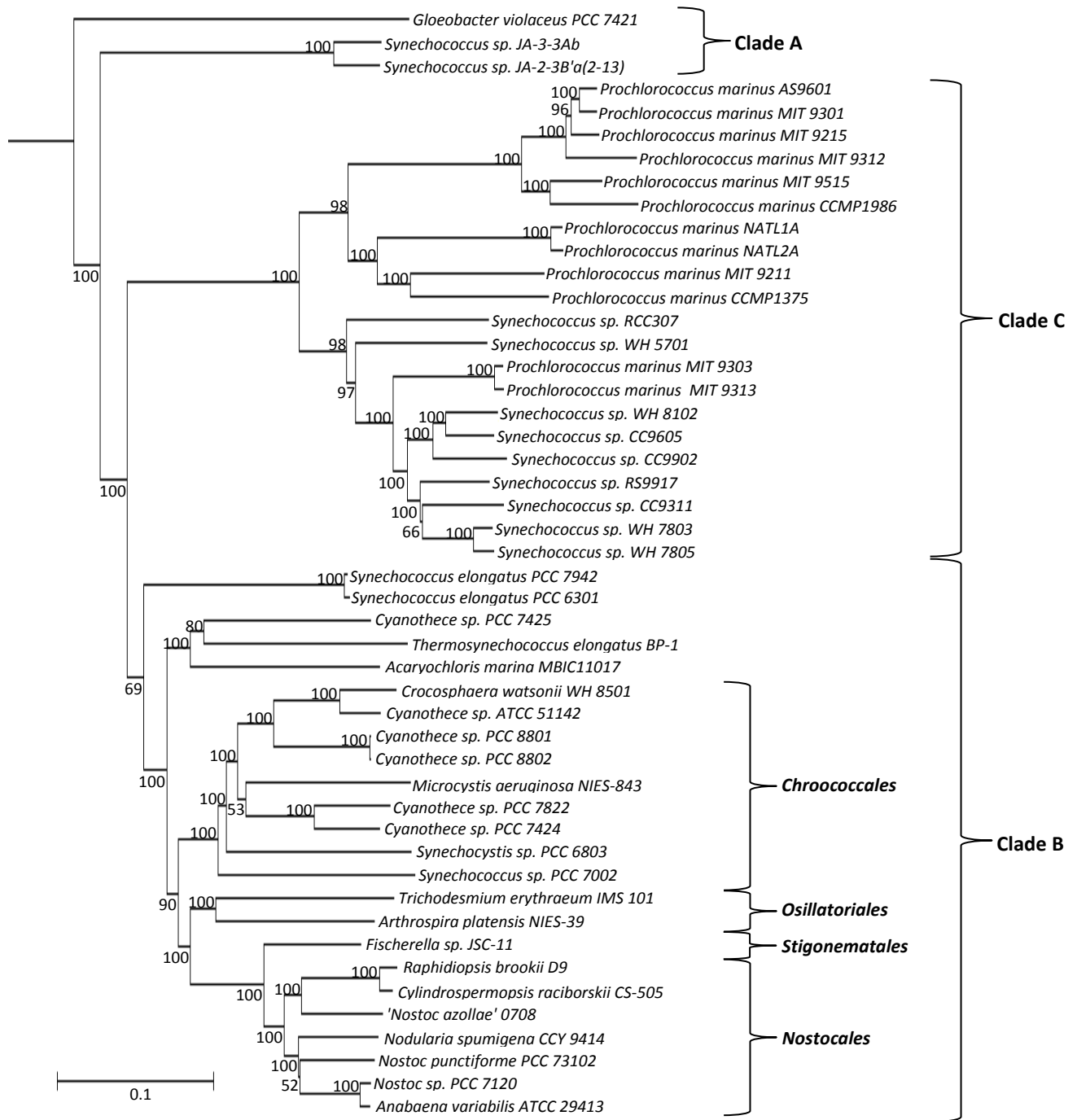
Protein lengths are associated with *Nostoc* sp. PCC 7120

Supplementary Table 2: CSIs specific for individual species/strains of Nostocales/Stigonematales.

Protein Name	GeneBank Identifier	Figure #	Indel Size	Indel Position
(A) CSIs specific for Fischerella sp. JSC-11				
lipid-A-disaccharide synthase	354565229	S. Fig # 71	6 aa ins	33-88
multi-sensor signal transduction multi-kinase	354568923	S. Fig # 72	10 aa ins	477-543
magnesium chelatase subunit H	354565228	S. Fig # 73	7 aa ins	759-818
Ferredoxin--nitrite reductase	354565504	S. Fig # 74	1 aa ins	194-259
Polysaccharide pyruvyl transferase	354566903	S. Fig # 75	48 aa ins	154-240
Major facilitator superfamily	354565836	S. Fig # 76	2 aa ins	291-346
Acetolactate synthase 3 catalytic subunit	354565331	S. Fig # 77	1 aa ins	57-105
Hypothetical protein FJSC11DRAFT_2574	354567198	S. Fig # 78	1 aa ins	29-89
(B) CSIs specific for R. Brookii D9				
HAD-superfamily hydrolase subfamily IIB	282896894	S. Fig # 79	2 aa ins	145-189
Elongation factor P	282896494	S. Fig # 80	3 aa ins	104-138
Hypothetical protein Aazo_1507	282895573	S. Fig # 81	5 aa ins	88-130
Phosphoenolpyruvate synthase	282897267	S. Fig # 82	6 aa del	531-583
Peptidyl-tRNA hydrolase	282898185	S. Fig # 83	2 aa ins	140-176
(C) CSIs specific for C. raciborskii CS-505				
UvrD/REP helicase	282898629	S. Fig # 84	4 aa ins	672-733
Hypothetical protein Aazo_4712	282898905	S. Fig # 85	3 aa ins	84-139
Hypothetical protein Aazo_4712	282898905	S. Fig # 86	4 aa ins	73-139
Protein serine/threonine phosphatase	282899315	S. Fig # 87	2 aa del	240-287
Nitrate reductase	282898751	S. Fig # 88	2 aa ins	6-64
Hypothetical protein FJSC11DRAFT_2719	282899487	S. Fig # 89	2 aa ins	83-122
Hypothetical protein Aazo_2379	282899431	S. Fig # 90	4 aa ins	37-100
Aspartate aminotransferase	282899590	S. Fig # 91	1 aa ins	63-112
(D) CSIs specific for 'N. azollae' 0708				
Glycerol-3-phosphate-transporting ATPase	298491822	S. Fig # 92	7 aa ins	59-116
tRNA/rRNA methyltransferase	298492725	S. Fig # 93	7 aa ins	32-100
Thiamine-monophosphate kinase	298490181	S. Fig # 94	5 aa ins	239-288
Glycosyl transferase family protein	298493172	S. Fig # 95	27 aa ins	202-287
Deoxyribonuclease V	298492125	S. Fig # 96	2 aa del	40-89
Carotenoid oxygenase	298492189	S. Fig # 97	7 aa ins	429-479
Glycyl-tRNA synthetase beta subunit	298491727	S. Fig # 98	6 aa ins	542-590
Glucokinase	298492461	S. Fig # 99	5 aa ins	33-90
DNA gyrase modulator peptidase U62	298490634	S. Fig # 100	2 aa ins	335-385
(E) CSIs specific for N. spumigena				
Hypothetical protein N9414_05744	119509958	S. Fig # 101	8 aa ins	237-293
O-succinylbenzoate synthase	119510629	S. Fig # 102	4 aa del	36-68
ATP-dependent RNA helicase	119510127	S. Fig # 103	1 aa ins	109-142

Twin-arginine translocation pathway signal	119511267	S. Fig # 104	10 aa ins	209-258
Hypothetical protein N9414_02991	119511526	S. Fig # 105	4 aa ins	213-267
Photosystem II chlorophyll-binding protein CP47	119509033	S. Fig # 106	1 aa ins	64-98
HAD-superfamily hydrolase subfamily IA, variant 3	119510625	S. Fig # 107	8 aa ins	76-116
Alpha amylase	119513191	S. Fig # 108	3 aa ins	104-153
Hypothetical protein N9414_11444	119513308	S. Fig # 109	4 aa ins	40-81
Transaldolase	119511494	S. Fig # 110	7 aa ins	198-293
Phosphoenolpyruvate carboxylase	119512252	S. Fig # 111	2 aa ins	859-899
Thiosulphate-binding protein	119512923	S. Fig # 112	1 aa ins	125-159
Serine/threonine kinase	119508927	S. Fig # 113	9 aa ins	480-540
Phenylalanyl-tRNA synthetase	119509505	S. Fig # 114	3 aa ins	586-644
Photosystem II protein D2	119512824	S. Fig # 115	1 aa ins	211-254
Glutamine amidotransferase	119510861	S. Fig # 116	5 aa ins	1254-1303
Cobalamin biosynthesis precorrin-3 methylase	119512863	S. Fig # 117	5 aa ins	394-436
UDP-N-acetylenolpyruvylglucosamine reductase	119511606	S. Fig # 118	3 aa ins	170-229
Hypothetical protein N9414_08994	119512228	S. Fig # 119	7 aa ins	397-440
Serine/Threonine protein kinase	119511124	S. Fig # 120	2 aa ins	42-99
Two-component response regulator	119509039	S. Fig # 121	2 aa ins	74-106
Alpha-glucosidase	119513109	S. Fig # 122	36 aa del	280-360
Serine/threonine kinase	119508927	S. Fig # 123	3 aa del	1588-1650
Peptidase C56	119509804	S. Fig # 124	8 aa ins	215-373
Folylpolyglutamate synthase	119510017	S. Fig # 125	9 aa ins	196-249
(F) CSIs specific to N. punctiforme PCC 73102				
Methyltransferase type 11	186683804	S. Fig # 126	8 aa ins	26-71
Adenylate kinase	186682826	S. Fig # 127	23 aa ins	26-84
Precorrin-4 C11-methyltransferase	186684080	S. Fig # 128	1 aa ins	71-117
Glycosyl transferase family protein	186686042	S. Fig # 129	8 aa ins	286-370
Cell envelope-related transcriptional attenuator	186685192	S. Fig # 130	3 aa ins	192-227
Methionyl-tRNA synthetase	186680640	S. Fig # 131	1 aa ins	356-417
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	186684135	S. Fig # 132	7 aa ins	89-135
Hypothetical protein Npun_F3493	186683647	S. Fig # 133	1 aa ins	108-151
Methionine synthase	186680613	S. Fig # 134	2 aa ins	34-80
Aldo/keto reductase	186681282	S. Fig # 135	9 aa ins	219-269
Primosome assembly protein	186680838	S. Fig # 136	5 aa ins	127-183
23S rRNA (adenine(2503)-C(2))-methyltransferase	186684082	S. Fig # 137	8 aa ins	104-141
Histidyl-tRNA synthetase	186685474	S. Fig # 138	8 aa ins	58-110
Hypothetical protein Npun_R2352	186682667	S. Fig # 139	2 aa ins	190-250
Hypothetical protein Npun_F4258	186684380	S. Fig # 140	4 aa ins	74-116
Glyoxalase/bleomycin resistance protein/dioxygenase	186684740	S. Fig # 141	9 aa ins	50-95
rRNA SAM-dependent methyltransferase	186684748	S. Fig # 142	12 aa ins	117-191

Amidase	186682798	S. Fig # 143	7 aa ins	297-349
Hypothetical protein Npun_R4999	186685090	S. Fig # 144	1 aa ins	64-121
Bifunctional folylpolyglutamate synthase/ dihydrofolate synthase	186685094	S. Fig # 145	5 aa ins	360-400
(G) CSIs specific for Nostoc sp. PCC 7120				
30S ribosomal protein S3	17231701	S. Fig # 146	7 aa ins	203-260
Elongation factor G	17231875	S. Fig # 147	13 aa ins	147-197
Aspartate aminotransferase	17228534	S. Fig # 148	7 aa ins	270-327
(H) CSIs specific for Anabaena variabilis ATCC 29413				
Putative glycerol-3-phosphate acyltransferase	75908940	S. Fig # 149	1 aa ins	160-194
Glycoside hydrolase	75910649	S. Fig # 150	1 aa ins	896-926



Supplementary Figure 1

A Neighbor-Joining Distance Tree based on 45 concatenated sequences for 48 species of Cyanobacteria. The tree was rooted using Sequences for *Staphylococcus aureus* subsp. *Aureus* Mu50 and *Bacillus subtilis* str. 168. Numbers located at nodes indicate the bootstrap values out of 100. Major clades and subclades of Cyanobacteria resolved in the tree are indicated.

		484	531
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17231417	RYSCEIERYLKE RLEKL RYEFPKTLVVNSHTTNTPGHAAVDFVAAIN
	Anabaena variabilis ATCC 29413	75907993	-----
	Nostoc punctiforme PCC 73102	186685021	----- -Q-- -RQ-----S-----E---I----
	'Nostoc azollae' 0708	298492899	S-----H----- -L-- -K---N-----S-----IT---
	Nodularia spumigena CCY9414	119509214	-----S----- -Q-- -S--IS-R--S-----GD--
	Raphidiopsis brookii D9	282897691	-----D--QN --Y-- KRQ--HS--S-Q-----A-P-IECLS---
	Cylindrospermopsis raciborskii CS-505	282900956	-----D--QD --Y-- KRQ--HS--S-Q-----A-P-IECLS---
	Fischerella sp. JSC-11	354568775	-----AD----- -Q-- QKR-A-----S-L-----E-A--ITS--
	Cyanothece sp. PCC 7425	219883216	--PDKRI-RID-ESVAD-Q-P-YGI-EQL-
	Cyanothece sp. PCC 8802	257059993	NWGTQ-L-S---K QFPDA-I-RID-ESLTD-N---YQCICKL-
	Cyanothece sp. PCC 8801	218246932	NWGTQ-L-S---K QFPDA-I-RID-ESLTD-N---YQCISQL-
	Microcystis aeruginosa NIES-843	166367098	KWGTI-L-S--RK QFPQK-I-RID-ESLPDY----YQAIGNL-
	cyanobacterium UCYN-A	284929649	KWGTQ-L-L--QQ QFPEA-I-RID-ESLID-N-S-YKCIDHLD
	Cyanothece sp. CCY0110	126660034	KWGTQ-L-S---K QFPDA-I-RID-QSLTN-N-D-YQCITQL-
	Other Cyanobacteria	Synechococcus sp. PCC 7002	170076845
Crocospaera watsonii WH 0003		357262293	KWGTQ-L-L---K QFPTANI-RID-QSLTD-N-P-YQCITKL-
Crocospaera watsonii WH 8501		67923848	KWGTQ-L-L---K QFPTANI-RID-QSLTD-N-P-YQCITKL-
Microcystis aeruginosa PCC 9443		389730945	KWGTI-L-Y--RK QFPQK-I-RID-ESLQDYS-D-YEIGNL-
Cyanothece sp. PCC 7424		218439140	QWGTL-L-A--NK LFPDL-I-RLD-ESLAD-H-P-YNCITRL-
Cyanothece sp. PCC 7822		307151280	QWGTLSL-A--K QFPTL-I-RLD-ESLAD-TQ--YNCITHLD
Cyanothece sp. ATCC 51142		172036781	KWGTQ-L-S--NK LFPDA-I-RID-QSLTD-H-D-YQCITQL-
Cyanothece sp. ATCC 51472		354554593	KWGTQ-L-S--NK LFPDA-I-RID-QSLTD-H-D-YQCITQL-
Synechocystis sp. PCC 6803		16332075	AWGTRNL-A---K QFPDRI-RIDAESLSD-H-P-HGSLTNL-
Lyngbya sp. PCC 8106		119484516	QWGTQ-L-T---T QFPNL-I-RID-ESLSD-S-P-YGCINQL-
Arthrospira platensis NIES-39		291568873	KWGTS-L-A-F-T QFPQL-I-RID-QSLAD-H-P-YGCINRL-
Arthrospira platensis Paraca		284050328	KWGTS-L-A-F-T QFPQL-I-RID-QSLAD-H-P-YGCINRL-
Microcoleus chthonoplastes PCC 7420		254411821	QWGTQ-L-A--RS -FPEA-I-RID-ESLTD-T-P-YGCMRDLD
Arthrospira sp. PCC 8005		376002382	KWGTS-L-A-F-A QFPQL-I-RID-QSLAD-H-P-YGCINRL-
Arthrospira maxima CS-328		209526706	KWGTS-L-A-F-A QFPQL-I-RID-QSLAD-H-P-YGCINRL-
Trichodesmium erythraeum IMS101		113474847	KWGTRAL-A---K QFPKL-I-RID-ESLAEVN-P-YGCIKSL-
Moorea producta 3L		332705148	QWGTQ-L-A---K QFPHLRI-RID-ESLGETN-P-MGCINNL-

Supplementary Figure 2

Partial sequence alignment for the hypothetical protein Npun_R4929, showing a 5 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		379	438	
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17231510	LSSTNPQANCQGVICSETGGCMRMETHGGAQ A AGLFQQVSSLSSEQKAEALLSQVQRWGRE	
	Anabaena variabilis ATCC 29413	75907904	-----D-----S-----T-----	
	Nostoc punctiforme PCC 73102	186684152	-----D-----S-----T-----	
	'Nostoc azollae' 0708	298490390	-----D-----S-----T-----	
	Nodularia spumigena CCY9414	119511492	----LE-D-----G-----	
	Raphidiopsis brookii D9	282897264	---S-TE-----T-----Q---S-----N--A--P-F-----	
	Cylindrospermopsis raciborskii CS-505	282899816	---S-TE-----T-----Q---S-----N--A--P-F-----	
	Fischerella sp. JSC-11	354568217	-----L--T-----Q---S T-----T-----V-----H-	
	Oscillatoria sp. PCC 6506	300866082	-N--EK-D---L-----S-----T-RIN--TP-FD-QT-Q---HL----D	
	Moorea producta 3L	332710465	-G---L--D-C--L--Q-T-----AG---SCRI----P-FD--T-F-----L-----	
	Microcoleus chthonoplastes PCC 7420	254413372	-G---MD-D-C--L--Q-T-----A----CRI---TPMFD--T-Q---L--S-D	
	Microcoleus vaginatus FGP-2	334118083	-N--EK-D---L--V-----K---TDR-N--P-FD-VT-H--G-HL-S---D	
	Arthrospira platensis Paraca	284052912	-T---R--C-IL---T-----AG-S--SCQTE--T-VTD---Y-MA--L-----D	
	Lyngbya sp. PCC 8106	119486356	-T-S--N--C-IL---T-----SG-S--SCRTE--TAT-D---S-MT--L-----D	
	Arthrospira platensis NIES-39	291565751	-T---R--C-IL---T-----AG-T--SCQTE--T-VTD---Y-MA--L-----D	
	Arthrospira maxima CS-328	209528064	-T---K--C-IL---T-----AG-S--SCQTE--T-VTD---Y-MA--L-----D	
	Arthrospira sp. PCC 8005	376004853	-T---K--C-IL---T-----AG-S--SCQTE--T-VTD---Y-MA--L-----D	
	Trichodesmium erythraeum IMS101	113474534	-D-S--S-D-C-IL---T---L-AG-S--SCRTE--A--D---Y-MA--L-----	
	Other Cyanobacteria	Cyanothece sp. PCC 7424	218438479	-----L--D-C--L-----T-----AS---SCRI--T--AD-NT-Q--G--LR---D
		Cyanothece sp. PCC 8802	257061610	-----L--D-C--L--G-T-----AS---CRIE--T--AD-NT-Y--GR-L---TD
Cyanothece sp. PCC 8801		218248542	-----L--D-C--L--G-T-----AS---CRIE--T--AD-NT-Y--GR-L---TD	
cyanobacterium UCYN-A		284929726	-----K-D-C--L--G-T-----AS---CRIE--T--AD--TDN-IQR-L---TD	
Crocospaera watsonii WH 0003		357262533	-----L--D-C--L--G-T-----AS---CRIE--T--AD-QTDQ-IQK-L---TD	
Crocospaera watsonii WH 8501		67922964	-----L--D-C--L--G-T-----AS---CRIE--T--AD-QTDQ-IQK-L---TD	
Cyanothece sp. ATCC 51142		172037448	-----L--D-C--L--G-T-----AS---CRIE--T--AD--TDQ-IEK-L---PD	
Cyanothece sp. PCC 7425		220908800	-A---LE-D-C-I---K-----AG---SC-TY--P-A--RS-V---EL---D	
Cyanothece sp. CCY0110		126659462	-----L--D-C--L--G-T-----AS---CRIE--TP-GD--T-H-IEK-L---PD	
Microcystis aeruginosa PCC 980		389788935	-----L--D-C--G-T-----AS---SCRIE--H-ED--T-T--GK-L---QD	
Cyanothece sp. PCC 7822		307153983	-----L--D-C--L---K-----AS---SCRIE--T--AD-NT-Q--G--LR---QD	
Microcystis aeruginosa NIES-843		166367106	-----L--D-C--G-T-----AS---SCRVE--H-ED--T-T--GK-L---QD	
Microcystis sp. T1-4		390442057	-----L--D-C--G-T-----AS---SCRVE--H-ED--T-T--GK-L---QD	
Synechocystis sp. PCC 6803		16330360	-G---T--Q---L--G-V-----AG---NYRV--TA-DD-NT-Q--GR-L-----D	
Microcystis aeruginosa PCC 7806		159027869	-----LH-D-C--G-T-----AS---SCRIE--H-ED--T-T--GK-L---QD	
Synechococcus elongatus PCC 6301		56751778	-T-E--D-D-C-IL---T-----AG---SCRTE--APP-NETS-S--AM-M---D	
Synechococcus sp. PCC 7335		254424330	-E---HD--N-IL--K-A-----SG-S--SS-VEE-T---D-RSDL--E--L-S--ED	
Synechococcus sp. PCC 7002		170078067	-T-N-LD-D-C-----S--H--SG---SCRIE--TP-AD-AT-H--G--L---KD	
Thermosynechococcus elongatus BP-1		22298082	-T---LE-D-C-I---R-----AG-S--LCRTYH-TP--D-T--I---EL---PD	
Acaryochloris marina MBIC11017		158336019	-A---L--D-C--L---R-----AG---SCRI-H-TP-QD-SS-T-----L--DQD	
Acaryochloris sp. CCMEE 5410	359460906	-A---L--D-C--L---R-----AG---SCRI-H-TP-QD-SS-T-----L--DQD		
Synechococcus sp. JA-3-3Ab	86605107	-T-S-QE-DAC--L---T-----AR---RCVVR--AP-EQESL-T--AAEL--P-PD		
Synechococcus sp. JA-2-3B'a(2-13)	86610338	-T-S-QE-DAC--L---IT-----AM---RCVVR--AP-EQDSDLT--AAEL--L-PD		

Supplementary Figure 3

Partial sequence alignment for an outer membrane adhesion protein (OpcA), showing a 1 aa insert that is specific for all Nostocales and Stigonematales.

			123	186	
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii	282899760	VHFHEVGAVDAIVDVGTCLSLDWLGIDS	DETGLPL VYCSAFPTGGGTVRAAHGRMPVPVAVL	
	Raphidiopsis brookii D9	282895922	-----	-Q-----	
	'Nostoc azollae' 0708	298493015	-N-----	G----- NQE-- I--P-----Q-----	
	Nostoc punctiforme PCC 73102	186681196	-----I-----	G-----E- NDE-W- L-----Q-A-----	
	Anabaena variabilis ATCC 29413	75908904	-----G-----	A- -NA-F- L-----Q-A-----	
	Nostoc sp. PCC 7120	17227696	-----G-----	A- -NA-F- L-----Q-A-----	
	Nodularia spumigena CCY9414	119509690	-----G-----	A- NDQ--- L---L-----E-A-----	
	Fischerella sp. JSC-11	354565251	-----G-----	----- L-----Q-A-----	
	Cyanothece sp. PCC 7822	307153962	-----T-----	G-----Q L---M-----K---L-----	
	Cyanothece sp. PCC 7424	218438455	-----T-----	G-----Q L---M-----K---CL-----	
	Cyanothece sp. PCC 8802	257061455	-----I--I-----	G-----EA L--EM-----W---L-----I	
	Cyanothece sp. PCC 8801	218247211	-----I--I-----	G-----EA L--EM-----W---L-----I	
	Cyanothece sp. PCC 7425	220909106	-----T--L-----	G-----EE L---L-----K---L-----	
	Cyanothece sp. CCY0110	126660838	-----T-----	G--Y-D-E L---L-----W---L-----	
	Cyanothece sp. ATCC 51142	172036647	-----T-----	S--G--Y-E-E L---L-----W---L-----	
	Synechococcus elongatus PCC 79	81301160	-----T-----	G--Y--EA ---L-----K---QLS---	
	Crocospaera watsonii WH 8501	67920200	-----T-----	G--Y-D-QT L---L-----W---L-----	
	Crocospaera watsonii WH 0003	357266554	-----T-----	G--Y-D-QT L---L-----W---L-----	
	Other Cyanobacteria	Microcystis aeruginosa NIES-84	166363888	-----T-----	G-----SVSE L--SL-----Q---L-----V
		Synechococcus sp. PCC 7002	170078317	-----T--L-----	G-----E--E L--APM-----K---IL-----
Acaryochloris marina MBIC11017		158338240	-----T-----	G--H--VEE MF---L-----K---KLS---	
Thermosynechococcus elongatus BP-1		22299839	-----L-----	G--Y-DVRQ ---L-----K---QL-----	
Synechocystis sp. PCC 6803		16329325	-----T-----	G--Y---Q C-W--L---S-----DL-----	
Synechococcus sp. PCC 7335		254425164	-----T--I-----	G-----N--H LV--PL-----CE--LL-----	
Synechococcus sp. JA-3-3Ab		86605652	-----L-----	VG---QVER -IS-PH-I--W-DTE--KLA-----I	
Synechococcus sp. BL107		116072860	-----L--V--V-AA--	Q-NPSR IC--PL-A-R--QT---LL-----	
Synechococcus sp. CC9311		113952851	-----I--I-SL--V--V-AGIEN-SPS-	I---P-A-H-R-KT---VL-----	
Synechococcus sp. CC9902		78185261	-----L--V--V-AA-HQ-NPSR	IC--PL-A-R---T---LL-----	
Synechococcus sp. CC9605		78212216	-----I--L--V--V-AAI-D-NPAR	IV--PL-A-S---ST---LL-----	
Synechococcus sp. RCC307		148242740	-----I-SL--VI-V-AG-LH--VQ-	LWATPP-A-H-Q--T---VL--A----	
Synechococcus sp. JA-2-3B'a(2-13)		86607683	-----L-----	IG---QVER -LS-PH-V--W-STE--KLA-----I	
Arthrospira platensis Paraca		284053292	-----T-----	G---K--R L-----V---F-----IL--A----	
Microcoleus chthonoplastes PCC		254413429	-----T--L-----	G-----R L---PL-----Q---L-----	
Arthrospira maxima CS-328		209522643	-----T-----	G---K--R L-----V---F-----IL--A----	
Arthrospira sp. PCC 8005		376007570	-----T-----	G---K--R L-----V---F-----IL--A----	
Trichodesmium erythraeum IMS101		113474503	-----T---I-----	G---D--L L---PL-I---K---L-----	
Arthrospira platensis NIES-39		291568166	-----T-----	G---K--R L-----V---F-----IL--A----	
Moorea producta 3L		332708685	-----T---I-----	G-----Q I---PL-----W---L-----	
Oscillatoria sp. PCC 6506		300868278	-----T-----	G---D--E L---L-----W---LA--T---	
Microcoleus vaginatus FGP-2		334121011	-----A---I-----	G---D-QE F-F--M-----K---KLA--T--M	
Prochlorococcus marinus CCMP1986		33861044	-----I--I-SL--I-V-T-IEY-KPKK	-F-NEPTL-R-FIK-D--KLS--S--I	
Prochlorococcus marinus MIT 9301		126695852	-----I--I-SL--VI-V-AA-NY-NPKK	--NEPIL-K-F-QTE--KLS--A--I	
Prochlorococcus marinus AS9601		123968080	-----I--L-SL---I-V-AA-NY-NPKR	--NQPML-K-F-QTE--KLS--P--I	
Gloeobacter violaceus PCC 7421		37520111	-----T--L--V-----	G--Y-N-EA LH---L-V---L-H---V---T---V	

Supplementary Figure 4

Partial sequence alignment for the hypothetical protein DUF111, showing a 7 aa insert that is specific for all *Nostocales* and *Stigonematales*.

			12	58
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227740	YQEAKNAYTQPQAVIVLGGST RNL	EREKFTASFARKHPNLPWIWISGG
	Anabaena variabilis ATCC 29413	75908946	-----	-----Q-----
	Nostoc punctiforme PCC 73102	186685641	-K-VQ-QFV-----V-----	-R-----E-V-Q---I---T--
	'Nostoc azollae' 0708	298492064	-K-VQ-QLLH----L-----	-R-----N--N---I-----
	Nodularia spumigena CCY9414	119512189	---S-IEELP----L-----	SH-----N---Q--S-----
	Raphidiopsis brookii D9	282897433	-R-VQ-QFSI----L-----S	KH---R-A-N--K---I---T--
	Cylindrospermopsis raciborskii CS-505	282900742	-R-VQ-QFSI----L-----S	KR---R-A-N--K---I---T--
	Fischerella sp. JSC-11	354565955	-K-IQGQF-----IL-----	KK-----D-----
	Microcoleus vaginatus FGP-2	334119335	--QV-SEFQR----L----A-	---V-A-K---DY-E---V-S-
	Oscillatoria sp. PCC 6506	300867952	FNQI-SYWE----LF----AA	---V-A-K---E-Q---V-S-
Other Cyanobacteria	Trichodesmium erythraeum IMS101	113475490	--QISGTIKP---LL----AI	---A-A-E---Q---D---V-S-
	Microcoleus chthonoplastes PCC 7420	254409604	-KIQSYLI--E-IL----EE	---L-A-D--Q--D-H--V-S-
	Moorea producta 3L	332709823	-KRIESYLV--KVAL----E	S--RYA-K--I---D-N--V-S-
	Crocospaera watsonii WH 0003	357262463	-KQLQSHFV--E-IF----HE	D--R-A-QL-L--D---V-S-
	Cyanothece sp. CCY0110	126654814	-KQLQSYFA--E-IF----HE	D--R-A-KL-LE-----V-S-
	Crocospaera watsonii WH 8501	67923734	-KQLQSHFV--E-IF----HK	D--S-A-QL-L--D---V-S-
	Cyanothece sp. ATCC 51142	172037429	-KQLQAYFV--E-IF----HE	D--R-A-KL-LE--D---V-S-
	Cyanothece sp. PCC 8801	218246780	-KQVQSYRVK-E-IF----HE	---R-A-QL-KD--T---V-S-
	Microcystis aeruginosa NIES-843	166368105	-HDLRQQWLK-E-IF----HA	D--R-A-KL-KEY-D---V-S-
	Cyanothece sp. PCC 7424	218440733	-KIQSYLVK-E-IL----HE	---R-A-HL-A-N-Q---V-S-
	Cyanothece sp. PCC 7822	307151691	-N-IQSYLVK-E-IV----HE	---RYA-HL-S-N-T---V-S-
	Microcystis aeruginosa PCC 7806	159027601	-LDLRQQWLK-E-IF----HA	D--R-A-KL-KEY-D---V-S-
	Synechococcus sp. PCC 7002	170078552	--RIQG-LRPAK-IF----HE	---R-A-Q--QE--D-KV-V-S-
	Synechococcus sp. PCC 7335	254423350	-TRYQQFLT-PVAL----AP	---R-A-Q--KT--TVE----S-

Supplementary Figure 5

Partial sequence alignment for the hypothetical protein Npun_R5589, showing a 3 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		124	166
<i>Nostocales/ Stigonematales</i>	Nostoc sp. PCC 7120	17231636	WPTVAQKARENPELAAVKIVPLV E KPLEKLLQAWGIEAIAPVG
	Anabaena variabilis ATCC 29413	75906983	---A-----
	'Nostoc azollae' 0708	298489753	---A-----Q-E---L--- - ---DQ-----V---S--
	Nostoc punctiforme PCC 73102	186683868	---A---Q---Q----- Q -----V-----
	Nodularia spumigena CCY9414	119512851	---A---Q---Q---I----- Q ---R---E---Q-----
	Cylindrospermopsis raciborskii CS-505	282902006	F--A-H-V-A--Q-E-I--L--- - -----ST---Q---S--
	Raphidiopsis brookii D9	282897666	F--A-H-V-A--Q-E-I--L--- - -----ST---Q---S--
	Fischerella sp. JSC-11	354566554	F--A-H-----Q-S--N-I--- - ---R---Q-----S--
	Oscillatoria sp. PCC 6506	300867934	---A-S--Q---D-S---LL--L R-I-Q---E--V---T--
	Lyngbya sp. PCC 8106	119490897	----VY-VKQ-SK-P---LL--L R-I-R-M-Q--V---S--
	Trichodesmium erythraeum IMS101	113474729	----IY--KQ--S---I--L--L R-V---MGQ-NV-SL----
	Microcoleus vaginatus FGP-2	334117607	---A-S--Q--HN-R-Q-LL--L R-V-Q--EH--V-S-----
	Moorea producta 3L	332709042	---A-YA-QQ-EQ-P--RLL--- --V-Q--EK--V---S--
	Arthrospira platensis Paraca	284052604	--K-T--VQ---QIP-KN-L--- --V-R-V-Q--V-I---IA
	Arthrospira platensis NIES-39	291572096	--K-T--VQ---QIP-KN-L--- --V-R-V-Q--V-I---IA
	Arthrospira maxima CS-328	209523931	--K-TE-VE---QIP-KN-L--- --V-R-V-Q--V-I---IA
	Arthrospira sp. PCC 8005	376003763	--K-TE-VE---QIP-KN-L--- --V-R-V-Q--V-I---IA
Other Cyanobacteria	Microcoleus chthonoplastes PCC 7420	254413436	Y--ACRMIE-R-D-P-KNL-A-F T--N-I-N--Y-P-GQPW
	Microcoleus vaginatus FGP-2	334119165	Y-AARKM-EAK-DMP-KNLSA-F T--N--SS-DF-M-GSAW
	Acaryochloris sp. CCMEE 5410	359457342	---A-YA--K--DAP---LL--L S-----AS--VSP-GE--
	Cyanothece sp. PCC 8801	218248074	---A-VV-KK--Q-S---LL--- --ITE--KQ-NV-TN-T--
	Cyanothece sp. PCC 7822	307154346	---A-LS-QK--K-P-I-LL--L --MAI--KQ--V---S--
	Synechococcus sp. PCC 7335	254421896	---I-KR-QDR---S-A--L-F R-V-Q-MTE--V-P--A-D
	Cyanothece sp. PCC 7425	220906960	--NAVYA--N--Q-P--NLI--T R-I-N---H---P-GE--
	Acaryochloris marina MBIC11017	158338689	---A-YA--K--DAP-I-LL--- S-----GS--VIP-GE--
	Microcystis aeruginosa NIES-843	166366761	---AEAV--K--D-----LS-- --IMQ--ER--VQP-DS--
	Cyanothece sp. CCY0110	126654869	---A-IV-QK---IS---LL--- --IFE--KQ--L-MKGS--
	Cyanothece sp. ATCC 51142	172037614	---A-IV-QK---IS---LL--- --IFE--KQ--L-MKGS--
	Synechocystis sp. PCC 6803	16331470	---A-AA-QQ--QWP-Q-LL--T --IAT---Q-QV-V--T--
	Thermosynechococcus elongatus BP-1	22297705	---A-YA-HQ--TAP-QR-L--- --VFQ--EQ--VRL-GE--
	Synechococcus sp. JA-3-3Ab	86605062	Y--REI--RK-DWP-QNVTA-F RS-DA---Q--Y---GSPL
	Cyanothece sp. PCC 7425	220907852	Y-S-SRMSQAK-D-P-KTL-S-F T--N-TRS--Y-P-GS-W
	Synechococcus sp. JA-2-3B'a(2-13)	86608971	Y--R-I-QRK-DWP-QNVTA-F RS-DA---H--Y-S-GSPL
	Cyanothece sp. PCC 7424	218437983	---A-VS-QK--K-P-L-LL--L --MAM--KQ--V---S--
	Synechococcus sp. PCC 7002	170079229	-SAA-KA-QD-AQFP-RTL-A-T Q-F-Q--AS-DVTP-GT--
	Cyanothece sp. CCY0110	126659174	Y-SAIKM-QHK-D-P-KNL-A-F T--GN--SD--Y-T-GEPW
	Gloeobacter violaceus PCC 7421	37523372	A--LRRA-QG--T-G-AGALR-L S--G--AD--F-P-GA-W

Supplementary Figure 6

Partial sequence alignment for an XRE family transcriptional regulator protein, showing a 1 aa insert that is specific for all *Nostocales* and *Stigonematales*.

	250			
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282897379	AVYVAVRQYIISKDLTIQQNL LTVQQNI ITQQQTIDSYFQGISDLVLD	
	Cylindrospermopsis raciborskii	282901389	-----	
	Nostoc punctiforme PCC 73102	186683859	-----V-----	
	'Nostoc azollae' 0708	298489761	-----V-----	-V-----
	Anabaena variabilis ATCC 29413	75906976	-----V-----	-----V-----
	Nostoc sp. PCC 7120	17231644	-----V-----	-----V-----
	Nodularia spumigena CCY9414	119511359	-----V-----	-----V-----
	Fischerella sp. JSC-11	354566565	--I-----	-----L-----
	Microcystis aeruginosa NIES-84	166364511	-----E-V-----R	-----A--
	Microcystis sp. T1-4	390440663	-----E-V-----R	-----A--
	Cyanothece sp. ATCC 51472	354553243	-----A-V-----L----	-----T-----
	Cyanothece sp. ATCC 51142	172036460	-----A-V-----L----	-----T-----
	Synechococcus sp. PCC 7335	254425132	-----R--V----	-----E----
	Synechococcus sp. PCC 7002	170078803	-----E-V-R-----R	-----T-----A-N
	cyanobacterium UCYN-A	284928681	-----A-V-----L--R	-----T-----
Cyanothece sp. PCC 7424	218441416	-----Q-V-R-----R	-----A--V-----	
Cyanothece sp. PCC 7822	307150130	-----Q-V-R-----R	-----A--V-----	
Cyanothece sp. PCC 8801	218248119	--I--A-V-----L----	-----T-----	
Cyanothece sp. CCY0110	126660151	-----A-V-----L--R	-----T-----	
Crocospaera watsonii WH 0003	357260345	-----A-V-----L--R	-----T--V-----N	
Thermosynechococcus elongatus	22298380	-A-----V-R-----R	-----A--I--	
Crocospaera watsonii WH 8501	67926006	-----A-V-----L--R	-----T--V-----N	
Synechocystis sp. PCC 6803	16330156	-----Q-V-----R	-----V-A-----A-S	
Acaryochloris marina MBIC11017	158339057	-L-----V--E--E--R	-----A-----IN	
Cyanothece sp. PCC 7425	220906859	-A-----V-R--M--R	-----A--I--N	
Acaryochloris sp. CCME 5410	359457157	-L-----V--E--E--R	-----A-----IN	
Cyanothece sp. PCC 7425	220906858	TA-----V-R--M--R	-----L--A----	
Synechococcus elongatus PCC 63	56751484	-LVI-----A-D-R--T--Q	--A----FI-----IS-	
Synechococcus elongatus PCC 79	81298833	-LVI-----A-D-R--T--Q	--A----FI-----IS-	
Synechococcus sp. JA-2-3B'a(2-	86608830	-A-I-----R-----I	-----A--E--N--S	
Synechococcus sp. JA-3-3Ab	86605799	-A-I-----V-R-----I	-----A--E--N--S	
Trichodesmium erythraeum IMS10	113474636	-----V-----R	-----A--V--AM-	
Moorea producta 3L	332709543	--FI-----V-----M	-----A--V--A--	
Microcoleus vaginatus FGP-2	334121461	G-----V-----R	-----A--V--A--	
Oscillatoria sp. PCC 6506	300866592	G-----V-----R	-----A--V--A--	
Lyngbya sp. PCC 8106	119484310	-----V-R-----	-----A--V--A-G	
Arthrospira platensis str. Par	284052740	-----V-R-----R	-----A--V--A-G	
Arthrospira maxima CS-328	209527058	-----V-R-----R	-----A--V--A-G	
Microcoleus chthonoplastes PCC 7420	254412044	--I-----V-----T	-----A--C--A--	

Supplementary Figure 7

Partial sequence alignment for a pentapeptide repeat protein, showing a 7 aa insert that is specific for all *Nostocales* and *Stigonematales*.

			150	191
<i>Nostocales/ Stigonematales</i>	Nostoc sp. PCC 7120	17232408	KSTTDFTPLDQDGKVGLLNVSE D	NVSIYALDGQHRLMGVQGL
	Anabaena variabilis ATCC 29413	75908425	-----	-----
	Nostoc punctiforme PCC 73102	186684604	-----I---K-----I--	D-T-----
	Nodularia spumigena CCY9414	119510075	-----K-----I--	E--T-----S----
	Cylindrospermopsis raciborskii CS-505	282898769	-A-----AN-HI-----A-	E D-N-----
	Raphidiopsis brookii D9	282895759	-A----I-S-ANSHI-----A-	E DT---V-----
	Fischerella sp. JSC-11	354565831	---V-----E-N---SIA-	K D-T-----
Other Cyanobacteria	Microcoleus chthonoplastes PCC 7420	254412761	Q-AA--I---NQT---D--P	S---F-----
	Trichodesmium erythraeum IMS101	113475600	--A--FS--SQ-----DLRL	E-AVF-----I---
	Microcoleus vaginatus FGP-2	334118575	--A-E--SF-KNENL-----K	EL--F-----
	Moorea producta 3L	332711306	Q-AAE--S--K--T---D--D	H--F-----I-A-
	Oscillatoria sp. PCC 6506	300864427	--SA--LAI-KNET--F--IKD	--VF-----I---
	Arthrospira platensis Paraca	284054012	C-AAEYI---EH--I-I-DL-S	P--VF-----I--V
	Arthrospira sp. PCC 8005	376003672	C-VAQYI---EH--I-I-DL-S	S--VF-----I--V
	Arthrospira maxima CS-328	209526890	C-AAQYI---EH--I-I-DL-S	S--VF-----I--V
	Lyngbya sp. PCC 8106	119491185	E-AIA-EG--SQ-QL---HLTS	D-AVF-----I---
	Cyanothece sp. PCC 8802	257060278	RTVA-----QNRL----IG-	-F--F-----I---
	Cyanothece sp. PCC 8801	218247201	RTVA-----QNRL----IG-	-F--FV-----I---
	Cyanothece sp. CCY0110	126656075	V-VDE--S---NDS-----IGK	-Y--F-----I---
	Cyanothece sp. ATCC 51142	172035450	V-VDN--S--KNDSI----IG-	-Y--F-----I---
	Crocospaera watsonii WH 0003	357261885	V-VD--S--KNDSI----IGK	-Y--F-----I---
	Cyanothece sp. PCC 7822	307154544	IPAVN----PQQNL---D---	-F--F-----I---
	Cyanothece sp. PCC 7424	218441732	T-AVN----PQ-NL---DI-D	-Y--F-----I---
	Cyanothece sp. PCC 7822	307155131	ATAMN-LS--DQESF-R-DIP-	S-P-F-----IK--
Cyanothece sp. PCC 7424	218441726	SPAIN-LS--EQESF---ELPD	S-R-F-----IK-	

Supplementary Figure 8

Partial sequence alignment for hypothetical protein Npun_R4490, showing a 1 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		203	264
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227812	GALHLEMPSPWEALIAFTKSGGYTAFR F KQLAEIMSPTLILWGDTKILGTEDGKRFKRAIP
	Anabaena variabilis ATCC 29413	75910904	-----I-----A-R-----
	Nostoc punctiforme PCC 73102	186683027	-----Q---NQ-----S--- -KISQ-LQQ-----S-----K-AM-----
	Nodularia spumigena CCY9414	119511478	-----N-TQ-----S--K A ---S--VQ-----V---K-QQ---
	'Nostoc azollae' 0708	298492243	-D--VK--N-NQ-----S--K - N---Q-RQ-----G-AQK--K---
	Raphidiopsis brookii D9	282896104	RD--I--AN-K-S--T--Q---Q--K L Q--GK-GQ-----S-R---K--EK-RQ---
	Cylindrospermopsis raciborskii CS-505	282898677	RD--I--AN-K-S--T--Q---Q--K L EE-GK-GQ-----S-R---K--DK-RQ---
	Fischerella sp. JSC-11	354565585	A-----C-N-QQ----- -K-G--QQ-----S-R---K-A--N----
	Microcoleus vaginatus FGP-2	334119253	A-----N-NQ-----GG-G EK-SQ-QQ-----KQ-R---A-AEK-A--A
	Moorea producta 3L	332711810	A-----G-NQ-----GD-G DK--Q-SQ-----H-E---AAASK--H--A
	Microcoleus chthonoplastes PCC 7420	254410223	A-----G-HR---S-----PP-G QK-TQ-QQ-----KD-R---KYAHK-EA---
	Lyngbya sp. PCC 8106	119484283	A-----L-NR---S---N--GS-- -M-HK-QQQ-----EN-R---A-AQK-QQ--V
	Arthrospira platensis NIES-39	291570622	A-W--QS-L-NRG--E-----G--G DR-NT-QQ-----KN-----K-AHK-RT---
	Arthrospira platensis Paraca	284050533	A-W--QS-L-NRG--E-----G--G DR-NT-QQ-----KN-----K-AHK-RT---
	Arthrospira maxima CS-328	209523662	A-W--QS-L-NRG--E-----G--G DR-NT-QQ-----NN-----K-ADK-RT---
	Trichodesmium erythraeum IMS101	113476588	AR-----G-NK---S-----GS-K QK-HL-QQQ-----EN-N---VG-ADK-VK--A
	Cyanothece sp. PCC 7424	218440099	A---KCGQ-NQ---S-----GS-V A---QLIQ-----EN-Q---Q-A-L-STL--
	Cyanothece sp. PCC 7425	220907352	A---C-G-R---VY--Q---GS-A Q---HLQA-----RN-R---A-AA--EQI--
	Microcystis aeruginosa NIES-843	166368955	AS---NC-H-S---S-----GS-L P--SQ-DRE---I--EN-Q---K-A-K-QQ-L-
	Cyanothece sp. PCC 8802	257059639	A---QC-N-NK-----GS-K R--PQLKPE---I--QN-Q---K-ATV--EL--
Cyanothece sp. PCC 8801	218246596	A---QC-N-NK-----GS-K R--PQLNPE---I--QN-Q---K-ATV--EL--	
Microcystis aeruginosa PCC 7806	159030698	AN---NC-H-SA---S-----G--L QK-SQ-NRE---I--EN-Q---K-A-I-QQ-L-	
Crocospaera watsonii WH 0003	357263214	A---KCEQ-SQ---S-----GS-K QEIINLKQE-IVI--EN-----K-ANK--EL--	
Acaryochloris sp. CCME 5410	359462238	AS---D---HR--VG-----NFLY DTIKD-PH-----EA---M-VK-ADK-DQT-A	
Acaryochloris marina MBIC11017	158338347	AS---D---HR--VG-----NFLY DTIKD-PH-----KA---M-VK-ADK-DQT-A	
Cyanothece sp. PCC 7822	307151502	A---KCTG-SQ---S-----GC-S E--PK-KQ-----EN-Q---IK-AA--QA-LG	
Cyanothece sp. ATCC 51142	172037253	A---NCER-SQS--S-----GS-K EEMVN-KQE---I--EN-----K-ADK-NEL--	
Crocospaera watsonii WH 8501	67923067	A---KCEQ-SQ---S-----GSLK QEIINLKQE-IVI--EN-----K-ANK--EL--	
Synechococcus sp. PCC 7002	170077488	-M--VQCDR-Q---Q---G--GS-Y PK-KQ-QQ-----EQ-R---KAAH--QQGL-	
Other Cyanobacteria	Synechococcus elongatus PCC 7942	81301093	A---QQ-G-R-G-RR--R---GSM- SR-P-LRQ--QL--RQ-Q---K-AAV-Q-LL-
	Synechocystis sp. PCC 6803	16330122	A-A--TC-G-S-G---S-----GS-A E--GQ-TL-S--I--KQ-----VRAAE--QNLL-
	Cyanothece sp. CCY0110	126660002	A---NCER-SQG--S-----GS-K QE-VNLQHE-T-I--EN-----K-ADK--AL--
	Synechococcus sp. PCC 7335	254423048	AS---VQ--N-AKG--S-----NFLK DKITLVSQE--V---RQ-Q---K-AT--EQSLS
	Synechococcus elongatus PCC 6301	56751825	A---QQ-G-R-G-RR--R---GSM- SR-P-LRQ--QLV--RQ-Q---K-AAV-Q-LL-
	Synechococcus sp. RCC307	148241885	S---ST-G-AA--AC-AR--FAGVG APLPAA-IHV--AQ-R--RAPQ --AAE-LL
	Synechococcus sp. RS9916	116073047	AS-P-QV-G-GPS-A--AR--FA-CG APLPSQTLQV--NQ-R--RAPQ --AAQELL
	Synechococcus sp. CC9311	113953479	AS--KV-G-ARS-A--AR--FAGCG SPLPSQ-LHV--EQ-R--RAPQ --SAQELL
	Synechococcus sp. BL107	116072471	AS---QC-G-A---A--AR--FAGSG IPLPSQ-LHVI--AQ-R--R-PQ--AALE-L
	Synechococcus sp. WH 7805	88808915	AS---QC-G-S---A--AR--FSGSG HPLPQQ-MHV--NN-R--RAGQQALSALLE
	Synechococcus sp. WH 7803	148239873	AS---QC-G-S---A--AR--FSGSG HPLPQQ-LHV--NN-R--RAGQQALSALLK
	Synechococcus sp. CC9902	78184960	AS---QC-G-AQ--A--ARG--FAGSG SPLPSQ-LHVI--AQ-R--R-PQ--AALE-L
	Synechococcus sp. WH 8016	352093664	AS---KV-G-ARS-A--AR--FAGCG DPLPPO-LHV--EQ-R--RAPQ --SAQELL
	Synechococcus sp. CB0101	318040238	S---AC-N-AG--G--AR--FAGCG DPLPPO-LQV--QN-R--RPPQ --AAQ-LL
	Synechococcus sp. WH 8109	260435687	S---QC-G-A---A--AR--FAGCG TPLPSQSLHVI--K-R--RAPQ -QALQ-LL
	Synechococcus sp. CC9605	78213150	AS---QC-G-A---A--AR--FAGSG DPLPSQSLHVI--N-R--RAPQ -QALQ-LL
	Cyanobium sp. PCC 7001	254430710	AS---AT-G-G---RR-AR--FAGCG QPLPPL-IQV--AN-R--RPPQ --AAL-LL
	Synechococcus sp. WH 8102	33865474	S---QC-G-A---A--AR--FAGCG EPLPPO-LHVI--AD-R--RAPLKQAAESLLQ
	Prochlorococcus marinus MIT 9515	123966410	ASI--GCYG-RNS-AS-A---FAGSY KYMKNISIKTIC-EN-R---KEI-KI-Q -
	Prochlorococcus marinus CCMP1986	33861642	ASI--GCYG-RNS-AS-A---FAGTY KYMKNISIKTIC-EN-R---KKEINKI-Q -
	Prochlorococcus marinus MIT 9211	159903640	ASI--INV-G-SSS-A--AR--VANCG KPVPNQ-IKVI--K--NK-I -ESSMK-L
	Prochlorococcus marinus AS9601	123968725	ASI--GCKG-RNS-AS-A---FAGTQ KYIQNI-IKT-C-EN-R---KTEL-KIGN -
	Gloeobacter violaceus PCC 7421	37522627	A---A--G-EQ-I---R---APLG EK-PALSP-----ED-R--DPR-AHK-YK---

Supplementary Figure 9

Partial sequence alignment for the protein 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase, showing a 1 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		15	49
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282896270	EVQEI LN LA I ARQ SK DLSQEF SYQQ ILE IAKELQI
	Cylindrospermopsis raciborskii	282901647	-----
	'Nostoc azollae' 0708	298492349	--Q-Q-Q----- TD -GA----- I---A---
	Nostoc punctiforme PCC 73102	186682283	D-Q-Q-H----- AN -KNT----E-----A--E-
	Nodularia spumigena CCY9414	119513508	D-R-Q-Q----- AD -QDK-----L-----D-
	Anabaena variabilis ATCC 29413	75910215	D--R-Q-Q----- AD -QDK----E-L----T--E-
	Nostoc sp. PCC 7120	17228870	D-R-Q-Q----- AD -QDKD---E-L----T--E-
	Fischerella sp. JSC-11	354564846	DI-K-H----- A- -QEK----E-L----T--E-
	Moorea producta 3L	332711211	DI-Q-----L--- EMVE---RE-LV---S--G-
	Microcoleus chthonoplastes PCC 7420	254410564	-L-Q-----V-- ANGG---RT-LV---A--G-
Other Cyanobacteria	Trichodesmium erythraeum IMS101	113477975	D-Q-Q-Q-LVNR SEGG--TKV-L---Q-MGV
	Oscillatoria sp. PCC 6506	300864692	DA-Q-Q-Q----- R EETG-M-RT-LF-V-S--G-
	Lyngbya sp. PCC 8106	119488066	DA---KI-F-KK -ENG-LTRP-LM---I--G-
	Arthrospira maxima CS-328	209527444	DA---QI-M--G QETG-LTRT-LE-M-M--G-
	Arthrospira sp. PCC 8005	376003370	DA---QI-M--G QETG-LTRT-LE-M-I--G-
	Cyanothece sp. PCC 7822	307154767	DI-Q-H-L--R NDQE-L-RE-LW---A--E-
	Cyanothece sp. PCC 7424	218441995	DI-Q-Q----- TDKE-L-RE-LW---A--E-
	Synechococcus elongatus PCC 6301	56752424	D----QR----S TAQD---A--LQ-M-A--G-
	Cyanothece sp. PCC 7425	220906465	D-Q---I---HD TDKE---RT-L---A--G-
	Acaryochloris marina MBIC11017	158339009	Q-Q-----Q- -YEG---HA-L---E--A-
	Acaryochloris sp. CCME 5410	359461391	Q-Q-----Q- -YEG---HA-L---E--A-
	Cyanothece sp. PCC 8802	257061346	D----H-----K TDVE-L-RA-LW---A--D-
	Cyanothece sp. PCC 8801	218247319	D----H-----K TDVE-L-RA-LW---A--D-
	Synechococcus sp. PCC 7335	254421339	-A-Q-QI---KE TE-G-LTRL-LS---A--N-
	Crocospaera watsonii WH 8501	67921458	-----H-----K TEVE-L-RT-LW---A--D-
	Microcystis aeruginosa NIES-843	166364903	D----Y---S-- GDKG-ITR--L----DD-A-
	Microcystis aeruginosa PCC 7806	159030188	D----Y---S-- GDRG-ITR--L----DD-A-
	Cyanothece sp. ATCC 51142	172035381	-----H-----K TEVE-L-RT-LW---A--D-
Cyanothece sp. CCY0110	126660395	-----H-----K TEVE-L-RT-LW---A--D-	
Synechococcus sp. JA-2-3B'a(2-13)	86609907	D-Q-QR----- PRLG--TRS-LQ-M-A--G-	

Supplementary Figure 10

Partial sequence alignment for the hypothetical protein Aazo_3898, showing a 2 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		126	184
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii cs-505	282900401	EGTLKPANTKVLDTSCIIDGRIELLETGFLGEG
	Raphidiopsis brookii D9	282896777	-----
	Nostoc punctiforme PCC 73102	186681617	-----F-----SM--A-----
	Anabaena variabilis ATCC 29413	75908439	-----F-----M-A-----A-----
	Nostoc sp. PCC 7120	17232449	-----F-----M-A-----A-----
	'Nostoc azollae' 0708	298492588	-----M-----T-----
	Nodularia spumigena CCY9414	119512057	-----F-----M--H-----A-----
	Fischerella sp. JSC-11	354566808	-----M-----S-----T-----
	Oscillatoria sp. PCC 6506	300866347	-V-----F-----SLD-VL- A ---F---P-----N-----
	Trichodesmium erythraeum IMS101	113474964	-V-V-----F-----SLDAVL- A -----ST--I-----S--K-----
	Lyngbya sp. PCC 8106	119486128	-V-----AF---L---SL-SMLI A -----SS-I-----V-K-----
	Microcoleus vaginatus FGP-2	334119393	-VS-T----AF-----KSFV-VL- A ---F---SA-V-----N-----
	Arthrospira platensis Paraca	284050004	-V-----AF-Q---SL--TLL A -----S-H-----P-----
	Arthrospira platensis NIES-39	291571392	-V-----AF-Q---SL--TLL A -----S-H-----P-----
	Arthrospira maxima CS-328	209523296	-V-----AF-Q---SL--TLL A -----S-H-----P-----
	Moorea producta 3L	332709826	-VS-----TF-Q---SL-SVL- A -----S-SS-II---S-----A-----I--
	Microcoleus chthonoplastes PCC 7420	254414739	-IS-----TF-----SLD-ML- A -----TS-----E--N--V--
	Cyanothece sp. PCC 8801	218247389	-IS-----TF-----SI---L- A ---Q--S--I-----Q-V-----I--
	Cyanothece sp. PCC 7424	218441363	-IS--M---TF-----SI-SML- A Q---Q-S--II-----Q-D--I--
	Cyanothece sp. PCC 7822	307153453	-VS--M---TF-----SI-SML- A Q---Q-A--II-----Q-D--I--
	Cyanothece sp. ATCC 51142	172037003	-VS-----TF-----SI---L- A ---QA-S--V-----Q-H--I--
	Cyanothece sp. CCY0110	126657580	-VS-----TF-----SI---L- A ---QA-S--IV-----Q-H--I--
	Crocospaera watsonii WH 8501	67921336	-VS-----TF-----SI---L- A ---QA-S--V-----Q-H--I--
	Synechococcus sp. PCC 7335	254422819	-S--E-N--A-F---SY---LI S -----SKS-----S-----D--K--I--
	Cyanothece sp. PCC 7425	220907268	-TTV-----A---V---S---LL A -----AA-----E--N--M--
Microcystis aeruginosa PCC 7806	159030191	-VS-----TF-----SM-SIL- A ---QA-A--V-----Q-D--I--	
Microcystis aeruginosa NIES-843	166364900	-VS-----TF-----SM-SIL- A ---QA-A--V-----Q-D--I--	
Other Cyanobacteria	Synechocystis sp. PCC 6803	16330795	-IS-----TFM----SI-MLL A -----IPP-V-----P-A--I--
	Synechococcus elongatus PCC 7942	81301192	-IT-----T-I--V--GS--SALL A -----IS--I-----A-I-----
	Synechococcus elongatus PCC 6301	56751731	-IT-----T-I--V--GS--SALL A -----IS--I-----A-I-----
	Thermosynechococcus elongatus	22297552	-TT---S--PA-----A-ASSLL A ---M---RA-----A--NL-V--
	Synechococcus sp. PCC 7002	170079416	-VS---V---TF-----SLD--L- A ---R-SP--II-----L-E--K--I--
	Acaryochloris marina MBIC11017	158334463	-LT---I--QTI-----SI--MLL A ---Q-SP-----G--L---D--
	Synechococcus sp. CB0101	318040430	-Y---EV---T---F---ST-A-L- A D-V-M--SA-I---V---VRG--DS-L--
	Synechococcus sp. CB0205	317970579	-Y---EV---T---F---ST-A-L- A D-V-M--SA-I---V---VRG--DS-L--
	Synechococcus sp. JA-3-3Ab	86605617	-VS-----PA----F---Y ALQAA A --SVVA-SP-----S---VAQ-I-----
	Cyanobium sp. PCC 7001	254432606	-Y---EV---T---FS-G-T-A-L- A --V-R--SA-I---V---RG--DS-L--
	Synechococcus sp. WH 5701	87302967	-Y---EV---T---F---SST-A-L- A D-V-Q--TA-I---V---RG--S-L--
	Synechococcus sp. RCC307	148241696	-Y---EV---T---F---SST-A-L- A D-V-Q--SP-IV--L---LRG---SEI---
	Synechococcus sp. RS9917	87125530	-Y---EV---T---F---S-A-A-L- A D-V-T--SA-I---V---VRG--AC-L--
	Synechococcus sp. CC9902	78185114	-S---EV---T---F---AST-A-L- A D-V-T--TA-I---V---RGM-AC-L--
	Synechococcus sp. CC9311	113954728	-Y---EV---T---F---TST-A-L- A D-V-T--TP-I---V---RGM-AC-L--
	Synechococcus sp. WH 8109	260434968	-S---EV---T---F---AST-A-L- A D-V-T--TA-I---V---RGM-AC-L--
	Synechococcus sp. CC9605	78212386	-S---EV---T---F---AST-A-L- A D-V-T--TA-I---V---RGM-AC-L--
	Synechococcus sp. WH 7805	88809168	-Y---EV---T---F---TST-A-L- A D-V-T--TA-I---V---RGMIGC-L--
	Gloeobacter violaceus PCC 7421	37522655	--T-S-V---T---RF-L-SLNAGSE S GQFRGD-RP-IV-----L-E--G-----
	Prochlorococcus marinus NATL1A	124026438	-Y---V---TV---F---ST-S-L- A D-V-T--SA-----QA--RF-LI--
	Prochlorococcus marinus MIT 9211	159903901	-Y---EV---TF---L---ST-A-LI A --I-T--SG-I---V---QG--KC-LI--
	Prochlorococcus marinus CCMP1375	33240835	-Y---EI---TF---L---SA-A-L- A --I-T--SA-II---V---KS-IDF-LI--
	Prochlorococcus marinus AS9601	123969043	-YK-----T---F---NTDAYL- N --I-PA-SP-I---V---NG--SC-L---

Supplementary Figure 11

Partial sequence alignment for a PilT domain-containing protein, showing a 1 aa deletion that is specific for all *Nostocales* and *Stigonematales*.

		342	400
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899023	PEGLSGEIEIPLESRLALLMEFQREI NQQH KNPQSMEDMFAL ALSKCKQQEQTRFDPEL
	Raphidiopsis brookii D9	282897020	-----I----- --S-----I--- -F-----
	Nostoc punctiforme PCC 73102	186684031	-A--A-D-----AD--WRV --LK SSN--R-QI-TQ --DE---QSN----K-
	'Nostoc azollae' 0708	298490092	-A--A-D-----AAS--WQV SK-L YSET-EK-V-TQ --NE-R--QS----K-
	Nodularia spumigena CCY9414	119510882	-AS-T-D-----VAD--WRV -NIK SSQKTRQEI--Q --DE-R--QS----K-
	Nostoc sp. PCC 7120	17231796	-A--A-D-----VAD--WQV --KR AASKNRQEI-VQ --EE-R--A N----K-
	Anabaena variabilis ATCC 29413	75907478	-A--A-D-----VAD--WQV --KR TASENRQEI-VQ --EE-R--A N----K-
	Fischerella sp. JSC-11	354566134	-A--A-D-----IV--WRV H-KQ SSQL-R-EIINQ --EE-R--SR ----K-
	Microcoleus chthonoplastes PCC 7420	254415859	-D--AYDA-----MTT--QRL -DY--Q-TPL-K --ND-QASAG-V---K-
	Moorea producta 3L	332706650	-D--AYDN-----R-MAY--QQL NTHE-E--ALSI --RS-QALSGEA---K-
	Microcoleus vaginatus FGP-2	334120838	---RA-T--VA---G-VAD--ELT -PRG-RP--SLSE--EQ-RELSG----A-
	Trichodesmium erythraeum IMS101	113477705	-D--KE----I-----G-VSY--ELT IPRGKRP--SLGES-E--SARKG-Y---T-
	Arthrospira platensis Paraca	284051591	-D-QK--Q-NIQA---G-VSY--ELT -PRGSRP--TLTD--NE-QKYSG----A-
	Arthrospira maxima CS-328	209525585	-D-QK-KQ-NIKA---G-VSY--ELT -PRGNRP--TLKD--NE-EKYSG----AV
	Other Cyanobacteria	Lyngbya sp. PCC 8106	119486129
Arthrospira sp. PCC 8005		376005448	-D-QK-KQ-NIKA---G-VSY--ELT -PRGNRP--TLKD--NE-EKYSG----AV
Gloeobacter violaceus PCC 7421		37522761	---A--A---A---G-AAAY--EWL VARGERPVCTPSE--L-ETGAGG-W--Q-
Microcystis aeruginosa PCC 7806		159027367	-N--K---S-----G-SAY--ELT -PRG-RL--DLGE--ER-QNYSGI----A-
Microcystis aeruginosa NIES-843		166368416	-N--K---S-----G-SAY--ELT -PRG-RL--DLGE--ER-QNYSGI----A-
Cyanothece sp. PCC 7424		218437557	-D--K---SI-A---G-VAY--DLT PA-PRG-RYS- GE--D--LAYSQ----A-
Cyanothece sp. PCC 7822		307152109	-N--K---SI-A---G-VSY--ELT -PRG-RA-YNLTE--EQ-QNYSQ----A-
Cyanothece sp. CCY0110		126659318	-D--K--DNGI-A---G-VAY--ELT -PRG-RDS-SLGE--E--ESYKE-W---T-
Cyanothece sp. PCC 8801		218246774	-D--K---A--A---G-VSY--ELT -PRG-RA--GLGE--ES-E-YRGSW---T-
Cyanothece sp. PCC 8802		257059816	-D--K---A--A---G-VSY--ELT -PRG-RA--GLGE--ES-E-YRGSW---T-
Crocospaera watsonii WH 0003		357261605	-S--K---NGI-----G-VSY--ELT -PRG-RPP-SLGE--E--ESRRE-W---N-
Crocospaera watsonii WH 8501		67924474	-S--K---NGI-----G-VSY--ELT -PRG-RPP-SLGE--E--ESRRE-W---N-
Cyanothece sp. ATCC 51142		172036471	-N--K--NSI-A---G-VAY--ELT -PRG-RP--SLGE--E--QSYQGSW---T-

Supplementary Figure 12

Partial sequence alignment for a putative sensor protein, showing a 4 aa insert that is specific for all *Nostocales* and *Stigonematales*.

			127		205
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282896617	GTPVRSQFLNRK A NSPLELPIDEA	PLIIVVFGSQGAVGVNQLVLR	
	Cylindrospermopsis raciborskii CS-505	282899703	----- I G--D-----V	-----A-----	
	Nostoc sp. PCC 7120	17227973	-----LG N --Q-D-A--GDV	-V-----A-----	
	Nostoc punctiforme PCC 73102	186682723	----G--DGA I ---D-A--GV	-----A--K---	
	Anabaena variabilis ATCC 29413	75909100	-----LS N --Q-D-A--GGV	-V-----A-----	
	Nodularia spumigena CCY9414	119513711	-----DQG L -TS-D---NGV	-----A--K---	
	'Nostoc azollae' 0708	298492134	-----HPA V DNA-D-L--NGV	-----A--K---	
	Fischerella sp. JSC-11	354568482	S-----E--EPG V A-S-D-G--ENV	-V-----A--K---	
	Oscillatoria sp. PCC 6506	300864360	-----TPQ -D---AENV	---AM-----I----	
	Microcoleus vaginatus FGP-2	334119986	-----A-Q -N-----SV	-V-L-M-----A--K---	
	Microcoleus chthonoplastes PCC 7420	254410009	----AD-YKPQ -D-----DV	----V-----S-----	
	Arthrospira sp. PCC 8005	376002127	----AD-T-ES E--F----H-	--V-L-----A--R---	
	Arthrospira maxima CS-328	209526539	----AD-T-ES E--F----H-	--V-L-----A--R---	
	Arthrospira platensis Paraca	284050421	----AA-T-ES E--F----H-	--V-L-----A--R---	
	Moorea producta 3L	332711627	----DS-QQSQ S-----P-V	----V-----A--K---	
	Lyngbya sp. PCC 8106	119488862	VGTPVRS AFLAE TE--SFS--EN-	----V-----A-----	
	Trichodesmium erythraeum IMS101	113474108	----EE--FPQ S-D---ENV	-V-IV-----AI-----	
	Cyanothece sp. PCC 7822	307155003	S-----TPQ A-D-----	----V-----A--E---	
	Cyanothece sp. PCC 8802	257061896	S-----P-YTPQ -D---ENV	---A-I-----IA-----	
	Cyanothece sp. PCC 8801	218248834	S-----P-YTPQ -D---ENV	---A-I-----IA-----	
	Cyanothece sp. PCC 7424	218440437	S----A--TPQ -D--LEED-	F--V-----A-----	
	Synechococcus sp. PCC 7002	170079178	----AE--TPQ ---D--GDR	--LAM-----I-----	
	Cyanothece sp. ATCC 51142	172036199	S-----YTSQ -D-N--NDV	---II-----S-----	
	Cyanothece sp. CCY0110	126660036	S-----YTPQ -D-N--VNV	---T-I-----S-----	
	Crocospaera watsonii WH 8501	67920992	S-----YTPQ S-D-N--NV	---II-----S-----	
Cyanothece sp. PCC 7425	220906690	----D--HPP A-D--L-EAE	-V--M-----A-----		
Microcystis aeruginosa NIES-843	166364915	S-----IPQ S-D---ET-	F--A-----AL-----		
Microcystis aeruginosa PCC 7806	159029109	S-----ED S-D---ED-	F--A-----AL--I--		
Acaryochloris sp. CCME 5410	359463874	----PA--TPC E-D---P--	T---A-----S--K---		
Acaryochloris marina MBIC11017	158337764	----PA--TPC E-D---P--	T---A-----S--K---		
Synechocystis sp. PCC 6803	16332235	S---E--RQPQ S-D---PNR	S---A-----T--Q---		
Synechococcus sp. JA-3-3Ab	86605551	----AE-YQPQ PL-AD---EGD	----M-----R-L-R--		
Synechococcus sp. JA-2-3B'a(2-13)	86609599	----AE-YTPQ PF-AD---GD	----M-----R-L-RM-		
Other					
Cyanobacteria	cyanobacterium UCYN-A	284929013	---I-E--SISQ Y-D-N--EK-	F--I-I-----SI-N---	
	Synechococcus elongatus PCC 7942	81301121	---L---YQPS QHG---ENV	-VLL-M-----AI-R---	
	Synechococcus elongatus PCC 6301	56751799	---L---YQPS QHG---ENV	-VLL-M-----AI-R---	
	Thermosynechococcus elongatus BP-1	22299352	V-GTPVRPDILH PGN-----KAV	----M-----AI-R---	
	Synechococcus sp. CC9902	78185714	----G--AQ -L--WVPSGAG--L-L-M-----I-L-RM--		
	Synechococcus sp. BL107	116071994	----T--AAQ -L--WVPSGAG--L-L-M-----L-L-RM--		
	Synechococcus sp. WH 8109	260434743	----S--APQ -L-GWVPSGEG--L-L-M-----L-L-RM--		
	Synechococcus sp. CC9605	78213971	----S--TPQ -L-SWVPHGAG--L-L-M-----L-L-RM--		
	Synechococcus sp. WH 5701	87300803	----AE--QPT -L-SWVPAGDG--VL-M-----L-L-RM--		
	Synechococcus sp. RCC307	148243324	----ED--KPA AC--WVPAGDG--LL-I-----L-L-RM--		
	Synechococcus sp. RS9917	87123530	----RE--EAQ AL--W-PHGDG--L-L-M-----L-L-RM--		
	Synechococcus sp. WH 8016	352095579	----AA--QQH SL-EWVPQGDG--LL-I-----L-L-RMT-		
	Synechococcus sp. RS9916	116074045	----TA--APQ GL-TWVPQGPGL--L-I-----L-L-RM--		
	Cyanobium sp. PCC 7001	254431216	----PE--TEA -P--WLPPGEG--LL-M-----L-L-RM--		
	Synechococcus sp. CC9311	113954576	----AA--QQQ SL-TWVPQGSGL--L-I-----L-L-RMT-		
	Synechococcus sp. CB0101	318040696	----RE--EPA -L-AWVPAGSG--L-L-M-----L-L-RM--		
	Synechococcus sp. WH 7805	88809805	----S--QTQ TL--WVPQGPGL--L-M-----L-L-RM--		
	Synechococcus sp. WH 7803	148240697	----NS--QTQ TL--WVPQGPGL--L-M-----L-L-RM--		
	Synechococcus sp. WH 8102	33866863	-M--AD--QSQ -C-AWVPEGRG--L-L-I-----L-L-RM--		
	Synechococcus sp. CB0205	317968180	----RE--QAA -L-NWVPAGAG--LL-M-----L-L-RM--		
	Prochlorococcus marinus MIT 9303	124024479	----A--SPQ -L-HWVPCGDG--L-L-I-----L-L-RM--		
	Prochlorococcus marinus CCMP1375	33239675	-M--KS-SLKN -L-IWVPKGLE--L-M-----L-L-RM--		
	Prochlorococcus marinus NATL1A	124024986	----TE--FE- SL-SWVPLGEGV--L-M-----IKM-EM--		
	Prochlorococcus marinus AS9601	123967752	----L-E--YKFN FL-EWVPKGNG--LI-M-----KAI--IL		
	Gloeobacter violaceus PCC 7421	37521240	----PE--ISP N-LL-VPVPEA--L-M-----RAI--M-		

Supplementary Figure 13

Partial sequence alignment for the protein N-acetylglucosaminyltransferase (MurG), showing a 1 aa insert that is specific for all *Nostocales* and *Stigonematales*.

			177		237
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282901218	DRQRLEDALEKYATL VADIIRMGVNSESL	VRNYPPPGPVLETMNLGRSATPPAALQLLVDL	
	Raphidiopsis brookii D9	282895501	-----I- -----I-----	-----I--	
	'Nostoc azollae' 0708	298493171	-----E-----S- ----V-T----D--	A-A----A-----G--L--F--I--	
	Nodularia spumigena CCY9414	119509770	-----E-----A- L--VV-----YDT-	A-A----A-----P----G-F--I--	
	Nostoc sp. PCC 7120	17231942	-----EG---G-- L---VKV-L-YD--	A-A----A-----P---QG-F-----	
	Anabaena variabilis ATCC 29413	75907545	-----EG---G-- L---VKV-L-YD--	A-A----AL-----P---QG-F-----	
	Nostoc punctiforme PCC 73102	186685031	-----EG---A- L--TV-T---YD--	A-A----SA-----T---PT--QG-F-----	
	Fischerella sp. JSC-11	354564734	--H--E-----A- L---V-V-L-YD--	A-A----A-----QG-----	
	Microcoleus vaginatus FGP-2	334116976	--T-----RF---	GEEATHRT-A---LAS-E--EN-E--F-----	
	Oscillatoria sp. PCC 6506	300869627	--T-----RF-L-	GEEATHRT-A---LAS---E--E--F-----	
	Trichodesmium erythraeum IMS101	113477198	--I--II--F--H	GEESHRSQAM--LAT-KVPE-SE--FK-----	
	Arthrospira platensis NIES-39	291568149	--A-----RF-LY	GEEASHT--A---LSS-N-PQ--S-SF-----	
	Arthrospira maxima CS-328	209528143	--T-----RF-LY	GEEASHT--A---LSS-N-PQ--S-SF-----	
	Lyngbya sp. PCC 8106	119488090	--P-----RL-LN	GEEASHATSAF--LAT-E-PQ-ASG--E-----	
Other Cyanobacteria	Microcoleus chthonoplastes PCC 7420	254417643	--T-FEL--RFVLY	PENTSRAAQDILAT-D-PQ--Q--FD--I--	
	Moorea producta 3L	332712186	--T-F-----RFVSE	PDK-SRA-Q--LEA-K-PQN-EN-FD--IN-	
	Cyanothece sp. PCC 7822	307150006	--T--E---FILQ	PEQKHSQAQDIL-AVD--Q--E--E--IS-	
	Microcystis sp. T1-4	390439312	--I--ES---FILQ	PEQKYPAAAMDILSL---Q--E--FE-----	
	Microcystis aeruginosa NIES-843	166367429	--I--ES---FILQ	PEQKYPAAAMDILSL---Q--E--FE-----	
	Cyanothece sp. PCC 7424	218438216	--P--E---LVLQ	PEQKNQAQDILLSS---T--E--E--IE-	
	Cyanothece sp. CCY0110	126660660	--T---S---FVLH	PEQNHRAQDILLKE--QAT-AE--FE---A-	
	Cyanothece sp. PCC 8801	218246504	--N--ES--RLILQ	PEQTHRQAQ-LLSEI---Q--E--FE---E-	
	Cyanothece sp. PCC 8802	257059537	--N--ES--RLILQ	PEQTHRQAQ-LLSEI---Q--E--FE---E-	
	Cyanothece sp. ATCC 51142	172038259	--T--EL--FVLQ	PEQNHRAQDILLKE--QGT-AE--FD---A-	
	Cyanothece sp. PCC 7425	220906337	--L--ET--RF-VF	GEDSNQRVQAQ-LLASA--GS-SQD-FR----	
	Synechococcus sp. PCC 7002	170077199	----FES---FVLY	PEQSARAADHLLLE----P-TEQ-FR---S-	
	Acaryochloris marina MBIC11017	158337924	--TY-Q---F-A-	GDEATTRSTAT-VLST-K-GTS-D--F----V	
	Thermosynechococcus elongatus BP-1	22298194	--K--EQ--R--IA	IESSAQHQAAIDLLSE-K-PTQ-LG-F-----	
	Synechocystis sp. PCC 6803	16329795	--L--E---FILF	PEQNHQAQ-DILQT--KPGRTDETQN--IE-	
	Acaryochloris sp. CCMEE 5410	359463693	--TY-Q---F-A-	GDEATTRSTAT-VLST-K-GTS-D--F----V	
	Synechococcus elongatus PCC 6301	56750434	--PW-K---QW-LF	SEDGNNTRAAAT-VLTA---Q-SAQ--HD---A-	
	Synechococcus sp. PCC 7335	254423969	--P--EIV-R--L-	GDESSQKSAAQ-LLTE-NQDKS-E--FDF--A-	
	Crocospaera watsonii WH 8501	67924092	--S--ES---FVLQ	PEQSHRNAQDILLKE--HKTI-DS-FE---A-	
	Gloeobacter violaceus PCC 7421	37521185	--L-----R--LW	GDEASDRASAQDVLVLSQ--HTTRSED-IA--A-	

Supplementary Figure 14

Partial sequence alignment for the protein Ribonuclease II, showing a 14 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		150	188
Nostocales/ Stigonematales	Anabaena variabilis ATCC 29413	75909548	WYVRRFRQLFVNSDLGKTI AE SPLIQLISSFFNVNWT
	Nostoc sp. PCC 7120	17231973	-----Q-----
	Nodularia spumigena CCY9414	119510108	-----Q-----
	Nostoc punctiforme PCC 73102	186684639	-----T-----
	'Nostoc azollae' 0708	298489626	-----S-----
	Cylindrospermopsis raciborskii	282899262	--I-----LA-----A-S-----V-T-L-----
	Raphidiopsis brookii D9	282897934	--I-----LA-----A-S-----V-T-L-----
	Fischerella sp. JSC-11	354564838	-----A-S-----T-----I---S
	Oscillatoria sp. PCC 6506	300867643	--I-----V-AD-----SL--FV--VV-TVL--D-SG
	Microcoleus chthonoplastes PCC	254414832	--F-----I-TE-----A--F--VT-AIL--T--P
	Trichodesmium erythraeum IMS10	113474690	--I-----VILD---Y-A--FV--VSNAIL--D--G
	Arthrospira sp. PCC 8005	376003701	--I-----V-LD-----AL--FV--IA-TVLS-DISG
	Arthrospira maxima CS-328	209523873	--I-----V-LD-----AL--FV--IA-TVLS-DISG
	Arthrospira platensis Paraca	284053117	--I-----V-LD-----AL--FV--IA-TVLS-DISG
	Lyngbya sp. PCC 8106	119486305	-----V-LD--FW-A--FA--VS-AVLS-DLS-
	Moorea producta 3L	332710553	--S-----VMAD-----AL--F--VT-AIL--S--S
	Microcoleus vaginatus FGP-2	334121333	--I-----AL-D-----AL--F--VV-TVLT-DLSG
Cyanothece sp. PCC 7425	220908177	-----V-AD-----AL>A-F--IT-TVL---S-	
Acaryochloris sp. CCME 5410	359459621	--A-----VIAD-----AL--FV--VTAAV-T--LAG	
Acaryochloris marina MBIC11017	158339106	--A-----VIAD-----AL--FV--VTAAV-T--LAG	
Synechococcus sp. PCC 7002	170077619	--I-----EV-QA---RAV--F--VT-AIL--S--F	
Cyanothece sp. PCC 7822	307152899	--I-----VLTE--I--IL--F--IT-AIL--S---	
Cyanothece sp. PCC 7424	218438224	--A-----V-TD--I--VL--FV--ITAAI--MS---	
cyanobacterium UCYN-A	284929490	--LE---I-AE-----L>V-F-----TIL-TS--L	
Cyanothece sp. PCC 8802	257059735	--L---G-ILQE--I--AL--F--IT-AIL--S-S-	
Cyanothece sp. PCC 8801	218246693	--L---G-ILQE--I--AL--F--IT-AIL--S-S-	
Cyanothece sp. CCY0110	126656655	-----EVLTE--V---L>A-FV--IT-AIL--S-S-	
Crocospaera watsonii WH 0003	357263608	-----QKVLTE--V---L>A-FV--IT-AIL--S-S-	
Crocospaera watsonii WH 8501	67922712	-----QKVLTE--V---L>A-FV--IT-AIL--S-S-	
Cyanothece sp. ATCC 51142	172036145	-----EVLTE--V--AL>A-FV--IT-AIL--S-S-	
Microcystis aeruginosa PCC 980	389788061	--S-----V-SD-EI-R-L--FF--IAAAVL-FSFNP	
Microcystis sp. T1-4	390437842	--S-----V-SD-EI-R-L--FF--IAAAVL-FSFNP	
Microcystis aeruginosa NIES-843	166364942	--L-----V-SD-EI-R-L--FF--IAAAVL-FSFNP	
Synechocystis sp. PCC 6803	16329406	--W---QKVLAD-EVVQ-L--FV--VA-AVLS---SP	
Synechococcus sp. PCC 7335	254423462	--Y---K-V-QD-QFYRS--F-G-VTGAV-SSGLSG	
Synechococcus sp. JA-3-3Ab	86606917	--W--ASKA-LG--A--L>R-FVE--LR-VST-EFSS	
Cyanothece sp. PCC 7425	220908387	--F-----AME-L-IA-IA>AS-GG-IA-ALLSA-IDS	
Synechococcus sp. JA-2-3B'a(2-	86607941	--W--AS-A-LG--A--L>R-FAE--LR-VST-EFSS	
Gloeobacter violaceus PCC 7421	37520796	--T--LGEA-QK-AI-QVL--FLE-IARAILT--IST	

Supplementary Figure 15

Partial sequence alignment for the protein arsenite efflux ATP-binding protein (ArsA), showing a 2 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		268	308
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282901375	IPRLKIAIAENYQSLG PQNPS LLQPALNSYQEAYIIAWQLQ
	Raphidiopsis brookii D9	282895320	---I-----V ---- F-----
	'Nostoc azollae' 0708	298492846	--A-----Q-----A RE--N ---E-F-K-----VT---S-
	Anabaena variabilis ATCC 29413	75908794	L-T-QL---A--ET-A RE--A -IEA-FKN-----T-----
	Nostoc sp. PCC 7120	17232832	L-T-QL---A--ET-A RE--A --EA-FKN-----T-----
	Nostoc punctiforme PCC 73102	186683811	--E-L---A-E--A KKD-N --LE-FQN-----TT----R
	Nodularia spumigena CCY9414	119513457	--E-L--GA--E--A QE-S- --TQ-F-N-----MM-VRSR
	Fischerella sp. JSC-11	354566601	L-A-L---SD-E--A KE--N -I-E-F-N--Q--T-----
	Moorea producta 3L	332710739	--A-----SD-N--N QPEK-SQT----SL--S--
	Microcoleus vaginatus FGP-2	334118683	--PI-M---SD--TI- QINL-AQY-----NL-VPI-
	Oscillatoria sp. PCC 6506	300864500	V-P--ME--S--EV-- Q-NL-NQY-----AS-LVV-
	Microcoleus chthonoplastes PCC 7420	254409564	--A--G-GAD-EA-N QPEK-SQ-----AL--SS-
	Lyngbya sp. PCC 8106	119486370	L-D--LS-GLD-EK-- Q--Q-SQN-----TV-TTI-
	Trichodesmium erythraeum IMS101	113475114	VSEI--S-----EE-- R-NL-SQY-----S--QSI-
	Other Cyanobacteria	Arthrospira platensis NIES-39	291570022
Arthrospira platensis Paraca		284050571	-ADQ-----T-HEQ-- QF-Q-GEA--Q--QF-INI-
Arthrospira maxima CS-328		209527005	--D-----T-HEQ-- QF-Q-GQA--Q--QF-INI-
Arthrospira sp. PCC 8005		376005648	--D-----T-HEQ-- QF-Q-GQA--Q--QF-INI-
Microcystis aeruginosa NIES-843		166368640	--A--Q-GLD--A-K DANK-SQNF----SL-FA--
Microcystis aeruginosa PCC 7806		159030405	--D---Q-GLD--A-K DVNK-SQNF----SL-FA--
Cyanothece sp. CCY0110		126657841	V-P---D-GKD-ET-D QPEL-SQN-----AL--S--
Crocospaera watsonii WH 0003		357261708	V-P---E-GQD---D QPEL-SQN-----AL--S--
Crocospaera watsonii WH 8501		67924390	V-P---E-GQD---D QPEL-SQN-----AL--S--
Cyanothece sp. PCC 7822		307150085	--DI--L-GQD-DT-N QPEK-SQNF---FSL--A--
Cyanothece sp. PCC 7424		218439980	--DI-VL-GLD-DA-N QPEK-SQNF---FSL--S--
Cyanothece sp. ATCC 51472		354556658	V-P---D-GKD-ET-D QPEL-SQN-----AL--S--
Cyanothece sp. ATCC 51142		172037896	V-P---D-GKD-ET-D QPEL-SQN-----AL--S--
Cyanothece sp. PCC 8802		257060454	--A--S-GQD-EA-N QPEM-SKN-----SLS-S--
Cyanothece sp. PCC 8801		218248221	--A--S-GQD-EA-N QPEM-SKN-----SLS-S--
cyanobacterium UCYN-A		284929663	--S--VD-GQD--L-D QPEI-SKN-----TL--K-K
Synechocystis sp. PCC 6803		16329967	V-DIQ--GDH-LE-E QPEN-SQA--K--TL--SIK
Synechococcus sp. PCC 7335		254424809	L-N-LVV--Q---AIN QPNN-ITY-RS--VT-QK-G

Supplementary Figure 16

Partial sequence alignment for a TPR repeat protein, showing a 5 aa insert that is specific for all *Nostocales* and *Stigonematales*.

		17	78
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17231801	MGLSLLAYMLLATTFWMMWRMRNSRS IP SFMMFAGGRSQVTLHLIMGVSMVGLVLLLLAIG
	Anabaena variabilis ATCC 29413	75907483	-----V-----I---LT V- ---V-----S-----
	Nodularia spumigena CCY9414	119510885	-----L--V--I--F-E-TLPH F- --PW-I--REVRS--YT--IT--S-----V-
	'Nostoc azollae' 0708	298491592	-A--I---AF-LM--LCI-QV--TGR S- IGVVLPRVN-IVK-F-YTI-IT--S-----V--
	Raphidiopsis brookii D9	282897022	---AIFV-TI--I-AY-L-QV-TNGR S- VGVVVPKVNLLVK-F-YFL-ITLIF-----L--
	Cylindrospermopsis raciborskii CS-505	282899021	---AIFV-TI--I-AY-L-QV-TNGR S- LGLVAPKVNLLVKAF-YLL-ITLIF-----L--
	Nostoc punctiforme PCC 73102	186684034	-----LM--I--V--F-A-T-QQ F- -TP-M--NREVRS--YL--I--S-----
	Fischerella sp. JSC-11	354566137	-----V--V--LI-IR-L-A-TLWQ
	Synechococcus sp. PCC 7335	254422408	--F-IA--L--VS-AGLFYT-PDQR RLDQQSRRP-WLRLA--TT-IT--M-----S--
	Cyanothece sp. PCC 7822	307154981	-----T--TI-GIS---YNR-INKQ
	Cyanothece sp. PCC 7424	218438547	-----T--TV-GIS-I--YNR-TNKQ
	Crocospaera watsonii WH 0003	357264612	-----F--L--GMS-G--FYR--TQA
	Crocospaera watsonii WH 8501	67921777	-----F--L--GMS-G--FYR--TQA
	Other Cyanobacteria	Cyanothece sp. CCY0110	126655035
Microcystis aeruginosa PCC 7806		159030632	--FT-F--F--VS-VN-GYR-LAGQ
Microcystis aeruginosa PCC 9701		389881414	--FA-F--F--VS-VN-GYR-LAGQ
Cyanothece sp. ATCC 51142		172037641	-----F--V--GLS-G--FYG-HRQV
Cyanothece sp. PCC 8801		218246885	-----TC-AI-GLS-G-IFYR-QQGA
Cyanothece sp. PCC 8802		257059926	-----TC-AI-GLS-G-IFYR-QQGA
Microcoleus vaginatus FGP-2		334119252	--V-----L--V--T--FAT-STKQ
Arthrospira maxima CS-328		209523208	----IT--LI--M--G--FTCSQQNA
Arthrospira platensis NIES-39		291571296	----IT--LI--M--G--FIC-QQNA
Lyngbya sp. PCC 8106		119487872	--F--F--LM--V--S--FYT--VKK
Arthrospira platensis Paraca		284050994	----IT--LI-GM--G--FIC-QQNA
Moorea producta 3L		332706647	-----F--V--I--SIVFSQ-H-GR
Trichodesmium erythraeum IMS101		113474491	--F--FT-S--VS-G--FFA--NKQ
Microcoleus chthonoplastes PCC 7420		254415845	-----A--S--V--SV-FTK-R-QH
Gloeobacter violaceus PCC 7421	37523959	-----AV-A--AS-L-VLIR-PGGR	
			RES-WLRS--YKI-ICL-S-----L--
			PRPQWLRP--YLI-GI--F-----S--
			PRPKWLRP--YTI-AI--F-----S--
			FSPLWLRPA-Y-T-AI--A--I--V--
			FSPLWLRPA-Y-T-AI--A--I--V--
			SSPVWLRV--Y-I-SI--I-----V--
			SRPKWLRSF--LS-LI--A--I--G--
			SRPKWLRPF--LS-LI--A--I--G--
			SRPVWLRW--Y-I-SI--I-----V--
			NRPQWLRPF-YT--GI-----I--
			NRPQWLRPF-YT--GI-----I--
			PRPKWLRPF--AT-AG-AT--I--G--
			SRPLWLRP--YAI-MI--V-----V-
			SRPLWLRP--YAI-MI--I-----TV-
			PRPPGLR-A-YVI-GT--A-I-----
			SRPLWLRP--YAI-MI--I-----TV-
			PRPTWLRP--Y-I-WI--A--V--G--
			KRPKWLRPF-Y-I-GV--F--V--I--
			PKPKWLRP--Y-I-WV--S-----G--
			SGRALRR--GA--AL--TS--G--

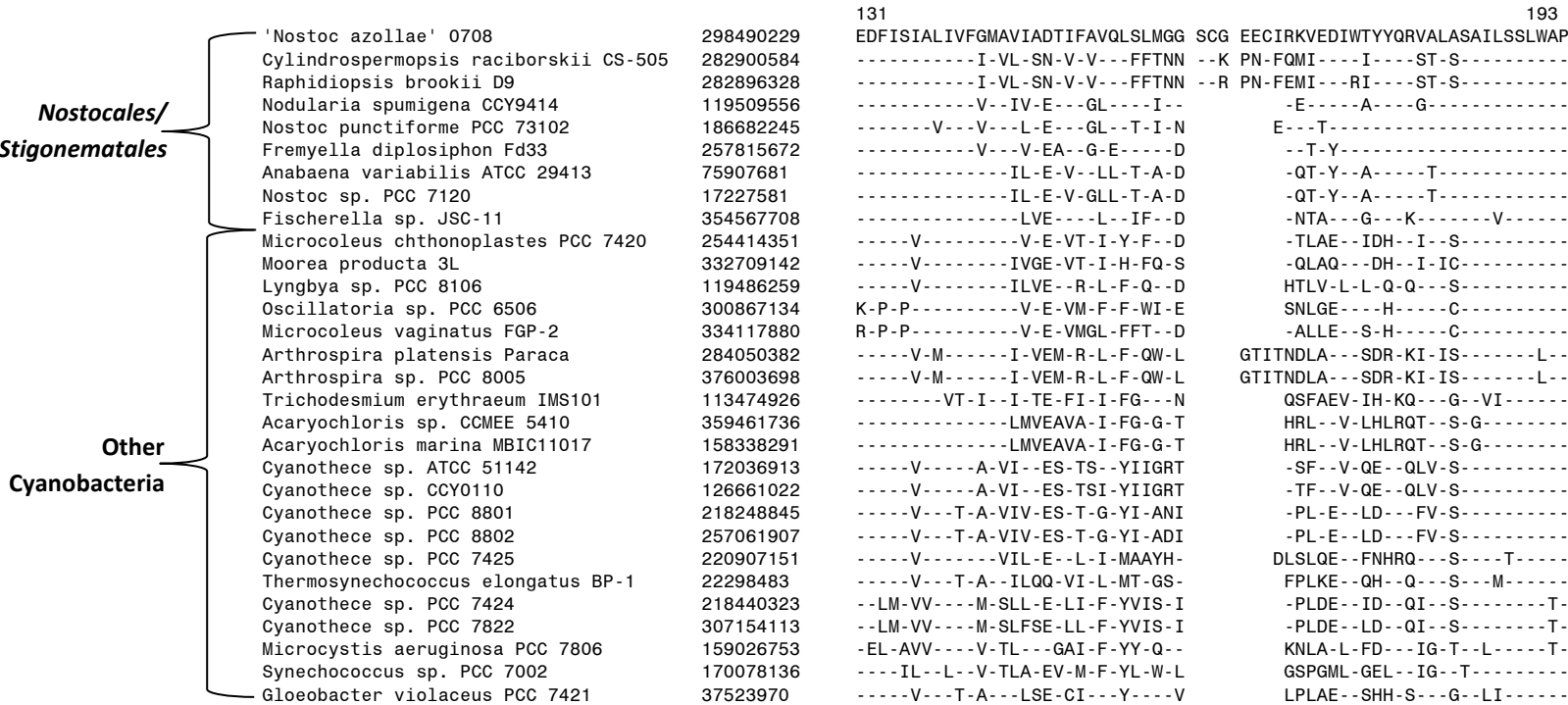
Supplementary Figure 17

Partial sequence alignment for the hypothetical protein Npun_F3904, showing a 2 aa insert that is specific for all *Nostocales*.

		148	204
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i> CS-505	282900218	LGVIINEVPPEELNTANDLLRPFL
	<i>Raphidiopsis brookii</i> D9	282897389	-----I-QMMN-----
	' <i>Nostoc azollae</i> ' 0708	298491926	-----G-----AL---KKQ-DDAC GRQRQ---
	<i>Nostoc</i> sp. PCC 7120	17232157	--S--S--I--SL--KRV-D--
	<i>Nostoc punctiforme</i> PCC 73102	186684840	--S--K-----AL-A-KGRV-D--
	<i>Anabaena variabilis</i> ATCC 29413	75908219	--S--S--I--SL--KRV-D--
	<i>Nodularia spumigena</i> CCY9414	119510802	--A--N---A--L---KQLV-D--
	<i>Fischerella</i> sp. JSC-11	354567400	--S-----SL--KQL-K--
	<i>Cyanothece</i> sp. PCC 7822	307154451	I-V--T-P-V--SL-K-KQEL-D--
	<i>Cyanothece</i> sp. PCC 7424	218440289	I-V--T-P-V--SL-KVKREL-D--
	<i>Synechococcus</i> sp. JA-3-3Ab	86605505	--IS--H-P-A-DSL-F-HRQL--H-
	<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	86608813	--IS--H-S-A-DSL-FVHRQL---
	<i>Cyanothece</i> sp. ATCC 51142	172035738	--A--H-P-V-DSLVAVKLAL-D--
	<i>Cyanothece</i> sp. PCC 7425	220905869	--IA--T-P-I--SL-L-REQL-H--
	<i>Synechococcus elongatus</i> PCC 6301	56750686	--AN-WSETG--DSL-A--QSL-D--
	<i>Crocospaera watsonii</i> WH 8501	67921047	--A--H-P-V-DSLEGVKLAL-D--
	<i>Cyanothece</i> sp. CCY0110	126656182	--A--H-P-V-DNL-NVKLAL-DS-
	<i>Synechococcus</i> sp. PCC 7335	254423693	--A--H-T-V-DKL-A-KKKL-DK-
	<i>Cyanothece</i> sp. PCC 8801	218247892	--IV--G--G--DSL-E-KKQL-D--
	<i>Cyanothece</i> sp. PCC 8802	257060787	--IV--G--G--DSL-E-KKQL-D--
	<i>Acaryochloris marina</i> MBIC11017	158337695	--S-FHN-AMIDAL-N-KA-L-N--
	<i>Synechocystis</i> sp. PCC 6803	16332002	--G--H-P-V-DGL-K-RQTL-DS-
	<i>Synechococcus</i> sp. WH 8102	33866351	--LW-DSC--DAL-A-KQTL-N--
	<i>Acaryochloris</i> sp. CCMEE 5410	359458351	--S-FHN-AMIDAL-N-KATL-N--
	<i>Synechococcus</i> sp. WH 8109	260436412	--LW-DSR--DAL--KQTL-D--
	<i>Synechococcus</i> sp. CC9605	78212198	--LW-DSR--DAL-A-KQ-L-D--
	<i>Synechococcus</i> sp. PCC 7002	170077367	--IP-HHK-L-DSL-TIYRFL-D--
	<i>Microcystis aeruginosa</i> NIES-843	166364185	--A--S-P-IADSL-K--QELNNQ-
	<i>Thermosynechococcus elongatus</i> BP-1	22299569	--V--E-P-V--SL-L-KSEL-D--
	<i>Synechococcus</i> sp. CC9902	78185278	--LW-DSR--PL-A-KQSL-DH-
	<i>Synechococcus</i> sp. BL107	116072842	--LW-DSR--PL-A-KQSL-EH-
	<i>Synechococcus</i> sp. RS9916	116073436	--LW-DSR--DAL-A-RDQL-E--
	<i>Synechococcus</i> sp. WH 7805	88807623	--LW-DSC--DAL-A-K-QL-D--
	<i>Synechococcus</i> sp. CC9311	113952774	--LW-DSR--DAL-A-HQL-D--
	<i>Synechococcus</i> sp. WH 7803	148240165	--LW-DSC--DAL-A-KAQL-E--
	<i>Synechococcus</i> sp. CB0205	317969193	--AWDDSS--PL-A-RDQL-QL-
	<i>Synechococcus</i> sp. WH 5701	87302846	--LWEDSR--DPL-A-RNVLQE--
	<i>Synechococcus</i> sp. WH 8016	352094892	--LW-DSR--DAL-A--HQL-D--
	<i>Synechococcus</i> sp. RS9917	87123777	--LL-DSR--AL-A-REQL-D--
	<i>Cyanobium</i> sp. PCC 7001	254432153	--QPWTDNR--DSL-A-RSEL-EH-
	<i>Synechococcus</i> sp. RCC307	148242759	--A-AWEDAG--PL-Q-KALL-DQ-
	<i>Microcoleus chthonoplastes</i> PCC 7420	254413093	--S-FH-V-L---L---KHQL-D--
	<i>Arthrospira platensis</i> NIES-39	291567706	--V--H-HSLLDEV--KERL-T--
	<i>Arthrospira platensis</i> Paraca	284052011	--V--H-HSLLDEV--KERL-T--
	<i>Moorea producta</i> 3L	332712366	--A-FH-V-LADAL-F-KQRL-ES-
<i>Oscillatoria</i> sp. PCC 6506	300868666	--V-PI-QF--DAL--KRR-L-D--	
<i>Lyngbya</i> sp. PCC 8106	119493958	--E-LE-SAAIDGF-V-KAQL-D--	
<i>Microcoleus vaginatus</i> FGP-2	334120294	--A-PID-Y IDDL--KQRL----	
<i>Trichodesmium erythraeum</i> IMS101	113476342	--T-PN-HSYIDSL--KQRL-NS-	
<i>Arthrospira</i> sp. PCC 8005	376007760	--V-CH-HSLLDEV--KERL-D--	
<i>Arthrospira maxima</i> CS-328	209524794	--V-CH-HSLLDEV--KERL-D--	
<i>Prochlorococcus marinus</i> MIT 9303	124022396	--LW-DSR--DAL-A-KEQL-DH-	
<i>Prochlorococcus marinus</i> AS9601	123968071	I-NLWEDCK--DSL-D-KKQL-DK-	
<i>Prochlorococcus marinus</i> NATL1A	124025246	--A-FW-DSR--AL-E-KNQL-DH-	
<i>Prochlorococcus marinus</i> CCMP1986	33861036	I-NLWEDSK--DAL-D-KKQL-DHF	
<i>Gloeobacter violaceus</i> PCC 7421	37519824	--C--EED-VIDQL-ILRSQ-L-N--	

Supplementary Figure 18

Partial sequence alignment for the hypothetical protein CRC_01609, showing a 7 aa insert that is specific for all members of a *Nostocales* subclade which includes *Cylindrospermopsis raciborskii* CS-505, *Raphidiopsis brookii* D9 and '*Nostoc azollae*' 0708.



Supplementary Figure 19

Partial sequence alignment for the rod shape-determining protein (MreD), showing a 3 aa insert that is specific for all members of a *Nostocales* subclade which includes *Cylindrospermopsis raciborskii* CS-505, *Raphidiopsis brookii* D9 and '*Nostoc azollae*' 0708.

			262		310
			DMSAVILTRLRRGKLPWVAD	G	HSHLHRRLLRVGLSHRLTVLFIYSLTLW
Nostocales	'Nostoc azollae' 0708	298492981	-TTQ-FIK--LA--N-LSTP	-	KD-----AW-F-Q-RAS-ILW----F
	Cylindrospermopsis raciborskii	282901407	-TTQ-FIK--LA--N-LSTP	-	KD-----AW-F-Q-RAG-ILW----F
	Raphidiopsis brookii D9	282897363	-----S-I-----		KR-----QA----W-----
	Anabaena variabilis ATCC 29413	75908298	-----S-I-Q-----		KR-----QA----W-----
	Nostoc sp. PCC 7120	17232299	-----D--H--S-FI--		QR-----QA--Q-----
	Lyngbya sp. PCC 8106	119492197	-----D--N--S-FI--		KR-----EA--Q-----
	Oscillatoria sp. PCC 6506	300869191	-----D--N--S-FI--		KR-----QA----FA-----
	Microcoleus vaginatus FGP-2	334117512	-----IE--H--S-FA--		NR-----KA--Q-----T----
	Arthrospira maxima CS-328	209523599	-----IE--H--S-FA--		NR-----KA--Q-----T----
	Arthrospira sp. PCC 8005	376004363	-----IE--H--S-FA--		NR-----KA--Q-----T----
	Arthrospira platensis Paraca	284051970	-----D--N--S-FI--		KR-----DA--Q----F-----
	Trichodesmium erythraeum IMS101	113476266	-----TL--CH-SS-FQ--		QR-----KA-----L--A-----
	Moorea producta 3L	332710310	--LV--IA-ISD--S-FY--		R-----NR--Q-----Y--T----
	Lyngbya sp. PCC 8106	119484905	--T--IS--K--S-F--		KR-----KA-I-----A----
	Microcystis aeruginosa NIES-843	166366620	--T--IS--K--S-F--		KR-----KA-I-----A----
	Microcystis aeruginosa PCC 7806	159027006	--V--IS-IAQ--S-FIG-		NR-----IQA-I-----A----
	Other Cyanobacteria	Crocospaera watsonii WH 8501	67921811	--T--I--SQ-RS-FK--	
Cyanothece sp. PCC 7424		218440404	--V--IS-ITQ--S-FIG-		NR-----IQA-I-----A----
Cyanothece sp. ATCC 51142		172036273	--V--IS-IAQ--S-FIG-		NR-----IQA-I-----A----
Crocospaera watsonii WH 0003		357264634	--I--IS-ISQ--S-FIG-		NR-----IQA-I-----A----
Cyanothece sp. CCY0110		126658334	--T--IS-ISQ-RS-FK--		KR-----KA-I-----A----
Cyanothece sp. PCC 7822		307151351	-----L--Q-LS-F--		KR-----NA-F--F--M--V-A--
Synechococcus sp. PCC 7335		254423341	--V--S-VIK--S-FK--		KG-----INA-I-----A----
Synechocystis sp. PCC 6803		16329559	--V--IS-ISQR-S-FI--		KR-----IQA-I-----T----
cyanobacterium UCYN-A		284929111	-----IS--SQ--S-FTG-		K-----MNA-I-----A----
Cyanothece sp. PCC 8801		218244968	-T-T--FS-----QS-FA--		KR-----KA--Q-V-----C----
Thermosynechococcus elongatus BP-1		22299671	-T--VVD--S-QS-FA--		KR-----KA--Q-AIA--V----
Cyanothece sp. PCC 7425		220907096	-T-T-V-N-----S-FS--		KG-----DA--Q-RS--L--AIVF-
Acaryochloris marina MBIC11017		158335098	--TT-VFS-ISS-SS-FKP-		KR-----AA-F--T-----G----
Synechococcus sp. PCC 7002		170079026	-----K--A--S-FHP-		KG-----KA--Q-HS--L--AM---
Synechococcus sp. JA-3-3Ab		86606658	--VT--D--S-QS-FK--		KR-----DA-F--R--VL--G--G---
Synechococcus elongatus PCC 6301		56751241	-TAS--Q--T-DS-FKP-		KR-----KA----SV-IV--A----
Gloeobacter violaceus PCC 7421		37521921	---T--K--A--S-FHP-		KG-----KA----HS--L--AM---
Synechococcus sp. JA-2-3B'a(2-13)	86610235	--AY--YR--KR-S-FI--		RE-I----MNS--TEIG--NL-GISQ-	
Prochlorococcus marinus AS9601	123968929				

Supplementary Figure 20

Partial sequence alignment for the family 4 glycosyl transferase, showing a 1 aa insert that is specific for all members of a *Nostocales* subclade which includes *Cylindrospermopsis raciborskii* CS-505, *Raphidiopsis brookii* D9 and '*Nostoc azollae*' 0708. A similar insertion was also identified in two members of Chloroflexi; *Roseiflexus castenholzii* DSM 1 (156740045) and *Roseiflexus sp. RS-1* (148658600).

			195		245
			PFDVNREGLVLGEGGAVFVLESQESIEQRQ	LTP	ENIYGEILGFGLTADAFH
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282896136	---I-----E-----	---	-----
	Cylindrospermopsis raciborskii	282901127	---	---	---
	'Nostoc azollae' 0708	298490359	---LH-----ADLAK---	TK-	-----
	Nostoc punctiforme PCC 73102	186683777	---LQ-----A-----A-LAK---		AKV-----N--Y-
	Anabaena variabilis ATCC 29413	75907461	---LQ-----A-----RQLA-K-K		AK---Q-----N--Y-
	Nostoc sp. PCC 7120	17231778	---LQ-----A-----RQLAKK-K		AK---Q-----IN--Y-
	Fischerella sp. JSC-11	354568940	---LH-----MLI---A--AK---		AKV--K-----T--Y-
	Moorea producta 3L	332707766	---KQ-Q---A---S-L---EDLARR-A		AP---KL-----S--Y-
	Lyngbya sp. PCC 8106	119486389	---RQ---FI-----I-----M--AKR-S		AR---QV---L---S--SS
	Microcoleus chthonoplastes PCC 7420	254411977	---QR-----A---L---A-LALS-G		VP---Q---M---SC--Y-
	Oscillatoria sp. PCC 6506	300864195	---SQ---F---A---L---A-LAKR-G		AK---QV-----N--CY
	Microcoleus vaginatus FGP-2	334119361	---E---F-----L---A-LARR-G		AK---R-S-L---N--C-
	Trichodesmium erythraeum IMS101	113474377	---RA---F---A---ILL---ANFAKK-S		AKV--Q-RS---N-GI-
	Arthrospira platensis Paraca	284050656	--SRR-D--A---A-LL---A-LAHS-H		GK---S---V-V--CY
	Arthrospira maxima CS-328	209525892	--SRR-D--A---A-LL---A-LAHR-H		GK---S---V-V--CY
Arthrospira sp. PCC 8005	376002150	--SRR-D--A---A-LL---A-LAHR-H		GK---S---V-V--CY	
Other Cyanobacteria	Cyanothece sp. PCC 7424	218437582	---KH-----EALAQG-G		AS---R-----SC--Y-
	Cyanothece sp. PCC 7822	307151580	---HK-----EANAA--G		AL---RV-----C--Y-
	Acaryochloris marina MBIC11017	158334259	---Q---F---A---IL---EESVAQK--		AK---QV---V---V--Y-
	Acaryochloris sp. CCMEE 5410	359462000	---Q---F---A---IL---EESVAQK--		AK---QV---V---V--H-
	Cyanothece sp. PCC 8801	218246032	---Q-----IL---A-MAYH-G		AS---Q-----C-GD-
	Cyanothece sp. PCC 8802	257059082	---Q-----IL---A-VAYH-G		AS---Q-----C-GD-
	Synechococcus sp. PCC 7002	170077623	---QR-----I-L--TL--AQ--N		AK-----WSF-C--H-
	Microcystis sp. T1-4	390438816	---IA-----IL---E-LALS-N		AP---Q---Y-F-C--D-
	Microcystis aeruginosa NIES-843	166367225	---RA-----IL---E-LALS-N		AA---Q---Y-F-C--D-
	Cyanothece sp. PCC 7425	220906675	---RQ-Q-----I---TA-LAQR-G		AAL--QV-----A--E-
	Microcystis aeruginosa PCC 9806	389791324	---RA-----IL---E-LALS-N		AP---Q---Y-F-C--D-
	Synechocystis sp. PCC 6803	16330682	---RQ-Q-----LL---TR-LAQK-K		AR-----W-FSC--L-
	Synechococcus sp. PCC 7335	254421719	---E---F-----LM---ENIAVA-G		AR---KVS-I-I-N-SH-
	Thermosynechococcus elongatus BP-1	22299414	---L--D-F---AGILI--EL-FALA-G		AR--A-MV-Y---C--Y-
	Synechococcus elongatus PCC 6302	56750993	---QG-D-F-M---AGIL---L-HAQA-G		AH-----V-Y-M-C--Y-
Prochlorococcus marinus MIT 9312	78780087	---AE-D-F-I---SGIL---TL-NAQK-D		AR-----I-Y-T-C--H-	
Prochlorococcus marinus AS9601	123969351	---AE-D-F-I---SGIL---TL-NAQK-N		AR-----V-Y-T-C--H-	

Supplementary Figure 21

Partial sequence alignment for the protein beta-ketoacyl synthase, showing a 3 aa insert that is specific for all members of a *Nostocales* subclade which includes *Cylindrospermopsis raciborskii* CS-505, *Raphidiopsis brookii* D9 and '*Nostoc azollae*' 0708. A similar insertion was also identified in four members of Fungi; *Candida albicans* SC5314 (68465653), *Lodderomyces elongisporus* NRRL (149241322), *Candida orthopsilosis* (380353104) and *Candida tenuis* ATCC 10573 (344230450).

			16	43
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282898069	SPGPVV PG	LEFGPFVIRWYGLLIASAVL
	Cylindrospermopsis raciborskii	282898913	-----	-----
	Nodularia spumigena CCY9414	119509993	-----I	V-I--IT-----T---
	Anabaena variabilis ATCC 29413	75908188	-----L	V-I--IT-----T---
	Nostoc sp. PCC 7120	17232191	-----L	V-I--IT-----T---
	'Nostoc azollae' 0708	298490254	-----IL	VKI--II-----T---
	Nostoc punctiforme PCC 73102	186684079	-----IL	VKL--LS-----T---
	Fischerella sp. JSC-11	354565943	-----IF	VKI--IT-----T---
	Crocospaera watsonii WH 8501	67921422	-----IL	F-A---IV-----T---
	Synechococcus sp. CC9902	78185302	-----EL	FQ----L-----A---
	Synechococcus sp. BL107	116072817	-----EL	FQ----L-----A---
	Cyanothece sp. ATCC 51142	172034971	-----IL	I-L--L-----F-----
	Cyanothece sp. CCY0110	126658441	-----IL	F-A--L-----F-----
	Synechococcus sp. RCC307	148241801	-----M-	F----LAV-----L---
	Other Cyanobacteria	Synechococcus elongatus PCC 6301	56750329	-----TL
Synechococcus elongatus PCC 7942		81300039	-----TL	V-I--LS-----T---
Synechococcus sp. WH 8102		33866375	-----L	FQ---LTL-----M---
Cyanothece sp. PCC 7822		307150514	-----I	F-I--IAV----F-----
Cyanothece sp. PCC 8801		218246018	-----IF	F-I--L-----SA---
Synechococcus sp. CC9311		113955493	-----L	FQ---ITL-----L---
Cyanothece sp. PCC 7424		218441250	-----F	F-I--IAV----F-----
Synechococcus sp. WH 7803		148240188	-----EL	FQL----L-----L---
Synechococcus sp. RS9917		87123754	-----EL	VQL----L-----L---
Synechococcus sp. WH 7805		88807599	-----EL	FQL----L-----L---
Synechococcus sp. WH 5701		87302098	-----LI	FQL----SL-----L---
Synechococcus sp. PCC 7002		170078171	-----IL	F-V-N-A-----I---
Synechococcus sp. CB0205		317970495	-----LL	FQL----SL-----I---
Synechococcus sp. PCC 7335		254421887	-----DI	VSI--LTL-----L---
Synechococcus sp. WH 8016		352094871	-----L	FQL--LTL-----L---
Cyanobium sp. PCC 7001		254431109	-----L-	FQL--LSL-----L---
Synechocystis sp. PCC 6803		16329453	-----M	FQI-G-AL----F-----I
Microcystis aeruginosa NIES-843		166368928	-----IL	W-I--ISV----F---M---
Synechococcus sp. JA-3-3Ab		86605781	-----I-	FSW---SL-----GL-TV
Arthrospira platensis Paraca		284054257	-----TI	--V--IS-----T---
Moorea producta 3L		332704986	-----II	V-L--LT-----T---
Arthrospira maxima CS-328		209527319	-----TI	--V--IS-----T---
Microcoleus vaginatus FGP-2		334120776	-----II	FQL--VA-----T---
Microcoleus chthonoplastes PCC 7420		254415315	-----IL	VKL--LT-----T---
Oscillatoria sp. PCC 6506		300866400	-----II	FSL--LS-----T---
Trichodesmium erythraeum IMS101	113475245	-----II	FQL--LI-----A-L-	
Lyngbya sp. PCC 8106	119485993	-----IL	IRL--IA----F-----	
Prochlorococcus marinus str. MIT 9303	124022372	-----EL	FQL--AL-----I---	
Prochlorococcus marinus str. NATL2A	72383628	-----EL	FQL--SL-----IS--	

Supplementary Figure 22

Partial sequence alignment for the prolipoprotein diacylglycerol transferase, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		97	145
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898786	EFYLAVVLDTVISRPVLLGCRY IDWQSPGENMQYVVVEQEFSPPFYARRL
	Raphidiopsis brookii D9	282895774	-----
	'Nostoc azollae' 0708	298490424	-----ALC-----KE SD ---E---K-----
	Nostoc punctiforme PCC 73102	186685262	-----AVC-----KE AD ---E-A--K-HH-----
	Nodularia spumigena CCY9414	119510291	-----AVC-----STE PD ---E-A--K-HH-----
	Nostoc sp. PCC 7120	17231406	-----AVC-----STE AD --LELADQK-----
	Anabaena variabilis ATCC 29413	75908004	-----AVC-----STE AD --LELADQK-----
	Fischerella sp. JSC-11	354569073	--F-----AVC--I---KE AD ---E-A-KK--H-----
	Trichodesmium erythraeum IMS101	113476345	-L-----PIVR-----SKQ GG --TETAIKS--K---N-----
	Microcoleus vaginatus FGP-2	334120290	-L---L--P-AR-----S-Q GG M-VEGSI-Q-Q-A-D-----
	Oscillatoria sp. PCC 6506	300868661	-L---L--P-AR-----SKQ GG -NVEAAIAQ--Q-I-D-D-----
	Lyngbya sp. PCC 8106	119493952	-----T--SSSR--M--STK GG VAVEEAISEV-Q---D---S-----
	Microcoleus chthonoplastes PCC 7420	254410665	-----I--YTLR-----SCQ GG -HVENVMAQ--R---ED-----
	Arthrospira sp. PCC 8005	376007764	-V---I--P-TR--L---SVQ GG VNVEATMFHI-K---D-----
	Arthrospira maxima CS-328	209524798	-V---I--P-TR--L---SVQ GG VNVEATMFHI-K---D-----
Other Cyanobacteria	Arthrospira platensis Paraca	284052711	-V---I--P-TR--L---SVQ GG VNVEATMFHI-K---D-----
	Cyanothece sp. CCY0110	126659214	--F-SIL--YQLQ---M-SAK GG --VETLLKHT-K--LD-G-----
	Cyanothece sp. ATCC 51142	172037272	--F-SIL--YQLQ---M-SAK GG --VETLLKH-HK-ILD-G-----
	Crocospaera watsonii WH 8501	67925570	--F-SI--YQLQ---M-SAK GG --VETLLKHT-K--LH-G-----
	Cyanothece sp. PCC 7822	307153608	--F-----YQLQ-----SVK GG M-VEALL-Y--R--I-E-----
	Cyanothece sp. PCC 7424	218437307	--F--I---YQLQ-----SPK GG MNIEALL-Y--R--I-D-----
	Synechococcus sp. PCC 7002	170077513	-IF-----YELQ-----SAH GG MEIDQLLA-LHQ--IKDF---Q--H-
	Microcystis aeruginosa NIES-843	166365727	-IF--IL--YHLQ-----ASQ GG M-VD-LLAT--K--I-ER---YLC-Q-
	Microcystis sp. T1-4	390442147	-IF--IL--YHLQ-----ASQ GG M-VD-LLAT--K--I-AR---YLC-Q-
	Cyanothece sp. PCC 8802	257058438	--F-C-LF-YQLQC-----SSQ GG -EVENLLSHIEK--LDG-----
	Cyanothece sp. PCC 8801	218245402	--F-C-LF-YQLQC-----SSQ GG -EVENLLSHIEK--LDG-----
	Synechocystis sp. PCC 6803	16329536	-LF--I---YQRQC--M-SSE GG --VETLL-Q--S-SLRTN---YL----

Supplementary Figure 23

Partial sequence alignment for the protein succinyl-CoA synthetase, showing a 2 aa deletion that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		273	325
		DVAAAKARYSLWIEGNPPRLIHP	IP GGSSEIITLRQLRHPLLVWQKHEQG
			-----Q-----
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i>	282900091	-I-T-----F-GA---FVNRO
	<i>Raphidiopsis brookii</i> D9	282896735	-M-T-R-----LKA---F-DRQ
	' <i>Nostoc azollae</i> ' 0708	298491014	-M-T-R-----LKA---F-DRQ
	<i>Nostoc</i> sp. PCC 7120	17135027	-L-V-----GS---F-NRQ
	<i>Anabaena variabilis</i> ATCC 29413	75908863	-L-T-R---F-LGA---F-NRE
	<i>Anabaena circinalis</i> 90	9715729	-L-T-RS---Y-LGA---F-QRQ
	<i>Nodularia spumigena</i> CCY9414	119512341	-L-T-----F-LQA---F-NRQ
	<i>Nostoc punctiforme</i> PCC 73102	186680652	-L-L-----Y-LKA---F-D-G
	<i>Fischerella</i> sp. JSC-11	354567326	-L-L-----Y-LQA---F-D-G
	<i>Arthrospira platensis</i> Paraca	284051189	-L-L-----Y-LQA---F-D-G
	<i>Arthrospira</i> sp. PCC 8005	376003750	-L-L-----Y-LKA---F-D-G
	<i>Arthrospira maxima</i> CS-328	209523920	-L-T-----L-A---F-DRN
	<i>Arthrospira platensis</i> NIES-39	291572107	-L-T-----LQA---F-D-N
	<i>Moorea producta</i> 3L	332712286	-L-T-R-----L-AT--QFTDLD
	Other Cyanobacteria	<i>Microcoleus chthonoplastes</i> PCC 7420	254412960
<i>Trichodesmium erythraeum</i> IMS101		113476819	-L-T-R---F-LG-A-A-F-EDT
<i>Cyanothece</i> sp. PCC 8801		218248506	-L-T-R---F-LGA---F-QDK
<i>Cyanothece</i> sp. PCC 8802		257061571	-L-T-RS---F-LG-A-A-F-EDK
<i>Cyanothece</i> sp. ATCC 51142		172036803	-L-V-RS---W-LK---F-E-H
<i>Crocospaera watsonii</i> WH 8501		67923383	-L-T-R-----L-----F-DG-
<i>Cyanothece</i> sp. CCY0110		126657853	-L-V-RS---W-LQA---F-DST
<i>Cyanothece</i> sp. PCC 7822		307152450	-L---RS---T-L---R-HF-KVH
<i>Microcystis aeruginosa</i> PCC 7806		159028850	-L-T-R-----L-----HF-DG-
<i>Cyanothece</i> sp. PCC 7424		218441797	-L-MTR-----LD-H--KF-A-D
<i>Cyanothece</i> sp. PCC 7425		220908146	-L-T-R---Y-L-A-R--FVNRA
<i>Microcystis aeruginosa</i> NIES-843		166365089	-L-T-R---Y-LGA-R-QFVNRA
<i>Synechococcus</i> sp. PCC 7002		170079175	-L---R---G--M---A-FVNQE
<i>Acaryochloris marina</i> MBIG11017		158336591	---V-R-----LQ---QFVSQT
<i>Acaryochloris</i> sp. CCMEE 5410		359461621	-L-V-R---G-HL-A-R--FTAS-
<i>Synechococcus</i> sp. PCC 7335		254425115	-L-V-R---G-HL-A-R--FTAS-
<i>Thermosynechococcus elongatus</i> BP-1		22300014	-L-T-RV---G-LGAH--QWLT-G
<i>Synechococcus elongatus</i> PCC 6301		56750922	-A-L-R---GA-LGAVR-D-EADP
<i>Synechococcus elongatus</i> PCC 7942		81299422	-L-L-RG--GR-LG-VE-Q-EAA-
<i>Synechocystis</i> sp. PCC 6803		16330262	-L-L-RG--GR-LG-VE-Q-EPAA
<i>Cyanobium</i> sp. PCC 7001		254432766	-LTL-R---GQ-LG-V---EADP
<i>Synechococcus</i> sp. WH 8109		260434383	---L-R---GQ-LGAVR-E-SADP
<i>Synechococcus</i> sp. CC9605		78212059	-A-L-R---GA-LGAVR-E-VADP
<i>Synechococcus</i> sp. RS9916		116073767	-LSLTR---GQ-LGAV--Q-EADP
<i>Synechococcus</i> sp. WH 5701		87301320	-L-L-R---A-LGAVR-E-EADP
<i>Synechococcus</i> sp. CB0101		318042364	-LGL-R---GA-LGAIK-K-GTE-
<i>Synechococcus</i> sp. RS9917		87125334	-L-L-RG--GRFLG-TA--MEASA
<i>Synechococcus</i> sp. CB0205		317968912	-L-L-RG--GQ-LGAV---DAQP
<i>Synechococcus</i> sp. RCC307		148241412	-L-L-RG--GR-LGAV--H-LEDP
<i>Synechococcus</i> sp. WH 8102		33866463	-L-L-RG--GQ-LGAV---EAEA
<i>Synechococcus</i> sp. WH 8016		352095971	-L-L-RG--GQ-LGAV---ESAV
<i>Synechococcus</i> sp. CC9902		78185392	-L-L-RG--GQ-LGAV---ESAV
<i>Synechococcus</i> sp. CC9311		113953514	-L---R---AS-LGAVR--FGERG
<i>Synechococcus</i> sp. WH 7805		88807280	EF-L-R---N-LG-VA-QIRDEE
<i>Synechococcus</i> sp. WH 7803		148240433	-L-L-RG--GQ-LG-V--A-HAEA
<i>Gloeobacter violaceus</i> PCC 7421		37523206	EF-LTRS---K--G-KT-TFEKNP
<i>Prochlorococcus marinus</i> CCMP1375		33239697	EH-LIRS---N--G--A-IVENNP
<i>Prochlorococcus marinus</i> MIT 9303		124022099	
<i>Prochlorococcus marinus</i> NATL1A		124025008	
<i>Prochlorococcus marinus</i> AS9601		123967774	
<i>Prochlorococcus marinus</i> CCMP1986		33860777	

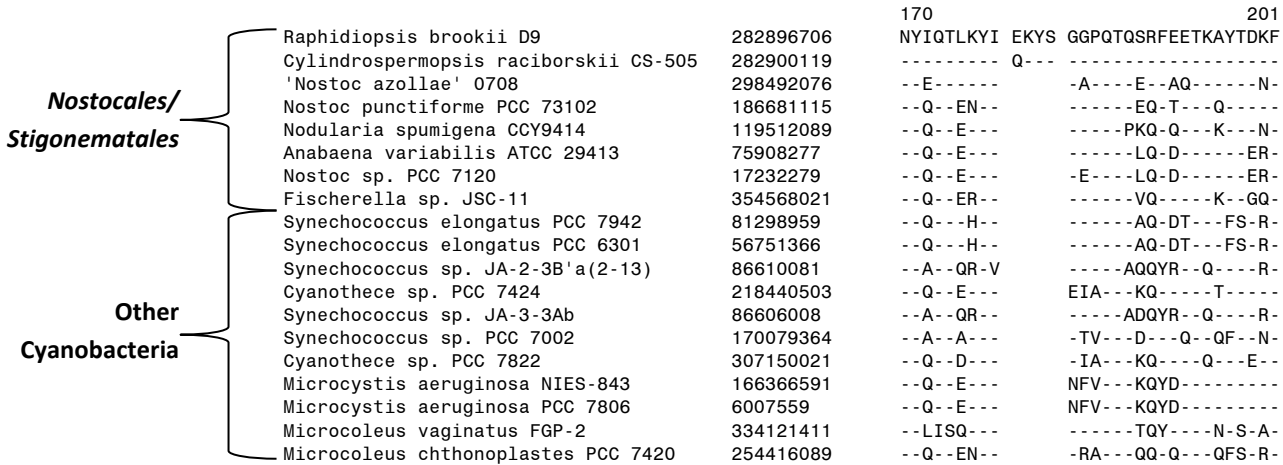
Supplementary Figure 24

Partial sequence alignment for the DNA mismatch repair protein, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		282899194	177		212
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899194	SGFTTPKVIATRSQAPHLIQ	K	WLG L P G G D P F Y A V I A
	Raphidiopsis brookii D9	282895956	-----AK-----	Q	----T-----
	Nostoc sp. PCC 7120	17232182	--S--EI--RK-T--NFL-		---A-----A--
	Anabaena variabilis ATCC 29413	75908198	--S--EI--RK-T--NFL-		---A-----A--
	Nodularia spumigena CCY9414	119508930	---V-E--VNK-T--NFL-		---AG-----
	Nostoc punctiforme PCC 73102	186683804	----AE---HK-TL-N-F-		---AA-----L-
	Fischerella sp. JSC-11	354565712	----E---RQ-NI-AFL-		---AG-----
	Lyngbya sp. PCC 8106	119488436	--SI-E---EQ-SV-SIL-		M---G-----
	Arthrospira maxima CS-328	209527362	--SV-E--INQ-SV-PVL-		M---G-----
	Oscillatoria sp. PCC 6506	300869331	--SQ-E--VKQ-NV-SFL-		M--MG-----
	Microcoleus chthonoplastes PCC 7420	254417036	--A--E-M-PS-PV-S-F-		M--VG-----
	Arthrospira platensis Paraca	284053987	--SV-E--IHQ-SI-PVL-		M---G-----
	Other Cyanobacteria	Trichodesmium erythraeum IMS101	113475706	--SQVE--SQK-NV-SFL-	
Moorea producta 3L		332708804	--SS-E-V-RP-PL-TFLL		Q--ITL-----L-
Microcoleus vaginatus FGP-2		334117918	--SK-E--VRKPNL-SFM-		M-RIAS-----V-
Synechococcus sp. PCC 7335		254424825	--GNIETVINK-E--AIL-		MM--G-----
Crocospaera watsonii WH 8501		67923500	---SK-E-VMNV-TL-SFL-		M--MM-A-----S
Cyanothece sp. ATCC 51142		172039040	---SK-E-VMNV-PL-SFL-		M--MM-A-----S
Cyanothece sp. ATCC 51472		354552675	---SK-E-VMNV-PL-SFL-		M--MM-A-----S
Crocospaera watsonii WH 0003		357262105	---SK-E-VMNV-PL-SFL-		M--MM-A-----S
Cyanothece sp. CCY0110		126659705	A--SK-E--MNV-PL-SFL-		I--MI-A-----S
Cyanothece sp. PCC 7822		307153322	---SQ-E--VVRQPSL-AFL-		MF--A-V-----
Cyanothece sp. PCC 8802		257060733	--SQ-E--VRS-PL-SFL-		M--IM-A-----
Cyanothece sp. PCC 8801		218247943	--SQ-E--VRS-PL-SFL-		M--IM-A-----
Synechococcus elongatus PCC 7942		81300139	A--SE-E---EV-P-LGWQ-		W L G L - Q - - - - - L -
Cyanothece sp. PCC 7425		220906445	---E-E---QPGLK-I-SW		F N S G - S - - - - -
Synechococcus sp. PCC 7002		170078393	--DK-T-VKQ--DQ-L-L-		M---GHR-----Y-
Acaryochloris marina MBIC11017		158334735	-LSE-Q-VSNQTPQST-LS		-M--G-S---F--VS
Acaryochloris sp. CCMEE 5410		359461165	-LSE-Q-VSNQTPQST-LS		-M--G-S---F--VS
Cyanothece sp. PCC 7424	218438896	---SK-E--VVRQPSI-SFLG		MF--ASV-----	

Supplementary Figure 25

Partial sequence alignment for the protein methyltransferase type 11, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.



Supplementary Figure 26

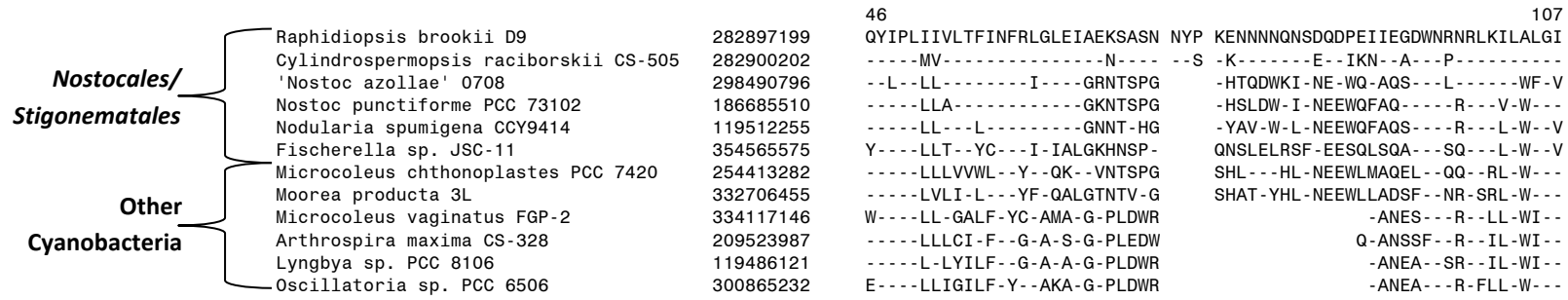
Partial sequence alignment for the prolipoprotein diacylglyceryl transferase, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

426

	492						
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899923	LFSTSIRNNISYGLDVLTEQEE	LGSAANLAKNRLGRFEQER	VESVANLVHIAAEINNFPQQYDTLVG		
	Raphidiopsis brookii D9	282896294	-----T-K---R---PIS---Q	-----N-----G--K-----	I---K-AK-ES--KS-----E---		
	Nostoc punctiforme PCC 73102	186684268	----T-Q---R---P-RK--N	---VT--R-AQME-----E-I---	--FA-K-AQ-D---S---H--E-I---		
	'Nostoc azollae' 0708	298491214	----T-K---R-S-P-SPP-N	---Y--K-AQ-HP--I-----E-I---	--Y--K-AQ-HP--I-----E-I---		
	Nodularia spumigena CCY9414	119510284	----T-K---R-S-P-SPP-N	---Y--K-AQ-HP--I-----E-I---	--Y--K-AQ-HP--I-----E-I---		
	Anabaena variabilis ATCC 29413	75908176	----T-K---R-S-P-SPP-N	---Y--K-AQ-HP--I-----E-I---	--Y--K-AQ-HP--I-----E-I---		
	Nostoc sp. PCC 7120	17232207	----T-K---R-S-P-SPP-N	---Y--K-AQ-HP--I-----E-I---	--Y--K-AQ-HP--I-----E-I---		
	Fischerella sp. JSC-11	354566660	----T-K---R-S-PLA--Q-	---LA-KTQ-HP--L-----E-I---	--Y--KQ-HP--L-----E-I---		
	Cyanothece sp. PCC 7424	218441387	---STV---R---PFK--P-	---Y--KQ-HP--L-----E-I---	--Y--KQ-HP--L-----E-I---		
	Cyanothece sp. PCC 7822	307153468	----T-E---R---PLK-RF-	---YA-KQSQ-YE--L-----E-I---	--Y--KQSQ-YE--L-----E-I---		
	Cyanothece sp. PCC 8801	218248700	----T-E---R---PLK-RF-	---YA-KQSQ-YE--L-----E-I---	--Y--KQSQ-YE--L-----E-I---		
	Cyanothece sp. PCC 8802	257061763	----L---R---EPLA--PQ	---NA-KQ-HP--L-----E-I---	--NA-KQ-HP--L-----E-I---		
	Cyanothece sp. PCC 7425	220910281	----L---R---EPLA--PQ	---NA-KQ-HP--L-----E-I---	--NA-KQ-HP--L-----E-I---		
	Cyanothece sp. ATCC 51142	172038090	----T-L---R---PLK-PF-	---YA-KQ-HP--L-----E-I---	--YA-KQ-HP--L-----E-I---		
	Cyanothece sp. CCY0110	126656899	----T-L---VR---PL--AF-	---NA-KQ-HP--L-----E-I---	--NA-KQ-HP--L-----E-I---		
	Synechococcus sp. PCC 7335	254424398	----T-Q---LR---PA-GTP-	---K-A-KEAQLHS--F-----E-I---	--K-A-KEAQLHS--F-----E-I---		
	Acaryochloris marina MBIC11017	158337512	----LK---R--NPLSDAS-	---Y--KQ-HP--L-----E-I---	--Y--KQ-HP--L-----E-I---		
	Acaryochloris sp. CCMEE 5410	359460776	----LK---R--NPLSDAS-	---Y--KQ-HP--L-----E-I---	--Y--KQ-HP--L-----E-I---		
Other							
Cyanobacteria	cyanobacterium UCYN-A	284929521	---N-T-L---R---PQKDIY-	---VNI-KKQ-HP--L-----E-I---	--VNI-KKQ-HP--L-----E-I---		
	Thermosynechococcus elongatus BP-1	22298197	----T-K---R--EPDATELK	---IAA-SQAQ-HK-VL-----E---	--IAA-SQAQ-HK-VL-----E---		
	Synechococcus sp. JA-2-3B'a(2-13)	86608710	---ATV-D--R--KPDA-DW-	---WA-KAA--H---L---KG----	--WA-KAA--H---L---KG----		
	Microcystis aeruginosa PCC 7806	159028055	----T-ED--R---PLLNFSA	---K-AVKQAR-EE--E---K-G----	--K-AVKQAR-EE--E---K-G----		
	Synechococcus sp. PCC 7002	170079375	---SQ--D--A--LPGSD-P-	I--S-QQAQ-HH--L---R--N----	--I--S-QQAQ-HH--L---R--N----		
	Crocospaera watsonii WH 8501	67923363	----T-L--LR---PLK--F-	T-IS-KQAE-HK--L-----E-V--	--T-IS-KQAE-HK--L-----E-V--		
	Crocospaera watsonii WH 0003	357262800	----T-L--LR---PLK--F-	T-IS-KQAE-HN--L-----E-V--	--T-IS-KQAE-HN--L-----E-V--		
	Synechococcus sp. JA-3-3Ab	86604918	---ATV-D--R--KPDA-DW-	---WA-KAA--H---LT--KG----	--WA-KAA--H---LT--KG----		
	Microcystis aeruginosa NIES-843	166364085	----T-ED--R---PLLNFSA	---A-KQAR-QE--E---K-G----	--A-KQAR-QE--E---K-G----		
	Synechocystis sp. PCC 6803	16331814	----T-AE--A-S-PSQTDTA	I-QA-HQA-LDQ--A---H-Q----	--I-QA-HQA-LDQ--A---H-Q----		
	Synechococcus elongatus PCC 6301	56750894	---LP--D--R--NPEVSEA-	IIAA-QQAQ-DR--L---RR-----	--IIAA-QQAQ-DR--L---RR-----		
	Synechococcus sp. CC9311	113953155	---TS-LAD-LR---PEAGMDR	---AA-DQARLL-DVKG--DGF----	--AA-DQARLL-DVKG--DGF----		
	Synechococcus sp. WH 8016	352094239	---TS-LAD-LR---PEAGMDR	---AA-DQARLL-DVKG--DGF----	--AA-DQARLL-DVKG--DGF----		
	Synechococcus sp. WH 7803	148239394	---TN-LAD-LR---EPQASMNR	---A-EQARLLSDVRG--DGM-----	--A-EQARLLSDVRG--DGM-----		
	Synechococcus sp. WH 7805	88808422	---TNTLAD-LR---PQASLDR	---A-EQARLLDGVKG--DGMN----	--A-EQARLLDGVKG--DGMN----		
	Microcoleus chthonoplastes PCC 7420	254412139	----LK---G--APGS--PQ	I-QA-KQAQ-HS-VL-----E-M---	--I-QA-KQAQ-HS-VL-----E-M---		
	Moorea producta 3L	332705098	----K---C--IPIS--HS-	---YA-SQSQ-HP--M-----E-I---	--YA-SQSQ-HP--M-----E-I---		
	Trichodesmium erythraeum IMS101	113477802	----T-K---C---PLK--K-	I-FA-KQAQ-HT--M--S---E-I---	--I-FA-KQAQ-HT--M--S---E-I---		
	Oscillatoria sp. PCC 6506	300869554	----T-K---R--NPLA--P-	I-EA-KQAQMS--L-----K-V---	--I-EA-KQAQMS--L-----K-V---		
	Microcoleus vaginatus FGP-2	334120026	---G-T-K---R--NPLA--PQ	I-AA-KQAQ-H---L-----K-V---	--I-AA-KQAQ-H---L-----K-V---		
	Lyngbya sp. PCC 8106	119488737	----T--E--R--NPLADAS-	I-TA-KTAQ-YG--L-----I---	--I-TA-KTAQ-YG--L-----I---		
	Arthrospira sp. PCC 8005	376006691	----T-A---C--NPSALRSQ	I-A--KIAQ-HD--T---KG---I---	--I-A--KIAQ-HD--T---KG---I---		
	Arthrospira maxima CS-328	209525287	----T-A---C--NPSALRSQ	I-A--KIAQ-HD--T---KG---I---	--I-A--KIAQ-HD--T---KG---I---		
	Arthrospira platensis Paraca	284049847	----T-A---C--NPSALRSQ	I-A--KIAQ-HD--A---KG---I---	--I-A--KIAQ-HD--A---KG---I---		
	Prochlorococcus marinus MIT 9312	78779229	---T---SE-LRF-EPNASKGT	---RKS-TKAGLIDD-----KFK-I---	--RKS-TKAGLIDD-----KFK-I---		
	Prochlorococcus marinus CCMP1375	33240192	---T-TLK---RF-VPDETIDK	---KL--SQA-LTED-KG--DGF-----	--KL--SQA-LTED-KG--DGF-----		

Supplementary Figure 27

Partial sequence alignment for an ABC transporter transmembrane region protein, showing a 20 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.



Supplementary Figure 28

Partial sequence alignment for the membrane bound O-acyl transferase, showing a 3 aa insert that is specific for *Cylandrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

			1007		1072
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898570	NWVVEANKFAPTLNVIQFGANT	RISDRTLSDKDLGDEETPVISS	RQKLLDQLQPLDMLVCSYGLLQ
	Raphidiopsis brookii D9	282898428	-----	-----M-P-----I-----	-----
	Nostoc sp. PCC 7120	17232392	---S-Q-----I---TGAN		-----G-----
	Nostoc punctiforme PCC 73102	186686080	---S-Q-----I---SGAN		-----G-----
	Nostoc punctiforme PCC 73102	186687055	---S-Q-----I---AGAN		-----G-----
	Nodularia spumigena CCY9414	119512991	---S-QR-----IV--AGAN		-----E-----
	Cyanothece sp. PCC 7425	220910355	--IS--QR-----L---SGD		-----H----L-----
	Microcystis aeruginosa PCC 7806	159028978	--IS--ER-----I--L-TGD		--A--R--F-L-----
	Cyanothece sp. CCY0110	126659998	--ID-CF-----P-L--SGN		--EI--N--F-L-I-----
	Cyanothece sp. CCY0110	126660775	--ID-CF-----P-L--SGN		--EI--N--F-L-I-----
Other Cyanobacteria	Crocospaera watsonii WH 0003	357266728	--ID-CF-----PLL--SGN		--EI--N--F-L-I-----
	Acaryochloris sp. CCME 5410	359457581	--IS--QR----RPL--TSQ		--E-----F-L-----
	Acaryochloris marina MBIC11017	158334343	--IS--QR----RPL--TSQ		--E-----F-L-----
	Cyanothece sp. CCY0110	126660084	--ID-CF-----P-L--SGN		--EI--N--F-L-I-----
	Acaryochloris marina MBIC11017	158341292	--HK--KRY----TPL--DNSN		--Q--EH-K-F--I-----
	Acaryochloris sp. CCME 5410	359462769	--NK--KRY----TPL--DNSN		--Q--EH-K-F--I-----
	Nitrosomonas sp. Is79A3	339482074	--QS-IQR-----QIL-GSD		--A-IEG-R-M-L-I---A---
	Pyramidobacter piscicolens W5455	282857036	--IE--AR-----DYRSAG		-E-V-AA-K-F-VVLA-----

Supplementary Figure 29

Partial sequence alignment for the protein SNF2-related helicase, showing a 22 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9. Sequence information for '*Nostoc azollae*' 0708 has been removed as it contains an insertion near the insertion of interest.

		21	66
	Raphidiopsis brookii D9	282896286	FGLVGIILLVIPNAELI FEEV LNLSEVGQTAMQWSMAGGGVAYMILG
	Cylindrospermopsis raciborskii CS-505	282901343	-----
	'Nostoc azollae' 0708	298489923	-S-----V-----
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227640	-----V---VL FH--F-----V----
	Anabaena variabilis ATCC 29413	75907735	-----V---VL FH--F-----V----
	Nostoc flagelliforme Sunitezuoqi	308205780	-----V---QV- F--K-----E--V-----
	Nodularia spumigena CCY9414	119512889	-----V---V- FHV--F----I----S---V----
	Nostoc punctiforme PCC 73102	186686043	----M-----DI- F--Q-E-I-L-----VDI----
	Fischerella sp. JSC-11	354567715	-----V-H--I- ---A-I-----L----V-----
	Synechocystis sp. PCC 6803	16330710	---A-L---L-HP-F- AH-PPI--V-F-----L---
	Microcystis aeruginosa PCC 7806	159027728	---A-L---L-HP-FV A--PDF-K--FA-----V--V--
	Microcystis aeruginosa NIES-843	166368829	---A-L---L-HP-FV A--PDF-K--FA-----V--V--
	Cyanothece sp. PCC 7822	307151824	---A-L-I-L--PD-- AK-P-I-K--FA--L-----V--L--
Cyanothece sp. PCC 7424	218440041	---A-L-V-L--P--V AQ-PDI-K--FS--L-----I--L--	
Cyanothece sp. ATCC 51472	354552086	---A-L---L--PDF- AS-PPI-K--FA-----M--L--	
Cyanothece sp. ATCC 51142	172039661	---A-L---L--PDF- AS-PPI-K--FA-----M--L--	
Cyanothece sp. CCY0110	126657355	---A-L---L--PDFV AS-PPI-K--FA-----M--L--	
Crocospaera watsonii WH 8501	67920230	---A-L---L-HPDF- AS-PPI-K--FA-----M--L--	
Cyanothece sp. PCC 7425	220907917	---A-L---L-HP-F- -S-PSQ--QLF-----V-IL--	
Other Cyanobacteria	Synechococcus sp. PCC 7002	3063601	---A-L---L---DF- AS-PDF--V-FRL---N---M--L--
	Synechococcus sp. PCC 7002	170078636	---A-L---L---DF- AS-PDF--V-FRL---N---M--L--
	Thermosynechococcus elongatus BP-1	22297957	---A-L---V-HP-F- -A-PVW--QLF----ES---V-IV--
	Synechococcus sp. JA-3-3Ab	86606197	-----L-W-V-HP-WV AS-PPI-LQ-FS---GN-----VS-
	Synechococcus sp. JA-2-3B'a(2-13)	86608031	---A-L-W-V-HP-WV AS-PPI-IQ-FSL--GN-----VF-
	Arthrospira platensis Paraca	284052466	---A-L-I-L--P--- I--PPI--SFF-L-----IVF-
	Arthrospira maxima CS-328	209526495	---A-L-I-L--P--- I--PPI--SFF-L-----IVF-
	Microcoleus chthonoplastes PCC 7420	254414983	---A-L---L-HP-W- MS-PPQ--LF-----V-I---
	Lyngbya sp. PCC 8106	119484671	---A-L-IIL--PQI- A--PSF--A-FG-----IVF-
	Microcoleus vaginatus FGP-2	334119541	---A-L--IL-RP--V MS-PPA--LF--G-----V-I-F-
	Oscillatoria sp. PCC 6506	300866876	---A-L-V---RP--- MS-PPA--VLF--G-----V-I-F-

Supplementary Figure 30

Partial sequence alignment for the hypothetical protein Aazo_0426, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		79	140	
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282898291	GLACLGLVLAQLFSSQNPF	LVLVAPLSGIFYW
	Cylindrospermopsis raciborskii CS-505	282901472	-----	-----
	'Nostoc azollae' 0708	298491545	--T--S--S-----S-YL	LA ---LS-----
	Nostoc punctiforme PCC 73102	186684156	-IV--S--VYL--PNLS-IV	LA M--LS----L---
	Nostoc sp. PCC 7120	17231472	--V--S--F-L--PRFGNI-	LG ---F-----L---
	Anabaena variabilis ATCC 29413	75907942	--V--S--F-L--PRFGNI-	LG ---F-----L---
	Nodularia spumigena CCY9414	119510140	-ML--A--WL--PNFG-L	LG ---LS----L---
	Fischerella sp. JSC-11	354568911	-IL--S--S--PNLSS--	IG M--LS-----
	Oscillatoria maxima PCC 6506	300863785	-ML-----SM-VPKGGI-	LL ---S--A-----
	Microcoleus vaginatus FGP-2	334121187	-AL-----GM-VPQ-GQI-	LA ---LS--A-V---
	Moorea producta 3L	332705415	-IV--S--SI--PTLTKV-	LG --AL--AA-L---
	Lyngbya sp. PCC 8106	119487841	--L--A--SM-VQDNSEW-	LG --ILS--T-L---
	Arthrospira maxima CS-328	209525214	-SL-----SMAIPASGYG-	LA ---FS--A-F---
	Arthrospira platensis Paraca	284052178	-SL-----SMAIPASGYG-	LA ---LS--A-F---
	Trichodesmium erythraeum IMS101	113477358	-AL-F--AISM-VPESKYL	LW ---S--A-----
	Microcoleus chthonoplastes PCC	254413039	-IL--S--SV-YPDLSQV-	LG --FL--VA-----
	Cyanothece sp. PCC 8801	218247821	--L--FAF--SF-YQPLT-L-	FA TI-LC--V-----
	Microcystis sp. T1-4	390439979	--F-FS--SV--PDT--YI	FA --GLS--A-V---
	Cyanothece sp. PCC 8802	257060859	--L--FAF--SF-YQPLT-L-	FA TI-LC--V-----
	Microcystis aeruginosa NIES-84	166363593	--F-FS--SL--PDT--YI	FA --GLS--A-V---
Synechococcus sp. PCC 7002	170078561	--V-IA---L-YPNIGFA-	LG IC-L---A-L---	
cyanobacterium UCYN-A	284929410	--L---I-SF-YQPLT-M-	LS -GILS-IT-L---	
Crocospaera watsonii WH 8501	67923215	-FF--I--SF-YPALT-L-	FA -E-LS-IA-----	
Synechococcus elongatus PCC 6301	56750863	--G----V-IATG-SPQLA	LI PATL--IA-RL--	
Synechococcus elongatus PCC 7942	81299486	--G----V-IATG-SPQLA	LI PATL--IA-RL--	
Synechococcus sp. JA-3-3Ab	86606676	-----A--QM-QPAWGAWP	WT -L-AS-WA-WY---	
Cyanothece sp. PCC 7424	218441139	--L--A--SL-YAPLTGL-	LC -I-L--VA-V---	
Cyanothece sp. ATCC 51142	172038242	--L--V--SFVYPLT-L-	FA -E-F--VA-----	
Cyanothece sp. ATCC 51472	354554405	--L--V--SFVYPLT-L-	FA -E-F--VA-----	
Acaryochloris marina MBIC11017	158335265	ATL--S--FSI--P-VGFG-	TV ITVIS-VA-W---	
Acaryochloris sp. CCMEE 5410	359462081	ATL--S--FSI--P-VGFG-	TV ITVIS-VA-W---	
Cyanothece sp. CCY0110	126655473	--L--V--SFVYPLT-L-	FA -E-LS-VA-----	
Cyanothece sp. PCC 7822	307152558	--LS-S--SL-YPPLTGL-	LC -TFL--A-----	
Synechococcus sp. WH 8016	352096551	-AG-----RE-NP-LSWWP	LS -I-LG--A-WV-S	
Synechococcus sp. WH 8102	33866570	-GT-F----R--APQLSWWP	LL -I-LG-AA-SV-S	
Cyanothece sp. PCC 7425	220907204	--F-----SI-LPQWGGGC	LA V--LT-IAP----	
Synechococcus sp. RCC307	148243180	-A-S-----R--LPQLGFWP	LL -A-LG-MA-LL-R	
Synechococcus sp. CC9311	113954006	-AG-----RE-NP-LGWWP	VS ---LG--A-WV-S	
Synechococcus sp. WH 7803	148238799	-AGS----IR-VNAGLGWWP	LL IAGLG--A-WV-T	
Synechococcus sp. WH 8109	260435503	-S--F----R--APQLHWWP	LL -IGLG--A-AV-T	
Synechococcus sp. CB0205	317968784	-A-----C--LPALAWWP	LL -T-LG--A-VV-R	
Synechococcus sp. CB0101	318042595	-AG-----R--IPPLHWWP	LL -IGLG--A-VL-R	
Synechococcus sp. CC9605	78211957	-S--F----R--APQLHWWP	LL -IGLG--A-AV-T	
Prochlorococcus marinus CCMP1375	33241188	-G-----IC--YA-L-WWP	LV ---LG-FA-F--R	
Prochlorococcus marinus MIT 9211	159904245	-GG-----IC--PVIAWWP	LV --GLG--A-L--R	

Supplementary Figure 31

Partial sequence alignment for the hypothetical protein Aazo_2731, showing a 2 aa deletion that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

**Nostocales/
Stigonematales**

Raphidiopsis brookii D9	282897472
Cylindrospermopsis raciborskii CS-505	282901303
'Nostoc azollae' 0708	298489622
Anabaena variabilis ATCC 29413	75908513
Nostoc sp. PCC 7120	17232529
Nostoc punctiforme PCC 73102	186686130
Nodularia spumigena CCY9414	119509796
Fischerella sp. JSC-11	354569217
Lynngbya sp. PCC 8106	119489333
Arthrospira platensis Paraca	284053725
Arthrospira sp. PCC 8005	376001709
Arthrospira maxima CS-328	209523121
Trichodesmium erythraeum IMS101	113478005
Microcoleus chthonoplastes PCC 7420	254416854
Moorea producta 3L	332712006
Microcoleus vaginatus FGP-2	334118565
Oscillatoria sp. PCC 6506	300866095
Cyanothece sp. PCC 8801	218247517
Cyanothece sp. PCC 7822	307151756
Cyanothece sp. PCC 8802	257061147
Crocospaera watsonii WH 8501	67923409
Crocospaera watsonii WH 0003	357262751
Synechocystis sp. PCC 6803	16331276
Cyanothece sp. PCC 7424	218438329
Cyanothece sp. PCC 7425	220909022
Cyanothece sp. CCY0110	126657224
Microcystis aeruginosa PCC 7806	159028625
Microcystis aeruginosa NIES-843	166364264
Acaryochloris marina MBIC11017	158338721
Cyanothece sp. ATCC 51142	172038189
Synechococcus sp. PCC 7002	170079005
Acaryochloris sp. CCME 5410	359458502
Synechococcus elongatus PCC 6301	56750742
Synechococcus sp. JA-2-3B'a(2-13)	86609848
Synechococcus sp. JA-3-3Ab	86606883
Cyanobium sp. PCC 7001	254430249
Synechococcus sp. WH 8016	352093529
Synechococcus sp. WH 7805	88809046
Synechococcus sp. WH 7803	148239999
Synechococcus sp. BL107	116071904
Synechococcus sp. CC9605	78214071
Synechococcus sp. RCC307	148243398
Synechococcus sp. CC9902	78185788
Synechococcus sp. RS9917	87125450
Synechococcus sp. WH 8109	260435064
Synechococcus sp. CC9311	113953607
Synechococcus sp. RS9916	116073205
Synechococcus sp. WH 5701	87302528
Synechococcus sp. WH 8102	33866932
Prochlorococcus marinus NATL1A	124025793
Prochlorococcus marinus MIT 9303	124022299
Prochlorococcus marinus CCMP1375	33240637
Prochlorococcus marinus AS9601	123968621
Gloeobacter violaceus PCC 7421	37521903

**Other
Cyanobacteria**

SPQAAKVVVEKLGYYHQVKSLLHGESFT	LEDTLN	LKDQLEIKALPGSPVGPVNVRENGY
-----Y---T-----	-----ST	-----
--N-----KNLIT-N---T--		--N-V----F----I--T-L----
--N----QA---KS-TT-A-----		FNN-I--R-F----I--T-V--S-
--N----QA---TS-TT-V-----		FNN-I--R-F----I--T-V--S-
--N----QQ---T--TT-A---T--		--N-V----F----I--TLV--S-
--N----LLPG---TS-T--A--AT-N		--NQ-V--T-V---T--Y-LV----
--N----QQ---T-ITV-T---T--		--NTSV----F----I--TLV----
--S----AQE---FT--RA-E--STYF		--DN-V----V----I--TLV----
--N----RE---T-IH---R-T-K		--GSS---R-V----I--TLV----
--N----RE---T-H--N---T-Q		--GASI--R-V---I--TLV----
--N----RE---T-H--N---T-K		--GAS--R-V--F-I--TLV----
--A----AQE---S--IT-E--HTYA		FENSI--R-V--L--TLV----
-EN--T--KE---T--TP-AF-QT--		-A--I--R---A-I--FLQ--A-
-TS----KQ---T--TP-TP--TVA		-GNH--R---A-I--FQQ--A-
--T---LSS---TNIT--ANWQELN		F-EK-Q-T-V--AEI--GQE----
--T-T--LSS---KSIA--SNWQE-I		Q-K-Q-I-V--AEIQ-GQV----
--N----KA---T--T--S---Y-		--DNAIS---V--L--QLI----
--N----KA---T--TT-D-HQ-Y-		--ANS----V----I--QLI----
--T----KG---T--T--S-W--Y-		--DNAIS---V--L--QLI----
--N----E---TDIHT-S---TY-		--EET-A-----LI--QLV--A-
--N----E---TDIHT-S---TY-		--EET-A-----LI--QLV--A-
--K--E-ATE---ET-TG-P-NQK-V		--N-RV--L-----I--TLV--A-
--N-E---KG---T-IT--A-HQTY-		FANR-Q---V----I--QLV----
--A----S---S-II--Q--Q-Y-		--NH--T-Q-TV---L--Q-V----
--N-T---KE---TDIRT-E--D-Y-		--DETID---F---L--QLV--A-
--N-E---RA-N-THLIP-T--S-YI		FDNAI--T-V---I--TLV----
--N-E---RA-N-THRIP-N--S-YI		FDNAI--T-V---I--TLV----
-EN-T---E---TD-TT-A---T--		WA-R-QVQ-----I--FLV--AL
--N-T---KE---TDIRT-D--D-Y-		--DETVEF---F---L--QLV--A-
--N----S---QTIHT-E-HQKY-		I-N---L--K-----TTL----
-EN-T---E---TD-TP-A---T-A		WA-R-QVQ-----I--FLV--AL
--A---AR---FS--TV-SP---V-		IGN--LQ-TK-A--QV----
--A--R--KS--FAT-QA-DPWQ-VR		WGD-RVT-I--A-L--L--L--
--T--R-ARS--FMR-QA-DPWQ-VQ		WGD-G-T-V--A-L--L--L-F
--T--RR-RQ--FQH-TA-AP---HS		LGD-G-R-TA-A--QV----
--AA--N--NR--FE-IET-KP--TC-		IKG-S-E-TA-AA--NV----
--AA---AKR--FTS-QT-NP---TE		RGP-Q-R-TA-AA--AV----
--AA---AKR--FTA-DT-NP---TE		RGP-A-R-TA-AA--AV----
-VA-TQ--K-M-FES-QA-KP--CT-		HKG-SVR-TS-A--TV----
-VA--R--DR--FTS--A-SP--RTN		HQG-QVR-SA-A--MV----
-AA-VQARRF--FT-SET-RP--RLQ		RGS--Q-TA-A--QV----
-VA-TQ--K-M-FET-QA-KP--CT-		HKG-TVR-TA-A--TV----
-AA---AKR--FSS-QA-KP-DRTE		RGA-R-R-TA-AA--AV----
-VA-TL--R--FTS--A-SP---TN		HQG-QVR-SA-A--MV----
-TA----KR--FEHIEP-KP--KC-		VKG-N-E-TA-AA--NV----
--A---A-G--FSD-VA-RP-DTHA		HH--N-Q-TA-AR--LV----
-AA--TK-RQ--FQA-N--SP--RLH		HGE-T-T-SA-A--QV----
-AS--R--RS--F-TITT-KP--TTN		HQG-TVR-TA-A--MV----
-QA-SN--S-I-FTEI-T-RP--A-K		--NNIN-Q-TS-AS--NI----
--S---L-GR--FK--TP--P---TN		IAG-T-R-TA-AM--SI----
-YS-GR-L---FSNKT--KP-Q-KN		L--TF-E-TE-AA--GL----
PIS-VETLK-I-FSTI-V-KPT-KTN		LFN-SFE-TA-A--QI----
--A--R-LRE-QFSS-QA-TNF-Q-R		LGN-RVT-V--AE-QF EQ----

Supplementary Figure 32

Partial sequence alignment for the protein Zn-dependent hydrolase, showing an 8 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		117	155
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i> CS-505	282898549	IRNLVNEKLEMVGLSA E IANLYPSELSSGMRKRVSFARA
	<i>Raphidiopsis brookii</i> D9	282896221	-----
	<i>Gloeobacter violaceus</i> PCC 7421	37523683	-GE--R-----DP V L-RQM-A-----Q-----
	<i>Nodularia spumigena</i> CCY9414	119511436	-QQI-K-----G -SH-----
	<i>Anabaena variabilis</i> ATCC 29413	75908885	--E-A-----RG -GH-----
	<i>Nostoc punctiforme</i> PCC 73102	186681882	--D-K-----PS -SD--A-----
	' <i>Nostoc azollae</i> ' 0708	298490315	--Q-----D---PG -GH-----
	<i>Nostoc</i> sp. PCC 7120	17227676	--E-A-----RG -GH-----
	<i>Fischerella</i> sp. JSC-11	354567663	V-K-E-----PG -GDR--A-----
	<i>Moorea producta</i> 3L	332711284	--K---QS-----G VGDR--A-----
	<i>Microcoleus chthonoplastes</i> PCC 7420	254411531	--E-D-----PG --DQ--A-----R-----
	<i>Oscillatoria</i> sp. PCC 6506	300865775	--A-D-----PG -GDR--AQ-----
	<i>Trichodesmium erythraeum</i> IMS101	113475809	-QE--Q-----GN -GDR--A-----
	<i>Microcoleus vaginatus</i> FGP-2	334118793	--E-E-----PR -GER--AQ-----
	<i>Lyngbya</i> sp. PCC 8106	119493275	-QK-D-----PG -GDRF--A-----
	<i>Arthrospira platensis</i> Paraca	284052670	--Q-E-R-----PN -CDR--A-----
	<i>Arthrospira maxima</i> CS-328	209524887	--Q-E-R-----PN -CDR--A-----
	<i>Arthrospira</i> sp. PCC 8005	376007644	--Q-E-R-----PN -CDR--A-----
	<i>Arthrospira platensis</i> NIES-39	291568101	--Q-E-R-----PN -CDR--A-----
	<i>Cyanothece</i> sp. PCC 7425	220909177	--E-E---V---PG -GD-----
<i>Acaryochloris marina</i> MBIC11017	158335470	--D-----V---PG -GH-F-A-----	
<i>Cyanothece</i> sp. PCC 8801	218245390	--Q-DA--A---QG --DK--Q-----	
<i>Acaryochloris</i> sp. CCME 5410	359457319	--D-----V---PG -GH-F-A-----	
<i>Crocospaera watsonii</i> WH 8501	67922746	--E-EA--A---QG MSDR--Q-----	
<i>Cyanothece</i> sp. PCC 7822	307150088	--Q-E-----PG -SHQ--A-----	
<i>cyanobacterium</i> UCYN-A	284928694	--K-DK--A---HG -SDR--Q-----	
<i>Cyanothece</i> sp. PCC 7424	218439977	--Q-E-----PG -SHQ--AQ-----	
<i>Synechococcus</i> sp. PCC 7002	170079313	--E-E-R-D---G VGDR--Q-----	
<i>Thermosynechococcus elongatus</i> BP-1	22297871	--EI-S---A---G MED--AQ-----	
<i>Synechococcus</i> sp. PCC 7335	254424603	-KL-DKA-AV---PD --D--A-----	
<i>Cyanothece</i> sp. ATCC 51142	172037581	--E-EA--A---QG MSDR--Q-----	
<i>Cyanothece</i> sp. CCY0110	126656904	--E-EA--A---QG MSDR--Q-----	
<i>Microcystis aeruginosa</i> NIES-843	166363331	--E-EKS-----QG -SD-----A-----	
<i>Synechococcus</i> sp. JA-2-3B'a(2-	86608564	--E-AA--A---AG KEHNL-E-----A-----	
<i>Microcystis aeruginosa</i> PCC 7806	159026050	--E-EKS-----QG -SD-----A-----	
<i>Synechocystis</i> sp. PCC 6803	16329341	--AI-E-N-L---PG -GDRF--A-----L---	
<i>Synechococcus elongatus</i> PCC 6301	56751172	-EHY-ACS-DS---PN --H-M-----	
<i>Synechococcus</i> sp. BL107	116071572	--ER-ETC--A---YD V-H---G-----Q-----	
<i>Synechococcus</i> sp. CC9902	78183818	--ER-ETC--A---YE V-H---G-----Q-----	
<i>Synechococcus</i> sp. WH 5701	87300575	--DR-GR--A---G -EERL-G-----Q-----	
<i>Synechococcus</i> sp. WH 8109	260435991	--DR-HAC--E---YD V-H--AG-----Q-----	
<i>Synechococcus</i> sp. CB0205	317968583	--QR-A-A-A---G TED--G-----Q-----	
<i>Synechococcus</i> sp. WH 7805	88808078	--ER-QQC--A---HE V-DK--GQ-----Q-----	
<i>Synechococcus</i> sp. WH 7803	148238592	--ER-QQC--A---HE V-EK--GQ-----Q-----	
<i>Synechococcus</i> sp. WH 8102	33864747	--ER-EAC--A---FD V-HQ--G-----Q-----	
<i>Synechococcus</i> sp. CC9605	78211759	--DRAHAC--A---YD V-H--AG-----Q-----	
<i>Synechococcus</i> sp. WH 8016	352095152	--ER-MAC--A---YE V-HQ--G-----Q-----	
<i>Cyanobium</i> sp. PCC 7001	254430174	--RK-A-A-A---GG -EERL-G-----Q-----	
<i>Synechococcus</i> sp. CC9311	113953288	V-DR-MAC--A---YE V-HQ--G-I---Q-----	
<i>Synechococcus</i> sp. RS9916	116075558	--ER-MAC--A---HD V-DK--GQ-----Q-----	
<i>Synechococcus</i> sp. RS9917	87123071	-HTR-MQC-DA---ND V-DK--GQ-----Q-----	
<i>Synechococcus</i> sp. RCC307	148241318	-KER-ARC--A---AG TED-M-A-----Q-----	
<i>Prochlorococcus marinus</i> CCMP1375	33239773	A-EI-ISC--E---YN --DKF-GQ-----Q-----	
<i>Prochlorococcus marinus</i> AS9601	123967850	-HEI-R-C-AE---FN VE-KL-N-----Q-----	

Supplementary Figure 33

Partial sequence alignment for the ABC transporter-like protein, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		282900497	564	601
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282900497	GKLQMSYDLFKLLNTFQED GR	KTTEFDLLSLWPFLVEN
	Raphidiopsis brookii D9	282896766	-----R-----H-V---	--R-----L---
	Nodularia spumigena CCY9414	119511418	----A----S--H-GKVN	PLN-----L-L--
	Nostoc sp. PCC 7120	17232312	----A----S--K-EESS	-LN-----L-L--
	Anabaena variabilis ATCC 29413	75908310	----A----S--K-EESS	-LN-----L-L--
	Nostoc punctiforme PCC 73102	186684255	----G-E--S---IANPN	PLN-E-----L-L--
	Fischerella sp. JSC-11	354568281	----A----L--S-DKTL	KL---A---L-L--
	Microcoleus chthonoplastes PCC 7420	254415648	----A----S--SAEDTQ	MP---R---SL-LD-
	Lyngbya sp. PCC 8106	119494006	----G----SV--PGEK	NRRN----T---L-L--
	Arthrospira platensis Paraca	284050988	----G----C--SSDRTK	SPN---A---L-LN-
	Arthrospira sp. PCC 8005	376005317	----G----C--TSDRTK	SPN---A---L-LN-
	Arthrospira maxima CS-328	209523770	----G----C--TSDRTK	SPN---A---L-LN-
	Trichodesmium erythraeum IMS101	113476012	----A----S--SEDTNL	GS---A---L-L--
	Other Cyanobacteria	Microcystis aeruginosa NIES-843	166364001	----A-E-SS---ENKS
Microcystis aeruginosa PCC 7806		159025953	----A-E-SS---PENKS	---A---I-L--
Cyanothece sp. PCC 7822		307150638	----A---N---NKTS	N---A---L-L--
Cyanothece sp. PCC 7424		218441213	----A---S---PNNTQ	KNK---A---L-L--
Crocospaera watsonii WH 8501		67920994	----A---C-FDSDEVN	LKQ-----L-LD-
Crocospaera watsonii WH 0003		357265640	----A---C-FDSDEVN	LKQ-----L-LD-
Acaryochloris sp. CCMEE 5410		359457099	----G-E--T--AEPDAH	PV--F-----L-L-T
Synechocystis sp. PCC 6803		16332102	----LA----N--AADND-	-S---T---L-LD-
Acaryochloris marina MBIC11017		158335137	----G-E--T--AEPEAH	PV--F-A---L-L-T
Cyanothece sp. ATCC 51142		172038060	----A---C-FEDDDHH	QK-----L-LDT
Cyanothece sp. CCY0110		126655911	----A---C-FEDDDNK	QK-----L-LDT
Cyanothece sp. PCC 7425		220910122	----A-E--S--ADQTPN	RP-----L-LD-

Supplementary Figure 34

Partial sequence alignment for the small GTP-binding protein domain protein, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		228	266
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899881	PEPLPKYISEISAEIKN YTE IHQQNRSTLTVGLVFDGD
	Raphidiopsis brookii D9	282897077	-----C----- --Q--F-K-----
	'Nostoc azollae' 0708	298489641	-----L-RLF---T -GES-N-G-----
	Nodularia spumigena CCY9414	119513312	-----L-RLFST--T YR-TDK-S-A-----
	Nostoc sp. PCC 7120	17232581	-----L-LLFET-R- -RET-P-S-S-----
	Nostoc punctiforme PCC 73102	186684582	-----L-KLFEV--T -RETDK-G-----
	Anabaena variabilis ATCC 29413	75908584	-----L-LLFET-RK -RET-P-S-S-----
	Fischerella sp. JSC-11	354568900	-----L-RLFQVM-T -REH-PTG---A-----
	Microcoleus chthonoplastes PCC 7420	254409746	----R-L-QLFRQ-RT -RR-TEG--R-----
	Lyngbya sp. PCC 8106	119488420	-----P-LFRQ--S -KRSQDS-A--F-----
	Trichodesmium erythraeum IMS101	113474765	----R-L-KLFKR--S YRAA-EGS-AI-----
	Arthrospira platensis Paraca	284053188	----R-L--LFRQ--S -TPKANGV-----
	Arthrospira platensis NIES-39	291570817	----R-L--LFRQ--S -TPKANGV-----
	Arthrospira maxima CS-328	209524936	----R-L--LFRQ--S -TPKANGV-----
	Oscillatoria sp. PCC 6506	300864526	----R-L--LFRV-RT --RAEGG-AI-F-----
	Microcoleus vaginatus FGP-2	334118948	----R-L-QLFRVMRT --RKS-G-SI-F-----
	Moorea producta 3L	332708999	----G-L--LFEALQS --REPGVE-A--A-----
Other Cyanobacteria	Cyanothece sp. ATCC 51142	172035686	----R-L-KLFTS--T AAS-GNNS-R-----
	Crocospaera watsonii WH 0003	357262623	----R-L-KLFLT--T AIS-GDNS-K-----
	Crocospaera watsonii WH 8501	67923543	----R-L-KLFLT--T AIS-GDNS-K-----
	Cyanothece sp. CCY0110	126658857	----R-L-KLFNN--T AAS-GDDS-R-----
	Synechococcus sp. PCC 7335	254425346	-----PNLFEA--K NQTNG----I----
	cyanobacterium UCYN-A	284929674	----C-LTKLLTTV-K FNAEKS-G-SI-----
	Cyanothece sp. PCC 8801	218246599	----R-L-KLFEA--T AASEGTNG-R-----
	Microcystis aeruginosa NIES-843	166367171	----RNLR--IDKLAH SA-LAP-R-----
	Microcystis aeruginosa PCC 7806	159026931	----RNLT--IDKLAH SA-LAP-RL-----
	Cyanothece sp. PCC 7424	218441826	----R-L-DLL-TV-A APGDG-K-----
	Cyanothece sp. PCC 7425	220907634	---IAR-LQSTC-TLHQ FR-THPDQ-S-----
	Cyanothece sp. PCC 7822	307153357	----R-L--LLETV-T APGEG-K-----
	Synechococcus sp. WH 8102	33864771	---AP-L-DLITAVQH STAAG-PA-----
	Synechococcus sp. PCC 7002	170078101	---E-HL-QLQRT-RA A-N-DLEAIQ--F-----
	Synechocystis sp. PCC 6803	16330692	-----VPKLL-ALQT AAKAGGPE-R--I----
	Synechococcus sp. RCC307	148243397	---AP-L--LI-QV-A STAAGQPA--I-----
	Acaryochloris marina MBIC11017	158337679	-----LGKSIQA-QS HRQ-APQEEAI--F-----
	Synechococcus sp. WH 8109	260436430	---AR-LGALI-AV-A TAAG-PA-----
	Acaryochloris sp. CCMEE 5410	359459165	-----LEKSIQA-QS HRQ-APQEEAI--F-----
	Prochlorococcus marinus str. NATL1A	124025071	---KA-L-QLIQ-VQD ESQAGK-SM-----

Supplementary Figure 35

Partial sequence alignment for the protein Phosphoglucomutase, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		12	49
	Raphidiopsis brookii D9	282898311	GIDAVKVGVEVDISGG G LPGIILGLPDSAIQESKERV
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899606	-----V--- -V-----
	Nodularia spumigena CCY9414	119510783	-----V--- -V-----
	Nostoc sp. PCC 7120	17231580	-----VV-----V-----
	Anabaena variabilis ATCC 29413	75907039	-----VV-----V-----
	Nostoc punctiforme PCC 73102	186682819	-----V--- -VV-----R---
	Fischerella sp. JSC-11	354568632	-----V--- -VV-----A-V-----
	Lyngbya sp. PCC 8106	119483342	-----K-VV-----T-V---R---
	Arthrospira platensis Paraca	284053280	-----L--- -K-V-----T-V---R---
	Arthrospira sp. PCC 8005	376004564	-----L--- -K-V-----T-V---R---
	Arthrospira maxima CS-328	209524574	-----L--- -K-V-----T-V---R---
	Oscillatoria sp. PCC 6506	300863843	-----V--- -K-VVV-----T-V---A----
	Microcoleus chthonoplastes PCC 7420	254409927	-----V--- -VVV-----T-V---R---
	Moorea producta 3L	332707070	-----V--- -VVV-----T-V---R---
	Microcoleus vaginatus FGP-2	334116671	-----A-V--- -K-VVV-----V---A----
	Other Cyanobacteria	Trichodesmium erythraeum IMS101	113478131
Microcystis aeruginosa PCC 9808		389826188	---I-I-----V--- -AVV-----T-V-----
Microcystis aeruginosa NIES-843		166364485	---I-I-----V--- -AVV-----T-V-----
Cyanothece sp. PCC 7425		220906543	--E-R-----V--- -VVV-----V-V---R---
Acaryochloris sp. CCMEE 5410		359464225	--E-I-----V--- -V-V-----T-V---R---
Acaryochloris marina MBIC11017		158334534	--E-I-----V--- -V-V-----T-V---R---
Cyanothece sp. PCC 7424		218441221	-----A---V--- -S-T-V-----T-----R---
Cyanothece sp. PCC 8801		218247696	----I-----VT-- -S-T-V-----T-V-----
Cyanothece sp. PCC 8802		257060967	----I-----VT-- -S-T-V-----T-V-----
Cyanothece sp. PCC 7822		307153388	-----V--- -A-TVV-----L-V---R---
Synechocystis sp. PCC 6803		16331658	----I-----V-A- -A-AVV-----T-V---R---
Cyanothece sp. CCY0110		126661166	-----VA-- -A-T-V-----T-V---R---
Cyanothece sp. ATCC 51142		172038974	-----VA-- -A-T-V-----T-V---R---
Crocospaera watsonii WH 0003		357261924	-----VA-- -A-T-V-----T-V---R---
Synechococcus sp. PCC 7002		170076988	-V--LR-----V--- -KMMVV-----T-V---R---
Thermosynechococcus elongatus BP-1		22298137	---IP-----V--- -RTVVV-----AGV--AR---
Synechococcus elongatus PCC 6301		56750173	-L--P-C--A-V-A- -AMVLV-----A-----
Synechococcus sp. RS9917		87125304	-L--RP-T---LAP- -LQLV-----A-----R---
Synechococcus sp. WH 7805		88807217	---RP-T---LAP- -LQLV-----T-----R---
Synechococcus sp. CC9311		113953395	-ME-LP-T---LAP- -LQLV-----T-----R---
Synechococcus sp. RS9916		116073812	-LE-CP-T---LAP- -LQLV-----A-----R---
Synechococcus sp. BL107		116071391	-LE-KP-T---LAP- -QLV--A-R-----R---
Synechococcus sp. JA-3-3Ab		86606663	-LK--G-T---VG-- -Q-TLV---T-V--A----
Synechococcus sp. WH 7803		148240473	---RP-T---LAP- -LHLV--A-T-----R---
Synechococcus sp. WH 8102		33864863	-LQ-HP-T---LAP- -QLV---K-----R---
Synechococcus sp. WH 8109		260436583	SLE-RP-V-DLAPGLP -VQLV---K-----R---
Synechococcus sp. PCC 7335		254422783	---L-----L--- -VVVV-----T-V---R---
Synechococcus sp. CC9902		78183991	-LE-KP-T---LAP- -QLV--A-K-----R---
Synechococcus sp. WH 5701		87301349	-LE-RE-A---AP- -LQ-V--AEA-VR--R---
Synechococcus sp. CB0205		317967977	--E-YP-Q---LAP- -ALV-V--A-A-V---R---
Cyanobium sp. PCC 7001		254430161	-L--HQ-T---GP- -ALQMV--A-A-V---R---
Synechococcus sp. CB0101		318040511	-LE-RA-A---AP- -LQVV--A-A-V---RQ--
Prochlorococcus marinus MIT 9303		124022024	--E-LA-T---LVP- -LQLV---A-----R---
Gloeobacter violaceus PCC 7421		37520045	---LG-T---VGF- --QTTLV---A-V---R--I

Supplementary Figure 36

Partial sequence alignment for the Mg chelatase-related protein, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		198		242
		RALPIAGWMSGLELAYLRL DS	NRNNQGDRLVLETGGTESWILANL	
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282901672	-----P-----	-----
	Raphidiopsis brookii D9	282896250	-----M-----F	DTS-----I-----A-----V---I
	'Nostoc azollae' 0708	298490971	-----F--V	DTS--T-----A-----I
	Anabaena variabilis ATCC 29413	75908378	-----F--V	DTS--M-----A-----I
	Nostoc sp. PCC 7120	17232380	-----C-NF	DTK--Q-----A-----I
	Nodularia spumigena CCY9414	119510299	----L-----F--F	DTSEEA--L----VN----V--I
	Nostoc punctiforme PCC 73102	186682051	----L-----D--W--F	DDSP--G--L----A-----
	Fischerella sp. JSC-11	354566488	--V-L-----GF-H-	QK-PRSS-----VSD---V-D-
	Synechococcus sp. PCC 7002	170077740	----L-----M---K-	ETSSRPL-R----ASD----V-V
	Microcystis aeruginosa PCC 7806	159028601	----LS-----M--K-	ETGSRPL-R----ASD----V-V
	Microcystis aeruginosa NIES-843	166363644	----L-----VC-KV	QEKPRPI-S----LSD-----
	Cyanothece sp. CCY0110	126659192	----L-----MG--K-	EESDRPI-R----VSD---VI-V
	Cyanothece sp. PCC 7424	218437072	----F-A-L-----G--QY	NTDPRPIMR----ASD---V--V
	Synechocystis sp. PCC 6803	16330318	-ST-L-----C-K-	EMDPKQ-L-----LSDR-VI-Y-
	Acaryochloris marina MBIC11017	158336667	-ST-L-----C-K-	EMDPKQ-L-----LSDR-VI-Y-
	Acaryochloris sp. CCMEE 5410	359459949	----L-----MG--KF	EESDRPIVR----VSD---VI-V
	Cyanothece sp. PCC 7822	307150318	--K-L-----C-T-	ETEPVPQ-I----VSDR---R-
	Cyanothece sp. PCC 7425	220906218	----L-A-----GF-KF	EE-PRPIVR----TSD----V-I
	Cyanothece sp. PCC 8802	257062177	----L-A-----GF-KF	EE-PRPIVR----TSD----V-I
	Cyanothece sp. PCC 8801	218249090	----L-----VS-KF	QTKPFPN-C----LSDN-----
	Cyanothece sp. ATCC 51142	172036928	--M--A-L---P-W-TY	DSPAKQ-L-----S-R-T--A-
	Synechococcus elongatus PCC 6301	56750056	----L-A-----VA-KF	KSKPLPI-S----LSD-----
	Crocospaera watsonii WH 8501	67922272	----L-A-----VA-KF	KSKPLPI-S----LSD-----
	Crocospaera watsonii WH 0003	357264130	----A-L---F--VS--F	DRKNSST-Y----LNDKSV-I--
	cyano bacterium UCYN-A	284928976	----L---L---P-S-SF	EETPQPL-I----ASDR---
	Thermosynechococcus elongatus BP-1	22299529	----A-L---M-P--HV	AEG--LF-A-LNDCYLF-Q-
	Synechococcus sp. JA-3-3Ab	86605930	--AL--LG--PVR--I	D-RQ-I-DA-QDD--LVTD-
	Synechococcus sp. RS9916	116075331	--AL--LG--PVR--M	NAQQ-I-DA-QDD--LVSD-
	Synechococcus sp. WH 7803	148238987	--AL--LG--PVR--V	SNQQ--DA-QDD--LVSD-
	Synechococcus sp. WH 7805	88807699	--AL--L---PVR-EV	C-QQ--A-LEDR-LVSD-
	Synechococcus sp. RCC307	148242688	--AL--LG--PVR--V	E-CQ--DA-QDDA-LVTD-
	Synechococcus sp. RS9917	87123919	-S-A---IA---P-R-EM	E-TS---A-LEDR-R--T-
	Synechococcus sp. WH 5701	87302524	--AL--LA---PVR-EM	VDRQ---A-LEDR-L--T-
	Cyanobium sp. PCC 7001	254432298	--AV--LG--PVR-LV	E-RQ--A-QDDR-LVSD-
	Synechococcus sp. WH 8102	33866273	-S-A---IA---PVR-EV	S-NQ--A-LEDR-L-G--
Synechococcus sp. CB0101	318041062	-S-AM--LG--PVR-LV	E-RQ-I--A-QDDR-LVSD-	
Synechococcus sp. CC9902	78185205	-S-AM--LG--PVR-LV	E-RQ-I--A-QDDR-LVSD-	
Synechococcus sp. BL107	116072198	-S-A---V---PVR-EV	C-QQ--A-QEDR-L-GQ-	
Synechococcus sp. CB0205	317969607	-S-AM--LG--PVRMIV	EDRQ--A-QDDR-LVSD-	
Synechococcus sp. CC9605	78212273	-SVAM--LG--PVRMIV	EDRQ--A-QDDR-LVSD-	
Synechococcus sp. WH 8109	260434334	--K-L-A-----SFV--	DTTDDKPK-L----AND-----	
Moorea producta 3L	332712125	--T-L-A-----F--V	NKGDPRS-L----END-----	
Lyngbya sp. PCC 8106	119490556	--I-L-A-----F---	VETPRPN-I----EN-----	
Arthrospira maxima CS-328	209528431	--I-L-A-----F---	LETPRPS-I----EN-----	
Arthrospira platensis Paraca	284051219	--TAL-----F-KF	ESGPPA--D--ANDR-----	
Microcoleus vaginatus FGP-2	334120908	--TAL-A-----FVTF	DGGPPA-----ASDR-----	
Oscillatoria sp. PCC 6506	300863927	--KAL-A-----FVKF	QGGVTP--L----ASD--A----	
Microcoleus chthonoplastes PCC 7420	254413499	--QAL-A-----FIKF	SPASPA--L-N--NDC-----	
Trichodesmium erythraeum IMS10	113478101	-S-AL-A-LG--P-K-LI	E-TQII--A-QADR-LVTDV	
Prochlorococcus marinus NATL1A	124025296	--AL--LG--PVC-AV	D-TQ-M--A-QDDR-LVTD-	
Prochlorococcus marinus MIT 9313	33863502	----LSA-L---PVR-LV	ENNQ-L--S-QESL-LVTDM	
Prochlorococcus marinus CCMP1375	33239980	-DRDL-S-LAAR-PVS-QY	TRAEI-G-I--A-LD-R----TF	
Gloeobacter violaceus PCC 7421	37523856			

Supplementary Figure 37

Partial sequence alignment for the hypothetical protein Aazo_1936, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		48	102
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282900791	DIRCLAIPKTDMRDWNLLTTNVLIDLIMELT TGS FKRGNRLVYLCGESFGACLAM
	Raphidiopsis brookii D9	282896533	-----R----- -T- V-----S-----
	Anabaena variabilis ATCC 29413	75908937	-V-----RQ-LTS-DV--N-----HA--E -SSQ-A-----G----
	Nostoc sp. PCC 7120	17227731	-V-----RQ-LTS-DV--N-----HA--E -SSQ-A-----G----
	'Nostoc azollae' 0708	298492949	-V-----RK-LTT-EV--S-----HA--E RSCH-P-----G----
	Nodularia spumigena CCY9414	119512205	-V-----RQ-LTT-DE-SN-----HA--E -SSH-P-----G----
	Nostoc punctiforme PCC 73102	186680642	-V-S--L-RK-LNT-DV--KS-----A--E -SSH-S-----G----
	Fischerella sp. JSC-11	354567133	-V-----RE-LTS-EV--N-----HA--E -SSQ-P-----G---Q
	Oscillatoria sp. PCC 6506	300867844	-----P--AS-DV-VSETIA--EA-KA AGNH-S-----G----
	Trichodesmium erythraeum IMS101	113475476	-----S-QN-LST-EQ-SEKTIN--KI-QK AAPK-P-----G---L
	Lyngbya sp. PCC 8106	119487034	-----S-LQ-RSN--T--ARTIA--KA--- SNPE-G--V-----G---V
	Arthrospira platensis Paraca	284052454	NL---T--PE-VN--DT--DRTV---QK--S LNL-QDI-----G----
	Microcoleus vaginatus FGP-2	334118813	-----TL-AD-TPS-DV-VGETIA--EA-KQ AGKPK-P-----G----
	Arthrospira maxima CS-328	209528059	NL---S--PE-LNH-DG--DRTV---QK--S LNL-QDI-----G----
	Arthrospira platensis NIES-39	291571485	NL---T--PE-VN--DT--DRTVA--QQ--S LNL-QDI-I-----G----
Microcoleus chthonoplastes PCC 7420	254417287	-V---V--PHYLSN-QD-ANQ-AS--AT--S Q-RSQS-----G---L	
Moorea producta 3L	332707291	-V-Y---TD-LSNSQK-AGV-VN--EA--N QGASQE-----K-K-S	
Other Cyanobacteria	Cyanothece sp. PCC 7425	220908712	NV-----PD-LSS-DR-AEQ-IA--EA--A RMSCSQ-----G---L
	Cyanothece sp. ATCC 51142	172035003	S---S--SN-QS--ST-V--TVT-MRK--E SHPHSS-----I
	Microcystis aeruginosa NIES-843	166368139	-L---V-SPQYLG--EE-S-L-IA-LEQ--K -QS-K-----G---L
	Cyanothece sp. PCC 8801	218246897	Q-Y--S--QN-CS---T-AKKTIT--EK-RQ NHLLSSP-----G---L
	Cyanothece sp. PCC 7822	307592423	-V--FV--ADNLT--EH-SSQAIK--TRS--E QKPQAT-----G---L
	Microcystis aeruginosa PCC 7806	159026978	-L----SPQYLG--EE-SDC-IA-LAQ--K QQS-K-----G---L
	Crocospaera watsonii WH 8501	67920575	A---S--SY-QS--ST-VHKTVT--RK--E AHPHSS--I-----G----
	Crocospaera watsonii WH 0003	357266117	A---S--SY-QS--ST-VHKTVT--RK--E AHPHSS--I-----G----
	Cyanothece sp. CCY0110	126659795	S---S--SN-QS--PT-VKKTVT--RK--E DHPHSSA-----G---I
	Synechococcus sp. WH 5701	87302256	-----SM-SD-LTG-ET-VEQIAR--AN-QH QIQS-PT-I-----G---L
	Synechococcus sp. RS9917	87124699	NV---SMSPS-RSG-GE--ERTAT--K--QD RNPGQMTII-----G---L
	Gloeobacter violaceus PCC 7421	37521810	-VTA-S--VD-RGE-PD-IAR-DE--AAHPG -RII-----G----

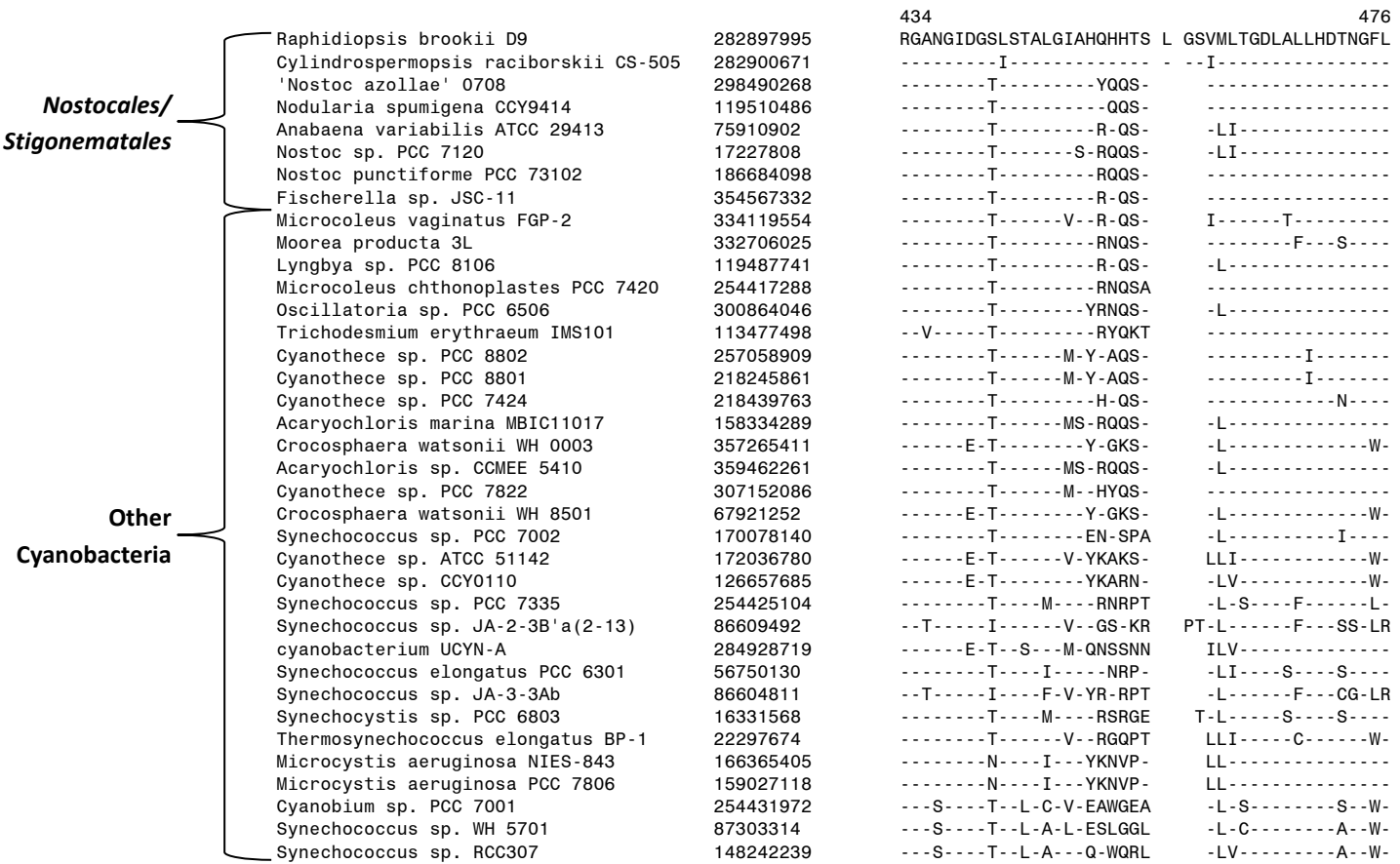
Supplementary Figure 38

Partial sequence alignment for the Alpha/beta hydrolase fold protein, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		54	101
	Raphidiopsis brookii D9	282896230	DSYGTGKLNGRSELLGKFTREY EHS DP
	Cylindrospermopsis raciborskii CS-505	282898556	-----R-----SQ--
	Nostoc sp. PCC 7120	17227796	-----R-----Q-AQ--
Nostocales/ Stigonematales	Anabaena variabilis ATCC 29413	75909269	-----R-----S-----
	'Nostoc azollae' 0708	298491921	-----R-----R-Q--
	Nodularia spumigena CCY9414	119508846	-----R-----R-N--
	Nostoc punctiforme PCC 73102	186681282	-----R-----R-----
	Fischerella sp. JSC-11	354564636	-----S-----Q-S--
	Lyngbya sp. PCC 8106	119486839	-----Q-----Q-SQQ-
	Microcoleus vaginatus FGP-2	334121131	-----R-----Q--R-SSD-
	Microcoleus chthonoplastes PCC 7420	254411837	-----R-----M---SQG-
	Trichodesmium erythraeum IMS101	113475801	-----Q---T---Q-Y-D-
	Cyanothece sp. PCC 7424	218437204	-----Q-----R-FTA-
Cyanothece sp. PCC 8802	257057936	-----Q-----R-FTA-	
Cyanothece sp. PCC 8801	218244911	-----T-K---T---Q--	
Cyanothece sp. ATCC 51142	172037425	-----K---T---E--KA-	
Crocospaera watsonii WH 0003	357264942	-----K---T---E--KA-	
Crocospaera watsonii WH 8501	67921562	-----S---S---Q-S-N-	
Cyanothece sp. PCC 7822	307152159	-----T-K---T---SH--	
Cyanothece sp. CCY0110	126654819	-----R-Q-K-S---Q-S-S-	
Cyanothece sp. PCC 7425	220910486	-----R-Q-Q-K---Q-SQA-	
Synechocystis sp. PCC 6803	16330597	-----R-S-Q-T---Q-SAS-	
Synechococcus sp. PCC 7335	254421749	-----Q---K---R--G--	
Microcystis aeruginosa NIES-843	166366476	-----R-S---S---Q-SQA-	
Synechococcus sp. PCC 7002	170077364	-----Q---K---R--G--	
Microcystis aeruginosa PCC 7806	159026972	-----R-----Q-L-AM	
Synechococcus sp. CC9311	113953419	-----R-----M---Q-AMRA	
Synechococcus sp. CB0205	317968769	-----R-----Q--R-AARL	
Synechococcus sp. WH 8102	33866563	-----R-----R---Q-L-AM	
Synechococcus sp. WH 8016	352096559	-----R-----T---R-AMAA	
Synechococcus sp. CB0101	318042609	-----S-----A--D-MA-L	
Synechococcus sp. WH 8109	260435169	-----RFT---Q---G-LQAL	
Synechococcus sp. WH 5701	87300949	-----RF---R---R-CAGL	
Cyanobium sp. PCC 7001	254430457	-----S-S---A---D-IA-L	
Synechococcus sp. CC9605	78211964	-----R-----S---G-IEAL	
Synechococcus sp. CC9902	78185484	-----L-----E---A-LSAM	
Synechococcus sp. WH 7805	88807847	-----R-----S---G-IDAL	
Synechococcus sp. BL107	116072604	-----F-----E---R-LSAM	
Synechococcus sp. WH 7803	148238806	-----RF---A---A-AIL-AR	
Synechococcus sp. RS9916	116075804	-----R-----F---QGLDGL	
Synechococcus sp. RS9917	87124924	-----AYA---R---QAIAAL	
Synechococcus sp. RCC307	148243173	-----E-----LLDT	
Prochlorococcus marinus MIT 9515	123965294	-----E-----LLDT	
Prochlorococcus marinus CCMP1986	33860612	-----N-K-----I---LLNT	
Prochlorococcus marinus AS9601	123967599	-----S-F-Q--K-I-D-LE-L	
Prochlorococcus marinus NATL2A	72383000	PKRRLKVKVT-----PF---	

Supplementary Figure 39

Partial sequence alignment for the protein aldo/keto reductase, showing a 5 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.



Supplementary Figure 40

Partial sequence alignment for the menaquinone biosynthesis protein, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		298	350
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898767	LPLQLISNLLSQVQIQGKELRLQIQIQL QNQV ARDILITRDKRLVSLHQLNL
	Raphidiopsis brookii D9	282895756	-----L--Q-----K----- ----T-----L--N----
	Anabaena variabilis ATCC 29413	75908427	---A--P-----T--A--F-Y----- ---V-LE-----LQW-NK-AI
	Nostoc sp. PCC 7120	17232411	---A--P-----L-T--A--F-Y----- ---V-LE-----LQW-SE-AI
	Nostoc punctiforme PCC 73102	186684606	---A--P---N---T-----F-H--V-I SK-L-LK--Q--LAW-N-VEI
	Nodularia spumigena CCY9414	119510077	---A--EP--N-AYL--Q--F-T--AKV -YNL-LE--H--IELIK--TI
	Fischerella sp. JSC-11	354565833	---V--QP--T---T-----FHAR-A-I VQ--LE--Q--LDLIGK-DI
	Cyanothece sp. PCC 7424	218441738	---G--TP--Q--NL--QD-KS--FKM -QS--KE--E--L-Y-N----
	Cyanothece sp. PCC 7822	307154551	---E--NP--Q-AKN---Q-IKI--AKI -KKL-QE--QK-L-Y-SE-A-
	Microcystis aeruginosa NIES-843	166366417	---K--FP--EAAK-----KF-SS-V -QSV-ES---K-LAY-E----
Other Cyanobacteria	Microcystis aeruginosa PCC 7806	159027115	---K--FP--EAAK-----KF-SA-V -QSV-ES---K-LAY-K--S-
	Cyanothece sp. PCC 8802	257060274	---G--FP--T-AKE-A-T--KQ--AKI -KA--QE--Q--LVY-EE-Q-
	Cyanothece sp. CCY0110	126656081	---A--TP--TEAK---T-VKQ--AI- -KTV-QQ---K-I-Y-ET-Q-
	Acaryochloris sp. CCMEE 5410	359463118	---A--QP--QD--E-LQT-V-YE-F-G -K-L-QE-NQH-L--VET-QI
	Cyanothece sp. PCC 8801	218247196	---G--FP--T-AKE-A-T--KQ--AKI -LA--QE--Q--LVY-EE-Q-
	Crocospaera watsonii WH 0003	357261876	---A--TP--TEAKE---S-VKQ--AIT -KTV-QQ----LNY-ET-Q-
	Cyanothece sp. ATCC 51142	172035129	---A--TP--TEAKK---I-IKQ--AI- TKTV-QQ--QK-LDY-ES-E-
	Crocospaera watsonii WH 8501	67925635	---A--TP--TEAKE---S-VKQ--AIT -KTV-QQ-N--LNY-ET-Q-
	Lyngbya sp. PCC 8106	119491182	---N--P--H-AET-A---KR--SMA -EV-QE--S--INYIN--S-
	Moorea producta 3L	332711303	---S--EP--T-A-A-A---S-AS-AK- ----QQ--Q--IDL-TA-K-
	Trichodesmium erythraeum IMS101	113475598	---IA--P--E-ASS--DL-ILRHEAEI -L-AFQE-----LNY-DD-SI
	Arthrospira sp. PCC 8005	376003674	---A--QP--ERGKV-----I-EKKVRI -EVVAE--R--LEY-KT-D-
	Oscillatoria sp. PCC 6506	300864432	---G--P--E-AK--A-T-VNQF-AKI -KTAVEQ-SD--LKY-T--S-
	Arthrospira maxima CS-328	209526893	---A--QP--ERGKV-----I-EKKARI -EVVAE--R--LEY-KT-D-
	Arthrospira platensis Paraca	284054014	---G--QP--HRAKL-----ISDHKARI -EVVAE--R--FDY-KT-D-
	Arthrospira platensis NIES-39	291572177	---G--QP--HRAKL-----ISDHKARI -EVVAE--R--FDY-KT-D-
	Microcoleus vaginatus FGP-2	334118568	---A--P--TEAK--A-T-ASQ--AKI -QNVIKQ-SD--LNYIAEIS-
Microcoleus chthonoplastes PCC 7420	254412664	---S--P--E-AKA-A-Q-QKQH--K- T-EV-KE--A--IQYIKKIS-	

Supplementary Figure 41

Partial sequence alignment for the DNA sulfur modification protein, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		282898911	348	389
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898911	PILEQKGYFTAGKYEQLQ NNK NNIPVSALLKALGNLLQQVLT	
	Raphidiopsis brookii D9	282898072	-----Q-----	
	'Nostoc azollae' 0708	298491722	--VR-R---IK---D-F-	R---F--FVQ-FRD-ME-L--
	Nostoc punctiforme PCC 73102	186682607	--A--R---IK--FD-FN	R---F--FVQTFRD-VG-L-S
	Nostoc sp. PCC 7120	17230174	--VR-R---IK--FD-F-	R---L--FVQ-FRD-IA-L--
	Anabaena variabilis ATCC 29413	75909015	--VR-R---IK---D-F-	R---F--FVQ-FRD-MG-L--
	Fremyella diplosiphon Fd33	48374205	--VR-R---IK--FD-FN	R---F--FVQ-FKD-MG-L--
	Nodularia spumigena CCY9414	119508927	--R-RS--IK---D-FK	RE--L--GFIQ-FRD-IA-I-L
	Oscillatoria sp. PCC 6506	300868404	--VR-R---IK---D-F-	R---F--FVQ-FRD-MG-L--
	Microcoleus vaginatus FGP-2	334116788	--VR-R---IK---D-F-	R---F--FVQ-FRD-MG-L-S
	Moorea producta 3L	332706581	--VR-R---IQ--FD-F-	R---F--IVQ--RD-ME-L--
	Trichodesmium erythraeum IMS101	113476169	--TA-H---IQ--FD-FN	Q-V-L--F-Q--R--ME---F
	Arthrospira platensis Paraca	284050878	--VR-R---IK---D-FN	RH--LA-FVQ-FRD-MG-I-S
	Microcoleus chthonoplastes PCC 7420	254415737	--VR-R---IS--FD--K	R---YAS-TQ-FQE-I--L--
	Other Cyanobacteria	Arthrospira platensis S6	206585493	--VR-R---IK---D-FN
Arthrospira sp. PCC 8005		376003847	--TR-Q---IQ--FD--N	HH--L--FVQ--RD-MG-F-S
Arthrospira maxima CS-328		209523991	--TR-Q---IQ--FD--N	HH--L--FVQ--RD-MG-F-S
Arthrospira platensis NIES-39		291566816	--VR-R---IK---D-FN	RH--LA-FVQ-FRD-MG-I-S
Lyngbya sp. PCC 8106		119487723	--VR-R-D-IK--FD-FN	R---F--FVQ-FRD-MG-L--
Acaryochloris marina MBIC11017		158338239	--VRRR---IQ--FD---	QSM-FA---Q-FQD-C--L-M
Acaryochloris sp. CCME 5410		359457475	--VRRR---IQ--FD---	QSM-FA---Q-FQD-C--L-M
Cyanothece sp. PCC 7822		307151273	--V-R--F-VS--FD---	R-V-YAS-IQ-FQE--R-I--
Cyanothece sp. PCC 7425		220909333	--TQ-----I---FD---	RER-YG---Q-FRS-I--I--
Cyanothece sp. CCY0110		126659478	--VRKR---IK---D-FN	R---F--FVQ-FRD-MG-L-S
Microcystis aeruginosa NIES-843		166366081	A-V-AR---IR--F---K	R---YL--IQ-FQE-IR-L--
Cyanothece sp. ATCC 51142		172039010	--TQ-R---I---FD-FK	R---YD-VIS-FKS-IN-L--
Cyanothece sp. PCC 7424		218442716	--SQNR---I---FD-F-	R---Y--IVF-FQS-VK-L--
Prochloron didemni P1-Palau		335387279	-LSQ-Q---IS--FD-F-	R---Y--IAA-FQS-VR-L-

Supplementary Figure 42

Partial sequence alignment for the serine/threonine protein kinase and signal transduction histidine kinase, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		142		194
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282898356	ALARWGATAEQKTLILLSEI G	QSTETISLSARNVENIKVIAADQLNVFDLLHA
	Cylindrospermopsis raciborskii CS-505	282899953	-----D-----	-N-----
	Nostoc punctiforme PCC 73102	186684509	-----VVP-S-A-----	AD-DNVY-----I--L-L--N-----
	Nodularia spumigena CCY9414	119511178	-----AP-N-A--MPI-	S-EDNVY-----IQ-L-L-----Y-----
	'Nostoc azollae' 0708	298491399	-----AQ--S-----	A-NVL-----I--L-L-----Y-----
	Anabaena variabilis ATCC 29413	75906915	-FT----EP-S-I-----F	P-NVY-----I--L-L-----Y-----
	Nostoc sp. PCC 7120	17231706	-FT----EP-S-I-----F	P-NVY-----I--L-L-----Y-----
	Fischerella sp. JSC-11	354565795	-I----DP-S---V--K	N-NVL-----L-L--N--Y-----
	Thermosynechococcus elongatus BP-1	22297626	-Q---IEP--V--LAAD-	--VA-----PTL-LLR-----Y-
	Cyanothece sp. PCC 7822	307154841	-----V-EKD-V-----	P-NVG-----IC-V--K--S--Y--T-
	Microcystis sp. T1-4	390438093	--S---SP-E-I---T--	P-NVY--G--IC-L-I-R--S--Y-V-L-
	Microcystis aeruginosa NIES-843	166368483	--S---SP-E-I---T--	P-NVY--G--IC-L-I-R--S--Y-VIL-
	Cyanothece sp. PCC 8801	218245139	-T---IDGN--VV--V---	-DN-Q-----PY--LLK--G--Y--V-
	Cyanothece sp. PCC 7424	218440624	-T---V-EKD-V-----	P-N-----IC-V-L-R--SI--Y-I-S-
	Acaryochloris marina MBIC11017	158337811	-MG--VN-DS-----ADK	HD--Y-----KL-L-L-SN--Y--A-
	Synechococcus sp. RCC307	148243216	-E---VP-GS-V-MV-AD-	SDNVR--V--LPQL-L-----N-
	Synechococcus sp. PCC 7002	170077684	-MT--VASGE-V--V-A-M	-DNVY-----IASV-MVP--G--I--I-N
	Synechococcus sp. WH 8016	352096520	--G-L-VS-TS-V-I--TNP	S-AVQR-I--L-TV-L-----
	Cyanothece sp. PCC 7425	220906720	-I----VQPN-D-V--VADK	QD--Y-----IATV-L-Q-TS-----N-
	Synechococcus sp. WH 5701	87300912	-D--DVP-GS-V--V-DAP	GDVRRR-V--L-KV-L-----
Cyanothece sp. CCY0110	126657989	-G--L-VEPN--VA-V-V-T	D-N-Y-----ICYL-IMR--S--Y--D-	
Cyanobium sp. PCC 7001	254430668	--T-L-IE-GA-V-V-DNA	S-AVRK-V--L-KV-L-----	
Crocospaera watsonii WH 8501	67923584	-----VEPSE-VA-V-L-A	GDN-Y-----LCFV-ILR--S--Y--N-	
Synechocystis sp. PCC 6803	16329940	-T---KP-KRV----D--	P-NVF--G--IPYL-ILR--N--IY-V-V-	
Synechococcus sp. CBO205	317968818	--S-F-IA-GS-V-VV-DAP	S-VVRR-V--L-KV-L-----	
Synechococcus sp. CC9311	113952777	-G-LDVS-TS-V-I--TNP	S-AVRR-I--L-TV-L-----	
Cyanothece sp. ATCC 51142	171700382	-T--VEPT--VA-V-V-T	DAN-Y-----LCHV-ILR--S--Y--D-	
Synechococcus sp. WH 7803	148238769	-G-L-IE-DS-V-IV-GTP	S-VLRR-V--L-KV-L--N-----	
Synechococcus sp. CBO101	318042563	-G-F-IADGA-V-VV-DAP	S-VVRR-V--L-KV-L-----	
Synechococcus sp. WH 8102	33866600	-G-L-IA-DA-V-IV-T-P	SDVRRR-V--L-KV-L--N-----	
Synechococcus sp. WH 7805	88807885	-G-L-IE-DS-V-IVVGNP	S-VLRR-V--L-KV-L--N-----	
Synechococcus sp. RS9916	116075766	-G-L-IA-DA-V--V-TNP	S-VVRR-V--L-KV-L-----N-	
cyanobacterium UCYN-A	284929562	-T---VESRK-VA--L-T	DNN-Y--V--LSYV-ILR--S--IY--N-	
Synechococcus elongatus PCC 6301	56751875	-T---VEPGV-V--TADR	DQNVV--V--LP-V-L-T-TN--I--N-	
Synechococcus sp. WH 8109	260436231	-G-L-VA-GS-V-IV-TNP	SDVRRR-V--L-KV-L-S-----	
Synechococcus sp. CC9605	78211927	-G-L-VA-GS-V-IV-TNP	SDVRRR-V--L-KV-L-S-----	
Synechococcus sp. RS9917	87124963	-G-L-IA-DS-V-I--NP	S-VVRR-V--LDKV-L-----N-	
Synechococcus sp. PCC 7335	254421500	-T--EID-SA-V-MVF--R	NDNVY--L--P-VRM-D-TN--IY-I--	
Synechococcus sp. BL107	116072563	-G-L-IA-DT-V-IV-TNP	S-MVRR-V--L-KV-L-S-NH-----	
Synechococcus sp. CC9902	78185523	-G-L-IA-DT-V-IV-TNP	S-MVRR-V--LDKV-L-S-NH-----	
Synechococcus sp. JA-3-3Ab	86605865	-L---VMED-SA--VA-R	S-AVERAV--IARV-L-GL-----NV	
Microcoleus chthonoplastes PCC 7420	254412295	-I----IDPK--V---PQP	AQQVY-----PS--L-Y--S--Y-V-N-	
Oscillatoria sp. PCC 6506	300865403	-----NP-SRI----P-K	Q-N-Y--G--PTL-I-L-GS--IY-I-A-	
Lyngbya sp. PCC 8106	119484832	-MS--IEP-S-V--V-P-K	Q-NVY--G--IARMRICL--S--Y-I---	
Arthrospira platensis Paraca	284052649	-E---VSSA-V--V-P-K	Q-NVY--G--KMRICL--S--Y-V-A-	
Arthrospira maxima CS-328	209528358	-----VP-A-V--V-P-K	Q-NVY--G--AK-RICL--S--Y-V-A-	
Moorea producta 3L	332712168	-IG--IEP-S-V---AQP	E-KVY--V--LATV-L-Y-TN--TY-I-A-	
Microcoleus vaginatus FGP-2	334120608	-IS---IEPDT-V---P-P	QPNVY--G--I-LVN--L-TS--Y-V-A-	
Trichodesmium erythraeum IMS101	113476578	-IT---IDIQA-IV-V-P-K	Q-NVY--G--IAKL-I-L-NN--IY-I-A-	
Prochlorococcus marinus MIT 9211	159904218	--S-F-IADDA-V-I--AQP	S-I-IR-V--I-RV--G-----N-	
Prochlorococcus marinus NATL1A	124026707	L-K-V-IDSDV-I-I--NKP	S-I-KR-IK-L-KV-L-S-----N-	
Prochlorococcus marinus AS9601	123969296	G--L-IQKTE-V-V--DSP	SDI-KK-IN-I-KV-L-----I-N-	
Prochlorococcus marinus CCMP1986	33862113	G-S-L-IEKTE-V-V--DNP	SDV-KK-IN-L-KV-L-----I-N-	
Gloeobacter violaceus PCC 7421	37519655	-FE--VDMASQSI-LILRE	RQTNTY-----LP-V--T-GN--R--AT	

Supplementary Figure 43

Partial sequence alignment for the ribosomal protein L4/L1e, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		282896194	371	416
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282896194	QTLAVRHRRLREVLEQVRVALGNMERE ITGG RPLVDLVFNHGHACLE	
	Cylindrospermopsis raciborskii CS-505	282901420	-----G--	-----
	Anabaena variabilis ATCC 29413	75908442	-----W-FD--AET--SLVK-	--I---I-----
	Nostoc punctiforme PCC 73102	186685702	-----W-F--AEN--SLIK-	--I---IL-----
	Nostoc sp. PCC 7120	17232439	-----W-FD--AEN--ALVK-	--I---I-----
	Nodularia spumigena CCY9414	119511001	-----D---S--DA-AAEV-DLIQG	-----L-----
	Fischerella sp. JSC-11	354565855	-----R--W-FDS-AQT--SATQG	--I---IL-----
	Acaryochloris marina MBIC11017	158335669	----I-RH--Q--D--QQ-V-EQTAG	-----IL-----
	Acaryochloris sp. CCME 5410	359461030	----I-RH--Q--D--QQ-V-EQTAG	-----IL-----
	Cyanothece sp. PCC 7425	220908997	----I-G---Q--DA-AAEVQLSGKP	I---I-----
Other Cyanobacteria	Cyanothece sp. PCC 7822	307153725	E---I-R----FD-IDKEVKQLRGN	-----LC-----
	Cyanothece sp. ATCC 51142	172039644	-----Y-FNE-AQT---IPQG	SGI----LS-----
	Cyanothece sp. PCC 7424	218441647	E-F-I-R---R--DE-DKTV-QLRGD	-SI----LS-----
	Cyanothece sp. CCY0110	126657878	-----Y-FNE-AQII-KIPQG	SGI----LS-----
	Synechococcus sp. PCC 7335	254423624	-----TH-----DD-QREIKEYT-P	-----LS-----F-
	Synechococcus sp. WH 5701	87300747	-----RQ--G--D--AAEV--PTPQ	---LN--LS-----
	Synechococcus sp. RS9916	116074520	D-K-I-NH--N-FDD-AS---DLNS	---N--LS-----M-
	Synechococcus sp. BL107	116072519	D-H-I-RQ--A-FDG-ARS--RDLKS	AKP----LS-----M-
	Synechococcus sp. WH 8016	352096891	D-K-I-RQM-D-FDA-VL---EDLGQ	ARP----LS-----
	Microcoleus chthonoplastes PCC 7420	254413443	-W---R---Q--DA-ATEV-DLAQG	-----LT-----
	Microcoleus vaginatus FGP-2	334116649	-----L-----DG-AD-V-DVA-G	-----LT-----
	Lyngbya sp. PCC 8106	119483446	---I-DH--Q-FNT-SQTI-ERPQN	Q-I---I-----
	Moorea producta 3L	332710294	DAI---GH--Q-FA--AA---SLPEK	-----FILS-----F-
	Gloeobacter violaceus PCC 7421	37522079	-----R---R--DA-HRQV-ERAKG	-----AI-T-----

Supplementary Figure 44

Partial sequence alignment for the protein metallophosphoesterase, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

			490		546
			LLLIGAYRDHEVYSGHPLTVTLENLRKVEPKI AI	NEIDLQPLDKNQLNLLISGNLAC	
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898911	-----R-----S---I-----		
	Raphidiopsis brookii D9	282898072			
	Fremyella diplosiphon Fd33	48374205	-----N--PPA---ML--GD-Q-KQAQ-	HT-T-K--IQ-IV-H-VADT-K-	
	'Nostoc azollae' 0708	298491722	-----N--PP---ML--S-IQ-TQAT-	-T-S-V--SQVEV-K-VADT-KS	
	Nostoc sp. PCC 7120	17228381	-----N--S-----IL--QDIQ-HAEN-	ET-T-K--NTLD--Y-VADT-K-	
	Anabaena variabilis ATCC 29413	75910701	---L-----PA---ML---AIE-TGAT-	-T-T---SLTS--Q--TDT-N-	
	Nostoc punctiforme PCC 73102	186682607	-----N--FAA---ML-I-EIA-AKVT-	-I-T-A--SAAS--Q-VTDT-S-	
	Nodularia spumigena CCY9414	119508927	--I-----N--NNA---HLAINEIK-QASTL	-Y-R-K--TQAD--Q--ADS-RW	
	Fischerella sp. JSC-11	354568923	I-----N--SRT---YL--NEIE-NRGI	-T-T-K--NQ-D--R--ADT-H-	
	Microcoleus vaginatus FGP-2	334117204	-----N--FPA---ML--AEIQ-TQA-V	-T-T-A--RAGV-C-VADT---	
	Oscillatoria sp. PCC 6506	300864567	-----N--PA---IL--SEIQ-NQAT-	-T-T-A--SQVKV-Q-VADT-K-	
	Lyngbya sp. PCC 8106	119487570	-----S--N--S-S---IH-ID-IK-LGVPN	K-T---SQSA--Q-VADT-N-	
	Moorea producta 3L	332706581	-----N--NPT---ML--SEMAN-GVSL	-T-T-D--S-SD--H--ADT-T-	
	Trichodesmium erythraeum IMS101	113476169	---L-----S-V---ILA--E-KQLGTTV	-T-T--S-AFRDM-Q-VADT-H-	
	Microcoleus chthonoplastes PCC 7420	254415737	--M-----N--NAT---IQ--QIQDTGEQV	ST-N----I-HV-Q-V-DT-N-	
	Arthrospira sp. PCC 8005	376003335	--IL-----SPV--FIL-VDEIIQSGATV	-T-T-S--PELDI-Q-VADT-N-	
	Arthrospira maxima CS-328	209526937	--IL-----SPV--FIL-VDEIIQSGATV	-T-T-S--PELDI-Q-VADT-N-	
Arthrospira platensis Paraca	284052405	---L-----N--SPV--FML-V-E-K-LGKTV	-T-T-T--AFEDS-Q-VADT---		
Arthrospira platensis NIES-39	291571924	---L-----N--SPV--FML-V-E-K-LGKTV	-T-T-T--AFEDS-Q-VADT---		
Other Cyanobacteria	Acaryochloris marina MBIC11017	158338239	-----T--N--LDA---MQ--RR-EQS-LT-	SRLH-TA-QAKD--A--DT---	
	Acaryochloris sp. CCME 5410	359458099	--VL-----N--FPA---ML-VDE-K-QHAAV	HT-T-A--GQLHV-Q-VADT-L-	
	Cyanothece sp. CCY0110	126659478	--IL---NN--FPA---MLS-DE-Q-SQAVV	-T-T-K--TFK-I-Q--ADT-V-	
	Cyanothece sp. PCC 7424	218442716	---L-----N--DAN--TILII-R-K-QGVRV	---N-L--KRDDVTQ-LTQT-L-	
	Cyanothece sp. PCC 7425	220906228	--IL-----N--SPS----H--RQIQETGTIA	RN-V----EITHV-Q-VADT-HN	
	Cyanothece sp. PCC 7822	307150544	--I-----N--S-T---MY--DQIK-THTR-	-Q-T-G--AI-HV-E-V-DT-L-	
	Microcystis aeruginosa NIES-843	166365021	-F--S--SN-IN-T---A--DK-EQGGII-	-QVI-R--GPA-V-Q--ADT-NH	
	Microcystis aeruginosa PCC 7806	159029461	F-I--S--NN--SLT---IE--DKIKQAKVT-	S--A-Y--QLRDI-Q--ADT-I-	
	Cyanothece sp. ATCC 51142	172039010	---L-----N--SLH--TMIMINR--EQGMI-	-D-V-D--NLEEIEQ-LADS-Y-	
	Cyanothece sp. ATCC 51472	354554545	-FI-----N--DLA--VRL--NDIHNQGVI-	E-LV-NN-TVS--EG--KDTINL	
	Prochloron didemni P1-Palau	335387279	---L-----N--DEN--TMLAIDR-K-QGVTV	-Q-V-T--KP-DVKQ-LAET-YR	

Supplementary Figure 45

Partial sequence alignment for the protein serine/threonine protein kinase and signal transduction histidine kinase, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

			44		98
			LLRMSFTYWL	VYCI	AFTCEKILIS S
			EW	EGLAMTLK	ITTALAYFLT
			FSCILSLPLY		
			-----TV--R-----		
			-F-----L--GI--MILP		---SVV-----S-----V----H
			--T--A-----V--GI--MVL		D---TT----R--S-----H
			'Nostoc azollae' 0708		298491121
			-F---S-----V--GL--LVFP		NY-PMI-----S-----H
			-F---S-----V--GL--MVFP		NY-PMM-----S-----H
			-F---S-----L--GL--LILP		D--TVL-SV-M-I--S-----V----H
			-F--T-C-----V--GI--LVLP		H--AIQ--R-----S-----V----H
			-IK-----I---S-FGL-LHLP		-DSW-LLS--L--V-S-L--A-S-----
			----I-A--I---L-V-LQ-VISP		D-DS-LLS-RL--I-S-----SVIC---H
			-W--I-V-----G-RGLQIAVPL		-GTD-GQMC--AL--FL--V-SL-C--MH
			-FA-I-A--M---ASVAIVRFGL		-G-FVIRS-RA-SV---L--WASV-----H

Supplementary Figure 46

Partial sequence alignment for the hypothetical protein Npun_R4078, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505* and *Raphidiopsis brookii D9*.

		167	218
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282897025	RALVTQPKILLLDQPISALDPVETSWLLEKITE GISD LCPDVKTTVLMVTYQ
	Cylindrospermopsis raciborskii CS-505	282899018	-----R-----E-----R-I-----A-----
	'Nostoc azollae' 0708	298491590	-T--I-----E-TF---GKA-R-IILK- -TQIH--I--NH-
	Nostoc punctiforme PCC 73102	186684036	---LI-----E-T---AGTA-H-MQVLNQ -SQTHQ--I--NH-
	Fischerella sp. JSC-11	354566139	---II--Q-----E-T-T--VGTA-RVMAILNQ -AQRQ-----NH-
	Cyanothece sp. PCC 8802	257061827	-G-IM--QL-----E-T-----DRAHH-MDLIN -SQTQ--IM-INH-
	Cyanothece sp. PCC 8801	218248769	-G-IM--QL-----E-T-----DRAHH-MDLIN -SQTQ--IM-INH-
	Cyanothece sp. CCY0110	126657130	-G-MM--TL-----E-T---RDKALQ-I-VLNT ITTNT-M-----NH-
	Crocospaera watsonii WH 0003	357261403	-G-MM---L---E-T---HDKALQ--VLKT ITIES--A-I--NH-
	Crocospaera watsonii WH 8501	67924625	-G-MM---L---E-T---HDKALQ--VLKT ITIES--A-I--NH-
Other Cyanobacteria	Cyanothece sp. ATCC 51142	172038438	-G-MM--TL-----E-T---RDKALQ-I-VLKT ITT-TQM-I--NH-
	Cyanothece sp. PCC 7822	307152377	-G-MM--QV-----E-T---VGNGNH-M-VLIN -SHSSE-AII--NH-
	Cyanothece sp. PCC 7424	218441302	-G-MMH-Q-----E-T---IGKGNH-I-VLI- VSDTTQ-AII-INH-
	Synechocystis sp. PCC 6803	16330818	--TMLS-R-----E-TA---ALAQRV-TQLKL INQENG--MV--NH-
	Synechococcus sp. PCC 7335	254423214	---I---V-----E-T--Q-VGYSEF--TRLA- -THQN-LAIV-ANH-
	Cyanothece sp. ATCC 51142	172036081	-----E-EVI---E-T-----IA-NSIE-L-SQ LQE-V-MII--H
	Cyanothece sp. CCY0110	126660071	-----E-EVI---E-T-----IA-NSIE-L-SQ LQE-V-MII--H
	Microcoleus vaginatus FGP-2	334118898	-GIIL-----E-T---AGKAAH--QVL-Q LTNG--IV--NH-
	Moorea producta 3L	332709782	----V--R-V---E-T---AGIA-QVIRVF-D -ARKEQI-I--NH-
	Microcoleus chthonoplastes PCC 7420	254412577	---IC---I---E-T---VGTA-HF-GVLAD MANQDQI-----NH-
Oscillatoria sp. PCC 6506	300863593	--VAI--Q-----E-T---AGRSHQV-QVLAQ -ATSDRM-I--NH-	

Supplementary Figure 47

Partial sequence alignment for the ATP-binding protein of ABC transporter, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

			27		46
			AIVAIHNTNLGS	GV	ALGGTRLLPYSTEEDALKDALRLS
			-----	--	-----V----
Nostocales/	Raphidiopsis brookii D9	282897902			
Stigonematales	Cylindrospermopsis raciborskii CS-505	282900699			
	Nostoc punctiforme PCC 73102	186681743	--I--D-T--P		-M-A--M--VN--A-----
	Nodularia spumigena CCY9414	119509380	--I--DRT--P		-M-A--F--VS--A-----
	Nostoc sp. PCC 7120	17227922	--I--D-T--P		-M-A--Y--IN--A--R-----
Other	Chlorogloeopsis sp. Cgs-089	223414368	--I--D-T--P		-M-A--F--VN--A-----
Cyanobacteria	Cyanothece sp. PCC 7424	218439239	--I-----A		-M-A--W--PS-A--R-V----
	Cyanothece sp. PCC 7822	307152271	--I-----		-M-A--W--LS-A--R-V----
	Lyngbya sp. PCC 8106	119485315	--I--DSR--A		-M-A--W--ES-VA--T-----

Supplementary Figure 48

Partial sequence alignment for the protein leucine dehydrogenase, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		365	422
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898976	SGPILNRGAIAWNYQGEFYFGRSLSLNETLIVDQDNK QTS
	Raphidiopsis brookii D9	282896074	-----D-----IN---- K-T
	Nostoc punctiforme PCC 73102	186684009	-----DS-Q-----T-E---TANNQR
	Nodularia spumigena CCY9414	119510862	-----DS-Q-----T-Q---AANNQR
	'Nostoc azollae' 0708	298490069	-----G--DS--Y-----T-----TSNNQP
	Anabaena variabilis ATCC 29413	332705926	-S-----DA-Q-----Q---VTSSNLR
	Nostoc sp. PCC 7120	17231835	-S-----DA-Q-----Q---ATSSNLR
	Fischerella sp. JSC-11	354566106	-S-----DS-Q--MD--A-L-N--TPNNO-
	Moorea producta 3L	332705926	-----DR-HVKL-----T---VTSNQKR
	Microcoleus chthonoplastes PCC 7420	254414756	-----TD-HQ-K---F--Q---TANGER
	Microcoleus vaginatus FGP-2	334117203	-----DR--AID--T-Q--V-TPANQP
	Oscillatoria sp. PCC 6506	300866383	-----DN-DIKI----Q-S--TSKGER
	Arthrospira sp. PCC 8005	376005801	-----M--DNN-RVRM---K-V---VSSTGQR
	Lyngbya sp. PCC 8106	119485450	-S-----G-TD--I-LD--TRF---TATG-R
	Other Cyanobacteria	Arthrospira platensis NIES-39	291567428
Arthrospira maxima CS-328		209523066	-----M--DNN-RVRM---K-V---VISTGQR
Arthrospira platensis Paraca		284053738	-----M--DDT-RVRM---K-V---IVTSTGQR
Cyanothece sp. PCC 7424		218437310	-----DR-QVKM---R-Q--V-TNGG-R
Cyanothece sp. PCC 7425		220909195	-S-----G--D---TL---R-QQ---TASGQS
Cyanothece sp. PCC 8802		257059905	-----G-----DE-RMKI---WQ---VTSSGQR
Synechocystis sp. PCC 6803		16331087	-----D---QTT-----S-IITGSGQR
Cyanothece sp. PCC 8801		218246863	-----G-----DE-RMKI---WQ---VTSSGQR
Cyanothece sp. PCC 7822		307153597	-----G-----EK--VKI---R-Q---TTNNG-Q
Cyanothece sp. CCY0110		126660043	-S---G---G---EK-QVFMN--A-Q---TATG-Q
Cyanothece sp. ATCC 51142		172038643	-S---G---G---DN-QV-LN--A-Q---LTRAG--
Acaryochloris marina MBIC11017		158338832	-S-----VV---P--Q-QM---N-QQV-STSGGKR
Acaryochloris sp. CCMEE 5410		359460708	-S-----VV---P--Q-QM---N-QQV-STSGGKR
Synechococcus sp. CC9605		78211877	-----V---GAS-DLQ---R--Q--R-NNGRR
Synechococcus sp. WH 8109		260436474	-----V---GSS-DLQ---R--Q--Q-NNGRR
Synechococcus elongatus PCC 7942		81299698	-----M---D---VLID--G-Q---QLSSGTR
Synechococcus elongatus PCC 6301		56750660	-----M---D---VLID--G-Q---QLSSGTR
Crocospaera watsonii WH 8501		67921288	-S---G---G---DDK-QV-LN--A-R--ITTATG-N
Crocospaera watsonii WH 0003		357260041	-S---G---G---DDK-QV-LN--A-R--ITTATG-N
Synechococcus sp. WH 8102		33866665	-----VV--GDNDQLQ---R-DQQ-Q-NGGRR
Microcystis sp. T1-4		390438749	-----DNR-KIRID---E---TATGQR
Synechococcus sp. CC9311		113953681	-----G-DRN-PLI---Q-IQDMT-AGRNR
Synechococcus sp. CB0205		317968697	-----VG-GDGELPQ-D--R-Q-SVLDPRQRR
Synechococcus sp. WH 8016		352096473	-----G-KRS-PLM---Q-IQEMT-FGRRR
Synechococcus sp. WH 5701		87301861	-----G-QPG-LPS---A-Q-Q--DERGQR
Microcystis aeruginosa NIES-843		166368177	-----DNR-KIRID---E---TATGQR
Synechococcus sp. PCC 7335		254424407	-----G--M--DS--LVID-FA-S-SVTTGVGEA
Synechococcus sp. WH 7803		148238720	-----VG-DGP-PLL-D--R-DQE-R-NGGRR
Synechococcus sp. BL107		116072135	-----V---SR-SPLQ---E-LQE-Q-NSSKR
Synechococcus sp. RS9917		87125017	-----G-DPGERLA---R--QE--LSGGER
Synechococcus sp. WH 7805		88807939	-----VG-DGP-PLL-D--R-DQEMR-NGGRR
Cyanobium sp. PCC 7001		254430401	-----VA---GREVPS---L-E-WV-GPDRAR
cyanobacterium UCYN-A	284929601	-S---G---VG--EN-K-FIDH--K-F--LNNGER	
Synechococcus sp. CC9902	78185584	-----V-G-SRNSTLH---E-LQE-Q-NSSKR	
Thermosynechococcus elongatus BP-1	22299393	-----G-DDR-QIVV---QQRVRTPTGTV	
Synechococcus sp. RS9916	116075716	-----VG-SNDNRLQ---R-YQE-R-EGGR-	
Synechococcus sp. RCC307	148243264	-----D-----AP-ELPR-S-VR-----S-AGRR	
Synechococcus sp. CB0101	318042648	-----AG-KAGELPR-D--M-V--IEDSNRQR	
Synechococcus sp. PCC 7002	170078987	-----V--DNA-RWE-D--KVETDIVAGNGER	
Prochlorococcus marinus MIT 9313	33864045	-----G-GPT-HLS---Q-QQE-QGSKGQR	
			LPVLFLLNSGYVQNGIARYT
			---S-----S-----
			--I--V-----S-----
			--SI-----
			V-I-A-----
			V-I-A-----
			--I-T-----S-----
			-----KA-----
			F-S-----A-S---
			---SH---A---S---G---
			-AIAH---A---A---G---
			FE-DL--T---KA-----
			F-IQH-----EA-S---
			FE-DL--T---KA-----
			FE-DL--T---KA-----
			FEIDL--T---KA-----
			---Y-----S-M---
			--INT-D--F--K-----
			--IR-----EG-M---
			-TANY-----R-----
			--IR-----EG-M---
			I-IGLV---I-G-M---
			V--S-----I-A-VS---
			I-IA-----I-A-VS---
			-SIVS-D---P-K-----
			-SIVS-D---P-K-----
			WSLMA-----K-LSL--
			WSLMA-----K-LSL--
			I-LVA-----RA-A---
			I-LVA-----RA-A---
			I-IA-----IKA-VS---
			I-IA-----IKA-VS---
			RGLSY-----R-LS---
			F-LTH---AF-AA-CS---
			W-LGM-----R-LS---
			WSLIS-----K-L---
			W-LGM-----R-LS---
			W-LSS-----R-L---
			F-LTQ---AFLTA-CS---
			F-I-AV---KA--G---
			WGLG-----R-LS---
			WRLG-----K-LS---
			WGLGL---F--R-LS---
			WGLG-----R-LS---
			--IVVV-----R-VS---
			ISIQS-----I-A--S---
			WRLG-----K-LS---
			-IVTF-----A-L-L--
			WGLGY-----R-FS--D
			IQLNAI--W-SK-V-Q-N
			WT-TSV-----K-L---
			VG-ELI-----KA-A-L-D
			WNLET-----R-LS---

Supplementary Figure 49

Partial sequence alignment for the hypothetical protein Npun_F3878, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		86	140
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282897472	AAKVVVEKLG YHQV KSLHHGESFT LEDTLN NN LKDQLEIKALPGSPVGPNNVRENGY
	Cylindrospermopsis raciborskii CS-505	282901303	-----Y---T-----ST-----
	'Nostoc azollae' 0708	298489622	-----KNLIT-N---T--
	Anabaena variabilis ATCC 29413	75908513	-----QA---KS-TT-A-----
	Nostoc sp. PCC 7120	17232529	-----QA---TS-TT-V-----
	Nostoc punctiforme PCC 73102	186686130	-----QQ---T---TT-A---T--
	Nodularia spumigena CCY9414	119509796	-----LLPG---TS-T--A--AT-N
	Fischerella sp. JSC-11	354569217	-----QQ---T---ITV-T---T--
	Cyanothece sp. PCC 8801	218247517	-----KA---T---T---S---Y-
	Cyanothece sp. PCC 7822	307151756	-----KA---T---TT-D-HQ-Y-
	Cyanothece sp. PCC 8802	257061147	-----KG---T---T---S-W--Y-
	Crocospaera watsonii WH 8501	67923409	-----E---TDIHT-S---TY-
	Crocospaera watsonii WH 0003	357262751	-----E---TDIHT-S---TY-
	Synechocystis sp. PCC 6803	16331276	--E-ATE---ET-TG-P-NQK-V
	Cyanothece sp. PCC 7425	220909022	-----S---S-II-Q-Q-Y-
	Cyanothece sp. PCC 7424	218438329	-E---KG---T---IT--A-HQTY-
	Cyanothece sp. CCY0110	126657224	-T---KE---TDIRT-E--D-Y-
	Microcystis aeruginosa PCC 7806	159028625	-E---RA-N-THLIP-T--S-YI
	Acaryochloris marina MBIC11017	158338721	-T---E---TD-TT-A--T--
	Microcystis aeruginosa NIES-843	166364264	-E---RA-N-THRIP-N--S-YI
Synechococcus sp. PCC 7002	170079005	-----S---QTIHT-E-HQKY-	
Cyanothece sp. ATCC 51142	172038189	-T---KE---TDIRT-D--D-Y-	
Acaryochloris sp. CCME 5410	359458502	-T---E---TD-TP-A--T-A	
Synechococcus elongatus PCC 6301	56750742	----AR---FS--TV-SP--V-	
Synechococcus sp. JA-2-3B'a(2-13)	86609848	--R--KS--FAT-QA-DPWQ-VR	
Synechococcus sp. JA-3-3Ab	86606883	--R-ARS--FMR-QA-DPWQ-VQ	
Thermosynechococcus elongatus BP-1	22299976	--N-ARS--FET-I--SPHQTH-	
Cyanobium sp. PCC 7001	254430249	--RR-RQ--FQH-TA-AP--HS	
Synechococcus sp. WH 7805	88809046	----AKR--FTS-QT-NP---TE	
Synechococcus sp. WH 7803	148239999	----AKR--FTA-DT-NP---TE	
Synechococcus sp. RCC307	148243398	-VQQARRF-FT-SET-RP--RLQ	
Synechococcus sp. WH 8016	352093529	--N--NR--FE-IET-KP--TC-	
Synechococcus sp. RS9916	116073205	----A-G--FSD-VA-RP-DTHA	
Synechococcus sp. BL107	116071904	-TQ--K-M-FES-QA-KP--CT-	
Synechococcus sp. RS9917	87125450	----AKR--FSS-QA-KP-DRTE	
Synechococcus sp. CC9605	78214071	--R--DR--FTS--A-SP--RTN	
Synechococcus sp. WH 8109	260435064	-TL---R--FTS--A-SP---TN	
Synechococcus sp. CC9902	78185788	-TQ--K-M-FET-QA-KP--CT-	
Synechococcus sp. CC9311	113953607	----KR--FEHIEP-KP--KC-	
Synechococcus sp. CB0205	317969612	---K-QQM-FE--TG-RP--TTA	
Synechococcus sp. WH 5701	87302528	--T-RQ--FQA-N--SP--RLH	
Synechococcus sp. WH 8102	33866932	--R--RS--F-TITT-KP--TTN	
Synechococcus sp. CB0101	318041068	--AK-RG--FET-TA-RP--QCR	
Arthrospira sp. PCC 8005	376001709	----RE---T--H--N---T-Q	
Arthrospira platensis Paraca	284053725	----RE---T-IH---R-T-K	
Lyngbya sp. PCC 8106	119489333	----AQE--FT--RA-E--STYF	
Arthrospira maxima CS-328	209523121	----RE---T--H--N---T-K	
Trichodesmium erythraeum IMS101	113478005	----QE---S--IT-E--HTYA	
Microcoleus chthonoplastes PCC 7420	254416854	--T--KE---T--TP-AF-QT--	
Moorea producta 3L	332712006	----KQ---T--TP-TP--TVA	
Microcoleus vaginatus FGP-2	334118565	----LSS---TNIT--ANWQELN	
Oscillatoria sp. PCC 6506	300866095	-T--LSS---KSIA--SNWQE-I	
Prochlorococcus marinus NATL1A	124025793	-SN--S-I-FTEI-T-RP--A-K	
Prochlorococcus marinus MIT 9303	124022299	---L-GR--FK--TP--P---TN	
Prochlorococcus marinus CCMP1375	33240637	-GR-L---FSNKT--KP-Q-KN	
Prochlorococcus marinus AS9601	123968621	-VETLK-I-FSTI-V-KPT-KTN	
Gloeobacter violaceus PCC 7421	37521903	--R-LRE-QFSS-QA-TNF-Q-R	
Other Cyanobacteria			-NN-V---F---I--T-L----
			FNN-I--R-F---I--T-V--S-
			FNN-I--R-F---I--T-V--S-
			-NN-V---F---I--TLV--S-
			-NQ-V--T-V---T--Y-LV----
			-NTSV---F---I--QLV-----
			-DNAIS---V---L---QLI----
			-ANS-----V---I--QLI----
			-DNAIS---V---L---QLI----
			-EET-A-----LI--QLV--A-
		-N-RV--L-----I--TLV--A-	
		-NH--T-Q-TV---L--Q-V-----	
		FANR-Q---V---I--QLV----	
		-DETID---F---L---QLV--A-	
		FDNAI--T-V---I--TLV----	
		WA-R-QVQ---I--FLV--AL	
		FDNAI--T-V---I--TLV----	
		I-N--L--K-----TTL----	
		-DETVHF--F---L---QLV--A-	
		WA-R-QVQ---I--FLV--AL	
		IGN--LQ-TK-A--Q- ----	
		WGD-RVT-I--A-L--L--L-F	
		WGD-G-T-V--A-L--L--L-F	
		YRD-T-Q-TK-ASI--TQQ----	
		LGD-G-R-TA-A--Q- ----	
		RGP-Q-R-TA-AA--A- ----	
		RGP-A-R-TA-AA--A- ----	
		RGS--Q-TA-A--Q- ----	
		IKG-S-E-TA-AA-- ----	
		HH--N-Q-TA-AR-PLV----	
		HKG-SVR-TS-A--T- ----	
		RGA-R-R-TA-AA--A- ----	
		HQG-QVR-SA-A--M- ----	
		HQG-QVR-SA-A--M- ----	
		HKG-TVR-TA-A--T- ----	
		VKG-N-E-TA-AA-- ----	
		FGE-R-E-TA-A--Q- ----	
		HGE-T-T-SA-A--Q- ----	
		HQG-TVR-TA-A--M- ----	
		HGE-Q-T-TA-A--Q- ----	
		-GASI--R-V---I--TLV----	
		-GSS--R-V---I--TLV----	
		-DN-V---V---I--TTV----	
		-GAS--R-V--F-I--TLV----	
		FENSI--R-V---L---TLV----	
		-A--I--R---A-I--FLQ--A-	
		-GNH--R---A-I--FQQ--A-	
		F-EK-Q-T-V--AEI--GQE----	
		Q-K-Q-I-V--AEIQ-GQV----	
		-NNIN-Q-TS-AS--I ---	
		IAG-T-R-TA-AM--SI ---	
		L--TF-E-TE-AA--GL ---	
		LFN-SFE-TA-A--QI ---	
		LGN-RVT-V--AE-QFE Q---	

Supplementary Figure 50

Partial sequence alignment for the protein Zn-dependent hydrolase (beta-lactamase fold protein), showing a 8 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		214	283
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i> CS-505	282899881	RDPLFGGKPP-EPLPKYISEISAEIKN YTE IHHQQRNSTLTVGLVFDGDDGRIAA
	<i>Raphidiopsis brookii</i> D9	282897077	-----C----- --Q--F-K-----
	' <i>Nostoc azollae</i> ' 0708	298489641	-----E-GA-----L-RLF---T -GES-N-G-----C----
	<i>Nostoc</i> sp. PCC 7120	17232581	-----GA-----L-LLFET-R- -RET-P-S-S-----C----
	<i>Anabaena variabilis</i> ATCC 29413	75908584	-----GA-----L-LLFET-RK -RET-P-S-S-----C----
	<i>Nostoc punctiforme</i> PCC 73102	186684582	-----GA-----L-KLFEV--T -RETDK-G-----C----
	<i>Nodularia spumigena</i> CCY9414	119513312	-----E-GA-----L-RLFST--T YR-TDK-S-A-----C----
	<i>Fischerella</i> sp. JSC-11	354568900	---T-E-GA-----L-RLFGVM-T -REH-PTG---A-----C--V--
	<i>Cyanobacterium</i> UCYN-A	284923674	-----N- ----C-LTKLLTTV-K FNAEKS-G-SI-----C----
	<i>Cyanothece</i> sp. ATCC 51142	172035686	-----E-GA----R-L-KLFTS--T AAS-GNNS-R-----S----
	<i>Cyanothece</i> sp. CCY0110	126658857	-----E-GA----R-L-KLFNN--T AAS-GDDS-R-----S----
	<i>Crocospaera watsonii</i> WH 0003	357262623	-----E-GA----R-L-KLFLT--T AIS-GDNS-K-----S----
	<i>Synechococcus</i> sp. PCC 7822	307153357	-----Y- ----R-L--LLETV-T APGEG-K-----S----
	<i>Crocospaera watsonii</i> WH 8501	67923543	---CE-GA----R-L-KLFLT--T AIS-GDNS-K-----S----
	<i>Cyanothece</i> sp. PCC 8801	218246599	-----E-GA----R-L-KLFEA--T AASEGTING-R-----S----
	<i>Cyanothece</i> sp. PCC 7424	218441826	-----GQ-P---R-L-DLL-TV-A APGDG-K-----S----
	<i>Synechocystis</i> sp. PCC 6803	16330692	-----GASP-----VPKLL-ALQT AAKAGGPE-R---I---A----
	<i>Cyanothece</i> sp. PCC 7425	220907634	-----GGAP--IAR-LQSTC-TLHQ FR-THPDQ-S-----S----
	<i>Microcystis aeruginosa</i> NIES-843	166367171	-----GGSP----RNLR--IDKLAH SA-LAP-R-----S--V--
	<i>Microcystis aeruginosa</i> PCC 7806	159026931	-----GGSP----RNLT--IDKLAH SA-LAP-RL-----S--V--
	<i>Acaryochloris marina</i> MBIC11017	158337679	-----E-GA P-----LGKSIQA-QS HRQ-APQEEAI--F-----S----
	<i>Synechococcus</i> sp. PCC 7335	254425346	A-----A-GA P-----PNLFEA--K NQTING----I---S---G
	<i>Acaryochloris</i> sp. CCMEE 5410	359459165	-----E-GA P-----LEKSIQA-QS HRQ-APQEEAI--F-----S----
	<i>Synechococcus</i> sp. RCC307	148243397	-----GH-P---AP-L--LI-QV-A STAAGQPA--I-----S----
	<i>Synechococcus</i> sp. JA-3-3Ab	86604742	-----GF-P---QNLQTSAQI-AA DPAP-KM-----A----
	<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	86607515	-----GF-P---QNLQASAQI-AA DPAP-KM-----A----
	<i>Synechococcus</i> sp. PCC 7002	170078101	P----GR-P---E-HL-QLQRT-RA A-N-DLEAIQ--F-----S----
	<i>Synechococcus</i> sp. WH 8109	260436430	-----GH-P---AR-LGALI-AV-A STAAG-PA-----S----
	<i>Synechococcus</i> sp. WH 8102	33864771	-----GN-P---AP-L-DLITAVQH STAAG-PA-----S----
	<i>Synechococcus</i> sp. CBO101	318042767	-----GN-P---AP-LQ-LI--VRA STLAGRPA--I-----S----
	<i>Synechococcus</i> sp. CBO205	317968600	-----GN-P---AP-LQQLI--VRA STLAGRPA--I-----S----
	<i>Synechococcus elongatus</i> PCC 6301	56750292	-----GGAP---AANVVLGCRQLSD GVTTTP-A--F-----S----
	<i>Synechococcus</i> sp. WH 8016	352095129	-----GCAP---AP-LG-LI-AL-D SKAAGRDA-----S----
	<i>Synechococcus</i> sp. RS9917	87123048	-----GH-P---AP-LGDLI-AV-Q ATAAGQHA-----S----
	<i>Synechococcus</i> sp. RS9916	116075584	-----GN-P---AP-LQ-LITAV-A AGSSGQPA-----S----
	<i>Synechococcus</i> sp. CC9605	78211780	-----GH-P---AP-LGGLITAV-A STAAG-PA-----S----
	<i>Synechococcus</i> sp. BL107	116071550	-N---GN-P---AP-LQ-LI-AVRA STAAG-PA-----S----
	<i>Synechococcus</i> sp. CC9902	78183839	-N---GN-P---AP-LQ-LITAVRA STAAG-PA-----S----
	<i>Synechococcus</i> sp. CC9311	113954476	-----GCAP---APHLG-LI-AV-A SKAAGRDA-----S----
	<i>Cyanobium</i> sp. PCC 7001	254432700	-----GN-P---AP-LQ-LITAVRA STAAGQPPAM-I-----S----
	<i>Thermosynechococcus elongatus</i> BP-1	22299063	-----SGAP---IGR-LAATQE RLR --HSA-P--C-----A--L--V
	<i>Synechococcus</i> sp. WH 5701	87300706	-----GH-P---AP-LQ-LI-AVRA SGEAGRPA--I-----S----
	<i>Synechococcus</i> sp. WH 7805	88808048	-----GC-P---APHLKQLI-AVQA STQSGHPA-----S----
	<i>Synechococcus</i> sp. WH 7803	148238614	-----GC-P---APHLQQLI-AVQT STQAGQPA-----S----
	<i>Arthrospira platensis</i> Paraca	284053188	-----GGAP---R-L--LFRQ--S -TPKANGV-----S----
<i>Arthrospira platensis</i> NIES-39	291570817	-----GGAP---R-L--LFRQ--S -TPKANGV-----S----	
<i>Arthrospira maxima</i> CS-328	209524936	-----GGAP---R-L--LFRQ--S -TPKANGV-----S----	
<i>Trichodesmium erythraeum</i> IMS101	113474765	---F- DGGAP---R-L-KLFKR--S YRAA-EGS-AI-----S----	
<i>Microcoleus chthonoplastes</i> PCC 7420	254409746	-----GGAP---R-L-QLFRQ-RT -RR-TEG--R-----S----	
<i>Oscillatoria</i> sp. PCC 6506	300864526	-----GGAP---R-L--LFRV-RT --RAEGG-AI-F-----S----	
<i>Lyngbya</i> sp. PCC 8106	119488420	-----DGGAP-----P-LFRQ--S -KRSQDS-A--F-----S----	
<i>Microcoleus vaginatus</i> FGP-2	334118948	-----GGAP---R-L-QLFRVMRT --RKS-G-SI-F-----S----	
<i>Moorea producta</i> 3L	332708999	-----GKSP---G-L--LFEALQS -REPGVE-A--A--S-----S----	
<i>Gloeobacter violaceus</i> PCC 7421	37519957	H----GT-P--IGR-LMPLFQAVRD YRGNRPV--AL--A-----S----	
<i>Prochlorococcus marinus</i> MIT 9211	159902846	-----GN-P---SQ-LGQLISSV-A SSAAG-QSL-----S----	
<i>Prochlorococcus marinus</i> NATL1A	124025071	---S- GN-P---KA-L-QLIQ-VQD ESQAGK-SM-----S----	
<i>Prochlorococcus marinus</i> CCMP1375	33239762	-----GNAP---A-NLDQLIVNTQE LALEGHPVL--A-----S----	
<i>Prochlorococcus marinus</i> AS9601	123967838	A--F- -GK-P---LN-ADDLQKTLMK NSTNEVK-L-II-----S----	

Supplementary Figure 51

Partial sequence alignment for the protein phosphoglucomutase, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		282897719	323	364
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282897719	YVYIIVIPLLGFATIKLLQR N	VVFNSKIQAGKRVQNSQCINC
	Cylindrospermopsis raciborskii CS-505	282902014	----V----F-----	-----
	Nodularia spumigena CCY9414	119510183	----L-----V--Q-F-K	-IL-T-A--VR---D-R----
	Nostoc punctiforme PCC 73102	186684759	-A--LL--II--G--QFF-K	----A-V--AS---K-R-V--
	Anabaena variabilis ATCC 29413	75907614	----LL---I--G---FF-K	I-L-P-L--VN-I-K-R-VK-
Other Cyanobacteria	Nostoc sp. PCC 7120	17231927	----LL---I--G---FF-K	I-L-P-L--VN-I-K-R-VK-
	Microcystis aeruginosa NIES-843	166365728	----LL---F-----F--K	F---T-S--A----K-----
	Microcystis aeruginosa PCC 9432	389675958	----LL---F---V--F--K	F---T-S--A----K-----
	Synechocystis sp. PCC 6803	16331877	-L--LL---V--GV--FF--	-----S--A--F-K-S-VK-
	Synechococcus sp. WH 5701	87303343	-L-LLAV--V---L-----A	L---PRL--AE---RGR-LH-
	Synechococcus elongatus PCC 6301	56750162	-L--FL--A--G---A-A	WI--PR---AN---K-R--K-
	Synechococcus sp. PCC 7335	254424906	-LQ-LL--AI--GV--FF--	F---TRL--AN---SGR-V--
	Cyanothece sp. PCC 7425	220909504	-LL-VI-----GL--F---	FI--PQV--KG-IRK-L--Q-
	Thermosynechococcus elongatus BP-1	22299540	---LLV-AI-LGM--VA-V	LFL-PQR--AG-I-KQR-VR-
	Microcoleus chthonoplastes PCC 7420	254411895	-A--LL---I--G---FF-K	I---P---AR---K-R--R-
Moorea producta 3L	332707815	----LF---I--G---FF-K	I---P-L--AS--KKLL----	

Supplementary Figure 52

Partial sequence alignment for the hypothetical protein N9414_10373, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505* and *Raphidiopsis brookii D9*.

Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282901339	107	QGMKLLYCDRQTYRQRHTQEVQEEFKKQL PRSS	162	NETFIIPEGGGLNGVRCMEIM
	Raphidiopsis brookii D9	282896290		-----H-----		-Q-----A-----
	'Nostoc azollae' 0708	298489927		---Q-V--N-T-----TQL--DL-QRF		D-V-----C---M---T---
	Nodularia spumigena CCY9414	119511926		---Q-V-LN-E---K--K-L--DL-QRF		G-V-----S-----T--V
	Anabaena variabilis ATCC 29413	75907595		---Q-V-V--E-----NSATLHADLR-RF		G-V-----S-----L--I
	Nostoc sp. PCC 7120	17231939		---Q-V-V--E-----NSTLHTDLR-RF		G-V-----S-----L--I
	Nostoc punctiforme PCC 73102	186681102		---Q-V-TN-EM----N-PV-E-YLQGRF		G-V-----S-----I

Supplementary Figure 53

Partial sequence alignment for the protein 1-aminocyclopropane-1-carboxylate deaminase, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		209	257
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i> CS-505	282901667	PIMLDFYLENAPLHTGAQE NLEGEEG NVGNSIGNWRIRCTINGESFVLD
	<i>Raphidiopsis brookii</i> D9	282896254	-----G-----S-----
	<i>Nodularia spumigena</i> CCY9414	119509212	--L----T----LV-ED -PNDTVSD-----D---I--
	<i>Anabaena variabilis</i> ATCC 29413	75908620	--L----T----LV-K- LPNDQFSD-----F-
	<i>Nostoc</i> sp. PCC 7120	17232645	--L----A----LA-K- QSNDAQFSD-----F-
	<i>Nostoc punctiforme</i> PCC 73102	186683948	--L----T----LVDK- -TNEGFSD---V---D---F-
	' <i>Nostoc azollae</i> ' 0708	298489998	-----T----IA-ED -PDDT-SD-----I--
	<i>Fischerella</i> sp. JSC-11	354565183	--L----T----LL-KD -PDNPFAD-----
	<i>Cyanothece</i> sp. PCC 7822	307154771	--L----T----LI--- -PEDE-AD---V-----L--
	<i>Cyanothece</i> sp. PCC 7424	218437411	-----T----LI--- -TEDD-VD---VIV---L--
	<i>Cyanothece</i> sp. CCY0110	126660056	-----T----FV--S DPNDE-VD---KA-----LI-
	<i>Cyanothece</i> sp. PCC 8801	218246997	-----T----QV--A SNDDSVAD---A-----L-
	<i>Cyanothece</i> sp. PCC 8802	257060072	-----T----QV--A SNDDSVAD---A-----L-
	<i>Cyanothece</i> sp. ATCC 51142	172038656	-----T----FI--S DPNDE-ID---KA-----LI-
	<i>Crocospaera watsonii</i> WH 8501	67920985	-----T----LV--S DPNDD-VD---KA-----N-LI-
	<i>Synechococcus</i> sp. PCC 7002	170078166	-----T----FI--A DNSDDVND---I-----I--
	<i>Synechocystis</i> sp. PCC 6803	16331025	--L--Y--T----IA-TN DADLSD---V-V-D---LV-
	<i>Synechococcus elongatus</i> PCC 6301	56752052	-V-----T----SL--A DDE-ED-Q--V-V----RTE
	<i>Synechococcus elongatus</i> PCC 7942	81300859	-V-----T----SL--A DDE-ED-Q--V-V----RTE
	<i>Cyanothece</i> sp. PCC 7425	220910028	--L----T----QV--- -PNDE-RD-Q--V-----S--EV-
	<i>Microcystis aeruginosa</i> PCC 7806	159030974	--L----S----RLSNTA -GDN-LQD---V-V-----L--
	<i>Synechococcus</i> sp. PCC 7335	254424462	-LL----T----FL--- DAEEELL-D-K--GIV--KD--I-
	<i>cyanobacterium UCYN-A</i>	284928785	-----IS-LSSDPLKNS -INKKID--VKA-----A-I-N
	<i>Microcystis aeruginosa</i> NIES-843	166367496	-LL----S--S--RLPNTA -GDNSLQ---V-V-----L--
	<i>Synechococcus</i> sp. JA-3-3Ab	86605924	--LV--W-A--VRESLLT DLPRD---Y-L--Q-GL--
<i>Thermosynechococcus elongatus</i> BP-1	22298654	-----S--L--DQLFNN EKV-QV-----Q--TV-	
<i>Acaryochloris marina</i> MBIG11017	158335606	TVLI--F-S-V-ETGNSG --KAQVSV--SK-RI--	
<i>Acaryochloris</i> sp. CMEE 5410	359457798	TVLI--F-S-VSEAGPNA- --QAQVSV--SQ-RI--	
<i>Synechococcus</i> sp. CBO205	317968702	-VLI--WL-ID--QHLRDD DAR--L-L-VSV--D--LV-	
<i>Synechococcus</i> sp. CBO101	318042643	-VLI--WL-LD--QNLRGD DAR--L-L-ISO--D--LV-	
<i>Synechococcus</i> sp. WH 5701	87300872	-VLI--WL-ID--QRLDGGSH--L-V-L-D--D--QV-	
<i>Cyanobium</i> sp. PCC 7001	254431427	-VLI--WL-RD--QGLREG DDG--L-I-V--D--LV-	
<i>Synechococcus</i> sp. WH 7803	148238725	-LF--WLIW---QNLREG DAR--L-L-ISO--D--LV-	
<i>Synechococcus</i> sp. WH 8102	33866657	-L--WLIW---QNIRDD-R---ISVD-D--LT-	
<i>Synechococcus</i> sp. CC9605	78211882	-LLV--WLIW---QNLREG D-R--L-ISO--D--LV-	
<i>Synechococcus</i> sp. BL107	116072159	-LL--WIIW---QNLRDGD-R--L-LSVD-D--LV-	
<i>Synechococcus</i> sp. RS9917	87125008	-LL--WLWV---QNLREG D-R--L-L-L--D-VL--	
<i>Synechococcus</i> sp. WH 7805	88807929	-LL--WLIW---QNLREG D-R--L-ISO--D--LV-	
<i>Synechococcus</i> sp. WH 8109	260436252	-LLV--WLIW---QNLREG D-R--L-ISO--D--LV-	
<i>Synechococcus</i> sp. CC9902	78185563	-LL--WIIW---QNLRDGD-R--L-ISO--D--LV-	
<i>Synechococcus</i> sp. RS9916	116075722	-LL--WLWV---QNLREG D-R--L-L-V--D--LV-	
<i>Synechococcus</i> sp. RCC307	148243258	AVPINWF-F---QHLRDGD EQ--L-LSLD-S-V---	
<i>Synechococcus</i> sp. CC9311	113954099	-LL--WLWV---QNLRAG DAR--L-L-I-V-ED--V-	
<i>Synechococcus</i> sp. WH 8016	352096478	-LL--WLWV---QNLRAG DAR--L-L-I-V-ED--V-	
<i>Microcoleus chthonoplastes</i> PCC 7420	254417164	--L--Y--T----LV--- -PEDE-VD-Q--V-----T--	
<i>Lyngbya</i> sp. PCC 8106	119486206	-VL-----A----LV--- -PEDE-VD---V-V--D--I-A	
<i>Oscillatoria</i> sp. PCC 6506	300866916	--G-----T----LV--- -PQDE-VD-K--V-----LT-	
<i>Arthrospira platensis</i> Paraca	284053388	--L-----D----LI--- DPEDE-ID---V-V-----IIE	
<i>Arthrospira platensis</i> NIES-39	291567741	--L-----D----LI--- DPEDE-ID---V-V-----IIE	
<i>Microcoleus vaginatus</i> FGP-2	334116912	--L-----T--FP-EV--- -SQDE-LD-K--A-V-----AI-	
<i>Trichodesmium erythraeum</i> IMS101	113478050	-----Y--T--IQ-PSIGG YPQDE-TDGH-L-V-V--T--IT-	
<i>Arthrospira</i> sp. PCC 8005	376007874	--L-----D----LI--- DPEDE-ID---V-V-----IIE	
<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	86607587	--LV--W-A---VRESLLS DLPRD---Y-L--Q-GL--	
<i>Prochlorococcus marinus</i> MIT 9313	33864038	-LL--WLIW---IQNLKEG D-R--LQVSV-----QIT	

Supplementary Figure 54

Partial sequence alignment for the hypothetical protein Aazo_0549, showing a 7 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		199	243
	Cylindrospermopsis raciborskii CS-505	282900458	QFIADNIEEQLKHIPFVG
	Raphidiopsis brookii D9	282896745	DELERVTEGMKRLAMSKLGAVFEEELGF
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298492661	-----
	Anabaena variabilis ATCC 29413	75910729	--LS-G-----ESIS P---IK-----VP-V-----
	Nostoc sp. PCC 7120	17228095	--K--F--F-----ESLS P--G-IK-----VP-V-----
	Nostoc punctiforme PCC 73102	186685520	--K--F--F-----ESLS P--G-IK-----VP-V-----
	Nodularia spumigena CCY9414	119510231	--K--FQ--F-Q----ESLS P--G-IK-----VP-V-----
	Fischerella sp. JSC-11	354567520	--FK-SF--F-Q----ESLS P--G-IK-----VP-V-----
	Microcoleus vaginatus FGP-2	334119058	--LK--L--F-----ETFT P-M--LK-----V--V--I-----
	Oscillatoria sp. PCC 6506	300867681	--LK--L--F-----ETFT P-M--LK-----VP-V--I-----
	Arthrospira sp. PCC 8005	376004243	--LK--LR--F-Q----ENFT P-M--LK-----VP----II-----
	Arthrospira platensis Paraca	284049961	--LK--LR--F-Q----ENFT P-M--LK-----VP----II-----
Arthrospira maxima CS-328	209528155	--LK--LR--F-Q----ENFT P-M--LK-----VP----II-----	
Lyngbya sp. PCC 8106	119490660	--LK--LR--F-Q----ETFT P-M--LK-----VP-V--I-----	
Trichodesmium erythraeum IMS101	113476604	--LQ--L--F-Q----ETFT P-M-GLKG-----V--V--I-----	
Moorea producta 3L	332712209	--T--L--F-NL----ESLS PD-Q-LK-----VP-V-----	
Microcoleus chthonoplastes PCC 7420	254415673	--L--L--F--F-LF--ESLT P--Q--K-----VP-V--L-----	
Microcystis aeruginosa PCC 7806	159030454	-----L-----L--F--DSLTL P-M--K-----VP-V--I-----	
Cyanothece sp. PCC 8802	257058388	--L--L--F--L--F--DSLTL P-M--K-----V--V--I-----	
Cyanothece sp. PCC 8801	218245352	--L--L--F--L--F--DSLTL P-M--K-----V--V--I-----	
cyano bacterium UCYN-A	284928885	--S-----F--L--F--DSLTL P-M--LR--V-----P-V--I-----	
Crocospaera watsonii WH 8501	67920284	--LS--L--F--L--F--ESLTL PDM--K-----P-V--I-----	
Thermosynechococcus elongatus BP-1	161485769	--T--L--F-----F--ENLTL P-MQ-LK-----VP-V-----	
Synechocystis sp. PCC 6803	16329681	--LT--L--I--L--F--DSLTL P-M--K-----VVP-V--I-----	
Cyanothece sp. ATCC 51142	172036316	--LS--L--F--L--F--ESLTL P-M--K-----P-V--I-----	
Cyanothece sp. CCY0110	126660199	--LS--L--F--L--F--ESLTL P-M--K-----P-V--I-----	
Synechococcus sp. PCC 7002	170077789	--LT--L--V--L--L--DSLTL P-M--LKDS-----V--V--I-----	
Microcystis aeruginosa NIES-843	166369006	-----L--F-NL--F--DSLTL P-M--K-----VP-V--I-----	
Acaryochloris sp. CCMEE 5410	359461735	--LS--F-----QL--EQVA PD-K-LKG-----V--V-----	
Acaryochloris marina MBIC11017	158338290	--LS--F-----QL--EQVA PD-K-LKG-----V--V-----	
Cyanothece sp. PCC 7822	307150889	--LS--L--F--L--F--D VT P-MD-LK-----VP-V--I-----	
Cyanothece sp. PCC 7424	218441470	--LS--L--F--L--F--D VT P-MD-LK-----VP-V--I-----	
Other Cyanobacteria	Synechococcus sp. PCC 7335	254423782	--LS--L--F--QL--ETLTL P--S--K-----VP-V-----
	Synechococcus sp. WH 7803	148239559	--LSGSV--SMR--L--M--GDLP P--N-LK-S-R--VP-V-----
	Synechococcus sp. RCC307	148242489	--LTGSA--AVR--L--M--GQVP A--N-IK-----VP-V-----
	Synechococcus sp. RS9916	116074807	--LSGSV--SV--L--M--GELP P--N-LKDS-R--VP-V-----
	Cyanothece sp. PCC 7425	220909494	---TE--L--I--DSIT P--G-LKG-----VP-V-----
	Synechococcus sp. WH 7805	88808594	--LSGSV--SMR--L--M--GDLP P--N-LK-S-R--VP-V-----
	Synechococcus sp. WH 8016	352093970	--LSGSV--SVR--L--M--GEIP P--N-LK-S-R--VP-V-----
	Synechococcus sp. RS9917	87124420	--LSGSV--SM--L--M--GELP A--N-LKGS-R--VP-V-----
	Synechococcus sp. CC9311	113954753	--LSGSV--SVRNL--M--GEIP P--N-LK-S-R--VP-V-----
	Synechococcus sp. WH 5701	87302729	--LSGSA--AF--L--MH--GELP P--L-K-S-R--VP-V-----
	Cyanobium sp. PCC 7001	254432745	-----A--AI--L--MH--GELP P--KNLK-S-R--VP-V-----
	Synechococcus sp. CB0101	318041518	--LQ--A--AI--L--LH--GELP P--KNLK-S-----VP-----
	Synechococcus sp. CC9605	78212960	--LSGSV--SVR--L--M--GEIP A--N-LKGS-R--VP-V-----
	Synechococcus sp. BL107	116070508	--LSGSV--SVR--L--M--GELP A--N-LKGS-R--VP-V-----
	Synechococcus sp. CB0205	317970000	--LQ--A--AI--L--MH--GELP P--KNLK-S-----VP-----
	Synechococcus sp. CC9902	78184642	--LSGSV--SVR--L--M--GELP A--N-LKGS-R--VP-V-----
	Synechococcus sp. WH 8109	260436475	--LSGSV--SVR--L--M--GEIP A--N-LKGS-R--VP-V-----
	Synechococcus sp. WH 8102	33865826	--LSGSV--SVR--L--M--GELP A--N-LKGS-R--VP-V-----
	Synechococcus elongatus PCC 6301	6249535	--LLT--GLTQGMQQ-----GAIT QGF-P-K-----SY--I-----
	Synechococcus elongatus PCC 7942	81299241	--LLT--GLTQGMQQ-----GAIT QGF-P-K-----SY--I-----
	Prochlorococcus marinus NATL1A	124025679	---S--SVQ--SV--NL--M--DAMQ E--FKSL--GSVR--VPSV-----
	Prochlorococcus marinus MIT 9312	78712776	--LS--SVQ--SV--N--LI--KDIP E--KNIKGSVR--VP-V-----
	Prochlorococcus marinus CCMP1375	33240378	---S--SVQ--GVRN--I--KEIP Q--IKSL--GSV--VP-V-----
	Prochlorococcus marinus AS9601	123968487	--LS--SVQ--SV--N--LI--KDIP E--KNIKGSVR--SVP-V-----

Supplementary Figure 55

Partial sequence alignment for the protein deoxyxylulose-5-phosphate synthase, showing a 4 aa deletion that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		271	320
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282895973	VRFQSIESSVQQQLQRCSTMGKELGLTSTVYL GNE TKQEASLWQQLPELI
	Cylindrospermopsis raciborskii CS-505	282901683	-----S--E-----S VD- -----L---
	Nodularia spumigena CCY9414	119511523	A-----GE--KE-SN-LLAV-EK---DG-I-V NEH-G----R-R--M
	'Nostoc azollae' 0708	298489871	T-----SE--KE-SN-LLSV-EQ---HGAI-S GEN-VE----R--Q-
	Anabaena variabilis ATCC 29413	75910753	A-----SE---E-SN-LLEV-EK---GV-VS GAD-SH--ER-S-Q-
	Nostoc sp. PCC 7120	17228120	A-----SE---E-SN-LLEV-EK---GV-VS GAD-SH--ER-Q-Q-
	Nostoc punctiforme PCC 73102	186681256	A-----SE--KE-SN-VLEV-QK---DGAIFA DAD-----R-Q-R-
	Fischerella sp. JSC-11	354566265	-----AE--KE-SN-LAEV-QN---QKAI-A GAD--D--KR-R-S-
	Microcoleus chthonoplastes PCC 7420	254416663	-----VTK---E-SS-LLDV-QQ---QG-L-Q DSD-VT---S---Q-
	Microcoleus vaginatus FGP-2	334118731	-----AE--K--SA-LLEV-EK---QG-SCC END-HQ---R---TM
	Oscillatoria sp. PCC 6506	300864919	-----PE-IE--SS-LLEV-YK---KG-ICT END-TNF--R-QSI-
	Trichodesmium erythraeum IMS101	113475707	---TVID--KE-SKKLLEV-NS-E-N--Y-H DD--E--NK-KQK-
	Moorea producta 3L	332710469	-----VAD--KE-CA-IQSVAEQ---RNSF-S NGDDT---ES-ANS-
	Arthrospira platensis Paraca	284050109	-----RE--EL-SD-LVELANQ-H-N--IIR ENDGDF---KQH-
	Arthrospira platensis NIES-39	291571945	-----RE--EL-SD-LVELANQ-H-N--IIR ENDGDF---KQH-
Other Cyanobacteria	Arthrospira maxima CS-328	209524004	-----PE--EL-GD-LVEV-NQ-N-N--IIR ENDGDF---KQQ-
	Lyngbya sp. PCC 8106	119487235	-----LPE--D--AT-LLEL-EK---VGDR-S ED-HH----KQ--
	Cyanothece sp. PCC 8802	257059819	-----PE---E-INQV-AIAQK---KTVF-Q KEI-EN----QDF-
	Cyanothece sp. PCC 8801	218246777	-----PE---E-INQV-AIAQK---KTVF-Q KEI-EN----QDF-
	Cyanothece sp. PCC 7424	218442237	L-----PE-IA--FNTVKMI-QQ---ING--E G-E-NH---R-Q--M
	Cyanothece sp. PCC 7822	307151329	L---G-PE-IT--SNQVKMI-Q---IDSI-R EDE-NP--KR-Q---
	Cyanothece sp. ATCC 51472	354554040	-----PE--DE-TKQVQAI-AEK-N-KTDF-E D-I-QN--KR-Q---
	Cyanothece sp. ATCC 51142	172038620	-----PE--DE-TKQVQAI-AEK-N-KTDF-E D-I-QN--KR-Q---
	Cyanothece sp. CCY0110	126657366	-----PK-IKE-SKQVQAI-AEK-DIKT-FCQ DQI-LN---R-Q---
	Microcystis aeruginosa NIES-843	166367173	L-----PE-ISE-S-QV-YL-QQ-DCQTNY-R DEE-I----RYRKT
	Crocospaera watsonii WH 8501	67921006	-----PE--NE-S-QVQAI-AEK-SIKVIFCK DEI-NN--K--QK--
	Microcystis aeruginosa PCC 7806	159028576	L-----PE-ISK-S-QV-YL-QQ-DCQTNY-R DEA-V---RYRKT
	Synechococcus sp. PCC 7002	170077333	L--EA-RE-ITA---ELK-I-QH---V-QPEH LD--A--R-AL

Supplementary Figure 56

Partial sequence alignment for a FAD linked oxidase-like protein, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		282899854	63	112
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899854	CPPLAQIHELIRSPLSGSNDFH GDE	LTEFIISPMETPVKTEIVPDAATC
	Raphidiopsis brookii D9	282898409	-----H-----Y-----	T-- -I----A-----A-----
	Aphanizomenon sp. 10E6	295812107	-----R-K-----YL-EF--N	T-V--R-VNSV-----S----S-
	'Nostoc azollae' 0708	298492330	--A--K-N--V--RYL-EF--D	D-V--H-INNTM--QVTA-----
	Nostoc punctiforme PCC 73102	186680891	-----K-NQ-T-I-YE-EFN-N	N-V--N-VNNAI-----
	Nostoc sp. PCC 7120	17228189	-----K-N----T-YQ-ELN--	D-V--H-VSGV-R---T-----
	Anabaena variabilis ATCC 29413	75910798	-----K-N--V-T-YQ-ELN--	D-V--H-VSGV-R---S-----
	Nodularia spumigena CCY9414	119509241	----GK-Q--T-VNVE-DE--T	D-V--R-VSSRN---T----I-
	Fischerella sp. JSC-11	354567360	-----R-NQ-T-KRYEKEFL--	D-V--H-ISSA-----S-----
	Other Cyanobacteria	Lyngbya majuscula CCAP 1446/ 4	46981399	P----K-ESIS---YT-KQS-Y
Microcoleus chthonoplastes PCC 7420		254417261	----K-ESID-T-YP-DRT--	D-T-TQ-VNST-Q-Q-P-----
Lyngbya sp. PCC 8106		119484432	P----K-ESIH-Y-YT-EQS-S	N-T-VS-KSNKIR-K-A-----
Trichodesmium erythraeum IMS101		113474623	R----D-ETIT-QH-ENA-FLD	---TS-NNNS-R-K-A-----
Arthrospira maxima CS-328		209522937	K--ISH--S-EAVW-PLAGYS-	FEIRASQSTIQGAKSAIVL--L---
Arthrospira sp. PCC 8005		376001807	K--ISH--S-EAVW-PLAGYS-	FEIRASQSTIQGAKSAIVL--L---
Arthrospira platensis NIES-39		291567263	K--ISH--S-EAVW-PLVGYS-	FEIRASQSTLQGAKSAIVL--L---
Arthrospira platensis Paraca		284052977	K--ISH--S-EAVW-PLVGYS-	FEIRASQSTLQGAKSAIVL--L---
Cyanothece sp. PCC 7425	220905911	P----R-ET-E-QTIPVDEINA	PG-R-TH-VN-I-R---S-----	

Supplementary Figure 57

Partial sequence alignment for the hydrogenase maturation protein, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

			1		48
	Raphidiopsis brookii D9	282898405	MKYVDEFNRNPEKAQGLQKEIAQLSLQISR	NS	HKNKHLKIMEVCGGHTH
	Cylindrospermopsis raciborskii CS-505	282899836	-----	-----	-----
Nostocales/ Stigonematales	Aphanizomenon sp. 10E6	295812110	-----A-Q-TK-Q-D-	-----	-----
	'Nostoc azollae' 0708	298492332	-----K-EA-TK-R-DK	-----	-----
	Anabaena variabilis ATCC 29413	75910800	-----E-EA-RR-EK-Q-LDK	-----	-I-----
	Nostoc sp. PCC 7120	17228191	-----E-EA-RR-EK-Q-LDK	-----	-I-----
	Nostoc punctiforme PCC 73102	12232076	-----E-EA-CR-TK-H-LEK	-----	PI-----
	Nodularia spumigena CCY9414	119509239	----N-E-AA-RNQ-T-N-LNK	-----	P-L-----
	Trichodesmium erythraeum IMS101	113474625	----K-D-D-VR-QKLG	-----	TTEHP-----
	Lyngbya majuscula CCAP 1446/4	72537498	-----D-Q-L-A-MG-NR-T-ILE-	-----	TSEHP-----
	Microcoleus chthonoplastes PCC 7420	254417253	-----D-Q-KA-LR-QT-TT-FGK	-----	TASQP--V-----
Other Cyanobacteria	Lyngbya sp. PCC 8106	119484435	-----D-Q-L-A-MG-NR-T-ILE-	-----	TPEHP-----
	Microcystis aeruginosa NIES-843	166368920	-----D-V-A-IQQ-E-IR-E	-----	KRE-VT-----
	Cyanothece sp. PCC 7425	220905914	----N-D-T-A-IAQ-R-GS-LYS	-----	Q--QVI-----
	Microcystis aeruginosa PCC 7806	159030398	-----D-V-S-IQQ-E-TR-E	-----	KRQ-VT-----

Supplementary Figure 58

Partial sequence alignment for the hydrogenase formation protein, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		282898911	691	EYS	KNRDKKELEEKIFIIANQLN	742
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i> CS-505	282898911	HDRVQQAAYLLIPDNEKESTHWHLGQLIY	EYS	KNRDKKELEEKIFIIANQLN	
	<i>Raphidiopsis brookii</i> D9	282898072	---I-----	-----	-----	
	<i>Nostoc</i> sp. PCC 7120	17229750	-----A---EH-KEI-LNI-R-LL	---	-I-SL-----D-----	
	<i>Anabaena variabilis</i> ATCC 29413	75906308	-----A---EH-KGI-LNI-R-LL	---	-IN-SL-----D-----	
	<i>Nostoc punctiforme</i> PCC 73102	186682607	-----F---ED--QT--LK---LL	---	N-ISAD-RQ-----Q-V----	
	<i>Fremyella diplosiphon</i> Fd33	48374205	-----S--A-EQ-QT--LQI-R-LL	---	Q-S-PS-R--RL-D-V-HC--	
	<i>Nodularia spumigena</i> CCY9414	119508927	-----A---ESK-Q-I-LRI-K-LL	---	N-TS-S-R-DL--ELV--F-	
	' <i>Nostoc azollae</i> ' 0708	298491722	-----S---DQ-QT--LKI---LL	---	S-APLR-QN-NS-E-V----	
	<i>Fischerella</i> sp. JSC-11	354568923	-----S---EDN-K-I-LKI-KOLL	---	-KIPIQDW--N--SLV--F-	
	<i>Microcoleus vaginatus</i> FGP-2	334119626	-----S---EQ-KE--LKI---LL	---	QSTTPEDRK-N--ALV----	
	<i>Microcoleus chthonoplastes</i> PCC 7420	254415737	-----A---GDI-KEI-LKV-H-LW	---	--TKSD-----D-V----	
	<i>Oscillatoria</i> sp. PCC 6506	300868806	-----S---TKQKA--LKI---LL	---	Q-TNQEAE--NV-D-VS---	
	<i>Lyngbya</i> sp. PCC 8106	119487723	-----F-----Q-QI--YQI-K-LL	---	QQISPEVI-DR--ELV----	
	<i>Trichodesmium erythraeum</i> IMS101	113476169	-----S---DQ--AN-Y-I---LL	---	EQISPETR-RY--E-V----	
	<i>Moorea producta</i> 3L	332706581	-----F---EDQ-QL--LTI---LL	---	--IPTAKR-----D-V----	
Other Cyanobacteria	<i>Arthrospira</i> sp. PCC 8005	376003335	---I---S-S---HQ-QA--LQI---LQ	---	H-LSPT-V---L-D-VGH--	
	<i>Arthrospira maxima</i> CS-328	209526937	---I---S-S---HQ-QA--LQI---LQ	---	H-LSPT-V---L-D-VGH--	
	<i>Arthrospira platensis</i> Paraca	284052405	-----S---EQQ-QT--YRI---LL	---	-QIPIESR-DY--ELISK--	
	<i>Arthrospira platensis</i> NIES-39	291571924	-----S---EQQ-QT--YRI---LL	---	-QIPIESR-DY--ELISK--	
	<i>Arthrospira platensis</i> S6	206585493	---I---S-S--H-DQ-QA--LQI---LQ	---	H-LSPT-A---L-D-VGH--	
	<i>Cyanothece</i> sp. PCC 7822	307153463	-----S---AH-KE--LKI---LR	---	--TSFE-QK-N--TLV----	
	<i>Cyanothece</i> sp. CCY0110	126659478	-----S---EEQ-QE--LKI---LL	---	NSTPAT--D-TL-T-V----	
	<i>Cyanothece</i> sp. PCC 7425	220906228	-----S---ATQR-V--LKI-R-LL	---	Q-TSPEKQK-T--SLV----	
	<i>Cyanothece</i> sp. PCC 7424	218439689	--Q---S-S---PQ--PL--VKI-R-LL	---	E-IPSEFK--R--E-V-H--	
	<i>Acaryochloris marina</i> MBIC11017	158338239	---I-----S---SGQ-RA--LMI--QLL	---	-LVPDM-RD-Q--ELV----	
	<i>Acaryochloris</i> sp. CCMEE 5410	359457475	---I-----S---SGQ-RA--LMI--QLL	---	-LVPDM-RD-Q--ELV----	
	<i>Cyanothece</i> sp. ATCC 51142	172039010	-----A--D-RQ-TAV-LSI-R-LR	---	DSMNEKDT-EL-A-VDH--	
	<i>Microcystis aeruginosa</i> NIES-843	166366081	-----S---ES-RKAN-LQI--FLL	---	-KTTTQAQ-N-L-D-V----	
	<i>Microcystis aeruginosa</i> PCC 7806	159029461	-----S--TEE--KR--LRI-R-LL	---	NSVREE--RDNL-K-V-H--	
	<i>Cyanothece</i> sp. ATCC 51472	354554545	-----S-LSKQKQ-EI-LKI-K-LL	---	N-TP-IKQ-----D-V--I-	

Supplementary Figure 59

Partial sequence alignment for the protein serine/threonine protein kinase and signal transduction histidine kinase, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		217	263
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282895647	DFVVDGILSGTDRQLYMEEFLLDGD CPFLGQTLQKAKLRSQTGALVLAIRR
	Cylindrospermopsis raciborskii CS-505	282899654	-----S-----NPA N --V--S-----
	'Nostoc azollae' 0708	298491740	-----T-A-----PA F --V--S-----R---S-----
	Nostoc punctiforme PCC 73102	186685052	-----T-A-----PA L --V--S-----S-----
	Anabaena variabilis ATCC 29413	75909935	-----T-A-----PA L --V--S-----S-----
	Nostoc sp. PCC 7120	17228574	-----FI-----F-----PA V Y-YV-H--RQ-R---S---I----
	Nodularia spumigena CCY9414	119509893	-----I-----PK V --VV-M--GQ-R--A----I--V--
	Fischerella sp. JSC-11	354564956	-----T---SF-----V-P- S --YS---RD---S-----
	Oscillatoria sp. PCC 6506	300867338	-----IT-A--SF-----I-PS S --LI-K--RE-H-----
	Microcoleus chthonoplastes PCC 7420	254409601	-----T---SSF-----PE T --CV-K--RE-R-----
	Moorea producta 3L	332708976	--L--V-T---SSF-----V-PN S -VYV---RN---S-----V--
	Microcoleus vaginatus FGP-2	334119086	-----T---AF-----P- V --CV--P-RD-R--L-----
	Cyanothece sp. PCC 7822	307151194	-----T-S--TF-----PK F --CV---SE-R--L-S-----
	Crocospaera watsonii WH 0003	357265619	-----T-S--TF-----PK F --CV---SE-R--L-S-----
	Crocospaera watsonii WH 8501	67921005	-----T-S--TF-----PK F --CV---SE-R--L-S-----
Other Cyanobacteria	Cyanothece sp. CCY0110	126659308	-----T-S--TF-----PK F --CV---SE-R--L-S-----
	Cyanothece sp. ATCC 51142	172036480	-----T-S--TF-----PK F --CV---SE-R--L-S-----
	Cyanothece sp. PCC 7425	220907925	-----T-S--SF-----ISAH T --HV--S-RD---S-----
	Cyanothece sp. PCC 7424	218442042	-----T---AF-----PE V --CV--P-KE-R--L-S-----
	Cyanothece sp. PCC 8802	257058276	-----T-S--TF-----PQ F --CI---SE-R--LKS-----
	Cyanothece sp. PCC 8801	218245244	-----T-S--TF-----PQ F --CI---SE-R--LKS-----
	Microcystis aeruginosa NIES-843	166364314	-----T---SF-----I-PR T --CV-LS-SQ-H---RS-----
	Microcystis aeruginosa PCC 7806	159027662	-----T---SF-----I-PR T --CV-LS-SQ-H---RS-----
	Synechocystis sp. PCC 6803	16329750	S-----T-A--SF-----RIGAE D --YI---RE-Q-A-S---I-----
	Synechococcus sp. JA-3-3Ab	86606261	-----R--TF-L--YE-AAQ E --YV---G--N--ARS---V----
	Synechococcus sp. JA-2-3B'a(2-13)	86609264	-----A-R--TF-L--YE-AAQ E --YV---R--D--NRS---V----
	Synechococcus sp. PCC 7002	170076953	-----T-AE-SY-L--IRV-SN C GGYI-KS-SE-R--V-S-----
	Acaryochloris sp. CCMEE 5410	359462365	N-M-VA-T-SEPGF-I--L-I-A- N --YV---GV-DV-----SI-----
	Synechococcus sp. PCC 7335	254425027	--M--V-SG--TY----F--PL V -AQI---KESQ--AL--V-IV----
	Acaryochloris marina MBIC11017	158339286	-----FA-G-QTY----F--PQ V SSKI---EESQ---L--V-IV----
Synechococcus elongatus PCC 6301	56750855	--I-DLT-ASE-SV-ID---EQP R SR-I-R--RE-D--GRS---I-----\	

Supplementary Figure 60

Partial sequence alignment for the protein neurotrophic tyrosine kinase receptor type 1, showing a 1 aa deletion that is specific for *Cylindrospermopsis raciborskii* CS-505 and *Raphidiopsis brookii* D9.

		17227730	57	115
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227730	PIRFACHHYMGQVPILREIVTQ	LGCFPLEENQHRQQSFFVQSQKLLQTKQMVGVFPEG
	Anabaena variabilis ATCC 29413	75908936	-----	-----
	Nostoc punctiforme PCC 73102	186680641	-----VM-----	-----T-N-----S---L---S-----
	Nodularia spumigena CCY9414	119512206	-----LM-----	-----DTPS-----E---II--SQ-----
	Fischerella sp. JSC-11	354567132	-----M-----A-	-----A--Q-----K-A-L--KS--L-----
	Oscillatoria sp. PCC 6506	300867845	-----VM--V--T	F-A-----PT----H--K-ATT--AGE-----
	Microcoleus vaginatus FGP-2	334118812	S-----VM----T	F-A---APE----H--S-ATA--SGE-----
	Lyngbya sp. PCC 8106	119487033	-----V-KDV--T	--A--Q-PTQ-S-H-LH-ASQ--NQE-----
	Moorea producta 3L	332707289	--PI--D--I---VM-VGCPA	D-L--SRKTRTYTVDV-TTGKSVSR-
	Cyanothece sp. PCC 7822	307151235	-V-I-----SKT-----FIQL	---I---SSGS---G--E-AKS---SH-W--I----
Other Cyanobacteria	Synechococcus sp. PCC 7335	254421829	-VH-----S--VMSD--DA	----DGPQK-GRT--K-ATHF-RSH-SI-----
	Cyanothece sp. PCC 7424	218437103	-V-I-----S-T-L--F-QL	---I--QPS-KS-KT--E-ANQV-HSQ-W--I----
	Microcystis aeruginosa PCC 7806	159029237	TLPI-----KT-F--LIGE	----FDTPEK-HRT--R-ATDF--SG-WI-L----
	Synechococcus sp. WH 5701	87302257	---V--Y--SR--G--M-SA	M-A--DAP-Q-PKE--QKAVS--HE-EV--I----
	Microcystis aeruginosa NIES-843	166367027	TLPI-----KT-L--LIGE	----FDTPEK-HRT--R-ATDF--SGRWI-L----
	Synechococcus sp. JA-3-3Ab	86605889	-----Y-LS--L--VALA	---I--QQGSGP-LA--R-A-AT-AAGGG--I----
	Synechococcus sp. RS9917	87124698	-V-----Y--S---V-HQATML	M-A---SGRQC-A---R--INI--SN-L-----
	Synechococcus sp. JA-2-3B'a(2-13)	86609758	-----Y-LA--L--ALA	-D-I--KQGSRQ-IH--R-A-AS-AAGVG--I----
	Cyanothece sp. PCC 7425	220908711	S-H---P--T-----DVIEQ	---L--SSQPE-N--QRASL--NS-A--I----
	Cyanothece sp. PCC 7822	307152251	-VAYMAKEELF-I-L-L-KPLIEL	Y-AY-IKRGSG DL-A-RAAR-A-AHGWA--I-L--
	Cyanothece sp. PCC 8802	257059173	-VAYMAKEELF--W--GIRL	Y-AY-VKRGSGDRGAIIRAAL-A--EGWI--I-L--
	Cyanothece sp. PCC 8801	218246124	-VAYMAKEELF--W--GIRL	Y-AY-VKRGSGDRGAIIRAAL-A--EGWI--I-L--
	Prochlorococcus marinus MIT 9211	159902693	--A-MAKQELFRI--GW-IRA	C-AY-VSRGASDREAIR-A-ER-H-GWAT--LD--
	Gloeobacter violaceus PCC 7421	37521809	-V-LV----L----VVN-L-RQ	--G-H-GPAGQGWSTL-R-ASRY-LSGTH-AI----

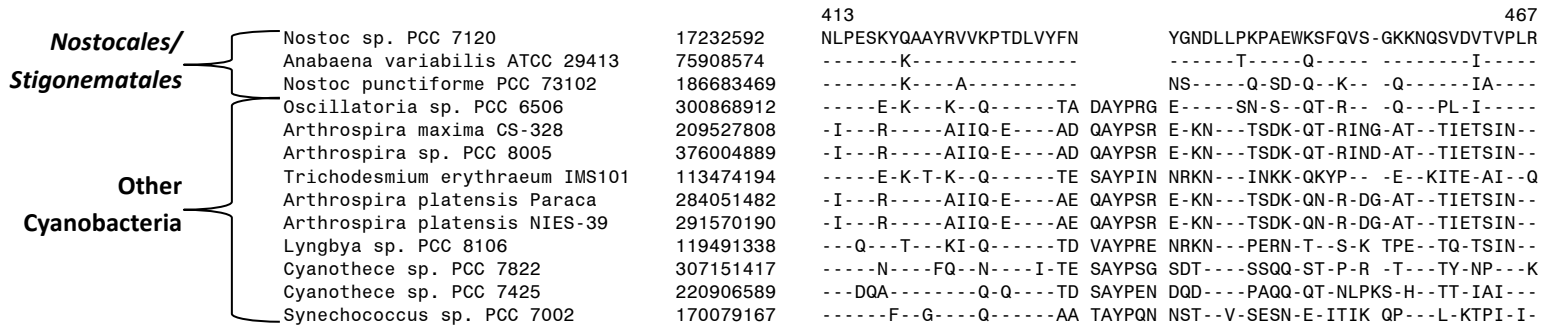
Supplementary Figure 61

Partial sequence alignment for the protein phospholipid/glycerol acyltransferase, showing a 1 aa insert that is specific for all members of a *Nostocales* subclade which includes *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413*, *Nostoc punctiforme PCC 73102* and *Nodularia spumigena CCY9414*.

			76		122
			LIGHSAGGWISRIYLGEKAYAAR	GKA	TSLWNAHSLVATLVTLGTPH
			-----P-----	---	-----P-----
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227750	-----M---P-MVG	SN-	P---K--P-----I-----
	Anabaena variabilis ATCC 29413	75908954	-----V-----DQP-L-P	-QV	SCC-K-----S-I-----
	Nostoc punctiforme PCC 73102	186683066	-----DQP-DGQ		V-AG-QW-Q--I-----
	Nodularia spumigena CCY9414	119510315	-----W---DRP-GPT		QQA-QGRDRT-V-IC-----
	Thermosynechococcus elongatus BP-1	22299525	-VA-----A----SVP-GGQ		V-QGREH-S--IS--S--
	Synechococcus elongatus PCC 6301	56752264	IV-----M-AEP-CDQ		I-Q-QPKIH---S-----
	Synechococcus sp. JA-3-3Ab	86606815	IV-----M-AEP-CDQ		V-Q-QPKIH---S-----
	Acaryochloris sp. CCME 5410	359459632	-VA-----A----SVP-GGQ		V-QGREQ--A-IS--S--
	Acaryochloris marina MBIG11017	158339118	-----AADFTNL V-DRP-GPT		KQACQERDRTPV-IC-----
	Synechococcus sp. JA-2-3B'a(2-13)	86607996	-----S--VML-L--SAD-FSG-		CYGGAQFCNR-I---S--
	Synechococcus elongatus PCC 7942	24251238	-----S--VML-P--ADQTFGL-		RF-GAA-CNR-I---S--
	Synechococcus sp. RS9916	116074771	-----S--VML-L--A-RGLEGA		SL-GRQ-CNR-IC--S--
Other Cyanobacteria	Synechococcus sp. WH 8109	260434414	-----S--VML-L--SDDPFEG-		IYAGSKRCNR-----S--
	Synechococcus sp. RCC307	148242538	-V---S--VML-L--SDD-FQG-		RYGGSDRCNR-----S--
	Synechococcus sp. WH 8016	352096383	-----S--VML-L--SREPF-G-		VYAGAERCNR-I---S--
	Synechococcus sp. WH 7803	148239521	-----S--IML-LF-DDAPFQG-		RYDGKA-ASA--M--S--
	Synechococcus sp. RS9917	87124387	-----S--IML-LL-DDAPFQG-		SYDGKA-AD---M--S--
	Cyanobium sp. PCC 7001	254431182	-----S--VML-L--SDEPFDGS		IYAGSKKCDR-----S--
	Synechococcus sp. WH 5701	87303536	-V---S--VML-L--SDVEFQGC		RYGGSDRCNR-----S--
	Synechococcus sp. CC9311	113955246	-----S--VML-L--DQPFSG-		IYSGYKSCNR-I---S--
	Synechococcus sp. WH 7805	88808632	-V-----A-LL--D-E-GR-		T--YRDR--S-I-----
	Prochlorococcus marinus MIT 9303	124023266			
	Gloeobacter violaceus PCC 7421	37520555			

Supplementary Figure 62

Partial sequence alignment for the hypothetical protein N9414_16197, showing a 3 aa insert that is specific for all members of a *Nostocales* subclade which includes *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413*, *Nostoc punctiforme PCC 73102* and *Nodularia spumigena CCY9414*.



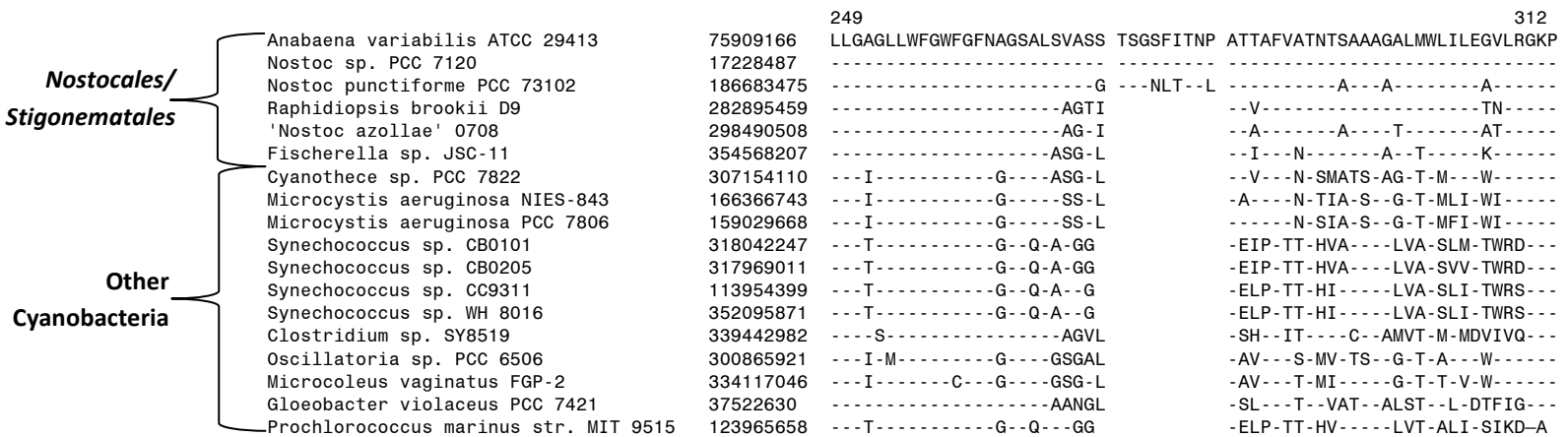
Supplementary Figure 63

Partial sequence alignment for the hypothetical protein all5100, showing a 6 aa deletion that is specific for all members of a *Nostocales* subclade which includes *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413* and *Nostoc punctiforme PCC 73102*.

		419	476	
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17228462	YENVLSPMYELSSLFHQKFDHLLLNDS	
	Anabaena variabilis ATCC 29413	75909176	-----N-----N-Y-	
	Nostoc punctiforme PCC 73102	186681495	-----Q---S-----Q-QQT--TE P-	
	Nodularia spumigena CCY9414	119512507	-----I-----V-NI-PTN--E	
	Fischerella sp. JSC-11	354566062	-----D--KK--VM--E-Y-S-VQQ-PLV	
	Oscillatoria sp. PCC 6506	300864973	-----G--TQ--AA--K-Y-P--QQ--V	
	Arthrospira platensis Paraca	284052893	-----E--NQ--AD--R-Y-P-VKQ-PTT	
	Arthrospira platensis NIES-39	291565660	-----E--NQ--AD--R-Y-P-VKQ-PTT	
	Arthrospira maxima CS-328	209527972	-----E--NQ--AD--R-Y-P-VNQ-PTT	
	Microcoleus vaginatus FGP-2	334118830	-----E--NL--AA--R-Y-P-VQQ-A-V	
	Microcoleus chthonoplastes PCC 7420	254414177	-----E--NQ--AA--L-Y-P-VQQ--V	
	Cyanothece sp. PCC 7425	220907377	-----E--SQ--AA-----AQVQNPAT	
	Cyanothece sp. PCC 7822	307153691	-----G--IQ--AD--Y-A-VQQ-P-V	
	Cyanothece sp. PCC 7424	218438121	-----G--TQ--VD-----P-VQQ-P-V	
	Synechocystis sp. PCC 6803	16330052	-----E--NQ--AK-----R-VAE-P-V	
	Acaryochloris marina MBIC11017	359462030	-----E--NQ--A-----Y-P-VQA-P-T	
	Microcystis aeruginosa NIES-843	166368522	-----E--NQ--AK--E-Y-L-VQA-P-F	
	Cyanothece sp. ATCC 51142	172038359	-----G--N--AG-----SRVKE-GNV	
	Cyanothece sp. CCY0110	126658314	-----G--N--AG-----SRVKE-GNV	
	Cyanothece sp. PCC 8801	218245201	-----E--NQ--AM-----Y-P-VQA-P-T	
	Synechococcus sp. PCC 7002	170077544	-----E--NQ--AD--R--S-VQQNPGV	
	Microcystis aeruginosa PCC 7806	159028313	-----E--N--AK--E-Y-R-VAE-P-V	
	Crocospaera watsonii WH 8501	67924295	-----E--NQ--AQ--E-YYP-VQA-P-F	
	Other Cyanobacteria	Synechococcus sp. WH 7805	88809672	-S---E--QS--AQ--R--AAVAT-P-V
		Synechococcus sp. WH 7803	148239177	-S---D--QA--AR--H--AAVAA-P-V
Synechococcus sp. PCC 7335		254422699	-SD--T--E--NQ---IAATGEP I E I L	
Synechococcus sp. RS9917		87123824	-D---E--Q---AR--HH--ASVAS--AI	
Synechococcus sp. CC9902		78184545	---I-E-INQ--AQ--KR--EQVS-SPEI	
Synechococcus sp. BL107		116070788	---I-E-INQ--AQ--KR-EQE-AANPGI	
Synechococcus sp. RCC307		148242151	-S---E--KQ--AQ--RR-GE-NSSSP-	
Cyanobium sp. PCC 7001		254432537	-A---E--QQ--A---V--P-VAA-PAV	
Synechococcus sp. WH 8109		260436802	-A---G--NQ--AQ--AL--TSVAA-HTT	
Synechococcus sp. WH 8102		33865968	-----E--N--AR--R--AAVNSEP-L	
Synechococcus sp. CB0101		318041601	-T---E--Q---AR--R--PAVAD-PGT	
Synechococcus sp. WH 8016		352094492	-D---E--NQ--AD--R--QQTAK-P-I	
Synechococcus sp. WH 5701		87300658	-A---E--NA--AR--L--P-VAQ-PGV	
Synechococcus sp. RS9916		116075120	-S---T--HQ--AA--AR--AAVQA-P-L	
Synechococcus sp. CC9605		78212604	-A---A--NQ--AQ--AR--PSVAA-PTT	
Synechococcus sp. CC9311		113953436	-D---E--NK--AA--H--KQTAE-P-F	
Synechococcus sp. CB0205		317970332	--T--E--QQ--AA--RYPTA-PS-P-	
Prochlorococcus marinus str. MIT 9303		124022809	-----E-INS--AK--L--P-VTK-PAI	
Prochlorococcus marinus str. NATL2A		72382331	-N---E-ISK---Y--ET---VAQNP-L	
				SI TKQADYRQALLYNLLLQTSCFRYWGQG
				---P-----
				-R-SH--NL--H-----
				--NL--H-----
				-QRP--QE-----
				---PE-Q---M-----
			-R-P--Q-----	
			-R-P--Q-----	
			-R-P--Q-----	
			---LE-Q-----	
			-QRS--Q-----T--E-----	
			-Q-P--Q-----	
			-RR--QE-----	
			--TPA--E-----	
			-QRE--QN----L--VE-----	
			-QR--QE---V--VE-----	
			-RR--LE-----V-----	
			-RTPE-QE---TM-VE-----	
			-RTPE-QE---TM-VE-----	
			-QRP--QE---V--VE-----	
			-RTPE-QE-----VE-----	
			--TEA-QRS---L--VE-----	
			-RRP--LE-----V-----	
			-RTTE-QE---TM-VE-----	
			-ATPA-Q---HV--E-----	
			-ASSA-----HV--E-----	
			R--PR--N--H-----	
			-ST-H-QE---HL--E-----	
			-QTEH-Q---VHL--SE-----	
			-QAPH-Q---VHL--TE-----	
			-H---HV--E-----	
			-GSPA-QE---HL--E-----	
			-GTRA-QE---HV--E-----	
			-GTSS--D---HL--E-----	
			-RTPA-QN---HL--E-----	
			-TGES--N---HL--E-----	
			-RTPH-Q---HL--E-----	
			-AN-H--DC--HL--E-----	
			-RTRA-QES--HL--E-----	
			-TSDR-QK---HL--E-----	
			-Q---HL--E-----	
			-QTPH-QE---L-M-E-----	
			--THS-QK---L--E-----	

Supplementary Figure 64

Partial sequence alignment for the hypothetical protein Npun_R1017, showing a 2 aa insert that is specific for all members of a *Nostocales* subclade which includes *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413* and *Nostoc punctiforme PCC 73102*.



Supplementary Figure 65

Partial sequence alignment for the protein ammonium transporter, showing a 9 aa insert that is specific for all members of a *Nostocales* subclade which includes *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413* and *Nostoc punctiforme PCC 73102*.

			135		189
<i>Nostocales/</i>	Nostoc sp. PCC 7120	17231389	GGRIIAITYSPSGRTGSW QP	WMAMGTAKAALES	LCRYFAVALGSRGITVNGISPG
<i>Stigonematales</i>	Anabaena variabilis ATCC 29413	75908021	-----	-----	-----
	alpha proteobacterium HTCC2255	114770156	--T-LTL--QG-N-VTPF	YNV--V-----AV--L-ND--PQ--R--A----	
Other bacteria	Roseiflexus castenholzii DSM 1	156743345	-AS-VTL--HG-QQVAQN	YNV--V-----ASV--L-ND--P--R--A--A-	
	Thermosynechococcus elongatus	22299236	--S--TL--LGGV-VVPN	YNV--I-----MNV--L-AE--P-N-R-----A-	
	Paracoccus denitrificans PD122	119384388	--S--TL--AG-N-VTPF	YNV--V-----SV--L-ND--PE--R--A----	

Supplementary Figure 66

Partial sequence alignment for the protein short-chain dehydrogenase/reductase, showing a 2 aa insert that is specific for *Nostoc sp. PCC 7120* and *Anabaena variabilis ATCC 29413*.

		17227628	388	424
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227628	RGVTPDGKLYNIARNILN GG	GSEFAGACFSPDGRTL
	Anabaena variabilis ATCC 29413	75907723	-----E-----LV--NFQ --	S-----Q---
	Nostoc punctiforme PCC 73102	186684422	I----E-D--RF-A-AI-	-----H-M-
	Nodularia spumigena CCY9414	119511526	V--NKK-EV-QF-M-NIT	SG-L-----A---M-
	Microcoleus vaginatus FGP-2	334116687	V----N-E--RL-K-AI-	D-----M-
Other Cyanobacteria	Moorea producta 3L	332710206	V--N-K-E--QL---A--	S-----N--IM-
	Cyanothece sp. CCY0110	126656794	I----E--Y-KL---AY-	N-----V--A---K-M-
	Cyanothece sp. PCC 7425	220910401	L-I--A--I-VLG--A--	R-----Q---
	Cyanothece sp. ATCC 51142	172038491	I---LE-QY-KL---AY-	N-----V-----K---
	Crocospaera watsonii WH 0003	357260160	I----E-QY-RL---AY-	N-----V----N-Q-M-
	Cyanothece sp. PCC 8801	218247513	V----Q-EI-QLG--A--	N-----V-----QK-M-
	Cyanothece sp. PCC 8802	257061151	V----Q-EI-QLG--A--	N-----V-----QK-M-
	Acaryochloris marina MBIC11017	158336680	--INQQ-Q--SF---A--	DK-L-----TNPL---
	Acaryochloris sp. CCMEE 5410	359459939	--INQQ-Q--AF---A--	DK-L-----TNPL---
	Thermosynechococcus elongatus	22297708	V-I--Q-EI-TF---A--	DQ-L--I-----Q-M-
Gloeobacter violaceus PCC 7421	37522016	--I--E---SF---A--	-F--T--S-----K-F-	

Supplementary Figure 67

Partial sequence alignment for the hypothetical protein Npun_F4300, showing a 2 aa insert that is specific for *Nostoc sp. PCC 7120* and *Anabaena variabilis ATCC 29413*.

			127			171
	Nostoc sp. PCC 7120	17231457	EGFPKILIEQKDDRFRLYAAMDCVFIESD	AH	HTLHILDFKTGKA	
	Anabaena variabilis ATCC 29413	75907957	-----	N-	-----	
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186686205	-----VV--E-YC-----		S-----S	
	'Nostoc azollae' 0708	298490840	-----V---ANYC-----		S-----S	
Other Cyanobacteria	Nodularia spumigena CCY9414	119509391	-----FV--E-VH-----	I-R--	R-----S	
	Fischerella sp. JSC-11	354567118	-----D--HYC-----	I----S	GN-----S	
	Microcystis sp. T1-4	390441547	--Y-E--V---GNY--N----	F--I-R-A-	D-I----L---QS	
	Microcystis aeruginosa NIES-843	166364681	--Y-E--V---GNY--N-S--F--	I-R-A-	D-I----L---QS	
	Microcoleus vaginatus FGP-2	334116570	--V-QA-I-K-GNYL-NF---	I--I-V-P-	GR-----S	

Supplementary Figure 68

Partial sequence alignment for the hypothetical protein Npun_R6172, showing a 2 aa insert that is specific for *Nostoc sp. PCC 7120* and *Anabaena variabilis ATCC 29413*.

		383	421	
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17232677	MPWVITYDPEVAHVGLYESQ	DVETIKI-PFSSVDRAIADG
	Anabaena variabilis ATCC 29413	75908647	-----I-----	-----A
	Nostoc punctiforme PCC 73102	186683625	-----I---M--HE	AQKLG I E-T-----
	Nodularia spumigena CCY9414	119509922	-----I---M--QE	AQKQGF EIN-----
	'Nostoc azollae' 0708	298490605	I-R-----I---I-ADE	AERLGI EI----- G-NN-----T-S
	Fischerella sp. JSC-11	354568874	-----I---M--QE	AQEKGI --N-----
	Moorea producta 3L	332709354	---A-----I---M--HE	AQAKGF N-N----- -T-----
	Microcoleus chthonoplastes PCC	254410978	-----I---M--DE	AQEKGM K-S----- -NT-----
	Oscillatoria sp. PCC 6506	300864852	---C-----I---M--QE	AKEKGI EIN-FF- -Q----VT--
	Microcoleus vaginatus FGP-2	334118137	I--C---N--I-R---GDRE	LQAAGI -ID-FL- -DK-----I--
	Lynghya sp. PCC 8106	119484786	I-SCI-----Q--IT-TE	ILRRPE R-K-FYV -L-EI---CI-R
	Trichodesmium erythraeum IMS10	113476862	I--A-F---L-R---T-Q	AKERYG -DVCVLKQ--G---Q-EA
	Arthrospira platensis NIES-39	291567161	I--A-F---L-R---T-A	ARGHYG -DVI-LKQQ--D---Q-E-
	Arthrospira platensis str. Par	284051886	I--A-F---L-R---T-A	ARGHYG -DVI-LKQQ--D---Q-E-
	Arthrospira sp. PCC 8005	376001890	I--A-F---L-R---T-A	AKAAYG -DVI-LRQQ--D---Q-E-
Arthrospira maxima CS-328	209526830	I--A-F---L-R---T-A	AKAAYG -DVI-LRQQ--D---Q-E-	
Acaryochloris sp. CCME 5410	359457842	-----F---I---M--GE	AQDQGM --N----- L-DD---L---	
Microcystis aeruginosa PCC 780	159028642	-----I-R--MD-T-	AQAQGL TTN----- -L-I-----T--	
Acaryochloris marina MBIC11017	158333369	-----F---I---M--GE	AQDRGM --N----- L-DD---L---	
Microcystis aeruginosa NIES-84	166365002	-----I---MD-T-	AQARGL ATN----- -L-I-----	
Cyanothece sp. PCC 7425	220907391	---T---L-----PQE	AACQNI P----CV -MTD-----A	
Synechocystis sp. PCC 6803	16329673	-----I---N-TM	AEALDI GYKI--- -Q-----AD	
Synechococcus elongatus PCC 63	56750440	I-R--F-----S-TA	ARHQGI AIA--T- -LDQ---TVL--	
Synechococcus elongatus PCC 79	81299927	I-R--F-----S-TA	ARHQGI AIA--T- -LDQ---TVL--	
Cyanobium sp. PCC 7001	254432211	I-RT-F-----T---T-AE	AAAQGI P--VTRF -LHEL---VES	
Synechococcus sp. WH 5701	87301512	I-RT-F-----G---S--E	AARQGV A--VTRF -LHEL---VES	
Cyanothece sp. PCC 7822	307153633	I--A-F---L-RI--T-AE	ARKKYD -IYVL-Q A-AD---Q-E-	
Synechococcus sp. PCC 7335	254425099	I--A-F-E--L-R---T-R-	ARDRYG -DIY-LKQG-DG----L-ES	
Cyanothece sp. PCC 8801	218245798	I--A-F---L-R---S--E	AREKYG QDICVLKQD-TN---Q-E-	
Cyanothece sp. PCC 8802	257058844	I--A-F---L-R---S--E	AREKYG QDICVLKQD-TN---Q-E-	
Synechococcus sp. PCC 7002	170077910	I--A-F-----R--MT-E-	ARKRYG --LIL-V --A-----Q-ED	
Cyanothece sp. CCY0110	126658328	I--T-F---L-R---T-T-	AKEKYG -NIYVLKQ--E---Q-E-	
Cyanothece sp. PCC 7424	218441833	I--A-F---L-R---T-AE	ARKKYN --DVL-Q E-ADI---Q-E-	
Cyanothece sp. ATCC 51472	354555007	I--A-F---L-R---T-E-	AKQRYG -AVHVLKQN--E---Q-E-	
Cyanothece sp. ATCC 51142	172035772	I--A-F---L-R---T-E-	AKQRYG -AVHVLKQN--E---Q-E-	
Gloeobacter violaceus PCC 7421	37520133	---C-----I-----G-DE	ARKRGI AFD--RL -LAES---LT--	
Burkholderiales bacterium JOSH	375104905	I--T-F-----R---N-AD	AKAQQV AF-VTRY GLDDL-----S	
Phaeodactylum tricornutum CCAP	219117666	V-A-M--E--Y-T---SEA	NADKQE IEVDTYRAGLEHN----LE-	
Prosthecochloris aestuarii DSM	194334052	I--CS--E--I---ET-AE	LHARHA GH-VYRF --NRI---TED	

Supplementary Figure 69

Partial sequence alignment for the protein mercuric reductase, showing a 6 aa deletion that is specific for *Nostoc sp. PCC 7120* and *Anabaena variabilis ATCC 29413*.

		884	929
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17231890	SGKLRLEEMLEEVLAEEN TYGVAGAG RALIFTQFAEWGKLLKPHL
	Anabaena variabilis ATCC 29413	75909484	-----
	'Nostoc azollae' 0708	298490375	-A-----E-
	Cylindrospermopsis raciborskii CS-505	282899415	-A----Q----I--KS
	Nodularia spumigena CCY9414	119510052	-A----D---TVA-E-GD
	Fischerella sp. JSC-11	354566873	----L-Q-----S-GD
	Fischerella sp. MV11	225593133	----L-Q-----S-GD
	Microcoleus vaginatus FGP-2	334119521	-----V-----
	Oscillatoria sp. PCC 6506	300868337	-----D-----SAED
	Arthrospira sp. PCC 8005	376005480	-----N-----LIS-GD
	Arthrospira maxima CS-328	209525614	-----N-----LIS-GD
	Arthrospira platensis NIES-39	291569923	-----N-----LIS-GD
	Arthrospira platensis Paraca	284053180	-----N-----LIS-GD
	Lyngbya sp. PCC 8106	119489573	-----FK----L-S-GD
	Moorea producta 3L	332708786	----R-----AIS-GD
	Trichodesmium erythraeum IMS101	113475235	-----GA---IIS-EE
	Microcoleus chthonoplastes PCC	254416145	----L-Q---D-AIS-GD
	Synechococcus sp. PCC 7335	254425442	-----D-----V--GD
	Microcystis aeruginosa PCC 9806	389804454	----L---A-V---I--GD
	Microcystis sp. T1-4	390438684	----L---A-V---I--GD
	Microcystis aeruginosa NIES-843	166362977	----L---A-V---I--GD
	Synechocystis sp. PCC 6803	16332119	----I--A-----IIS-GD
	Acaryochloris marina MBIC11017	158339003	----K--Q----L-D-GD
	Acaryochloris sp. CCME 5410	359461387	----K--Q----L-D-GD
	Other Cyanobacteria	Synechococcus sp. WH 5701	87300749
Synechococcus sp. CB0205		317968221	-A-V-----I---MEAGD
Synechococcus sp. RCC307		148243480	-A-----I---MEAGD
Synechococcus sp. CC9311		113955603	-V-----I-D--VEAGD
Synechococcus sp. WH 8016		352095643	-V-----I-D--VEAGD
Synechococcus sp. WH 7805		88809531	-V-----I---I-AGD
Synechococcus sp. RS9916		116074100	-A-----I-D--IEAGD
Synechococcus sp. RS9917		87123478	-M-----I---IDAGD
Synechococcus sp. WH 7803		148240736	-V-----I-D--IEAGD
Synechococcus sp. CC9605		78214026	-A-----I-D--IEAGD
Synechococcus sp. WH 8102		33866892	-A-----I-D--IEAGD
Synechococcus sp. WH 8109		260434360	-A-----I-D--IEAGD
Thermosynechococcus elongatus BP-1		22299156	-----I---QALQEVGD
Crocospaera watsonii WH 0003		357262043	-H--E--G-----I--GD
Crocospaera watsonii WH 8501		67924074	-H--E--G-----I--GD
Synechococcus elongatus PCC 6301		56750675	-A----L---A-LTDAGD
Synechococcus elongatus PCC 79		81299683	-A----L---A-LTDAGD
gi 66361321 pdb 1Z5Z A Chain A		66361321	---MI-TM-II--A-D-GD
Microcystis sp. T1-4		390438857	-H--E--G-----IL-GD
Cyanothece sp. CCY0110		126660084	---LFQ-I---L-DNRH
Prochlorococcus marinus str. M		159902591	-T-----L-DVIF-TED
Prochlorococcus marinus subsp.		33239499	-S-----I-Q--KESHD
gi 66361330 pdb 1Z6A A Chain A		66361330	---MI-TM-II--A-D-GD
Gloeobacter violaceus PCC 7421		37521835	---S--G-----DEE
Oscillochloris trichoides DG6		309790429	---A--D-----V-AED
Actinoplanes missouriensis 431		383775390	---A---LA--IVEDGD

Supplementary Figure 70

Partial sequence alignment for the protein SNF2-like helicase, showing an 8 aa insert that is specific for *Nostoc sp. PCC 7120* and *Anabaena variabilis ATCC 29413*.

		33		88
Nostocales/ Stigonematales	Fischerella sp. JSC-11	354565229	NGPGEITTWVRPVVKALRQKLGWER FSEAKD	IGTHLRISVVLV PCPHASGKEADIA
	Anabaena variabilis ATCC 29413	75907920	-----V-----E--DD-	VQV----I-- ---N-----A--
	Nostoc sp. PCC 7120	17231497	-----V-----E--SDD-	VQV----I-- ---N-----A--
	Nostoc punctiforme PCC 73102	186683891	-----V-----R---QF-DD-	NQV----- ---SN-----A--
	'Nostoc azollae' 0708	298489992	-----V-----EQ--ND-	S-V-----I-- --L-----S--
	Nodularia spumigena CCY9414	119509080	-----V-----E--DNP	AVV----- ---N-T---GA--
	Cylindrospermopsis raciborskii CS-505	282901223	-----V---L---I---EQ--NDI	SQV----I-- ----T---R--
	Raphidiopsis brookii D9	282895507	-----V-----T---EQ--NDI	SQV----I-- ----T---E--K--
	Oscillatoria sp. PCC 6506	300864554	-----V-----Q---Q--CD-	-LV-----I-- ---N---S--A--
	Microcoleus chthonoplastes PCC 7420	254412560	-----V-----Q---Q--CD-	AVF----I-- ---N-T---Q-V
	Microcoleus vaginatus FGP-2	334117292	-----LA---A-Q---EQ--NAG	-QA----- ---T---E--
	Lyngbya sp. PCC 8106	119486017	-----L---S---E---KQ--DQT	EKV---LI-- ----T---VA--
	Trichodesmium erythraeum IMS101	113476757	-----LV-----QE---QANIDM	---L--- ---T---VA--
	Arthrospira platensis NIES-39	291566131	-----L-----LE-I-SQNSDRH	KD---LI-- ----T---VE-G
	Arthrospira platensis paraca	284052197	-----L-----LE-I-SQNSDRH	KD---LI-- ----T---VE-G
	Arthrospira sp. PCC 8005	376006836	-----L-----LE-I-SQNSHLD	-T-LI-- ---N-T-N-VE-G
	Arthrospira maxima CS-328	209527614	-----L-----LE-I-SQNSHLD	-T-LI-- ---N-T-N-VE-G
	Synechococcus sp. JA-2-3B'a(2-13)	86610274	-----L-----ME--Q-PQA-	L-L--- ---Q-MK--
	Cyanothece sp. PCC 7822	307155180	-----VS-----Q--ND-	-SV----I-- -T--M-T--A--
	Cyanothece sp. CCY0110	126659054	-----V-----KQ--DDT	AQT---M-- ---ST-E--A--
	Cyanothece sp. PCC 8802	257060536	-----V-----EIF-KD-	SQV---M-- ----T-Q--A--
	Cyanothece sp. PCC 8801	218248137	-----V-----EIF-ED-	SQV---M-- ----T-Q--A--
	Cyanothece sp. ATCC 51142	172038303	-----V-----R---KQ--DDP	AQT---M-- ---ST-Q--T--
	Cyanothece sp. ATCC 51472	354554348	-----V-----R---KQ--DDP	AQT---M-- ---ST-Q--T--
	Cyanothece sp. PCC 7424	218439719	-----VS-----H--DD-	-SV----I-- -T--M-T-SA--
	Synechococcus elongatus PCC 7942	1399817	-----VA-----Q--DN-	DR--L-L-A --SNGT-Q-AA-
	Synechococcus elongatus PCC 6301	56750187	-----VA-----Q--DN-	DR--L-L-A --SNGT-Q-AA-
	Microcystis aeruginosa NIES-843	166367227	-----S-----R---N-TED-	SL-----I-- -T--M-A-TEVV
	Cyanothece sp. PCC 7425	220905706	-----L-----Q---S--ED-	QQV---L-- --SN-T-E-VT--
	Crocospaera watsonii WH 0003	357266044	-----L-----TE--TDP	QQT---M-- ---ST-Q--T--
	Crocospaera watsonii WH 8501	67920633	-----L-----TE--TDP	QQT---M-- ---ST-Q--T--
	Synechocystis sp. PCC 6803	16330951	M-----Q---L--Y	S-----M-- ---M-R--EVV
	Synechococcus sp. PCC 7335	254422209	-----VA---K---RS---APNY-	----- ---PT-L
	Synechococcus sp. PCC 7002	170076996	-----V---L---Q-V-AHW-SLP	-----A --S-ST-Q-A--
	Synechococcus sp. JA-3-3Ab	86605575	-----LL-----LE---Q-PQA-	L-L--- ---Q-MK--
	Acaryochloris sp. CCMEE 5410	359457713	-----LS-----LQV-SRRPQFS	-V-IL-- ---G--QAV-
	Acaryochloris marina MBIC11017	158336303	-----LS-----LQV-SRRPQFS	-V-IL-- ---G-H-QAV-
	Thermosynechococcus elongatus BP-1	22299093	-----LA---Y-F-TR--RVKQGV	-V--MV ---Q-E--
	Synechococcus sp. RS9917	87125314	-----LA-----LAD--HRS-VLRP	RAPWSP--LR-VLV--N-T-T-QV--
	Synechococcus sp. RS9916	116073802	-----LA-----LAEQ-HRS-QLRP	RAQSSL--LR-VLV--N-T-Q--A--
	Synechococcus sp. RCC307	148241439	-----L-----LAER-H-RALPST	-LQ-VLV--N-T-T--V--
	Synechococcus sp. WH 5701	87301339	-----L-----LAQR-HRQ-ELQP	VVPGAAC-LE-VLV--N-T-S--RV--
Synechococcus sp. CC9605	78213771	-----L-----LAER-HSS-RLRP	RSSEAPA-LH-VLV--N-T-Q-RAA--	
Synechococcus sp. WH 8102	33864873	-----L-----LAER-HAT-PLRP	RTSGTAA-LQ-VLV--N-T-Q-RAA--	
Cyanobium sp. PCC 7001	254431819	-----L-----AQR-QRR-LTPE	SOASLALY-VLV--N-GT-R-HRA--	
Synechococcus sp. CB0205	317967967	-----LA-----LAQR-HRE-PLRP	LTPGAPMALQ-VLV--N-T-I-HQV--	
Synechococcus sp. WH 8016	352096038	-----L-----LAEQ-HRR-LMRP	RAPGSLL-LR-VLV--N-T-T--SA--	
Synechococcus sp. CC9311	113955036	-----L-----LAEQ-HRR-LMRP	RAPGSLL-LR-VLV--N-T-T--AA--	
Synechococcus sp. BL107	116071381	-----LS-----LAER-HSLQPLRP	RVPSSPA-LQ-VLV--N-T-Q-HRA--	
Synechococcus sp. CC9902	78184001	-----LS-----LAER-HSLQPLRP	RVPSSPA-LQ-VLV--N-T-Q-HRA--	
Synechococcus sp. CB0101	318040500	-----LS-----LAQR-HRE-AMRP	VAAQSPMALR-VLV--N-T-S-HRV--	
Synechococcus sp. WH 7805	88807227	-----LA-----LAEQ-HRR-LMRP	RAPGSPV-LR-VLV--N-T-T--EA--	
Synechococcus sp. WH 7803	148240463	-----L-----LAEQ-HRR-LLRP	RAPASP--LH-VLV--N-T-T--EA--	
Gloeobacter violaceus PCC 7421	37522281	-----LN-----AE-VRRWPQA-	L--I-A -DGNS---EQM-	
Prochlorococcus marinus MIT 9303	124022034	-----LA-----IAES-HRSILMRP	RAASSPVDLR-VLV--N-T-Q--A--	
Prochlorococcus marinus CCMP1986	33860619	-----LS-----DY-NKINDTSK	NIDK-NY-IR-VLV--N-T--YEV--	

Supplementary Figure 71

Partial sequence alignment for the protein lipid-A-disaccharide synthase, showing a 6 aa insert that is specific for *Fischerella sp. JSC-11*.

		477	543
Nostocales/ Stigonematales	Fischerella sp. JSC-11	354568923	FLDDLQWADAASLKFIQSLTTSR QSPLLCSTKS EDTREYKAEIIGSNQGSILLIGAYRDNEV
	Nodularia spumigena CCY9414	119508927	-----FT---L--L-NSAS N-SENIAMSATLEK-K-L-I-----
	Anabaena variabilis ATCC 29413	75909801	-----SG---L--L-MNDT -YL-----
	Nostoc sp. PCC 7120	17231183	-----SG---L--L-MNDT -YL-----
	Nostoc punctiforme PCC 73102	186683708	-----V-N---L--L-A-VT KGQ-L--M-----
	Fremyella diplosiphon Fd33	48374205	-----S---L--L-MADT CHL-----
	'Nostoc azollae' 0708	298491722	-----S---L--L-MADT SHL-----
	Cylindrospermopsis raciborskii CS-505	282898911	-----PT---L--F-IKEP DTKYL-----H--
	Oscillatoria sp. PCC 6506	300868404	-----S---L--L-IANT SYLF-M-----
	Microcoleus vaginatus FGP-2	334117410	-----LM-L-MSES DARYL-----
Other Cyanobacteria	Lyngbya sp. PCC 8106	119487723	-----LM-L-MADT -YLFI-----
	Trichodesmium erythraeum IMS101	113475755	-----T---LMEL-MSDK -SQHL-----
	Moorea producta 3L	332708490	-----R---EL--L-M-AA DNKYLFF-----
	Arthrospira sp. PCC 8005	376003847	-----L--ELMKL-MEEK TYL--L-----
	Arthrospira maxima CS-328	209523991	-----L--ELMKL-MEEK TYL--L-----
	Arthrospira platensis NIES-39	291571924	-----L--ELMKL-IEEK TYL--L-----
	Microcoleus chthonoplastes PCC 7420	254411958	-----V-L-T--L--LIM-DS EIHLYF-----
	Arthrospira platensis S6	206585493	---W--S-S--NLLKM-MQET --YL-IL----H--
	Arthrospira platensis Paraca	284052405	-----L--ELMKL-IEEK TYL--L-----
	Prochloron didemni P1-Palau	335387279	-----FGT--L-ELMV-DE QVE-L--L-----
	Cyanothece sp. PCC 7822	307151273	-----L-----FM-M-AS DTQYL-----
	Cyanothece sp. CCY0110	126658537	-----I-----LFKH-MSES KDCHL-----
	Cyanothece sp. ATCC 51472	354554545	-----S-T--QL-R--I-TL DDQ-LFI-----
	Cyanothece sp. ATCC 51142	172036732	-----S-T--QL-R--I-TL DDQ-LFI-----
	Acaryochloris sp. CCMEE 5410	359458100	-----NLLKLIMEGV GLDI-TKHL-MV-----I
	Acaryochloris marina MBIC11017	158336697	-----S---L-KLMM-D-QMR-L-----
	Microcystis aeruginosa NIES-843	166366081	-----I-S---LL-LIV-DK RGKYLFF-----
	Cyanothece sp. PCC 7425	220906228	-----LP--RL-EL-V-NP ESQYL-IL-----
Cyanothece sp. PCC 7424	218442716	-----F-T--L-ELMMEN-EI-YL--L-----	
Microcystis aeruginosa PCC 7806	159029461	-I-----FS--SL-KE-ISDP DIQYF-I--S--N---	

Supplementary Figure 72

Partial sequence alignment for the protein multi-sensor signal transduction multi-kinase, showing a 10 aa insert that is specific for *Fischerella sp. JSC-11*.

		759	818
Nostocales/ Stigonematales	Fischerella sp. JSC-11	354565228	LPCGLHVIGKPPSAEEAVATLVNIIAGLDRPE GVEALNL TSLQSLPRIANSIAGLRDIDEIY
	Anabaena variabilis ATCC 29413	75909518	-----T---I-----S---Q- EEVL-----I-----
	Nostoc sp. PCC 7120	17231857	-----T---I-----S---Q- EEIL-----I-----
	'Nostoc azollae' 0708	298489993	-----T---I-----S---S- EEIF-----I-----
	Nostoc punctiforme PCC 73102	186683218	-----I-----S---Q- DGI-G--G-----N--D--
	Raphidiopsis brookii D9	282896763	-----T-Q-----S---T- EEIL-----Q-----LE--
	Nodularia spumigena CCY9414	119509081	-----GS---Q- EEIL-----S-I-----D--
	Cyanothece sp. PCC 7424	218438418	-----Q-T---I----- EEIL-----I-----
	Cyanothece sp. PCC 8801	218249027	-----I-----A--E- EGIT-----V--E-E-V-
	Microcystis aeruginosa PCC 9701	389881880	-----L-----S---E- DD-L-----I-----E-V-
	Cyanothece sp. PCC 7822	307152707	-----Q-T---I-----E- EEIL-----I--N-E---
	Microcystis sp. T1-4	390437881	-----L-----S---E- DD-L-----I-----LE-V-
	Microcystis aeruginosa NIES-843	166363960	-----L-----S---E- DD-L-----I-----LE-V-
	Thermosynechococcus elongatus BP-1	22297815	-----TT-----S----- DGIK---L--E-----
	Cyanothece sp. ATCC 51472	354552460	-----V---T---I-----S---E- EGIT-----D--D---E-V-
	Cyanothece sp. ATCC 51142	172039271	-----V---T---I-----S---E- EGIT-----D--D---E-V-
	Crocospaera watsonii WH 0003	357264272	-----I---T---I-----S---E- EEIT-----D--D---E-V-
	Crocospaera watsonii WH 8501	67922189	-----I---T---I-----S---E- EEIT-----D--D---E-V-
	Synechococcus sp. PCC 7335	254423414	-----V-E-T---I----- EEQK---Q---E-IN---Q-
	Cyanothece sp. CCY0110	126661027	-----V---T---I-----S---E- EGII---L-D--D---EQV-
	Synechococcus sp. PCC 7002	170077639	-----T---I-----GS---E- EGIW---T---K-I---NME---
	Synechocystis sp. PCC 6803	16329632	-----Q-----I-----S---ED EGIWA--TL--E-I---NME---
	Acaryochloris marina MBIC11017	158338660	-----T---I-----S---G- DNILG---L---IN--L---F
	Cyanothece sp. PCC 7425	220910329	-----A--I-----S---E- EGIKG---LL-E---LE---
	Acaryochloris sp. CCMEE 5410	359461828	-----T---I-----S---G- DNILG---L---IN--L---F
	Synechococcus sp. JA-3-3Ab	86605997	-----A-T---I---S-SI--- EGILG---Q-----
	Synechocystis sp. PCC 6803	992631	--VV---Q---I-----S---ED EGIWA--TL--E-I---NME---
	Synechococcus elongatus PCC 6301	56751964	-----T---I-----S---D- EGITG--TL--R---EME--F
	Synechococcus elongatus PCC 7942	81169827	-----T---I-----S---D- EGITG--TL--R---EME--F
	cyanobacterium UCYN-A	284928943	-----TG--V-----S---E- EDII---SL--T-IE---KD---
	Synechococcus sp. WH 5701	87302937	-----T---T---I-----A-E-E- EGYR---ALL-E---T--V-
	Synechococcus sp. RCC307	148242069	-----T---T---I-----A-E-ED DG-R---SLL-E-I---S---V-
	Synechococcus sp. WH 8109	260435732	-----T---T---I-----A-E-E- DG-R---GLL-EAM--S-ED--
	Synechococcus sp. CC9605	78213349	-----T---T---I-----A-E-E- DG-R---GLL-EAM--S-ED--
	Synechococcus sp. WH 8102	33865354	-----T---T---I-----A-E-E- DG-R---GLL-EAM--S-ED--
	Synechococcus sp. RS9916	116072918	-----T---T---I-----A-E-E- DG-R---GLL-EAM--T-ED--
	Synechococcus sp. WH 7803	148239774	-----T---T---I-----A-E-E- DG-R---GLL-EAM--S-ED--
	Synechococcus sp. WH 8016	352093779	-----T---T---I---S-A-E-E- EG-R---GLL-E-I--K-ED--
	Synechococcus sp. WH 7805	88808822	-----T---T---I-----A-E-E- DG-R---GLL-EAM--T-ED--
	Synechococcus sp. CC9311	113953972	-----T---T---I---S-A-E-ED DG-R---GLL-E---K-ED--
	Synechococcus sp. RS9917	87125687	-----T---T---I-----A-E-E- EGIR---GLL-EA---K-GD---
	Synechococcus sp. CC9902	78184402	-----T-R-T---I---S-A-E-E- DG-R---GLL-E-I---S-ED--
	Cyanobium sp. PCC 7001	254431261	-----T---T---I-----A-E-E- EGIR---ALL-ECR--T-A-V-
	Synechococcus sp. CBO101	318041717	-----T---T---I-----A-E-E- DGIR---ALL-E-K---T-TD--
	Synechococcus sp. BL107	116070940	-----T-R-T---I---S-A-E-E- DG-R---GLLGEAI--S-ED--
Microcoleus chthonoplastes PCC 7420	254410035	-----T---I-----S---E- EEIL-----N-----	
Oscillatoria sp. PCC 6506	300864031	-----T---I-----S---E- DN-L-----V---NDVF	
Microcoleus vaginatus FGP-2	334117283	-----T---I-----N---E- DG-I-----G-V-	
Moorea producta 3L	332708185	-----T---I-----N---S- DEIIG-----IE-----	
Arthrospira platensis paraca	284054016	-----T---I-----S---E- EGIV-----IN---E-L-	
Arthrospira platensis	99033850	-----I-----S---E- EGIV-----IN---E-L-	
Arthrospira platensis NIES-39	291567418	-----I-----S---E- EGIV-----IN---E-L-	
Trichodesmium erythraeum IMS101	113476754	-----T---I-----S---E- EGITG-----N---E-L-	
Arthrospira sp. PCC 8005	376001692	-----I-----S---E- EGIL-----IN---E-L-	
Lyngbya sp. PCC 8106	119487056	-----T---I-----S---E- DG-L---H---D-IN-NLE--	
Arthrospira maxima CS-328	209523055	-----A---I-----S---E- EGIL-----IN---E-L-	
Prochlorococcus marinus 9313	33863087	-----T---T---I---S-A-E-ED DG-R---GLL-EAID-K-ED--	
Prochlorococcus marinus CCMP1375	33240353	-----T-M-T---I-S--S-A-E-EQ EKIR---GLL-EAI-K--E---	
Prochlorococcus marinus NATL2A	723822220	-----T---AT---I---S-SIE-ED DGIR---GLL-E-K-KT-E-V-	
Prochlorococcus marinus AS9601	123968572	-----T--EA-T-----S-E-EQ EG-R---GLL-E-I-LT-EQ--	
Gloeobacter violaceus PCC 7421	7522191	V--D--V-T--T-T--ID--GV-SF----	KEMEG-D-ML-TAR-W--EALQ

Supplementary Figure 73

Partial sequence alignment for the protein magnesium chelatase subunit H, showing a 7 aa insert that is specific for *Fischerella sp. JSC-11*.

194

259

Nostocales/ Stigonematales	Fischerella sp. JSC-11	354565504	DNSVHAEINDLAFVPAFHEYGEVGRLDGGEKDFPT P	YNSSSKVFGFNILVGGFFSAKRCEAAIPL
	Nodularia spumigena CCY9414	119510725	-----Q-GSGEWG-V-S-WGVGSG	EKPEAS-----
	Nostoc sp. PCC 7120	17228103	-----K-GIGDWW--N--ESS-YQ	-----V-----
	Anabaena variabilis ATCC 29413	75910736	-----K-GIGDWW--G--ESS-HQ	-----V-----
	Nostoc punctiforme PCC 73102	186681964	-----K-ANGQEFKSKPP	I-----
	Cylindrospermopsis raciborskii CS-505	282898756	---I-----K-GF	---V-----G-----
	Raphidiopsis brookii D9	282898029	---I-----K-AF	---V-----G-----
	Cyanothece sp. PCC 7425	220909931	--A-----I-----R-EQL	---V-----
	Microcystis aeruginosa NIES-843	166364582	-----I-----YR-GEL	---VV-----I
	Cyanothece sp. ATCC 51142	172036138	-----I-----YKNGEL	---VV-----
Cyanothece sp. ATCC 51472	354552938	-----I-----YKNGEL	---VV-----	
Crocospaera watsonii WH 0003	357263580	-----I-----YKNGEL	---VV-----S--D----	
Synechocystis sp. PCC 6803	16331650	-----V-----YR-GIL	---VV-----SR-----	
Cyanothece sp. CCY0110	126656647	-----I-----YKDGNL	---VV-----	
Crocospaera watsonii WH 8501	67922714	-----I-----YKNGEL	---VV-----S--D----	
Cyanothece sp. PCC 8801	218247263	-----I-----YKDGQL	---V-----E-----	
Synechococcus sp. JA-2-3B'a(2-13)	86607536	-----KNGRL	---V-----R--AE--G-	
Synechococcus sp. JA-3-3Ab	86605327	-----KNGRL	---V-----R--AE--G-	
Acaryochloris sp. CCMEE 5410	359458074	-----I-----YKDGKL	---V-----AP--SM	
Acaryochloris marina MBIC11017	158336122	-----I-----YKDGKL	---V-----AP--SM	
Thermosynechococcus elongatus BP-1	22298892	-----V-----YRQGR	---V-----R--A-----	
Synechococcus sp. PCC 7335	254425090	ED-I-S---I--I---KDG-D-SSE-	---V-----R--S--A-	
Synechococcus elongatus PCC 6301	56750319	---I-----T--YQDGT	---VW-----ST-VAP---	
Cyanothece sp. PCC 7822	307153401	---I-----V-----YKNGEL	---V-----YL--Q--AE--M	
Cyanothece sp. PCC 7424	218438658	---I-----I-----YKQGE	---V-----YL--Q--AE-VKM	
Synechococcus sp. PCC 7002	170078433	---S-V---IG---KDGQF	---V-----Y---Q-QAEG-SI	
Synechococcus sp. CB0205	317968299	--FLLH --I--H-VERNGEM	--GVWI--IL-SQMNAY---	
Synechococcus sp. WH 8109	260435045	--FLLH --IV-H-VERDGV	--GVWI--VL-SQMNAY-V--	
Synechococcus sp. RCC307	148243581	--FLLH --IV-H-VERDGV	--GVWI--IL-SQMNAY----	
Synechococcus sp. CC9605	78214157	--FLLH --IV-H-VERDGV	--GVWI--VL-SQINAY----	
Synechococcus sp. CB0101	318040833	--FLLH --IV-H-VERNGEL	--GVWI--IL-SQMNAY----	
Synechococcus sp. WH 7803	148240828	--FLLH --IV-H-VE-NGEM	--GVWI--IL-SQMNAY----	
Synechococcus sp. RS9917	87123393	--FLLH --IV-H-VE-NGEL	--GVWI--IL-SQMNAY-L--	
Synechococcus sp. BL107	116071845	--FLLH --II-H-VE-NGVM	--GVWIA-IL-SQLNAY-L--	
Synechococcus sp. CC9902	78185851	--FLLH --II-H-VE-NGVM	--GVWIA-IL-SQLNAY-L--	
Synechococcus sp. RS9916	116074235	--FLLH --II-H-VENNGEL	--GVWI--VL-SQMNAY-L--	
Synechococcus sp. WH 7805	88809407	--FLLH --IV-H-VERDGE	--GVW--IL-SQMNAY----	
Synechococcus sp. WH 5701	87300718	-SFLLLH ---L--H-GGEL	--TVM-----Q-N-L----	
Synechococcus sp. CC9311	113953367	--FLLH --IV-H-VERDGV	--GVWI--IL-SQMNAY----	
Synechococcus sp. WH 8103	3152707	--FLLH --IV-H-VENNGVM	--S-WI--IL-PQMNAY-F-M	
Synechococcus sp. WH 8102	33867007	--FLLH --IV-H-VENNGVM	--S-WI--IL-PQMNAY-F-M	
Synechococcus sp. WH 8016	352095755	--FLLH --IV-H-VKRDGV	--GVWI--IL-SQMNAY-L-M	
Cyanobium sp. PCC 7001	254432737	-SFLLLH ----L--H-NGRL	--TVM-----Q-N-L----	
Lyngbya sp. PCC 8106	119485484	-----I-----KDET	---V-----D-----	
Trichodesmium erythraeum IMS101	113474849	-----I--I--YKNDKF	---V-----	
Oscillatoria sp. PCC 6506	300866335	-----I--I---KDGTF	---V-----D-----	
Moorea producta 3L	332711101	-----I-----YKDGQL	-L-V-----V----	
Microcoleus vaginatus FGP-2	334118443	-----I-----YKDGNI	---V-----	
Leptolyngbya boryana	1113082	-----I-----KDGTL	-----G-----	
Geitlerinema sp. PCC 8501	2499192	-----I-----YKDGKL	---V-----V----	
Phormidium laminosum	208968954	-----I-----YKDGKL	---V-----V----	
Microcoleus chthonoplastes PCC 7420	254416977	---I-----I-----YRDGEL	---V-----YL--Q--VE----	
Arthrospira sp. PCC 8005	376002065	---I-----YRDRIL	---V-----YL--Q-QVE-V--	
Arthrospira platensis NIES-39	291568672	---I-----YRDDVL	---V-----YL--Q-QVE-V--	
Arthrospira maxima CS-328	209526361	---I-----YRDRIL	---V-----YL--Q-QVE-V--	
Arthrospira platensis paraca	284052863	---I-----YRDDVL	---V-----YL--Q-QVE-V--	
Prochlorococcus marinus str. M	33864503	---FLLH --II-H-VE-DGVM	--GVWL--IL-SQMNAY-L--	
Prochlorococcus marinus str. N	124026876	K--FLLH ---I-H-VKINGVL	--SVWI--VL-SVMN-Y-V--	
Physcomitrella patens subsp. p	168046727	H-LFE-PH---Y---TKNGVF	-----S--AE----	

Supplementary Figure 74

Partial sequence alignment for the protein ferredoxin-nitrite reductase, showing a 1 aa insert that is specific for *Fischerella sp. JSC-11*.

Group	Strain	Accession	Protein	Sequence	Reference
Nostocales/ Stigonematales	Fischerella sp. JSC-11	354566903	ADDPVWALQSKPVETVH	GTSLQHGTSLQHGTSLQHGTSLQHGTSLQHGTSLQHGTSLQK	DLSAPRVAVTLRSHPLLTPTRL
	Nostoc punctiforme PCC 73102	186680603	-----E----PGLW		--P-----Q--E---
	Nostoc sp. PCC 7120	17232001	-----A---FPFLA		--PT--I-----N--Q--AA--
	Nodularia spumigena CCY9414	119509456	-----E----PGLW		--P-----N--Q--Q--
	Anabaena variabilis ATCC 29413	75909581	-----A---FPFLA		N-PT--I-----N--Q--AP--
	Cylindrospermopsis raciborskii CS-505	282899344	-----G-A-ES-PELA		G-PK--I-----K--Q--QD--
	Raphidiopsis brookii D9	282896025	-----G-A-ES-PELA		S-PK--I-----K--Q--EG--
	Oscillatoria sp. PCC 6506	300869101	-----EAL--AGLW		--P-----N---ST-----
	Microcoleus vaginatus FGP-2	334117996	-----ENS--SGLW		--P-----C---ST---A--
	Microcoleus chthonoplastes PCC	254415955	-----D-T--PGLW		--K-----H--QE--
	Arthrospira maxima CS-328	209527480	-----DP---PGLW		--P-----W-----K--
	Arthrospira sp. PCC 8005	376007165	-----DP---PGLW		--P-----W-----K--
	Arthrospira platensis str. Par	284051496	-----DP---PGLW		--FP-----W-----K--
	Cyanothece sp. PCC 7424	218441336	-----E---QGLW		--P-----N--A---AKK--
	Cyanothece sp. PCC 7822	307153897	-----E---QGLW		--P----LN---Q--Q--
	Cyanothece sp. ATCC 51472	354553355	-----TA-F-PGLA		--P-----N-----Q--
	Other Cyanobacteria	Cyanothece sp. ATCC 51142	172036577	-----TA-F-PGLA	
Cyanothece sp. PCC 8801		218246539	-----KA-S-PKLS		--P-----N--K--Q--Q--
cyanobacterium UCYN-A		284929101	-----R--DISDFI		QT-Q-I--IN--N--S--IKK--
Cyanothece sp. CCY0110		126656383	-----TA-S-SGLA		--P--K--N-----SE--
Cyanothece sp. PCC 7425		220909839	-A-----AAS--PALA		--PL-----V--A--Q--S--
Microcystis sp. T1-4		390437907	-----EAQS-KGLS		--P--I--N--P-----AK--
Microcystis aeruginosa NIES-84		166363599	-----DAQS-KGLS		--P--I--N--P-----AK--
Acaryochloris sp. CCME 5410		359462282	-----N-AA---GWVD		K-PS-K--LV-----Q---P--
Acaryochloris marina MBIC11017		158334911	-----N-AA---GWVD		K-PS-K--LV-----Q---PH--
Synechococcus sp. PCC 7002		170076666	-----AA-SFLEMP		G-PR--I--N---RT---QKI
Synechococcus elongatus PCC 63		56752524	-----EPL-LDQPL		SNQEI--C--P--E--E-S
Synechococcus sp. JA-3-3Ab		86607454	-----G-P--TLPEWE		RWPQ--I--V--Q--Q--A--
Synechococcus sp. PCC 7335		254421373	-----DYE--KGLW		--P----V--A--P-SQ-SVS--
Synechococcus sp. JA-2-3B'a(2-		86607636	-----S-P--TLPEWE		TWPQ--I--V--P--Q--GAC-
Gloeobacter violaceus PCC 7421		37520314	TC-----EPASNP-AQ		A-AP----V--P--D-DVR-Q

Supplementary Figure 75

Partial sequence alignment for the protein polysaccharide pyruvyl transferase, showing a 48 aa insert that is specific for *Fischerella sp. JSC-11*.

		291	346
	Fischerella sp. JSC-11	354565836	NRSLWVFGALQALSNIAYFILAQ PQ FSQNYQLLVLTINIENFCGGLGTAAFVAFMLM
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119510082	-----L-----V--L--YA--N VG---PA-----
	'Nostoc azollae' 0708	298489950	-----L-----V--L--LL--- VGK---V-L-----A-----
	Nostoc punctiforme PCC 73102	186684609	-----I-AI--V-G-L-F-A--V VGK-FY-MLAAV---Q--A--E-----
	Nostoc sp. PCC 7120	17232025	-----I--I--L---LG-YA--V AGK--S----AV-----SA--V-V-T-----
	Anabaena variabilis ATCC 29413	75907622	-----I--I--L---LG-YA--I AGK--S----AV-----SA--V-V-T-----
	Raphidiopsis brookii D9	282895754	--A-----I--L---LG-YT--I VGKD-S--AIA-M---TSA--V-V-T--Y--
	Cylindrospermopsis raciborskii CS-505	282898765	--A--I--I--L---LG-YT--I VGKD-P--AIA-M---SSA--V-V-T--Y--
	Microcystis sp. T1-4	390438868	-----I--G---I--L-----N LG---PFM--A-----V-----
	Microcystis aeruginosa NIES-843	166363044	-----I--G---I--L---T--N LG--PFM--A-----V-----
	Cyanothece sp. PCC 7424	218440896	-----I--G-----L--WG-SL VGK--PAMI-----IA--G-----
Other Cyanobacteria	Cyanothece sp. PCC 7425	220906647	-----I--G-----L--LL--I GGKS-P-M-----S-----GL-----
	Cyanothece sp. PCC 7822	307154293	-----G-----A--WV-SL VGKDSTTM--A-----IA--G-----
	Synechococcus sp. RS9917	87125372	--A---Y-VIG-F--L--WA--R -GGGF-G-LVAVSV-----S-MVVS--L-L--
	Synechococcus sp. WH 5701	87301289	-----L-ALIG-IG-LS-WA--- -GGGQTA---AVGM--IG--MVG-V---L--
	Thermosynechococcus elongatus BP-1	22298325	H-A--IM-G---S--LS--V--N AGA-PTVM-AA-S-D---A--AI--LT-L--
	Gloeobacter violaceus PCC 7421	37521237	-----A--LG--L--L AGKSPL-FAAVTV--G-A-----

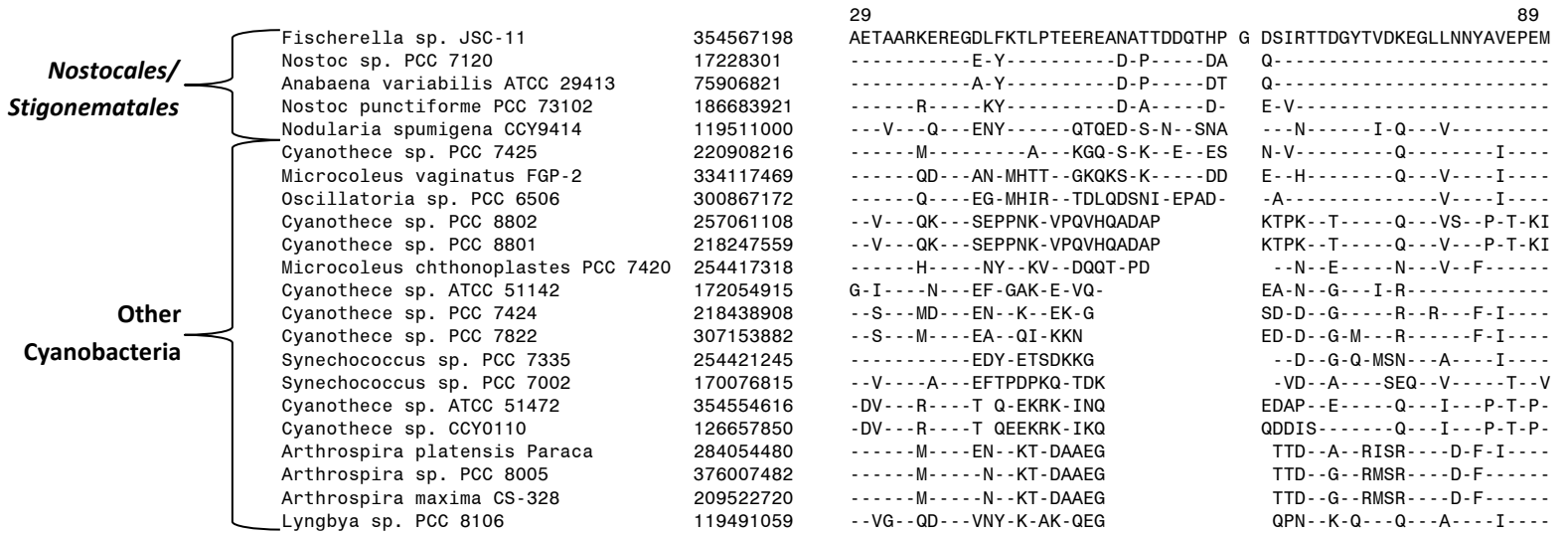
Supplementary Figure 76

Partial sequence alignment for AmpG-related permease, showing a 2 aa insert that is specific for *Fischerella sp. JSC-11*.

		57	105	
Nostocales/ Stigonematales	Fischerella sp. JSC-11	354565331	IFGYPGGAILPIYDDLYKAEASGD G IKHILVRHEQGAHAADGYARATG	
	Nostoc sp. PCC 7120	17232105	-----V--T-S L-----	
	Anabaena variabilis ATCC 29413	75908969	-----V--T-S L-----	
	Nodularia spumigena CCY9414	119509049	-----V-ST-A -----	
	Nostoc punctiforme PCC 73102	186682147	-----V--T-A -----	
	'Nostoc azollae' 0708	298492509	-----V---I -----S-----	
	Raphidiopsis brookii D9	282896701	-----V-ET-T -----S-----	
	Cylindrospermopsis raciborskii CS-505	282900124	-----V-ET-S -----S-----	
	Arthrospira platensis paraca	284054082	-----E--R---E-G -Q-----S-----	
	Arthrospira sp. PCC 8005	376005312	-----E--R---E-- -Q-----S-----	
	Arthrospira maxima CS-328	209523765	-----E--R---E-- -Q-----S-----	
	Arthrospira platensis NIES-39	291569589	-----E--R---E-G -Q-----S-----	
	Trichodesmium erythraeum IMS101	113476429	-----E--R---A-- LQ-----S-----	
	Oscillatoria sp. PCC 6506	300865814	-----E--R---E-G -Q-----G-----	
	Microcoleus vaginatus FGP-2	334118943	-----L--E--R--ET-A -----	
	Microcoleus chthonoplastes PCC 7420	254412830	-----E--QW-ER-E -Q-F-----	
	Moorea producta 3L	332710392	-----E--HW-DA-K -Q-----S-----	
	Arthrospira platensis	124375	-----SN----E-I-R--QA-E --Y-----S--	
	Other Cyanobacteria	Cyanothece sp. PCC 7822	307150286	-----E--RS--R-- -Q-----
		Cyanothece sp. PCC 7425	220905985	V-----E--R---R-- LQ-----
Cyanothece sp. PCC 7424		218441880	-----E--RF--R-- -Q-----	
Acaryochloris marina MBIC11017		158336223	-----EV---QR-- L-----S-----	
Acaryochloris sp. CCMEE 5410		359457747	-----EV---QR-- L-----S-----	
Cyanothece sp. CCY0110		126660706	-----E--R---R-- -Q-F-----S-----	
Cyanothece sp. ATCC 51142		172039391	-----E--R---R-- -Q-F-----S-----	
Thermosynechococcus elongatus BP-1		22298839	-----E--R--SK-I -Q-----	
Synechocystis sp. PCC 6803		16330569	-----E--RF--A-E -E-----S-----	
Crocospaera watsonii WH 0003		357265634	-----E--R---R-E LQ-F-----S-----	
Crocospaera watsonii WH 8501		67920987	-----E--R---R-E LQ-F-----S-----	
Synechococcus sp. PCC 7335		254421963	-----A-HH--K-Q VQ-----S-----	
Synechococcus sp. PCC 7002		170079383	-----A-HPV-ER-E -----S-----	
Cyanobium sp. PCC 7001		254430689	-----E-H--SR-W L-----GS---A-----	
Synechococcus elongatus PCC 6301		56751375	-----E-HR---A-R VQ-----V---A-S-----	
Synechococcus elongatus PCC 7942		81298950	-----E-HR---A-R VQ-----V---A-S-----	
Synechococcus sp. CB0101		318041131	-----E-H--SR-W L-----GT---A-----	
Synechococcus sp. CB0205		317969616	-----E-H--SR-W L--V-----GT---A-----	
Synechococcus sp. WH 5701		87302611	V-----A-H--SR-W L--V-----GT---A-----	
Synechococcus sp. CC9311		113952806	-----A-HI--SE-W L-----GT---A-----	
Synechococcus sp. WH 8016		352094724	-----A-HI--SE-W L-----GT---A-----	
Synechococcus sp. WH 7805		88807693	-----A-HV--SE-W LR-V-----GT---A-----	
Synechococcus sp. WH 7803		148238981	-----A-HV--SE-W LR-V-----GT---A-----	
Synechococcus sp. CC9605		78212267	-----A-HI--SE-W V-----AGT---A-----	
Synechococcus sp. WH 8109		260434708	V-----A-HI--SE-W V-----AGT---A-----	
Synechococcus sp. WH 8102		33866278	-----A-HI--SE-W V-----AGT---A-----	
Synechococcus GOM 3M9		90655440	-----A-HI--SE-W VR-----AGT---A-----	
Synechococcus sp. RS9916		116075324	-----A-HV--EQ-W VR-----AGT---A-----	
Prochlorococcus marinus MIT 9211		159903070	-----SV---KQ-W L-----S--	
Prochlorococcus marinus CCMP1375		33239977	-----AV---KA-W V-----G-----S--	
Prochlorococcus marinus AS9601	123968117	-----AVH---QE-W L--YM-----GS-----S--		

Supplementary Figure 77

Partial sequence alignment for the protein acetolactate synthase 3 catalytic subunit, showing a 1 aa insert that is specific for *Fischerella* sp. JSC-11.



Supplementary Figure 78

Partial sequence alignment for the hypothetical protein FJSC11DRAFT_2574, showing a 1 aa insert that is specific for *Fischerella sp. JSC-11*.

		145	189	
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282896894	YVERCGVTPLAIGDLRQLLTDHDS SP GPTKVLALSHDVLQIQLL	
	Cylindrospermopsis raciborskii CS-505	282901619	----S-----GG-	
	'Nostoc azollae' 0708	298489725	-AQ-----IPVS---V-SNE	
	Anabaena variabilis ATCC 29413	75907973	-AQ-S-I--I-V-----T--NA	
	Nostoc sp. PCC 7120	17231442	-AQ-S-I--I-V-----T--NA	
	Nodularia spumigena CCY9414	119512518	-AQ-S-IN-I-V-----V-SNE	
	Nostoc punctiforme PCC 73102	186686228	-A--SAII-IPV-----T--NE	
	Fischerella sp. JSC-11	354565961	-A--S-IK-I-V-----V--TY	
	Microcoleus chthonoplastes PCC 7420	254415892	-AQ-ADIQ--SV-----V--NE	
	Oscillatoria sp. PCC 6506	300864302	-AQ-SEIQ-I-V-----SV-DTE	
	Microcoleus vaginatus FGP-2	334121385	-A--S-IE-MV-K--RD-PGE	
	Gloeobacter violaceus PCC 7421	37520402	-G--T---QVV---RV--AE	
	Cyanothece sp. PCC 7822	307150695	-AQ-S--EAI-V---NVIAET	
	Cyanothece sp. PCC 7424	218440162	-AL-S--EAVEV---EVI-ET	
	Other Cyanobacteria	Synechococcus sp. JA-2-3B'a(2-13)	86609548	-A--SQ--FQMFSSA-YGRSP
Synechococcus sp. JA-3-3Ab		86605068	-A--SQ--FQMLSPA-HGKSP	
Cyanothece sp. PCC 8802		257061786	-Q--S-IE-VVV---NI-ALK	
Cyanothece sp. PCC 8801		218248724	-QQ-S-IE-VVV---NI-ALK	
Microcystis aeruginosa NIES-843		166363376	-RQ-S-IIAQ-V---KVAHKT	
Microcystis sp. T1-4		390439914	-RQ-S-IVAE-V---KVAHKT	
Cyanothece sp. CCY0110		126661066	-RQ-S-IE-VLVK---SI-EQE	
Cyanothece sp. ATCC 51472		354556048	-RQ-S-IE-VLVE---SI-KQE	
Cyanothece sp. ATCC 51142		172035904	-RQ-S-IE-VLVE---SI-KQE	
Crocospaera watsonii WH 8501		67924677	-RK-S-TE-VLVK---SI-EQE	
Crocospaera watsonii WH 0003		357261329	-RK-S-TE-VLVK---SI-EQE	
Synechocystis sp. PCC 6803		16331052	-M--S-AMAAQ-S---SIIELG	
Synechococcus sp. PCC 7002		170077775	-Q--T-CKANLLT---DIVAQA	
				-----CD-AN---A--
				---I---CD-TEV-NN--
			---I---CD-TEV-SN--	
			---I---CD-TEA-N---	
			---I---D-T---DR--	
			E---I---CD-T-I-DN--	
			-----CQ-TTV-DQ-F	
			---I---CD-PEA-AN--	
			-----CENPGVLDG--	
			---I-G-TPGEATTDR--	
			PT-----CQ-G-IA-Q--	
			PT-----CQ-A-TA-K--	
			--I--V-E-TEFL--Q	
			--I--V-E-TEFL-Q-Q	
			T-----IGKP N-MNQ--	
			T-----IGKP N-MNQ--	
			--I--MGE-KQF-A-IW	
			--I--MGE-KQF-A-IW	
			T--I--IG EPQY-GK--	
			T--I--IG EPNY-GK--	
			T--I--IG EPNY-GK--	
			T--I--IG EPKS-GN--	
			T--I--IG EPKS-GN--	
			ST--M--I--RNAP-MAQ-M	
			---I--VC-TETVPKT--	

Supplementary Figure 79

Partial sequence alignment for the HAD-superfamily hydrolase, showing a 2 aa insert that is specific for *Raphidiopsis brookii* D9.

		104	138
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282896494	YLKEGMEVNVTFWMKQEE KQV LKQVLEVELPNSVV
	Cylindrospermopsis raciborskii CS-505	282899242	-----S-E-N I-----
	Nodularia spumigena CCY9414	119511600	-----IR-G E-----
	'Nostoc azollae' 0708	298490179	-----A--IR-G E-----
	Nostoc sp. PCC 7120	17232550	-----IR-G E-----A----
	Anabaena variabilis ATCC 29413	75908532	-----IR-G E-----A----
	Microcystis aeruginosa PCC 7806	159028172	-----IL--N E-----T--I
	Thermosynechococcus elongatus	22298837	-----IVK-G E-----
	Synechococcus sp. CB0205	317968392	-----VY-N GKP-----
	Synechococcus sp. CB0101	318042998	-----VY-N GKP-----
	Synechococcus elongatus PCC 6301	56751554	-----A--IT-N G--I-----
	Acaryochloris marina MBIC11017	158337061	-----IQ-G E-----T---
	Microcystis aeruginosa NIES-843	166368434	-----IL--N D--DI---T--I
	Synechococcus sp. CC9311	113955533	-----VT-N EKP-----
	Synechococcus sp. WH 8016	352096751	-----VT-N DKP-----
	Synechococcus sp. WH 7805	88809333	-----VS-N DKP-----
	Cyanobium sp. PCC 7001	254430413	-----VS-N GKP-----
	Other Cyanobacteria	Synechococcus sp. RS9917	87123318
Synechococcus sp. WH 5701		87301580	-----VS-N GSP-----
Synechococcus sp. WH 8102		33864569	-----VS-N GSP-----
Cyanothece sp. PCC 7822		307153833	---Q---IL--N D-----T--I
Synechococcus sp. CC9902		78183613	-----VS-N DTP-----
Synechococcus sp. CC9605		78211588	-----VS-N DTP-----
Synechococcus sp. WH 8109		260434331	-----VS-N DTP-----
Synechococcus sp. RCC307		148241130	-----VS-K DNPI-----
Cyanothece sp. PCC 7425		220906117	-----AS-IR-G E--M-----
Synechococcus sp. RS9916		116074306	-----S-VS-N EKP-----
Synechococcus sp. PCC 7002		170076688	--M-E----L--N S--ID----T--
Cyanothece sp. PCC 7424		218441974	-I--N----LY-G E--NI---T--I
Cyanothece sp. PCC 8801		218245423	---E--ADIV--G D----Q--T--I
Fischerella sp. JSC-11		354566434	-----A--IR-G E-----
Microcoleus chthonoplastes PCC 7420		254415365	--S-----VT-N D-----
Trichodesmium erythraeum IMS101		113475050	--N-----IM-G E--I-----
Lyngbya sp. PCC 8106		119487326	--T-----IR-N D-----
Arthrospira maxima CS-328		209527755	--I-----LY-N D-----
Oscillatoria sp. PCC 6506		300868625	--T-----ILR-G E-I-----A--
Microcoleus vaginatus FGP-2		334117257	--N----SILR-N D-IM-----
Moorea producta 3L		332711075	--T-E---ILT-N D-I-----
Prochlorococcus marinus NATL1A		124024738	-----VS-N EKP-----
Prochlorococcus marinus CCMP1375		33239478	-----VS-N EKP-----
Prochlorococcus marinus MIT 9303		124021745	-----VS-N GNP-----
Prochlorococcus marinus AS9601		123967562	-----I-HN GK-----I-IT

Supplementary Figure 80

Partial sequence alignment for the protein Elongation factor P, showing a 3 aa insert that is specific for *Raphidiopsis brookii* D9.

Nostocales/ Stigonematales	Raphidiopsis brookii D9	282895573	AYQIADSTGKIWVVI	NQNSG	QNQNQLQLGQEVVVIKGGKVKYKGI
	Cylindrospermopsis raciborskii CS-505	282901611	-----R-----		---K-----R--I-H-S-
	'Nostoc azollae' 0708	298490661	-----L-----T		--T--V--DA-F--RI--S-
	Nodularia spumigena CCY9414	119509105	----N-----IT		D-P--KE-EQ-----R--S-
	Nostoc punctiforme PCC 73102	186683900	----N-----IT		--K--AK-AQ-----IR-QS-
	Anabaena variabilis ATCC 29413	75908914	V--N---T---T		K---KP-ES--FQ---R-ES-
	Nostoc sp. PCC 7120	17227707	V-K-N---T---LT		K---KP-ES--F----Q--S-
	Fischerella sp. JSC-11	354569183	--LVR-E-----IT		--T--RV-DQ----QIR-RS-
	Microcoleus vaginatus FGP-2	334119747	---LE-G--S--ILT		T-DLP---D--LL--E-R--S-
	Lyngbya sp. PCC 8106	119486290	---LE-N--TV--FS		TDDLPEK---ILVR-Q-N-ET-
	Microcoleus chthonoplastes PCC 7420	254409457	---VQ-T-----R		--ELPNV-DQ---G-Q-QFQS-
	Arthrospira platensis Paraca	284051539	---LQ---T--FT		T-PLPSI--QMF-R-Q-N-EQ-
	Trichodesmium erythraeum IMS10	113477589	--ELT---S--ILT		RAELPKI-AKIT----IH-QS-
	Arthrospira maxima CS-328	209526713	---LQ---AT---FT		T-PLPSV-EQML-R-Q-N-EQ-
	Moorea producta 3L	332706103	--ELK-A-ST---S		QGTLPNI-DQ-L-H--AQFQS-
Other Cyanobacteria	Cyanothece sp. PCC 7822	307151608	--RLQ-H--SV--RT		---LPKP--AL----Q-D-QS-
	Microcystis aeruginosa NIES-843	166367416	--LLQ-KS---IRT		NSNKLPRQ-E-I-----IDFEA-
	Microcystis aeruginosa PCC 7806	159030389	--LLQ-GS---IRT		NSNKLPRQ-E-I-----IDFEA-
	Cyanothece sp. PCC 7424	218437606	---LQ-T--AV--RT		--SLP-T-TQLM--QLD-QS-
	Crocospaera watsonii WH 0003	357265269	S--LQ-E--TV--LT		-GTLP-T-ET-Q-T---T-QS-
	Crocospaera watsonii WH 8501	67921353	S--LQ-E--TV--LT		-GTLP-T-ET-Q-T---T-QS-
	Synechococcus sp. PCC 7002	170077354	--ELQ----R---IT		ETPLPEP-A-LA-A-QLQFHEL
	Cyanothece sp. ATCC 51142	172035728	S--LQ-Q--TV--LT		-GPLP-T-ET-Q-T---T-QS-
	Cyanothece sp. PCC 8801	218246428	S--LR-E--TV--LT		-GALPRS-DRIM--Q-E-QS-
	Cyanothece sp. CCY0110	126658295	S--LE-Q--TV--LT		-GPLP-T-DR-Q-T---T-QS-
	Synechocystis sp. PCC 6803	16332144	---Q-K--RV--KT		DRQLPRK-TVLS-R-E-AFEP-
	Synechococcus sp. PCC 7335	254422245	I---K-D--SL--LT		R-SDPNV-EIATVS-I--HEA-

Supplementary Figure 81

Partial sequence alignment for the hypothetical protein Aazo_1507, showing a 5 aa insert that is specific for *Raphidiopsis brookii* D9.

		531	583	
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282897267	EFIIANHIKAHPLALLHFDELEEDLTKYKIVEL	GIDSISLNPDSVMKTMLEIA
	Cylindrospermopsis raciborskii CS-505	282899812	-----DE---H--A--	RFLVEL -----I-----
	Nostoc sp. PCC 7120	17228131	---N-----I-----DE-A---A--	RFLVEQ -----L--L----
	Anabaena variabilis ATCC 29413	75910765	---N-----I-----DE-A---A--	RFLVEQ K-----L--L----
	Nostoc flagelliforme Sunitezuoqi	308205536	----Q-QI--M--I-Y-L-KDEFA-A--A-I	LVEQ -----L-----V-
	Nostoc punctiforme PCC 73102	186680567	----Q-QI--M--INVEL-KDEFA-A--A-I	LVEQ -----L-----V-
	Anabaena variabilis ATCC 29413	75910045	----Q-QV----I-Y-K-ADAEQ-A-VAAI	LVEE -----L-----
	Fischerella sp. JSC-11	354568885	-----I-----DE-A---A--	RFLVEE -----L--L----
	Arthrospira maxima CS-328	209524357	-----KY-Q--DESV-RE-G--	RFLVEL ----M-----I--I-D-G
	Arthrospira platensis NIES-39	291571801	-----KY-Q--DESV-RE-G--	RFLVEL ----M-----I--I-D-G
	Arthrospira sp. PCC 8005	376007376	-----KY-Q--DESV-RE-G--	RFLVEL ----M-----I--I-D-G
	Oscillatoria sp. PCC 6506	300865669	-----M--IK--DEAV-NE-A--	RFLVEL ----L-----L--L-D-
	Lyngbya sp. PCC 8106	119484375	-----I-Y----DE--REVA--	RFLVEL -----I--I-D-G
	Microcoleus chthonoplastes PCC 7420	254415120	-----V----N--K--DATV-RE-YKM	LVEL ----M-----L-MV-
	Microcoleus vaginatus FGP-2	334120710	-----I-----VDESV-EE-AA-	RFLVEL -----L--L-DV-
	Moorea producta 3L	332709390	-----M--I--K--DPSV-EE-AQ-	LVEL -----L--L-D-
	Arthrospira platensis NIES-39	291566502	----Q-RV----I--Q-QDQAAEQE-EK-	RFLVDL -----LL--L--V-
	Arthrospira maxima CS-328	209528186	----Q-RV----IN-YQ-QDETARQE-DR-	RFLVNL -----LL--L--V-
	Arthrospira maxima CS-328	209524136	----Q-RV----IN-YQ-QDETARQE-DR-	RFLVNL -----LL--L--V-
	Arthrospira sp. PCC 8005	376004060	----Q-RV----IN-YQ-QDETARQE-DR-	RFLVNL -----LL--L--V-
	Trichodesmium erythraeum IMS101	113477623	--V----GV----N--VMDEVE-A--AK-	RLLVEL -----L--RF-V-
	Lyngbya sp. PCC 8106	119484520	--L-N-I--V--M--I--KITNYSIRHE-S-M	LVEA -----V-VKRRV-
	Crocospaera watsonii WH 8501	67924504	-----MK----DEAV-EQ-A--	LVEL ----M-----L--L-MV-
	Cyanothece sp. PCC 7425	220907425	-----MN--S-QDKAA-WE-A--	LVEQ -----L--L-A--
	Crocospaera watsonii WH 0003	357261548	-----MK----DEAV-EQ-A--	LVEL ----M-----L--L-MV-
Cyanothece sp. PCC 8802	257060661	-----IK----DESV-DQ-S--	LVEL ----M-----L--L-MV-	
Cyanothece sp. PCC 8801	218248015	-----IK----DESV-DQ-S--	LVEL ----M-----L--L-MV-	
Cyanothece sp. PCC 7822	307151509	-----IK----DESV-AQ-AQ-	LVEL ----M-----L--I-MV-	
Cyanothece sp. ATCC 51142	172039605	-----MK----DEGV-EQ-T--	LVEL ----M-----L--L-MV-	
Thermosynechococcus elongatus	22298307	-----Q-----Y--KDEAA-RQ-A--	LVAQ -----L-----Q--	
Cyanothece sp. CCY0110	126660019	-----MK----DEAV-EQ-A--	LVEL ----M-----L--L-MV-	
Cyanothece sp. PCC 7424	218438806	---N-----K--Q--DENA-RN-AQ-	LVEL -----L---MV-	
Acaryochloris sp. CCME 5410	359461132	-----I-----Q--DP-AQQE-AH-	LVEQ -----L--RFT--	
Synechococcus elongatus PCC 6301	56750765	-----Q-----Q--DLFVQDQ-AK-	LVEQ -----L--L-A--	
Synechococcus elongatus PCC 7942	81299592	-----Q-----Q--DLFVQDQ-AK-	LVEQ -----L--L-A--	
Synechococcus sp. PCC 7335	254425819	-----V----Q--Q-TSEQD-RA-AQ-	LVEE -----I--T-A--	
Synechocystis sp. PCC 6803	16331869	-----MK----DP-A-AE-A--	LVEL -----L--V-R--	
Acaryochloris marina MBIC11017	158334808	-----I--Q-DDP-AQQE-TH-	LVEQ -----L--RFT--	
Synechococcus sp. PCC 7002	170076879	-----K--I-----DLMV-KE-A--	LVEL -----LL--K-A--	
Microcystis aeruginosa NIES-843	166363003	-----Q--S--K-H--GDV-DQ-A--	LVEL -----L--L-MV-	
Microcystis aeruginosa PCC 7806	159030393	-----Q--S--K-H--GDV-DQ-A--	LVEL -----L--L-MV-	
Cyanobium sp. PCC 7001	254430235	-----V----QPE-VA-AAERQA-A--	LVAE -----A-LQ-R-SV-	
Synechococcus sp. JA-3-3Ab	86605451	-----QV----V-Y--EAV-EERFLV	QQ -----LL--L-VV-	
Synechococcus sp. CB0205	317970224	-----Q-RV--M----LNRVVSQEEQQA-LQ-	LVKE -----A-I--R-A--	
Gloeobacter violaceus PCC 7421	37521554	-----QV-----EQ-P-GDE-ARLSA-	RFLVER ----L-----AAL--R-LV-	

Supplementary Figure 82

Partial sequence alignment for the protein Phosphoenolpyruvate synthase, showing a 5 aa deletion that is specific for *Raphidiopsis brookii* D9.

		282898185	140	176
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282898185	FPRLRIGIGKPKNLPNN SN	NDDGGTVSHVLGKGFSPQE
	Cylindrospermopsis raciborskii CS-505	282898618	-----	---D-----
	'Nostoc azollae' 0708	298490956	----K-----GAA-E	-NTD-----R--AT-
	Nostoc sp. PCC 7120	17228339	-----AV-G	-NSE-----AT-
	Anabaena variabilis ATCC 29413	75910645	-----AV-G	-NSE-----AT-
	Nodularia spumigena CCY9414	119508860	-----NDS-D-S	ESI-----AA-
	Nostoc punctiforme PCC 73102	186684439	-----GAAD-	--SN-----R--SA-
	Fischerella sp. JSC-11	354568542	-----R--GAT-G	-EN--Y-----TT-
	Trichodesmium erythraeum IMS101	113476455	-----ITAE	RGEAK-I-----K-
	Moorea producta 3L	332711589	-----SSQK-DT	I-----
	Oscillatoria sp. PCC 6506	300864784	-----QSSS-D	KE-I-----LE-
	Microcoleus vaginatus FGP-2	334119291	-----Q-DSAK	NEKD-I-F-----SE-
	Lyngbya sp. PCC 8106	119484739	-----D--DPASS	QT-E-I-----S-
	Arthrospira platensis Paraca	284051522	----L--D--QPPASS	S-RE-I-----P-N-
	Microcoleus chthonoplastes PCC 7420	254415168	-----SDPEK-T	-G-----D-
	Arthrospira maxima CS-328	209527283	----L--D--QAAVL-	H-RE-I-----P-N-
	Crocospaera watsonii WH 0003	357265569	-----SNQVD-T	I-----R--E-
	Cyanothece sp. ATCC 51142	172036669	-----SGQ-D-T	-----R-A-E-
	Crocospaera watsonii WH 8501	67925084	-----SNQVD-T	IP----R--E-
	Cyanothece sp. PCC 8801	218248401	----L---SNQSKET	I-----A---
	Cyanothece sp. CCY0110	126658601	-----SNQ-D-T	-----R-A-E-
	Synechococcus sp. PCC 7335	254423495	----V-V-SADKSGDR	--AV-----R--SD
	Cyanothece sp. PCC 7822	307150148	-----SNGKKET	I-----R--AD-
	Cyanothece sp. PCC 7424	218441393	-----SNQKKET	I-----R--AD-
	Cyanothece sp. PCC 7425	220908888	----L--N-RQAGGE	KDAIP-----A-
	Synechococcus elongatus PCC 6301	56751295	----L--SL-RSQS-D	RHDA-----AVS-
	Microcystis aeruginosa NIES-843	166363429	-A-----STEKATI	-----R-T---
	Microcystis aeruginosa PCC 7806	159027229	-A-----SSEKATI	-----R-T---
	Synechococcus sp. PCC 7002	170078845	-----TQLKGDR	P-I-----A-A-
	Acaryochloris sp. CCMEE 5410	359462044	V-----S-PPERGK	D-I-F--R-A-P-
	Acaryochloris marina MBIG11017	158338610	V-----S-PPERGK	D-I-F--R-A-L-
	Synechococcus sp. JA-3-3Ab	86605936	----L-V-R--GNQDR	-G---S-E-A-
	Synechococcus sp. JA-2-3B'a(2-13)	86608210	----L-V---GSQDV	-G---G-A-A-
	Synechococcus sp. WH 8016	352095092	-A-----A-GQN-SE	RKAR-----S--RE-
	Synechococcus sp. WH 7805	88808009	-A-----A-GRS-EE	RRAR-----A--RE-
Synechocystis sp. PCC 6803	16331675	----L---SDGSKDT	IA-----S-	
Synechococcus sp. BL107	116072067	-----A-AEN-AE	RRAR-----P--RA-	
Synechococcus sp. WH 8102	33864807	-----A-ADN-AE	RRAR-----S--RA-	
Synechococcus sp. CB0101	318042729	-----A-ALN-VE	RKQR--G---R-DAA-	
Synechococcus sp. WH 7803	148238651	-A-----A-GRS-EE	RRAR-----A--RE-	
Synechococcus sp. RS9916	116075623	-A-----A-GRSAEE	RRAR-----S-HQA-	
Synechococcus sp. RS9917	87125090	-A-----A-GRTAEE	RRAR-----R--QA-	
Synechococcus sp. CC9311	113953631	-A-----A-GQN-SE	RKAR-----S--RD-	
Synechococcus sp. CC9902	78185646	-----A-AEN-AE	RRAR-----P--RA-	
Synechococcus sp. WH 8109	260436438	-----A-AEN-AE	RRAR-----S--KA-	
Synechococcus sp. CC9605	78211817	-----A-AEN-AE	RRVR-----P--KV-	
Synechococcus sp. WH 5701	87301942	-A-----A-ALD-AL	RKAR-I----R-EAD-	
Cyanobium sp. PCC 7001	254431920	-----A-AEN-AE	RRAR-IG---R-AA-	
Synechococcus sp. RCC307	148243439	-A-----A-GTSSEE	RKE--G---N-H-A-	
Synechococcus sp. CB0205	317968636	-----A-ALN-VE	RKQR--G---R-AVA-	
Thermosynechococcus elongatus BP-1	22298927	---KV--AFGDR-QQQ	TGPRNA-PF--H--AT-	
Prochlorococcus marinus MIT 9313	33864102	-C-----A-GCTSEE	RRAR-----V--R--	
Prochlorococcus marinus NATL1A	124025041	-K-K---A-S-DQQE	RKSK-----R--KE-	
Prochlorococcus marinus AS9601	123967809	-K-I----S-PITNGK	NHFN-I-----NI-LD-	

Supplementary Figure 83

Partial sequence alignment for the protein peptidyl-tRNA hydrolase, showing a 2 aa insert that is specific for *Raphidiopsis brookii* D9.

		672	733
Nostocales/ Stigonematales	<i>Cylindrospermopsis raciborskii</i> CS-505	282898629	LGDFSLAEVARAQIRTALHQGTET MNNS QQNVNHNNSNLSLFLKQAWEEAKDLKLAEEYRLL
	<i>Raphidiopsis brookii</i> D9	282898173	-----
	' <i>Nostoc azollae</i> ' 0708	298492614	-----T-S-----AG--Q--
	<i>Nostoc punctiforme</i> PCC 73102	186684554	-----T-S-----A--GESTI
	<i>Nodularia spumigena</i> CCY9414	119511881	-----T-S-----I-GESNL
	<i>Anabaena variabilis</i> ATCC 29413	75907647	-----T-S-----A--GE-GI
	<i>Nostoc</i> sp. PCC 7120	17227661	-----T-S-----A--GE-GI
	<i>Fischerella</i> sp. JSC-11	354569011	-----T-S-----A--NQGSL
	<i>Cyanothece</i> sp. PCC 7425	220909979	-----T-----HR--QSQS
	<i>Microcystis aeruginosa</i> PCC 7806	159030183	-----A-V--RYLN
	<i>Cyanothece</i> sp. PCC 8801	218247595	-----A-S-----A-V-SQYIN
	<i>Microcystis aeruginosa</i> NIES-843	166363409	---A-T-----A-V--RYLN
	<i>Cyanothece</i> sp. CCY0110	126659573	---N-T---S---A-V-HNYIN
	<i>Cyanothece</i> sp. PCC 7822	307152601	-----T-----A-V-SQYQN
	Other Cyanobacteria	<i>Cyanothece</i> sp. ATCC 51142	172037191
<i>Cyanothece</i> sp. ATCC 51472		354553922	---N-T---S---A-V-HNYIN
<i>Crocospaera watsonii</i> WH 0003		357262486	---E-T---S---A-V-HQYIN
<i>Crocospaera watsonii</i> WH 8501		67923682	---E-T---S---A-V-HQYIN
<i>Cyanothece</i> sp. PCC 7424		218441857	-----T-S-----A-V-SQYQN
<i>Synechocystis</i> sp. PCC 6803		16329962	---Q-T-----VV-HRQQQ
<i>Acaryochloris</i> sp. CCMEE 5410		359460789	---P-----ANV-KQFPL
<i>Acaryochloris marina</i> MBIC11017		158337497	---P-----ANV-NQVPL
<i>Synechococcus</i> sp. PCC 7335		254424367	---K---D-----L--YT-GVYAQ
<i>cyanobacterium</i> UCYN-A		284929239	---N-----I--T--ALS-EKYLS
<i>Synechococcus</i> sp. PCC 7002		170078704	---YN--Q-----HFV-GQDQG
<i>Synechococcus elongatus</i> PCC 7942		25019676	---PQPWIAI--L-L-R--QAE-GA
<i>Synechococcus elongatus</i> PCC 6301		56751600	---PQPWIAI--L-L-R--QAE-GA
<i>Synechococcus</i> sp. JA-3-3Ab		86605069	-E-VDMRAE-L-L-QLSRRSCW
<i>Arthrospira maxima</i> CS-328		209523619	-----T-S-----AN--NSQQL
<i>Lyngbya</i> sp. PCC 8106		119491302	---E-T-S-----AN--QHPL
<i>Trichodesmium erythraeum</i> IMS101		113476147	-----D-T-----AS-Q-ENFP
<i>Arthrospira platensis</i> Paraca		284050462	-----T-P-----AY--NYQQL
<i>Microcoleus chthonoplastes</i> PCC 7420		254416448	-----T-----AS--EHPLP
<i>Microcoleus vaginatus</i> FGP-2		334118608	-----T-----AS--EKYPL
<i>Oscillatoria</i> sp. PCC 6506	300865249	-----T-----AS--QFPL	
<i>Moorea producta</i> 3L	332710143	---E-T-----AY--GKPLP	
			K--A-----
			IPNVS---Q--Y--M-----
			PDVT---Q--H--IS-----
			PHVS---Q--H--TS-----
			P-VTS---Q--H--T-----
			P-VTS---Q--H--T-----
			PDVT---Q--Y--T-----
			AHQPTGRSP-PDTAT--QR-A--I--F--
			PD-IPIPPPLS--L--GQ--K-----
			QGIISQIR-PET--Q--GQ--K-----
			PD-IPIPPPLS--L--GQ--K-----
			P-DSLIIPNPED---QQ--K-----
			PDILPRIPDPLK---NK--E--F--
			P-DPLIIPNPEA---QQ--K-----
			P-DPLIIPNPEA---QQ--K-----
			PENTLIIPNPDN--Q--QQ--K-----
			PENTLIIPNPDN--Q--QQ--K-----
			SDILPRIPHPVD---NK--W--F--
			KPQAIA-PVTAT--Q--RR--E-----
			PNMEE--QR-S--R--F--
			PNVEE--QR-S--S--F--
			DRKPAPINIED--QQ--TN--R--F--
			YEN-ICTN-I--V--QS-L--K-----
			R-PTAPATITTL-Q--QQ--TE-----
			GVEPFTD-AA-VS--QA-RE-----
			GVEPFTD-AA-VS--QA-RE-----
			EQQADS-T-Q-FLDLA--RL--
			PDITT---Q--Y--T-----
			PNVM---Q--GY--T-----
			DISK---Q--GY--M--F--
			PDITT---Q--Y--N-----
			TALE---K-GQ--T-----
			PN-GE---Q--GY--T--F--
			PDIAT---L-GY--MS--F--
			MAPE---Q-AQ--T--F--

Supplementary Figure 84

Partial sequence alignment for the protein UvrD/REP helicase, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505.

		84	139
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898905	GWIGESSIVPLIEAWLQSLSLPGNLGN LSN VVVAHSLSIPLTFFLIAYLQIVLDEL
	'Nostoc azollae' 0708	298492858	-----T--V-M-E--K-WTV-IS-S-
	Nodularia spumigena CCY9414	119511111	-----V-VKS--KTWP--LEMS-
	Nostoc punctiforme PCC 73102	186684590	-----T--V-VN---R-WP--SGMT-
	Anabaena variabilis ATCC 29413	75908578	-----AVVMRW-IK-WP--A-VNH
	Nostoc sp. PCC 7120	17232588	-----AVVMRW-IK-WP--A-VNH
	Fischerella sp. JSC-11	354568930	-----T--V-VKS---WP--GINI
	Acaryochloris marina MBIC11017	158336061	-----HTVAV--IS--RQSA-SDSTQV
	Acaryochloris sp. CMEE 5410	359462657	-----HTVAV--IS--RQSA-SDSTQV
	Cyanothece sp. CCY0110	126658843	-----TMAK-VMEV-SK-P--PLVA-
	Cyanothece sp. ATCC 51142	172036207	-----TMAK-VMEG-SK-P-SPSVA-
	Synechococcus elongatus PCC 6301	56751524	--V--T-AV--RQA-EQ-P--AIGPE
	Cyanothece sp. PCC 8802	257061886	-----GTMAV-VRYL-KH-P-SDQWT-
	Cyanothece sp. PCC 8801	218248824	-----GTMAV-VRYL-KH-P-SDQWT-
	Other Cyanobacteria	Synechococcus sp. PCC 7002	170077826
Synechococcus sp. PCC 7335		254421990	-----HTMAA--IH--AR-PIQDRFSE
Crocospaera watsonii WH 8501		67924320	-----T-AK-VLEI-ST-P-SEVFLK
Cyanothece sp. PCC 7822		307150928	-----GTMAV-LKQLIVNIPM-SGTLD
Crocospaera watsonii WH 0003		357261792	-----T-AK-VLEI-ST-P-SEVFLK
Cyanothece sp. PCC 7424		218437698	-----DT-AVVLKQLIIT-PIASETLD
Microcystis aeruginosa NIES-843		166364246	-----TMTVVRKL-SF-PFSL-WQ-
Microcystis aeruginosa PCC 7806		159028903	--T--TMTVVRKL-SF-PFSL-WQ-
Microcoleus vaginatus FGP-2		334119232	--L--GAT-HW--P--DK-PIGHLSAV
Microcoleus chthonoplastes PCC 7420		254410841	-----NTVAVFVA-V-TK-P--PEMVQ
Arthrospira platensis Paraca		284052758	-----TMAV--RQI-TKIP-HQFQVA
Arthrospira platensis NIES-39		291566993	-----TMAV--RQI-TKIP-HQFQVA
Moorea producta 3L		332706606	-----TTMAT-VA-GIAK-PITPA-TP
Microcoleus vaginatus FGP-2		334120963	-----T-AV-VRMSIGY-P--IYIKS
Arthrospira sp. PCC 8005		376007425	-----TMAVV-RQI-NQIP-YQFQ-W
Arthrospira maxima CS-328		209522818	-----TMAVV-RQI-NQIP-YQFQ-W
Trichodesmium erythraeum IMS101		113475751	----AKTM-EIVTNAIKTIP--IHI-Q
Gloeobacter violaceus PCC 7421		37520354	-----GT-AA--PI-SGF-IAAGSAL
			-L-----V-I-----
			FL-----IA-----
			FL-----IA-----I---
			F-----IA-----
			F-----IA-----
			AI-----IA--F-----
			FI--T-AV-IA-LV-----
			FI--T-AV-IA-LV-----
			SIS--IA--IA-LAL-----
			RIS--IA--IA-LAL-----
			PLS-V-A--A-A-LV-----
			TLS--FA--IA--AL-----
			TLS--FA--IA--AL-----
			PF-V-I---FA-----
			FFS--IAL-SA-L-L-----
			NL--IA--VA-LSL-----
			YL--GIA--VA-V-L-----
			SL--IA--VA-LSL-----
			YI--GIA--A-I-LV-----
			LLS-GF-LC-A-LIV-----F---
			LLS-GF-LC-A-LIV-----F---
			LT--IAV-AA-I-VT-VE-----
			GM---A--VG-----
			IAHSLAIPV---ML-----
			IAHSLAIPV---ML-----
			-IT--A-LVA-----
			IAIAHFSIPIV--L-L-----
			LAHLSISIV---ML-----
			LAHLSISIV---ML-----
			GNDDSVAITIV--LIL-----
			-TI-TIVA-L--T-MH-----

Supplementary Figure 85

Partial sequence alignment for the hypothetical protein Aazo_4712, showing a 3 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505.

	139								
	Cylindrospermopsis raciborskii CS-505	282898905	QLGITLSSLALGWIGESSIVPLIEAWLQSLSLPGNLGN	LSNV	VVAHLSLSIP-LTFFLIAYLQIVLQV				
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298492858	-----T--V-M-E--K-WTV-IS-S-		FL-----V- I-----				
	Nostoc punctiforme PCC 73102	186684590	-----T--V-VN--R-WP--SGMT-		FL----- IA-----I---				
	Nodularia spumigena CCY9414	119511111	-----V-VKS--KTWP--LEMS-		FL----- IA-----				
	Anabaena variabilis ATCC 29413	75908578	-----V-----AVVMRW- IK-WP--A-VN-		F----- IA-----				
	Nostoc sp. PCC 7120	17232588	-----V-----AVVMRW- IK-WP--A-VNH		F----- IA-----				
	Fischerella sp. JSC-11	354568930	-----T--V-VKS---WP--GINI		AI----- IA--F-----				
	Microcoleus chthonoplastes PCC 7420	254410841	-----NTVAVFVA-V-TK-P--PEMVQ		GM----A-- VG-----				
	Oscillatoria sp. PCC 6506	300866192	-----GTMAI-VRRF-AY-P--A-ISE		AI----G-SI----L-----				
	Lyngbya sp. PCC 8106	119486081	-F-----GTLAVA-R-LIV--PI-L--QD		FI----A--IF--V-----				
	Microcoleus vaginatus FGP-2	334120963	-----T-AV-VRMSIGY-P--IYIKS		AI---F---IV--L-L-----				
Other Cyanobacteria	Arthrospira platensis Paraca	284052758	-F-----TMAV--RQI-TKIP-HQFQVA		I---A--IV--ML-----				
	Arthrospira platensis NIES-39	291566993	-F-----TMAV--RQI-TKIP-HQFQVA		I---A--IV--ML-----				
	Moorea producta 3L	332706606	-----TTMAT-VA-GIAK-PITPA-TP		-IT---A-L VA-----				
	Microcoleus vaginatus FGP-2	334119232	TGTTAG-L- ---L-GAT-HW--P--DK-PIGHL SAV		LT---IAV- AA-I-VT-VE---				
	Arthrospira sp. PCC 8005	376007425	-F-----TMAVV-RQI-NQIP-YQFQ-W		L-----SIV---ML-----				
	Arthrospira maxima CS-328	209522818	-F-----TMAVV-RQI-NQIP-YQFQ-W		L-----SIV---ML-----				
	Trichodesmium erythraeum IMS101	113475751	-----FT----AKTM-EIVTNAIKTIP--IHI-Q		GNDD-VA-TIV--L-LIL-----				
	Acaryochloris marina MBIC11017	158336061	-I-----HTVAV--IS--RQSA-SDSTQV		FI--T-AV- IA-LV-----				
	Acaryochloris sp. CCMEE 5410	359462657	-I-----HTVAV--IS--RQSA-SDSTQV		FI--T-AV- IA-LV-----				
	Cyanothece sp. CCY0110	126658843	-----TMAK-VMEV-SK-P--PLVA-		SIS--IA-- IA-LAL-----				
	Cyanothece sp. ATCC 51142	172036207	-----TMAK-VMEV-SK-P--SPSVA-		RIS--IA-- IA-LAL-----				
	Cyanothece sp. PCC 8802	257061886	-----GTMAV-VRYL-KH-P--SDQWT-		TLS--FA-- IA--AL-----				
	Cyanothece sp. PCC 8801	218248824	-----GTMAV-VRYL-KH-P--SDQWT-		TLS--FA-- IA--AL-----				
	Synechococcus sp. PCC 7002	170077826	-----DTMATV-AQF-VG-P--DDWIQ		PF-V-I--- FA-----				
	Synechococcus elongatus PCC 6301	56751524	-----A-----V--T-AV--RQA-EQ-P--AIGPE		PLS-V-A-- -A-A-LV-----				
	Synechococcus sp. PCC 7335	254421990	-----HTMAA--IH--AR-PIQDRFSE		FFS--IAL- SA-L-L-----				
	Crocospaera watsonii WH 8501	67924320	-----T-AK-VLEI-ST-P-SEVFLK		NL---IA-- VA-LSL-----				
	Cyanothece sp. PCC 7822	307150928	-----GTMAV-LKQLIVNIPM-SGTLT		YL--GIA-- VA-V-L-----				
	Crocospaera watsonii WH 0003	357261792	-----T-AK-VLEI-ST-P-SEVFLK		SL---IA-- VA-LSL-----				
	Microcystis aeruginosa NIES-843	166364246	-----TMTVFVRKL-SF-PFSL-WQ-		LLS-GF-LC -A-LIV-----F---				
Microcystis aeruginosa PCC 7806	159028903	-----T---TMTVFVKKL-SF-PFSL-WQ-		LLS-GF-LC -A-LIV-----F---					
Cyanothece sp. PCC 7424	218437698	-----A-----DT-AVVLKQLIIT-PIASETLD		YI--GIA-- -A-I-LV-----					
Gloeobacter violaceus PCC 7421	37520354	-----IA-----GT-AA---PI-SGF-IAAGSAL		--TI-TI VA-L--T-MH-----					

Supplementary Figure 86

Partial sequence alignment for the hypothetical protein Aazo_4712, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505.

		240	287
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899315	GVDAKIAYSRLPDAYQLTQALGPRD GVYPDISFLEVCEDTLFLVSDGL
	Nodularia spumigena CCY9414	119511123	--E-S---A-----QH SIN--VE-FQIT--S-I-----
	'Nostoc azollae' 0708	298492439	--ESA--A-----I----AN F-H-EVD-F-IS---L-A---
	Nostoc punctiforme PCC 73102	186681193	--E-S---A-----EN SIN--VG-F-IN---L-A---
	Anabaena variabilis ATCC 29413	75906579	--E-S---A-----ET AIN--VE-FDIN--S-LV-A---
	Nostoc sp. PCC 7120	17228043	--E-S---A-----ET AIN-SVE-FDIN--S-LV-A---
	Fischerella sp. JSC-11	354568421	--EPSV--T-----EH SISA-VQ--DIA----I-A---
	Crocospaera watsonii WH 8501	67923494	---P-L--A-----HN NK Y-Q---L---LA--S---C---
	Crocospaera watsonii WH 0003	357262665	---P-L--A-----HN NK Y-Q---L---LA--S---C---
	Cyanothece sp. PCC 8802	257060152	--EPE---A-----H NS CIQ---R--D-Q-----C---
	Cyanothece sp. ATCC 51142	172035066	--IEP-L-A-----H NK YIQ---R-IDLA-----C---
	Cyanothece sp. PCC 8801	218247082	--EPE---A-----H NS CIQ---R--D-Q-----C---
	Cyanothece sp. CCY0110	126660989	--EP-L--A-----HE NK YLQ---R-IDLV-----C---
	Synechococcus elongatus PCC 7942	81300324	---PDV--A--E-H-----N ND WLD-EVR---R---V---C---
	Synechococcus elongatus PCC 6301	6689332	---PDV--A--E-H-----N ND WLD-EVR---R---V---C---
	Cyanothece sp. PCC 7424	218437039	--EPQ---G-----HE SQ YIK---KY-DLH---L-C---
	Other Cyanobacteria	Synechococcus sp. PCC 7002	170079294
Cyanothece sp. PCC 7822		307152030	--IAPE-----HE NQ Y-K---KY-DLH---L-C---
Thermosynechococcus elongatus		22299187	--LEPQ---G-----N NE AIH-EVTD--IE---I-C---
Acaryochloris marina MBIC11017		158338458	--T-PAV--A-----E NE ALE---QL-TAS---L-IC---
Acaryochloris sp. CCMEE 5410		359459059	--T-PAV--A-----E NE ALE---QL-TAS---L-IC---
Cyanothece sp. PCC 7425		220906655	--ER---A---H-----KR QE SLH-S-QS--LE--SVL--C---
Synechococcus sp. PCC 7335		254423007	--PV---E-----S QR DLV-GVTY-NFS---L-C---
Synechocystis sp. PCC 6803		16329363	--L-PTL--A-----HS SS YLQ--VAIF-IE--C-I--C---I
Arthrospira platensis Paraca		284052402	---PET--N-----GK YLK---E-FDIT---L-A---
Arthrospira platensis NIES-39		291570747	---PET--N-----GK YLK---E-FDIT---L-A---
Microcoleus vaginatus FGP-2		334116884	---PET--A-----I---SN FLK--VQ--DL-----LI-A---
Microcoleus chthonoplastes PCC 7420		254412389	--EEA-----S DQ F-T---N---LN-----I-C---
Moorea producta 3L		332707025	---PEL-----S EQ FLN--VQ-IDIA---I---C---
Arthrospira sp. PCC 8005		376002957	---PE---T-----SK YLK--VR-WDIT---L-A---
Arthrospira maxima CS-328		209528170	---PE---T-----SK YLK--VR-WDIT---L-A---
Oscillatoria sp. PCC 6506		300866685	---PQT-----NN F-K-EVQ-I-LL---LI-A---
Lyngbya sp. PCC 8106		119485434	--SPDV--A-----RN F-K---Q---Q---V-I-A---
Trichodesmium erythraeum IMS101	113474474	--ELQ---G-----N EN F-R---Q--NLKN---LI-A---	

Supplementary Figure 87

Partial sequence alignment for the protein serine/threonine phosphatase, showing a 2 aa deletion that is specific for *Cylindrospermopsis raciborskii* CS-505.

6

64

Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282898751	KTLCPCYCGVCGGLEVSSPTQHPSFTTDN	TG	DQETVNWVRVGRGDKSHPSQGMVCKGATV
	Raphidiopsis brookii D9	282898022	-----L-P-----		-----I
	Nostoc punctiforme PCC 73102	186681962	-----P-A-LGKA-NRD		G-GNPT-----A-----I
	Nodularia spumigena CCY9414	119510727	-----P-A-LAKP-NRD		S-GNPI-----A----K-----I
	Nostoc sp. PCC 7120	17228108	-----TP-A-LNKA-NRD		S-GNPT-----A-----I
	Anabaena variabilis ATCC 29413	75910741	-----TP-A-LNKA-NRD		S-GNPI-----A-----I
	Fischerella sp. JSC-11	354565510	-----P-A-PGKA-HRD		G-GNPI-----A-----I
	Arthrospira maxima CS-328	209525537	-----LP-A-PGKA-NRD		S-GNPI-Q----RT----L-K-----I
	Arthrospira sp. PCC 8005	376001959	-----LP-A-PGKA-NRD		S-GNPI-Q----RT----L-K-----I
	Microcoleus chthonoplastes PCC 7420	254414982	-----TP-A-SGKA-HRD		SEG-PI-----RD-----K-----
Trichodesmium sp. WH 9601	6007444	--A-----P-ATPGKV-NRD		SKGNPT-Q----R-----K-L-----	
Arthrospira platensis Paraca	284050142	-----LP-A-PGKA-NRD		SLGNPI-Q----RT----L-K-----I	
Oscillatoria chalybea	899356	-----P-ARSGKPVNRD		S-GYPV-K-Q--RE-----L-K-----I	
Arthrospira platensis NIES-39	291567083	-----LP-A-PGKA-NRD		SLGNPI-Q----RT----L-K-----I	
Moorea producta 3L	332711104	-----LP-AGSGRA-YRD		S-GNPI-K----R----N-----	
Oscillatoria sp. PCC 6506	300868377	-----KPGQS		-K-K--RE---M-----	
Lyngbya sp. PCC 8106	119485488	-----D-LP-ALPGKA-NRD		K-GNPT-Q----RN-----K-----	
Trichodesmium erythraeum IMS101	113474851	--A-----P-ATPGKV-NRD		SKGNPT-Q----R-----K-L-----	
Microcoleus vaginatus FGP-2	334118437	-----VPDKKQS		-K----RN-----I-----	
Cyanothece sp. PCC 8802	257061406	-----LP-A-PGKP-NRD		S-G-PI-Q----R-----K-----	
Cyanothece sp. PCC 8801	218247258	-----LP-A-PGKP-NRD		S-G-PI-Q----R-----K-----	
Synechococcus elongatus PCC 6301	56750324	-----AVP-A-PGRA-VRD		REG-PI-QI---RQ-----	
Synechococcus elongatus PCC 7942	81300044	-----AVP-A-PGRA-VRD		REG-PI-QI---RQ-----	
Cyanothece sp. PCC 7822	307151189	-----P-ASAGRA-HRD		S-GNPI-K-TA-RA----K-----	
Cyanothece sp. PCC 7424	218440394	-----P-ASTTHA-HRD		S-GKPI-K-M--R----K-----I	
Microcystis aeruginosa NIES-843	166368137	--Q-----T-T--ASAGQA		-K-F-----	
Microcystis aeruginosa PCC 7806	159027137	--Q-----T-T-ASAGQA		-K-F-----	
Crocospaera watsonii WH 0003	357263191	--I-----LP-A-PGKAVNRD		SEGNI-Q----R----K-KI-I--G--	
Crocospaera watsonii WH 8501	67923091	--I-----LP-A-PGKAVNRD		SEGNI-Q----R----K-KI-I--G--	
Cyanothece sp. ATCC 51142	172036129	-----LP-A-PGKPVNRD		SEG-PI-Q----RN----L-K-----	
Acaryochloris marina MBIC11017	158336125	-----LP-ALPGKPVNRD		SAG-PI-----QA----K-Q-----	
Cyanothece sp. PCC 7425	220909924	-----LTSETNAVAI		KQ---R-----I	
Cyanothece sp. CCY0110	126656640	-----LP-A-PGKPVNRD		SEG-PI-Q----RN----L-K-----	
Synechococcus sp. JA-2-3B'a(2-13)	86607542	-----AV-TASGG		-KI--RE---L-----	
Thermosynechococcus elongatus BP-1	22298898	SS-----AVPQGGKT		YKI---RQ-----I	
Synechocystis sp. PCC 6803	16330080	R-----AVASP-KSVVDAGH		AHKI--RQ-----	
Synechococcus sp. JA-3-3Ab	86605334	-----AVATASGD		-KM--RE---L-----	
Synechococcus sp. WH 5701	87300726	-AQ-----LKP-AADANSPDLP		-AAEAI-T----RQ----L-Q-----	
Synechococcus sp. RCC307	148243588	TSQ-----LLP-A-PGRPVRD		GEGHPM-SA---RA-----Q--I----	
Synechococcus sp. RS9916	116074222	RSQ-----LLP-A-PGQAVKGD		EEGNPM-TA---R-----Q--I----	
Synechococcus sp. PCC 7335	254425255	-----FP-ARPRKI-RD		S-GRPI-QAV-N-A---K-Q--I--SI	
Synechococcus sp. CB0101	318040841	RSQ-----LMP-A-PGQAVKRD		AEGNPV-TA---RQ-----Q--I----	
Synechococcus sp. PCC 7002	170077929	-----EVIKNEQP-		LK-M--RQ----L-K-----I	
Synechococcus sp. WH 8102	33866994	RSQ-----LLP-AVKQAVKRD		AEGNPM-TA---RE----L-Q--I----	
Synechococcus sp. CC9311	113954017	RSQ-----DLQP-AKKGEAVRRD		ADG-PM-TA---RQ----L-Q--I----	
Synechococcus sp. WH 8016	352095743	RSQ-----DLRP-A-KGEAVRRD		AEG-PM-TA---NRL----L-Q--I----	
Synechococcus sp. BL107	116071858	RSQ-----LLP-AEKGRAVRRD		AEGNPM-TA---RQ----L-Q--I----	
Synechococcus sp. WH 8109	260435619	RSQ-----LLP-AVKGEAVKRD		AEGNPM-TA---RE----L-Q--I----	
Synechococcus sp. CC9605	78214144	RSQ-----LLP-AVKGEAVKRD		AEGNPM-NA---RE----L-Q--I----I	
Synechococcus sp. CC9902	78185839	RSQ-----LLP-AEKGRPVRD		AEGNPM-TA---RQ----L-Q--I----	
Synechococcus sp. WH 7803	12007388	QSQ-----DLLP-GEAGKAVKRD		AEGNPM-SA--HRE---H-Q--I----	
Synechococcus sp. WH 7805	88809420	QSQ-----DLLP-GEAGKAVKRD		AEGNPM-SA--HRE---H-Q--I----	
Gloeobacter violaceus PCC 7421	37521140	-----TADARVRGD		RA---TL-----	

Supplementary Figure 88

Partial sequence alignment for the protein nitrate reductase, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505.

		83	122
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899487	FARQFFSPLHNNFQIAA QQ PKPKVVTCLLIQSLIIECFAI
	Raphidiopsis brookii D9	282897458	-----
	'Nostoc azollae' 0708	298491975	--KE--G--Q--T--
	Nostoc punctiforme PCC 73102	186682129	--KE--G--Q--T--
	Nodularia spumigena CCY9414	119508944	--QK--G--E--K--
	Nostoc sp. PCC 7120	17232775	--KE--AG--G--K--
	Anabaena variabilis ATCC 29413	75908747	--KE--AG--G--K--
	Nostoc sp. KNUA003	359308046	--KE--A--D--KA--
	Nostoc sp. PCC 6720	371444172	--KE--A--D--KA--
	Fischerella sp. JSC-11	354567344	--K--AQ--H--V--
Other Cyanobacteria	Trichodesmium erythraeum IMS101	113475918	--KK--E--E--T--
	Microcoleus vaginatus FGP-2	334118953	--K--D--R--T--
	Microcoleus chthonoplastes PCC 7420	254413406	--KEY--D--Q--T--
	Arthrospira maxima CS-328	209525004	-G-E--AK--G--K--
	Arthrospira platensis NIES-39	291570759	-G-E--AQ--G--K--
	Arthrospira platensis Paraca	284049999	-G-E--AQ--G--K--
	Oscillatoria sp. PCC 6506	300864520	--KD--AA--S--E--
	Lyngbya sp. PCC 8106	119485190	Y-H--Q--Q--KD--
	Cyanothece sp. CCY0110	126657419	Y-Q--Q--T-K
	Cyanothece sp. PCC 8801	218245336	Y-QE--S--G--K-K
	Crocospaera watsonii WH 8501	67921181	Y-ER--Q--G--T-K
	Cyanothece sp. ATCC 51142	172035694	Y-Q--AE--G--K-K
	Cyanothece sp. PCC 7425	220905840	--Q--AD--E--K--
	cyanobacterium UCYN-A	284929141	Y-KE--Q--E--K
	Acaryochloris sp. CCMEE 5410	359460551	--R--D--K--D--
	Thermosynechococcus elongatus	22298856	--KA--EK--RA--R-L
	Acaryochloris marina MBIC11017	158337165	--R--D--K--D--
	Synechocystis sp. PCC 6803	16331419	Y-QE--AG--G--H-F
	Microcystis aeruginosa NIES-843	166368050	Y--E--S--Y
	Synechococcus sp. WH 8102	33866270	--KK--E--G--A-L
	Synechococcus sp. RS9917	87123922	--KE--A--G--A-L
	Synechococcus sp. JA-2-3B'a(2-13)	86609877	--T--A--R--K--
	Synechococcus sp. WH 7805	88807702	--KT--A--G--T-L
	Microcystis aeruginosa PCC 7806	159029404	Y-KE--S--Y
	Synechococcus sp. WH 7803	148238990	--KE--G--G--T-L
	Synechococcus sp. WH 5701	87302521	--KE--G--A-L
	Synechococcus sp. WH 8016	352094715	--QE--A--G--S-L
	Synechococcus sp. BL107	116072201	--KE--G--G--V-L
	Synechococcus sp. CC9605	78212276	--KK--E--G--A-L
	Synechococcus elongatus PCC 7942	81300402	--QK--ER--E--KA--
Synechococcus sp. RS9916	116075334	--T--A--G--K-M	
Synechococcus sp. WH 8109	260434321	--KK--E--G--A-F	
Synechococcus sp. CC9311	113955280	--QE--A--G--S-L	
Synechococcus sp. RCC307	148242685	--KE--D--G--RS-L	
Synechococcus sp. CC9902	78185202	--KD--G--G--V-L	
Synechococcus sp. JA-3-3Ab	86605134	--A--A--RD--S--	
Synechococcus elongatus PCC 6301	56750059	--QK--ER--E--KA--	
Synechococcus sp. CB0205	317969604	--KE--A--G--K-L	
Synechococcus sp. PCC 7335	254424350	--KE--AQ--D--T-L	
Synechococcus sp. CB0101	318041059	--KE--A--TT-L	
Cyanobium sp. PCC 7001	254432794	--KE--A--G--R-L	
Prochlorococcus marinus MIT 9303	124022488	--E--A--RD--T-L	
Prochlorococcus marinus AS9601	123968123	--KK--K--G--V-L	
Prochlorococcus marinus CCMP1986	33861089	--KE--K--G--V-L	
Prochlorococcus marinus NATL1A	124025299	--KT--K--LE-F	
Gloeobacter violaceus PCC 7421	37522715	--AY-EQ-DG--Q--	

Supplementary Figure 89

Partial sequence alignment for the hypothetical protein FJSC11DRAFT_2719, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505.

37

100

Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899431	MAYSLCINSLGILAILLLQIGYWLGLGEFSVTD GNFT	GGDLNWAQLAVRHTPLISWLVFVPLAY
	'Nostoc azollae' 0708	298491266	-----SL-F--V--V-V-----I--R--WSSP	I-----R-----MS-----L-----
	Nodularia spumigena CCY9414	119513322	-----SLS-----M--I-I-----I-R---L-SA	--W--VR-V---M--L--L-----
	Anabaena variabilis ATCC 29413	75909869	-----SL-----M--I-V-----A---DLWYSS	-EWG--R-----M-----L-----
	Nostoc sp. PCC 7120	17228502	-----SL-----M--V-V-----A---DLWYSS	-EWG--R-----M-----L-----
	Nostoc punctiforme PCC 73102	186685973	--F-----S-I-VE-----T--EDLWYAT	-E-T-SR-V-Q-L--LA-LV-----
Other Cyanobacteria	Oscillatoria sp. PCC 6506	300868601	-----RLT---VV---MGL-----VFDWSSK	DEFS-LDIL-Q-F-VVACL-----
	Trichodesmium erythraeum IMS101	113475286	-----QLA---VI---MGL--LI--SSLFYPG	EVS-L--IIEYMS-VGVL--I----
	Arthrospira platensis NIES-39	291568852	-----KLR--SV---I--GL--FQ-I-IALSAR	N-SAIDILLQYM-IA-AI--I----
	Microcoleus vaginatus FGP-2	334116560	-----RLT---V-SA--VGL-----AVLDW-SKV	EFS-LEVL--QM-IAASIM-----
	Arthrospira maxima CS-328	209526724	-----KLR--SV---I--GL--FQ-INIALSAR	--SAIDILLQYM-IA-AI--I----
	Arthrospira sp. PCC 8005	376002363	-----KLR--SV---I--GL--FQ-INIALSAR	--SAIDILLQYM-IA-AI--I----
	Arthrospira platensis Paraca	284050935	-----KLR--SV---I--GL--FQ-I-IALSAR	N-SAIDILLQYM-IA-AI--L----
	Moorea producta 3L	332708209	-----RLTR--M--MV-IGL--VR-ISQPEFLA	ITE-S-LR-IIQ-M-VFAGL--I----
	Microcoleus chthonoplastes PCC 7420	254410489	-----RLTG---SA--IT---IQ-Q-QPEFLA	I-EIG-LR-M--FM-VVAG-M-I----
	Lyngbya sp. PCC 8106	119489849	-----RLV-----S---G---S-VSNIF-PA	E-TGLEFLLV-MA-V-AI-----
Cyanothece sp. PCC 7822	307152064	-----RLSQM-----I-VGL--LS--FRS-IWE	GSSILE--ITQM--LTSL--I----	
Acaryochloris sp. CCME 5410	359457115	-----RLI-----WM-IS---LWSNFAPLLGG	TMATD-INTLHG-MG--FG-M-----	

Supplementary Figure 90

Partial sequence alignment for the hypothetical protein Aazo_2379, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505.

		63	112
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899590	LLSALTQKLAENFGIINDHQAS
	Nostoc punctiforme PCC 73102	186681093	--T-IAG--QA---E--GENC
	Nodularia spumigena CCY9414	119510165	----AE--QT--K-A--QDNC
	Nostoc sp. PCC 7120	17228534	--N---E--ST--N-E--TTDNC
	Anabaena variabilis ATCC 29413	75909900	--N---D--ST--KME--TTDNC
	Fischerella sp. JSC-11	354566095	--T-I-E--QT--N-E--TNNS
	Microcystis aeruginosa PCC 7806	159027561	--TVIR----D-R-E-S-K--
	Microcystis aeruginosa NIES-843	166365788	--AVIR----D-RVE-S-K--
	Synechocystis sp. PCC 6803	16331997	-I---E--RRD-D-NLSSD--
	Acaryochloris sp. CCMEE 5410	359460353	-VDSI----QAE-QLNFHRD-I
Other Cyanobacteria	Cyanothece sp. ATCC 51142	172036036	-IEQIVT--Q-D-K-T--NYRS
	Cyanothece sp. CCY0110	126660477	--KQIVT--Q-D-Q-N--NNRS
	Cyanothece sp. PCC 7424	218437441	--REIIAE--SQE---S-TPNN-
	Cyanothece sp. PCC 7822	307154621	--AT-EE--KIE-R-T-HPNR-
	Acaryochloris marina MBIC11017	158333381	-VD-IA--QAE-QLRFHRD-I
	Cyanothece sp. PCC 8801	218246048	-INII---QQD-N-NL-N-K
	Cyanothece sp. PCC 7425	220906998	--RQ-IQT--QQE---GADLE-I
	Cyanothece sp. PCC 8802	257059097	-INII---QQD-N-NL-N-K
	Synechococcus sp. CBO205	317970439	--RAE-RHW-EQQHQ-DCSASEI
	Oscillatoria sp. PCC 6506	300863973	--E-IET--KAD--E--SKNC
	Arthrospira maxima CS-328	209524645	--EIIAD--QTE-H-NLS-R-K
	Arthrospira sp. PCC 8005	376002655	--EIIAD--QTE-H-NLS-R-K
	Microcoleus chthonoplastes PCC 7420	254413374	---TIAA--HTD-R-TL-Q-NC
	Lyngbya sp. PCC 8106	119483394	-QEVIAA--QAE-Q-L-T-KNY
	Arthrospira platensis Paraca	284050147	--EIIAD--QTE-H-NLS-R-K
	Arthrospira platensis NIES-39	291568533	--EIIAD--QTE-H-NLS-R-K
	Moorea producta 3L	332709618	--RE-IEA--QSD-D-E--SANR
	Microcoleus vaginatus FGP-2	334120122	-QD-IAA--KTD-D-E--SKNC
	Trichodesmium erythraeum IMS101	113474185	--RE-IAH--SQE--LNYKAENI
	Prochlorococcus marinus MIT 9303	124022748	--ELIN--MIQ--LDLAESMV
			VVVTAGSNMGFMNAVLAITNPGDEIIL
			I-----I-----
			I-----A---I---S-----
			I-----A---I---S-----
			I-----A---I---S-----
			I-----L--I-----I-----
			IF-----A---I---S-----
			IF-----A---I---S-----
			-----A---L-----EV-----
			M---G--A-----LD---V--
			II-----A---I---A-----
			I-----A---I---A-----
			-----G-----I---S-----I
			-----G--A---I---S--E---I
			M---G--A-----LD---V--
			I---G--A-I--L---QA-----
			-----V-----Q-----
			I---G--A-I--L---QA-----
			LI-----A-QAV-Q-LCD---V--
			I-----I---S-----I
			I-----I---S-----I
			I-----I---S-----I
			I-----GI---S-----I
			I-----GI---S-----I
			I-----A---I---T-----
			I-----TH-I---SAR--V-I
			I--N-GKHSFL-LM--LIE---V-I
			M-----A-HAIAQVLCD---V--

Supplementary Figure 91

Partial sequence alignment for the protein aspartate aminotransferase, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii* CS-505.

		59		116
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298491822	GCGKSTLLRLIAGLEVMTAGNIWVGD	SEADQGY RLINDLLPKERDIAMVFQNYALYPH
	Nostoc punctiforme PCC 73102	186685978	-----G-----	-----P-----
	Anabaena variabilis ATCC 29413	75908271	-----G-----	-V--P-----
	Nostoc sp. PCC 7120	17232273	-----G-----	-V--P-----
	Cylindrospermopsis raciborskii CS-505	282900193	-----G---I-N	D---P--A-----
	Raphidiopsis brookii D9	282897192	-----G---I-N	H---P-A-----
	Nodularia spumigena CCY9414	119509512	-----A-G--T--	---E-P-----
	Nodularia spumigena CCY9414	119510235	-----I-T---SA-S--LFI--	--V-NIPARQ--
	Fischerella sp. JSC-11	354565728	-----G-----	--V--P-----
	Moorea producta 3L	332708939	-----I---EL--K---N	-QV--P-----
	Microcoleus vaginatus FGP-2	334118495	-----A---IL-----	--L--P--D-N-----
	Microcoleus chthonoplastes PCC 7420	254409975	-----V---EL-G-K----	-MV--P--A-N-----
	Arthrospira platensis NIES-39	291570602	-----TP--K--I--	-V--CP--N--V-----
	Arthrospira platensis Paraca	284051442	-----TP--K--I--	--V--CP--N--V-----
	Arthrospira sp. PCC 8005	376004531	-----TP--K--I--	--V--YP--N--V-----
	Arthrospira maxima CS-328	209523646	-----TP--K--I--	-V--YP--N--V-----
	Lyngbya sp. PCC 8106	119489754	-----SL-G---I-E	A-V-G-P-A-----
Cyanothece sp. PCC 7425	220908906	-----TL-G-----	-RV--P-A-----S-----	
Cyanothece sp. PCC 8802	257060154	-----TV-S-Q---N	-EVTH-P-----	
Cyanothece sp. PCC 8801	218247084	-----TV-S-Q---N	-EVTH-P-----	
Other Cyanobacteria	Cyanothece sp. ATCC 51142	172038995	-----M---TA-S-----E	ENVTP-Q--Q-----
	Synechocystis sp. PCC 6803	16330100	-----TV-G---LI--	-RV--P-A-----S-----
	Synechococcus sp. PCC 7335	254423032	-----S---TL-G-E-----	-NVDA-P--Q-----S-----
	Thermosynechococcus elongatus BP-1	22297706	-----DEP-T-R-----	-V-G-P-A-----S-----
	cyanobacterium UCYN-A	284929460	-----TI-S-EVLI-E	KNVGS---Q-----
	Acaryochloris sp. CCME 5410	359459050	-----DL-G-S-RI--	-K--H-P-A-----S-----
	Acaryochloris marina MBIC11017	158338467	-----DL-G-S-RI--	-K--H-P-A-----S-----
	Crocospaera watsonii WH 8501	67921854	-----M---TA-S---F--E	ENVTH-P--Q-----
	Cyanothece sp. CCY0110	126660925	-----M---TA-S-H-L--E	KNVTQ-P--Q-----
	Crocospaera watsonii WH 0003	357264765	-----M---TA-S---F--E	ENVTH-P--Q-----
	Synechococcus sp. PCC 7002	170079148	-----TV-T-Q-AI--	-PV--PS-Q-----
	Rhodothermus marinus DSM 4252	268315938	---T---V---TVSR-T-RI--	QVV--VP-RD-----

Supplementary Figure 92

Partial sequence alignment for the protein glycerol-3-phosphate-transporting ATPase, showing a 7 aa insert that is specific for 'Nostoc azollae' 0708.

		34	100
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298492725	HLLLEEACAANYPLEVVCCTLQWQQAHPLWSEVCSRCD ASGGGSH RAEVSEEILAAIATTVQPDGV
	Nostoc sp. PCC 7120	17227671	-----MD-----PE-EAN-TA-DDA----- -Y-I---V-Q-----
	Anabaena variabilis ATCC 29413	75907658	-----MD-----PE-EAS-TA--E-A-N--- -C-I---V-Q-----
	Nostoc punctiforme PCC 73102	186685767	-----V-----T----SD--A--SS--EKAS-KS- -T-I---V-N-----
	Nodularia spumigena CCY9414	119513673	---Q---VV---VT---PE--T--E--EAA--HS- -V-I--K-----
	Cylindrospermopsis raciborskii CS-505	282898547	--IT---SV-----T----MD--NS-SE--QN--NA-H -T-I--Q-V---M-----
	Raphidiopsis brookii D9	282896219	--IA---SV-----T----ID--NS-SE--QN--NA-H -V-I--Q-V---M-----
	Fischerella sp. JSC-11	354567183	-----V-----A-----Q---IS--M--E-A-QKS- --I-----VM-----
	Oscillatoria sp. PCC 6506	300869138	-----QKH--AT--Y-SE--ER----KK--LQAR -V-L--DRV-E-----A
	Microcoleus vaginatus FGP-2	334119481	-----K-Q-AT---PE-LDQ-GK--EQ-GNLA- -T-L--AKV-ES-----I
	Moorea producta 3L	332710004	---A---VD--VTL-Y-SE-LE-----SQDASV-SQ -V---QSV-K-----E---
	Lyngbya sp. PCC 8106	119493555	--I---L-TQ--LT--Y-A--NRY-R--ES--QAAS -S-E-N-RV-Q-----N---I
	Trichodesmium erythraeum IMS101	113476712	-----KR---IT--H-FE--EK-KF-YEL-SQKAQ -I-L--TK-IQSL---S---
	Arthrospira platensis Paraca	284053061	--V-A--V-WDIMA---P--EER-RD--ELAVK-A- -T-Q--DRLI--MT--N---
	Arthrospira maxima CS-328	209523883	--I-A---V-WEIITL---P---ER--E--ELAAKTA- -T-Q--DRLIV-MT--N---
Arthrospira sp. PCC 8005	376003714	--I-A---V-WEIITL---P---ER--E--ELAVKTA- -T-Q--DRLIV-MT--N---	
Cyanothece sp. PCC 7424	218438771	---T--SV-SS-D---P---TN-K---EKAQ-LAQ -T---TP-V-----IN---	
Cyanothece sp. PCC 7822	307150993	---T--KVKSS-M---PD--NRY---ELASQKAQ ---I-TP-----MT--N---	
Acaryochloris marina MBIC11017	158334661	---QS-WEQ-WSF-I--Y-DR--AKY--I-QQIQD-TT -L-R--P-V-Q-T---H---	
Acaryochloris sp. CCMEE 5410	359462395	---QS-WEQ-WSF-I--Y-DR--AKY---QQIQG-TV -L-Q--P-V-Q-T---H---	
Cyanothece sp. PCC 7425	220910387	---L--VETG-----I-F-P--AKQAPD-LAQLEQLT -L-----QV-----N---	
Cyanothece sp. PCC 8802	257060931	---T--EV-CS-DCL-Y-A--SRY---ELATQKAQ -I-L--P-V-Q-L---N---	
Cyanothece sp. PCC 8801	218247732	---T--EV-CS-DCL-Y-A--SRY---ELATQKAQ -I-L--P-V-Q-L---N---	
Other Cyanobacteria	Synechococcus elongatus PCC 6301	56750360	---Q--I-T-W--LQ--A-PA--NQ-----QQAIQAAP EQT--DDL--YLC--N---
	Synechococcus elongatus PCC 7942	81300008	---Q--I-T-W--LQ--A-PA--NQ-----QQAIQAAP EQT--DDL--YLC--N---
	Crocospaera watsonii WH 8501	67925525	--V-T-YE--CS-MTL-Y-E--RTRY---EN-SKKVQ -L-L--L-V-T-MT--IN---
	Cyanothece sp. ATCC 51142	172035318	--V-T-YKE-CS-ITL-Y-EE--ARY---ES-SK-VK -S-L--ADV--SMT--N---
	Synechococcus sp. PCC 7335	254422310	--I-A--EVR---QTL-Y-EA--VR-KR-CE-AIAQSH QV-L--P-V-KS--S-IT---A
	Cyanothece sp. CCY0110	126656411	--V-T-YQEGCS-MTL-Y-E--RDRY---EN-SK-VK -S-L--I-V-SN-T--N---
	Microcystis aeruginosa NIES-843	166368966	---T-L-L-CS--T--Y-EK--RQAE-AEFQNAQAK -V-I--P-V--SL---N---
	Microcystis aeruginosa PCC 7806	159030692	---T-L-V-CSP-T--Y-EK-R-RQG--AEKLQNAQAK -V-I--P-V--SL---N---
	Synechococcus sp. JA-3-3Ab	86605783	-----LGVGW--RA--Y-FD-AER--R-LASLPKHLR RQA--L-VM--LS--ES-E--
	Synechococcus sp. PCC 7002	170077725	N-I-A-I-KGFSFQT-LF-AA--TKY-A-A-QLNGDRQ TL-----V-QKV---N---I
	cyanobacterium UCYN-A	284929697	--MT--EE-CS-LTLLF-EK-KIENFK--EIASRLTK -Y-T--P-V-KS-T--N---
	Synechococcus sp. JA-2-3B'a(2-13)	86609412	--I--LRVGV--QAL-Y-PD-ASR--H-LADLPKGLR RQP--S-VM--LC--ES-E-A
	Synechocystis sp. PCC 6803	16330947	---V-L-QGKGFNHG-F-AM--EKN-V-ADRLMAQSV HSYL--G-V--KM-S--N----
	Synechococcus sp. CB0101	318040410	-Q-Q-LLRLGLQP-QLLA-PA-IER-GA-LADQDLPLQ P-G--VME-V--DH----
	Gloeobacter violaceus PCC 7421	37520123	-PVAA-L-GGWE--F--A-EA-VGE--AMHTALSG-AG -Y-T-APAVIE-L-D-H-S---
	Prochlorococcus marinus NATL1A	124026423	-----LKT-FLPTE-IA-PE-CDQ-EEILKKIA--TL LTL-NKNV-ETSLS--T----

Supplementary Figure 93

Partial sequence alignment for the protein tRNA/rRNA methyltransferase, showing a 7 aa insert that is specific for 'Nostoc azollae' 0708.

Nostocales/ Stigonematales	'Nostoc azollae' 0708	298490181	QICRASRVGAILEQSKIPLPSAFHKWLTGPGQ WLHPP QAWLNYALYGGEDF	
	Anabaena variabilis ATCC 29413	75908534	----T-G---VI-SDQ----T--NH----E-	A-----
	Nostoc sp. PCC 7120	17232553	----T-G---VI-SDQ----A--DH----E-	A-----
	Nodularia spumigena CCY9414	119511602	---HT-G---V--AK----AS-NH----EK	A-E-----
	Cylindrospermopsis raciborskii CS-505	282902042	-----Q---V-H-D---VPISTNN----LR	--D-----
	Raphidiopsis brookii D9	282897748	-----Q---V-H-D---APISNY----PR	--D-----
	Nostoc punctiforme PCC 73102	186680969	-----G---V--RRQ-SM-A--DH---EE-	A-A-----
	Fischerella sp. JSC-11	354568848	-----N---VV-RTQ--F-P--AH----E-	A-E-S-----
	Cyanothece sp. PCC 7425	220905935	-----G-----DSDRL---A--RHTQDLTP	TQA-DWG-----
	Acaryochloris sp. CCMEE 5410	359458632	-----GM--VIQ-TQ--I--KSLRNAQVLTS	EQAIEW-----
	cyanobacterium UCYN-A	284928967	K--QM-N---EI-----MIAKLS--IPKEE	A-RWV-----
	Synechococcus sp. PCC 7002	170077870	---QR-G---VV-AE---SP-LRRYQD-ET	A-RW-----
	Acaryochloris marina MBIC11017	158333796	-----G---EIL-TQ--I-MS-NNTQVLTS	EQAIEW-----
	Cyanothece sp. PCC 8802	257061042	---M-G---KI-R-Q-S--PSLLALVS-E-	A-QWG-----
	Cyanothece sp. PCC 8801	218247623	---M-G---KI-R-Q-S--PSLLALVS-E-	A-QWG-----
Cyanothece sp. ATCC 51142	172035040	---AM-E---QIQL-QL-CV-MLSQ-FSSA-	A--WV-----	
Synechococcus elongatus PCC 6301	56751552	-----Q---VI-ALPL-ATTS-DRDRLIQA	-----	
Synechocystis sp. PCC 6803	16329394	-----QM--E-WAE-L-I-QGLTD-VGLET	AREWT-----	
Cyanothece sp. CCY0110	126658426	---AM-G---QIQL-CL-CV-IISQ-FSSE-	A--WV-----	
Cyanothece sp. PCC 7424	218437202	---EV-G---Y--YD---IAPQLKLLGSFEE	I-DWV-----	
Thermosynechococcus elongatus BP-1	22299909	---Q-G---V-WADQ--IVDDGDHQQALNW	-----	
Cyanothece sp. PCC 7822	307152161	----F-G---R-ASE---SPDLKLLGSFEE	M-EWV-----	
Microcystis aeruginosa PCC 7806	159027604	----C-G---EI-RIPLHPTLKEYVGAEKAL	EW-----	
Microcystis aeruginosa NIES-843	166368108	----C-G---EI-RIPLHPTLKEYVGAEKAL	EW-----	
Synechococcus sp. JA-3-3Ab	86606546	-V-QS-G-A-RIAAER--ISV-LRQAFPQA	-EW-----	
Microcoleus chthonoplastes PCC 7120	254415394	-----G---VV-RDA--I--VLPQLTS-E-	VWDWV-----	
Trichodesmium erythraeum IMS101	113477461	---Q-QI--RI-RNQ--I--HSLT-FVSDE-	A--W-----	
Microcoleus vaginatus FGP-2	334117254	-----G---KVDRT--I-NI-SEI-STEK	A-DW-----	
Lyngbya sp. PCC 8106	119492217	---Q-Q---RI-RHQ---SK-LQQLVS-SE	AI-W-----	
Oscillatoria sp. PCC 6506	300867160	---Q-N---KI-RNQ-SI-TTISRLVS-EK	A--WG-----	
Moorea producta 3L	332711072	---G-G---EIDPNR-T--PS-S-LVS-E-	A-DW-----	
Arthrospira platensis NIES-39	291566626	--SQ-G---RI-AQ-L-HSNSLE-LAPAET	I-DWV-----	
Arthrospira sp. PCC 8005	376003100	--SQ-G---RI-AQ-L-HSNSLFQLAPAET	V-DWV-----	
Arthrospira maxima CS-328	209525715	--SQ-G---RI-AQ-L-HSNSLFQLAPAET	V-DWV-----	
Other Cyanobacteria				

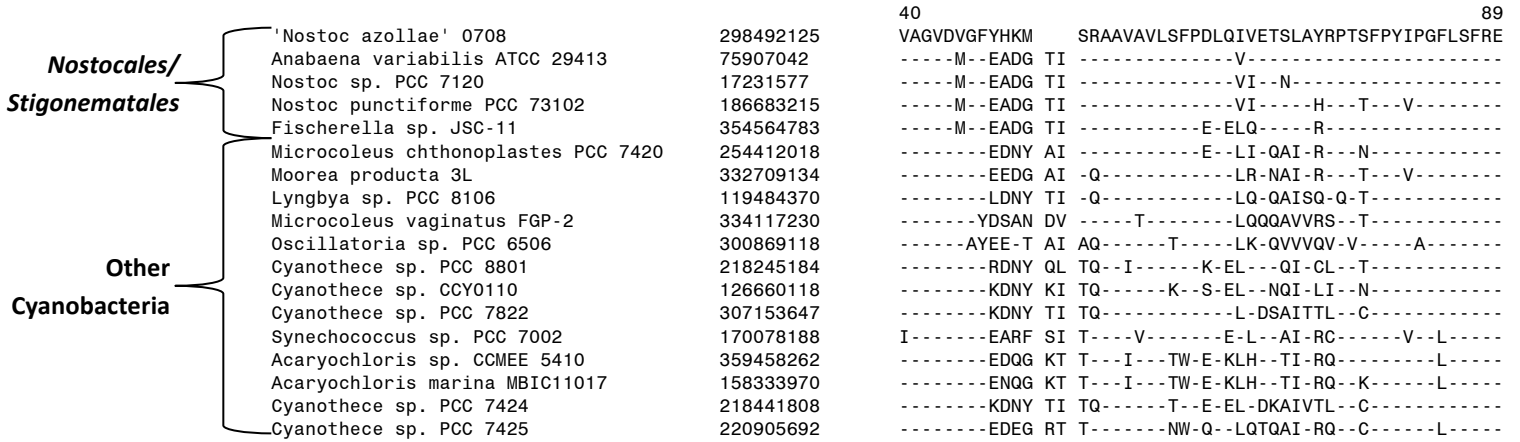
Supplementary Figure 94

Partial sequence alignment for the protein thiamine-monophosphate kinase, showing a 5 aa insert that is specific for '*Nostoc azollae*' 0708.

		202	297
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298493172	YVCSKLGALYVEMGKINEGMELLNQGLSQM IGNQLNQSNNQVHKDKIRLRGFONSQS RKGVNSLSQDIKETNYDILYELHYHLGIA
	Nostoc punctiforme PCC 73102	186680726	-----T---T-----RH-I- VA-E-----
	Nodularia spumigena CCY9414	119511308	-----QT---TQ-LK--QK-FN
	Nostoc sp. PCC 7120	17232349	-----T---DK----KR-IN
	Anabaena variabilis ATCC 29413	75908350	-----T---TDK----KR-IN
	Fischerella sp. JSC-11	354566523	-----T---S--V---AR-VT
	Microcoleus chthonoplastes PCC 7420	254412415	-L-----QI-QVG--IA--KR-FT
	Oscillatoria sp. PCC 6506	300866172	-----I-EVQR-IA--KR--T
	Microcoleus vaginatus FGP-2	334119111	-A-----T-Q-AR-IN--TS--T
	Trichodesmium erythraeum IMS10	113476885	-----L-VDEREKA---FR--Q
Other Cyanobacteria	Lyngbya sp. PCC 8106	119489474	--A-----D--E-AK-LQ--TK--K
	Moorea producta 3L	332709827	-----IQNNQVQ-L---KR--T
	Arthrospira platensis Paraca	284053631	-LA-----QL--LSA-ID--EK--K
	Arthrospira maxima CS-328	209525507	-LA-----QL--LSA-ID--EK--K
	Cyanothece sp. PCC 7424	218438950	-----LQI-QEK--IK--K--K
	Cyanothece sp. PCC 7822	307152415	-----LQI-QEK--IK--K--K
	Cyanothece sp. ATCC 51472	354556402	-T-----AQI--EKQAIK-FK--K
	Cyanothece sp. PCC 8802	257059608	-T-----LQI--EKD-IK--KK--K
	Cyanothece sp. ATCC 51142	172038052	-T-----AQI--EKQAIK-FK--K
	Crocospaera watsonii WH 8501	67924935	-T-----AQI--EKQAVK-FK--K
	Cyanothece sp. PCC 8801	218246568	-T-----LQI--EKD-IK--KK--K
	Acaryochloris marina MBIC11017	158334175	-L-----D--QVEP--H--Q--N
	Crocospaera watsonii WH 0003	357261055	-T-----AQI--EKQAVK-FK--K
	Acaryochloris sp. CMEE 5410	359463169	-L-----D--QVEP-IH--K--N
	Cyanothece sp. CCY0110	126655343	-T-----AQI--EKQAIK-FK--K
	cyanobacterium UCYN-A	284928862	-T-----SQA--EKQAVK-FK--K
	Microcystis aeruginosa NIES-843	166366084	-----G--LKI-RDK--FK--KH--K
	Microcystis aeruginosa PCC 7806	159029464	-----G--LKI-RDK--FK--KH--K
	Synechococcus sp. PCC 7002	170076814	-----GI-LGE-LPKQALKI-K--K
	Cyanothece sp. PCC 7425	220908908	-L-----G--LQ--DVQQ-LQ--Q--A
	Synechocystis sp. PCC 6803	16331831	-L-----FSL-DEK--VK--K--K
	Synechococcus sp. PCC 7335	254425006	LIAN-----QSRWAKALT--H-A-KAS
	Synechococcus sp. RS9916	116073900	-A-A----E-AS-QRER-IN--Q--AA
	Synechococcus elongatus PCC 7942	81299081	-L-----G-L-QE-DLKAAQRW-K--K-G
	Synechococcus elongatus PCC 6301	56751252	-L-----G-L-QE-DLKAAQRW-K--K-G
	Synechococcus sp. RS9917	87125210	-A-A----EIAE-HQGR-LS--R--AS
Synechococcus sp. RCC307	148241378	-A-A----G-E-DQ-NTRT-LA--EM--EH	
Synechococcus sp. CC9311	113952955	-A-A----E-AD-DRTH-IA--RE--AN	
Synechococcus sp. WH 8016	352096139	-A-A----E-AD-DRVQ-IA--RE--AN	
Prochlorococcus marinus MIT 9303	124024305	-T-A----E-DE-QKER-IQ--R--K-	
			QANSLHQGITPV-Q--E----- ACQEE-E----- ACQEE-E----- TA-D-----Y----- DSQVDASV-F-----N- AISIDTYV-----A-- AAAIDDS-V--N----- DS-VDKTV----- QLSLYSKY-EA-ADDSLF----- GT-LEPAV-F-----N- QASGIEGGTM----- QASGIEGGTM----- SNQASPVQ-F-----N- SNKGSQEV-F-----AN- SNKADTHV-----NL SNKADAHV-F-----AN- SNKADTHV-----NL SNKADTHV-----NL SNKADAHV-F-----AN- LDPPEAS-R-----HT SNKADTHV-----NL LNPPEAS-C-----HT SNKADTHV-----NL FNPK-PH--F-----NI CHGS-PH-----Y--ANY CHGS-PH-----Y--ANY YSANESANLR-----N- ADPAESAVR--I-----L- SNTASIPVRF-----AN- AITSRLQPLTE--AY-----V LKDAGERAAER--LL--A-- RPEPAVR--L--ALL RPEPAVR--L--ALL LEADPNANTPTDNTSVSER--LL--A-- CQSDQHRER--LL----- LGEGDE-AAER--LL----- LGEG-EHAAER--LL----- NENTNA-PIER--LL-----

Supplementary Figure 95

Partial sequence alignment for a glycosyl transferase family protein, showing a 27 aa insert that is specific for '*Nostoc azollae*' 0708.



Supplementary Figure 96

Partial sequence alignment for the protein deoxyribonuclease V, showing a 2 aa deletion that is specific for '*Nostoc azollae*' 0708.

		429	479
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298492189	WVLTVVYDGN TDSSEVWLFNSN TLGGGSP RLDEEPVCKIGLPSVIPHSFHG
	Nostoc punctiforme PCC 73102	186681270	-----A-SHT----I-D-D
	Nodularia spumigena CCY9414	119509181	-----SHN----V-E-D
	Anabaena variabilis ATCC 29413	75908387	-----SHG-Q--IYD-E
	Nostoc sp. PCC 7120	17232387	-----SH--Q--IYD-D
	Cylindrospermopsis raciborskii	282898791	-----NH--TICI-E-H
	Raphidiopsis brookii D9	282895779	-----SH--KICI-D-H
	Fischerella sp. JSC-11	354568699	-I-----C-R-Q--V-DAH
	Moorea producta 3L	332710907	-I-----S-T--MV-GR-
	Oscillatoria sp. PCC 6506	300868107	-I--L-----S-T----I-D-A
	Microcoleus vaginatus FGP-2	334118227	-IV--I----S-----I-D-D
	Microcoleus chthonoplastes PCC	254410750	-LI-F-H-H-S-T--LVVL-AQ
	Trichodesmium erythraeum IMS10	113477252	-L--F--EKEKH--LLVIDAQ
	Lyngbya sp. PCC 8106	119486151	-LV-F-H-E-QNQ--LVIVDAQ
	Synechocystis sp. PCC 6803	16331057	-II-----DRHG-----WDAD
	Acaryochloris marina MBIC11017	158334451	-I-----QAASHT----ILAAH
	Acaryochloris sp. CCME5 5410	359457881	-I-----QAASHT----ILAAH
	Other Cyanobacteria	Cyanothece sp. CCY0110	126654815
Crocospaera watsonii WH 0003		357262462	-----NH--LRIYD-E
Crocospaera watsonii WH 8501		67923716	-----NH--LRIYD-E
Cyanothece sp. ATCC 51142		172037428	-----EHR--LRIYE-E
Cyanothece sp. PCC 7424		218442207	---M--N--S-R--IRIYD-D
Microcystis aeruginosa PCC 780		159028973	-LIV-----NH--Q-RIYD-Q
Cyanothece sp. PCC 8801		218244907	-I-----SAHR--RIYD--
Synechococcus sp. PCC 7335		254424445	-L-I--NSLNSR--L-I-AAK
Microcystis aeruginosa NIES-84		166368801	-LIV-----NH--Q-RIYD-Q
Cyanothece sp. PCC 7822		307592164	-I--E-F--EQVK----I-DAL
Cyanothece sp. PCC 7425		220909329	-I-----HHQ-T-QIYPAD
Synechococcus sp. PCC 7002		170078798	-L---FNAPANR--LRIYGGD
Synechococcus sp. JA-3-3Ab		86605763	-IV---F-AASQR-Q-VILDAS
Thermosynechococcus elongatus		22297559	-L-Q---ASCHK-QLLI-DAA
Synechococcus elongatus PCC 63		56751324	---SLI--AARH--ALVILSAQ
Synechococcus sp. CC9311		113954610	---DL-WN-AR-A-DLYILDAR
Gloeobacter violaceus PCC 7421		37522343	---S-I--AGLHR--L-I-DAD
Oryza sativa Japonica Group		108863037	YLIFF-H-E--GK--NVIDAK
Coccomyxa subellipsoidea C-169		384247857	YLA-Y---EQ-----FMVYDAC
Burkholderia gladioli BSR3		330820166	---AC--RAA--T-D-VVLDAR

Supplementary Figure 97

Partial sequence alignment for the protein carotenoid oxygenase, showing a 7 aa insert that is specific for '*Nostoc azollae*' 0708.

		542	590	
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298491727	LQDFFIQRIRTLFQEA E VQIFID EIDYDLVNAVLGENDPEYTERALQDL	
	Nostoc sp. PCC 7120	17231603	-NE--L-----L--EK Q-----L-----	
	Nostoc punctiforme PCC 73102	186684136	--E--L-----L--EK Q-----T-K--	
	Cylindrospermopsis raciborskii CS-505	282899452	--E-----L--EK H-----Q----I-----	
	Nodularia spumigena CCY9414	119509794	--E--L-----L-DEK Q-----R-----	
	Anabaena variabilis ATCC 29413	75907017	-NE--L-----L--EK Q-----D-----	
	Raphidiopsis brookii D9	282897614	-RE-----L--EK H-----Q----I-----	
	Fischerella sp. JSC-11	354566384	--E--L-----L--EK Q-----Q-----	
	Oscillatoria sp. PCC 6506	300868195	-S--L-----L--EK G-----	
	Microcoleus vaginatus FGP-2	334120052	-YE--L-----L--EK NV-----R--	
	Moorea producta 3L	332709830	--E--L-----L--EN K-----D----QN---I-V	
	Microcoleus chthonoplastes PCC 7420	254414262	--E-----L-DEQ K-----R---Q---K-V	
	Lyngbya sp. PCC 8106	119487891	-YE--L-----L-DEK N-----D---K--	
	Arthrospira platensis Paraca	284051092	-E--L-----L-DEK S-----DE-I--K-----	
	Arthrospira sp. PCC 8005	376003506	-E--L-----L-DEK A-----DE-I--K-----	
	Arthrospira maxima CS-328	209524276	-E--L-----L-DEK A-----DE-I--K-----	
	Other Cyanobacteria	Trichodesmium erythraeum IMS101	113477880	-SN--L---K--ILL-EK N-----D--A--K--G--K--
		Cyanothece sp. PCC 7424	218441712	-KS----LS--L-DDL H-----R--
Cyanothece sp. PCC 8801		218247137	-KG-----Q--L--EL KL-----D----Q----E--	
Cyanothece sp. PCC 7822		307153543	-KS--L--LS--L-DDL K-----AQ---E--	
Cyanothece sp. CCY0110		126657519	--T-L---Q--L-DEL N-----A-A---D---N-	
Cyanothece sp. ATCC 51142		172037905	--T-----Q--L-DEF N-----A-AQ--D---N-	
Synechococcus sp. PCC 7002		170079051	--S-----Q--L-DEK G-----D-A-----T--	
Microcystis aeruginosa NIES-843		166367631	IK---L--LQ--LIDDL R-----KS---EN-	
Cyanothece sp. PCC 7425		220909945	-E--L---NTLD-D- Q----I--I---A----Q---ASI	
Crocospaera watsonii WH 0003		357263801	--T-----QNIL-DDF N-----A-A--M----N-	
Crocospaera watsonii WH 8501		67922573	--T-----QNIL-DDF N-----A-A--M----N-	
Microcystis aeruginosa PCC 943		389677864	IK---L--LQ--LSDDL R-----KS---EN-	
Microcystis sp. T1-4		390439523	IK---L--LQ--LIDDL R-----L--KS---EN-	
Synechococcus elongatus PCC 7942		81298829	-T---G--L---L-DEQ G-----DG-AN-LL--IA-P	
Synechococcus elongatus PCC 6301		56751488	-T---G--L---L-DEQ G-----DG-AN-LL--IA-P	
cyanobacterium UCYN-A		284928750	-RN--A---QS-L-DEM G-----DVES--I---F--I	
Synechococcus sp. PCC 7335		254423532	-I---D-----L-DNL -----IF-Q -KH-QL-S-T--	
Acaryochloris marina MBIC11017		158333790	-R---G--L-S-L-DEQ T-----KG -H-Q-V-V-V	
Acaryochloris sp. CCME 5410		359458626	-R---G--L-S-L-DEQ T-----KG -H-Q-V-V-V	
Thermosynechococcus elongatus BP-1		22299735	-RQ--S--L---L--DL G-E-----IP-DQ--LQI--R-V	
Synechococcus sp. JA-2-3B'a(2-13)		86609951	--E-LA---Q--L--EK R-----R--V-DPESVK--DPQKAE	
Gloeobacter violaceus PCC 7421		37521215	-C-W-A--LYG-LV-RD GL-----L-----AV-TG--R--AN-	

Supplementary Figure 98

Partial sequence alignment for the protein glycyl-tRNA synthetase beta subunit, showing a 6 aa insert that is specific for '*Nostoc azollae*' 0708.

33

79

Nostocales/ Stigonematales	'Nostoc azollae' 0708	298492461	TPQPATPEAVLFAMVDAIA	TLTKG	NAQIDPDNQITIAIGVGTGPADA
	Nostoc sp. PCC 7120	17232494	-----FLV-----		-----E-----
	Anabaena variabilis ATCC 29413	75908482	-----FLV-----		-----E-----
	Nodularia spumigena CCY9414	119510207	-----S-----AV-----		-V--Y--AV-----K
	Nostoc punctiforme PCC 73102	186681167	A---T-----AVL-----		---N-E-F-----S--
	Cylindrospermopsis raciborskii CS-505	282899910	--K---Q---ST-LEG-N		SL-EQNGTIL-L-----V--
	Fischerella sp. JSC-11	354568075	-----IAT-----		-L-LSEAV-----
	Microcoleus chthonoplastes PCC 7420	254409429	-----S--K--VEV-A----		-L-TNHSV-----
	Oscillatoria sp. PCC 6506	300868925	-----A--A---T		ELNPTA-SVK-----I-----
	Microcoleus vaginatus FGP-2	334121289	-----AT-T---L		QLNPVENAVSAIG-I-----
Other Cyanobacteria	Arthrospira platensis Paraca	284049937	-----L-A-ILN--S		QITA-VESWA-----L---V-G
	Moorea producta 3L	332709202	-----T--IE-I-A--S		RL-TREDV-G--L-C----V
	Arthrospira maxima CS-328	209528190	-----A-ILN--S		QITA-VESWA-----L---V-G
	Arthrospira platensis NIES-39	291570066	-----L-A-ILN--S		QITA-VESWA-----L---V-G
	Lyngbya sp. PCC 8106	119486099	---A-----FSVLLE--L		QITKNVKSIGG--I-L---V--
	Crocospaera watsonii WH 0003	357266148	-----VKS--E---		-LT-TGNCL-L-----K
	Cyanothece sp. PCC 7425	220907956	-----TA-I-A-Q		TL--EQ--V-----R
	Crocospaera watsonii WH 8501	67920530	-----VKS--E---		-LT-TGNCL-L-----K
	Cyanothece sp. PCC 8801	218248396	-----K--KSI---VK		E-NSNQDCL-----E
	Cyanothece sp. PCC 7424	218442238	-----Q---ESI-TVVK		-LNQ-HNCL--L-----
	Cyanothece sp. CCY0110	126655036	---S-----VKTIIE---		-LTQEONCL-L-I-----K
	Cyanothece sp. PCC 7822	307155366	-----Q---ESI-T-VK		-LNR--NCL--L-----
	Microcystis aeruginosa NIES-843	166364586	-----KE-ALTIHQG-C		-VNL-KSCQ-----
	Cyanothece sp. PCC 51142	172037435	---S-----VET-IQ---		-LT-NQNC--L-----K
	Microcystis aeruginosa PCC 7806	159027064	-----KE-ALTIHQG-C		-VNL-KSCQ-----V--T--
	Synechococcus sp. PCC 7002	170079033	---S--T--AE-IANLVK		ELNS-HSCK-V-I-I---V-G
	Synechococcus sp. CC9311	113955336	-----G--CI-L-E--E		AL---RRALIV-I-L---M-V
	Synechococcus sp. WH 5701	87303291	--R--L-G--CT-L-E--T		AV--GR-AGRV-I-L---M-R
	Synechocystis sp. PCC 6803	16331233	-----L-K--YAQLAHG-D		LLKNSACV-----M-----
	Synechococcus elongatus PCC 7942	81300920	--T-S--SN-RA-L-E--R		-L--QAEL---I-----
	Acaryochloris sp. CCMEE 5410	359462651	---S---P-VD-IAT-VH		TL---KTLPV--I-M----N
	Synechococcus sp. WH 8016	352094573	-----G--CV-L-E--E		AL---RRASIV-I-L---M-V
	Acaryochloris marina MBIC11017	158336055	---S---P-VD-IST-VH		AL---KTLPV--I-M----N
	Synechococcus elongatus PCC 6301	56751990	--T-S--SN-RA-L-E--R		-L--QAEL---I-----
	Synechococcus sp. CC9902	78184501	---P-V-G--TM-LCE--E		-L---HLAALV---L---M--
	Thermosynechococcus elongatus BP-1	22299898	---P-F--R-IK-ILQG-Q		DL-GED-AL--M-I---V--
	Synechococcus sp. CB0101	318041203	---T-M-G--VI-I-E-VQ		-L--ERLASCV-L-L---M--
	Synechococcus sp. BL107	116070833	---P-V-G--TM-LCE--D		-L---HLADLV---L---M--
	cyanobacterium UCYN-A	284928791	--R-----VT-ITQIVK		VLKK-CDS--L---I---V-C
	Synechococcus sp. WH 8109	260435487	---V-G--TM-LCE--E		---EHAEE-V-I-L---M--
	Synechococcus sp. CC9605	78212559	---V-G--TM-LCE--E		---GHAAG-V-I-L---M--
	Synechococcus sp. RCC307	148242024	---P-A-G-ITT-L-E--E		AL--ER-AD-V---L---M-R
	Synechococcus sp. CB0205	317969683	---M-GG-VT-I--V-		AL---RRASRV-I-L---M--
	Synechococcus sp. WH 8102	33866021	---V-G--TM-LC--VE		EL---GAAALV---L---M--
	Synechococcus sp. RS9916	116075438	---M-G--CM-ICE--E		RL---AREAADV-I-L---M--
	Synechococcus sp. RS9917	87124038	---V-G--CM-LCE--E		AL---RCAAIV-I-L---M--
	Synechococcus sp. WH 7805	88809581	-----GS-CMEL-E--D		TL--ERKAALV-I-L---M--
	Synechococcus sp. WH 7803	148239099	-----GS-CMEL-E--D		TL--QRTAERV-I-L---M--
	Cyanobium sp. PCC 7001	254432865	---V-G--TM-IAE-VE		A---ERRADRIV-I-H---S-R
	Prochlorococcus marinus MIT 9303	124023544	---S-V-G--TV-LCE--E		TV--NHAAAFV---L---M--
Gloeobacter violaceus PCC 7421	37523199	--R-S-----A-LSALV-		VL--GREAV-L---V--LV-V	

Supplementary Figure 99

Partial sequence alignment for the protein glucokinase, showing a 5 aa insert that is specific for 'Nostoc azollae' 0708.

		335		385
	'Nostoc azollae' 0708	298490634	GAKVSVGPNFYHVFAAATPD SL	HGLPMQELSLETAENVILIDDLQALHAGV
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii CS-505	282899189	----TIS-----N-TE	K-F--Q-----
	Raphidiopsis brookii D9	282895951	----T-S-----N-TE	E-F--Q-V-----
	Nostoc punctiforme PCC 73102	186685486	-----S-----TT--E	E----D-----F-----
	Nodularia spumigena CCY9414	119509954	----IS-----TS-E	-D--D-----
	Nostoc flagelliforme Sunitezuoqi	308205941	-----S-----ST-N-E	E-----F-----
	Nostoc sp. PCC 7120	17232710	-----S-----SSGDSE	-K-N-----H-----
	Anabaena variabilis ATCC 29413	75908976	-----S-----SGNSE	EKF----D-----H-----
	Fischerella sp. JSC-11	354568204	----T-S-----GNSPA	--F--N-----
	Acaryochloris sp. CCME 5410	359463925	----T--H-----S--E	-TY-----F--V-----
	Acaryochloris marina MBIC11017	158335615	----T--H-----S--E	-TY-----V--V-----
	Microcystis aeruginosa PCC 7806	159027557	----T-S-H-----S-NQP	KSY-----V--S-----
	Cyanothece sp. PCC 7822	307150726	----IS--Y--L-GE-NG	ES--DK--V--S-----
	Microcystis aeruginosa NIES-843	166366110	-S--T-S-H-----S-NQP	KSY-----V--S-----
	Cyanothece sp. PCC 7424	218438331	----T-S--Y--V-GEE-SG	ESY--V--VY--N-----
	Other Cyanobacteria	Synechococcus sp. JA-3-3Ab	86606884	----T-AAHY----G-P-E
Cyanothece sp. ATCC 51142		172035569	----T--SH-----SNSETN	P-YN-DE----V--K-----
Synechococcus sp. JA-2-3B'a(2-13)		86609847	----T-AAHY----G-P-E	STYD-A----VW--E-----
Cyanothece sp. PCC 8801		218246947	----T-SSH-----GSEETP	-A-T-D-----S-----
Cyanothece sp. PCC 8802		257060012	----T-SSH-----SEENP	-A-T-D-----S-----
Synechococcus elongatus PCC 7942		81300098	----T-S-H-GV-E--EARS	--D--S-----
Synechocystis sp. PCC 6803		16331645	-S-IT--AH---YSEK-AE	KSY--DQ---VY--K-----
Cyanothece sp. PCC 7425		220909982	----IT-S-H----LPGERV-	ANFD-A-----FV-----
Synechococcus elongatus PCC 6301		56750273	----T-S-H-GV-E--EARS	--D--S-----
Cyanothece sp. CCY0110		126661345	----T-SSH---I-SNSQAE	PQYN-DE-----K-----
Crocospaera watsonii WH 0003		357263212	----T--SH-----SNSEAN	PQYN-DE-----F--Q-----
Crocospaera watsonii WH 8501		67923060	----T--SH-----SNSEAN	PQYN-DE-----F--Q-----
Synechococcus sp. PCC 7335		254423121	----I-L-YIE-LPGE-A-	DNYQ-DK-DG-V-----
cyanobacterium UCYN-A		284929189	-S--NIR-HY--ISNQAN	F-CN-NMVK--VF--K-----
Thermosynechococcus elongatus BP-1		22298624	----T-SGH--D-P-GE	RPDGLVW--E-H-----
Synechococcus sp. WH 8102		33865730	-----DWFV-GTTPGQT	SGQELDHTRE-DTFV--E--S-----
Synechococcus sp. BL107		116070615	-----DWFVLTSEGCS	SGNNLDQSTEKDTFV--E--S-----
Synechococcus sp. WH 7803		148239436	-----DWFV-HRSEGAT	LPADHLDRHT-SDSFV--ES-N-----
Synechococcus sp. WH 7805		88808727	-----DWFV-HRSEGCR	PPADHLDRHT-QDSFVM-ES-N-----
Synechococcus sp. CC9902		78184740	-----DWFVLTSEGCS	SGNTLDQSTEKDTFV--E--S-----
Synechococcus sp. CC9605		78212838	-----DWFV-GSNPQVS	SGNSLDHRTESGPFV--E--S-----
Synechococcus sp. RS9916		116074908	-----DWFV-GTSEGTS	SGTTHLHHAAT-T-PFV--ES-N-----
Synechococcus sp. WH 8109		260434581	-----DWFV-GSNPQVS	SGNSLNHRTEGPFV--E--S-----
Synechococcus sp. CC9311		113955164	-----DWFV-SSSEGSS	RPVEHLDRHN-RDTFV--EG-N-----
Synechococcus sp. RCC307		148242573	-----S-DWFEIG-TPGAG	CGDAS-NRFQ-DG-VW--S-S-----
Synechococcus sp. CB0101		318041667	-----DWFVIGTTPGS-	-GQAGLNRFKAS-PIVW--S-S-----
Synechococcus sp. WH 5701		87302397	-----HWFEIGP-GSS-	EQTGARG-DRFSQDGIWV--S-S-----
Synechococcus sp. RS9917		87125803	-----DWFV-GTSDGMS	SGNSHLHARSTDTFV--EG-N-----
Cyanobium sp. PCC 7001		254430985	-----DWFVIGPS-GS-	-GQQGLDRFN-D-PLVW--S-S-----
Synechococcus sp. WH 8016		352093854	-----DWFV-SSSQGLI	SPSPHLDRS-Q-PFV--ES-S-----
Synechococcus sp. CB0205		317970133	-----DWFVIGPTPGTS	-GQGLDRFKADQPI-W--S-S-----
Moorea producta 3L		332710225	----T-S-----PGQSAE	--Y--DQ----W--EVN-----
Arthrospira platensis Paraca		284052628	-----S---FC-APGSAEN	GGSYNID--D--W--E-H--S--
Arthrospira platensis NIES-39		291571935	-----S---FC-APGSAEN	GGSYNID--D--W--E-H--S--
Arthrospira maxima CS-328		209523996	-----S---FC-APGTGEN	SESYNIN--D--W--E-H--S--
Arthrospira sp. PCC 8005	376003855	-----S---FC-APGTGEN	SESYNIN--D--W--E-H--S--	
Prochlorococcus marinus NATL2A	72382205	-----S-DWL-V-SKSESEM	DKDESLSIKN-IKEY----E-S-I-S-	
Prochlorococcus marinus 9313	33863043	-----DWFV-SKTEGQN	SANTHLQHSRSK-TFV--EG-N-----	
Prochlorococcus marinus CCMP1375	33240446	-----S-DWPVIYKTKGET	AKYPDLQFQH-K-EFV-VEN-H-----I	

Supplementary Figure 100

Partial sequence alignment for the protein DNA gyrase modulator peptidase U62, showing a 2 aa insert that is specific for '*Nostoc azollae*' 0708.

		237	283
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119509958	VERILRLFPPEMILEVRD SASADLSV RGLLYYQLGEYHQAAADDLQSYL
	Nostoc punctiforme PCC 73102	186684449	I----L---NLT--L-- ---IS----N-P--VN---H--
	Anabaena variabilis ATCC 29413	75908837	-----L----ATS-L-- -----Y-P--TN--ET--
	Nostoc sp. PCC 7120	17227520	-----L----TTS-L-- -----H-P--TN--EL--
	Fischerella sp. JSC-11	354564814	-----L---GFA----- ---C----RFS---N--ET--
	Oscillatoria sp. PCC 6506	300865543	I----L---DAAG-R-- --I----F-RWQESRQ--EM--
	Trichodesmium erythraeum IMS101	113474370	ID--L---NAVM-K-- --V----DIWAK-RE--E--I
	Moorea producta 3L	332705524	---M-L---NAPVQL-- ---I-S---YLSE-CQ--ER--
	Lyngbya sp. PCC 8106	119483290	I----LI---APN-Q-- --I----ERWTE-RQ--EN--
	Microcoleus vaginatus FGP-2	334119364	I----L---DALG-R-- --I----RFVE-RH--EL--
	Arthrospira sp. PCC 8005	376007000	---VI-S-DD-K-K-- ---I----YRWSE-RC--ED--
	Arthrospira maxima CS-328	209525446	---VI-S-DD-K-K-- ---I----YRWSE-RC--ED--
	Arthrospira platensis Paraca	284055021	---VI-S-DD-K-R-- ---V----YRWSE-RC--ED--
	Cyanothece sp. PCC 7424	218439838	I-Q--LI--NHP--L-- K--I----QWEK-SQ--KI--
	Other Cyanobacteria	Crocospaera watsonii WH 0003	357260958
Crocospaera watsonii WH 8501		67925007	I-LL-Q---NHP--I-- -----RQWEK-TT---L--
Cyanothece sp. ATCC 51142		172036086	I-LL-Q---NHPI-L-- -----RQWEK-TA---L--
Cyanothece sp. PCC 7822		307152968	L-Y--LI--NHP--L-- ---I--H---TQK-SQ--GF--
Cyanothece sp. CCY0110		126660076	I-LL-Q---NHPI-I-- -----RQWEK-TA---L--
Cyanothece sp. PCC 8801		218247269	I-LL-II--NHP--L-- -----H--QWQ--TK--NL--
Cyanothece sp. PCC 8802		257061395	I-LL-II--NHP--L-- -----H--QWQ--TK--NL--
Microcystis aeruginosa NIES-843		166368563	I-LL-L-I-NHP--L-- ---V---I-QFS--QQ--GF--
Microcystis aeruginosa PCC 7806		159028679	IDLL-L-I-NHP--L-- ---V---I-QLS--QQ--GF--
Acaryochloris marina MBIC11017		158337334	S-Q--LVD-KCYS-L-- --I---HV-RWSE-RQ---DF--
Acaryochloris sp. CCMEE 5410		359463572	S-Q--LVD-KCYS-L-- --I---HV-RWSE-RQ---DF--
Synechococcus sp. PCC 7335		254421427	LDKL-L-R-LDT-QL-- ---HH--QNDRL-KQ--E---
Gloeobacter violaceus PCC 7421		37522062	---L-VI--NTPA-W-- --I-H-R--HPTA-RA--EN--

Supplementary Figure 101

Partial sequence alignment for the hypothetical protein N9414_05744, showing an 8 aa insert that is specific for *Nodularia spumigena* CCY9414.

		36	68
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119510629	HGIWHTREGIIVCL GGCYGEIAPISWFGSETL
	Nostoc punctiforme PCC 73102	186686474	--K-DI-----LR- TDET LRV-W-----
	Nostoc sp. PCC 7120	17227530	----EV--A--LR- TDTT -KV-W-----
	Anabaena variabilis ATCC 29413	75908848	----EV--A--LR- TDTT -KV-W-----
	'Nostoc azollae' 0708	298490585	--V-EI--SV-IR- VDIT -KVSW--S--H-----
	Raphidiopsis brookii D9	282898201	--V-DM-----IR- VDRT -KV-W--S--A-----
	Cylindrospermopsis raciborskii CS-505	282901080	--V-DM---VMIR- VDRT -KV-W--S--A-----
	Lyngbya sp. PCC 8106	119487537	----EK-----IH- TNEQ EKS-W-----L-----
	Oscillatoria sp. PCC 6506	300864052	--D-NI-----IR- TGET -KFSL-----N-L---SF
	Arthrospira platensis Paraca	284052034	--L-EV-----IE- MDEK -KV-W-----L-----
	Arthrospira sp. PCC 8005	376006280	--L-EV-----IE- MDER -KV-W-----L-----
	Arthrospira maxima CS-328	209525648	--L-EV-----IE- MDER -KV-W-----L-----V
	Trichodesmium erythraeum IMS101	113477491	----RI-----LK- TNKT -EI-L-----L-F-----
	Microcoleus vaginatus FGP-2	334119558	----DI-S---LRV VGEN -RI-W-----L-----SF
	Microcoleus chthonoplastes PCC 7420	254415285	----KV-----LR- VDEK -RV-W-----LP-----
	Moorea producta 3L	332709413	--K-TV-----LS- SDEI -NI-W-----P-----
	Other Cyanobacteria	Fischerella sp. JSC-11	354568380
Cyanothece sp. PCC 7822		307592178	--L-KV-K---R- TDTK -RK-L-----LE-----SF
Cyanothece sp. PCC 7424		218437913	--L-RI-----IS- KDDR -KI-W-----P-----
Cyanothece sp. PCC 8801		218246059	--V-K-----IR- TDTT -KI-W-----LP-----S-
Cyanothece sp. PCC 8802		257059108	--V-K-----IR- TDTT -KI-W-----LP-----S-
Cyanothece sp. PCC 7425		220907643	--V-SV-QS-L-Q- RDGD -RS-W-----L-----
Synechococcus sp. PCC 7002		170079236	--L-TV-----LS- RDER -RITQ--V-LP---T---
Acaryochloris sp. CCME 5410		359461059	Y-Q-SI-R-LV-T- EDDA -NI-Q-----L-----W
Acaryochloris marina MBIC11017		158336885	Y-P-SI-R-LM-M- EDDA -NI-Q-----L-----W
Synechococcus sp. PCC 7335		254421298	Y---KH-----LK- SDHQ -RI-F--V--LP---T--Q
Microcystis aeruginosa NIES-843		166368583	--L-QI-----IR- EAED -RI-W-----LPE-----
Microcystis aeruginosa PCC 7806		159027129	--L-EI-----IR- EAED -RI-W-----LPE-----
Thermosynechococcus elongatus BP-1		22298717	Q-V-RS-S--YLR- EDEQ -QV-----LPGW-----
Synechocystis sp. PCC 6803		16331451	--H-QR-R-L--E- TDQR -NR-R-----LP---T-D-

Supplementary Figure 102

Partial sequence alignment for the protein O-succinylbenzoate synthase, showing a 4 aa deletion that is specific for *Nodularia spumigena* CCY9414.

			109		142
	Nodularia spumigena CCY9414	119510127	SMVMFGGVSINPQKRLK E	GRVDILVSTPGRLLD	
	Nostoc sp. PCC 7120	17232210	-----QR--	-----A-----	
Nostocales/	Anabaena variabilis ATCC 29413	75908173	-----QR--	-----A-----	
Stigonematales	Anabaena variabilis	4104820	-----QR--	-----A-----	
	'Nostoc azollae' 0708	298490698	--A-----	SG---V-----	
	Nostoc punctiforme PCC 73102	186683738	-----G--L--QR--	N-----A-----	
Other Bacteria	Synechococcus sp. PCC 7335	254422154	-T-I---K---ISK-R	QG---IA-----	
	Janthinobacterium sp. Marseill	152981148	-A-I---G---IKM--	HG---A-----	
	Nematostella vectensis	156312380	-Q-V---G---MMN-R	KGA-V--A-----	

Supplementary Figure 103

Partial sequence alignment for the protein ATP-dependent RNA helicase, showing a 1 aa insert that is specific for *Nodularia spumigena* CCY9414.

			209		258
<i>Nostocales/</i>	Nodularia spumigena CCY9414	119511267	AHGVSVEISK FGWSGLFSID	RSQKSWSDRNSRYGRRITGNTEMRVSGP	
<i>Stigonematales</i>	Anabaena variabilis ATCC 29413	75908286	-----T-	N-ND-TYN---P----V-A--Q-----	
	Nostoc sp. PCC 7120	17232286	-----Q-T-	N-ND-TFN---EF-----A--P--I---	
	Fischerella sp. JSC-11	354565101	-----	K-SD-TYN---S-----A-----	
Other Bacteria	Oscillochloris trichoides DG6	309791015	-----RR	AYDG--YV-D-Q-N-----E-PI-I---	
	Nitrosomonas sp. AL212	325981098	---INI-T-D-	EHHQ-G-N-H--LN-----LCKI---	

Supplementary Figure 104

Partial sequence alignment for a twin-arginine translocation pathway signal protein, showing a 10 aa insert that is specific for *Nodularia spumigena* CCY9414.

		213	267
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119511526	MGRFNHEAVAVDPKTYGYVYETEDRGDSCFYRFVFN VRPT KVGDLKGGSLYALKV
	Nostoc sp. PCC 7120	17227628	-----A--I--Q-----L----L--
	Nostoc punctiforme PCC 73102	186682595	-----A---N---I-M---Q-NGL----I--
	Anabaena variabilis ATCC 29413	75907723	-----I---A--I--Q---S---L----I--
Other Cyanobacteria	Cyanothece sp. PCC 8802	257061151	----R---I-----I-Q---QV-----R--
	Cyanothece sp. PCC 8801	218247513	----R---I-----I-Q---QV-----R--
	Cyanothece sp. CCY0110	126656794	----R---I-----I-Q---QN-----R-K
	Cyanothece sp. ATCC 51142	172038491	----R---I---N---I-Q---QN-----R-H
	Crocospaera watsonii WH 0003	357260160	----R---I-----I-Q---QN-----R-H
	Thermosynechococcus elongatus BP-1	22297708	----Q---I-I---NII-Q-----L----I-K
	Acaryochloris marina MBIC11017	158336680	-----DSNI--Q-----GL----R-Q
	Acaryochloris sp. CCMEE 5410	359459939	-----DSNI--Q-----GL----R-Q
	Cyanothece sp. PCC 7425	220910401	----Y---I---RS-II-L---LT-GR---I-A
	Gloeobacter violaceus PCC 7421	37522015	-----I-----N-EGL----L--
Moorea producta 3L	332710206	-----N--I--Q---M-GL----I-K	
Microcoleus vaginatus FGP-2	334116687	-----I--R--I-----QR--LL---I-F	

Supplementary Figure 105

Partial sequence alignment for the hypothetical protein N9414_05744, showing a 4 aa insert that is specific for *Nodularia spumigena* CCY9414.

		64	98
Nostocales/ Stigonematales	<i>Nodularia spumigena</i> CCY9414	119509033	PFMSRLGVTESWGGWSV V TGNAAADPGFWSFEGVA
	<i>Nostoc punctiforme</i> PCC 73102	186682779	-----Q-----N- --GP-T-----
	<i>Nostoc</i> sp. PCC 7120	17227634	---A---Q----- --GT-T-----
	<i>Raphidiopsis brookii</i> D9	282897575	---A----- --GT-V-----
	<i>Cylindrospermopsis raciborskii</i> CS-505	282900928	---A----- --GT-V-----
	' <i>Nostoc azollae</i> ' 0708	298492503	-----IK----- --GT-V-----
	<i>Fischerella</i> sp. JSC-11	354568684	---A---VG---N- --A-NY-----
	<i>Mastigocladus laminosus</i>	5759154	---A---VG---N- I-ASNY-----
	<i>Synechococcus elongatus</i> PCC 6301	56750842	---A---Q-----I --ET-V---Y-----
	<i>Microcystis aeruginosa</i> NIES-843	166366040	---A---G----- --ETGV-----
	<i>Synechocystis</i> sp. PCC 6803	16331660	---A---S--N--- --ETGL-----
	<i>Crocospaera watsonii</i> WH 8501	67926056	---A---G----- --ETGVN-----
	<i>Cyanothece</i> sp. PCC 7424	218438550	---A---G----- --ETGVN-----
	<i>Cyanothece</i> sp. PCC 8801	218247835	---A---G----- --ETGVN-----
	<i>Cyanothece</i> sp. ATCC 51142	172036752	---A---G----- --ETGVN-----
	<i>Cyanothece</i> sp. PCC 7822	307151151	---A---G----- --ETGVN-----
	<i>Cyanothece</i> sp. CCY0110	126659760	---A---G----- --ETGVN-----
	<i>Cyanobium</i> sp. PCC 7001	254430999	---A---G-----I --ETGV-----
	<i>Cyanothece</i> sp. PCC 7425	220906297	---A---G-----I --ETGVN-----
	<i>Synechococcus</i> sp. CB0101	318042632	---A---G-----I --ETGV-----
	<i>Thermosynechococcus elongatus</i> BP-1	22299073	---A---G--S---I --ETGI-----
	<i>Synechococcus</i> sp. RCC307	148241517	---A---D-----I --ATGVE-----
	<i>Synechococcus</i> sp. WH 8102	33866514	-----G-----I --ETGV-----
	<i>Synechococcus</i> sp. BL107	116072657	-----G-----I --ETGV-----
	<i>Synechococcus</i> sp. WH 8016	352096610	-----G-----I --ETGV-----
	<i>Synechococcus</i> sp. CC9311	113953324	-----G-----I --ETGV-----
	<i>Synechococcus</i> sp. CB0205	317968715	---A---D-----I --ATGV-----
<i>Synechococcus</i> sp. WH 7803	148238854	---A---D-----I --ATGV-----	
<i>Synechococcus</i> sp. WH 7805	88807801	---A---D-----I --ATGV-----	
<i>Synechococcus</i> sp. PCC 7002	170078366	---A---G--N--DI --AVGTN-----	
<i>Synechococcus</i> sp. CC9902	78185430	-----G-----I --ETGV-----	
<i>Synechococcus</i> sp. CC9605	78212020	-----G-----I --ETGV-----	
<i>Synechococcus</i> sp. WH 8109	260436168	-----G-----I --ETGV-----	
<i>Synechococcus</i> sp. RS9917	87124896	---A---G-----I --ETGV-----	
<i>Synechococcus</i> sp. WH 5701	87300978	---A---G--S---I --ETGV-----	
<i>Synechococcus</i> sp. RS9916	116075862	---A---G-----I --ETGV-----	
<i>Synechococcus</i> sp. JA-3-3Ab	86605358	---V---VN-----NI Y-ESVAN----T-----	
<i>Trichodesmium erythraeum</i> IMS101	113478046	---A---Q----- --EV--N-----	
<i>Microcoleus vaginatus</i> FGP-2	334121257	---T---G-----I --EGNV-----	
<i>Oscillatoria</i> sp. PCC 6506	300864994	---A---K-----I --ES-A-A-I-----	
<i>Arthrospira maxima</i> CS-328	209524018	---A---Q----- --EV-DN--L-----	
<i>Arthrospira</i> sp. PCC 8005	376003878	---A---Q----- --EV-DN--I-----	
<i>Arthrospira platensis</i> Paraca	284051286	---A---Q-----L --EV-DN--I-----	
<i>Moorea producta</i> 3L	332706537	---A---D-----I --ETVTNA-L-----	

Supplementary Figure 106

Partial sequence alignment for the protein photosystem II chlorophyll-binding protein CP47, showing a 1 aa insert that is specific for *Nodularia spumigena* CCY9414.

			76		116
	Nodularia spumigena CCY9414	119510625	NCPDWLPLVISINGE	SPTETLRE	RFRLDQGEILRYDRQLDL
	Nostoc punctiforme PCC 73102	186685044	-----IVIV--D		-----S-----
Nostocales/	Anabaena variabilis ATCC 29413	75909956	-----MI-A--D		--M-----Q-E-K--V
Stigonematales	Nostoc sp. PCC 7120	17228553	-----MI-A--D		--M-----V-H-E-K--V
	'Nostoc azollae' 0708	298489783	-----VM-DS-		-----K--E-N-K---
	Cylindrospermopsis raciborskii CS-505	282899806	----C-----T-E-D		--G----T--Q-N-E---
	Raphidiopsis brookii D9	282897273	----C-----T-E-D		--G----T--Q-N-E---
	Fischerella sp. JSC-11	354568748	---N---T-KVA--		--M-Y----S-E-R---
	Moorea producta 3L	332708741	-----I--VVE--		-----H-E-----
	Cyanothece sp. PCC 7425	220909284	-----LV---D		-----S-E---N-
Other Bacteria	Cyanothece sp. PCC 7822	307154677	-----G-I-T-DE-		-----S-E-S---
	Anaerolinea thermophila UNI-1	320161895	-A---E-HLFL---		F--M-R---H-R-T---
	Chloroflexus aggregans DSM 948	219849147	-A---AMQVLLD--		K-S-AT-T--A-R-E--M
	Chloroflexus aurantiacus J-10-f1	163847579	-A---ATQ-VLD--		K-S-AS-TV-A-R-E---

Supplementary Figure 107

Partial sequence alignment for the protein HAD-superfamily hydrolase subfamily IA, variant 3, showing an 8 aa insert that is specific for *Nodularia spumigena* CCY9414.

		104	153
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119513191	DVLENGVNSIYANYFDIAWNSP ISN GFGDSQQPILAPLLGNFYGESLENG
	Nostoc punctiforme PCC 73102	186685754	-I--H-PD-S-TD---LS--A- ---R-ER-----D---A----
	Anabaena variabilis ATCC 29413	75907649	----H-PD-S-TD---VC---- -AN--ER-----D-----K-
	Nostoc sp. PCC 7120	17227663	----H-PD-S-TD---VC---- -AN--ER-----D-----K-
	Fischerella sp. JSC-11	354565627	-----SD-Q-RDF---D-HHH YPEN-GKV---F--K-C-DC--S-
	Oscillatoria sp. PCC 6506	300868086	-----HD-D-FD---E--H- DENLKGR---M-----DC----
	Microcoleus vaginatus FGP-2	334118968	----H-PD-D-FD---E--HA YEHLRGRV-----C----
	Lyngbya sp. PCC 8106	119486153	-I--Y-PD-E-FEF-E-D-QH- YK-IRGRV-T----D---NC----
	Arthrospira sp. PCC 8005	376007053	----H-PD-E-FD--E-D-DQA YENIKGRV-----D---NC----
	Arthrospira maxima CS-328	209526663	----H-PD-E-FD--E-D-DQA YENIKGRV-----D---NC----
Other Cyanobacteria	Arthrospira platensis Paraca	284053893	----H-PD-E-FD--E-D-DQA YENIKGRV-----D---NC----
	Arthrospira platensis NIES-39	291565708	----H-PD-E-FD--E-D-DQA YENIKGRV-----D---NC----
	Microcoleus chthonoplastes PCC 7420	254413906	E----P--Q-KD---N-EHH YEGIRERV---F--K-C-DC--S-
	Cyanothece sp. PCC 8802	257059293	-----PY-D-FD---D-EH- YT- IKGK--T---D--STC---N
	Cyanothece sp. PCC 8801	218246240	-----PY-D-FD---D-EH- YT- IKGK--T---D--STC---N
	Synechococcus sp. PCC 7335	254424500	----Y-PD-E-VD---E-EHS YE-I-GKV--M--D--DRC--R-
	Cyanothece sp. PCC 7822	307154096	-----KD-E-Y-F---E-EHH YE-FREKV-T-M--D---NC----
	Synechococcus sp. JA-3-3Ab	86605118	----S-P--DFID--VN-DHS YLGIKGR---F--D--NKC--S-
	Cyanothece sp. PCC 7424	218441521	----H-KD-E-YTF---E-EHH YE-LREKV-T-M--D---NC----
	Synechococcus sp. JA-2-3B'a(2-13)	86609605	----S-P--DFVD--VN--HS YLGIKGR---F--D--DKC--S-

Supplementary Figure 108

Partial sequence alignment for the protein Alpha amylase, showing a 3 aa insert that is specific for *Nodularia spumigena* CCY9414.

			40		81
	Nodularia spumigena CCY9414	119513308	YVQGF	FLAVAQMPYQPIEHGWIELSE ATES	DFIVSIIDPSLPH
	Nostoc punctiforme PCC 73102	186685662	-----	TFVGK--R----S----GD	ASE-R----T---
Nostocales/	Anabaena variabilis ATCC 29413	75908479	-----	I-GK--K-K-----GD	-V---T---
Stigonematales	Nostoc sp. PCC 7120	17232497	-----	--GK--K-K-----VGD	-V---T---
	Fischerella sp. JSC-11	354566348	-----	TKGK---M-YA-----D	R-V--T---
	Raphidiopsis brookii D9	282895589	-----	LTGH--K----A-----ED	V-----FAY
Other	Cylindrospermopsis raciborskii CS-505	282899365	-----	VLTGH-----A-----ED	V-----FTY
	'Nostoc azollae' 0708	298491431	-----	VVFTGQ--K-V--A----QD	V----TF-Y
Cyanobacteria	Microcoleus vaginatus FGP-2	334120024	-----	PGV--K---YS----D-	R-V--TF-

Supplementary Figure 109

Partial sequence alignment for the hypothetical protein N9414_11444, showing a 4 aa insert that is specific for *Nodularia spumigena* CCY9414.

		231	268
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119511494	AKLQAVRGKVA LAKLSVG IANAKIAYQEYKIIQSDRW
	Anabaena variabilis ATCC 29413	75907902	---V--K----
	Nostoc punctiforme PCC 73102	186684150	---K--K----
	Nostoc sp. PCC 7120	988292	---MV-K----
	'Nostoc azollae' 0708	298490388	-V-RG-----
	Fischerella sp. JSC-11	354568215	---K-IK----
	Lyngbya sp. PCC 8106	119486354	---EK-K----
	Oscillatoria sp. PCC 6506	300866090	---E--K----
	Microcoleus vaginatus FGP-2	334118085	---E--K--I-
	Trichodesmium erythraeum IMS101	113474532	---E--K----
Other Cyanobacteria	Microcoleus chthonoplastes PCC 7420	254413190	---KD-K----
	Moorea producta 3L	332710463	---KGIE----
	Lyngbya sp. PCC 8106	119486005	-R-SQIK----
	Cyanothece sp. PCC 7425	220908802	-R-E--K----
	Cyanothece sp. PCC 7424	218438477	VR-REIK----
	Cyanothece sp. PCC 7822	307153981	V--K-IK--I-
	Acaryochloris sp. CCMEE 5410	359460904	TQIAQ-K----
	Synechococcus sp. PCC 7002	170078069	S--IGLK----
	Cyanothece sp. CCY0110	126657827	VR-RELK-DI-
	Cyanothece sp. ATCC 51142	172036291	MR-RELK-D--
Acaryochloris marina MBIC11017	158339390	VL--SLQ----	

Supplementary Figure 110

Partial sequence alignment for the protein transaldolase, showing a 7 aa insert that is specific for *Nodularia spumigena* CCY9414.

		859	899
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119512252	RFLPSWYGIGTALQGFLE EE EESEEHKLLRYFYLKWPF
	Nostoc punctiforme PCC 73102	186685513	-----V-----E--N A-P-----M----V-----
	'Nostoc azollae' 0708	298490793	-----L-----E--N -QP-----V-----
	Nostoc sp. PCC 7120	161353735	-----V-----E-FN --P-----M----V-----
	Anabaena variabilis ATCC 29413	75908353	-----V-----E-FN --P-----M----V-----
	Raphidiopsis brookii D9	282897196	-----V-----NE--Q -QP-----M----I-----
	Cylindrospermopsis raciborskii CS-505	282900199	-----V---DE--Q -KP-----M----I-----
	Fischerella sp. JSC-11	354565578	-----V-----E--N --P--N-Q-----M-----
	Microcoleus chthonoplastes PCC 7420	254413284	-----V-A--K----- --P--N-----F-----
	Microcoleus vaginatus FGP-2	334120182	-----V-----E--L --P--M--Q-----
	Oscillatoria sp. PCC 6506	300867228	-----V-----E--L -KP-----Q-----
	Arthrospira platensis Paraca	284050999	-----V---IEQ-IR --P--NI-----I-----
	Trichodesmium erythraeum IMS101	113474661	----A--V-----E-VE K-P-----Q---V-----
	Moorea producta 3L	332710484	-----V-----S-V QAP--NIN--C--S-----
	Other Cyanobacteria	Lyngbya sp. PCC 8106	119487042
Arthrospira maxima CS-328		209528283	-----V---IEQ-IQ --PQ-NI-----I-----
Cyanothece sp. PCC 8801		218248219	----A--V-----S-V --P--N-----
Cyanothece sp. PCC 7424		218441147	----A--V---EQ-IN R-P-----
Cyanothece sp. PCC 7822		307152504	----A--V---EQ-IS T-P-----
Acaryochloris marina MBIC11017		158337676	----A--V---D--- -K-A--T--Q--G-----
Acaryochloris sp. CCMEE 5410		359459163	----A--V---D--- -K-A--T--Q--G-----
Cyanothece sp. CCY0110		126658238	----A--V---EN--Q Q-PD-N-----
Crocospaera watsonii WH 8501		67923850	----A--V---E--Q Q-P--N--FK-----
Crocospaera watsonii WH 0003		357262297	----A--V---E--Q Q-P--N--FK-----
Cyanothece sp. ATCC 51142		172038735	----A--V---ES--Q Q-PN-N-----
Cyanothece sp. ATCC 51472		354556129	----A--V---ES--Q Q-PN-N-----
Thermosynechococcus elongatus BP-1		22299455	----A--V---KE--E -KPA--S-----Y-----
Synechococcus sp. PCC 7335		254422587	-----V---KE-I AQPT--I-Q----A-----
Synechococcus elongatus PCC 6301		56751855	-----V---E--Q -RP-QN-N-----E-----
Synechococcus elongatus PCC 7942		81301061	-----V---E--Q -RP-QN-N-----E-----
Synechococcus sp. PCC 7002		56417897	--V-A--V-----FE -DPV-N--M--S-----
Cyanothece sp. PCC 7425		220907631	----A--V---I-D--A -KPV--S--Q--F-----
Microcystis aeruginosa PCC 7806		159027003	----A--L-----M--- DSPNRN-E---H--H-----
Synechocystis sp. PCC 6803		16330985	----A--V---KS-V QDPVKNM-----F-----
Microcystis aeruginosa NIES-843		166366623	----A--L-----M--- DSPSRN-E---H--H-----
Synechococcus sp. JA-2-3B'a(2-13)		86609298	----A--V-----EYVA RG-HN-NH-QHL-RE----
Synechococcus sp. RS9917		87124941	-----F-V---ADEVA NDQ-Q-E---RLHQR----
Synechococcus sp. WH 5701		87300934	-----F-V-A---EE-- DDPGQ-E---LL-QR----
Cyanobium sp. PCC 7001		254431278	-----F-V-A---REE-- HDP-Q-E---LL-QR----
Synechococcus sp. JA-3-3Ab		86605542	----A--V---EEYAA KG-HN-NH-QHL-RE----
Synechococcus sp. WH 8102		33866579	-----F-V---AEEVN DDP-Q-D---RLHQR----
Synechococcus sp. WH 7805		88807863	-----F-V---HEE-E NDPDQMS---TLHQR----
Synechococcus sp. WH 8016		352096542	-----F-V-S---SEE-E ADPDQ-T---TLHQR----
Synechococcus sp. CC9311		113954224	-----F-V-S---SEE-E ADPDQ-T---TLHQR----
Synechococcus sp. RS9916		116075789	-----F-V---AEEVA SDP-Q-DV---RLHQR----
Synechococcus sp. RCC307		148243189	-----F-V-A---IREE-G DDG-Q-DT---TL-QR----
Synechococcus sp. WH 7803		148238790	-----F-V---HEE-V NDPDQMS---TLHQR----
Synechococcus sp. WH 8109		260436235	-----F-F---SEEVG GDS-Q-D---RLHQR----
Synechococcus sp. CC9605		78211948	-----F-F---TEEVG GDS-Q-D---RLHQR----
Synechococcus sp. CB0101		318042586	-----F-V-A---EE-- GDPGQ-E-FQQL-QR----
Synechococcus sp. CB0205		317968794	-----F-V---EE-S GDPGQME-FQQL-QR----
Synechococcus sp. BL107		116072588	-----F-F---AEEVK ADP-Q-D---RLHQR----
Synechococcus sp. CC9902		78185499	-----F-F---AEEVK ADPDQ-D---RLHQR----
Prochlorococcus marinus CCMP1375		33241179	-----F-V---EEE-K SDPD-IE---MLNQR----
Prochlorococcus marinus MIT 9303	124023967	-----F-V---AAEVE SDADQ-D---RLHQR----	
Prochlorococcus marinus CCMP1986	33862131	-----F-V---SVE-K SDP-QIE---VLHQR----	
Prochlorococcus marinus AS9601	123969314	-----F-V---SSE-N LDPQQIE---VLHQR----	
Prochlorococcus marinus NATL2A	72382985	-----F-V---ATE-K ADPDQMEM---MLNQR----	
Gloeobacter violaceus PCC 7421	37519983	----QA--L---D--IR CNR-RN-AE--SM-RQ----	

Supplementary Figure 111

Partial sequence alignment for the protein phosphoenolpyruvate carboxylase, showing a 2 aa insert that is specific for *Nodularia spumigena* CCY9414.

		125	159
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119512923	AGLIKPGWEKTPN G DGIVHRVGVIVTREGNPKN
	Anabaena variabilis ATCC 29413	75910911	-----QEA-- NS--T---VALE-----
	Nostoc sp. PCC 7120	17227818	---Q---EL-- N---SK--AA-II-P---D
	Raphidiopsis brookii D9	282897042	---Q---EA-- NA--SQ--AAL---P-----
	Nostoc punctiforme PCC 73102	186685453	---Q---EV-- N---SK--AAL---P---G
	'Nostoc azollae' 0708	298492152	---Q---EF-- NS--SK--AAL-----G
	Cylindrospermopsis raciborskii CS-505	282901328	---QL---EA-- NA--SQ--AAL---P-----
	Fischerella sp. JSC-11	354564754	---E---EL-- NAV-S---AALI--P-----
	Arthrospira platensis NIES-39	291570705	---E---NEA-- NA-----AA-I--P-----
	Arthrospira platensis Paraca	284051317	---E---NEA-- NA-----AA-I--P---Q-
	Arthrospira maxima CS-328	209527649	---E---NEA-- NA-I---AA-I--P-----
	Microcoleus vaginatus FGP-2	334117932	---Q---EV-- E---SK---A-----G
	Oscillatoria sp. PCC 6506	300869606	---E---ES-- ---SK--AAL-----G
	Lyngbya sp. PCC 8106	119486222	---E---QEL-- -S---K--VAF-L-DQ- S-
	Cyanothece sp. PCC 7822	307151272	-----EA-- GA--T---ASLM--K----K
	Microcystis aeruginosa PCC 7806	159029008	---Q---EA-- -S---K--VAF-P-DA-I-
	Other Cyanobacteria	Synechococcus sp. JA-2-3B'a(2-13)	86607738
Cyanothece sp. PCC 7425		220908865	-----D--R--- S---T---VAL---P---G
Cyanothece sp. PCC 7424		218440387	---Q---EF-- -S---K--AAL-K-DDSV-
Microcystis aeruginosa NIES-843		166365896	---Q---EAT-- -S---K--VAF-P-DA-I-
Synechococcus elongatus PCC 6301		56752418	---Q---QQRV-- N--ITN--VAL--Q-----G
Synechococcus elongatus PCC 7942		81300490	---Q---QQRV-- N--ITN--VAL--Q-----G
Synechococcus sp. JA-3-3Ab		86605189	---R---SR--- Q---T---AA---Q---RG
Cyanothece sp. PCC 8801		218247536	---D---EL-- ES---K--VAF-P-DEKV-
Cyanothece sp. PCC 8802		257061128	---D---EL-- ES---K--VAF-P-DEKV-
Synechocystis sp. PCC 6803		16332111	---Q---QEA-- GS--TN--IAF--KASDNIK
Cyanothece sp. ATCC 51142		172038743	---D---TEL-- ES---K--VAF-P-NEEI-
Cyanothece sp. ATCC 51472		354556121	---D---TEL-- ES---K--VAF-P-NEEI-
Thermosynechococcus elongatus BP-1		22297818	---D-D-TSRA-- NS--TT--CALI-----
Synechococcus sp. PCC 7002		170077417	---D---TEA-R -SVITS--IA--P-DP-L-

Supplementary Figure 112

Partial sequence alignment for the protein thiosulphate-binding protein, showing a 1 aa insert that is specific for *Nodularia spumigena* CCY9414.

		480	540
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119508927	PLVIFLDDLQWADFTSLKLIQLLNSAS NDSENIAMS ATLEKQKSLLIIGAYRDNEVNNNAHP
	Anabaena variabilis ATCC 29413	75909801	-----SG-----MNDT GY--L-----P---
	Nostoc sp. PCC 7120	17231183	-----SG-----MNDT GY--L-----P---
	Nostoc punctiforme PCC 73102	186683708	-----V-NA-----ATVT -GQ--LM-----A---
	Fremyella diplosiphon Fd33	48374205	-----SA-----MADT CH--L-----PP---
	'Nostoc azollae' 0708	298491722	---M-----SA-----MADT SH--L-----PPG---
	Cylindrospermopsis raciborskii CS-505	282898911	-----P-----F-IKEP DT-Y--L-----H--YSG--
	Raphidiopsis brookii D9	282898072	-----S-----F-IKEP DT-Y--L-----H--YSR--
	Microcoleus vaginatus FGP-2	334117410	-----AA---M---M-E- DARY--L-----SA---
	Trichodesmium erythraeum IMS101	113475755	-----TA---ME--M-DK NSQH--L-----P---
	Oscillatoria sp. PCC 6506	300868404	-----SA---M---IANT SY-FLM-----P---
	Lyngbya sp. PCC 8106	119487723	-----AA---M---MADT GY-F-----P---
	Moorea producta 3L	332708490	-----RA--E---MT-A DN-Y-FF-----EV--
	Arthrospira sp. PCC 8005	376003847	-----LA--E-MK--MEEK TY--LL-----SPV--
	Arthrospira maxima CS-328	209523991	-----LA--E-MK--MEEK TY--LL-----SPV--
	Arthrospira platensis NIES-39	291571924	-----LA--E-MK--IEEK TY--LL-----SPV--
	Arthrospira platensis S6	206585493	-----W--S-SA--N-LKM-MQET -GY--L-----H--SPV--
	Other Cyanobacteria	Arthrospira platensis Paraca	284052405
Cyanothece sp. PCC 7822		307151273	---L-----LA--F-FM-MT-- DTQY--L-----PS---
Cyanothece sp. CCY0110		126658537	---M-----I-AA---FKH-M-E- -DCH--L-----S-L--
Cyanothece sp. ATCC 51142		172036732	-----S-TA--Q--RS-ITTL DDQ--F-----DL---
Cyanothece sp. ATCC 51472		354554545	-----S-TA--Q--RS-ITTL DDQ--F-----DL---
Acaryochloris sp. CCME 5410		359458100	-----AA--N-LK-IMEGV GLDINT-H--MV-----IFS---
Acaryochloris marina MBIC11017		158336697	---V-----SA-----K-MMTR QMR--L-----DAT---
Microcystis aeruginosa NIES-843		166366081	-----I-SA-----L--IVTDK RG-Y-FF-----SST--
Cyanothece sp. PCC 7425		220906228	---M-----LP--R--E--VTNP ESQY--L-----SPS---
Cyanothece sp. PCC 7424		218439689	---V-----I-AE-----EM-MTDD RI-Y-F-V-----IT--
Microcystis aeruginosa PCC 7806		159028712	-----LA--Q--E-IMLDG DMDY-FL--S--S--I-ST-

Supplementary Figure 113

Partial sequence alignment for the protein serine/threonine kinase, showing a 9 aa insert that is specific for *Nodularia spumigena* CCY9414.

		586	644	
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119509505	GRIFWQEEDGLQETDAIAGIMGVSV AH RRYRSIGKWSKGSSENPMTWFDAGILESVF	
	Anabaena variabilis ATCC 29413	75908438	-----K-----I--DT -L-----KDQ-L---E-----	
	Nostoc sp. PCC 7120	17232450	-----K-----I--DT -L-----K-Q-LA--E-----	
	Nostoc punctiforme PCC 73102	186681560	-Q--R--E--E-L--I--DR TV--SSGR-E-I--E-----	
	'Nostoc azollae' 0708	298493037	-----R--E--D-AEML-----D- TL---RDGR-QTL--E-----	
	Cylindrospermopsis raciborskii CS-505	282898724	-T--SR--E--E-SQV-----DT TV-R-FN-GR-----A-----N--	
	Raphidiopsis brookii D9	282897764	-T--SR--E--E-SQV-----DT TV-R-FN-GK-----A-----N--	
	Fischerella sp. JSC-11	354568807	-----N-----A-----L--DR RSN--TSSGR-Q--S--YE---T-D---	
	Microcoleus chthonoplastes PCC 7420	254413812	-----RD--FA-G-RV--L--DP -R--TR-GQDA--S--YE---I--L-	
	Moorea producta 3L	332710104	-----RP-E-FA-A-----L--NI TQ-N-VR-GHSE---YE---V----	
	Microcoleus vaginatus FGP-2	334119963	-----KDG-SMK-A--V--I--DP TVW--QQ-GRDR-L---E---V--A-	
	Lyngbya sp. PCC 8106	119486585	--V--K-TETF--K-----I--DT TQ-H-TR-GR-Q-I---E---AI-	
	Oscillatoria sp. PCC 6506	300865106	-----KDGEM--G-RV--I--DP VQ--QSGRDRAL---E-----	
	Trichodesmium erythraeum IMS101	113475716	-----N-VEKY--AE---VISNG- KS-TR-GQ-Q-LS--E-----Q--	
	Arthrospira maxima CS-328	209526097	-----RRDGEVY--E-----VI--DP -Q-R-VNSGKD--S--E---V---W	
	Arthrospira platensis Paraca	284052815	-----RRDGEVY--E-----VI--DP -Q-R-VNSGKD--S--E---V---W	
	Arthrospira sp. PCC 8005	376002739	-----RRDGEVY--E-----VI--DP -Q-R-VNSGKD--S--E---V---W	
	Arthrospira platensis NIES-39	291570936	-----RRDGEVYH-E-----VI--DP -Q-R-VNSGKD-A-S--E---V---W	
	Other Cyanobacteria	Cyanothece sp. PCC 7425	220905838	--V-----N--Y-A--L--I--DP FRGKVVRRSQQ-Q-LS-YE-----D---
		Cyanothece sp. PCC 7822	307151961	-----YKSNE-I--Q-SL--L--EV TPL-R-TR-GKSQ--S--YE-----
Cyanothece sp. PCC 7424		218441654	-----QKTK--K-E-LLS--L--NL TAG-R-T--GKPT-S-Y---L----	
Cyanothece sp. ATCC 51142		172036236	-----RRINGEIA-W-N-----DL LPQ-R-TRSGKAD---YE-----	
Cyanothece sp. CCY0110		126656357	-----RKIDGERA-W-N-----DL LPQ-R-TRSGKPK---YE-----	
Synechococcus elongatus PCC 6301		56750269	-----R----FF-A-RMG--L--DP -R-R-QR-GK-QAID-YA---V--EI-	
Synechococcus elongatus PCC 7942		81300102	-----R----FF-A-RMG--L--DP -R-R-QR-GK-QAID-YA---V--EI-	
Microcystis aeruginosa NIES-843		166366292	--V-RL-S-NIN-Y-----F--DF FPA-R-LRSRK-VA---YE-----A--	
Microcystis aeruginosa PCC 7806		159028409	--V-RL-S-NIN-Y-----F--DF FPA-R-LRSRK-VA---YE-----A--	
Crocospaera watsonii WH 0003		357262434	-----RKIDGEIA-W-N-----DL LPQ-R-TRSGKA---S--E-----R--	
Crocospaera watsonii WH 8501		67922517	-----RKIDGEIA-W-N-----DL LPQ-R-TRSGKA---S--E-----R--	
Thermosynechococcus elongatus BP-1		22298204	--V-G-D-E--W-N-RLG--I--DP WQGWVRRSQQPQ-L--YE-----	
Cyanothece sp. PCC 8802		257061049	-----RIIDGKD--W-E---F--DV FPQ-R-TRSGKSV---YE-----	
Cyanothece sp. PCC 8801		218247617	-----RIIDGKD--W-E---F--DV FPQ-R-TRSGKSV---YE-----	
Acaryochloris sp. CCME 5410		359463406	-QV--K-GKK-K-AR-----L--DP AQGKWMRTANQDYA---E---L--R-L	
Synechocystis sp. PCC 6803		16330982	--V--N-GDIG-A-HL---C--SQ ITE-T-PQ-GKPK--S-Y---L--AI-	
Acaryochloris marina MBIC11017		158337435	-QV--K-GEK-K-AR-----L--DP AQGKWMRTANQDYA---E---L--R-L	
cyanobacterium UCYN-A		284929696	-K--HSIDGQNR-S-R-S-----TF LVE-R-TT-GKVP---YKS--V-D-I-	
Synechococcus sp. JA-3-3Ab		86605030	-HV-AYAPPSSP-PYQE-EHL--IL GGDLPNDWRHHSR--D--E---L-V-AL	
Synechococcus sp. RCC307		148242322	-KV-ARDG-VFS-EERLC-VLA-ER RQ-L--TSGQPS-LSYHQGR-L-ALAM	
Synechococcus sp. WH 7805	88808526	-NV-FPDAS-VRQ-SRLG-VIC-ER RLSR-LNSGKPS-LSYVE-R-L-T-L-		
Synechococcus sp. CBO205	317969857	-NV-DAKAQ-SPQSHL--VIC-SR QAEL--QSGKPPAPDY-S-R-V-Q-AL		
Synechococcus sp. WH 5701	87303685	--V-QS-GEN-L-RTSL--VIC-ER RSEL--SAGKARSLDY-E-R-V-QOAL		
Gloeobacter violaceus PCC 7421	37522495	-VV-LRSDE-IF-SGRLGAV-C-EP AVGDWQKLTLP-FD--A--V-AAIL		
Prochlorococcus marinus AS9601	123968524	-NV-HKKT-FI--EVLNGA-Y-NKK F---LNSGKD-HLNYQ-R-K-KEAL		
Prochlorococcus marinus MIT 9202	254526671	-NV-QKKPEFI--EVLNGA-Y-NKK F---INSGRD-DLNY-Q-R-K-KEAL		

Supplementary Figure 114

Partial sequence alignment for the protein phenylalanyl-tRNA synthetase, showing a 3 aa insert that is specific for *Nodularia spumigena* CCY9414.

		211		254
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119512824	AIHGATVENTLFEDGEGS S	NTFFAFNPTQAEETYSMVTANRFWS
	Raphidiopsis brookii D9	282897939	-----	-----S-----
	'Nostoc azollae' 0708	298490362	-----	-----G-----
	Nostoc sp. PCC 7120	17231782	-----A	-----R-----S-----
	Anabaena variabilis ATCC 29413	75907465	-----A	-----R-----S-----
	Nostoc punctiforme PCC 73102	186683774	-----D-A	-----R-----S-----
	Fischerella sp. JSC-11	354568337	-----A	-----R-----
	Microcystis aeruginosa PCC 9432	389675664	-----	-----R--E-----
	Crocospaera watsonii WH 8501	67925969	-----QA	-----R--E-----
	Cyanothece sp. PCC 7425	220905664	-----T-	-----R-----
Other Cyanobacteria	Synechococcus sp. WH 8109	260436021	-----QA	-----K--E--E-----
	Thermosynechococcus elongatus BP-1	22297998	-----Q--A	S--R-----
	Microcystis aeruginosa NIES-843	166364539	-----	-----R--E-----
	Synechococcus sp. WH 8020	95115474	-----Q-	-----K--E--E-----
	Synechococcus sp. WH 8101	95115490	-----Q-	-----K--E--E-----
	Synechococcus sp. WH 8017	95115476	-----QA	-----K--E--E-----
	Synechocystis sp. PCC 6803	16330392	-----D-	-----R--E-----
	Moorea producta 3L	332710007	-----Q--DKA	-----R-----
	Microcoleus chthonoplastes PCC 7420	254412285	-----	-----R-----
	Microcoleus vaginatus FGP-2	334121061	-----	-----R-----S-----
Arthrospira maxima CS-328	209527136	-----	-----R--E-----	

Supplementary Figure 115

Partial sequence alignment for the protein photosystem II protein D2, showing a 1 aa insert that is specific for *Nodularia spumigena* CCY9414.

		1254	1303
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119510861	KTQGLNLDCLLQLPDAKND APKEP LRERSWLVEHE VHSNGAVVDDQLLAD
	Anabaena variabilis ATCC 29413	75907517	---AI--N-----TREN
	Nostoc sp. PCC 7120	17231836	---AI--N-----T-EN
	'Nostoc azollae' 0708	298490067	---SI--G--TK-----QN
	Nostoc punctiforme PCC 73102	186684008	---RS-----K---TRDN
	Raphidiopsis brookii D9	282896073	---SV-----TK---T-SN
	Cylindrospermopsis raciborskii CS-505	282898975	---SV-----TK---TRSN
	Fischerella sp. JSC-11	354566107	---A-----T-T-
	Microcystis aeruginosa NIES-843	166365652	---ES-----N---GRS-
	Microcystis sp. T1-4	390442177	---ES-----N---GRG-
	Crocospaera watsonii WH 0003	357265663	---S-----N---V-S-
	Acaryochloris sp. CCME 5410	359461148	---SQ-----T---TRT-
	Cyanothece sp. CCY0110	126656852	---SI-----N---V-S-
	Crocospaera watsonii WH 8501	67920972	---S-----N---V-S-
	Cyanothece sp. ATCC 51142	172039266	---SI-----N---V-S-
	Cyanothece sp. ATCC 51472	354552465	---SI-----N---V-S-
	Cyanothece sp. PCC 8801	218248470	---KA-----K---VTR-
	Acaryochloris marina MBIC11017	158334825	---SQ-----TH---TRT-
	Cyanothece sp. PCC 7425	220905967	---TS-----T---VRT-
	Cyanothece sp. PCC 8802	257061535	---A-----K---VSS-
	Synechocystis sp. PCC 6803	16332153	---N-T-----N---T-QN
	Synechococcus elongatus PCC 7942	81299699	---AA-----IK---TRS-
	Synechococcus elongatus PCC 6301	56750659	---AA-----IK---TRS-
	Thermosynechococcus elongatus BP-1	22298911	---GP---A---TK---TRS-
	Synechococcus sp. PCC 7002	170078988	---S--V---TN---VRDN
	Cyanothece sp. PCC 7822	307153958	---S-----N---VRS-
	Cyanothece sp. PCC 7424	218438449	---S--TVE---N---EV-T-
	Synechococcus sp. JA-3-3Ab	86607312	---S---E---RT-PGFDP
	Synechococcus sp. WH 7803	148238721	---VD-SS--APIQGSSE
	Synechococcus sp. WH 8102	33866664	---VD-SS--APIQGSSE
	Synechococcus sp. CB0205	317968698	---KA-D-S--DPIPQAA-
	Synechococcus sp. RCC307	148243263	---A-D-S--APIAGAE-
	Synechococcus sp. RS9916	116075718	---SVD-SS--APISGSE-
	Cyanobium sp. PCC 7001	254431541	---RS-D-S--IDPVADPA-
	Synechococcus sp. CC9605	78211878	---K-VD-SS-RAPISGAE-
	Synechococcus sp. PCC 7335	254423302	---KA-D-SAIKV---V-S-
	Synechococcus sp. WH 7805	88807938	---R-VD-SS--APIQGSSE-
	Synechococcus sp. RS9917	87125016	---AA-D-TT--APVPGHGE
Synechococcus sp. CC9311	113952811	---KCVD-SS--APVGDA--	
Synechococcus sp. WH 8016	352096474	---KCVD-SS--APVGDA--	
Synechococcus sp. WH 5701	87301860	---SA-D-T--APLPGSD-	
Synechococcus sp. BL107	116072136	---SCVD-SS--APITGSE-	
Synechococcus sp. CC9902	78185583	---SCVD-SS--APIAGSE-	
Synechococcus sp. CB0101	318042647	---KAID-S--DPIPQAA-	
Microcoleus vaginatus FGP-2	334117205	---AI-----TK---T-S-	
Oscillatoria sp. PCC 6506	300865984	---AV-----T---T-S-	
Leptolyngbya boryana	1369908	---L-----T---TRS-	
Lyngbya sp. PCC 8106	119485449	---SAI-----IN---TRS-	
Moorea producta 3L	332705923	---SM-----T---TRT-	
Microcoleus chthonoplastes PCC 7420	254414772	---STVSVK--TD---TSR-	
Arthrospira platensis Paraca	284053653	---K--D-GV-M-Q-V-PGN	
Arthrospira platensis NIES-39	291569328	---K--D-GV-M-Q-V-PGN	
Synechococcus sp. JA-2-3B'a(2-13)	86608729	---S---E---RT-PGFDP	
Arthrospira maxima CS-328	209526123	---K--D-GV-M-P-V-PGN	
Arthrospira sp. PCC 8005	376005066	---K--D-GV-M-P-V-PGN	
Prochlorococcus marinus MIT 9313	33864044	---T-D-S--DPIPGAV-	
Prochlorococcus marinus NATL2A	72382916	---KKVD-SS--KPI-NPT-	
Prochlorococcus marinus CCMP1375	33241117	---KEVD-SS--KPIPKVQ-	
Prochlorococcus marinus AS9601	123969248	---SNID-SS--NKKYLNT-	
Gloeobacter violaceus PCC 7421	37521077	---AR-D-TS--GG-EHAG	
Other Cyanobacteria			HRKPE--E--P-----P-L--AI--R
			---R-SAEA-G--PILE-----
			---R-SAEA-G--PILE-----
			---R-A-AA-G--VILE-----
			---Q-ASEA---PILE-----
			---R-SAEA-G--VILE-----G-
			---L--HDA-G--PILE-T----
			---R-SDEA-G--PILE-----
			---G--D-GG--N--N-L--EI---
			---R-SAEA-G--PILE-E-----
			---T--Q-AAEA-G--P-LE-----
			---K-SSEA-G--PILE-H----
		---K-SPEA-G--PILE-D-----	
		---R-DSQA-G--VILE---V---	
		---T-SATA-G--PILE-DF---	
		---T-SATA-G--PILE-DF---	
		---Q-DAQA-G--PILE-----	
		---F---P-----P-L--D----	
		---F---SD-----P-L--N----	
		---T--N--T-----L-----S-	
		---D--N--G-----P-L--V--E-	
		---N-KD-----P-L-----G-	
		---F--N--A-----P-L-----	
		---D--H--P-----L--SI---	
		---D--H--P-----P-L--SI---	
		---D--N--P-----P-L--SI---	
		---A--I-DAKA-D--PILE-----	
		---S--IEA---E-LENA--K---	
		---QIEP---N-LENA--K---	
		---K-SKAA---F-LE-EF-S---	
		---D--A-S-CA-G--P-L--LI-Q-	

Supplementary Figure 116

Partial sequence alignment for the protein glutamine amidotransferase, showing a 5 aa insert that is specific for *Nodularia spumigena* CCY9414.

		394	436
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119512863	GDPGIYAMAAAVFEVCDSEA LRRNR NPKPEWDSIDIHVAPGISAMQAAAA
	Anabaena variabilis ATCC 29413	75909075	-----L-YHL -----E-N-----
	Nostoc sp. PCC 7120	17227949	-----L-YHF -----E-N-----
	'Nostoc azollae' 0708	298489716	-----T-----L-RYN -----QN---Q-----T--
	Cylindrospermopsis raciborskii CS-505	282901772	-----T---D-L-KGQ YDQ---QT---E-----T--
	Raphidiopsis brookii D9	282897636	-----T-----LEKGQ YNQ---QT---Q-----T--
	Nostoc punctiforme PCC 73102	186684877	-----T-----RY- -----T-----
	Fischerella sp. JSC-11	354566023	-----L-KQ -----QG-E-Q-----
	Arthrospira platensis Paraca	284050581	-----S-----LETK- ----EA-A-Q-C-----
	Arthrospira platensis NIES-39	291568595	-----S-----LETK- ----EA-A-Q-C-----
Other Cyanobacteria	Oscillatoria sp. PCC 6506	300867406	-----F-----LER-- ----EGLE-R-----
	Arthrospira sp. PCC 8005	376002607	-----S-----LETK- ----EA-A-Q-C-----
	Arthrospira maxima CS-328	209524687	-----S-----LETK- ----EA-A-Q-C-----
	Microcoleus vaginatus FGP-2	334117117	-----F-----AI-FD- ----QG---R-----
	Microcystis aeruginosa PCC 7806	159027611	-----LEKK- --A-QQVA-Q-C-----
	Microcystis aeruginosa NIES-843	166366948	-----LEKK- --A--A-A-Q-C-----
	Synechococcus elongatus PCC 7942	81300663	-----LETAE NLD-QTL--RIC-----
	Synechococcus elongatus PCC 6301	56752249	-----LETAE NLD-QTL--RIC-----

Supplementary Figure 117

Partial sequence alignment for the protein Cobalamin biosynthesis precorrin-3 methylase, showing a 5 aa insert that is specific for *Nodularia spumigena* CCY9414.

		170	229	
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119511606	CIADMLVSAEVLSPDGTLET-LTPAQIGYSYRSSLQ NSE SPTDRRCQRIVQTQATFQLQPG	
	'Nostoc azollae' 0708	298490185	-----L-----R--- I---EL--Q---L-- -GK-----	
	Nostoc punctiforme PCC 73102	186680973	-----L-----R--- I---EL--Q---L-- GGD-----L--A--	
	Nostoc sp. PCC 7120	17232558	-----I---Q-----V--- EEL--A--T-L-- GSN-V-----	
	Anabaena variabilis ATCC 29413	75908538	-----I---Q-----I--- EEL--G--T-L-- GSN-V-----	
	Cylindrospermopsis raciborskii CS-505	282902045	-MGE-----L----- VNRS-L-T---L-- GNE-----	
	Fischerella sp. JSC-11	354568844	-----I---Q-----I--- SE-L-----T-I-- GSD-----RM---	
	Oscillatoria sp. PCC 6506	300869090	ST-EI--N-S----T-KV-I ---QEL--R--T-A-- GSN-L-----	
	Moorea producta 3L	3324106480	ST--L--FHTD----- SMAQ ---SQDL-FR--T-I-- GSD-----K--	
	Lyngbya sp. PCC 8106	119492220	ST-EI--Q-----RS-Q--V ---QDLA-R--T-N-- NSD-V-L-----A----	
	Arthrospira maxima CS-328	209525718	-T--I--NTHI-ERS--QV -P-EKL--R--T-N-- GSD-L-----	
	Arthrospira sp. PCC 8005	376003096	-T--I--NTHI-ERS--QV -P-EKL--R--T-N-- GSD-L-----	
	Microcoleus chthonoplastes PCC 7420	254410739	-A--L--V-TI--T-VV-R ---QEL-FR--T-I-- GGT-F-----	
	Arthrospira platensis Paraca	284051262	-T--I--INTHI-ERS--HV -P-QKL--R--T-N-- GGS-L-----	
	Arthrospira platensis NIES-39	291566623	-T--I--INTHI-ERS--HV -P-QKL--R--T-N-- GGS-L-----	
	Microcoleus vaginatus FGP-2	334119874	-T--I--N-R---G-KM-I ---QELE-R--T-V-- GGD-L--E-----	
	Trichodesmium erythraeum IMS101	113477458	-M--V--N-H-IL-N-EIDI -S-QDLE-N--Y-K-- GKSIL-A-----	
	Cyanothece sp. PCC 7424	218439184	----C---L-----V-R -SAS-LN----T-N-- GDR-M-IE-----	
	Cyanothece sp. CCY0110	126658805	-A--C---V-V---KV-- ---EDLN---T-A-- GG--L-IE-----	
	Cyanothece sp. PCC 51142	172037287	-A--C---V-V---KV-- ---EDLN---T-A-- GG--L-IE-----	
	Cyanothece sp. ATCC 51472	354553832	-A--C---V-V---KV-- ---EDLN---T-A-- GG--L-IE-----	
	Cyanothece sp. PCC 7425	220906334	-T--Y-RCVQ--EA---ITE ----LD-G---N-- GS--L-L-----	
	Thermosynechococcus elongatus BP-1	22297913	-T-QR---VI-E----AV VAAREL--A--T-N-- ET--L-L--W--E--	
	Cyanothece sp. PCC 7822	307154433	-V--C---L-----V-R ---Q-LN---T-N-- GDR-L-IE-----	
	Crocospaera watsonii WH 0003	357265827	-T--S---V-V---KV-- ---EDLN---T-A-- GGK-L-IE-----T-	
	Cyanothece sp. PCC 8801	218247348	-T--C-I--V-----K-- ---KDLN---K---- GQSKL-LE-----EM-	
	Crocospaera watsonii WH 8501	67920790	-T--S---V-V---KV-- ---EDLN---T-A-- GGK-L-IE-----T-	
	Microcystis aeruginosa PCC 7806	159030201	-V--R--R-L-N---Q--- -SKEDLN----- GD--L-IE-----EAT	
	Microcystis aeruginosa NIES-843	166365247	-V--R--R-L-N---Q--- -SKEDLN----- GD--L-VE-----EAT	
	Other			
	Cyanobacteria			
	cyano bacterium UCYN-A	284929363	----L--NVVI--Y--KVK- ---KDLG--T---- NGKHL-LE-----Q-	
	Synechococcus sp. PCC 7335	254424563	ET--I--ET--VD----S- ---DL-FR--T-N-- KHLGD-L-VD--K-T--	
	Acaryochloris marina MBIC11017	158334962	-T--V-TEVHA-DK--VQV -K-EHMAFQ---I-- QSPKP-LLGV--HAN	
	Synechococcus sp. PCC 7002	170076737	-V--Q-LWLK--N---IHQ -Q-TDLD-R--T-N-- GSDHL-LE-A--E--	
Acaryochloris sp. CCME 5410	359457392	-T--V-TEVHA-DR--IQV -K-EHMAFQ---I-- QSP-P-LLGV--HAN		
Synechocystis sp. PCC 6803	16332095	-T-ET--E-T-MR---G--V --NE-L-F---T-N-- KHLGD-L-VD--K-T--		
Synechococcus sp. JA-3-3Ab	86606523	AMS-V--EVQI-DEEQEPCR -E-VDLQFG--R-R-- DSPWT--G--L--L--		
Synechococcus sp. RS9916	116074301	-T--C-M-VK-IDTKAKDDA G-MRTLNNDALAFAYR HSVLQDSALM-VS-R--E--		
Synechococcus sp. WH 7803	148238363	-T-EQ-I-VD-IRFSDAQPS -ATLSRDELAF-YRHS ALQSNRHL-VA-E--E--		
Synechococcus elongatus PCC 6301	56752360	-M--I-Q-VQ-ITDQ- ---WSRE-LQ-D--H-V-- TGHC-VS-QL-----		
Synechococcus sp. RS9917	87123323	-T-ER--AVD---L--DTCG ARLRLEPNLDFDYRH SRLQNGHHL-IRGH--E--		
Synechococcus sp. RCC307	148241125	---ES-ID-T--D-SSGQTR RMSCNELDYDYRH-AL QSEALV-LS-R-R--A-		
Synechococcus sp. WH 7805	88809338	-T-EQ-I-VD-IRLSDPKP- -ANLSREDLAF-YRHS ALQTNPHL-VA-E--E--		
Synechococcus sp. CC9902	78183608	-T--S-I-VR-M PKQGG ESFDLHCDELDFAYRH SRLQADNFV-LS-R--E--		
Synechococcus sp. BL107	116071783	-T--S-I-VR-M PKQGG QSFDLHCDELDFAYRH SRLQADNFV-LS-R--E--		
Synechococcus sp. WH 8109	260434940	-T--W-E-VR-M PLEGG NCFELQRDQLDFAYRH SRLQEDALV-LS-R-R--		
Synechococcus elongatus PCC 7942	1841789	-M--I-Q-VQEI RDQGLETSREQLQDYDYRH SVLQGTGHC-VSPQ-----		
Synechococcus sp. WH 8102	33864564	-T--W-E-VR-A PLVGG VSFELSRDELDFDYRH SRLQDEELV-LS-R-R-E--		
Synechococcus sp. CC9605	78211583	-T-EW-E-VR-V PLEGG NCFELQRHQLDFAYRH SRLQEDNLV-LS-R-R--		
Cyanobium sp. PCC 7001	254432662	-T-EW-H-VR-- -PARPE QPFELARELEFAYRH SRLQQEPL--LS-R-R-E--		
Synechococcus sp. CB0101	318042993	-T-EV-E-VT-I EPHRPE QPF-LDASELDFAYRH SRLQNEPLV-LS-R-R--A-		
Synechococcus sp. CB0205	317968397	-T--I-E-VT-V -PEQPS Q-HTIPAAALDFDYRH SRLQAEPLV-LS-R-R-SS-		
Synechococcus sp. WH 5701	87301575	-T--W-QEVT-- -PNGGD SPFVLQ-SELEFAYRH SRLQ-EPLL-LS-R-R-SS-		
Prochlorococcus marinus MIT 9211	159902562	-T--R-H-VK-I-IKEGK-FE--KKDL-F---N-V- QNEEL--LS-R-E-E--		
Prochlorococcus marinus NATL2A	72383185	--SSY-E-ITT--LT- EYQRIKGDNLNFG--Q-L-- NEKL--VS-RLK-AS-		
Gloeobacter violaceus PCC 7421	37521886	QFS-T-E-V---TET- ERRRVAAGEL-LT----L-- QRDWV-LS-RLR-A-		

Supplementary Figure 118

Partial sequence alignment for the protein UDP-N-acetylenolpyruvylglucosamine reductase, showing a 3 aa insert that is specific for *Nodularia spumigena* CCY9414.

		397	440
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119512228	VQSEISYLEQVEAA IAMLAEG IAQIENYQAAADLQALEEIRDEL
	'Nostoc azollae' 0708	298492768	--T--D----- -S--DQ--T-----
	Anabaena variabilis ATCC 29413	75909510	--A--D----- ----D--T-E-----
	Nostoc sp. PCC 7120	17231866	--A--D----- -S--D--T-E-----
	Nostoc punctiforme PCC 73102	186684685	--T--E----- ----DT-KTSE--R-----E--
	Raphidiopsis brookii D9	282896726	-NR--E----- TN--D--TKQ--E-I-----
	Cylindrospermopsis raciborskii CS-505	282900099	-NR--E----- -N--DS--TKE--E-I-----
	Fischerella sp. JSC-11	354568385	-NA--E----- ----TT-T-----E--
	Cyanothece sp. PCC 7822	307150921	-V--N-----S LN--G--TNPE--K-----E--
	Cyanothece sp. PCC 8801	218245202	--R--D-----S LT-L-T--SPE-W-T-Q-----
	Cyanothece sp. PCC 8802	257058233	--R--D-----S LT-L-T--SPE-W-T-Q-----
	Microcystis sp. T1-4	390441004	-V--N-----RSS VS-L--SSGE----D--QE--
	Microcystis aeruginosa NIES-843	166368634	-V--N-----RSS LS-L--SSSE----D--QE--
	Crocospaera watsonii WH 0003	357265026	--E--N-----DS LQ--D--NSPE--T-Q--KE--
	Other Cyanobacteria	Crocospaera watsonii WH 8501	67921510
Cyanothece sp. ATCC 51142		172035331	--E--N-----DS LQ--DD--SLE--T-Q--KE--
Cyanothece sp. CCY0110		126658280	--E--N-----DS LQ--VD--YSLE--T--KE--
Acaryochloris sp. CCMEE 5410		359458594	-L--VA-----S -Q-L-QFED----LT-HD----
Cyanothece sp. PCC 7425		220909490	-EA--Q----- VQ--CDPDHGST--LT-M--E--
Acaryochloris marina MBIC11017		158338596	-L--VV-----S -Q-L-QFEDV---LT-HD----
Synechococcus sp. PCC 7002		170076778	-ET--H--HI--S LG--A--EN-T--E--Q--QA--
Synechococcus sp. PCC 7335		254422705	-KA--D--N--DS LL--VPA--SE---K---TYL--
Synechocystis sp. PCC 6803		16330193	-EQ-LA--T---TS VQALTEFDDL--W-T-T-----
Cyanothece sp. PCC 7424		218437705	-L---E-----S LN--R--NHPE---
Synechococcus elongatus PCC 6301		56750529	-EA-LR--DSIAT- L-PLQTA-TR--WL--QD--E--
Thermosynechococcus elongatus		22299601	AEG-LA--D--A-T LAT-TS-DV-----A--
Synechococcus sp. JA-3-3Ab		86604956	-AQ--R-----T LQHLGT-RDPQ--A--R-VET--
Synechococcus sp. JA-2-3B'a(2-13)		86609586	LGQ--L-----S LQ-L-T-RDPQ--TT-K--ET--
Oscillatoria sp. PCC 6506		300868392	--E--D----- LSHL-S-HNSE--T-----
Lyngbya sp. PCC 8106		119489516	--A--D----- ----T--TPE--T-A-----
Microcoleus chthonoplastes PCC 7420		254414319	--G-VE-----S LI--DR--TTPE-----Q---E--
Moorea producta 3L		332708323	--D--Q-----S LT-LNT-NTLD--T-Q---E--
Arthrospira platensis NIES-39	291570025	-RE--N-----S LS-VA--RKP----N-K---G--	
Arthrospira platensis Paraca	284052855	-RE--N-----S LS-VA--RKP----N-K---G--	
Arthrospira sp. PCC 8005	376005650	-RE--N-----S LY-VA--RKP---T-T---G--	
Trichodesmium erythraeum IMS101	113476422	--N--H-----F LS-LN--RSPE--T-----E--	

Supplementary Figure 119

Partial sequence alignment for the hypothetical protein N9414_08994, showing a 7 aa insert that is specific for *Nodularia spumigena* CCY9414.

		42	99
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119511124	LVSRRHLELRQVNSAKNKN VC CWQVISKGTNGTFLEGMLVLQCDLPDSGLLQLAQGGP
	Nostoc punctiforme PCC 73102	186681192	-----H-D--D-GS G-RL--Q-----N-V--I-SP---NS-----
	Nostoc sp. PCC 7120	17228044	-----I--S-GGGGS --V-----N-V--I-DA--NNA-----
	Anabaena variabilis ATCC 29413	75906578	-----I--S-GGGGS --L-----N-V--I-DA--NNA-----
	'Nostoc azollae' 0708	298492440	---Y---K-FGTG-KGD A-----Q-----N-V--TKST---NS-----
	Cylindrospermopsis raciborskii CS-505	282902057	---Y---KKLPETGDVD S--IV-R-----VN-V-TSK-R-M-NY-F-----
	Fischerella sp. JSC-11	354568420	-----IS-GYDGG S-----VN-V--T-SPV-NDS-----R---
	Trichodesmium erythraeum IMS101	113474475	---Y---HKISKSPSG- I-RLV-L-----VN-V-MS-GIVS-GS-I---R---
	Oscillatoria sp. PCC 6506	300867215	---Q---KI-NSPSNS T-LLVNQ-----VN-V-MS-GVVS-GS-I---R---
	Moorea producta 3L	332707024	I-----ASA-PSGPM --LVNQ-A----N-V--S-GV--EKA-I---K--
	Arthrospira sp. PCC 8005	376002958	---Y--D-QRSTTTQ-SR -MFN-R-----VN-I-AT-GP-T-GA--E--K--
	Microcoleus vaginatus FGP-2	334116882	---Q---QR--NGRASQ -YLVNS-----LVN-I-MS-GWIS-GA-I---R---
	Microcoleus chthonoplastes PCC 7420	254412315	---Y---VTVVQSGHK --LVNQ-----N-V--SREI--NQ--IR--RE--
	Arthrospira platensis NIES-39	291570746	---Y--D-QRSATTH-SR -IFN-R-----VN-V-AS-GP-A-GA--E--K--
	Arthrospira platensis Paraca	284052401	---Y--D-QRSATTH-SR -IFN-R-----VN-V-AS-GP-A-GA--E--K--
Lyngbya sp. PCC 8106	119493807	---Y--Q--KLSENPL-W MLVN-----VN-V-TS-AVVTEGI-IE--K---	
Arthrospira maxima CS-328	209523685	V-----WNKT-GWELV NF-A---YVNNQM-NKRRIV-GMVVR-GHT--	
Cyanothece sp. PCC 7425	220908104	V----V---RSG-LWEVV NL-----Y-D-KRIH-AP-T-G-IIR--RS--	
Thermosynechococcus elongatus BP-1	22299030	V----V---RHGLQWEVV NL-----Y-D-KRIQ-AT-T-G-I-R--RS--	
Acaryochloris sp. CCMEE 5410	359461331	V---Y-V---TGAQWEVV NI-----Y-D-KR-H-AP-A-GNIMR--RS--	
Acaryochloris marina MBIC11017	158335788	V---Y-V---TGAQWEVV NI-----Y-D-KR-H-AP-A-GNIMR--RS--	
Cyanothece sp. ATCC 51142	172035916	V-----I-YEENHWQVV NL-----YID-E-IS-KQAK-GMIIR--SS--	
Cyanothece sp. CCY0110	126656525	V-----I-YEDNHWQVV NL-----YIN-E-IS-KQAI-GMIIR--SS--	
Crocospaera watsonii WH 0003	357264721	V-----I-HE-NHWEIV NL-----YI-EK-IE-KQVT-GMTIR--SS--	
Crocospaera watsonii WH 8501	67921847	V-----I-HE-NHWEIV NL-----YI-EK-IE-KQVT-GMTIR--SS--	
Synechocystis sp. PCC 6803	16330968	V-----REEE-WVAI NL-A---YM-AEM-EKLV-K-GMV-R--SS--	
Cyanothece sp. PCC 8802	257058647	V-----I-REGQWEVV -L-----YVN-K-IKTARAV-GMIVH--AT--	
Cyanothece sp. PCC 8801	218245601	V-----I-REGQWEVV -L-----YVN-K-IKTARAV-GMIVH--AT--	
Synechococcus sp. PCC 7002	170077100	V-----V---LEGEWSIV NL-----YI-DQR-E--VAK-GLIFR--TS--	
Microcystis aeruginosa NIES-843	166365375	V--R---I-HNDDHDDWE LENL-A----VN-KPIEKTAHV-GMVVR--TS--	

Supplementary Figure 120

Partial sequence alignment for the protein serine/threonine protein kinase, showing a 2 aa insert that is specific for *Nodularia spumigena* CCY9414.

			74	106
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119509039	VPVILLTAKDEVSD	GH RVAGLDAGADDYVVKPF
	Nostoc punctiforme PCC 73102	186682106	-----	-----
	Nostoc sp. PCC 7120	17228689	I-----R-----	-----
	Anabaena variabilis ATCC 29413	75906838	I-----R-----	-----
	'Nostoc azollae' 0708	298489896	--I-----A	C-----
	Nostoc flagelliforme Sunitezuoqi	308205569	T--LF----TLD-	--Q-----L----
	Raphidiopsis brookii D9	282896786	I--LF----TLD-	--Q----G---L----
	Cylindrospermopsis raciborskii CS-505	282900388	I--LF----TLD-	--Q----G---L----
	Fischerella sp. JSC-11	354567463	M-----	-----
	Microcoleus chthonoplastes PCC 7420	254414181	--I-----	-----
	Moorea producta 3L	332711034	--I-----	--S-----
	Lyngbya sp. PCC 8106	119493257	--I-----	--T---S--N--II---
	Oscillatoria sp. PCC 6506	300863556	L--LM---R-RME-	K-----L----
	Arthrospira platensis NIES-39	291568591	T-ILF----TLD-	--K-----L----
	Arthrospira maxima CS-328	209524683	T-ILF----TLD-	--K-----L----
	Arthrospira platensis Paraca	284052127	T-ILF----TLD-	--K-----L----
	Other Cyanobacteria	Microcoleus vaginatus FGP-2	334118756	T--LF----T-D-
Acaryochloris marina MBIC11017		158334650	-----I--	--E-----M----
Cyanothece sp. PCC 7822		307151193	--I-----I--	--T-----I----
Cyanothece sp. PCC 7425		220908785	I-----R---G-	--T-----
Cyanothece sp. PCC 7424		218439203	--I-----I--	--T-----II---
Cyanothece sp. CCY0110		126658291	--I-----DI--	-----I----
Cyanothece sp. ATCC 51142		172035733	--I-----DI--	-----I----
Synechococcus sp. PCC 7335		254421256	M-----GIE-	-----
Synechococcus sp. JA-2-3B'a(2-13)		86607675	--V-M--R-----	-A-----
Crocospaera watsonii WH 0003		357265272	--I-----DIG-	-----I----
Crocospaera watsonii WH 8501		67921356	--I-----DIG-	-----I----
Synechococcus elongatus PCC 6301		56752230	T-I-----I--	--S--NS-----T---
Synechococcus sp. JA-3-3Ab		86607308	--V-M--S---A-	-----
Synechococcus sp. PCC 7002		170076628	T-IL-----I--	--S--N-----T---
Microcystis aeruginosa PCC 7806		159029102	A-I-I-----IP-	--T--N-----T---
Microcystis aeruginosa NIES-843		166365570	T--LF----TID-	-----L----
Synechocystis sp. PCC 6803		16330036	T--LF----TLD-	-----G---LI---
Cyanothece sp. PCC 8801		218248106	T--LF----TLN-	--L-----L----
Acaryochloris sp. CCMEE 5410		359462309	T--LF----TLD-	--Q-----L----
Synechococcus sp. CB0101		318042450	--L---R-GTA-	--E-----LS---
Synechococcus sp. RCC307		148241662	--LM--GY---K-	K-EA---V---L----
Synechococcus elongatus PCC 7942		81300213	--LF---R-T-D-	--Q-----L----
Synechococcus sp. RS9916		116073301	T--LM---H-ALDE	--A-----FLT---
Gloeobacter violaceus PCC 7421	37521380	--LM---R-TIDS	KL--FE-----L----	

Supplementary Figure 121

Partial sequence alignment for a two-component response regulator protein, showing a 2 aa insert that is specific for *Nodularia spumigena* CCY9414.

			304		343
Nostocales/	Nodularia spumigena CCY9414	119513109	VREVAEGFITHDPLSAIHLDIV	-----	NPGVKADRKSEL--FEEGRAQEVFCOLPN
Stigonematales	Nostoc sp. PCC 7120	17232823	L---VK--E-YNI-V---L-----D	VLDNFRAFTIDPDRFPHLPELAAELAAGIRLITII	----R-S--NK------D---K---
	Nostoc punctiforme PCC 73102	186682762	--KL-DE-RQRRI-CDV-----D	YMNSYRVFTWSPKRFADPQKLI GNLKQDGFVVITIV	D---YEPEADYKV-D--LKNNY-IRKT-
	Fischerella sp. JSC-11	354566180	--QLVQQLRQRRI-CDV-----D	YMFGRVFTWNPFRFPDPQKLLSDLAEDGIKVVTIV	D---FEPQADYTVYAQ-LEKDY-IRRAD
Other	Cyanothece sp. PCC 7424	218438873	--Q-Q-----N-----D	CLDDFRAFTIDPDRFPKIREFTQELKDKGVRLIII	----PH-DNK- ---IT--I--KTQD
Cyanobacteria	Cyanothece sp. PCC 7822	307150509	--Q--QE-R-REI-CDV--F--D	YMRGFRVFSWSPKRFNPNTGLLDLSQAGFKVVTII	D---YEPEADYTV-D--IQKDY-VRK--
	Synechococcus sp. JA-3-3Ab	86606938	-LA--KQ-RQRQI-CD--Y---D	YMRGYRVFTWDPFRFPEPAQLMAQLHEAGFRGVAIV	D---F-PEADYAV-D--L--DF-IRRAD
	Gloeobacter violaceus PCC 7421	37521104	--N--QA-EAN-I-V-----D	VMDDFKAFTIDPKRFPLLGEFTRDLARRGVRLVSIL	--AI--PDLPI -RD-MERKA-VTT--

Supplementary Figure 122

Partial sequence alignment for the protein alpha-glucosidase, showing a 36 aa deletion that is specific for *Nodularia spumigena* CCY9414.

1588

	1650				
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119508927	AHEINNPINFIHGNISHVNQYVQYLLDLIAVYQQED	ANLVQQKTAEIDLDFLSEDLPKILNSM	
	Nostoc punctiforme PCC 73102	186683961	-----N-AGE---D---LV-----Y	PHS SP--EE-YE---I-----D--	
	Nostoc sp. PCC 7120	17232179	-----T-ASD---D--N-V-I--R-C	PNY SDILAEES-Q--I---A-----	
	Anabaena variabilis ATCC 29413	75908201	-----A-ASE---D--N-VTI--R-C	PNY SDILAEFY-Q--I---A-----	
	Raphidiopsis brookii D9	282898074	-----V---YA-VD-AID-I-----LNI---Y	PTP HQIIVE-SKS-N----VS---V----	
	Cylindrospermopsis raciborskii CS-505	282898911	-----V---YA-VD-AID-I-----LNI---Y	PTP HQIIVE-SKS-N----VS---V----	
	'Nostoc azollae' 0708	298491080	-----M---Y--LNYATK---D-IY-VEC--E-P	SYS EVVLRRA-KRG-EIE-IM---RL-S--	
	Fischerella sp. JSC-11	354568923	-----T-AQE---D--E---I--H-L	PNP SAQ-AE-SE-----VQ---V-D--	
	Oscillatoria sp. PCC 6506	300866979	-----V---Y-----TE-TKD---VEL-E--Y	PAP NDRISEH-QA-----AK-----D--	
	Arthrospira maxima CS-328	209527324	-----V---S---CTKE---D--K--E---SHY	PNP DPDIEEFSS-----A-----	
Other Cyanobacteria	Microcoleus vaginatus FGP-2	334118334	-----V---S--LI-TGA-TRN--E-VEL---QY	PNP TPE--A-LQ---E-I-----Q--	
	Microcoleus chthonoplastes PCC 7420	254411506	-----VT-----L--IED-S-D---VEC---Y	PNP SPHL-N-ITHVN-----Q-F--V-V--	
	Lyngbya sp. PCC 8106	119487061	-----V---Y--LDPATEFT-D---VEL---Y	PQP TPKIVE-I-----Y-I-----S--	
	Arthrospira platensis Paraca	284051851	-----V---Y--LTYARD-TKDM-E-VEL--EQY	PEP LEVI-ERIE-----V-----S--	
	Arthrospira sp. PCC 8005	376004225	-----V---Y--LTYARD-TKD--E-VEL--EQY	PEP LEVI-ERIED-----V-----S--	
	Trichodesmium erythraeum IMS101	113477605	-----Y--VKPART-TKD-----NL---YY	PER -SEILEIETT-----IR--F-SL-D--	
	Moorea producta 3L	332710883	---M-----Y--LFFASD-IED--S-VEL---YY	PNP DPEIKNRLE---E--AQ---IA--	
	Arthrospira platensis NIES-39	291570657	-----V---Y--LTYARD-TKDM-E-VEL--EQY	PEP LEVI-ERIE-----V-----S--	
	Synechococcus sp. PCC 7335	254422974	-----VS--Y--VE-TE--IED-FT-LS---RCY	PEP LPA--SC-Q--E-E--L-----AMQ-	

Supplementary Figure 123

Partial sequence alignment for the protein serine/threonine kinase, showing a 3 aa deletion that is specific for *Nodularia spumigena* CCY9414.

		215		279
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119509804	LNTALSGERYSM EALEKYAEKESDNQVR AACSSASL	SLFQEMIGNR-QRHIQKLETYLDR-LGEKPS
	Nostoc punctiforme PCC 73102	186686270	--N--A-----L--F---L-----RE-N	-V---I-A-K -----IHH -----
	Anabaena variabilis ATCC 29413	75910693	-----G-----L-----T-----VEAK	A-----T-K ----EY-----T- -----
Other Cyanobacteria	Nostoc sp. PCC 7120	17228388	-----G-----L-----T-----T-AK	A-----T-K ----EY-----T- -----
	'Nostoc azollae' 0708	298492895	-----G-----L-----L-----EE-K	N-----T-KN- --K---S--H- FH----
	Oscillatoria sp. PCC 6506	300865887	-----T-----CQ-F-H-G--T--VEL-	--LHD-VH-K -H---L--AR-NA -----
	Microcoleus vaginatus FGP-2	334116827	-----A---AC--FQN----TQ--DLQ	--LR--Q-K -N--LA--KR-ND F-----
	Microcoleus chthonoplastes PCC 7420	254410130	-----A---GRK-F-Y--Q-A--LEL-	--L--I-E-K -H--IM--NR-RD -QAFE-
	Lyngbya sp. PCC 8106	119490012	----IV---TLA-FQE-SKQV--PDFV	I-LR-A--TK ---EL--AR-RD F--QVT
	Cyanothece sp. PCC 7425	220906019	----IA---GL-VF-H--GNAV-ESM-	DV--SICA-K ---L-V--SR-GQ ---T--
	Cyanothece sp. ATCC 51142	172038353	----IM--H-TIR-F-E-TQRA--P-LK	T--N-VCTTK -N---R--SR-GD FN--V-
	Cyanothece sp. PCC 7424	218440089	----T--N-TAK-F-L----AT-FGLK	TVLE-ILVTKR- --DL-KNR-SK ---NIT
	Acaryochloris marina MBIC11017	158340884	---TIV---TL--FRQ-SHRV-HQDL-	LFL-GISSTKHY -VER--AH-YDVF--RAT
	Acaryochloris sp. CCME 5410	359464142	---TIV---TL--FRQ-SHRV-HQDL-	LFL-GISSTKHY -VER--AH-YDVF--RAT
Synechococcus sp. PCC 7335	254425163	-D--II---TL-QIKQ-SYRARSPELV	-ILT-IVDSKHK ---Q-QDR-HSGF-LVA	

Supplementary Figure 124

Partial sequence alignment for the protein peptidase C56, showing an 8 aa insert that is specific for *Nodularia spumigena* CCY9414.

			196		249
	Nodularia spumigena CCY9414	119510017	IAREKAGIIKPGCPVVVGNLPPDAKKVVR	SALSGSASL	SRAQELQSLYITPQPA
	Anabaena variabilis ATCC 29413	75909888	--G-----L-----A---M-----E----		-----CPL-L----
Nostocales/	Nostoc sp. PCC 7120	17228521	--G-----L-----T--S--E----		-----CPL-L----
Stigonematales	Nostoc punctiforme PCC 73102	186685094	-----L-----Q--L--EE---		---L--KCPIF-----
	'Nostoc azollae' 0708	298489854	--G-----I-Q--A--E----		A-SL---CPI-A----
	Cylindrospermopsis raciborskii CS-505	282899100	-----A-----I-K--E----A-I		C-GV--E-PIMVVS--
	Raphidiopsis brookii D9	282897701	-----A-----M-K--E---RA-I		C-SI--E-PIFVVS--
	Fischerella sp. JSC-11	354567289	-----L-V---A-I-P-----EQ---		--VL---CPIY-----
	Synechococcus elongatus PCC 6301	56750607	--A-----SR-----P-L-E-RT-I-		N--EQ--A-M-AIE--
Other	Cyanothece sp. PCC 7424	218437546	--G-----L-K---A-I-----S-QA--T		--L-Q-NCPVWV-S--
Cyanobacteria	Synechococcus sp. PCC 7335	254421746	--G-----F--V-AI--P--E-VAI-E		-K-K-VGAPL-WV---
	cyanobacterium UCYN-A	284928824	-----S-F-FN---I-EV-NE--AIFI		--IKA-ECPSWV---
	Acaryochloris marina MBIC11017	158334901	--G----VL--QR-A---L--E-AA--Q		AKI-K-NCPA-W----
	Oscillatoria sp. PCC 6506	300863528	--G----L-QN--A-I-Q--LE--A-IE		K-LEQ-NCPA-W-K-

Supplementary Figure 125

Partial sequence alignment for the protein folylpolyglutamate synthase, showing a 9 aa insert that is specific for *Nodularia spumigena* CCY9414.

		26		71
		VTHVDEGFIQQLTDLYRDRLKF	ILRLRSVQ	DTRILDMSSWSVSHLP
		-----E--Q-		N-----
		-----R-----E-Q		N-----
		-----D-----S--Q		N-L-----
		-----D-----I--QY-Q-		K--F-----
		-----D-----I--QY-Q-		K-C-F-----
		-----S--G-----EH-R-		N-----
		-----E--Q-----		Q-----L-----
		-----D--N-----		N-----
		-----DR--N--E----		N--F--G-----
		-----D--E--G--E----		N-----
		-----D--VN--E--Q-		N-----L-----
		-----A--LD--Q--E----		H-----H-----
		-----DR-RG--S--SD		NS-V--L-----
		-----DR-RG--S--SV		NS-V--L-----
		-----D--N--E--Q-		H-----L-----
		-----D--N--E--Q-		Q-----L-----
		-----D--N--E--Q-		Q-----L-----
		-----D--N--E--N-		K-----L-----
		-----D--N--E--N-		K-----L-----
		-----A--K-----IPA		NS-VF-----
		-----Q--D--SN--QE--Q-		Q-----L-----
		-----Q--D--SN--QE--Q-		Q-----L-----
		-----A--G-----RY-Q-		Q-----L-----
		-----DR--A--E--I--S-		N-----L-----
		-----DR--A--E--I--S-		N-----L-----
		-----DQ--D--SE--EK--		H-----L-----
		-----D-----QH-Q-		GFT--L-----
		-----D-----QH-Q-		GFT--L-----
		-----DN--H--N--E--T-		N-P--L-----
		-----DA-LKR--Q--E--Q-		NSQV--L-----
		-----DA-LKR--Q--E--Q-		NSQV--L-----
		-----D--D--Q--GTL-Q-		ND-V--L-----
		-----DF-LAR-----QY-Q-		QM-V--L-----
		-H-L-A--RAR--Q--E-IP-		CAEV--L-----
		-H-L-A--RAR--A--E-IP-		CAQV--L-----
		-Q-L--A-RTR--Q--E-IP-		CATV--L-----
		-Q-L--A-RRR--Q--Q-IP-		CAVV--L-----
		-H-L-A--RAR--A--HE-IP-		CAQV--L-----
		-H-L--A-RAR--A--EQIP-		CATV--L-----
		-Q-L-DA-RAR--Q-----IPA		RAVV--L-----
		-Q-L-G--RGR--Q--E-IP-		CAVV--L-----
		-H-L-GA-RNR--A--E-IP-		CAVV--L-----
		-H-L-AP-RER--T--QK-P-		CAVV--L-----
		-Q-L-A--RGR--Q--SE-IPG		CAVV--L-----
		-Q-L--RTR--NF--EHIPS		GAVV--LG-----
		-H-L-QA-RTR--A--TE-IPK		NAVV--L-----
		-Q-L--RTR--SF--EHIPS		GAIV--LG-----
		-H-L-AA-RLR--Q--Q-IPS		CAVV--L-----
		--L-DL-LKE-----K-ISK		-SI--L-----I----
		-N-L-Q--RSR--Q-----ITS		GSVV--L-----
		-Y-L-AN-R-Y-SNV-KNEIAD		-STV--L-----D-Y--
		-Y-L-AN-R-N-S--EREIDN		NSTV--L-----D-Y--
		-Q-I-LS--AN-QEVF-QY-PE		NAT--T-----

Supplementary Figure 126

Partial sequence alignment for the protein methyltransferase type 11, showing an 8 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		26		84	
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186682826	IPLISTGEILREAIS	GDQPLPEFRWSEADPRTLSLVYA	SLSELGYHARPYIQKGELVDP
	Nostoc sp. PCC 7120	17227564	-TQ-----		H-----R--Q--MI-----
	Anabaena variabilis ATCC 29413	75908882	-TQ-----		H-----RQ-Q--MV-----
	'Nostoc azollae' 0708	298491417	--H----D---Q--K		DQTP--LK-QV-MD-----
	Nodularia spumigena CCY9414	119511160	--H----D---Q-MQ		EQTP--IK-QS-VNS-D----
	Cylindrospermopsis raciborskii CS-505	282899971	V-H----D---Q--T		DQTA--VK-QE-MD--D----
	Raphidiopsis brookii D9	282898338	V-H----D---Q--T		DQTA--VK-QE-MD--D----
	Fischerella sp. JSC-11	354566739	---A---L--Q--A		NFTN--RF-Q--MAT-----
	Microcoleus chthonoplastes PCC 7420	254409923	--VE--S-N---R--A		AQTD--KR-Q--VE-----
	Lyngbya sp. PCC 8106	119493584	--S-----V--N--A		NQTN--QK----VE-----
	Oscillatoria sp. PCC 6506	300867700	--C----D-----A		-QTN--RL-E--VE--D--R-
	Moorea producta 3L	332712149	--H-----G-V-		AQTL--QK-QS-MD-----
	Microcoleus vaginatus FGP-2	334118630	L--LAV-D-----A		EDTD--KR-VS-VER-----
	Trichodesmium erythraeum IMS101	113476560	V-H----D---SNVA		ER-P--IK-KD-MD--D----
	Arthrospira maxima CS-328	209527122	--H-----DHV-		KQTP--QQ-KT-MDL-D----
	Arthrospira platensis Paraca	284053322	--H-----DHV-		KQTT--QQ-KT-MDL-D----
	Other Cyanobacteria	Synechocystis sp. PCC 6803	16329235	L-V--M-G-----A	
cyanobacterium UCYN-A		284929580	--H----M-----A		GQTK--QE-QS-VD-----
Microcystis aeruginosa NIES-843		166368465	--H-----Q--T		EKT--QQ-QA-VE-----
Microcystis aeruginosa PCC 7806		159026035	--H-----Q--T		EKT--QQ-QA-VE-----
Cyanothece sp. CCY0110		126658008	--H----M-----A		KQTP--QK-QT-VD-----
Cyanothece sp. PCC 7822		307154823	--H-----Q--T		QSTP--LK-KI-MD-----
Crocospaera watsonii WH 0003		357262602	--H----M--A--A		KQTS--QK-QG-MD-----
Crocospaera watsonii WH 8501		67923558	--H----M--A--A		KQTS--QK-QV-MD-----
Synechococcus sp. PCC 7002		170077666	--H-----A--A		NQT--QQ-QGFVDR-----
Acaryochloris sp. CCME 5410		359457961	--H----DLF-A---		GKTP--LKVQS-LDQ-Q----
Cyanothece sp. PCC 8802		257058146	--H----M--Q--A		QQ-D--QK-QI-VDR-D----
Cyanothece sp. PCC 8801		218245121	--H----M--Q--A		QQ-D--QK-QI-VDR-D----
Acaryochloris marina MBIC11017		158337829	--H----DLF-A---		GKTP--LKVQS-LDQ-Q----
Cyanothece sp. PCC 7425		220906738	--H----D-F-A-VR		NQTP--QQVQA-LDS-R----
Synechococcus elongatus PCC 6301		56751893	LAH----L--A-VT		AQTP--IE-KG-MDR-----
Synechococcus sp. PCC 7335		254423030	V-H----D---A-VK		EKTD--LK-EGFMKA-----
Cyanothece sp. PCC 7424		218440642	--H----D---Q--T		DSTP--IK-QS-MD-----
Thermosynechococcus elongatus BP-1	22297644	--H----D---AERA		AGTL--QQ-QS-MDR-----	
Gloeobacter violaceus PCC 7421	37520961	L-QV---D---A-VK		EGTP--LE-QS-MNR-A----	

Supplementary Figure 127

Partial sequence alignment for the protein adenylate kinase, showing a 23 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		71	117
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186684080	AEIIRTANQTLLEEILPIMIDRVRSQ Q KSLVRLHSGDPSLSYSAIHEQM
	Nostoc sp. PCC 7120	17232190	---Q---K---I-L-A-----
	Anabaena variabilis ATCC 29413	75908189	---Q---K---I-L-A---H
	'Nostoc azollae' 0708	298490255	---K---K---SM--GK---H
	Nodularia spumigena CCY9414	119509992	---P---K---IEL-ME---H
	Raphidiopsis brookii D9	282898070	--V-G--TK-----VGTI-AA-Q-N
	Cylindrospermopsis raciborskii CCS-505	282898912	-QV-G--TK-----VGTI-AA-Q-N
	Fischerella sp. JSC-11	354565942	-----P---K-----L-Q-----
	Microcoleus chthonoplastes PCC 7420	254415290	-----G-K---D---L-E-----
	Oscillatoria sp. PCC 6506	300866399	-----Q---K-----A-VVE-----
	Lyngbya sp. PCC 8106	119485992	--L-A--K-----V-M-E---QG
	Trichodesmium erythraeum IMS101	113475244	-----T---K-----L-----KG
	Moorea producta 3L	332710496	-----SG-K---V-L-E---H
	Arthrospira platensis Paraca	284054258	---P---L---D---V---QKG
	Other Cyanobacteria	Arthrospira sp. PCC 8005	376003399
Arthrospira maxima CS-328		209527320	---P---L---D---V---QKG
Microcoleus vaginatus FGP-2		334120775	--V-Q--DK-----V-L-VE---GG
Cyanothece sp. PCC 7425		220907304	--R-P--K---D---L-E---H
Synechococcus sp. PCC 7335		254421973	-KVA--KM---A-V-L-E---G
Acaryochloris marina MBIC11017		158337698	--L-P--TK---D-V-M-----H
Acaryochloris sp. CCMEE 5410		359458354	--L-P--TK---D-V-M---EK---H
Cyanothece sp. PCC 7822		307150515	-QL-P-G-K---I-L---W-KNN
Synechocystis sp. PCC 6803		16329293	--CVA-G-K---T-V-L--TA--QG
Crocospaera watsonii WH 8501		67923836	C-L-A-G-K---V-MK-----NN
Cyanothece sp. CCY0110		126657344	C-L-P-G-K---D-MK--NQ-KNN
Cyanothece sp. PCC 7424		218441249	-QL-P-G-K---N-I---EG-KNN
Cyanothece sp. ATCC 51472		354553316	C-L-P-G-K---D-MK---N--KNH
Cyanothece sp. ATCC 51142		172036533	C-L-P-G-K---D-MK---N--KNH
Crocospaera watsonii WH 0003		357262318	C-L-A-G-K---V-MK---Y---NN
Synechococcus sp. WH 7805		88807598	C-Q--STL---QVI-L---SSQG
Synechococcus sp. WH 7803		148240189	C-R--STL---VI-L---ANQG
Synechococcus sp. RS9916		116073462	C-R--STL---DV--L---A-QG
Microcystis aeruginosa PCC 7806		159028916	--L-P--SSI-----I-L-----RG
Microcystis aeruginosa NIES-843		166365310	--L-P--SSI-----I-L-----RG
Synechococcus sp. WH 8109		260436008	C-R--SMT---LI-LL---Q-QG
Synechococcus sp. CC9605		78212175	C-R--STT---DLI-LL---Q-QG
Cyanothece sp. PCC 8801		218248952	-QL-P-G-K---Q-IDLI-KS-KNN
Synechococcus sp. BL107		116072816	C-K---STM---VI-LLV--QK-G
Cyanothece sp. PCC 8802		257062038	-QL-P-G-K---Q-IDLI-KS-KNN
Synechococcus sp. RS9917		87123753	C-R--SSTL---DVV-L-----QG
cyanobacterium UCYN-A		284929741	--L-P-G-K---S-VKLIK--M-G
Synechococcus sp. CB0205		317970496	C-R--STL---VMLL-L---KQG
Synechococcus sp. WH 8102		33866376	C-R--STL---VI-LL---Q-QG
Synechococcus sp. CC9311		113953234	C-T--SSTL---DV--L---AKKG
Synechococcus sp. WH 8016		352094870	C-K--SSTL---DV--L---AKKG
Synechococcus sp. CC9902		78185303	C-K--STM---DVI-LLVE-HK-G
Synechococcus sp. CB0101		318040334	C-RV--STL---VMAV-LE-TAAG
Synechococcus elongatus PCC 7942		81300038	--R-LP--HR-----LKE--AAG
Synechococcus elongatus PCC 6301		56750330	--R-LP--HR-----LKE--AAG
Synechococcus sp. RCC307		148241800	-RL-SG-L---SVMEA-L-GI-RG
Synechococcus sp. WH 5701		87302097	-R---STL---VMELVVE-A-AG
Synechococcus sp. JA-3-3Ab	86606171	-A-D-RSL---TL--Q-VE--KAG	
Prochlorococcus marinus str. M	33863592	C-R--SSM---V--LLVN-AQAG	
Prochlorococcus marinus subsp. uncultured Prochlorococcus mar	91070405	S-K-K-STL---K-TS---E-FNEG	
Prochlorococcus marinus str. A	123968052	S-K-K-SSLN--Q-TS---EKFKAG	
Gloeobacter violaceus PCC 7421	37520946	--R-D-RSL---A---DLLA--QTG	

Supplementary Figure 128

Partial sequence alignment for the protein precorrin-4 C11-methyltransferase, showing a 1 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		302	361
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186686042	WGDLWLLTSVQGLPLLIVLSFLLIS PLSYSYSLQ TRLNSSLAPLLLLLGLNLFVIRFA
	'Nostoc azollae' 0708	298489924	-----SA-----I-L-AY-PM- --PYL-IS-V----I--LF----
	Nostoc sp. PCC 7120	17227639	-----SA-----TGY---N SWQ-ML-VK-----V--L-----
	Nodularia spumigena CCY9414	119512888	-----SA-----V---CW--- LNP-----V--L-----
	Raphidiopsis brookii D9	282896287	-----FC-----VI-A---VP P-PYF-AS---A--V--LI----
	Anabaena variabilis ATCC 29413	75907734	-----SA-----TGY---G SWQ-LL-VK-----V--L-----
	Cylindrospermopsis raciborskii CS-505	282901342	-----FC-----PVI-A---V- P-PYL-GS---A--V--LI----
	Fischerella sp. JSC-11	354567716	-T---V---T-A--V--I--Y--L- PHLPIT-I-H-----LIV--G
	Microcoleus vaginatus FGP-2	334119540	-----LA-----VP---CY--FG S-PPVGLF-HFW-----V---AV-S-
	Oscillatoria sp. PCC 6506	300866875	-----LA-----IP-L--Y-FL- LPF-LS-S-S-F-----LA--V-
	Microcoleus chthonoplastes PCC 7420	254415028	-S-V--L-T-A---PVSIAV-VSL SLGISSPVH-PI-----G-L-M-W-
	Arthrospira maxima CS-328	209527264	-A--GF-LCC----FVLTIG-ALG L-SPHHNPFTC-WR--A--FI--V-
	Arthrospira sp. PCC 8005	376003785	-A--GF-LCC----FVLTIGFALG F-FPYHNPFTW-WR--A--FI--V-
	Arthrospira platensis NIES-39	291569218	-A--A--FCC----I-VLTIA-AFG F-FPYHNPFTW-WR--A--FI--V-
	Arthrospira platensis Paraca	284054903	-A--A--FCC----I-VLTIA-ALG RF-PL-N-L V-I-----WLTV---
	Lyngbya sp. PCC 8106	119484670	---GV-LAI-----P-LVLS-GCW G-GY-F-S -K-A-----L----
	Cyanothece sp. PCC 7822	307151823	---G--ICT-----LSVGSICL E EGLSNPF-SGW-D--V--LLV--G
	Other Cyanobacteria	Synechococcus sp. PCC 7002	170078635
Thermosynechococcus elongatus BP-1		22299562	---AYF-WA-----PLFVLG-VAL WGEDDFDLT-R-AIA--G--LLV-WG
Synechocystis sp. PCC 6803		16329636	-AEVIF-LL----I-F F-L--ML LSSLSIQ-AW----C--G--LG
Cyanothece sp. PCC 7425		220907916	----GF-LA-----PMS-CLG-AG -AGNHFLS-E-AIA--I--I----
Cyanothece sp. ATCC 51142		172039660	----I--LLT-A---FLLIV-VSFW SIGYHFLS-K-AIA--IV--I--S
Cyanothece sp. CCY0110		126657356	----A--LLT-A---LLIV--SFW IFGYDFLS-KIV-T--L-LL----
Microcystis aeruginosa PCC 7806		159026444	----A--ALT----II-TPLL-TAL -VGEHFL-L-Q-AIA--IL--I----
Crocospaera watsonii WH 0003		357266526	---FT--LLT-A---LLIVLASLW -VGEHFL-L-Q-AIA--IL--I----
Crocospaera watsonii WH 8501		67920231	---FT--LLT-A---LLIVLASLW -VGEHFL-L-Q-AIA--IL--I----
Cyanothece sp. PCC 7424		218440042	----G--LLT-----LSIALG-CL GLGYAFLT-K-A-----C--L----
Microcystis aeruginosa NIES-843		166368830	----A--ALT----II-TPLL-TAL IFGYNFLS-KIV-T--L-LL----
Synechococcus sp. JA-3-3Ab		86605657	---MA--VLT--I---A-GGLGIKA G-GFPFSLTEQV-----G-LAL-LG
Synechococcus sp. JA-2-3B'a(2-13)		86607677	---IA--VLT--I-GPV-VGL-GKA GMGFPFSLTEQV-----G-LIL-LG

Supplementary Figure 129

Partial sequence alignment for a glycosyl transferase family protein, showing an 8 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		192	227	
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186685192	VLSIPKDTMIAIPEKG EKG LTKVSDANAKGGPVLA	
	Nodularia spumigena CCY9414	119509636	-----R-----Q-----	
	Aphanizomenon sp. 10E6	295293761	L----RG-----Q-	
	Anabaena variabilis ATCC 29413	75908891	-----R-----	
	Nostoc sp. PCC 7120	17227683	-----R-----	
	'Nostoc azollae' 0708	298490835	-----RG---SL--Q-	
	Raphidiopsis brookii D9	282898388	-----G--VSL--D-	
	Cylindrospermopsis raciborskii CS-505	282899500	-----G--VSL--D-	
	Fischerella sp. JSC-11	1134567998	M----R--RVE-EGH-	
	Microcoleus chthonoplastes PCC 7420	254412010	M----R--QVEL--V-	
	Oscillatoria sp. PCC 6506	300869394	L----R--QVEL-NI-	
	Moorea producta 3L	332712341	M----R--R-E--GV-	
	Trichodesmium erythraeum IMS101	113475525	L--V-R--KVQ--GI-	
	Microcoleus vaginatus FGP-2	334119380	L--V-R--KVEV-GMS	
	Lyngbya sp. PCC 8106	119493916	--V-R--RTYVEGM-	
	Arthrospira maxima CS-328	209524282	-----R--RTFVKGR-	
	Arthrospira platensis NIES-39	291571670	-----R--RTFVKGR-	
	Arthrospira sp. PCC 8005	376003500	-----R--RTFVKGR-	
	Arthrospira platensis Paraca	284052728	-----R--RTFVKGR-	
Other Cyanobacteria	Cyanothece sp. PCC 7424	218439895	M----R--SRVE--GI-	
	Cyanothece sp. PCC 8801	218247317	M----R--SRVE--GV-	
	Cyanothece sp. PCC 8802	257061348	M----R--SRVE--GV-	
	Cyanothece sp. PCC 7822	307153778	M----R--SRVE--GV-	
	Microcystis aeruginosa NIES-843	166365076	M----R--SRVR--GH-	
	Cyanothece sp. ATCC 51142	172037120	M--V-R--SRVN--GR-	
	Cyanothece sp. ATCC 51472	354555884	M--V-R--SRVN--GR-	
	Microcystis aeruginosa PCC 7806	159025959	M----R--SRVR--GH-	
	Cyanothece sp. CCY0110	126658077	M--V-R--SRVN--SV-	
	Synechocystis sp. PCC 6803	16330199	L----R--RVF--DV-	
	Synechococcus elongatus PCC 6301	56750203	L----R--QV---GY-	
	Crocospaera watsonii WH 0003	357265428	M----R--SRVN--GV-	
	Crocospaera watsonii WH 8501	67921247	M----R--SRVN--GV-	
	Acaryochloris marina MBIC11017	158337663	I----R--RVS--DV-	
	Cyanothece sp. PCC 7425	220909223	--T--R--RVE--GY-	
	Acaryochloris sp. CCME 5410	359459152	I----R--RVS--DV-	
	Synechococcus sp. PCC 7335	254423552	-M--R--RVE--DY-	
	Synechococcus sp. PCC 7002	170076990	M----R--SQVR--NV-	
	Thermosynechococcus elongatus BP-1	22299301	-----R--RVE--GY-	
	Synechococcus sp. JA-3-3Ab	86605772	-----R--RVRL-SF-	
	Synechococcus sp. JA-2-3B'a(2-13)	86608480	-----R--RVRL-GF-	
	Synechococcus elongatus PCC 7942	81299770	-V-L-R--RTEL-GY-	
	cyanobacterium UCYN-A	284929255	I----R--QAI-KHEI	
	Gloeobacter violaceus PCC 7421	37523048	---V-R--RTD--GF-	
				-----E-----
				-----E--Q-----
				-----E-----
				-----E--Q-----
				-----E-----
				-S-I-E-----
			-S-I-E--T-----	
			IK-INS--VQ---A-S-	
			-A-IN---VE--AA---	
			Q--I-E---Q-QA-T-	
			II-IN---V---ST-T-	
			-A-INE--LY---K--K	
			FS-I-E---T--AA--T	
			M--IN---YY---A-S-	
			M--IN---YY---A-S-	
			M--IN---YY---A-S-	
			M--IN---YY---A-S-	
			FS-IN---VQ---A---	
			Y--IN---IH---A---	
			Y--IN---IH---A---	
			Y--IN---VH---A---	
			F--IN---VH---T---	
			Y--IN---VH---G---	
			Y--IN---VH---G---	
			F--IN---VH---T---	
			Y--IN---IH---G---	
			Y--IN---Y---Q---	
			T--INA--VF---A---	
			YR-IN---VH---AS---	
			YR-IN---VH---AS---	
			V--INS--WW---D-VT	
			ME-INA---E--AN--M	
			V--INS--WW---D-VT	
			FS-INE---E--AE---	
			I--INE---Y--GQ---	
			TS-INA--VY-DVE-TV	
			V--INA--LV---A-T-	
			VA-INA--LA---A-T-	
			ID-INA--VY-----	
			R-INE--YY---S---	
			AH-INV--VY---A---	

Supplementary Figure 130

Partial sequence alignment for the protein cell envelope-related transcriptional attenuator, showing a 3 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

Nostocales/ Stigonematales	}	Nostoc punctiforme PCC 73102	186680640	356	D NNVPSIGNEGISDENPLKTIGLRLGQVKEAKKELAF	417
		Nodularia spumigena CCY9414	119512207		NADLANDLGNLLNRTLSMVKKYCV	
		'Nostoc azollae' 0708	298492950		-----N-----S	-DGLT----A-N-----A-----G-----QS-QA---
		Cylindrospermopsis raciborskii CS-505	282900792		-----N-----ST	GK---DP-A-PN--A--A--G-----Q--TA---
		Raphidiopsis brookii D9	282896532		-----N-----A	GEI---APQD-PP--T-----GQ---K--Q--SV---
		Anabaena variabilis ATCC 29413	75908935		-----N-----A	GEI---APQD-PP--T-----GQ---K--Q--SV---
		Nostoc sp. PCC 7120	17227729		-----N-----A	-YAL--T--D-PA--T--AL-VD--AK--Q--EV---
		Fischerella sp. JSC-11	354567131		-----G-----A	-YAL--T--D-PA--T--AL-VD--AK--Q--EV---
		Moorea producta 3L	332707288		-----G-A---Q	G---P-TS-D-PNDS---S-----N--EA---
		Oscillatoria sp. PCC 6506	300867846		-----V-----AH---G	GI--GLTGS--QV---LE---D--S--AR--ES---
		Arthrospira platensis Paraca	284052455		-----MK-AA--FD	GLI-NVNA-D-APD---E---T-SDRTAQ--EN---
		Lyngbya sp. PCC 8106	119487032		-----R-AH--FE	GR--A-AVDN--AD-H--S--SS--NT-SQF-EK---
		Microcoleus vaginatus FGP-2	334117482		-E-----N-AR--G	GRI-HVSGAV-PED---SY-VT--QT-AQG-ER---
		Arthrospira sp. PCC 8005	376001331		-----MK-AA--FE	GS--NVLG-N-DGD-L--AMS-D--T--AGF-E---
		Other Cyanobacteria	}	Trichodesmium erythraeum IMS101	113475475	
Crocospaera watsonii WH 8501	67921186				-----G-T---K	GQI-K-NVQEVVTH---K--MG--K-TQ--TS-D---
Cyanothece sp. CCY0110	126656215				-----G-TQ---Q	GSL-ELTAELEPQ-H-----T--K-IK--ET-K---
Synechocystis sp. PCC 6803	16331573				-----G-----Q	GSL-E-TASELPQDH-----S--K-IK--EK-K---
Cyanothece sp. PCC 7424	218437104				-----G-Q---N	GQG-QVMATDLAPD---AL-SH--E-SS--ER-S---
Cyanothece sp. PCC 8801	218248716				-----G-I---Q	G-G-QLTSDSLDEN---I--QQ-S-R-S-SMES-Q---
Cyanothece sp. ATCC 51142	172036861				-----G-TQ---Q	GVI-PLII-ELAL-HS--S--E--DR-TQ--ET-R---
Synechococcus elongatus PCC 6301	56750417				-----A-R--Q	GSL-KMTATELEPQ-HS-----T--K-IN--ET-K---
Acaryochloris marina MBIC11017	158335466				-----S-----K-AV--N	LQI-LGSTA--AAD---RQ--EA-TDS--S--DR---
Synechococcus sp. JA-2-3B'a(2-13)	86609674				-----N-LR--QS	GQ--D-DIQ--PEDH--AQL-RKA-D--AK--QN-N---
Acaryochloris sp. CCME 5410	359457323				-----S-----K-AV--H	DYRVPVAV--PADH---LAE--APA-AD--ER---
Microcystis aeruginosa NIES-843	166367026				-----G-----Q	GQ--D-DIQ--VPEDH--VQL-RTA-D--AK--QN-N---
Cyanothece sp. PCC 7822	307151236				-----G-Q---K	-IP-QLTGVD-PND-R--NM-IN--GI-EQK-DA-Q---
cyanobacterium UCYN-A	284929295				-----D-----G-TQ---E	GKG-ALKATE-DNN----I--EG-S-R-R-CLEA-K---
Microcystis aeruginosa PCC 7806	159029236				-----G-----Q	GLL-FMSGIDLPK-H--K--N--N-IK--DN-K---
Synechococcus sp. JA-3-3Ab	86606068				-----N-L---QP	-IP-QLTGVD-PND-R--NM-TN--VI-EQK-DA-Q---
Synechococcus sp. PCC 7002	170079299				-----G-L---K	DYR-PAVE-PADH---LAE--APT-AD--ER---
Cyanothece sp. PCC 7425	220908709				-----D-----K-AF--Q	SEI-NLDLGT-PVDH---AL-ET--DRTC--DN-N---
Thermosynechococcus elongatus BP-1	22298292				-----K-AW--TD	GQI-TGTADSLPADH--QAL-QSV--R-AQS-A--D---
Synechococcus sp. WH 5701	87303680				-N---TI-----S--ARRWFD	GK--NVQGA--PR-H--RQLAEH-SQTYGQG-RQ---
Synechococcus sp. RS9916	116074880				-N---TI-----A--AR-WFG	QA--APS--AVGSSH--AQAA-LAR---I-GMER-D---
Synechococcus sp. CC9605	78212864				-N---TI-----S--AR-WFD	DA--QA--TA-ATDH--RQHAEEAIG--RD-MPDM---
Synechococcus sp. WH 7805	88808698				-N---TI-----A--SR-WFD	DG--PA--AASEADH--ALKAAETIST-M--MPO---
Synechococcus sp. WH 7803	148239464				-N---TI-----A--SR-WFE	ASL-PVS AAVRSDH---DTTE-TIDR-RQSIPAF---
Synechococcus sp. WH 8109	260435477				-N---TI-----S--AR-WFD	ASL-PVS PDVRS-H---QTEQTI-R-RQSIPA---
Cyanobium sp. PCC 7001	254431009		-N---TI-----S--AR-WFD	DV--PA--AAS-ADH--ALKAAETVST-I--MPQ---		
Prochlorococcus marinus CCMP1375	33240420		-N---TI-----S--SR-WFS	GA--PA--ASAGDH--AQATRLA---ACAGLER-D---		
Gloeobacter violaceus PCC 7421	37522592		-----GN--V---G	-C--N-HK-NENN-L--LLASKTISD-INY-EQ-S---		
				GVI--V A-DPTDA-RDKATG-AQVCA-HWEA-R---		

Supplementary Figure 131

Partial sequence alignment for the protein methionyl-tRNA synthetase, showing a 1 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

			89	135
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186684135	GIETI GEMELAWRNLR WRSQPQA SLPWVGITGTNGKTTTTTALIAAIF	
	Anabaena variabilis ATCC 29413	75907016	-----H-K	
	Nostoc sp. PCC 7120	17231604	-----H-K	
	Nodularia spumigena CCY9414	119509795	-----Y-Q	AS-----
	'Nostoc azollae' 0708	298491728	---I---H-Q	NI--A-----S-T---
	Fischerella sp. JSC-11	354566385	-----Y--	-I--A-----S--S-
	Synechococcus sp. JA-2-3B'a(2-13)	86607791	--PVV --A---IY-D	H-----S----V-EM-
	Synechococcus sp. JA-3-3Ab	86605662	---VV --A---PF-E	H-----S----V-EM-
	Synechococcus sp. PCC 7335	254424301	---L-----A--	DR-----V----
	Cyanothece sp. CCY0110	126660928	K-D-- --L----Y-Q	-S-----V----
	Synechococcus sp. WH 8109	260435255	--KVR ---D---DA-Q	QI--I-----V-H-LSHVL
	Synechococcus sp. WH 8102	33865066	--VAVD ---AV--DA-K	HI-----V-H-LSHVL
	Crocospaera watsonii WH 0003	357262566	N-D-- --L----Y-N	-S-----V----
	Crocospaera watsonii WH 8501	67923621	N-D-- --L----Y-N	-S-----V----
	Synechococcus sp. CC9902	78184101	-VQIR --I----EA-N	DI-----V-H-LSHVL
	Synechococcus sp. PCC 7002	170077472	--D-- --L----D	QK-----V-E---
	Synechococcus sp. WH 8016	352095857	-VTVR ---AV--QA--	HC--I-V-----V-H-L
Synechococcus sp. BL107	116071267	-VQVR --I---DA-N	DI-----V-H-LSHVL	
Synechococcus sp. CC9605	78213668	-VQVR ---DT--DA-Q	QI--I-----V-H-LSHVL	
Cyanothece sp. PCC 7424	218442054	---L-----Y-N	GCS-----V----	
Synechococcus sp. RCC307	148242946	-CRVV --V-V--QAMA	GI--I-----V-S-V-HLL	
Cyanothece sp. ATCC 51142	172035156	Q-D-- --L----Y-Q	-S-----V----	
Synechococcus sp. CC9311	113955473	-VTVR ---AV--QA-C	KC--I-----V-H-L	
Synechococcus sp. RS9916	116073647	D-AVC ---AV--DA-S	HR--I-----V-H-LQHVL	
Microcystis aeruginosa NIES-843	166363698	--D-- --L----Y-Q	-S--L-----C----	
Microcystis sp. T1-4	390439240	--D-- --L----Y-Q	-S--L-----C----	
Synechococcus sp. CB0101	318042232	--R-D --ITT--EASA	GV--I-----V-H--HLL	
Synechococcus sp. WH 5701	87301212	-VRVC ---SV--EASN	EI--I-----V-H-LSH-L	
Cyanothece sp. PCC 7822	307154774	--D-- --L----S-K	PSQ--V-----VT---	
Synechococcus sp. CB0205	317969026	--P-N --IST--EASG	-T--I-----V-H-V-HLL	
Synechococcus sp. RS9917	87123641	-VAV- --A--QF-N	RT--I-----V-Q-LQHVL	
Cyanobium sp. PCC 7001	254430906	-VRVQ ---VP--EAS-	PV--I-----V-H-VHHLL	
Cyanothece sp. PCC 8801	218248504	--DV- --L----S-Q	-S-----V----	
Synechocystis sp. PCC 6803	16330305	-V-VT --L---QY-H	AV-----S-VQ---	
Synechococcus elongatus PCC 6301	56752431	LQIETL--L-----T-S	DI--AV-----	
Synechococcus sp. WH 7803	148240320	-LRVR ---SV--DA-S	GI--I-----V-H-LQHVL	
Acaryochloris marina MBIC11017	158334692	---F --L---S-K	QY-----IVE---	
Synechococcus sp. WH 7805	88807416	-FRVR ---SV--DA--	GI--I-----V-H-LQHVL	
Acaryochloris sp. CCMEE 5410	359463258	---F --L---Y-K	QY-----IVE---	
Cyanothece sp. PCC 7425	220906241	-----I---H-Q	QF-----V--M-	
cyanobacterium UCYN-A	284928909	KVSI- --L--S--Y-N	LS--LI-----K-VES--	
Oscillatoria sp. PCC 6506	300868200	-----V---H-Q	-Q--I-----	
Microcoleus chthonoplastes PCC 7420	254414291	--D-- -----Y-K	-S-----	
Moorea producta 3L	332707401	-----I---Y-Q	-C--AV-----S----	
Arthrospira platensis Paraca	284051225	-----Q	DI-----K--	
Arthrospira platensis NIES-39	291569849	-----Q	DI-----K--	
Arthrospira maxima CS-328	209525298	-----Q	DI-----K--	
Microcoleus vaginatus FGP-2	334120051	-----A---Y-K	S-----	
Lyngbya sp. PCC 8106	119487890	---IV -----N	AV-----	
Prochlorococcus marinus MIT 9313	33863701	--DID ---AV---S-S	H--IA-----V-H-LNHVL	
Prochlorococcus marinus CCMP1375	33240888	--IVE S-IN---A-K	NT-----V-HML	
Prochlorococcus marinus AS9601	123969098	--RV- --INV--EI-K	DTN-I-----V-H-LSH-L	

Supplementary Figure 132

Partial sequence alignment for the protein UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase, showing a 7 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		108	151
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186683647	GPARSLLFVMPYGFMGVLLGATWYRR R VPWIVSITLGLTLLGLG
	'Nostoc azollae' 0708	298492012	--V--M-----F-L-----M-----
	Raphidiopsis brookii D9	282895674	--V-----I--FA-L-----S-H--F-----I-C---
	Cylindrospermopsis raciborskii CS-505	282900315	--V-----I--FA-L-----VS-H--V---S---I-C---
	Nodularia spumigena CCY9414	119512452	--V-----FA-----H-----
	Anabaena variabilis ATCC 29413	75909273	--V-----FA-----A-----
	Nostoc sp. PCC 7120	17227792	--V-----FA---V-----A-----
	Fischerella sp. JSC-11	354567780	--I-----FA-----A-----S--AV-----
	Microcoleus chthonoplastes PCC 7420	254413468	--T--I--LI---L---Q---L-K-QAN-LF--GI-----
	Trichodesmium erythraeum IMS101	113475248	--T--I---I-FAL---M---L-R-GKS-F---Y-S---SI-
	Microcoleus vaginatus FGP-2	334117938	--T--I-Y---F-I---A--GL-R-KAS-SL--SV-A-I--V-
	Moorea producta 3L	332706232	-VT--II-LV---L---Q---L-K--AS-LL--SM---I--I-
	Arthrospira platensis Paraca	284050780	--T--I-Y-I---W---V--GL-K-GAM-P-TVAI-S-I-AI-
	Lyngbya sp. PCC 8106	119490055	--T--II-LI---W---AM-GL-K-GAN-F---GV-AII-SI-
	Arthrospira sp. PCC 8005	376001833	--T--I-Y-I---W-A-V--GL-K-GAK-P-TVAIAS-I-AI-
	Arthrospira maxima CS-328	209522908	--T--I-Y-I---W-A-V--GL-K-GAK-P-TVAIAS-I-AI-
	Oscillatoria sp. PCC 6506	300868018	--T--I---I-F-LLS---GL-R-GAS-S---G--AIV-SF-
	Synechococcus sp. JA-2-3B'a(2-13)	86609748	--T--I-Y-I---LL-YWCARL-QQ-LS-Y--VVS-AV-SA--
	Cyanothece sp. PCC 7424	218439630	--T--IV-LI---V---Q---C-K-GA--QW-ML---II-VF-
	Cyanothece sp. PCC 7822	307153482	--T--IV-FI---V---Q---C-R-GAS-YL-ML---II-VF-
Synechococcus sp. PCC 7335	254423145	--T--V-----IL--Q---YL-R-KAG-GI--LT-G--V-F-	
Synechococcus sp. PCC 7002	170078396	--T--IIYTI---LI--Q--WM-R-Q-S-YW-V---I-----	
Microcystis aeruginosa PCC 7806	159026388	--T--IV-LI-H-L--IQ--LC-R-GAR-EF--LT-G---AF-	
Synechococcus elongatus PCC 6301	56751839	--T-GI--LL-P-CL----CC-Q--LS-GR--L--GG-E-I-	
Cyanothece sp. PCC 7425	220905783	--P---Q-LV-H-LL-I--HF-Q--AS-W---AW-A---GI-	
Acaryochloris sp. CCME 5410	359457479	--P--IQ-L---VV--M--GL-K-Q-S-AI-MGWSI-IMAI-	
Acaryochloris marina MBIC11017	158338243	--P--IQ-L---VV--I--GL-K-Q-S-AI-MGWSI-VMAI-	
Thermosynechococcus elongatus BP-1	22298945	--P---Q-LL-H-I---IF-GC-AKES-GR--LS-AVI--A-	
Crocospaera watsonii WH 0003	357263688	--T--IV-LI---L---Q---F-R-NL--EF-MIT-A----F-	
Crocospaera watsonii WH 8501	67922653	--T--IV-LI---L---Q---F-R-NL--EF-MIT-A----F-	
Cyanothece sp. ATCC 51142	172036337	--T--IV-LI---V---Q---F-R-G--EF-MIT-A----F-	
Cyanothece sp. CCY0110	126654855	--T--IV-LI---V---Q---F-R-GL--EF-MIT-A----F-	
Synechococcus sp. JA-3-3Ab	86605519	--T--I-Y-I---LL-YWCARL-R--LS-Y--VAS-AA-SAF-	
Cyanothece sp. PCC 8801	218245754	--T--IV-LI---L--IQ--YC-Q-GAN-GF-LVT-A-I--	
Cyanobium sp. PCC 7001	254431659	--I-GP-LLF---LLALW--WC-R--LN-WL-WS--SVI-AA-	
Microcystis aeruginosa NIES-843	190192140	--T--IV-LI-H-L--LQ--LC-R-GAR-EF--LT-G---A	
Synechocystis sp. PCC 6803	16330851	--T--VV---I---L-SLQ--FC-V-NAS-GF--FT-A-I--F-	
Synechococcus sp. RCC307	148241761	--V-GP-MLF---LLSLW--WC-R-QCS-WF-MAV-V-I-AA-	
Synechococcus sp. CB0205	317970523	--I-GP-VLF---LALW--WC-R-GLS-YL-WGV-V-I-AG-	
Synechococcus sp. RS9916	116073491	--V-GP-ILF---LPSLW--WC-L-GMS-WI-WG---I-AA-	
Synechococcus sp. WH 8016	352094842	--L-GP-MLF---MLSLW--WC-Q-K-S-WI-WG--VVI-AA-	
Synechococcus sp. CC9311	113954252	--L-GP-MLF---LLSLW--WC-HKKS-WI-WG--VVI-AA-	
Synechococcus sp. WH 5701	87303040	--I-GP-VLF---LLALW--WG-AE-RS-WLTSV-L-I-SA-	
Synechococcus sp. WH 7803	148240217	--V-GP-MLF---LLSLW--WC-QH-QS-WI-WG--VVI-AA-	
Synechococcus sp. RS9917	87123725	--V-GP-MLF---LLSLW--WA-S-SFS-WL-WGV-L---AI-	
Synechococcus sp. WH 8102	33866402	--V-GP-LLF---L-ALW--WS-GKGRS-WL-WSV-VV--SI-	
Prochlorococcus marinus MIT 9312	78778822	--V-GT--LF---ILAFW--WC-FKES-RL-L-I-VII----	
Prochlorococcus marinus AS9601	123968030	--V-GT--LF---ILAFW--WC-FKES-KL-L-G-VVI----	
Prochlorococcus marinus CCMP1986	33860995	--I-GT--LF---ILAFW--WS-FMKKN-WL-WCG-FI----	

Supplementary Figure 133

Partial sequence alignment for the hypothetical protein Npun_F3493, showing a 1 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		34	80
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186680613	AEDFGGPQYEGCNEYLVHTKPEA EA VAKVHRDFLAAGADVIEDTDFG
	Nodularia spumigena CCY9414	119511447	-----N-----
	Raphidiopsis brookii D9	282895748	-----E-----I-----
	Cylindrospermopsis raciborskii CS-505	282898760	-----E-----I-----
	Anabaena variabilis ATCC 29413	75909261	-----V-----
	Nostoc sp. PCC 7120	17227804	-----V-----
	'Nostoc azollae' 0708	298490772	-----A-----S-----
	Fischerella sp. JSC-11	354565463	-----E-----
	Cyanothece sp. PCC 7425	220908876	-----E-----
	Cyanothece sp. PCC 8802	257061375	-----E-----
	Cyanothece sp. CCY0110	126660254	-----E-----
	Cyanothece sp. PCC 8801	218247289	-----E-----
	Cyanothece sp. PCC 7822	307150703	-----E-----
	Acaryochloris sp. CCMEE 5410	359463590	-----E-----IQ--D-
	Cyanothece sp. ATCC 51142	172034994	-D-----E-----
Other Cyanobacteria	Crocospaera watsonii WH 8501	67922204	-----E-----I-----
	Crocospaera watsonii WH 0003	357264218	-----E-----I-----
	Cyanothece sp. PCC 7424	218440169	-----L-----
	Acaryochloris marina MBIC11017	158334435	-----E-----IQ--G-
	Synechococcus sp. PCC 7335	254424469	-A-----E-----E-----
	Microcystis aeruginosa PCC 7806	159029504	-----AE-----S-----
	Microcystis aeruginosa NIES-843	166364879	-----AE-----S-----
	Thermosynechococcus elongatus	22298571	-----KE-----ISQ--P
	Synechococcus elongatus PCC 7942	81300181	-----ET-----W-IR----
	Synechococcus sp. PCC 7002	170079061	-A-----AE-----Q-----
	Synechococcus elongatus PCC 6301	56750193	-----ET-----W-IR----
	Synechocystis sp. PCC 6803	16331640	-A-----AE-----
	Synechococcus sp. JA-3-3Ab	86607035	-A-----EL-----I--LS----
	Synechococcus sp. JA-2-3B'a(2-13)	86609654	-A-----EL-----M--L----
	Cyanobium sp. PCC 7001	254431763	-----AAL-----VSR-D-
	Synechococcus sp. WH 8102	33865772	-----EL-----N-AV-----
	Synechococcus sp. RS9917	87124476	-A-----EL-----N--V-R-D-
	Synechococcus sp. CC9605	78212885	-----DL-----N-AV--D-
	Synechococcus sp. BL107	116070564	-D-----DL-----N--I-R-D-
	Synechococcus sp. RCC307	148242308	D-----AAL-----N--V-R-D-
	Synechococcus sp. WH 8109	260436564	-D-----DL-----N-AV--D-
	Synechococcus sp. CC9311	113955505	-----EAL-----N--V-R-D-
	Synechococcus sp. CB0205	317969871	-D-----LAL-----N--F-R-D-
	Synechococcus sp. WH 8016	352093917	-D-----EAL-----N--V-R-D-
	Synechococcus sp. CC9902	78184696	-----LDL-----N--I-R-D-
	Synechococcus sp. WH 7803	148239613	-----EAL-----N--V-R-D-
	Synechococcus sp. CB0101	318041425	-----AAL-----N--F-R-D-
	Synechococcus sp. WH 5701	87302677	-----AAL-----N--F-R-D-
	Synechococcus sp. RS9916	116074862	-A-----EEL-----N--V-R-D-
	Synechococcus sp. WH 7805	88808540	-D-----EAL-----N--I-R-D-
	Trichodesmium erythraeum IMS101	113474854	-----KE-----D-----
	Moorea producta 3L	332707971	-----E-----
	Oscillatoria sp. PCC 6506	300865649	-----EY-N-----
	Arthrospira maxima CS-328	209523868	-----KE-----
	Arthrospira platensis NIES-39	291569987	-----KE-----
Arthrospira sp. PCC 8005	376005437	-----KE-----	
Arthrospira platensis	290771324	-----KE-----	
Arthrospira platensis Paraca	284052062	-----KE-----	
Microcoleus vaginatus FGP-2	334120138	-----KE-----	
Microcoleus chthonoplastes PCC 7420	254415779	-----AK-----	
Lyngbya sp. PCC 8106	119493973	-----KE-----	
Prochlorococcus marinus CCMP1375	159903428	-----S-F-----LISN-KS	
Prochlorococcus marinus NATL2A	72382173	-D-Y--ASL-----N--L-SVSS	
Prochlorococcus marinus AS9601	123968517	D-----KEL-----N--LSS--V	
Gloeobacter violaceus PCC 7421	37520046	-----AHL-----FS----	

Supplementary Figure 134

Partial sequence alignment for the protein methionine synthase, showing a 2 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		219	269	
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186681282	PKGIRGLLFRQILPGMR RRGASAGS SLLGCLREVAQARNKMTMSQVAINWC	
	Nodularia spumigena CCY9414	119508846	-----L---V-	
	Anabaena variabilis ATCC 29413	75909269	-R-V-----L---V-	
	'Nostoc azollae' 0708	298491921	---V--I-----IK	
	Nostoc sp. PCC 7120	17227796	-R-V-----L---V-	
	Cylindrospermopsis raciborskii CS-505	282898556	-----I---M---IK	
	Raphidiopsis brookii D9	282896230	-----I---M---IK	
	Fischerella sp. JSC-11	354564636	-----L---V-	
	Lyngbya sp. PCC 8106	119486839	-----V-----VK	
	Microcoleus vaginatus FGP-2	334121131	-----ACK-L----	
	Trichodesmium erythraeum IMS101	113475801	-----I--K-----I-	
	Microcoleus chthonoplastes PCC 7420	254411837	---L-R---K-L---IQ	
	Cyanothece sp. PCC 7822	307152159	-----KRL---I-	
	Cyanothece sp. PCC 8801	218244911	---L--I-C--L--KIK	
	Cyanothece sp. PCC 8802	257057936	---L--I-C--L--KIK	
Cyanothece sp. PCC 7425	220910486	-----F---L---Q		
Cyanothece sp. PCC 7424	218437204	---V--I--KRL---I-		
Crocospaera watsonii WH 8501	67921562	---F--IV-K-M--KI-		
Crocospaera watsonii WH 0003	357264942	---F--IV-K-M--KI-		
Cyanothece sp. ATCC 51142	172037425	-R-F---F--M--KI-		
Cyanothece sp. CCY0110	126654819	-G-F---F--M--KI-		
Other Cyanobacteria	Microcystis aeruginosa PCC 7806	159026972	-S-L-HF--G-LI-AIS	
	Synechocystis sp. PCC 6803	16330597	-P-----RL--KIQ	
	Synechococcus sp. PCC 7335	254421749	---L-RFALG---DVQ	
	Synechococcus sp. PCC 7002	170077364	---L-RRV-K-V--N-A	
	Synechococcus sp. RCC307	148243173	-R-P--A---RL--ALK	
	Synechococcus sp. WH 5701	87300949	---P-----RLQ-RIE	
	Synechococcus sp. CB0205	317968769	-P-T--S---RLESSLA	
	Synechococcus sp. WH 8016	352096559	ATML-SR--QRL--ASE	
	Synechococcus sp. CC9311	113953419	STVL-SR--QRL--ASE	
	Cyanobium sp. PCC 7001	254430457	-A-P--A---RLT-QLA	
	Synechococcus sp. WH 7805	88807847	LTPL--G---RL--ASE	
	Synechococcus sp. WH 7803	148238806	LTPL-TG---RL--ASE	
	Synechococcus sp. RS9917	87124924	CTRL-QG---RL---SE	
	Synechococcus sp. BL107	116072604	STWV-QR---RL--DSL	
	Synechococcus sp. CB0101	318042609	LT-T-RS--RLE-Q-H	
	Prochlorococcus marinus CCMP1375	33240642	ETFL-QRI-NRLI-SSI	
				---E-----AS-----L---
				---Q--QD--S-----L---
				GI--T---IGDT-----L---
				---Q--QD--S-----L---
			G---T-E-I-NS-G---G-----	
			G---T-E-I-NS-G---G-----	
			---A---I--F---S-----L---	
			P--S--K-ISDS-----VK	
			---E---I-AFK---V-----	
			P--L--EA--KS-----L---	
			P-----I-SSKE---A-----	
			P--Q--QA---Y-G--L---L---	
			A--D--Q-I-SL-----L---	
			A--D--Q-I-SL-----L---	
			P--N--SAI-T--Q--IA-I-L---	
			P--Q--EA---H-E---A---L---	
			P--QG-E-I-KF-E---A-----	
			P--QG-E-I-KF-E---A-----	
			P--QG-EDI-TQ-Q---A-----	
			P-IQG-E-I-TQ-Q---A-----	
			P-MS--EVI---K---I---L---	
			P--DT-QAI-RE-E---A---L---	
			P--N---AI-K-NE---A-T---	
			P-QS--Q-LMALHQ---A---L---	
			P-QALMA-L-ACHQISPAAL-LS--	
			P--ALM-NI-SEHGAQV-E--L---	
			P-RQTME-I-NG-PGGLGA--L---	
			G-RAVMV-I-AS-SAS-A---L---	
			D-RSVMV-I-SQ-RAS-A---L---	
			P--AAAMEAM-RP-STPLAA--L---	
			--RQTMVSI-E--GV-QV---L---	
			--RQTMVSI-E--GVSQV---L---	
			A-RQG-AAI-V-HGVSQA---L---	
			QIRQL-AAI--E-GVS-V---L---	
			Q-RQRME-I-AG-PGGLAA--L---	
			D-RKG-HKI-ND-GVSQT---L---	

Supplementary Figure 135

Partial sequence alignment for the protein aldo/keto reductase, showing a 9 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

			127		183
	Nostoc punctiforme PCC 73102	186680838	LPPGLLGRSQRRLVRGGRAAGS SYLLA	STNPSAFLTPTARQVWELLQTQPGDYS	
Nostocales/ Stigonematales	Nodularia spumigena CCY9414	119509983	-----L-DAEINRQ	ISSDLT--S-A-Q--LN---S-A----	
	Anabaena variabilis ATCC 29413	75907420	-----V--IKRQNEQL-	V--S-P-Q-ILQ---S-S----	
	Nostoc sp. PCC 7120	17231740	-----V--IKRENEQLF	VS--S-P-Q-ILQ---S-S----	
	'Nostoc azollae' 0708	298492421	-----L--KPENIPN-A	ENFLRSTA-L-ILKI--S-QD----	
	Raphidiopsis brookii D9	282895664	----I-E--S-L--K-ENIPP-G	EEFLGSQA--EILKV--S--D----	
	Cylindrospermopsis raciborskii CS-505	282900305	----I-E--S-L--K-ENIPP-G	EEFLRSQA--EILKI--S--D----	
	Fischerella sp. JSC-11	354565056	-----S-L--K-ENIPN-A	NE--SLP--EIIKI--S-AD----	
	Microcoleus vaginatus FGP-2	334117297	-----R-AV----IKDAISPNA	ET--N-A-N-ILQ---A-KH---T	
	Microcoleus chthonoplastes PCC 7420	254415732	-----M-----V--H-EALPD-A	DT-VGR----ILA--DA--E----	
	Moorea producta 3L	332708972	-----NSSLALPKEA	HVFINNSA-S-IL-I--AK-N----	
	Arthrospira platensis Paraca	284050398	-----NDAIPG-A	EE--Q-K-QRILQV-KQSAT---T	
	Arthrospira platensis NIES-39	291565766	-----NDAIPG-A	EE--Q-K-QRILQV-KQSAT---T	
	Arthrospira maxima CS-328	209525471	-----R-----TNAIPD-A	EE--Q-K-QRILQ--RQSAT---T	
	Arthrospira sp. PCC 8005	376006976	-----R-----TNAIPD-A	EE--Q-K-QRILQ--RQSAT---T	
	Other Cyanobacteria	Trichodesmium erythraeum IMS101	113476495	-----K-G-----IIEAAIPT-G	EI--NAN-QKIL-I--NSKTK--T
Lyngbya sp. PCC 8106		119493499	--T--R-----FVDKIPS-V	EN--NSA-LKIVN---AKKSN-L-	
Microcystis aeruginosa		1769412	--T--RS-----KPAAIIP-A	EV-C--LS--ILS--KN-KE----	
Microcystis sp. T1-4		390437728	--T--RS-----KPAAIIP-A	EV-C--LS--ILS--KN-KE----	
Microcystis aeruginosa NIES-843		166368030	--K--RS-----KPEAIIP-G	EV-C--LS--ILS--KN-KE----	
Synechocystis sp. PCC 6803		16331088	-----Q-----K-NGDRLPSDW	PLFL-QPSHQG---ILT---SSKE----	
Microcystis aeruginosa PCC 9443		389732112	--K--RS-----KPEAIIP-A	EV-C--LS--ILS--KN-KE----	
Cyanothece sp. PCC 7425		220906008	-----Q-----K-L-SDIP-DM	S-V-L-LLQV-H RGK---A	
Cyanothece sp. PCC 7424		218440062	-----EK-----L--KTEALPK-A	E--CSLS-Q--LS--NS-KE----	
Cyanothece sp. PCC 7822		307153455	-----EK-----L--NTEVLPE-A	E--CSR-P-QE-LK--KS-KD----	
Cyanothece sp. PCC 8802		257058078	-----L--KPDTIPP-A	ET-CS-V-G-LLR--RS-KE----	
Cyanothece sp. PCC 8801		218245054	-----L--KPETIPP-A	ET-CS-V-G-LLR--RS-KE----	
Cyanothece sp. ATCC 51142		172037442	---N--SK-----KNENIPPNA	EF-CR-A-A-ILS--KA-KD-N--	
Thermosynechococcus elongatus		22297958	---V-----V--RPPQGIPPL	SEQGQHLLRF---KGS----	
Crocospaera watsonii WH 8501		67923222	---K--SK--Q---KIDNIPP-A	ET-CS-A-ARILT---E-KD----	
Synechococcus sp. PCC 7002		170076993	P-GL-RRSQP-VRL-KTDLASIDI	D--P---Q-LFQ---RGSTQ---	
Crocospaera watsonii WH 0003		357263029	---K--SK--Q---KIDNIPP-A	ET-CP-A-ARILT---E-KD----	
Cyanothece sp. CCY0110		126658851	-----FTK-----RQDNIPP-A	EM-CRTA-AKILS--Q-KD----	
Synechococcus sp. PCC 7335		254423305	-----S-T--V--CQSISETV	KLSGYQSKLS-SAL--IETLNSET	
Acaryochloris marina MBIC11017		158341273	-----Q-----HTSINSVDV	HG-GTA-QDILT---QS-TQ--T	
Synechococcus sp. JA-3-3Ab		86606133	-----S-----GSPPGEC-	-A-QALLR--REK G-EL-	
Acaryochloris marina MBIC11017		158333512	-----Q-----NTTLHPHDI	NG-GAA-QGILA---QS-SQ--T	
Synechococcus sp. JA-2-3B'a(2-13)		86609671	-----S-----V--L-LPAEGEE	CS-A-QTLLQ---AK G-EL-	
cyanobacterium UCYN-A		284929135	---R--K-----ITEVVNYET	KD-HNSSIKILD---KTKSKS--	
Synechococcus elongatus PCC 6301		56750886	-----R-----ARS-TQPILR	SGA-QRL-P---KNLETG-	

Supplementary Figure 136

Partial sequence alignment for the primosome assembly protein, showing a 5 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		104	141
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186684082	DGQIIETVGIPT FAERGE GP KARLTVCVSTQVGC PMAC
	Nodularia spumigena CCY9414	119510469	----- AK-----
	Anabaena variabilis ATCC 29413	75909016	--E-V----- DK-----
	Nostoc sp. PCC 7120	17227851	--E-V-A---- DK-----
	'Nostoc azollae' 0708	298490256	--E-----S HK-----
	Raphidiopsis brookii D9	282895427	--E-----S DK-----
	Cylindrospermopsis raciborskii CS-505	282900428	--E-----S DK-----
	Synechococcus sp. JA-2-3B'a(2-13)	86608506	---L----- AK-----S-----
	Fischerella sp. JSC-11	354565937	--F----- EK-----
	Trichodesmium erythraeum IMS101	113476505	--N----- YK-----S-----
	Lynghya sp. PCC 8106	119488056	--NN----- EK-----S-----
	Oscillatoria sp. PCC 6506	300866402	----V-A---- EK-----S-----
	Moorea producta 3L	332708170	----- PK-----S-----
	Microcoleus chthonoplastes PCC 7420	254415410	--N----- AN-----S-----
	Microcoleus vaginatus FGP-2	334120777	----- DK-----A-----A-G-
	Arthrospira sp. PCC 8005	376003036	----- HD-----S-----
	Arthrospira platensis Paraca	284051696	----- HD-----S-----
	Arthrospira maxima CS-328	209526448	----- HD-----S-----
	Acaryochloris sp. CCMEE 5410	359460600	----- QK-----S-----G-
	Cyanothece sp. PCC 7425	220905656	----- NR-----S-----G-
Cyanothece sp. PCC 7822	307153286	--L----- EK-----S-----D-	
Acaryochloris marina MBIC11017	158337109	----- QK-----S-----G-	
Cyanothece sp. PCC 7424	218438587	--L----- EK-----S-----D-	
Microcystis aeruginosa NIES-843	166363996	--L---A----S EK-----S-----	
Microcystis aeruginosa PCC 7806	159025948	--L---A----S EK-----S-----	
Cyanothece sp. PCC 8801	218246284	--L----- AK-----S-----	
Cyanothece sp. PCC 8802	257059331	--L----- AK-----S-----	
Synechococcus sp. PCC 7335	254421960	--L---A----S DK-----S-I-----G-	
Synechococcus elongatus PCC 7942	81300567	--E---A---S GD-----S-L--A---	
cyanobacterium UCYN-A	284929286	--L----- SK-----S-----K-	
Synechococcus sp. PCC 7002	170077338	--L----- EK-----S-----A---	
Synechococcus elongatus PCC 6301	56752343	--E---A---S GD-----S-L--A---	
Thermosynechococcus elongatus BP-1	22299968	--ET----- SD-----S-----	
Synechococcus sp. JA-3-3Ab	86607188	--EL----- AR-----S-----	
Crocospaera watsonii WH 0003	357265606	--L----- AK-----S-----N-	
Crocospaera watsonii WH 8501	67921029	--L----- AK-----S-----N-	
Cyanothece sp. ATCC 51142	172037640	--L----- SK-----S-----N-	
Synechocystis sp. PCC 6803	16331844	--L----- SK-----S-----A-D-	
Cyanothece sp. CCY0110	126659786	-EL----- SK-----S-----N-	
Cyanobium sp. PCC 7001	254431713	--LS----- PG-----S-----	
Synechococcus sp. CB0101	318042314	--LS----- EG-----S-----	
Synechococcus sp. CC9902	78184186	--ET----- DQ-----I-S-----	
Synechococcus sp. BL107	116071168	--ET----- DQ-----I-S-----	
Synechococcus sp. WH 8102	33865151	--ET----- DQ-----S-----	
Synechococcus sp. WH 8109	260436262	--ET----- DQ-----S-----	
Synechococcus sp. RS9916	116073729	-AET----- DQ-----S-----	
Synechococcus sp. WH 5701	87301290	--LSL----- RD-----S-----	
Synechococcus sp. CB0205	317968945	-QLS----- EG-----S-----	
Synechococcus sp. RS9917	87125371	--EL----- DQ-----S-----	
Synechococcus sp. CC9605	78213579	--ET----- DQ-----S-----	
Synechococcus sp. RCC307	148243053	--LSL----- RD-----S-----	
Synechococcus sp. WH 8016	352095934	--ETL----- DQ-----S-----	
Synechococcus sp. WH 7805	88807336	--ETL----- DQ-----S-----	
Synechococcus sp. CC9311	113955511	--ETL----- DQ-----S-----	
Synechococcus sp. WH 7803	148240394	--ETL----- DQ-----S-----	
Prochlorococcus marinus MIT 9303	124022151	--ET--S---- QQ-----L-S-----	
Prochlorococcus marinus CCMP1375	33241085	--E----- GN---I---S-I---G-	
Prochlorococcus marinus AS9601	123969217	-NES--C---- EK---A-L-S-----D-	
Gloeobacter violaceus PCC 7421	37521648	--ETV----- AE-----S-----	

Supplementary Figure 137

Partial sequence alignment for the protein 23S rRNA (adenine(2503)-C(2))-methyltransferase, showing an 8 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		186685474	58		110
Nostocales/	Nostoc punctiforme PCC 73102	186685474	AKGNQGDNIYIGIDPILPPNRQAE	KDKSGETG	KDKSGETGSEARALKFDQTVP
	Anabaena variabilis ATCC 29413	75908477	-----		-----
	Nodularia spumigena CCY9414	119512876	-----		-E-A-----
	Nostoc sp. PCC 7120	17232504	-----LE-----		-----D-----
Stigonematales	Cylindrospermopsis raciborskii CS-505	282899366	-----E-----L-		R-----A-----
	'Nostoc azollae' 0708	298491438	-----		-----AL-----
	Raphidiopsis brookii D9	282895590	-----E-----L-		R-----A-----
Other	Fischerella sp. JSC-11	354566529	-----N-----		-E-A-S-----
	Microcoleus chthonoplastes PCC 7420	254412359	-----LQ-----		-EQA-S---E-----
	Trichodesmium erythraeum IMS101	113474598	-----V--LS-----		--QA-DS---E-G-----

Supplementary Figure 138

Partial sequence alignment for the protein histidyl-tRNA synthetase, showing an 8 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		190	250
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186682667	GTEFAELRNYRTGDDLRFIDWKATARR VG AYGNATPLVRVLEPEQEQTLLILLDRGLMTA
	Cylindrospermopsis raciborskii CS-505	114462346	-----CM-----L----- ----LS-----Q-----
Other Cyanobacteria	Raphidiopsis brookii D9	282897556	-----CM-----L----- ----VS-----Q-----
	Nostoc sp. PCC 7120	17232527	--D-----L----- VGV-----I-----
	Anabaena variabilis ATCC 29413	75908511	--D-----L----- VGV-----I-----
	Microcoleus chthonoplastes PCC	254411379	-----E-GV---T-L----- SR--I-----R---I-----
	Moorea producta 3L	332712244	-----D-GI---T-L----- SR-----K---I-----
	Synechocystis sp. PCC 6803	16329852	---R---D-QA-----YL---S--- GI--M-----R---VI-----
	Crocospaera watsonii WH 0003	357264014	---S---E-VS-E-T-L-N---S---G TH-IIK---E-R---I-----
	Crocospaera watsonii WH 8501	67920427	---S---E-VS-E-T-L-N---S---G TH-IIK---E-R---I-----
	Acaryochloris sp. CCME 5410	359457277	-----QE-NR---L----- GH----C---RD-P-----
	Acaryochloris marina MBIC11017	158335514	-----QE-NR---L----- GH----C---RD-P-----
Thermosynechococcus elongatus	22299625	-----E-HL-----LM---S--- GR---AM---RD-P-I-----	
Cyanothece sp. PCC 7425	220907133	---S---D-NL-----LM----- GR---LM---RD-P-I-----	

Supplementary Figure 139

Partial sequence alignment for the hypothetical protein Npun_R2352, showing a 2 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

			74		116
	Nostoc punctiforme PCC 73102	186684380	IPNPVMAGQQRVRDNLILYSSNSD SRAS	RFENNSAVRGNLVSNE	
Nostocales/	Nostoc sp. PCC 7120	17227964	----IA--EQ--KDI-FFN-RP-	K----Q-I-S-S-T--	
Stigonematales	Anabaena variabilis ATCC 29413	75909094	----IA--EQ--KDI-FFN-IP-	K--S-Q-I-S-S-T--	
	Nodularia spumigena CCY9414	119511822	-----E--KSL-QD- FFNPTT-	K----PV-QS-A-T--	
Other	Fischerella sp. JSC-11	354565683	----NA--EII--RT-W-YTSA-	K-D---T--A-----	
Cyanobacteria	Cyanothece sp. PCC 7822	307592047	L---IL--E--ANEIYFD-LSDK	-----EL-AQ-TE--	
	Cyanothece sp. PCC 7424	218440119	L---L--E--GNDILF--LSDK	----PV--AQ-IQ-	

Supplementary Figure 140

Partial sequence alignment for the hypothetical protein Npun_F4258, showing a 4 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		186684740	50	NDSPLGVYA	95
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186684740	LGIFKPKKTNESEFEAMS	NDSPLGVYA	KSKISLCLVSNLEDSIAH
	Nostoc sp. PCC 7120	17228241	-----Q-H----QNST		---V-----AA---
	Anabaena variabilis ATCC 29413	75910844	---Q-Q-H---HNST		---V-----AA---
	Nodularia spumigena CCY9414	119509303	---Q-Q-NHTL--ANSA		N-Q-----QA---
	'Nostoc azollae' 0708	298490203	-----QNY---AVNT		--SL-----D--SA---
	Cylindrospermopsis raciborskii CS-505	282901744	---R-Q-DH---VVHG		-NPL---I-----NA-SD
	Raphidiopsis brookii D9	282897400	---R-Q-DH---VVHG		-NPL---I-----NA-SD
	Fischerella sp. JSC-11	354565741	-----P-HYL---YST		-----R-----A---
	Cyanothece sp. PCC 7424	218441326	---S---SHQL--SDSA		--GM---F--E--DVA--L
	Cyanothece sp. PCC 7822	307151389	---S---SHQ---YHSA		-TRM---F--ED--AA--N-
	Cyanothece sp. CCY0110	126659877	---Q--GSH-T--NNSA		R-GL-----E--KA--L
	Cyanothece sp. PCC 8801	218248946	---C--SDHQ---GNSQ		G-G-----TD--AA--I
	Other Cyanobacteria	Microcystis aeruginosa NIES-843	166363949	--L-C--ESHDR--SDSR	
Cyanothece sp. PCC 8802		257062032	---C--SDHQ---GNSQ		G-S-----TD--AA--I
Synechococcus sp. PCC 7002		170077402	-----SPQH---TGT-		GAL-F---DD--EA--T
Microcystis aeruginosa PCC 7806		159029295	--L-C--ESHRE--SDSR		G-G-M-I-F--ES--KA-EY
Synechococcus sp. PCC 7335		254423340	-A----QAASS---A-S-		SGAA-I-----AD-AAA--R
Lyngbya sp. PCC 8106		119486058	---Q-NS-QVE--ADQA		-TGM-----A--DGA--E
Microcoleus vaginatus FGP-2		334118178	-----DSHK---ANRA		-TGM-----D--TA--R
Arthrospira sp. PCC 8005		376003438	---E-RSGDRQ--SSPK		GAM---Q--D--KA-EN
Arthrospira maxima CS-328		209526810	---E-RSGDRQ--SSPK		GAM---Q--D--KA-EN
Arthrospira platensis NIES-39		291566925	---E-RGGDRQ--SSPG		GAM---Q-CY--KA-EN
Arthrospira platensis Paraca		284050669	---E-RGGDRQ--SPQ-		GAM---Q-CY--KA-EN

Supplementary Figure 141

Partial sequence alignment for the protein glyoxalase/bleomycin resistance protein/dioxygenase, showing a 9 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		117		191
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186684748	VNGLLRQYLKAGE QGGRGAGGQRSR GAEEKDSSLVPPVQPPVSTQAPVLFNFDPLQLPENPVERLGLHSFPDW	
	Nodularia spumigena CCY9414	119512103	-----I-L-QK	LPE--K-----
	Anabaena variabilis ATCC 29413	75910625	-----I-L-EA	SP-F-E-----
	'Nostoc azollae' 0708	298492855	-----LIES	SSE--K-----
	Nostoc sp. PCC 7120	17228315	-----I-L-EA	SP-F-E-----
	Cylindrospermopsis raciborskii CS-505	282901368	---F---I-LLEK	TS---I---L-----Y---N-
	Raphidiopsis brookii D9	282895327	---F---I-LLEK	TS---I---L-----Y---N-
	Fischerella sp. JSC-11	354565440	-----F-VME-	EVGRTGGGVTGNTDEIFASPSLSFLTSTF-----
	Moorea producta 3L	332710036	-----N-A-L-LAN	SP---E--AGDM-----
	Microcoleus vaginatus FGP-2	334118913	-----I-LTDN	D--PLTVTNYKST-----E-
	Lyngbya sp. PCC 8106	119487592	---I--E---QKEK	QTQNRPIRLLED--LIQT--N-Y----
	Microcoleus chthonoplastes PCC 740	254415201	---V---A-LNTN	PQLD-----PQ-IQ-----Y----
	Arthrospira platensis Paraca	284050071	-----LVEA	GNSLESI-----S-----
	Arthrospira maxima CS-328	209524134	-----LLEA	GNSLESI-----S-----S
	Trichodesmium erythraeum IMS101	113474586	-----E-I-QNLT	IN-----QK--SY--N-
Cyanothece sp. PCC 7425	220906120	-----I-QSAG	G---S--TD--QQ-AL---Y---	
Acaryochloris sp. CCME 5410	359457510	-----T-A-AN	LN--E--PD-IQ-----Y-Y---	
Acaryochloris marina MBIC11017	158338705	-----A-A-AN	LN--E--PD-IQH---Y-Y---	
Microcystis aeruginosa PCC 7806	159030883	---M-----LTAN	GD---I---TAS-----	
Cyanothece sp. PCC 7424	218439986	---I---D-IKEQ	GK---KF-SD--NH--V---N-	
Cyanothece sp. PCC 8802	257058561	---I---I-L-Q-	-N--I--DD-IS---V-Y---F	
Cyanothece sp. PCC 8801	218245524	---I---I-L-Q-	-N--I--DD-IS---V-Y---F	
Cyanothece sp. ATCC 51142	172035456	---I-----SSET	DG---I--P--I-KIA-QY----	
Microcystis aeruginosa NIES-843	166367976	---M-----LTAN	GD---L---TAS-----	
Cyanothece sp. CCY0110	126656069	---I---I--SK-	SQ---I--S--I-KIA-QY--N-	
Cyanothece sp. PCC 7822	307150080	---I---D-IQAQ	GK---E--SDRIN--V---N-	
Crocospaera watsonii WH 0003	357261147	---I---I-STQG	-H--I--NDTLKK-A-QY----	
Synechococcus sp. JA-2-3B'a(2-13)	86609413	---I--A-Q-QVLD	QPEKDP-KPLEDT--AD-IAQ--R----	
Crocospaera watsonii WH 8501	67924848	---I---I-SIQG	-H--I--NDTLKK-A-QY----	
Synechococcus elongatus PCC 7942	81299771	--AI-----QSE-	R--P--SD-IAA-A-HQ----	
Synechococcus sp. JA-3-3Ab	86605784	---I--A-Q-Q-LD	GAEKDP-KPLQDT--TD-IAQ--R----	
Synechococcus elongatus PCC 6301	56750589	--AI-----QSE-	R--P--SD-IAA-A-HQ----	
Synechococcus sp. PCC 7002	170077132	---C--G---QRDK	GH-----AAA--ALQY---E-	
Synechococcus sp. PCC 7335	254423387	--AM---A-E-MP	-ANVIA--K--ERTR-----Y---	
Synechocystis sp. PCC 6803	16329552	---M--R-Q-AEEQ	GKNILDQEKIS--EQY----	
cyanobacterium UCYN-A	284928897	--AI--E-I-VKCK	ENCK-I-SKDF-NTTA-KY--N-	
Thermosynechococcus elongatus BP-1	22297753	---V--R-S-CTTD	PLEAFIAHLPL-SQ--CRY---E	
Synechococcus sp. RS9916	116073315	-----GA--AQEA	GEQ-T--DD-AL--ALQQ-L---	
Synechococcus sp. CC9311	113954299	---V--SA--AKEA	GET-PV-DQSA--ALC--L-V-	
Synechococcus sp. CC9605	78212348	-VNGVLRRAALR-R-	AGET---NDLPAQ-ALA--L---	
Synechococcus sp. RS9917	87125560	-----AA--ARQA	G-T-AM-ADAA--SLEQ-L-L-	
Synechococcus sp. RCC307	148241673	---V--AAV-AVEA	GES-PV-SD-QK--ALE--L---	
Synechococcus sp. CC9902	78185142	-----GV--AHEA	GEM-E-SAD-AI--AQQT-L---	
Synechococcus sp. BL107	116072280	-----GV--AHES	GET-E-SAD-AIQ-AQQA-L---	
Synechococcus sp. WH 8109	260434429	-VNGVLRRAALR-R-	AGET--S-NDLAAQ-AQSQ-L---	
Synechococcus sp. WH 7803	148240114	-----AA--AREA	G-G-VQ-TD-ALA-AQNO-L-L-	
Synechococcus sp. WH 7805	88809208	-----AA--ARDA	G-G-VR-SE-EME-AQDQ-L-L-	
Synechococcus sp. CB0205	317967915	---M--AF--RQEA	GEA-P-P-GD--AA-A-RQ-L---	
Synechococcus sp. WH 8016	352096314	---V--SA--AKEA	GET-AV-HQ-A--ALS--L-V-	
Synechococcus sp. WH 5701	87303174	-----AC--RRQP	LAPGAAPWE--P--SD-AAS-ALRG-L-P-	

Supplementary Figure 142

Partial sequence alignment for the protein rRNA SAM-dependent methyltransferase, showing a 12 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		186682798	297	RQSPPKV	349
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186682798	PEAHRARAAAFVITASEGANLHLDKLR	RQSPPKV	SRPQDFDPATRDRFLAGAL
	Nostoc sp. PCC 7120	17228513	-----	-----	T-----
	Nodularia spumigena CCY9414	119512906	-----	-----ES-K	-----
	Anabaena variabilis ATCC 29413	75909879	-----	-----	T-S-----
	Fischerella sp. JSC-11	354566978	-----	-----C-----E--	Q-----
	Oscillatoria sp. PCC 6506	300864621	-----	-----C-----AN--	-----
	Trichodesmium erythraeum IMS101	113477617	---D-----	---YI---A-----N--	I-----
	Microcoleus chthonoplastes PCC 7420	254413517	-----	-----YM---T-----FAN--	-----
	Lyngbya sp. PCC 8106	119491326	-----	-----K---YL---C---S---AN--	-----
	Microcoleus vaginatus FGP-2	334118075	---SD-----	---I---C---G---N--	-----T-
Other Cyanobacteria	Moorea producta 3L	332710185	---K-----	---YI---N---S---N--	K--D-----V-----
	Crocospaera watsonii WH 0003	357260813	-----	-----YI---C---S-F--EN--	-----F
	Crocospaera watsonii WH 8501	67925167	-----	-----YI---C---S-F--EN--	-----F
	Cyanothece sp. CCY0110	126657391	-----	-----YI---C---S-F--QN--	-----S-----F
	Synechococcus sp. JA-2-3B'a(2-13)	86610086	---Q-----	---L---A-----PN--	T--H-----
	Synechococcus sp. JA-3-3Ab	86606740	---Q-----	---YL---A-----PN--	T--H-----
	Cyanothece sp. ATCC 51142	172038459	-----	-----YI---C---S-F-IKN--	L-----T-----F
	Cyanothece sp. PCC 7822	307151358	--TA-----	--Y-----S---AN-K	T-----T-----
	Cyanothece sp. PCC 7424	218439121	-Q-D-----	-YI-----S---AN-K	T-L-----
	Microcystis aeruginosa NIES-843	166365160	--TV-----	--YI-----W--PR-Q	T-LE-----
Acaryochloris sp. CCMEE 5410	359460300	--SA-----	--YL-----SQ---PR-Q	T-L-----V---M	
Acaryochloris marina MBIC11017	158335190	--SA-S----	--YL-----SQ---QR-Q	T-L-----V---M	
Microcystis aeruginosa PCC 7806	159028871	--TA-----	--YI-S-A---W--PR-Q	T-LE-----	

Supplementary Figure 143

Partial sequence alignment for the protein Amidase, showing a 7 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		64	121	
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186685090	EHYVGYHLSFDLTEITPDLPSWLTNLQERMLIA V AESQTEELQPLKVLLEPTQQQIG	
	Nostoc sp. PCC 7120	17228517	-----T-----V- -VT--S--I-----	
	Anabaena variabilis ATCC 29413	75909884	-----T---N---A---V -LVA--P--QI-----	
	Nodularia spumigena CCY9414	119510012	-----S-----L-V DEVAD-P--I--A-----	
	'Nostoc azollae' 0708	298489859	-----HN--SD-KAQI AAIT-----T--Q-----	
	Raphidiopsis brookii D9	282897696	-N-----SRVL--ST--LE--Q-- -LIL--S-----	
	Cylindrospermopsis raciborskii CS-505	282899104	-N-----SRVS--SA--LE--QK- -LVL--S-----	
	Fischerella sp. JSC-11	354567285	-----V-----Y-K--T-SFAL- -QE-P-Y-----I-----	
	Microcoleus chthonoplastes PCC 7420	254417594	-----TVS-----KS--Q-LAE- SSQNPAQC----I-----I--	
	Microcoleus vaginatus FGP-2	334116942	-----A-T-----ESMK--FEE- VKE-PNQF-----A--	
	Oscillatoria sp. PCC 6506	300866886	-----I-DTAT--S---KSV--FES- -QA-PK-F-----M--	
	Moorea producta 3L	332706160	-----V----DSAT-----EC-K-SFKE- S-QNP-QF-----M--	
	Lyngbya sp. PCC 8106	119493548	-----SDG----Q--DD-K-QFEK- TQD-SDLY----I-----M--	
	Trichodesmium erythraeum IMS101	113477886	-----NI-----NA-K-KFDS- IIE-PKMF----I-----M--	
	Arthrospira platensis Paraca	284053476	-----D-----T-IHD-SD- IPS-----P----L-S	
	Arthrospira maxima CS-328	209527072	-----D-----T-IHH-SD- IPS-----P----L-S	
	Other Cyanobacteria	Cyanothece sp. PCC 8801	218247862	-----GDRVE--T---AMKQKFSQ- SQANLTQFS-----M--
		Cyanothece sp. PCC 7424	218439299	-----F-----DL-Q-----NQ-SALYAD- -LHNPQQIK---I---V--M--
Cyanothece sp. PCC 7822		307152546	-----V----DLAC--N--NQ-TAKYAD- -RLNPQQIK---I--D-M--M--	
Synechocystis sp. PCC 6803		16329767	-----I----DQVES--E--N--SQSFSQ- YQTNP-QIE---I--N-Q-----	
Cyanothece sp. CCY0110		126654792	-----SDRIS--TT---A--IQLTEL SQQDPTQFC--TI---E--M--	
Synechococcus sp. PCC 7002		170078550	-----VT--AD--Q NAPE-IT--K-Q--M--	
Crocospaera watsonii WH 8501		67922723	-----SDRIS--AL---A-KTKLSEL SQKDPSQFS--TI---Q--M--	
Microcystis aeruginosa PCC 7806		159030109	Q-----NVSN--E--EK-TNNFQQ- HQEDANLY--NI--P-Q--M--	
Cyanothece sp. ATCC 51142		172036438	-----SDRIS--TT--GA--TKFTEC SQQDPTQFC--TI---E--M--	
Microcystis aeruginosa NIES-843		166365744	Q-----NVSN--E--EK-TNNFQQ- HQEDADLY--NI--P-Q--M--	
Cyanothece sp. PCC 7425		220906049	-----ELGAGQ-LS-W--SLG DRLAQLDL---I--S--L-I--	
Synechococcus sp. PCC 7335		254425025	---T-----DEIES--V--GS-RTQS QATGGR-S-----Q-----	
Synechococcus elongatus PCC 6301		56750123	---T-----AL--N-SQ--DEF-V-L -AQGEA-AP---T-R-----	
Thermosynechococcus elongatus BP-1		22299287	-----SDHPDP-ET--NQVG-TL KAQSL-FEL--K----V--	
Acaryochloris sp. CCME 5410		359461072	-----I-V-DDLSADQTLS-W-EQWR -RLKAVSIT-----Y-MP-T--	
Acaryochloris marina MBIC11017		158336897	-----I-V-DDLLADQTLS-W-EQWR -RLKAVSIT-----Y-MP-T--	
Synechococcus sp. JA-3-3Ab		86605278	P-----P-LEIGPEWGS-ADY-QH-EAQV -TL--P-QC----	
Gloeobacter violaceus PCC 7421		37523628	---T---T--EVEDVPALVERLQQAGVPV -RP-----	
Synechococcus sp. JA-2-3B'a(2-13)	86609797	P-----PA-EVGPEWGS-ANY-EH-RAQV -TL--P-QL----		

Supplementary Figure 144

Partial sequence alignment for the hypothetical protein Npun_R4999, showing a 1 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

			361	400
Nostocales/ Stigonematales	Nostoc punctiforme PCC 73102	186685094	LLRPGDRLFLVPIP VESWT	GRTSADLDSLNLAYRLCPQL
	Anabaena variabilis ATCC 29413	75909888	----N-Q-Y--V-	DNN--TEY-GK--CDI--E-
	Nostoc sp. PCC 7120	17228521	----N-Q-Y--V-	DNN--NPEV-SK--SDV--K-
	Nodularia spumigena CCY9414	119510017	---SH-Q-Y--V-	DHS--NPLE--KI-REI--E-
	'Nostoc azollae' 0708	298489854	--KT--K-Y--V-	DNN--N--H-SK--NETY-D-
	Cylindrospermopsis raciborskii CS-505	282899100	--KTK--Y--V-	-SDY-N--Y-KA--LET--D-
	Raphidiopsis brookii D9	282897701	--KTK--Y--V-	-SDY-N--Y-KT--LET--D-
	Fischerella sp. JSC-11	354567289	----N-Q-Y--V-	NHS--NPEE--K--RDI--E-
	Oscillatoria sp. PCC 6506	300863528	V-AT--Y--V-	DHS--EPEE--K--LSI--T-
	Moorea producta 3L	332711130	-----S--V-	DHS--E--Y--T--QQI--T-
	Lyngbya sp. PCC 8106	119485334	---S--Y--V-	DHS--EPEQ--V-SRSI--N-
	Trichodesmium erythraeum IMS101	113475294	-----Y--V-	DHS--NQ-STI-QDI-TK-
	Arthrospira maxima CS-328	209524203	--HS--H--V-	DHS--IPEQ--H--HQI-----
	Arthrospira platensis Paraca	284051470	T-HS--H--V-	DHS--MPNG--H--HEI-----
	Other Cyanobacteria	Microcoleus chthonoplastes PCC 7420	254411511	---G--VY--V-
Microcoleus vaginatus FGP-2		334120435	--KK--Y--V-	DHN-SSPAE--A-QII--E-
Cyanothece sp. PCC 7822		307152203	-----Y--V-	EQAT--PAV--A--KEV--H-
Microcystis aeruginosa NIES-843		166364340	---K--S-Y--V-	DHS--NPET--E--ASI--E-
Synechococcus sp. PCC 7002		170077029	---D--H-H--VG	-HS--PQ--M--QVI--A-
Microcystis aeruginosa PCC 7806		159029336	---K--S-Y--V-	DHS--NPET--E--ASI--E-
Cyanothece sp. PCC 7424		218437546	-----EQ-Y--VV	-HS--PAI--T--DKI--N-
Acaryochloris sp. CCMEE 5410		359462292	--Q--T-S--V-	-HS--EPQA--E--KTV--R-
Synechocystis sp. PCC 6803		16330491	-----L--V-	DHD-G--PT-EQ--WQIM-D-
Cyanothece sp. PCC 8802		257058119	-F--D-TVY--V-	DEKT-NPEK-SE--IQ--E-
Cyanothece sp. PCC 8801		218245096	-F--D-TVY--V-	DEKT-NPEK-SE--IQ--E-
Synechococcus sp. PCC 7335		254421746	-----H-TAV	-HL--EPEA-KAI-QKV--A-
Crocospaera watsonii WH 0003		357264149	---AK-S-Y--V-	DHST-KP-Y--I--KQI----
Crocospaera watsonii WH 8501		67922269	---AK-S-Y--V-	DHST-KP-Y--I--KQI----
Acaryochloris marina MBIC11017		158334901	---Q--T-S--V-	-HS--EPQA--E--KTV--G-
Thermosynechococcus elongatus BP-1		22297734	---A--W--V-	DHA--N-EH-KD--TC--R-
Cyanothece sp. ATCC 51472		354552672	--KEN-S-Y--V-	DHST-NPHD--E--RQI--H-
Cyanothece sp. ATCC 51142		172039043	--KEN-S-Y--V-	DHST-NPHD--E--RQI--H-
Cyanothece sp. CCY0110		126657288	---KD-S-Y--V-	DHSTENP-N-K--DQI--K-
Cyanothece sp. PCC 7425		220909719	--QA--S-Y----	-HQG-EPEK--S-VKGMNLP-
Synechococcus sp. JA-2-3B'a(2-13)	86608851	-----CF-AL--A	-HSGT-PEQ--Q--QT-Q-H-	
Synechococcus sp. JA-3-3Ab	86606025	-----F-AL-LF	-HAGTHPQE--H--QI-Q-H-	

Supplementary Figure 145

Partial sequence alignment for the protein bifunctional folylpolyglutamate synthase/dihydrofolate synthase, showing a 5 aa insert that is specific for *Nostoc punctiforme* PCC 73102.

		203	260
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17231701	LGIKVWVFKGEIIPGQEVATPPPSPRD RDRDRGD RDREPRRRQQRRRQQFEDRSNEG
	Fremyella diplosiphon Fd33	48374164	-----E-A-A-TS-E -E-----
	Cylindrospermopsis raciborskii CS-505	282899958	-----DIVSQ-VT-E -----A
	Raphidiopsis brookii D9	282898351	--V-----DIVSQ-VT-E -----A
	'Nostoc azollae' 0708	298491404	-----ETPAQ-TT-E -----A
	Fischerella sp. JSC-11	354565800	-----E-PA--TT-E REP--R-----
	Microcoleus vaginatus FGP-2	334120603	-----QEAA-NAAVP -TK-K---N-----
	Oscillatoria sp. PCC 6506	300865408	--V-----DIPA-NAGPP -TKNKG--HK-----
	Lyngbya sp. PCC 8106	119484827	-----ETSNSANVGQ P-RK-P--KYD-----
	Arthrospira sp. PCC 8005	376007821	--V-----ETPA-NTRAP KSR RTP--KYD--SD
	Arthrospira platensis Paraca	284053309	--V-----ETPA-NTRAP KSR RTP--KYD--SD
	Microcoleus chthonoplastes PCC 7420	254412256	--V--I-----ST-DVA---TAQP -R-----
	Moorea producta 3L	332712163	-----V-----EKPT-NTAQP -R-PK-----D
	Crocospaera watsonii WH 0003	357262589	-----I---V---QEAMAAPAPT PRKK--P-----
	Crocospaera watsonii WH 8501	67923589	-----I---V---QEAMAAPAPT PRKK--P-----
	Cyanothece sp. ATCC 51142	172038933	--V-----DEQAMAAPAPT PRKK--P-QFEDRSNE
	Cyanothece sp. CCY0110	126657994	--V-----DEQAMAAPAPT PRKK--P-----
	Cyanothece sp. PCC 7424	218440629	--V-I-----EQA-AQPATT PKR-----
	Synechocystis sp. PCC 6803	16329935	-----I---V---A-IVA-PSQP RRKS----D---QD-
	cyanobacterium UCYN-A	284929567	--V-----EQ-TTSA-PT PKRN--K-----
	Thermosynechococcus elongatus BP-1	22297631	--V--I---VL---TE-V-REAT-R SP---LP---N---
	Cyanothece sp. PCC 7425	220906725	-----V-----EPPSNREAQP -Q-H--KK-----
	Synechococcus sp. PCC 7002	170077679	-----I-----ETSANAAPL PRRKS-----EQ
	Synechococcus sp. CB0205	317968823	-----VL--TQDKV-VGAAPK -RAS--P-----
	Synechococcus sp. CB0101	318042558	-----VL--NKDKV-VGAAPK -RAS--P-----
	Synechococcus sp. BL107	116072558	-----VLSE-SQTM-VGANPR -RAS--P-----
	Cyanobium sp. PCC 7001	254431944	-----VL--KEQL-VGAAPR -RTS-QP-----Q
Synechococcus sp. CC9902	78185528	-----VLNE-SQTM-VGANPR -RAS--P-----	
Synechococcus sp. WH 7805	88807890	-----VLGDEAQQQL-VGATP --RAG--P-----	
Synechococcus sp. RS9917	87124968	-----VLGDEARQQL-VGATP --RAG--P-----	
Synechococcus sp. WH 7803	148238764	-----VLGDEAQQQL-VGATP --RAG--P-----	
Synechococcus sp. RS9916	116075761	-----VLGDEARQQL-VGATP --RAG--P-----	
Synechococcus sp. WH 8016	352096515	-----VLSDDSQQQI-VGANP --RAG--P-----	
Synechococcus sp. CC9311	113955322	-----VLSDDSQQQI-VGANP --RAG--P-----	
Synechococcus sp. WH 5701	87300907	-----VL--QEKL-VGAAPR -RAN--P-----	
Synechococcus elongatus PCC 6301	56751880	-----R-----D-A-SNVGQP --N-----	
Microcystis sp. T1-4	390438088	--V--I-----EIAAAVPAQA P-R-----S-	
Synechococcus sp. RCC307	148243221	-----VL--KQQP-VGAAPR -RAS--P-E-----	
Microcystis aeruginosa NIES-843	166368478	--V--I-----EIAAAVPAQA P-R-----S-	
Synechococcus sp. WH 8109	260435557	-----VLSE-AQPM-VGAAPR -RAS--P-----	
Synechococcus sp. CC9605	78211922	-----VLSE-AQPM-VGAAPR -RAS--P-----	
Synechococcus sp. PCC 7335	254425252	--V-T-----V-----SQPEESRAQP --R-P-----DS	
Synechococcus sp. WH 8102	33866605	-----VLGDEAPLI-VGASPR -RAS--P-----	
Prochlorococcus marinus NATL2A	72382961	-----VL-KE-QPL-VG-SPR -SRGN--P-----	
Prochlorococcus marinus AS9601	123969291	-----VL-KE-QTI-VGASPK -KAS--P-----	
Prochlorococcus marinus MIT 9313	33864005	-----I---VLGDEAQT-M-VGASPR -RGN--P-----	
Prochlorococcus marinus CCMP1986	33862108	-----VL-KE-QTI-VGAI PR -KGS-KP-----	

Supplementary Figure 146

Partial sequence alignment for the protein 30S ribosomal protein S3, showing a 7 aa insert that is specific for *Nostoc sp. PCC 7120*.

		147		197	
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17231875	LHALKAVSSRPLVAHQYPIM	QREVGAASADPN	FGGEQLTGFIDMVSEQAY
	Anabaena variabilis ATCC 29413	75909501	-----		A-----
	Nostoc punctiforme PCC 73102	186685147	-----		N-----
	'Nostoc azollae' 0708	298489816	-E-----		Q-----
	Fischerella sp. JSC-11	354567150	-----		----I---L-T----
	Cyanothece sp. PCC 7425	220907054	-Q---T-----P-----G		Q--E-L---L-T----
	Crocospaera watsonii WH 0003	357260166	----D-T--VIPQ---R		QQRE-I-Y--L-----
	Cyanothece sp. ATCC 51142	172038017	----D-T--V-PQ---R		QDRE-I-Y--LIT----
	Cyanothece sp. PCC 8801	218246184	-Y---E-T--VIPQ---R		QQDE-I-Y--L-T----
	Cyanothece sp. PCC 8802	257059232	-Y---E-T--VIPQ---R		QQDE-I-Y--L-T----
	Cyanothece sp. CCY0110	126658056	----D-T--V-PQ---R		QDRE-I-Y--LIT----
	Acaryochloris marina MBIC11017	158333793	-D---Q-----L-----G		QN-D-Q---L-----F
	Cyanothece sp. PCC 7822	307154648	----E-T---PQ---R		SNHD-I-Y--LI-----
	Cyanothece sp. PCC 7424	218440682	----Q-----LPQ---R		QAHD-I-Y--LI-----
	Other Cyanobacteria	Microcystis sp. T1-4	390438493	-Q---S-----PQ---R	
Synechococcus sp. PCC 7335		254425615	-D---GL-----P-----G		Y--S-V---L-----
Synechocystis sp. PCC 6803		16331722	-Q---S-----IPQ---Y		K--E-Q-Y--LIT----
Acaryochloris sp. CCME 5410		359458629	-D---Q-----P-L-----G		QN-D-Q---L-----F
Microcystis aeruginosa NIES-843		166368342	-Q---S-----PQ---R		QNNETI----LIN----
Synechococcus sp. PCC 7002		170078682	-A--QEA-N---PQ---R		KDQQI-L-Y--L-----
Thermosynechococcus elongatus BP-1		22299902	-A-YRQA-----P----W		Q-DD-L-Y--L-T----
Microcoleus vaginatus FGP-2		334121329	ID---S-----I-P-----G		KN--II---L-T----
Moorea producta 3L		332706738	--NS--T---L-----G		K--H-----L-T----
Oscillatoria sp. PCC 6506		300866473	-N---S-----VIP-----A		K--AIV---L-T----
Microcoleus chthonoplastes PCC 7420		254411840	-D---G-----I-P-----G		KA---I---L-T---F
Trichodesmium erythraeum IMS101		113476366	-E---T--N--I-P-----A		KNN--I---L-T----
Arthrospira sp. PCC 8005		376007842	-N---T-----I-PQE---A		K-SE-I---L-T----
Arthrospira maxima CS-328		209526753	-N---T-----I-PQE---A		K-SE-I---L-T----
Arthrospira platensis Paraca		284052630	-N---T-----I-PQE---A		K-SE-I---L-T----
Lyngbya sp. PCC 8106		119486185	-A---T-----V-P-----A		Q-QE-V---L-T----
Gloeobacter violaceus PCC 7421		37522381	-A--QSI-----ALR-F--G		Q-DR-S---L-----W

Supplementary Figure 147

Partial sequence alignment for the protein elongation factor G, showing a 13 aa insert that is specific for *Nostoc sp. PCC 7120*.

		270	327
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17228534	SQYAALGALQAKPEYLQDHIGA LAQPAVG IAQVRQIVFDYLKQLQGLCNITPADGAFY
	Anabaena variabilis ATCC 29413	75909900	-----K-----N-----T-----
	Nodularia spumigena CCY9414	119510165	-----KN---
	Nostoc punctiforme PCC 73102	186681093	-----E---SN---
	Cylindrospermopsis raciborskii CS-505	282899590	-----H-NG-F-KTNLQS
	Fischerella sp. JSC-11	354566095	-----RNN-A-
	Moorea producta 3L	332709618	-----NTRKD-C-P--N-
	Oscillatoria sp. PCC 6506	300863973	-----GRN-WENQ---
	Microcoleus chthonoplastes PCC 7420	254413374	---V--A---MGMS-T-P--Q-
	Microcoleus vaginatus FGP-2	334120122	-----VGRG-CDNYVR-
	Arthrospira maxima CS-328	209524645	---V----KVGY--C-N--KT
	Arthrospira sp. PCC 8005	376002655	---V----KVGY--C-N--KT
	Arthrospira platensis Paraca	284050147	---V----KVGY--C-N--KT
	Arthrospira platensis NIES-39	291568533	---V----KVGY--C-N--KT
	Other Cyanobacteria	Lyngbya sp. PCC 8106	119483394
Trichodesmium erythraeum IMS101		113474185	A--G-IA--ESSQDCVEQMRQ-
Cyanothece sp. PCC 7424		218437441	-----KTGKS-C-E-LPN
Cyanothece sp. CCY0110		126660477	-----TVGKN-YKKY--E
Cyanothece sp. PCC 7822		307154621	-----KVGVT-CKK-LQD
Cyanothece sp. ATCC 51142		172036036	-----KVGKS-CKK-LEE
Microcystis aeruginosa PCC 7806		159027561	-----VGKS-PLEQLKT
Cyanothece sp. PCC 8801		218246048	----V---KQGKN-C-QYLPE
Cyanothece sp. PCC 8802		257059097	----V---KQGKN-C-QYLPE
Microcystis aeruginosa NIES-843		166365788	-----VGKS-PLEQLKT
Cyanothece sp. PCC 7425		220906998	----V----GVG-CREKLA-
Acaryochloris sp. CCMEE 5410		359460353	----V----VGRG-CHQ-LAT
Synechocystis sp. PCC 6803		16331997	-----AC-RVGKN-SAQFLPE
Acaryochloris marina MBIC11017		158333381	----V----VGRG-CHQ-LAT

Supplementary Figure 148

Partial sequence alignment for the protein aspartate aminotransferase, showing a 7 aa insert that is specific for *Nostoc sp. PCC 7120*.

		160	194
Nostocales/ Stigonematales	Anabaena variabilis ATCC 29413	75908940	QFAVIGSIYSQYVLGTGEP P KVGLLNIGEEDTKGN
	Nostoc sp. PCC 7120	17227734	-----
	'Nostoc azollae' 0708	298491028	---M-----MP-- -I-----
	Nostoc punctiforme PCC 73102	186680646	---M-A-----T-- -----S---
	Nodularia spumigena CCY9414	119512196	---M-T-----VPQ- D-----
	Cylindrospermopsis raciborskii CS-505	282899055	---M-----IS-- -----EN---
	Raphidiopsis brookii D9	282898513	---M-----IS-- -----EN---
	Fischerella sp. JSC-11	354567125	---M-V-----VP-- -----C---
	Microcoleus chthonoplastes PCC 7420	254413061	---M-TV-----IP-- -----C---
	Oscillatoria sp. PCC 6506	300867962	---M-T-----VT-- -----PS---
	Microcoleus vaginatus FGP-2	334119327	---LM-T---C---VA-- -I-----PS---
	Lyngbya sp. PCC 8106	119488099	---M-----VDQ- RI-----AS---
	Moorea producta 3L	332707284	---IM-V--K---ME-- -----EC---
	Arthrospira platensis Paraca	284053529	---M-----VDN- T-----SS--D
	Arthrospira maxima CS-328	209526113	---M-----VDN- T-----SS--D
	Arthrospira sp. PCC 8005	376002726	---M-----VDN- T-----SS--D
	Trichodesmium erythraeum IMS101	113475485	---M-T---ES--A-D-- -----PS---
	Thermosynechococcus elongatus BP-1	161485765	---M-----DVEN- -----C---
	Cyanothece sp. PCC 8801	218246637	---LM-T---E---VEQ- -----S---
	Cyanothece sp. PCC 7425	220907908	---LM-T---E---VEK- Q-----EC---
	Acaryochloris marina MBIC11017	158333842	---TM-T---E---EA- -----SC---
	Acaryochloris sp. CCMEE 5410	359457938	---TM-T---E---EA- -----SC---
	Cyanothece sp. ATCC 51142	172036500	---LM-T---M-VEQ- -----SS---
	Synechococcus elongatus PCC 7942	81300263	---M-----RD--QAN- R---V-----S---
	Synechocystis sp. PCC 6803	16330609	---LM-TV-----VDS- -----SN---
Synechococcus elongatus PCC 6301	56685071	---M-----RD--QAN- R---V-----S---	
Cyanothece sp. CCY0110	126660635	---LM-T---M-VN-- Q-----SS---	
Synechococcus sp. PCC 7002	170077151	---LM-T---K---NK-- Q-----PS---	
Crocospaera watsonii WH 8501	67922028	---LM-T---M-VEQ- -----SS---	
Cyanothece sp. PCC 7424	218438035	---LM-T---K---IE- -----SS---	
Crocospaera watsonii WH 0003	357264548	---LM-T---K---M-INT- -----SS---	
Other Cyanobacteria	cyanobacterium UCYN-A	284928702	---LM-T---K---MAVED- -----SS---
Cyanothece sp. PCC 7822	307152596	---LM-T---K---M-VE-- Q-----AS---	
Microcystis aeruginosa NIES-843	166364689	---FM-----R---ND-- -----KS---	
Synechococcus sp. PCC 7335	254421564	---RM--L-C-C---IEK- R-----PN---	
Synechococcus sp. JA-3-3Ab	86606993	---RM--L-C-C---IEK- R-----PN---	
Synechococcus sp. JA-2-3B'a(2-13)	86609781	---LL-N---RD--QVEQ- RI-----C---	
Synechococcus sp. WH 8102	33866777	---LL-N---RD--QVDR- RI-----EC---	
Synechococcus sp. WH 8016	352095482	---LL-N---RD--QVEQ- RI-----EC---	
Synechococcus sp. RCC307	148241348	---LL-N---RD--QVDR- RI-----EC---	
Synechococcus sp. RS9917	87125176	---LL-N---RD--QVDR- RI-----EC---	
Synechococcus sp. CC9311	113953618	---LL-N---RD--QVDR- RI-----EC---	
Synechococcus sp. CC9605	78213893	---LL-N---RD--QVDR- RI-----EC---	
Synechococcus sp. WH 8109	260434773	---LL-N---RD--QVDR- RI-----EC---	
Synechococcus sp. BL107	116071498	---LL-N---RD--QVEK- R-----EC---	
Synechococcus sp. CC9902	78183889	---LL-N---RD--QVEQ- R-----EC---	
Cyanobium sp. PCC 7001	254432863	---LL-N---RD--QVAR- RI--V-----C---	
Synechococcus sp. RS9916	116073934	---LL-N---RD--QVAQ- R-----EC---	
Synechococcus sp. WH 7803	148240585	---LL-N---RD--QVAQ- RI-----EC---	
Synechococcus sp. WH 7805	88807097	---LL-N---RD--QVAR- RI-----EC---	
Synechococcus sp. WH 5701	87301815	---LL-N---RD--VAR- S-----EC---	
Synechococcus sp. CB0101	318040601	-W-LL-N---RD--QVSK- RI-----EC---	
Synechococcus sp. CB0205	317968095	-W-LL-N---RD--QVQ-- RI-----EC---	
Prochlorococcus marinus CCMP1375	33239609	---LL-N---RD--QVQN- RI-----EC---	
Prochlorococcus marinus MIT 9313	33864260	---LL-N---RD--QVAR- RI-----EC---	
Prochlorococcus marinus AS9601	123967689	---LL-N---AKD--QVKK- RI-----EC---	
Prochlorococcus marinus NATL2A	72383337	---LL-N---RD--QVDK- RI-----SC---	

Supplementary Figure 149

Partial sequence alignment for the protein putative glycerol-3-phosphate acyltransferase, showing a 1 aa insert that is specific for *Anabaena variabilis* ATCC 29413.

		896	926
Nostocales/ Stigonematales	Anabaena variabilis ATCC 29413	75910649	LRLTLRSPTWPDPEADT G GIHEFAYTVYPH
	Nostoc sp. PCC 7120	17228343	-----T-----
	Nostoc punctiforme PCC 73102	186684435	-----N---S---K -F---T--L---
	Nodularia spumigena CCY9414	119508855	-----N-----T-R -F---T-AL-S-
	'Nostoc azollae' 0708	298490961	-----SN---SQ--R ------T-AL---
	Raphidiopsis brookii D9	282898181	M-----H---Q--R -E-Q-N-AL---
	Cylindrospermopsis raciborskii CS-505	282898622	M-----H---Q--R -E-Q-N-AL---
	Fischerella sp. JSC-11	354568538	-----N---K--R -V---T-AL---
	Cyanothece sp. PCC 7424	218441141	I-----I---N--R ------T-AL---
	Cyanothece sp. PCC 8801	218247240	-----R-----C-R -L-H-T-AI---
	Crocospaera watsonii WH 0003	357264300	-----K--R---TC-I -KQ--T-AI---
	Cyanothece sp. PCC 8802	257061425	-----L---KC-R -L-H-T-AI---
	Cyanobacterium UCYN-A	284929264	-----K--Q---NC-K ---Q-S-AI---
	Cyanothece sp. CCY0110	126656675	-----K--R---TC-I -Y---T-AI---
	Other Cyanobacteria	Cyanothece sp. ATCC 51142	172036118
Crocospaera watsonii WH 8501		67922172	-----K--R---TC-I -KQ--T-AI---
Acaryochloris marina MBIC11017		158339464	-----N---GC-R -H-Q-T-AI---
Acaryochloris sp. CCMEE 5410		359459314	-----N---GC-R -H-Q-T-AI---
Cyanothece sp. PCC 7822		307153393	-S-S-----N---Q--R ET-H-S-AL---
Cyanothece sp. PCC 7425		220909035	-----A-----T--R -L-H-S-AL---
Microcystis aeruginosa PCC 780		159027198	-----R-----L -KQ--T-AI---
Synechococcus elongatus PCC 63		56752397	-----GS-----Q--L -Q-Q---AI---
Synechococcus elongatus PCC 79		81300512	-----GS-----Q--L -Q-Q---AI---
Microcystis aeruginosa NIES-84		166366395	-----R-----L -KQ--T-AI---
Synechocystis sp. PCC 6803		16331225	M-----SV---S--R -K-K-S-YF---
Thermosynechococcus elongatus BP-1		22297834	-----TS-R -CQ-C-S-G-F--
Synechococcus sp. PCC 7335		254423409	I-----K-SI---T--R -T-R-T-AI---
Synechococcus sp. PCC 7002		170077868	-----K--L---K--R -M-HL--AL---
Trichodesmium erythraeum IMS10		113476728	-----GS---SK--V ----T-AI---
Microcoleus chthonoplastes PCC 7420		254415502	-----K--M--H-D--R ---H-T-AL---
Oscillatoria sp. PCC 6506		300864405	-----GS---EA--R ---Q-T-GI---
Moorea producta 3L		332709091	-----Q ----Q-T-AL---
Lyngbya sp. PCC 8106		119492081	I-----GS-----K--L ---H-T-AI---
Microcoleus vaginatus FGP-2		334118721	I-----GS---EQ--V -VS--TCA----
Arthrospira sp. PCC 8005		376002828	IG-S--GSL---K--V -V---S-GI---
Arthrospira maxima CS-328		209524907	IG-S--GSL---K--V -V---S-GI---
Arthrospira platensis Paraca		284050203	IG-S--GSL---K--I -V-Q-S-GI---
Arthrospira platensis NIES-39		291570844	IG-S--GSL---K--I -V-Q-S-GI---

supplementary Figure 150

Partial sequence alignment for the protein glycoside hydrolase, showing a 1 aa insert that is specific for *Anabaena variabilis* ATCC 29413.